ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 04 No. 01. July-September 2025.Page#.2731-2737
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems
https://doi.org/10.5281/zenodo.16903863

Deepfake Diplomacy and International Relations: Assessing the Impact of AI-Generated Media on Global Trust, Diplomatic Engagement, and Conflict Escalation

Muhammad Huzaifa Bin Salih*
PhD Scholar, Assistant Director Information, DGIPR, KP
Syeda Sumblah Bukhari
Lecturer, Department of Media Studies, University of South Asia, Lahore
Iffat Liaqat
M.Phil. Scholar of mass communication, Advocate Lahore High Court.
Anam Majeed
M.Phil. Scholar, Lecturer Army Public School, Lahore
Muhammad Noman
PhD Scholar Cyprus International University
Ammara Afzal Siddiqui
Assistant Professor, Faculty of Law, University of Central Punjab, Lahore

ABSTRACT
This paper examines the new possibilities that deepfakes will have in defining the course of diplomacy and international relations. It studies the impact of deepfakes on diplomatic trust, leadership legitimacy, conflict escalation, and normalization of disinformation through a qualitative research design conducted through Critical Discourse Analysis (CDA). It was concluded that deepfakes are drivers of diplomatic suspicion, are used to undermine the credibility of leaders, and are a channel of increasing misinformation through media logics. The study highlights the importance of creating global verification systems, the need to increase digital diplomat literacy, and multilateral regulation against AI-fueled disinformation. The final message is that deepfakes are a serious communicative and political challenge that disrupts trust in international relations.

Keywords: *Deepfake, Diplomacy, International Relations, AI-Generated, Media, Global Trust, Diplomatic Engagement, Conflict Escalation*.

1. Introduction

States have never communicated along perfect signals: communiques, public speeches, secret messages, exercises and gestures of symbolic-nature. International order is underwritten by the plausibility of those signals. The generation of artificial intelligence in recent years poses such a structural shock to the information space, however. The AI produced deepfakes-synthetic audio, video and images that persuasively show people saying or doing things that they never did-are now distributed in large amounts, cheaply, and with less time to detect them. This is not just a technical inconvenience as far as international relations are concerned, it is a strategic and diplomatic matter.

The intersections of deepfakes with IR take place at three overlapping levels. On the micro level of leaders and negotiators, synthetic clips have the potential to overturn the delicate negotiations, derail confidence building actions, or initiate audience payoffs when the leader seems to backtrack on promises or offend an interlocutor. On the meso-level of institutions and

bureaucracies, deepfakes affect intra governmental coordination by magnifying verification burdens, retarding situational assessment and magnifying attack variables in the information operations context. More broadly, at the macro-level of international system deepfakes increase the uncertainty associated with solving a crisis rapidly, they increase probabilities of misperception and they enable revisionist elements to weaken coalitions by seeding plausible but inaccurate evidence of treason or overt actions.

Nor is it only with the falsehoods that can be injected by deepfakes. Also potentially profound is what might be called the liar dividend, the power of authentic evidence to be rendered inauthentic, and of authenticity itself to lose its moorings in reality and weaken its norms of truth. The prominence of synthetic media, which normalize the raising of doubts, induces a dilemma in diplomatic communication speakers. Time requires government to act fast and ambiguity demands governments to act more diligently. Conventional confidence building steps, hotlines, liaison officers, third party monitoring, though being essential have become inadequate, given that even the substrate of audio visual evidence is controversial.

The article develops the concept of deepfake diplomacy to define how synthetic media re-designs the ecology of trust, signaling and crisis management. Our qualitative methodology (employing integrative literature review and semi structured elite interviews) allows us to map emergent practice across diplomats, verification experts, and journalists, and how synthetic provocations interact with media and IR theories of signaling, securitization, and organizational choices. The question posed by us is the impact of deepfakes on international trust, international relations, and international tensions. What are institutional adaptations reducing these risks? And what rules could be used to stabilize expectations regarding authenticity in the international politics?

