INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

AI AND MISINFORMATION: THE ROLE OF GENERATIVE MODELS IN SHAPING PUBLIC OPINION IN PAKISTAN

Younus Bashir [1], Saadia Bakhtawar [2], Jafar Ali [3]

DOI: https://doi.org/10.63544/ijss.v5i3.301

Affiliations:

[1] Lecturer, Department of Management Sciences, SZABIST University, Pakistan
Email: younusbashir@hotmail.com

[2] Senior Lecturer, Department of Management Sciences, SZABIST University, Pakistan
Email: saadia.bakhtawar@szabist.edu.pk

[3] School of Business, Yeungnam University, South Korea.
Email: jafarisali110@gmail.com

Corresponding Author's Email:

[1] younusbashir@hotmail.com

Copyright:

Author/s

License:

Article History:
Received: 29.04.2026
Accepted: 20.05.2026
Published: 31.05.2026

Abstract

The rapid proliferation of generative artificial intelligence (GAI) has introduced unprecedented challenges to information integrity worldwide, with developing nations like Pakistan facing heightened vulnerabilities. This paper explores the role of generative AI models in propagating misinformation and shaping public opinion within Pakistan's unique socio-political and digital ecosystem. Employing a qualitative, multi-method research design incorporating online content analysis and case study approaches, the study examines AI-generated misinformation across political, social, and geopolitical domains in Pakistan from 2023 to 2025. The findings reveal that deepfake technologies, large language models, and generative adversarial networks are increasingly being weaponized to create highly realistic, scalable false narratives that undermine electoral integrity, target vulnerable populations - particularly women politicians - and exacerbate geopolitical tensions, especially in India-Pakistan relations. The research identifies critical gaps in Pakistan's existing legal framework, including the Prevention of Electronic Crimes Act (PECA), which lacks specific provisions addressing deepfakes and AI-generated content. Furthermore, low digital literacy rates, rapid information dissemination through closed messaging platforms, and technological resource constraints significantly impede detection and mitigation efforts. Drawing on Information Overload Theory, the Social Amplification of Risk Framework, and Cultivation Theory, this paper demonstrates how GAI-driven misinformation erodes public trust, intensifies societal polarization, and threatens democratic processes. The study offers comprehensive policy recommendations encompassing legislative reforms, media literacy initiatives, and public-private partnerships to foster a more resilient information environment in Pakistan.

Keywords: AI, Misinformation, Public Opinion, Generative Artificial Intelligence, Digital Ecosystem, Civil Society, Media Outlets, Government Agencies, Pakistan

1. Introduction

1.1. Background: The Rise of AI and Misinformation

The concept of artificial intelligence (AI) has brought an unprecedented shift in the technical and technological sphere, altering many aspects of human life, such as medicine, finance, communication, government, and entertainment.

But there is a deeper and increasing problem, which comes with these developments: the amplification of misinformation and disinformation. False or inaccurate information rumoured without the purpose of

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

deceiving is known as misinformation, and any intentional fabrication of information and its distribution to mislead viewers is called disinformation (Wardle and Derakhshan, 2017).

In the world, the effects of the AI-created misinformation are already observed in the form of political manipulation, the polarization of social values, economic instability, and reduced confidence in the media and the governmental system (Knott et al., 2024).

It is against this backdrop that the current paper will focus on how generative AI is being employed to generate and spread fake news in the Pakistani online environment and how these messages are influencing people and making them distrust traditional media and institutions. It also explores the structural weaknesses that exist in the Pakistani digital ecosystem that allows AI-driven misinformation to spread, evaluates the sufficiency of the current policy and regulatory framework to handle them, especially in the context of deepfakes, and offers policy suggestions to key stakeholders.

This study holds importance in a number of ways. To begin with, it adds to the new literature body of research in the intersection of artificial intelligence, misinformation, and attitudes of the population in developing nations, where it provides a narrow analysis of Pakistan.

This research area will begin in 2023 and end in 2025, a time when there is a rapid pace of development and usage of generative AI technologies. Although international trends and comparison examples are also taken into account where possible, the major analytical emphasis is put on the socio-political and digital contexts of Pakistan. However, there are some weaknesses that should be taken into consideration.

1.2. Problem Statement: Generative Models and Public Opinion in Pakistan

The present situation in the digital sphere and the users of social networks in Pakistan, where the digital environment is rapidly expanding and the number of users on social networks is skyrocketing, makes it an especially sensitive area of the spread of AI-generated misinformation. The level of internet penetration and the adoption of social media in the country have been exploding in the past few years, and nowadays millions of citizens are its users to gain access to news, information, and socialize (Toor et al., 2024).

The socio-political context, usually permeated by tensions and polarization of speech, is the very ground where similar information can be manipulated with a certain purpose.

2. Literature Review

2.1 Understanding Misinformation and Disinformation

As the world develops into a global village, the concepts of misinformation and disinformation have become major concerns when discussing the integrity of information ecologies. Although the two terms are applied interchangeably, it is important to establish the exact definitions and differences in their meanings to have a full picture of their effects, particularly in generative AI.

2.1.1 Definitions and Typologies. Misinformation can be defined as false or inaccurate information that is not disseminated with an aim of deceiving. It may occur due to real accidents or errors, misunderstandings or misinterpretation of facts.

Disinformation: The phenomenon that is called disinformation, essentially, is proven false or misleading information generated and disseminated on purpose with the malicious aim of deceiving or controlling an audience.

In addition to these two main categories, researchers have presented more typologies in order to distinguish the subtleties of problematic information:

Satire or Parody: Satire and parody usually include misleading information but are not targeted to make people deceived by exaggerating the truth. But any other discourse out of context can be taken as real news.

False Connection: This is where there is no support of the headlines, visuals or caption to the content.

False context: There would be true content with false contextual information.

Imposter Content: On the off chance that authentic sources are deceived.

Manipulated Content: Real information or images are misused to mislead.

Fabricated content: New content is generated to create falsification

Generative AI is most likely to be predominant in generating and sharing routinely altered and fake content and thus presenting a far more scalable and convenient possibility of generating disinformation that is

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

highly convincing.

2.1.2 Historical Context of Misinformation Spread. False information is not a novel phenomenon; it has existed since the beginning of the human history, and it has changed and developed along with the changes in communication technologies. Whether in ancient times when word of mouth brought about rumours, or in wartime when propaganda leaflets were distributed, or yellow journalism in the 19th century, the ability of human beings both to get things wrong and to lie intentionally has been there.

In the 20th century, mass media, namely radio, television, and print, came to dominance, which led to the concentration of information distribution and, to a certain degree, the possibility to exert more control over the factual correctness of information due to the editorial work. These mediums were, however, also used as effective mediums of propaganda in the times of war and political instability.

The introduction of the internet and, even more importantly, the introduction of social media platforms in the 21st century was a paradigm shift.

Generative AI is the most recent (and possibly the most powerful) iteration in the ability to generate and spread disinformation in this historical arc. It transcends human error or manual creation to allow building of large volumes of automated and highly realistic content and thus presents a whole new order of challenge to information integrity.

2.2 Generative Artificial Intelligence (GAI)

Generative Artificial Intelligence (GAI) is a type of AI systems which creates new, original content, and do not just analyse or categorize already existing data. Recent scholarship has also shown that artificial intelligence is increasingly being used in strategic decision-making, communication systems, and institutional planning, but its rapid development also creates barriers related to ethics, reliability, governance, and responsible application (Iram et al., 2025).

2.2.1 Overview of GAI Technologies (LLMs, Deepfakes, etc.). Among the GAI technologies leading this revolution, there are several that are important as well:

Large Language Models (LLMs): OpenAI GPT series, Google Gemini, and others are large language models that have been trained on large text datasets in an attempt to learn and generate human-like language.

Deepfakes: Deepfakes are synthetic media, a portmanteau of deep learning and fake, in which someone in an existing image or video is substituted with someone else.

Generative Adversarial Networks (GANs): GANs involve two competing neural networks trained to defeat one another, where one network is trained to produce a lot of data, called a generator, and another one is trained to recognize what it is, called a discriminator. The generator synthesizes new data (e.g.

Diffusion Models: This is a more recent form of generative models that have demonstrated incredible success in generating high quality images and other data.

2.2.2 Capabilities of GAI in Content Generation. The GAI models can do extremely impressive stuff with content generation:

Scalability: GAI has the capacity to generate tremendous amounts of content at a cadence and level that could not possibly be created by human actors. This will allow information channels to be flooded with artificial narratives in extreme quantity, making it very difficult for factchecking to keep up, leaving users with little chance to figure out legitimate sources.

Realism: The content produced is sometimes of high realism to the extent that it is very difficult for an ordinary human being to distinguish it as fake. This realism applies to language, intuitive details, as well as to emotional displays, strengthening the force of falsity. The growing use of AI across institutional and educational contexts shows that AI tools are not limited to technical automation; rather, they are becoming part of knowledge production, communication, and decision-support systems, which makes their misuse in public information spaces more serious (Malik & Rafiq-uz-Zaman, 2025).

2.3. Impact of AI on Information Ecosystems Globally

The widespread adoption of AI and generative models, especially, has recalibrated the global information ecosystem immensely.

2.3.1. Case Studies from Other Regions. Electoral Interference (United States and Europe): Generative AI is suspected to be involved in election manipulation, including the use of AI-generated robocalls

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

during the 2024 U.S. presidential election.

Geopolitical Disputes (Ukraine-Russia War): The war between Ukraine and Russia has seen the confrontation between AI-powered information war and has become a battleground of the same. It is also claimed that both parties have used AI to produce and share propaganda, deepfakes of leaders, and doctored footage of battles to affect both the global and local perception.

This is indicative of the ability of GAI to increase hostile interplay and obfuscate facts in geopolitical situations of significant stakes.

Public Health Crises (COVID-19 Pandemic): Artificial intelligence-generated content was part of the infodemic, where public health issues during the COVID-19 pandemic spread misleading medical guidance, conspiracy theories, and anti-vaccine messages. Although not necessarily a deliberate malicious act, the amount of content, much of it convincing, that AI produced fed into the spread of inaccurate information, amounting to a challenge that public health authorities found hard to curb.

Financial Scams and Fraud: There are more convincing financial scams based on GAI to build convincing phishing mails, complex deepfake voice calls, and convincing fake webpages. Such frauds that have involved the use of AI are more difficult to detect, costing individuals and businesses large sums of money all around the world.

Such international case studies highlight a shared theme: GAI increases the scale, pace, and realism of misinformation, and thus conventional mechanisms of factchecking and content moderation become less effective. Capabilities to create content that is passed through human and automatized detection systems are increasingly worrying.

2.3.2. Challenges in Detection and Mitigation. The rapid evolution of GAI presents significant challenges for detection and mitigation efforts:

Refinement of Synthetic Content: The more advanced the models of GAI become, the harder it is to detect an artificial nature of the content they generate, even to professionals with the necessary knowledge and expertise. This makes the process of manual fact checking time-consuming and, on more than one occasion, incomplete (Zellers et al., 2019).

Scalability in Production: The amount of content generated by AI puts current moderators and detection systems to their knees. The volume of synthetic media exceeds the capacity of social media providers and news outlets to effectively combat it, resulting in a reactive rather than proactive stance against misinformation (Li & Yang, 2024).

Evasion Techniques: The attackers can change their methods and keep on acquiring tricks to counter algorithms.

Attribution Problems: Misinformation that is produced by AI cannot be traced easily.

No Global Harmony to Regulation: There is no single international law or regulatory framework that can address the issue of AI-generated fake news. In the Pakistani context, the challenge is further intensified by an AI policy vacuum, where adoption of artificial intelligence is increasing but clear regulatory, ethical, and institutional safeguards remain underdeveloped (Rafiq-uz-Zaman, 2025).

Technological Arms Race: Generative AI is typically developed further than the detection tools. This turns it into the arms race that will never come to an end as soon as the methods of detection will be outcompeted by superior generation techniques (Li & Yang, 2024). AI governance requires more than technical detection tools; it also needs ethical innovation, institutional accountability, and a shift from passive adoption to responsible governance frameworks (Ali & Rafiq-uz-Zaman, 2025).

The challenges outlined above point to the need of the multi-faceted nature of the approach to the dealing with the GAI challenges by incorporating the technology-based solutions, policy, media literacy, and international cooperation to safeguard the sustainability of information ecosystems against the pernicious impact of GAI.

2.4. Digital Landscape and Media Consumption in Pakistan

Over the past few years, Pakistan has been enjoying a digital boom that has largely been fueled by the fast proliferation of the internet and increase in the number of users to the social networks. Digital expansion in Pakistan has increased the importance of digital literacy because citizens now require stronger critical skills

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

to evaluate online information, identify unreliable sources, and respond responsibly to digital content (Rafiq-uz-Zaman, 2023).

2.4.1 Internet Penetration and Social Media Usage Statistics. The penetration of the internet in Pakistan is well disseminated, but a high disparity is still observed in the regions. By early 2025, the internet penetration rates were about 45-51 percent of the total population approximately 116 million people, which means that the adoption of the digital world has increased at a high rate in recent years. . However, studies show that barriers such as affordability, inadequate infrastructure, and low digital literacy continue to sustain a digital divide among individuals in the country.

The main statistics of the digital world of Pakistan are:

Total Internet Users (early 2025): about 116 million

Early-2025 Internet penetration rate: 45-51% of the population

Active social media population (early 2025): 66.9 - 71.7 million

Prevalent Social Media: Facebook

Gender Inequality: A high proportion of dominance of social media among men.

Such statistics indicate a high number of online users and a strong online presence, which means that Pakistan has become an important field of information distribution and, hence, misinformation.

2.4.2 Role of Social Media in Shaping Public Opinion in Pakistan. Social media sites are among the leading sources of news and information to Pakistani people and especially young citizens in shaping their opinion in the country. Generation Z is especially important in this debate because young digital users are deeply connected to online platforms, but their ability to manage misinformation depends on digital literacy, innovation awareness, and psychological well-being in digital learning environments (Rafiq-uz-Zaman et al., 2026). These platforms have the advantages of accessibility and immediacy, which makes information spread fast making citizens participate in political debate, opinions, and mobilizes on various issues.

The social media in Pakistan has been observed to affect the way people think when engaging in political practices, social gatherings etc. and crisis situations. The influence of social media on users' behaviour, attention, and academic or social outcomes indicates that online platforms can shape not only information access but also attitudes, perceptions, and everyday decision-making (Rafiq-uz-Zaman & Malik, 2025).

Consequently, the merging of generating generative AI on a large and active number of social media users, the powerful position of social media in shaping the national-level discourse and the precondition of media illiteracy establishes a suitable environment in which generative AI can generate considerable shift and influence within popular opinion in Pakistan.

2.5. Theoretical Frameworks for Understanding Misinformation

In order to thoroughly examine the process of generative models in forming the opinions of the population through misinformation, one must refer to the already existing theory, which explains the factors of information processing in individuals and the ways that the false story can be cultivated in society. The theories give a perspective of how the mechanisms of influence and vulnerability can be comprehended.

2.5.1. Information Overload Theory. The Information Overload Theory assumes that when people are presented with too much information, their level of processing such information drops.

Shallow Critical Evaluation: Information overload can cause one to use heuristics or shortcuts, including being influenced by headlines, source credibility signals (even when this is false), or even social signals (i.e., the number of shares that a post might get) instead of a careful assessment of material before reading.

Greater Vulnerability to Simple Narratives: Nuanced or complex information takes more mental effort to consume. Simpler, highly dramatized, or emotionally appealing stories, like those used by misinformation, are much easier to consume and retain in an overloaded environment.

Fatigue and Apathy: It is known that continued exposure to excessive information flow can result in information fatigue in which people maintain an indifference or a lack of interest in confirming the information, resulting in a higher risk of becoming susceptible to false reports.

Generative AI promotes information overload by facilitating the synthesis of tremendous amounts of

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

content at a rate that is even higher than before, resulting in an increasingly congesting sphere of information and possibly exceeding the limits of human ability to critically analyze the information.

2.5.2. Social Amplification of Risk Framework. The Social Amplification of Risk Framework (SARF) describes the processes through which risks (and therefore information about risks, both correct and incorrect) are amplified or dampened out by social, cultural and psychological mechanisms.

Social Stations: Information falls through different stations (e.g. media, social networks, opinion leaders), which interpret it, reinterpret the information and spread it, In emerging-market contexts, WhatsApp groups and similar informal digital networks operate as grassroots communication ecosystems, which can support collaboration but may also accelerate the circulation of unverified information when users share content without verification (Rafiq-uz-Zaman et al., 2025).and in many cases, inflate or tame the magnitude of perceived importance.

Cultural Filters: Societal values, beliefs and cultural customs are filters that influence the reception and re-transmission of any information. The misinformation, which fits into or reinforces existing cultural storylines, or prejudices, is more likely to be amplified.

Behavioural responses: Enhanced information may cause large behaviour responses such as a hammering outcry, a shift of policy, or even cohesive action, though the initial information may be false or exaggerated.

In the case of AI-generated misinformation, SARF can be used to understand how deceptive narratives, especially those that appeal to passionate feelings or that resonate with natural aspects of societal separation, such as political party divisions, can quickly gain traction on social media platforms in Pakistan to generate an outsize popular response and influence mass opinion.

2.5.3. Cultivation Theory. Cultivation Theory, which is mostly related to George Gerbner, argues that the world and reality feelings are created according to the long-term exposure to media and mostly television.

Mainstreaming: The process through which intensive viewing produces a narrowing of perspectives between various demographic categories as media representations suppress the existence of differences in the outlooks that may otherwise exist on other grounds.

Resonance: This comes in when the real-life experience of an individual resonates with the media portrayal thus, increasing the cultivation effect.

Although it was initially used in relation to legacy media, Cultivation Theory is very much applicable in the digital era and the effects of misinformation initiated by artificial intelligence.

2.6. Research Gap

Although the present state of knowledge on AI, misinformation, and its implications to society is growing, there are yet still several gaps in research as far as the particularities of the Pakistani context and the fragile position of generative are concerned models. These gaps need to be filled by ensuring that effective strategies are developed to contain the adverse effects of the AI-generated misinformation.

Based on the comprehensive literature review, the following research gaps have been determined according to the literature review:

This study will fill these research gaps and lend a well-contextualised, more nuanced understanding of the risk that AI-generated misinformation poses in Pakistan to aid more effective and context-specific interventions

3. Methodology

The study is qualitative, explorative, and descriptive, using a qualitative methodology to explore how generative AI, misinformation, and opinion are interacting together intricately in the context of Pakistan.

3.1. Research Design

The given research follows a multi-method qualitative research design by following the main features of content analysis and a case study strategy. It can be analysed as a whole phenomenon, but also elaborate on individual cases of AI-based misinformation through both general and specific analysis of online content.

Exploratory: The study is an exploratory study because it aims to reveal and comprehend the new tendencies and features of AI-generated misinformation in Pakistan, which is rather understudied. It tries to

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

find trend patterns, recurring issues, and the particular mechanism by which generic models are being relied on.

Descriptive: The article will report/ elaborate on the situation with AI-generated misinformation in Pakistan by stating how widespread it is, its forms, as well as the mediums it is distributed by. It shall also outline the realized effects on the opinion of the people and the current response measures of the policies.

Qualitative: The major focus here would be more on the how and why of the phenomena in place of what or how much. This entails meaning-making, storytelling, and contexts that are associated with misinformation created by AI.

3.2. Data Collection Methods

Data for this research will be primarily collected through two main qualitative methods:

3.2.1. Online Content Analysis (Social Media & News Articles). The online content will be analyzed systematically in an attempt to investigate publicly accessible online content in terms of AI-generated misinformation in Pakistan. This will entail:

Social Media Platforms: The content of the most popular social media in Pakistan, such as Facebook, X (previously Twitter), YouTube, and TikTok, will be analyzed.

News Articles and Reports: Analysis of news articles published by credible Pakistani and non-Pakistani media sources, and of articles published by fact-checking organizations and digital rights foundations (with presence in Pakistan). This will assist in recognizing instances of misinformation generated by AI, their sources, distribution, and first effects that were documented.

Academic Databases and Research Repositories: The review of academic articles, theses, and reports of local and global research organizations and institutes related to AI, misinformation and digital media in Pakistan. This will give us a background knowledge and bring out cases that have been documented.

The data will also be preferred to the period 2023-2025 in order to collect the latest tendencies and developments regarding GAI misuse.

3.2.2. Case Study Approach. In order to demonstrate the real-life consequences of AI-based misinformation and offer deeper understanding of the matter, a set of case studies should be created.

Political Disinformation: Cases of AI-created materials that have been employed in the context of elections, or political crisis to control any discourse or misuse it against political subjects.

Social and Cultural Misinformation: Inputs involving instances of GAI being employed to propagate misinformation and that affects social peace, religion or a particular population group (e.g. women politicians).

Geopolitical Misinformation: Case law of geopolitical and international misinformation caused by AI-generated content and applied to regions of conflict or inter-country relations with Pakistan.

Each case study will include a thorough analysis of the content, channels of its spread, the type of advanced AI tools that are speculated or proven to be used, the response of society and any governmental counteraction or factchecking.

3.3. Data Analysis Techniques

The qualitative data collected will be analysed using the following techniques:

3.3.1. Qualitative Content Analysis. The online content (social media posts, news articles) and case study materials will be analysed by qualitative content analysis. This organized procedure includes:

Coding: Pointing out and classifying relevant matters, patterns, and ideas in the data. This entails its forms of misinformation, generative AI-related tools involved, stories used, most likely audiences, and emotional appeals.

Categorization: Sorting codes into higher and lower categories in order to come up with a complete picture of an aspect of the phenomenon.

Interpretation: The analysis of connections between categories and themes to make relevant conclusions about the role of generative AI in influencing the opinion of the population.

3.3.2. Thematic Analysis. Analysis of patterns (themes) within the data and especially case studies will be identified, analysed and reported through thematic analysis. This involves:

Familiarization: Getting acquainted with the data in order to have a wide picture.

Initial Code Generation: The production of initial codes out of the data.

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

Strategy of searching Themes: Categorizing the codes into possible themes.

Themes Review: Considering themes and how to refine them so that they represent the data.

Defining and Naming Themes: Creation of the definitions which help to define themes and the names of the themes.

Reporting the Report: Creating the Report by integrating the thematic analysis within the entire story of the research paper.

3.4. Ethical Considerations

Paying needed attention to the ethical issues is required when conducting research on misinformation, particularly in relation to online content and to sensitive political situations:

Privacy and Anonymity: Privacy of the analysed work will be respected by keeping individual participants and sensitive information as anonymous where possible; particularly when including social media posts in a direct quote.

Minimizing Harm: The study will not take any steps that might end up resulting in the propagation of misinformation or harm to a group or individual. This involves the proper presentation of sensitive matters and not fuelling false stories.

Objectivity and Bias: When carrying out the research, the researchers will be and remain objective during the collection and analysis of the data presented, with the consideration of possible biases and reduction of its impact. The coding and thematic analysis will be carried out by multiple researchers to increase the inter-coder reliability and decrease the individual biasness.

Transparency: The analytical processes, the data used, and how the methodology was carried out will be straightforwardly explained so that the research can be replicated easily.

3.5. Research Framework

This research is guided by an integrated conceptual framework that combines elements from the Information Overload Theory, Social Amplification of Risk Framework (SARF), and Cultivation Theory, as discussed in the literature review. This framework helps to understand the multi-layered process through which AI-generated misinformation influences public opinion in Pakistan.

Explanation of the Framework:

1. Generative AI Models (A): This is the beginning, which signifies the ability of technology to generate synthetic content.

2. Content Generation (B): The means through which GAI uses to generate misinformation of different kinds (text, image, audio, video).

3. AI-Generated Misinformation/Disinformation (C): AI-Generated inaccurate information then released into the digital ecosystem.

4. Digital Ecosystem in Pakistan (D): The main mediums of such going viral especially the social media platforms, with high internet penetration and active internet population(s).

5. Information Overload (E): The number of available contents, which is increased by GAI, causes information overload of users therefore decreasing their level of critical thinking.

6. Social Amplification of Risk Framework (SARF) (D to E): Assignment of misinformation across the social network and algorithmic biases add to the phenomenon of information overload and shape popular opinion.

7. Formation of Public opinion (F): The main result of the effect of decreased discursive judgement in the limits of the information overload, repetitive views of synthetically made materials (Theory of Cultivation), and assimilation through cultural, linguistic lenses.

8. Cultivation Theory (C to F): Prolonged exposure to narratives by AI influences the perception of reality among users, and they come into line with the false information.

9. The last three filters (D-F) are Cultural/Linguistic Filters: Misinformation is perceived, understood, and believed differently based on a particular socio-cultural and linguistic background of Pakistan.

10. Effect on Trust, Polarization, Behaviour (G): The projection of the manipulation of the creation of public opinion (through misinformation) demonstrates a very real result or consequence of the phenomenon, namely, loss of trust in institutions, magnification of polarization in a society, and changes in the behaviour

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

of a population.

11. Policy & Regulatory Responses (H): Policy and regulatory measures are taken by governments and other stakeholders in the form of policies and regulations, which should circulate in a manner that can dampen down the transmission and effect of AI generated misinformation.

This framework allows us to compare and contrast the data, draw meaningful conclusions, and answer the research questions since it shows the inter-relationship between technological, social and cognitive factors related to AI-generated misinformation propagation and its effects on subject matter.

4. Findings and Case Studies

The section introduces the results of the online content analysis and full-fledged case studies with the purpose to demonstrate the ubiquity and diversity of manifestations of A.I. generated misinformation in Pakistan; The discussion shows that generative models are being used at an increasing rate to build very believable And expandable false narratives with impacts in various parts of the public life.

4.1. Overview of AI-Generated Misinformation in Pakistan

Pakistan has a set digital environment with a rich application of social media platform and diverse media literacy that provides a perfect environment on how AI-generated misinformation can spread. According to our review, the volume and even sophistication of synthetic content has exploded, particularly after 2023 when a host of sophisticated generative AI tools are more widely available.

One can note the following:

Deepfakes Pervasiveness: Deepfake technology, which often manipulates videos and sound, is one of the most widespread means of producing political and social disinformation. It includes altered speeches of prominent individuals and entirely made-up compromising personas, ultimately undermining public trust

Textual-based Misinformation: Large Language Models (LLMs): It is widely used to generate fake news, social media content, and propaganda-based narratives, which imitate the forms of authentic journalistic writing or conversations. This material typically exploits pre-existing social divides or political sensitivities.

Image-based Misinformation: In several cases, AI-generated images that are usually endowed with fabricated events or of fictional people are being used to give published text-based misinformation a band of is authenticity.

Rapid Dissemination: The information is rapidly distributed through very popular online social media sites like Facebook, X (previously twitter) WhatsApp and is often augmented by automated bots and innocent user.

Specific Narratives: You will find that misinformation campaigns can be done with specific groups in mind utilizing aspects of culture and language to make maximum effects and appeal to the audience that it is directed to.

The case studies below will demonstrate concretely how the generative AI has been used to influence the masses with respect to shaping the public opinion in Pakistan, showing the working process through the mechanisms of creation, diffusion, and influence.

4.2. Case Study 1: Political Disinformation (e.g., Election-related Deepfakes)

Context: Pakistan politics can be marked with a prevalent rivalry and abundant elections and thus, it is one of the targets of information manipulation. The 2024 general elections have been specifically marked by the overwhelming increase in the use of AI-generated content as a means of disinformation in politics

Incident Description: A rather salient incident was the spreading of deepfake video and audio files that were also used against important political figures.

Generative Models Employed: Although it is difficult to be more specific, based on the complexity of such deepfakes, the likelihood of advanced and GAN-powered video and audio editing systems was ensured.

Spreading and Influence: These deepfakes mainly spread in the most common social media such as X (formally Twitter) along with Facebook, and WhatsApp groups by taking advantage of their fast sharing features, and the increasing popularity of content based on algorithms.

Collapse of Trust: The spreading of such convincing counterfeits introduced a tendency towards

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

overall loss of trust in the information presented on the internet, thus making it more difficult of citizens to tell between genuine information and hoaxes. This introduced a situation of doubt and bewilderment.

Voter manipulation: The implication of such deep fakes was obviously to cause a distortion of public opinion and voter behaviors. Although it is hard to measure the direct influence on the election outcomes, the scale of disinformation campaigns and the targeted audience within them leaves little to be desired to define it as a substantial perception management operation.

Further Polarization: These deepfakes further polarized society because they made contentious claims, or performed contentious acts, and people on both sides of the political fence have intensified hatred towards each other.

Response: In Pakistan, fact-checking groups and certain reporting outlets were trying to dispel these deepfakes, but were in most cases, beaten to the punch by the information spreading out fast.

4.3. Case Study 2: Social and Cultural Misinformation (e.g., Targeting Women Politicians)

Background: Female representation in Pakistani politics has been a major issue where women have been judged mainly based on gender lines and online bullying. With the rise in generative AI, there is a new menace to women politicians as it can be used to ridicule and shut women down based on prejudices and cultural views that a society may have

Incident Description: An especially troubling pattern on display in Pakistan has been the use of AI-generated obscene explicit videos and photos, or other defamatory materials, to attack high profile women politicians.

Generative Models Involved: Deepfake video and image generation is very high in these malicious campaigns.

Dissemination and Impact: Dissemination works through the use of the main two avenues, which include through the use of a privately owned messaging service such as WhatsApp and public social media, and are usually boosted by users with anonymous profiles and networks that have been organised.

Response: Although there have been public protests and occasional legal redress taken with regards to existing cybercrime law (such as PECA), the absence of a specific law banning the act of deepfakes, as well as difficulties tracking down culprits on the anonymous internet environment, also makes the successful pursuit of remedy an uphill task . The cases demonstrate the necessity of strong laws and law enforcement to safeguard individuals and especially vulnerable segments of the population against gender-based violence and defamation against them, facilitated by an AI tool.

4.4. Case Study 3: Geopolitical Misinformation (e.g., India-Pakistan Conflicts)

Context: Geopolitical rivalries between India and Pakistan.

Incident Description: The increasing AI generated content in both countries has brought about the dramatic rise in information that tries to sway the domestic opinion when we are in a tensed situation or conflict e.g., in a border skirmish or when the countries are headed towards a diplomatic crisis. Examples include:

Generative Models Utilized: These geopolitical disinformation campaigns feature heavy usage of the combination of generative AI tools.

Dissemination and Impact: The dissemination plan tends to be multi-platform based on where the information will be distributed using main social media platforms (X, Facebook, YouTube), messaging based (WhatsApp) and to some degree even adopted by less ethical traditional media sources. The effect is massive:

An increase in Tensions: Artificially intelligent-generated misinformation may directly lead to an intensification in tensions between the two countries due to the proliferation of misinformation with provocative measures or the stretching of events that may result in misjudgments.

Mobilization and hate: The intention of false narratives to demonize the other side panders to public anger and hate given that it is easier to get diplomatic solutions to a conflict; this environment creates fertile ground that facilitates conflict.

Propaganda of the Real: This is because the stream of synthetic information blends truth and fiction into one another as it becomes hard to make objective judgments with actual generated information, so rational debate runs the risk of being eliminated in favor of the blurring of fact and fiction.

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

Response: In both India and Pakistan, organizations and individual reporters working in the field of fact-checking have been working to expose these AI-driven fake news.

4.5. Analysis of Generative Models Used and Their Impact

The above-given case studies demonstrate in a pictorial manner how generative AI models can be practically applied and immensely influential in creating the public opinion in Pakistan. A closer examination of the type of models that are used, the strategic utilization of its potential to maximize reach and persuasiveness of misinformation demonstrates some similarities.

Deepfake Technologies Monopoly:

In all the discussed case studies, especially political and geopolitical ones, the technologies of deepfake (mostly video and audio synthesis) come out as one of the prevailing tools.

Making less visible but critical contributions to the construction of the stories that are used to frame and situate the visual or audio material created through the use of AI, LLMs can be found. They use their capabilities to:

Scalable Content Production: The rate and amount of content LLMs are able to generate legislatively, allows disinformation actors to saturate informational vectors, overshadowing fact-checking initiatives and people with the perpetuated false narratives at all times.

Combining Models of Higher Impact: Some of the most effective misinformation campaigns have the synergistic fact that they used various generative models.

Impact on Public Opinion: The total impact of these generative models on the attitude of individuals in Pakistan is immense and giant multidimensional:

Weakening of Confidence in the Sources of Information: This is due to the pre-condition of more-like confusion between the true and the fake, and to the perennial exposure to very real semi-realistic fake contents, there is a virtually complete collaterating aura of distrust to all media, including traditional news and official pronouncements.

Control Over Political Discourse: Manipulating the political discourse directly, by using AI as a tool of spreading fake news about electoral processes and affecting the progress of policy-making by introducing false information into the political discourse becomes within the capabilities of malicious actors.

Generative AI models, in other words, have revolutionized the nature of misinformation, turning it into an affair that is cheap to scale, far more successful in its realism and growing harder to counter. The consequence of their influence into the minds of people in Pakistan can be outlined as the worsening crisis of trust, enhanced fragmentation of society and the major dilemma of the democratic process legitimacy.

5. Discussion and Implications

The conclusion of the case studies and the overall suggestion of the AI-generated misinformation in Pakistan report key effects of the state of public opinion, unity in the society, political stability and economic prosperity.

5.1. Impact on Public Opinion and Trust

The main short-term and widespread effect of misinformation created by AI is loss of trust in society.

Generalized Distrust: As the case studies have also shown, recurring experience with AI-generated fakes, particularly the rise of fakes that go viral before being debunked, creates a general situation of generalized distrust.

Increased mental load and fatigue: The Person is forced to constantly verify information and question its accuracy, thus taking a tremendous toll on the mental load.

Reduction of Seed Biases: AI-created misinformation is likely to tap into already established biases and views.

Reduced Public Discourse: Any complex form of discourse in the society would be impeded by the predominance of distrust and inability to identify truth in the public sphere.

The situation is especially dire in Pakistan, where media literacy levels are different and the fact a great number of people use social media to acquire their news means the sheer effect it has on the opinion of people and the way they choose to trust it is severe.

5.2. Societal Implications (Polarization, Social Cohesion)

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

In addition to the short-term effects on citizen opinion and feelings of trust, the spread of the misinformation products of AI has major implications to the society, specifically to social cohesion and polarization.

Raised Polarization in Society: Generative AI also allows extremely specific and emotive messages to be built that can further divide any society along the lines of politics, religion, ethnicity, or sects.

Loss of Social Cohesion: In a case whereby a huge proportion of the population is exposed to and believes different and, in most cases, contradictory realities generated by AI, it is hard to build social cohesion on important national matters.

Encouragement of Violence and Discrimination: The worst effect on the society is the way AI-created misinformation can lead to violence and discrimination in real life.

Delegitimising Democratic Processes: In an informed participatory process is the order of the day in a democratic society. Cornerstone.

Influence on Disadvantaged Populations: The sections of the population that are not so aware of the extent.

Conclusively, the free flow of AI-created fake news creates the threat of breaking the Pakistani society went up, still trying to cement a number of existing rifts, and destroying the ethos of social trust and collective action. Addressing such implications on the part of the society is the necessity to devise concerted efforts in order to become resilient in the information environment and allow citizens to manage the challenges of the digital age.

5.3. Political Implications (Electoral Integrity, Governance)

Given that electoral processes and the sanctity of governance are paramount to the political process, the political implications of AI-created misinformation are far reaching and directly threaten the integrity of elections, quality of governance as well as the stability of governance.

Electoral Manipulation: The manipulation of voter outcomes during election times is also a clear asset that could be used through AI-generated deep face use and fabricated stories as seen in Case Study 1. This encompasses development of counterfeit endorsement, dissemination of defamatory information about candidates, or elicitation of counterfeit news about voter-polling procedures (e.g.

Subverting Governance and Policy-Making: Even misinformation, particularly when boosted through AI could severely backfire on proper governance.

Destabilization and social unrest: the AI-induced geopolitical misinformation, as demonstrated by Case Study 3, can lead to intensification of international antagonisms and even social unrest within one of the countries.

Dissolution of Confidence in Political Institutions: Public trust in political institutions diminishes when, on an ongoing basis, political leaders or institutions are the object of a smear campaign, or when people feel that the political institutions lack the ability to counteract dissemination of fake news. This may give rise to political apathy, /dissociation, or a shift towards populist or extremist ideologies that offer to provide clear answers to difficult questions, frequently by appealing to popular resentment based on misinformation.

Difficulties Faced by the Law Enforcement and Security Agencies: The quick pace at which AI-generated misinformation is developing poses a new challenge to law enforcement and security agencies.

In Pakistan, where politics contains a lot of heated rhetoric and national faith in various governments is weak, the unregulated spread of AI-created rumors are a major threat to the democratic process and also the ability of effective and stable government. It requires a proactive and multi-direction approach in order to ensure protection of the political sphere.

5.4. Economic Implications (Damage to Reputation, Market Manipulation)

Economic consequences of AI-created misinformation tend to be less apparent than social and political ones but can have huge and long-term ramifications. These effects can expose individuals, companies and national economies.

Loss of corporate image, brand image: AI-powered automated content may create fake reviews, libelous articles, or spoofed images/video that circulates misinformation about a business, product, or service.

Manipulation of the market: Generative AI may be used to propagate fake financial publications,

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

rumors, or reports to manipulate the pricing of a stock, cryptocurrencies or other market values.

Costs to Specific Industries: It hits some industries especially hard.

Verification and Cybersecurity: Companies and government agencies will have to pay more in terms of expenses.

Loss of Investor Faith: This type of clouding with lies could frighten off foreign direct investment (FDI) and suppress economic growth. Investors would like to have a stable and predictable environment and a country that is tormented by far-reaching and well-structured disinformation can be regarded as excessively dangerous, and it shall be the loser in matters of economic prospects.

In the case of Pakistan, an aspiring new economy wanting to make capital investments and engender digital technology, the threat of misinformation attributable to AI is large. It is costly to the economy since it can pull down industry confidence, misrepresent economic forces, and it can waste resources that are needed elsewhere in economic activities that are productive.

5.5. Challenges in Detection and Counter-Measures in Pakistan

Contrary to the increased awareness of AI-generated misinformation, Pakistan has specific and considerable issues in terms of its detection and counter-measures. All these problems are the result of a blend of technological, socio-cultural, and regulatory issues.

Technological Lag and Resource Constraints: There are more sophisticated versions of AI detection tools deployed in the world, but in Pakistan, their availability and operation is impeded by the lag in technology and a lack of resources.

Linguistic and Cultural subtleties: Majority of AI detection models are trained utilizing huge amounts of information that is mostly in English.

Low Media Literacy: A considerable sector of the Pakistan population, mainly in rural settings, does not have good digital and media literacy.

Fast Transmission on Messaging Apps: People pay some attention to public social media, but a lot of misinformation circulates via locked messaging programs such as WhatsApp. They are harder to monitor and fact-check as these closed groups enable false narratives to spread fast in trusted social groups before they can be disproved.

Regulatory and Enforcement Gaps: Pakistani key law like the Prevention of Electronic Crimes Act (PECA) does handle fake news, but the application of such legislation to sophisticated AI models remains very weak, and in the case of deepfake models, is virtually non-existent.

Political interference/ freedom of expression issues: Regulation of misinformation may be construed as an attempt to limit free speech and crackdown on dissidents in some cases.

The Lack of Collaboration and Coordination: Unit since successful counter-misinformation should be enabled by effective interagency, corporation, civil society organizations, media outlets and academia collaboration and coordination. In Pakistan the coordinated efforts are in many cases isolated and therefore do not really make much of a difference.

These obstacles point to the fact that technological solution is not enough. In Pakistan, the resilience to AI-generated misinformation can only be achieved by a holistic strategy to combine policy, education, and cross-sector partnerships.

6. Policy and Regulatory Landscape

Addressing the challenges posed by AI-generated misinformation necessitates a robust and adaptive policy and regulatory framework.

6.1. Existing Laws and Regulations (PECA, etc.)

In Pakistan, the main law-enforcement tool dealing with cybercrimes, including the spread of false information, is the Prevention of Electronic Crimes Act (PECA) of 2016, having successive amendments.

Section 20 (crimes against the dignity of a natural person): This section is a crime against intentionally and publicly displaying or sending information that might damage the reputation or privacy of a natural person. Although not specifically talking about AI, it has been used to convict individuals who share defamatory information and this is likely to cover AI-generated information and content.

Section 37 (Unlawful Online Content): The Pakistan Telecommunication Authority (PTA) is granted

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

immense power to prohibit or take down any online information which is considered unlawful, including information, which is misleading, information which is deceptive, and even information that conflicts with the integrity, security or defence of Pakistan. This offers legal authority in taking action against misinformation irrespective of their source.

Other related amendments: PECA has also been amended, especially in early 2025, to augment sections that address the wilful distribution of fake or false information.

Other legislation, including the Defamation Ordinance 2002 and sections of the Pakistan Penal Code, can also be applied to those who share false information, although doing so in relation to online AI generated material can often be complicated.

Besides legislation, the Pakistani government has also started a wider ranged digital transformation policy, including the Digital Pakistan Policy, and debates on a National AI Policy.

Although these pre-existing frameworks give some legal protections against misinformation, how well these mechanisms, currently, stand up against the fast-changing generative AI technologies is raising one serious concern.

6.2. Gaps in Current Legislation Regarding GAI

Although the existing legal and regulatory frameworks do cover PECA and other laws dealing with related issues, there have been great gaps in legislation to tackle the challenge posed by generative AI-generated misinformation, especially the pernicious deepfakes. These loopholes compromise the capability of authorities to pre-empt, identify, and convict the abuse of GAI.

The Notion of Deepfake Lacking Specific Legislation: The largest gap is a lack of deepfake-specific legislation that explicitly defines deepfake and criminalizes its creation, distribution, and/or malicious use.

As an example, establishing that a person or organization was working to deceive or cause harm, a criterion in many existing laws, may be difficult in light of state-of-the-art AI-created content.

Issues of Attribution and Provenance: Existing legislation grapples with the technical intrigues of the attribution of AI-generated work to its creator.

Absence of a Comprehensive Data Protection Law: Even in the air of the discussion of a National AI Policy, Pakistan has no law of data protection.

Absence of a Comprehensive Data Protection Law: Even in the air of the discussion of a National AI Policy, Pakistan has no data protection law. This lack poses vulnerabilities as training data that informs the models of generative AI includes personal data, and existing privacy regulations do not appropriately protect against deepfaking practices (e.g., non-consensual intimate imagery) based on this data A healthier data protection structure is critical to the governance of the ethical design and application of AI.

Narrow Scope of Platform Liability: Although the PECA can also lead to the blocking or taking down of illegal content, it does not pay enough attention to making social media platforms and technology companies directly responsible and liable enough of the explosion and the algorithmic amplification of AI-generated misinformation. The existing system's focus on a user or the state to report and remove false information rather than nudging platforms to engage in proactive measures of detection and moderation.

Improper Technical Expertise among Regulatory Authorities: Regulation of AI-generated content needs a thorough knowledge of the same technology.

Finding a balance between the misinformation suppression and the primary rights, specifically the freedom of expression.

These legislative and enforcing loopholes are enough to prove that Pakistan should act fast in order to come up with more advanced, technology-intensive, and rights-congenial control systems to assist in curbing the special threats associated with generative AI-generated fake news.

6.3. Proposed Policy Recommendations

To effectively counter the growing threat of AI-generated misinformation and safeguard public opinion in Pakistan, a multi-pronged and adaptive policy approach is required. These recommendations aim to address the identified legislative gaps, enhance public resilience, and foster responsible technology use.

6.3.1. Legislative Reforms. An effective strategy that strikes a balance between combating the growing problem of AI-based misinformation and protecting the opinions of the populace in Pakistan is

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 5 Issue 3, 2026

required in a multi-pronged and evolvable fashion. These suggestions will be used to solve the legislative gaps, improve the civil resilience, and promote the civil use of technologies.

Specific laws: Pakistan should ensure it has specialized enforcement measures to combat deepfakes, specifically through legislation that formally defines and criminalizes the intentional publishing and production of deepfakes, especially those meant to defame, harass (especially gender-based), interfere with elections and incite violence.

Exhaustive Data Protection Law: The law on data protecting need to be in place hastily.

Increased Platform Accountability: A new law should be developed that allocates more responsibility on platforms and technological companies regarding the content that they are propagating on their platforms. This may include;

Mandatory Transparency: Platforms would be required to post their content moderation policies and how they use AI to amplify their content as well as meet the requirement provided by the data on the detected misinformation.

Proactive Moderation: Requiring platforms to invest in artificial intelligence-based systems to detect misinformation coupled with the hiring of human moderators who can screen and remove artificial intelligence-generated misinformation, particularly deepfakes, on a timely basis.

Technical Solutions: By using streaming or e-book search technologies, it is possible to trace content provenance and the origins of AI-generated works; this can be achieved through technical solutions like digital watermarking or cryptographic signatures of the generated content that it can be verified whether an item is AI generated.

Analysis and Revision of PECA: Although having the provision of PECA can offer a good start, the process should encompass evaluation and upgrading to embrace technological progress. This encompasses making the definitions clear that apply to AI-generated content and testing that enforcement measures are up to the task of dealing with the unfamiliarities of digital evidence.

6.3.2. Media Literacy Initiatives. National media literacy program: The development of a far-reaching national media literacy program that would be part of school curricula all the way up to the adult stage. This course ought to be directed at:

Critical Thinking Skills: Educating people about how to critically examine material, recognizing logical fallacies, and making informed inquiries of the source.

Digital Forensics Basics: Creating awareness/guides on what to look out for in AI-generated content (e.g. inconsistencies in deepfakes, unnatural language in LLM-generated text).

What is Algorithmic Amplification? Making sense of how social media algorithms function, and how they can accidentally boost misinformation.

Responsible Sharing: Discoursing about rightful sharing of information online, such as verifying what one shares.

Partnership with the Civil Society and Media: Cooperating and assisting civil society groups, fact checking and independent media outlets that have a proactive participation in media literacy and debunking activities. They can be much more far reaching and effective when they are given resources and technical assistance.

6.3.3. Collaboration with Tech Companies.

Public-Private Partnerships: Developing stronger relationships between the government, academia and major AI and technology companies. The cooperation must be devoted to:

Research and Development: collaboratively invest in research and development of high-tech AI detection technologies, especially those that fit within the social and cultural setting of Pakistan.

Information Sharing: The development of rapid information exchange both across platforms, government agencies, and fact-checkers when it comes to the emergent threats and patterns of AI-generated misinformation.

Capacity Building: Technology corporations can help to increase the technical capabilities of Pakistani regulation and law enforcement roles to improve their level of knowledge in these areas.

Ethical AI Development Guidelines: Stimulating and, where possible, requiring the adoption of

ethical AI development guidelines by the companies operating in Pakistan. To support safety, accountability, and transparency in the development of generative AI models, these guidelines should make a priority of the safeguards against its possible misuse.

Going by the proposed policy recommendations, Pakistan will be in a position to strengthen its information landscape, limit the negative effects of AI-propagated fake news, and utilize the positive potential of artificial intelligence in the national development process and protect its information space.

7. Conclusion and Recommendations

Combining resources and applying a multistakeholder approach is the necessary step to take to overcome the complicated problem of AI-generated misinformation. The recommendations that shall follow, are focused on different actors in Pakistan to develop a stronger and more knowledgeable civic sphere:

7.1 For the Government and Regulatory Bodies

Pass Deepfake-Specific Laws: Focus on the issuance and enactment of unambiguous, comprehensive laws focusing specifically on targeting malicious generation and sharing of deepfakes.

Enhance Data Protection Legislation: Improve the speed at which a strong data protection legislation is adopted and implemented, including procedures of how personal information is collected and dealt with, and so on the way they are used in AI models. This will provide a significant legal structure on which to build on privacy violations that come along with deepfake and other AI-generated media.

Increase Technical Capabilities: Invest in the technical capabilities and forensic skills of the law enforcement agencies, regulation (e.g., PTA) and judiciary in the discovery, investigation and prosecution of AI cases.

Encourage Responsibility: Introduce policies that keep the leaders of social media and technology accountable companies that are more responsible to the contents of their services.

Invest in Public-Private Partnerships: Form institutions of cooperation between governments, academia and key AI and technological companies. The spheres to which this partnership needs to pay attention are collaborative research on detection technologies, disseminating the information about the emerging threats and capacity building initiatives.

7.2 For Media Organizations and Journalists

Invest in fact-checking and verification: Invest and give priority on setting up fact-checking units with the help of latest AI detection tools as well as trained personnel. Establish expedited certification and debunking process on AI-created misinformation especially during critical times like these in elections or in case of a crisis.

Make the Administration respond: Diversify and enlarge the stewardship by media literacy programs which will assist the Administration in development and promotion of media literacy campaigns, educating the population on the ways of detecting AI-created content, learning about algorithmic amplification, and become more critical of theirs information sources. It can be done through the assistance of public service announcement, workshops and educational information.

Ethical Guidelines on AI Utilization: Draw and follow strict ethical guidelines on the application of AI technology in the field of journalism whereby the transparency of the use of AI tools in the production of journalistic content should be observed and journalistic integrity upheld.

Partner with Technology Firms: Cooperation with those in the tech industry will be crucial to speeding up the removal of false AI generated information with any debunked content and to provide access to data sets that would help identify how content is shared.

7.3 For Civil Society Organizations and Academia

Learn more through original research: Keep investigating through original research as the field of AI-generated misinformation in Pakistan evolves, placing particular emphasis on their social, cultural, and linguistic implications as well as the success of counter measures. Policy makers and awareness campaigns can use this research information.

Promoting Rights-Respecting Policies: Champion policies that are competent in counteracting misinformation and yet still respect key rights especially freedom of expression and privacy. Keep track of government activities so that regulations cannot be employed to muzzle justified protest.

Develop Locally Conceived Media Literacy Programming: Assess and develop locally conceived media literacy programming that is tailored to the needs and concerns of diverse societies within Pakistan as well as to the most vulnerable (and/or at-risk) groups most exposed to misinformation.

Investigate Fact-Checking Initiatives: Look into the work of independent fact-checking organizations to determine how they can be helped and supported, including which resources should be provided.

7.4 For the Public

Develop Critical Thinking: Critically evaluate information you see online and develop the skills of critical thinking when reading online. Question the source, seek out additional supporting evidence and be careful of emotionally charged or sensational headline.

Anything Worth Sharing is Worth Verifying: Avoid passing on content especially that which might seem to be either too good or too bad to be true, without making sure it is genuine with multiple credible sources.

Report Misinformation: Report possible AI-generated misinformation to the platforms, fact-checking organizations, or the authorities. Reporting is beneficial to the elimination of false content, and active involvement in reporting will facilitate the timely detection and elimination of false content.

Demand Quality Journalism: Find and support good journalism and individual journalists who are committed to high standards of journalism and factual reporting.

Combined, adopting these recommendations can help Pakistan create a more robust information environment, reduce the negative impact of fake news created by AI, and utilize the potential of artificial intelligence that can improve the life of its people dramatically.

7.5 Future Research Directions

Generative AI is fast-changing and requires constant research on how it is used in connection to misinformation. There is a need to conduct further research based on the identified research gaps that have been identified in this study, as well as the implications of this study.

Longitudinal Studies on Public Opinion Shifts: Conduct studies over a period of time to exactly identify the causal role AI generated misinformation plays in opinion shifts in Pakistan. This would deal with the changes in beliefs, attitudes, and behaviour over long time frames which would be reported accordingly with exposure to synthetic content.

Greater Cultural and Linguistic Adaptation: There is a greater knowledge deficit of how generative AI models are being specifically modified to existing cultural and linguistic nature of Pakistan. This will entail the investigation of the peculiarities of the content production in regional languages and the way cultures precondition the usefulness as well as distribution of disinformation.

Minimizing the Harms of Misinformation: A rigorous test of the efficacy of different classes of counter misinformation interventions in Pakistan, including media literacy interventions and fact-checking interventions, and regulatory interventions. Experimental or quasi-experimental designs will be appropriate in this study to establish the best technique in mitigating the impact of AI-generated misinformation.

AI and Counter-Misinformation: Research on the potential of counter-misinformation via AI itself, in Pakistan. This might involve the development of AI-based systems to detect synthetic content in its initial phase, the introduction of automated fact-checking depending on national backgrounds and AI-supported educational interventions.

Vulnerabilities and Resilience to Psychological Factors: Research the psychological challenges at work that are either increasing the susceptibility or stability of the people in Pakistan to the AI-generated misinformation. They can be the experimental studies on cognitive biases, emotional vulnerability, and the influence of social networks to support or justify falsehoods.

Economic Impact Quantification: Develop strategies that can be employed to further quantify the economic impact of the AI-generated misinformation in Pakistan such as losses to business start-ups, market manipulation and distinction of resources to address them.

International Cooperation and Best Practices: Research of the best practices in relation to the international cooperation to tackle the problem of cross-border AI-generated misinformation and find the most successful examples of cooperation with other countries and make the main conclusions to follow in Pakistan.

Relative research directions in the future will help to form a fuller picture of which threats and opportunities are posed by generative AI in the Pakistani information ecosystem and thus serve as a platform to develop more sustainable solutions to them.

Acknowledgments

The authors would like to thank the participants and the respective universities for their support and cooperation in conducting this study.

Contribution of Authors

All the authors participated in the ideation, development, and final approval of the manuscript, making significant contributions to the work reported.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of 1964 Helsinki declaration and its later amendments.

Data Availability

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

References

Ahmed, S., Masood, M., Bee, A. W. T., & Ichikawa, K. (2025). False failures, real distrust: The impact of an infrastructure failure deepfake on government trust. Frontiers in Psychology, 16, 1574840. https://doi.org/10.3389/fpsyg.2025.1574840

Akhtar, Z. Bin. (2024). Unveiling the evolution of generative AI (GAI): a comprehensive and investigative analysis toward LLM models (2021–2024) and beyond. Journal of Electrical Systems and Information Technology, 11(1), 1–21. https://doi.org/10.1186/S43067-024-00145-1

Ali, A., & Rafiq-uz-Zaman, M. (2025). Institutional inertia vs. ethical innovation: A comparative analysis of AI governance at The Islamia University of Bahawalpur and Cambridge University Press. Contemporary Journal of Social Science Review, 3(4), 91–102. https://doi.org/10.63878/cjssr.v3i4.1695

Ali, E., Afzal, U., Arshad, A., & Asim, R. M. (2025). Social media algorithms and their impact on political disinformation and misinformation. Journal of Media Horizons, 6(1), 335–345. https://jmhorizons.com/index.php/journal/article/view/127

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Arab News. (2025, January 6). 'Form of violence': From US to Pakistan, deepfake videos target women politicians. Arab News PK. https://www.arabnews.pk/node/2585378/pakistan

Azgın, B., & Kiralp, S. (2024). Surveillance, disinformation, and legislative measures in the 21st century: AI, social media, and the future of democracies. Social Sciences, 13(10), 510. https://doi.org/10.3390/socsci13100510

Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. Data & Policy, 3, e32. https://doi.org/10.1017/dap.2021.20

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Busch, E., & Ware, J. (2023). The weaponisation of deepfakes: Digital deception by the far-right. https://doi.org/10.19165/2023.2.07

Chaudhary, S. (2024). Women's politicians bearing the brunt of deepfake. Media Diversity Institute. https://www.media-diversity.org/womens-politicians-bearing-the-brunt-of-deepfake/

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1819. https://doi.org/10.15779/Z38RV0D15J

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. IEEE Signal Processing Magazine, 35(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202

Darnton, R. (2000). An early information society: News and the media in eighteenth century Paris. The American Historical Review, 105(1), 1–35. https://doi.org/10.1086/ahr/105.1.1

DataReportal. (2025). Digital 2025: Pakistan. We Are Social & Kepios. https://datareportal.com/reports/digital-2025-pakistan

Dawn. (2025, May 15). Fact check: Viral Daily Telegraph article praising PAF after India escalation is fake. Dawn News. https://www.dawn.com/news/1910765

De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. Frontiers in Public Health, 11, 1166120. https://doi.org/10.3389/FPUBH.2023.1166120

De Gaspari, F., Hitaj, D., Pagnotta, G., De Carli, L., & Mancini, L. V. (2022). Evading behavioral classifiers: A comprehensive analysis on evading ransomware detection techniques. Neural Computing and Applications, 34(14), 12077–12096. https://doi.org/10.1007/s00521-022-07096-6

Digital Rights Foundation. (2024). Platforms at the polls: Disinformation, political ads & accountability during the 2024 Pakistan general elections. https://digitalrightsfoundation.pk/wp-content/uploads/2024/12/Platforms-at-the-Polls.pdf

Digital Rights Foundation. (2025). DRF analyses Pakistan's National AI Policy 2025. https://digitalrightsfoundation.pk/wp-content/uploads/2025/08/DRF-Analyses-Pakistans-National-AI-Policy-2025-1.pdf

DW. (2025, January 28). Pakistan Senate votes to ban fake news on social media. DW. https://www.dw.com/en/pakistan-social-media-fake-news-law/a-71433026

Eppler, M., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. The Information Society, 20(5), 325–344. https://doi.org/10.1080/01972240490507974

European Commission, High-Level Expert Group on Fake News and Online Disinformation. (2018). A multi-dimensional approach to disinformation: Report of the independent High Level Group on fake news and online disinformation. Publications Office of the European Union. https://op.europa.eu/o/opportal-service/download-handler?format=pdf&identifier=6ef4df8b-4cea-11e8-be1d-01aa75ed71a1&language=en

Farooq, A., & de Vreese, C. (2025). Deciphering authenticity in the age of AI: How AI-generated disinformation images and AI detection tools influence judgements of authenticity. AI & Society, 1–12. https://doi.org/10.1007/S00146-025-02416-5

Farrukh, R. (2024). Woman politician in Pakistan's Balochistan province targeted with AI-generated clip. AFP Fact Check. https://factcheck.afp.com/doc.afp.com.36K79TY

Federal Communications Commission Consumer Advisory Committee. (2024). Report of the Consumer Advisory Committee: Uses of artificial intelligence for robocalls, robotexts, and other malicious calling practices. Federal Communications Commission. https://www.fcc.gov/sites/default/files/CAC_Final_Report_September2024.pdf

Full Fact. (2025, May 27). Video of Donald Trump talking about India and Pakistan is a deepfake. Full Fact. https://fullfact.org/conflict/donald-trump-india-pakistan-ceasefire-ai-deepfake/

Fusco, F. (2022). Artificial intelligence and fake news: Criminal aspects in Pakistan and Saudi Arabia. Pakistan Journal of Criminology, 14(4). https://www.pjcriminology.com/wp-

content/uploads/2023/05/2-Artificial-Intelligence-and-Fake-News.pdf

Gerbner, G., & Gross, L. (2000). Living with television: The violence profile. In J. R. Harris (Ed.), The fear of crime (pp. 169–196). Taylor & Francis. https://doi.org/10.4324/9781315086613-10/living-television-violence-profile-george-gerbner-larry-gross

Goodfellow, I. J., et al. (2014). Generative adversarial nets. In Advances in Neural Information Processing Systems 27. NeurIPS. https://proceedings.neurips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

Google Public Policy. (2024). An AI opportunity agenda for Pakistan. https://static.googleusercontent.com/media/publicpolicy.google/en//resources/pakistan_ai_opportunity_agenda_en.pdf

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.-F. Balcan, & H.-T. Lin (Eds.), Advances in Neural Information Processing Systems (Vol. 33, pp. 6840–6851). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

Hogan, L. (2024). Misinformation war rages online amid India-Pakistan tensions. ABC News. https://www.abc.net.au/news/2025-05-24/misinformation-online-war-kashmir-conflict-india-pakistan/105318696

Imran, A., Ali, M. I., & Rehman, A. (2025). The role of generative AI in undermining electoral integrity: A study on AI-driven election interference. Global Political Review, 10(1), 89–106. https://doi.org/10.31703/gpr.2025(x-i).08

Iram, A., & Khan, R. S. M. (2025). Misogynistic hate speech against female politicians on new media in Pakistan. Inverge Journal of Social Sciences, 4(1), 111–121. https://doi.org/10.63544/ijss.v4i1.137

Iram, S., Rafiq-uz-Zaman, M., & Malik, N. (2025). Advancing strategic decision-making through artificial intelligence: A critical review of methods, applications, barriers, and emerging trends. Contemporary Journal of Social Science Review, 3(3), 3025–3044. https://doi.org/10.63878/cjssr.v3i3.2289

Joseph, J., Jose, B., & Jose, J. (2025). The generative illusion: How ChatGPT-like AI tools could reinforce misinformation and mistrust in public health communication. Frontiers in Public Health, 13, 1683498. https://doi.org/10.3389/fpubh.2025.1683498

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux. https://psycnet.apa.org/record/2011-26535-000

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X., & Ratick, S. (1988). The social amplification of risk: A conceptual framework. Risk Analysis, 8(2), 177–187. https://doi.org/10.1111/J.1539-6924.1988.TB01168.X

Khan, A. (2024). Generative artificial intelligence and deterrence stability between India and Pakistan. Center for International Strategic Studies. https://ciss.org.pk/generative-artificial-intelligence-and-deterrence-stability-between-india-and-pakistan/

Knott, A., Pedreschi, D., Jitsuzumi, T., Leavy, S., Eyers, D., Chakraborti, T., Trotman, A., Sundareswaran, S., Baeza-Yates, R., Biecek, P., Weller, A., Teal, P. D., Basu, S., Haklidir, M., Morini, V., Russell, S., & Bengio, Y. (2024). AI content detection in the emerging information ecosystem: New obligations for media and tech companies. Ethics and Information Technology, 26(4), 1–14. https://doi.org/10.1007/S10676-024-09795-1

Kuźnicka-Błaszkowska, D., & Kostyuk, N. (2025). Emerging need to regulate deepfakes in international law: The Russo-Ukrainian war as an example. Journal of Cybersecurity, 11(1). https://doi.org/10.1093/cybsec/tyaf008

Łabuz, M., & Nehring, C. (2024). On the way to deep fake democracy? Deep fakes in election campaigns in 2023. European Political Science, 23(4), 454–473. https://doi.org/10.1057/s41304-024-00482-9

Li, F., & Yang, Y. (2024). Impact of artificial intelligence generated content labels on perceived accuracy, message credibility, and sharing intentions for misinformation: Web based, randomized, controlled experiment. JMIR Formative Research, 8(1), Article e60024. https://doi.org/10.2196/60024

López-Borrull, A., & Lopezosa, C. (2025). Mapping the impact of generative AI on disinformation: Insights from a scoping review. Publications, 13(3), 33. https://doi.org/10.3390/publications13030033

Mahmood, W., & Shahzad, M. (2024). Mediating role of continued influence of misinformation (CIM) between intolerance and user engagement on social media. Media and Communication Review, 4(1), 1–24. https://doi.org/10.32350/MCR.41.01

Malik, N., & Rafiq-uz-Zaman, M. (2025). The role of artificial intelligence in educational leadership and management for service quality in Pakistani higher education: A narrative review. Journal of Applied Linguistics and TESOL (JALT), 8(2), 1–18. https://doi.org/10.63878/jalt2080

Mazhar, S. (2025, May). When reality is manufactured: AI misinformation and the future of journalism in Pakistan. Digital Rights Foundation. https://digitalrightsfoundation.pk/when-reality-is-manufactured/

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys, 54(1). https://doi.org/10.1145/3425780

Momeni, M. (2025). Artificial intelligence and political deepfakes: Shaping citizen perceptions through misinformation. Journal of Creative Communications, 20(1), 41–56. https://doi.org/10.1177/09732586241277335

Nadeem, K., & Mumtaz, T. (2024). Manipulated minds: Political and social factors enabling deepfake impact in Pakistan. Regional Lens, 3(1), 145. https://doi.org/10.62997/rl.2024.31051

Nasir, S., Zia, A. Z., & Khalid, A. K. (2024). Mainstream media and social media news consumption patterns of youth in Lahore Pakistan. Journal of Media & Communication, 5(1), 16–39. https://doi.org/10.46745/ilma.jmc.2024.05.01.02

National Assembly of Pakistan. (2025). Prevention of Electronic Crimes (Amendment) Act, 2025. https://na.gov.pk/uploads/documents/679255ee36f45_595.pdf

O'Connor, C., & Weatherall, J. O. (2019). The misinformation age: How false beliefs spread. Yale University Press.

Outlook India. (2025, May 25). Fact check: Pak minister uses "fake" news report to praise country's Air Force in Senate; local media, PIB says "AI-generated". Outlook India. https://www.outlookindia.com/national/fact-check-pak-minister-uses-fake-news-report-to-praise-countrys-air-force-in-senate-local-media-pib-says-ai-generated

Papageorgiou, E., Chronis, C., Varlamis, I., & Himeur, Y. (2024). A survey on the use of large language models (LLMs) in fake news. Future Internet, 16(8), 298. https://doi.org/10.3390/fi16080298

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in Cognitive Sciences, 25(5), 388–402. https://doi.org/10.1016/J.TICS.2021.02.007

PTA. (2014). PTA (Re-constitution) Act, 1996. https://www.pta.gov.pk/assets/media/pta_act_consolidated_footnotes_11012022.pdf

Rafiq-uz-Zaman, M. (2023). The Impact of Digital Literacy on Students' Learning Outcomes: A Comprehensive Review. Inverge Journal of Social Sciences, 2(2), 194–205. https://doi.org/10.63544/ijss.v2i2.210

Rafiq-uz-Zaman, M. (2025). Between Adoption and Ambiguity: Navigating the AI Policy Vacuum in Pakistani Higher Education. Research Journal for Social Affairs, 3(6), 877–885. https://doi.org/10.71317/RJSA.003.06.0523

Rafiq-uz-Zaman, M., & Malik, N. (2025). The impact of social media use on students' behavioral health and academic outcomes: A narrative review (2016–2024). Journal of Applied Linguistics and TESOL (JALT), 8(1), 1–16. https://doi.org/10.63878/jalt2081

Rafiq-uz-Zaman, M., Köksal, O., & Akomodi, J. O. (2026). Empowering Generation Z: Digital Literacy, Innovation, and Well-Being in the Digital Learning Era. Inverge Journal of Social Sciences, 5(3), 106–124. https://doi.org/10.63544/ijss.v5i3.286

Rafiq-uz-Zaman, M., Malik, N., & Bano, S. (2025). Learning to Innovate: WhatsApp Groups as Grassroots Innovation Ecosystems Among Micro-Entrepreneurs in Emerging Markets. Journal of Asian Development Studies, 14(1), 1854–1862. https://doi.org/10.62345/jads.2025.14.1.47

Raza, M., Yaseen Moroojo, M., Abad, B., Javed Ali, P., & Shabbir, T. (2025). AI-driven journalism in

Pakistan: Legal constraints and future directions. ACADEMIA International Journal for Social Sciences, 4(2), 1929–1944. https://doi.org/10.63056/ACAD.004.02.0297

Rehman Uddin, M. H., & Noor ul Huda. (2025). Artificial intelligence and the fight against misinformation: Analyzing the role of digital media in shaping public perceptions in Pakistan. Advance Social Science Archive Journal. https://assajournal.com/index.php/36/article/view/720/1043

Schmitt, M., & Flechais, I. (2024). Digital deception: Generative artificial intelligence in social engineering and phishing. Artificial Intelligence Review, 57(12), Article 324. https://doi.org/10.1007/s10462-024-10973-2

Shahzadi, A., & Hussain, S. (2024). The role of culture mediating the social media in eliminating the gender inequality. Online Media and Society, 5(1), 75–84. https://doi.org/10.71016/oms/gx2ppb85

Shair, W., Waheed, A., Kamran, M. M., & Kubra, N. (2022). Digital divide in Pakistan: Barriers to ICT usage among the individuals of Pakistan. Journal of Economic Impact, 4(3), 196–204. https://doi.org/10.52223/jei4032206

Siraj, S. A., Gul, Q., & Inam. (2025). Towards an informed society: A critical analysis of media literacy in Pakistan. Journal of Social Sciences and Humanities, 33(1). https://ojs.aiou.edu.pk/index.php/jssh/article/view/2903

Toor, S. I., Zaidi, S. Z., Waheed, J., & Siddiqui, A. (2024). The public view on social media platforms: A case study of Pakistan. Pakistan Journal of Humanities and Social Sciences, 12(1), 1–15. https://doi.org/10.52131/pjhss.2024.v12i1.1898

Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. PLOS ONE, 18(10), e0291668. https://doi.org/10.1371/JOURNAL.PONE.0291668

Vinay, R., Spitale, G., Biller-Andorno, N., & Germani, F. (2025). Emotional prompting amplifies disinformation generation in AI large language models. Frontiers in Artificial Intelligence, 8, 1543603. https://doi.org/10.3389/frai.2025.1543603

Vincent, Z., Hashmat, S., & Shahwar, D. E. (2025). Impact of artificial intelligence (AI) on professional journalism in Pakistan. Online Media and Society, 6(2), 29–44. https://doi.org/10.71016/OMS/15153P68

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe. http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf

WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, & IFRC. (2020). Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. World Health Organization. https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation

Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). Defending against neural fake news. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, & R. Garnett (Eds.), Advances in Neural Information Processing Systems (Vol. 32, pp. 9051–9062). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf