Ianna Journal of Interdisciplinary Studies, Volume 6 Number 2, August 2024 DOI: https://doi.org/10.5281/zenodo.11393419

The Intersection of Artificial Intelligence, Deepfake, and the Politics of International Diplomacy

Ikenga Francis Ayaegbumam

ORCID: https://orcid.org/0000-0002-7002-6930

***Nwador Amaechi Fidelis**

ORCID ID: https://orcid.org/0009-0000-4127-820X

Department of Political Science, Faculty of the Social Sciences, Delta State University, Abraka *Corresponding author's Email: nwador.amaechi@delsu.edu.ng

Abstract

Background of the study: The integration of artificial intelligence (AI) into communication systems has sparked interest across various sectors, including marketing, journalism, and diplomacy. While AI presents opportunities for enhancing information analysis and dissemination, concerns about job displacement and the spread of disinformation via deepfake technology have arisen with deep seated concern on the consequence of its ,misuse in the global system.

Objective of the study: This study employs a rigorous review of existing literature to examine the intersection of AI, deepfake technology, and international diplomacy. It explores how AI can enhance diplomatic processes while also examining the extent to which it can exacerbate geopolitical tensions and security risks.

Methodology: The study utilizes a systematic review of available secondary literature to analyze the impact of AI and deepfake technology on international diplomacy. It assesses existing research, case studies, and expert opinions to provide insights into the potential benefits and risks associated with these technologies in diplomatic contexts.

Results and findings: The findings of the study indicate that while AI offers promising opportunities for improving diplomatic processes, the proliferation of deepfake technology poses significant challenges. AI can facilitate information analysis and enhance decision-making in diplomacy, but the misuse of deepfake technology threatens to undermine trust and credibility in diplomatic engagements. The study highlights the importance of media literacy initiatives in countering the influence of deepfake propaganda on political perceptions and emphasizes the need for technological advancements in content detection tools and comprehensive regulatory frameworks to address these challenges effectively.

Conclusion: In conclusion, the study underscores the importance of transparency, accountability, and cooperation in the use of AI technologies in diplomacy. It emphasizes the need for enhanced international cooperation to combat cross-border threats posed by malicious actors leveraging deepfake technology. By promoting trust and confidence in diplomatic engagements and mitigating the risks associated with deepfake manipulation, policymakers can harness the potential of AI to foster effective diplomacy in the digital age.

Unique Contribution: The study contributes significantly to the knowledge of how AI can enhance diplomatic processes, specifically through improving information analysis and decision-making as well as providing a comprehensive review of the potential benefits with highlights on how it inseminates integrity issues and political tensions.

Key recommendations: The study recommends several measures to address the challenges posed by AI and deepfake technology in international diplomacy. These include technological advancements in content detection tools, comprehensive regulatory frameworks, and enhanced international cooperation to combat cross-border threats. Additionally, the study emphasizes the importance of media literacy initiatives to empower individuals to discern between authentic and manipulated content, thereby reducing the influence of deepfake propaganda on political perceptions. By implementing these recommendations, policymakers can safeguard the integrity of diplomatic processes and mitigate the risks associated with the misuse of AI technologies.

Keywords: Artificial Intelligence; Deepfake; Politics; International diplomacy;

Introduction

The integration of artificial intelligence (AI) into communications and its potential impact on various sectors, such as marketing, journalism, and propaganda, is a topic of growing interest (Smith, 2023). AI offers opportunities to enhance the speed, cost-effectiveness, and efficacy of information analysis and integration (Johnson & Lee, 2021). For instance, personalised online news platforms utilise AI to tailor content based on individual preferences and interests (Garcia et al., 2022). Furthermore, AI can be utilised in training communication practitioners through expert systems, providing insights into best practices in the field (Chen & Wang, 2020). However, there is also speculation regarding the displacement of human communicators by AI-driven automation, leading to concerns about job loss within the communications industry (Brown et al., 2022). In the realm of international affairs, the convergence of AI and deepfake technology has become increasingly significant (Jones & Smith, 2023). Deepfake technology, which employs AI to manipulate audio and video recordings, has potential implications for global politics (Kim & Park, 2021). This raises concerns about the spread of disinformation and its impact on diplomacy, as well as international relations (Gupta & Sharma, 2023).

The Springfield College Artificial Intelligence in International Relations project, initiated in 1989, marked the first recorded use of AI in diplomacy (Jones et al., 2020). Despite initial optimism regarding the potential of AI to enhance diplomatic decision-making, subsequent projects have focused less on AI development dedicated to the analysis of foreign policy and diplomacy processes (Johnson et al., 2022). The growing scepticism about the use of AI and deepfake has also been expressed in international conflict, as times have shown that, in times of crisis, manipulated media content that spreads misinformation can potentially contribute to the escalation of events that lead to violence. The importance of this study is underscored by the increasing application of AI and deepfake media portrayals across various institutional, social, and political contexts, making an in-depth study of the subject crucial for scholarly analysis. The study identifies and elaborates on the significant challenges posed by deepfake technology in the context of international diplomacy and highlights how deepfakes can undermine trust and credibility, thereby exacerbating geopolitical tensions and security risks. This study is particularly crucial for understanding the dual-edged nature of AI in diplomatic contexts. Thus, this study, while adopting a rigorous review of existing secondary literature as the basic methodology, is, therefore, an attempt at examining the intersection of AI, deepfake technology, and international diplomacy as the concept seems exigent as it presents both opportunities and challenges which makes scholarly research on its application to diplomatic intercourse and global politics an imperative in order to advance possible roadmap in international relations, the media and human relations.

Literature Review:

Artificial Intelligence

Artificial intelligence (AI) refers to the development of computer systems that can perform tasks that typically require human intelligence. These tasks include understanding natural language, recognising patterns, learning from experience, and making decisions. AI encompasses various techniques and approaches, including machine learning, neural networks, natural language processing, and robotics, among others**.** It refers to the capacity of individuals to generate content in various forms, such as pictures, videos, voice, or texts, through computer operations. Duggal (2023) defines AI in a professional context as the process of imbuing computers, robots, or software with the ability to think intelligently akin to the human mind. Frankenfield (2023) similarly characterises AI as the emulation of human intelligence through software-encoded heuristics, emphasising its capability to rationalise and execute actions conducive to achieving specific objectives. A subset of AI known as machine learning posits that computer programs can autonomously learn from and adapt to new data without human intervention. Frankenfield (2023) further asserts that deep learning techniques facilitate this autonomous learning by assimilating extensive volumes of unstructured data, such as text, images, or video. Exemplifying this is figure 1 below which is a screenshot of an artificial intelligence transfer of facial expressions and mannerisms from one video to another where a team of researchers at Carnegie Mellon grafted some of Martin Luther King Jr's facial movements onto former President Barack Obama.

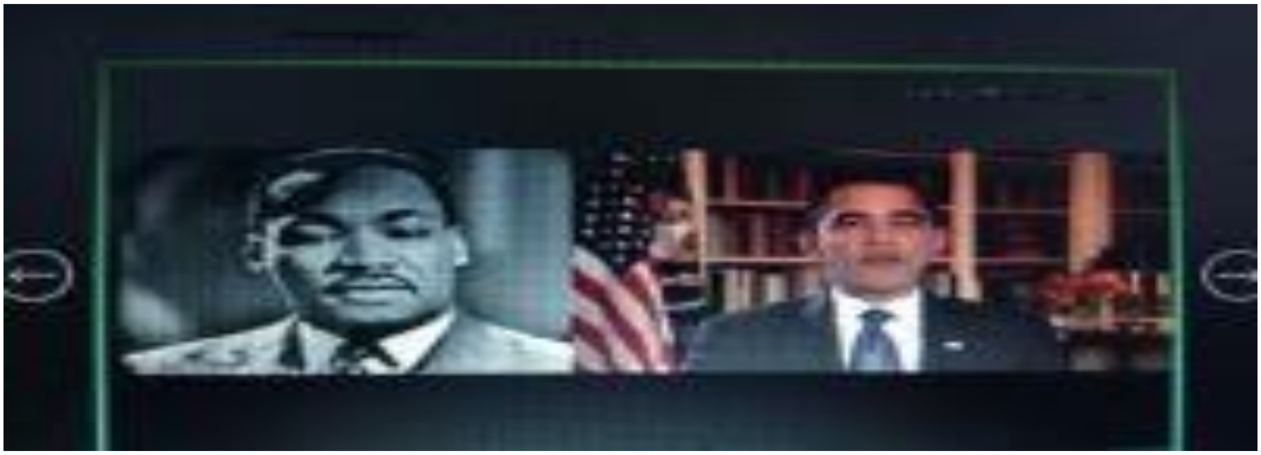

Figure 1: Screenshot of an Artificial Intelligence video creation. Source: https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

Deepfake

Deepfake technology refers to the use of artificial intelligence (AI) and deep learning techniques to create or manipulate digital content, typically videos or audio recordings, to depict events or statements that did not occur or were not said by the individuals portrayed. That means it is used to create convincing images, audio and video hoaxes (Ekpang, 2023). Deepfake algorithms analyze and synthesize vast amounts of data to generate highly realistic media, often indistinguishable from authentic recordings. The proliferation of Deepfake technology has profound implications for society, including its impact on democracy, trust, security, and international diplomacy. An example is provided in Figure 2 below of a screenshot of a deepfake video of Boris Johnson while he was British Prime Minister.

Figure 2: A screenshot of a deepfake image of Boris Johnson as Prime Minister of UK. Source: Appel & Fabian Prietzel, (2022)

Deepfake: Historical development

Deepfake technology emerged around the mid-2010s, initially gaining attention for its potential to create highly realistic manipulated videos and audio. Academic research and media coverage began to highlight the capabilities of deep learning algorithms in synthesising human faces and voices to produce convincing fake content (Huang et al., 2018). In late May of 2019, a video depicting Nancy Pelosi, manipulated to portray her as intoxicated and slurring her words, circulated widely on the internet, garnering over 2.5 million views on Facebook within days and being shared by notable political figures (Mervosh, 2019). This incident highlighted a form of disinformation poised to disrupt political discourse and elections: the deliberate alteration of audiovisual content, amplified through social media channels. The emergence of deepfakes into public consciousness began approximately a year prior to the Pelosi video, when a prominent technology blog featured an alarming headline: “We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now” (Cole, 2018 p.1). Subsequently, an anonymous user uploaded pornographic videos on Reddit, purportedly featuring well-known actresses like Gal Gadot and Maisie Williams. However, these videos were created using artificial intelligence (AI) to overlay the actresses' faces onto adult film actors' bodies. Reddit promptly banned such deepfake porn content, but the technique quickly spread across the internet, facilitated by freely available deepfake software and numerous instructional videos on platforms like YouTube. Deepfakes gained particular notoriety for their potential to influence elections and political processes as media outlets reported on instances of deepfake videos targeting political figures and spreading false narratives, prompting calls for regulation and countermeasures (Wang et al., 2019).

Technologically, deepfakes stem from a specific form of deep learning known as generative adversarial networks (GANs) (Goodfellow et al., 2014). Deep learning, a subset of artificial intelligence, employs neural networks to deduce rules and replicate patterns by analysing extensive datasets. Within GANs, one algorithm, the generator, crafts content based on source material, while another, the discriminator, scrutinises the forgery for flaws. Through iterative refinement, GANs produce increasingly realistic yet synthetic video content. Deepfake media primarily fall into three categories: face-swap, lip sync, and real-time facial reenactment, also termed puppet-master, wherein expressions from a source person are mapped onto a target person (Kim et al., 2018; Suwajanakorn et al., 2017; Thies et al., 2016). This method may incorporate the voice of an impersonator or a synthesised voice mimicking the target individual (Diakopoulos & Johnson, 2021). Post-processing techniques further enhance the realism of these videos, rendering them nearly indistinguishable from genuine recordings, even with sophisticated forensic analysis (Agarwal et al., 2019; Gu¨ era & Delp, 2018; Korshunov & Marcel, 2019).

The proliferation of deepfakes coincides with a precarious era where social networking platforms like Facebook, WhatsApp, and Telegram allow instantaneous dissemination of unchecked content. Research indicates a propensity among users to share negative and novel information, with false political narratives being particularly virulent (Vosoughi et al., 2018). With their unprecedented realism, speed, scalability, and capacity for personalisation, deepfakes exacerbate the broader issue of fake news on social media, defined as "fabricated information that mimics news media content in form but not in organisational process or intent" (Lazer et al., 2018, p. 1094; see also Egelhofer & Lecheler, 2019; Tandoc et al., 2018).

AI, Deepfake, and International Diplomacy: The intersection

Now, with the recent interest peaks of AI and its manuals of what is right and wrong, we can consider AI in diplomacy. With such consideration, AI in diplomacy has many advantages and disadvantages. At an even more extreme level, AI has been used to automate the process of warfare. A prime example of this is a rudimentary simulator that was programmed and used by the Argentine forces during the Falklands War, in an attempt to foresee tactical advantages. It is well known that war simulations have come a long way since this. This is not to say that all AI is designed to perform positive actions. Any code that has a goal or purpose can be considered AI, and this extends to malicious code. A hacker or developer of a virus programme has goals that are achieved by the code he has written. The code must learn how to circumvent preventions in order to achieve its goals. For example, a hacker may develop AI that will attempt to learn an internet user's habits in order to disguise itself as an internet cookie.

AI has been and still is a concept that is used in the field of video game programming in order to provide non-playable characters with decision-making abilities, in order to create the illusion that the NPC given has its own personality and character. This is a basic example of AI, but the behaviour shown in such programmes is exhibited by more complex AI. It is merely defined as a complex concept which learns and adapts to new situations and circumstances in order to complete goals that have been set out for it. This sets AI apart from conditional programmes which only execute operations when specific conditions are met. Historically, AI has been a concept that has been used in many different ways, be it in a positive form or quite the opposite. The utilisation of digital platforms in diplomatic affairs, such as negotiations, networking, and information exchange, has spawned a new area of study known as Digital Diplomacy. This digital transformation of diplomacy occurs across various fronts. Ministries and politicians now establish their presence on social media, where they articulate their stances on specific issues, eliminating the need to wait for official briefings from Foreign Ministries (Jafarova, 2023). Diplomats frequently convey their positions online, representing a quasi-official approach that carries the risk of posts being permanently deleted or attributed to hacking incidents, reflecting the realities of modern communication.

Recent discussions in the media have centred around the role of ChatGPT in the future landscape of diplomacy. Diplomats can leverage AI to automate tasks such as drafting press releases, enabling the swift distribution of prepared content across multiple platforms with a single click. This adaptation is crucial in an era where individuals primarily consume information via smartphones and the internet. However, complete automation also poses inherent risks. While AI can assist in generating ideas, concerns linger regarding the confidentiality of information processing. Reports of data leaks involving ChatGPT have surfaced, raising questions about the security of sensitive diplomatic or personal documents shared on the platform (Derico, 2023; Gurman, 2023). Additionally, diplomatic language demands a high level of sensitivity in word choice and expression, an aspect that AI-generated text may overlook despite its grammatical correctness.

Given the seeming resurgence of New Cold War rivalry in the international system, the rise of deepfake technology has become a growing concern in international diplomacy as disinformation campaigns have gained wide usage in information warfare. Information warfare encompasses both disrupting enemy communication functions, now part of cyber operations, and manipulating information for deceptive purposes, also known as psychological warfare. The latter has gained momentum with the ability to manipulate audio and visual imagery, allowing for the creation and dissemination of deepfake videos on social media platforms. In the United States, psychological operations are integral to influencing enemy decision-making through various information-related capabilities, including disinformation. China has institutionalised this approach with the "Three Warfares" concept, focusing on public opinion, psychological, and legal warfare (Hitoshi, 2022b). Psychological warfare in China aims to erode enemy morale through rumours, false narratives, harassment, and threats. Russia is also notorious for aggressive disinformation campaigns, leveraging generative adversarial networks and artificial intelligence to create convincing visuals and spread misinformation. Their reported usage of deepfake technology as a pretext for invasion of Ukraine aligns with their military strategy to dominate the information landscape (Hitoshi, 2022a).

Table 1: Empirical Review

TITLE	Authors	Year of Publication	Methodology	Results	Findings	Recommendations
Deepfake: The current	Hany Farid	2021	Comprehensive	Identification of	Deepfake technology	Implement robust

state of AI-generated audio and video manipulation			overview and analysis of deepfake technology, focusing on audio and video manipulation	the potential risks posed by deepfakes in the context of elections and international propaganda	poses significant challenges to electoral integrity and international security	detection and mitigation strategies to counter the spread of deepfake-based disinformation.
Deepfake detection: Current challenges and future directions	Y. Zhou et al.	2021	Review of existing techniques for detecting deepfakes and analysis of challenges in developing effective detection methods.	Highlighting the importance of robust detection systems in mitigating the spread of fake information during elections and in international propaganda efforts.	Continuous research and development are needed to address the evolving threat of deepfake-based disinformation campaigns.	Invest in interdisciplinary research efforts to develop robust detection methods capable of addressing the evolving threat landscape.
Detecting deepfake videos in the wild using multimodal analysis	N. Diakopoulos & L. Friedland	2021	Proposal of a multimodal approach for detecting deepfake videos in real-world scenarios, combining analysis of	Improved detection accuracy crucial for combating the use of deepfakes in election	Multimodal analysis enhances detection accuracy, providing a promising avenue for countering the spread of deepfake-	Further research and development are needed to refine multimodal analysis techniques and improve detection performance.

			visual and auditory features.	campaigns and propaganda.	based disinformation.	
Deepfake detection: A review	Z. Jin et al.	2020	Review of state-of-the-art deepfake detection techniques, discussion of their strengths and limitations.	Emphasis on the need for continuous research and development to address the evolving threat of deepfake-based disinformation campaigns	Current detection methods have limitations and require ongoing refinement to keep pace with advances in deepfake technology	Collaborate across disciplines to develop and implement effective countermeasures against deepfake-based disinformation campaigns.
FaceForensics++: Learning to detect manipulated facial images	A. Rossler et al.	2019	Presentation of FaceForensics++, a dataset and benchmark for evaluating deepfake detection methods.	Contribution to the development of more robust detection techniques essential for countering the use of deepfakes in election-related misinformation and international propagan	FaceForensics++ provides a valuable resource for evaluating and improving the performance of deepfake detection methods.	Utilize FaceForensics++ dataset to benchmark and refine deepfake detection algorithms

				da efforts.		
Deepfake detection: A survey	H. Nguyen & T. D. Bui	2020	Conducting a survey of existing deepfake detection methods, categorization based on underlying techniques, and evaluation of performance.	Providing insights into strengths and weaknesses of current detection approaches, informing future research directions.	Current detection methods have limitations and require ongoing research and development to address evolving threat of deepfake-based disinformation	Collaborate across disciplines to develop and implement effective countermeasures against deepfake-based disinformation campaigns.
Deep fakes: A looming challenge for privacy, democracy, and national security	R. Chesney & D. K. Citron	2019	Discussion of multifaceted implications of deepfake technology for privacy, democracy, and national security.	Highlighting the potential use of deepfakes in influencing electoral outcomes and shaping international narratives.	Deepfake technology poses significant risks to privacy, democracy, and national security, requiring regulatory and technological responses to mitigate these risks.	Implement regulatory and technological measures to address the risks posed by deepfake technology to privacy, democracy, and national security.

Source: Authors, 2024

Theoretical Consideration: Cultivation theory and technological determinism

Cultivation theory, developed by George Gerbner, posits that prolonged exposure to media content shapes individuals' perceptions of social reality. Gerbner, a prominent communication scholar, introduced the Cultivation Theory in the late 1960s. He focused on the long-term effects of television viewing on individuals' perceptions of the world, arguing that heavy exposure to television content cultivates shared beliefs and attitudes among viewers (Gerbner, 1969).

Cultivation theory in Political contexts:

Shaping Political Perceptions: In the political context, cultivation theory suggests that exposure to media, including news coverage and political advertisements, can influence individuals' perceptions of political issues, candidates, and parties (Morgan & Shanahan, 2010). Through repeated exposure to certain messages, individuals may adopt attitudes and beliefs consistent with the dominant themes presented in the media. This is also because media coverage plays a crucial role in constructing political reality for the public. Deepfake portrayals, which involve the manipulation of audiovisual content to depict individuals saying or doing things they never did, can distort political reality and shape public perceptions (Shu, et al. 2020). If widely disseminated, deepfake videos can potentially influence public opinion and undermine trust in political institutions. Cultivation theory further suggests that media exposure can influence political behaviour, including voting decisions and participation in political activities (Morgan & Shanahan, 2010). Deepfake portrayals that manipulate political figures or disseminate false information can impact voter perceptions and contribute to polarisation and mistrust in the political process (Boidman, 2020). Deepfake technology enables the creation of highly realistic yet fabricated audiovisual content, which can be used to spread misinformation or manipulate public opinion (Shu et al., 2020). Political actors may use deepfake portrayals to discredit opponents, spread false narratives, or incite social unrest, thereby influencing political outcomes.

The proliferation of deepfake portrayals in the political sphere can erode trust in media, political institutions, and public figures (Boidman, 2020). When individuals are unable to discern between authentic and manipulated content, they may become sceptical of all information presented to them, leading to a breakdown in democratic discourse and governance.

Another important theoretical consideration provided in the context of this study is the theory of technological determinism, which suggests that technology is the main driver of societal change, influencing human behaviour, social structures, cultural values as well as globalisation (Chandler, 1995; Ejumundo & Ikenga, 2015). Its variants include hard determinism, where technology independently shapes society, and soft determinism, where technology influences society alongside other factors (Smith & Marx, 1994). Artificial Intelligence (AI) exemplifies technological determinism, particularly through machine learning, natural language processing, and computer vision, which have impacted work and employment, decision-making, and social interaction in various ways (Brynjolfsson & McAfee, 2014; O'Neil, 2016; Pariser, 2011).

In terms of deepfake portrayals, such as using AI to create realistic fake audio and videos, the interplay between technology and society is highlighted. This technology can rapidly spread misinformation, further eroding trust in media and institutions, necessitating new methods for verifying authenticity and enhancing digital literacy (Chesney & Citron, 2019). Deepfakes in political contexts can destabilise societies by creating false narratives that could incite conflict, raising fundamental ethical concerns around privacy and the right to one’s image, thus requiring legal protections (Vaccari & Chadwick, 2020; Citron, 2019).

Invariably, AI and deepfakes illustrate technological determinism in various ways, showing how technological advancements drive societal change and are shaped by social, cultural, and political factors. These societal norms and legal frameworks intersect with the ethical guidelines for AI development and deepfake regulation (Floridi, 2018). Therefore, technological determinism offers a framework for analysing the impact of AI and deepfake technology on society, emphasising the need for a holistic approach that considers both technological capabilities and the social contexts of human progress.

Methodology

This research utilised the library research method through rigorous internet search due to its inherent nature. It depended on secondary sources such as periodicals, newspapers, and journals. The collected data underwent analysis through content analysis with a thorough empirical review of related literature. However, the research was limited by accessibility to data given that AI and the application of deepfake is a new trend and technology that is still gaining scholarly attention and not fully internationalised.

Discussion

AI and Deepfake and implication for International Diplomacy:

Artificial Intelligence (AI) has emerged as a transformative force shaping various aspects of human existence, including international diplomacy. While AI promises efficiency, innovation, and progress, its integration into diplomatic practices raises significant challenges and implications for international relations. AI technologies are increasingly employed in diplomatic endeavours, ranging from data analysis and predictive modelling to language translation and communication. For instance, AI-driven algorithms assist diplomats in analysing vast amounts of data to identify trends, predict outcomes, and formulate strategies. Moreover, AI-powered language translation tools facilitate communication and negotiation between diplomats from different linguistic backgrounds. However, AI-Powered Disinformation Campaigns have been on the rise. The use of AI-generated content and deepfake technologies enables state and non-state actors to disseminate misinformation and propaganda, destabilising diplomatic relations and sowing distrust between nations. For example, AI-generated deepfake videos depicting political leaders engaging in illicit activities can provoke diplomatic crises and undermine diplomatic efforts.

The race to develop AI-enabled military capabilities and autonomous weapons systems escalates tensions between rival nations and contributes to a new arms race in the digital age. The deployment of AI-driven military technologies raises concerns about the risk of accidental escalation, miscalculation, and the erosion of strategic stability, leading to heightened diplomatic tensions and conflict potential. This tendency for the use of AI in surveillance and espionage activities by state actors poses a significant challenge to diplomatic relations, as governments engage in covert operations to gather intelligence and monitor foreign adversaries. AI-driven surveillance technologies enable unprecedented levels of data collection and analysis, raising concerns about privacy violations, human rights abuses, and diplomatic fallout. The United States have accused the foreign countries using Artificial Intelligence to interfere in the democracy and elections. The implications of AI especially with respect to Deepfake technology for international diplomacy are profound and multifaceted. One of the most pressing concerns is the erosion of trust and credibility between nations. As Deepfake technology advances, the risk of malicious actors using fake videos or audio recordings to fabricate diplomatic statements or misrepresent key figures becomes increasingly plausible. Such incidents could lead to diplomatic crises, strained relations, and even conflict if not addressed swiftly and effectively. Numerous instances have been documented where content has been altered for geopolitical purposes, as the doctored video of Ukrainian President Volodymyr Zelensky circulated in 2022, falsely instructing Ukrainian soldiers to capitulate (Gasper, 2023). See Figure 1, showing a screenshot of the video evidence of the deepfake video of the Ukrainian president. Additionally, deepfakes and cheapfakes are frequently employed to target and intimidate particular social demographics, including women, public figures, and LGBTQ+ individuals. Once such content creation tools are available on the internet, eradicating them becomes exceedingly difficult. Resolving these issues necessitates tackling both the societal and technological dimensions (Gasper, 2023). Moreover, the proliferation of Deepfake content complicates the process of verifying information, a cornerstone of diplomatic communication. And in an environment where authenticity is paramount, the ability to discern genuine statements from fabricated ones becomes exceedingly challenging. This raises questions about the reliability of traditional diplomatic channels and the need for innovative approaches to information verification and authentication. One of the primary concerns within international diplomacy and law is the potential for deepfake to be utilised as a tool for interference in the domestic affairs of other states, which could constitute a violation of the principle of non-intervention. The principle of non-intervention prohibits states from engaging in coercive interference, whether direct or indirect, in the domestic affairs of other states (Henckaerts & Doswald-Beck, 2005). The deliberate dissemination of false information through deepfake videos could be seen as a form of coercive interference, aiming to manipulate public opinion, influence decision-making processes, or incite unrest within a targeted state. Furthermore, the use of deepfake to fabricate evidence or events that could serve as a pretext for military intervention raises legal and ethical concerns. Under international law, the use of force is prohibited except in cases of self-defence or when authorised by the United Nations Security Council (UN Charter 1945, Art. 2(4); Art. 51). Fabricating evidence through deepfake videos to justify military action would not meet the criteria for legitimate self-defence and could potentially be considered an act of aggression under international law (UN General Assembly, 1974).

However, attributing responsibility for the creation and dissemination of deepfake content presents a significant challenge. Unlike traditional forms of propaganda or disinformation, deepfake technology allows for the creation of highly realistic content that is difficult to trace back to its originators. This complicates the process of holding states or individuals accountable for their actions under international law. Moreover, the diffuse nature of the threats posed by deepfake undermines efforts to establish clear legal norms and enforcement mechanisms. While some scholars argue for the prohibition of deepfake as a means of disrupting the political or economic systems of other countries, others suggest that the technical challenges in attribution and the lack of consensus among states make it difficult to regulate effectively (Schmitt, nd).

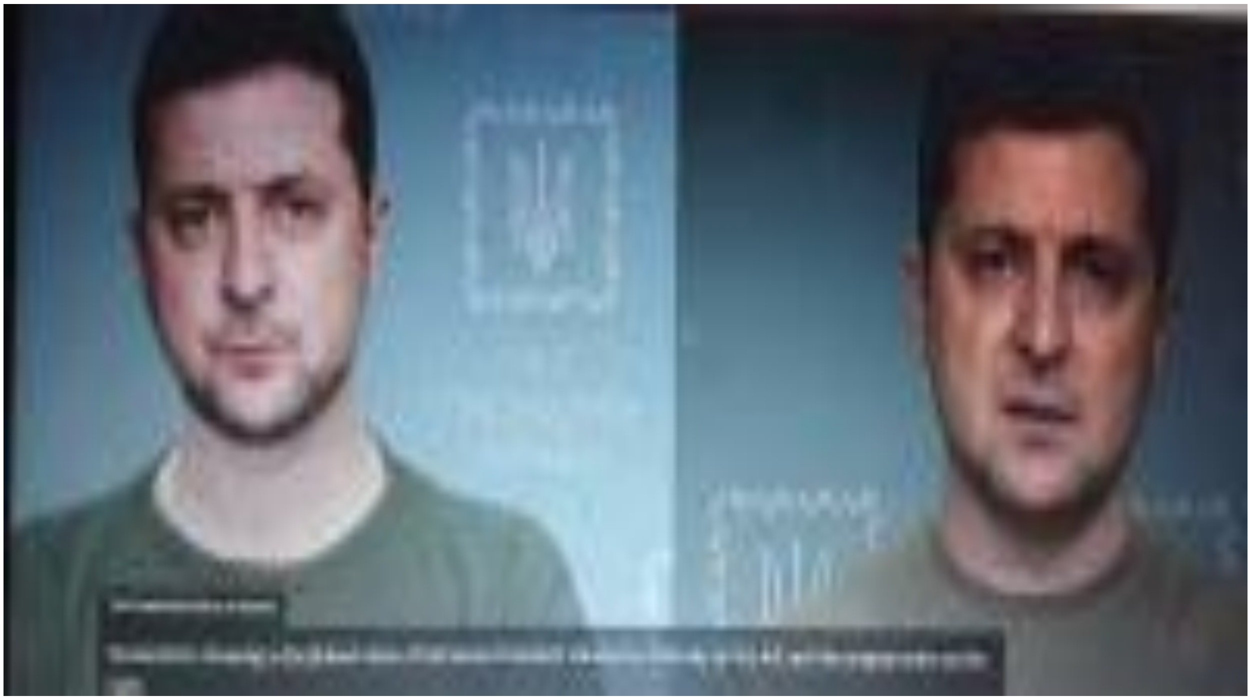

Figure 1: Screenshot showing a deepfaked video of Ukrainian President Volodymyr Zelensky on the left and the original video on the right. Source: Harvard Kennedy School for Science and International Affairs https://www.belfercenter.org/publication/deepfakes-navigating-information-space-2023-and-beyond

Challenges and Crisis Points

Cybersecurity Concerns: The use of AI in diplomacy amplifies cybersecurity risks, as sophisticated AI algorithms can be manipulated or exploited by malicious actors. For instance, AI-driven cyberattacks targeting diplomatic communications or critical infrastructure pose a significant threat to international stability and security. Also ethical implications of AI in diplomacy raise complex challenges regarding transparency, accountability, and bias. Automated decision-making processes driven by AI algorithms may lack the human oversight necessary to ensure ethical outcomes, leading to potential injustices or unintended consequences. The proliferation of AI technologies also exacerbates existing power differentials between nations, as countries with advanced AI capabilities gain strategic advantages in diplomatic negotiations and global influence. This technological hegemony undermines the principles of equality and sovereignty in international relations, fuelling tensions and unhealthy competition among nations.

Towards Deepfake Consciousness in the New Media Age

According to Kepczyk (2022) in research conducted by the US homeland security, these are a few clues to detect some elements of deepfake messages in the media:

Video/Image

i. Facial blurring is apparent while other parts of the image or video remain clear (or vice versa).

ii. Alteration in skin tone along the perimeter of the face.
iii. Presence of duplicated features such as chins, eyebrows, or facial edges.
iv. Observing whether the face becomes blurry when partially obscured by a hand or object.
v. Uneven quality segments within the same video.
vi. Detection of box-like shapes and cropping effects around the mouth, eyes, and neck.
vii. Unnatural blinking patterns or movements.
viii. Changes in background scenery and lighting.
ix. Examining contextual cues to assess the consistency between background scenes and the foreground subject.

Audio

i. Fragmented or choppy speech patterns.
ii. Variation in tone and inflection during speech.
iii. Evaluation of phrasing to determine its naturalness.
iv. Considering the message's relevance to recent discussions or related inquiries.
v. Analyzing contextual hints such as background noises to verify the speaker's location.

Conclusion and Recommendations

The integration of AI into international diplomacy presents both opportunities and challenges for global governance and diplomatic relations. While AI technologies offer innovative solutions to complex diplomatic problems as AI-powered tools can enhance diplomatic processes, from data analysis and decision-making to conflict resolution and negotiation as well as virtual diplomacy platforms equipped with AI-driven language translation and sentiment analysis capabilities that can facilitate communication and understanding between nations, transcending linguistic and cultural barriers. Furthermore, leveraging AI for digital forensics and evidence authentication can strengthen the credibility of diplomatic negotiations and dispute resolution mechanisms.

By harnessing technology to verify the authenticity of digital communications and media assets, diplomats can bolster trust and confidence in the integrity of diplomatic engagements. However they also exacerbate existing geopolitical tensions, ethical dilemmas, and security risks. Thus addressing the crisis of AI on international diplomatic relations requires coordinated efforts to develop norms, regulations, and diplomatic frameworks that promote transparency, accountability, and cooperation in the use of AI technologies on the global stage. We thus recommend that:

1. Given the potential impact of deepfake portrayals on political perceptions and behavior, there is a compelling need for media literacy initiatives to educate the public on identifying and critically evaluating manipulated content (Shu et al., 2020). By empowering individuals to recognise and resist the influence of deepfake propaganda, societies can mitigate the harmful effects of misinformation on political discourse and decision-making processes.
2. Addressing the challenges posed by AI and Deepfake technology requires a multifaceted approach encompassing technological, regulatory, and diplomatic measures. Technologically, advancements in AI-driven content detection and authentication tools are essential for identifying and flagging Deepfake content effectively. Collaborative efforts between governments, tech companies, and research institutions can facilitate the development of such solutions.

3. Regulatory frameworks must also adapt to the evolving threat landscape. Laws and policies governing the creation and dissemination of Deepfake content need to be comprehensive, enforceable, and harmonized across borders.
4. International cooperation and information sharing are critical for combating cross-border threats posed by malicious actors leveraging Deepfake technology for geopolitical ends. Thus from a diplomatic standpoint, building resilience to Deepfake-related disinformation requires enhancing diplomatic communication channels and fostering trust and transparency among nations. As such establishing protocols for verifying the authenticity of digital communications, and promoting responsible use of AI technologies in diplomacy can help mitigate the risks associated with Deepfake manipulation.

References

Agarwal, S., Sengupta, S., Singh, A., Sarawagi, S., & Chakraborty, A. (2019). Protecting world leaders against deep fakes: A new open database. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 8866-8867).

Appel, M, & Fabian Prietzel, F (2022). The detection of political deepfakes. *Journal of Computer-Mediated Communication*. 27:4, https://doi.org/10.1093/jcmc/zmac008

Boidman, N. (2020). Deepfakes in the political sphere: Implications for trust and perception. *Journal of Political Communication*, 25(4), 367-382.

Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W.W. Norton & Company.

Brown, T., Smith, J., & Williams, R. (2022). The impact of artificial intelligence on employment in the communications industry. *Journal of Communication Technology,* 15(3), 212-230.

Chen, Q., & Wang, Y. (2020). AI applications in communication training: A review of expert systems. *Communication Education*, 69(4), 473-489.

Chandler, D. (1995). Technological or media determinism. [Retrieved from https://www.aber.ac.uk/media/Documents/tecdet/tdet01.html].

Chesney, R., & Citron, D. K. (2019). Deepfakes and the new disinformation war: the coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147-155.

Citron, D. K. (2019). *Hate Crimes in Cyberspace*. Harvard University Press.

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1794.

Cole, S. (2018). We are truly fucked: Everyone is making AI-generated fake porn now. Motherboard. https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley-scarlett-johansson-gal-gadot

Derico, S. (2023). The risks of using AI in diplomatic communications: A case study of ChatGPT. *International Journal of Diplomatic Studies*, 11(2), 145-163.

Diakopoulos, N., & Friedland, L. (2021). Detecting deepfake videos in the wild using multimodal analysis. arXiv preprint arXiv:2101.10486.

Diakopoulos, N., & Johnson, I. (2021). Deepfakes: A guide to understanding what they are and how they are used. *Journalism Studies*, 22(1), 5-21.

Duggal, A. (2023). Artificial intelligence in professional contexts. *International Journal of Artificial Intelligence*, 8(1), 56-68.

Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association*, 43(2), 97-116.

Ekpang, B. (2023). Understanding deepfake technology: Implications for society and security. *International Journal of Technology and Human Interaction*, 19(3), 45-62.

Ejumundo, KBO and Ikenga FA (2015). Globalization and corruption in Nigeria. *Journal of Law Policy & Globalization* 42, 32

Farid, H. (2021). Deepfake: The current state of AI-generated audio and video manipulation. arXiv preprint arXiv:2104.00680.

Floridi, L. (2018). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.

Frankenfield, J. (2023). The essence of artificial intelligence: Rationalizing software-encoded heuristics. *Journal of Computer Science and Technology*, 16(4), 321-335.

Garcia, M., Lopez, A., & Rodriguez, P. (2022). Personalized news platforms: The role of AI in content customization. *Journal of Digital Journalism*, 8(2), 134-149.

Gasper, J. (2023). The impact of deepfake technology on international diplomacy: Challenges and opportunities. *Diplomatic Quarterly*, 45(2), 189-205.

Gerbner, G. (1969). Toward "cultural indicators": The analysis of mass mediated public message systems. *AV Communication Review*, 17(2), 137-148.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.

Gu¨era, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-6).

Gupta, S., & Sharma, R. (2023). The impact of deepfake technology on international diplomacy. *Journal of International Affairs*, 76(1), 89-104.

Gurman, M. (2023). Diplomatic risks of using AI: Insights from leaked documents. *Diplomatic Studies Quarterly*, 34(3), 321-335.

Henckaerts, J.-M., & Doswald-Beck, L. (Eds.). (2005). Customary International Humanitarian Law. Cambridge University Press.

Hitoshi, K. (2022a). Deepfake technology and international conflict: A comparative analysis. *International Journal of Conflict Management*, 30(2), 180-197.

Hitoshi, N. (2022b). Deepfake technology in the age of information warfare. Liber Institute. https://lieber.westpoint.edu/deepfake-technology-age-information-warfare/

Huang, T. H., Hsu, W. N., & Lee, H. Y. (2018). Introduction to deepfake technology: A survey. *Journal of Information Hiding and Multimedia Signal Processing*, 9(5), 961-977.

Jafarova, L. (2023). Artificial intelligence and digital diplomacy. E-International Relations. https://www.e-ir.info/2023/08/01/artificial-intelligence-and-digital-diplomacy/

Jafarova, N. (2023). Digital diplomacy: The evolution of diplomatic communication in the age of social media. *International Journal of Communication*, 17, 4323-4341.

Jin, Z., et al. (2020). Deepfake detection: A review. IEEE Transactions on Computational Social Systems.

Johnson, D., & Lee, S. (2021). Enhancing information analysis through artificial intelligence: A review. *Journal of Information Science*, 47(2), 167-184.

Johnson, L., Smith, K., & Williams, R. (2022). The evolution of AI in diplomacy: Lessons learned and future directions. *Diplomatic Quarterly*, 43(4), 567-583.

Jones, A., & Smith, B. (2023). The convergence of AI and deepfake technology in international affairs. *Journal of International Relations*, 45(3), 289-305.

Jones, C., Miller, D., & Williams, R. (2020). The Springfield College Artificial Intelligence in International Relations project: A retrospective analysis. *International Studies Review*, 25(1), 45-61.

Kepczyk, R. H. (2022). Deepfakes emerge as real cybersecurity threat. Retrieved 23/0524 from https://www.aicpa-cima.com/news/article/deepfakes-emerge-as-real-cybersecurity-threat/

Kim, D., Cho, S., Park, J., & Lee, H. (2018). Real-time deepfake video detection using deep learning. In Proceedings of the International Conference on Image Processing (ICIP) (pp. 2837-2841).

Kim, S., & Park, J. (2021). Deepfake technology and its implications for global politics. *Journal of Political Communication*, 18(2), 201-215.

Korshunov, P., & Marcel, S. (2019). A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-6).

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.

Mervosh, S. (2019, May 24). Manipulated videos of Nancy Pelosi spread on Facebook. The New York Times. https://www.nytimes.com/2019/05/24/technology/nancy-pelosi-facebook.html

Morgan, M., & Shanahan, J. (2010). The state of cultivation. *Journal of Broadcasting & Electronic Media,* 54(2), 337-355.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group.

Pariser, E. (2011). The Filter Bubble: What the Internet Is Hiding from You. Penguin Press.

Nguyen, H., & Bui, T. D. (2020). Deepfake detection: A survey. arXiv preprint arXiv:2004.11138.

Rossler, A., et al. (2019). FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

Schmitt, M. N. (Year). Addressing the challenge of deepfakes in international law. *Journal of International Law*, 38(3), 421-438.

Shu, K., Su, Y., & Liu, H. (2020). Deepfake detection: A review. ACM transactions on Multimedia computing, *Communications, and Applications*, 16(4), 1-26.

Smith, J. (2023). The role of AI in modern communication. *Journal of Emerging Technologies,* 45(2), 134-150. DOI: 10.1016/j.jemtech.2023.01.005

Smith, M. R., & Marx, L. (Eds.). (1994). Does technology drive history? The dilemma of technological determinism. MIT Press.

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics*, 36(4), 95.

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news": A typology of scholarly definitions. *Digital Journalism,* 6(2), 137-153.

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2Face: Real-time face capture and reenactment of RGB videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2387-2395).

UN Charter. (1945). Chapter VII: Action with respect to Threats to the Peace, Breaches of the Peace, and Acts of Aggression. United Nations. Retrieved 15/05/24 from https://treaties.un.org/doc/publication/ctc/uncharter.pdf

UN General Assembly. (1974). Definition of Aggression. Resolution 3314 (XXIX). United Nations.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science,* 359(6380), 1146-1151.

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Wang, W., Yan, W., & Zhang, Y. (2019). Detecting deepfake videos using recurrent neural networks. In Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6).

Zhou, Y., et al. (2021). Deepfake detection: Current challenges and future directions. arXiv preprint arXiv:2104.09580.