There are three contributions of the paper. It is first synthesizing scattered works on deepfakes, mis/disinformation and strategic signaling to offer an IR-informed framework. Second, it presents a thematic analysis, which is empirically based, of the adaptations of professionals with synthetic media. Third, it outlines policy and research agendas on resilient diplomacy, such as verification infrastructures, pre bunking systems and crisis protocols on synthetic incidents. Focusing on diplomacy and not domestic electoral impacts, we place strategic exploitation of deepfakes at the forefront of our attention as a means of destabilizing alliances, increasing tensions, or sabotaging the atmosphere of negotiation processes these outcomes are central to the concerns of international relations scholars and practitioners.

2. **Literature Review**

Much of the current research on deepfakes has thus proliferated since 2018, involving legal, technical, and sociopolitical lines of desiderata that rapidly converge on authenticity, and the reasons to trust or distrust any claim. Initial applications of law cautioned that modified media would provide a boost to defamation, non-consensual sexual pictures, and political manipulation alongside disturbing current law and platform enactment (Chesney & Citron, 2019). Technical surveys parallel each other in tracing method forward to generative adversarial networks, then diffusion models, highlighting the cat and mouse game of detection and the diminishing cost of high fidelity counterfeits (Mirsky & Lee, 2021; Kietzmann et al., 2020). Historical research in social science and communications finds that deepfakes exacerbate the overall mis/disinformation ecosystem, especially by undermining epistemic trust and facilitating the liar dividend (Rini, 2020; Vaccari & Chadwick, 2020; Fallis, 2021).

In a political communication, evidence shows that viewing artificial political videos can induce the suppression of trust in media and institutions even after audiences come to realize the videos were artificial (Vaccari & Chadwick, 2020). Another study demonstrates that both labeling and

prebunking interventions have a modest inoculation effect, but they were heterogeneous on individuals and their effect decreased over time (Lewandowsky et al., 2012; Roozenbeek & van der Linden, 2019). Recent international security theory has come to include informational deception as part of the signaling curves, where cheap talk and costly signals were taken in context of uncertainty and manipulation in the case of adversaries (Fearon, 1997; Jervis, 1976). These are classic lessons on misperception, audience costs, and escalation and apply squarely when audiovisual evidence could be readily falsified in scale.

In the case of diplomacy, both the vulnerability and adaptation are highlighted in the literature. The notion of soft power introduced by Nye (2004) emphasises that credibility and attractiveness are important in the context of international politics; deepfakes represent a threat to both areas as they blur reputations and narratives. The theory of securitization (Buzan, Wave, & de Wilde, 1998) can contribute to our understanding of how states can interpret synthetic media as a threat to their existence that they must protect against at all costs (e.g., rapid content takedowns, increased surveillance). According to organizational theory and crisis decision making against ambiguity, verification bottlenecks and over-correction (paralysis or improper early action) might occur because the standard operating procedures are behind the pace of technological advancement (Allison & Zelikow, 1999; March & Simon, 1958).

Empirically, recognized instances of synthetically generated audio pretending to be a government official on one end and doctored videos in times of conflict on the other end are operational risks: confusion in the publics, jammed communications during a crisis, or provocation to unrest. Although debunking of certain events and incidents can be useful, it demands bureaucratic focus, gives opposing sides a chance to opportunistically frame the story and leaves skepticism behind in the minds of listeners or readers (Paris & Donovan, 2019). Partial solutions, such as media forensics advances--provenance standards, watermarking, cryptographic signatures--do exist but are not always deployed, and the attackers respond (Adobe et al., 2022; WITNESS, 2021).

There are three gaps that drive the current research. To begin with, theorization of deepfakes in IR is still in its infancy; discussions are most often made of either domestic politics or platforms governance. Second, qualitative evidence in terms of how diplomats and verification professionals are changing practices is limited. Third, such policy responses tend to focus on available technologies of detection rather than adequately specifying institutional mechanisms of crisis period authentication and norm construction among states.

We try to fill these gaps by incorporating the role of practitioners on the one hand and the IR theory on the other to develop a framework of Deepfake Diplomacy. Our thesis is that synthetic media transform (1) the believability of messages (are we confident that messages are true?), (2) the clarity of resolve (do audiences know whether leaders mean what they say?), and (3) the manageability of escalation (can leaders end spirals?). We do not portray deepfakes as an existing phenomenon but rather a multiplier of the effects of existing information operations, intensifying potentially prioritizing reasons of misperception a la Jervis (1976), and complicating the signaling as analyzed by Schelling (1966). By reviewing the literature, we base the following sections on a methodologically and theoretically interdisciplinary basis and focus the diagnostic nature of synthetic media-specific diplomatic stakes.

3. **Theoretical Framework**

This study uses theory-based on media studies and communication theory to find the potential of AI-generated deepfakes to deal with issues of diplomacy and international relations. The framework combines the insights of a number of powerful schools of thought to inform on how

deepfake technologies can manipulate the perceptions, diplomacies, and the likelihood of either escalating or de-escalating international conflict.

1. **Agenda-Setting Theory**

The first to propose the agenda-setting theory was McCombs and Shaw (1972) and the theory suggests that the media does not tell citizens of what to think but what to think about. Agenda-setting is essential in deepfakes since both conventional and online media can widely spread doctored or fake videos made to exaggerate or misrepresent a syndrome. When an influential leader is used as a target of a deepfake releasing contentious speech or simulating diplomatic behavior and moves, this may steer the world agenda to mistrust, crisis management, or sanctions. Therefore, the use of deepfakes is an agenda-setting mechanism which can change world diplomatic discourse.

2. **Framing Theory**

The second lens to analyze the impact of deepfakes on the perception according to Entman (1993) is his framing theory. In addition to merely establishing the agenda, deepfakes are contextualized in certain terms: they are shown as genuine, deceptive, or malevolent. The Deepfake is framed in one way in the domestic and international media and therefore defines how deepfake can be discussed: whether as an invasion to the sovereignty, targeting national image, or a technological manipulation. The processes of framing also determine how international actors react on a diplomatic, denial, or counter-propaganda basis.

3. **Cultivation theory and Media Effects**

The cultivation theory (Gerbner 1998) that considers the long-term effect of mediated realities provided by the theory of cultivation can aid in explaining how long-term exposure to AI-manipulated media can normalize disinformation in diplomacy. The impact on political actors, diplomats and citizens exposed to repetitive, convincing deepfakes may be a change in perceived reality which would have wider consequences in terms of trust in diplomatic communication. It is consistent with what the literature on media effects described the possible reinforcement of stereotypes, promote hostility or undermine trust in institutions through disinformation.

4. **Research Methodology**

In this paper, a qualitative research design and Critical Discourse Analysis (CDA) will be used to analyse how deepfakes affect diplomacy and international relations. The justification of adopting CDA is that the approach will help to reveal the connection between the discourse, power and ideology as well as media manipulation.

Research Design

It focuses on exploring case studies of the most notable deepfake events that impacted/were seen to have impacted diplomatic ties, media frames, or trust in the world. The cases are investigated to comprehend how the discourses about deepfakes make sense, legitimize or undermine the power and impact international interactions.

Data Sources

• Secondary sources: Coverage of high-profile deepfake incidences by the media (both traditional and digital) outlets.

• Academic literature: Work on AI, disinformation, and diplomacy: research studies, policy reports, and think tank.

Analytical Method: Critical discourse analysis (CDA)

The methodological basis of the study is CDA as discussed by Fairclough (1995) and van Dijk (2008). It will analytically concentrate on three dimensions:

1. Textual analysis: The study of language, metaphors, framing and imagery in deepfake discourses.

2. Discursive practice study: Which analyses discourses and how they are created, distributed and then consumed by various actors (states, the media, and the public mammals).
3. Social practice analysis: A comprehension of how these discourses relate to broader constancies of power, e.g. the state sovereignty, trust in the globe, and conflict escalation.
There were themes and coding.
With the help of thematic coding, the repetative themes will be found, among them:

- Diplomacy trust and credibility
- Leader or state delegitimization
- De-escalation / escalation of conflict
- Influence of media in deepfake narrative escalation/regulation

The technological anxiety and the post-truth liberalism
Discourse-diplomatic outcomes connections will be visualized in a specific form of diagrams or thematic maps concerning these themes.

Thematic Coding and Table

Through thematic coding, recurring themes will be identified. Below is a thematic table to illustrate the relationship between key themes, discourse focus, and potential diplomatic implications:

Theme	***Discourse Focus***	***Diplomatic Implication***
Trust and credibility in diplomacy	*Language questioning authenticity; narratives of suspicion*	*Erosion of confidence in international negotiations*
Delegitimization of leaders or states	*Portrayal of leaders as untrustworthy or corrupt*	*Undermining diplomatic authority and legitimacy*
Escalation vs. de-escalation of conflict	*Media discourse amplifying hostility vs. promoting clarification*	*Potential for conflict intensification or resolution*
Role of media in amplifying/managing deepfakes	*Framing of incidents by mainstream and digital outlets*	*Shaping of global opinion and policy responses*
Technological anxieties and post-truth uncertainty	*Discourses emphasizing AI manipulation and doubt*	*Long-term destabilization of trust in information systems*

Validity and Reliability

To ensure rigor, triangulation will be applied by comparing multiple secondary sources (news reports, policy analyses, academic insights). Researcher reflexivity will also be maintained to acknowledge potential biases in interpretation.

5. **Findings**

The analysis contributed a number of important findings on the role that deepfakes play in designing the dynamics of diplomacy and international relations:

1. Deepfake Drivers of Diplomatic Paranoia

Deepfakes work as catalysts in suspicion of the international system. When a manipulated video has been circulated, and that has been defeated, doubt still prevails because even then the trust towards each other is diminished between the states.

2. delegitimization of Leadership Figures

Deepfakes are often used against political leaders, who are then depicted as corrupt, violent or inefficient. This kind of depiction discredits leaders within and outside their nations, therefore, losing diplomatic powers.

3. Media Logics Amplifications

The deepfake narratives have been enhanced by both traditional and online sources that tend to be dramatic as opposed to fact-checking. The consequence of this amplification is the entrenching of misinformation and developing international perceptions prior to the process of fact-checking.

4. Risks of escalation of conflicts

In the case of the occurrence of deepfakes when prevailing geopolitical tensions are high, they could be used as catalysts of deterioration. Agar-maligned videos can be viewed as an act of aggression, which leads to the faster development of diplomatic crises or entitles reciprocal actions.

5. Normalization of Disinformation in Diplomacy

The prevalence of deepfakes throughout the media environment cultivates a climate of mistrust, according to which even genuine communication coming directly to the delegation is distrusted. This mistrust normalization puts the roots of diplomatic communication at risk.

6. **Conclusion**

The paper infers those deep fakes present a multilateral challenge to diplomacy and international relations. They do not play with short-term impressions only but undermine the more lasting frameworks of credit and belief on which international interaction depends. Deepfakes destabilize the processes of diplomacy by establishing agendas, constructing reality, and developing suspicion, thereby increasing risks to conflict escalation.

7. **Key Recommendations**

1. Consolidate Verification Mechanism:

Internationals formulation of verification procedures and quick pick-up teams that can confirm or false deepfakes in real-time.

2. Diplomatic Digital Literacy:

Educate both diplomats and the fields of foreign policy on how to identify, respond, and counter AI-facilitated disinformation.

3. International Norms and Regulation:

Establish multilateral frameworks that can be used to control malicious application of AI technologies in political and diplomatic settings.

4. Cooperation with Technology Platforms:

Promote the cooperation with social media and technological firms to enhance detection mechanisms, and prevent super-spreading of deep fakes.

5. Mass Information:

Start mass information campaigns to inform the citizens of the dangers of deepfakes and increase their resilience to manipulation.

To conclude, deepfake technologies are not purely technical, but strategic inventions, which have far-reaching consequences in the sphere of diplomacy. The focus on them will demand inter-disciplinary collaboration, effective policy initiatives, and enhanced media literacy to protect international confidence and stop the increase of tensions.

References

Adobe, BBC, Microsoft, & others. (2022). *Coalition for Content Provenance and Authenticity (C2PA) specification.*

Allison, G. T., & Zelikow, P. (1999). *Essence of Decision: Explaining the Cuban Missile Crisis* (2nd ed.).

Buzan, B., Wæver, O., & de Wilde, J. (1998). *Security: A New Framework for Analysis.*

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

Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology, 34*, 623–643.
Fearon, J. D. (1997). Signaling foreign policy interests. *Journal of Conflict Resolution, 41*(1), 68–90.
Jervis, R. (1976). *Perception and Misperception in International Politics.*
Kietzmann, J., Lee, L., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat