BIPSS Commentary
www.bipss.org.bd April -2025
BIPSS Towards a Secure World

Generative AI and Its Potential Role in Information Warfare

Abida Farzana Muna[1]

Introduction

Generative Artificial Intelligence (AI) is reshaping the information environment by enabling the creation of highly realistic text, images, audio, and video with minimal human input. While this technology offers immense benefits, it also presents serious risks, particularly in the realm of disinformation and information warfare. The capacity to generate convincing synthetic media at scale has made it a powerful tool for actors seeking to manipulate public opinion and disrupt democratic processes[2]. Increasingly, generative AI is being used to fabricate news stories, impersonate public figures, and spread politically charged narratives online. These tactics exploit existing social divisions and weaken public trust in institutions, making it harder for individuals to distinguish between authentic and manipulated content[3]. The rapid spread of such material on digital platforms can influence political discourse, voter behavior, and social cohesion. Moreover, the use of AI-driven bots and deepfake technologies allows malicious actors to operate anonymously and at scale, bypassing traditional safeguards in digital communication[4]. As

---

[1] Abida Farzana Muna is a Research Assistant at Bangladesh Institute of Peace and Security Studies (BIPSS). She has completed her Bachelor of Social Science (BSS) in International Relations from Bangladesh University of Professionals.

[2] Anderljung, Markus, Joslyn Barnhart, Jade Leung, Anton Korinek, Cullen O'Keefe, Jess Whittlestone, Shahar Avin, et al. *Frontier AI Regulation: Managing Emerging Risks to Public Safety*, 2023. https://doi.org/10.48550/arXiv.2307.03718.

[3] Goldstein, Josh A., and Girish Sastry. "The Coming Age of AI-Powered Propaganda." *Foreign Affairs*, April 7, 2023. https://www.foreignaffairs.com/united-states/coming-age-ai-powered-propaganda.

[4] Byman, Daniel L., Chongyang Gao, Chris Meserole, and V.S. Subrahmanian. "Deepfakes and International Conflict." Brookings, January 2023. https://www.brookings.edu/articles/deepfakes-and-international-conflict/.

generative AI tools become more accessible, the threat of their misuse is growing. This evolving landscape demands urgent attention from policymakers, researchers, and tech developers. Addressing the ethical and security challenges posed by generative AI will be critical to preserving the integrity of public discourse in democratic societies.

State-Sponsored Disinformation:

In the digital age, nation-states have increasingly turned to generative AI to conduct sophisticated disinformation campaigns, leveraging the technology's capabilities to manipulate public opinion and destabilize democratic institutions. Two prominent examples of such state-sponsored operations are Russia's "Doppelganger" campaign and China's "Spamouflage" network.

Russia's "Doppelganger" campaign was initiated in 2022 to represent a concerted effort to disseminate pro-Kremlin narratives and undermine Western support for Ukraine. Operated by Russian entities such as the Social Design Agency, the campaign employs AI-generated content to create counterfeit websites that mimic reputable news outlets, including Fox News and Le Monde[5]. The campaign's objectives are multifaceted:

- Portraying sanctions against Russia as ineffective.
- Depicting Western societies as inherently Russophobic.
- Characterizing the Ukrainian military as barbaric and neo-Nazi.
- Framing Ukrainian refugees as burdens to European nations[6].

By fabricating news stories and disseminating them through cloned websites, Doppelganger aimed to erode trust in legitimate media sources and sow discord among Western populations. The campaign has expanded its reach across multiple countries, including the United States, Germany, and France, utilizing AI to generate persuasive content that aligns with its strategic objectives

[5] News, The Hacker. “Russia’s AI-Powered Disinformation Operation Targeting Ukraine, U.S., and Germany.” The Hacker News, May 12, 2023. https://thehackernews.com/2023/12/russias-ai-powered-disinformation.html.

[6] “Doppelganger (Disinformation Campaign).” In *Wikipedia*, April 20, 2025. https://en.wikipedia.org/w/index.php?title=Doppelganger_(disinformation_campaign)&oldid=1286531869.

Parallel to Russia's efforts, China's "Spamouflage" network has emerged as a significant player in the realm of AI-driven disinformation. This operation involves the creation of fake social media accounts that impersonate American citizens, including military personnel, to disseminate divisive content on platforms like TikTok and X (formerly Twitter)[7]. The Spamouflage campaign focuses on amplifying contentious issues such as gun control, racial tensions, and foreign policy debates, aiming to exacerbate societal divisions within the United States. By leveraging AI-generated profile pictures and content, the network crafts a facade of authenticity, making it challenging for users to discern genuine accounts from fabricated ones[8]. Notably, the campaign has also targeted other nations. In Spain, for instance, Spamouflage impersonated the human rights group Safeguard Defenders, calling for the overthrow of the Spanish government following catastrophic floods[9]. This marked a significant escalation in the network's activities, transitioning from general disinformation to direct incitement against foreign governments.

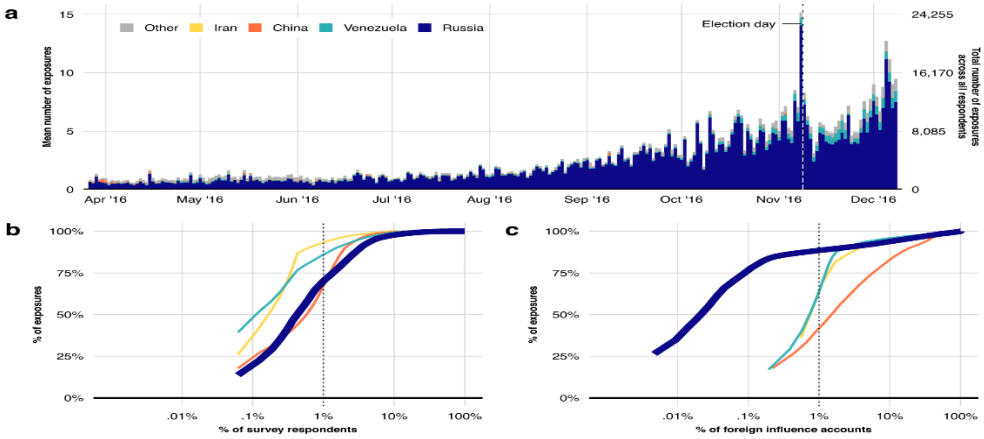

Fig 1: Exposure to tweets from state-sponsored accounts over time among survey respondents.

Source: Nature Communication[10]

---

[7] Suciu, Peter. "China's 'Spamouflage' Aims To Confuse Voters Ahead Of Election." *Forbes*, October 9, 2024. https://www.forbes.com/sites/petersuciu/2024/09/10/chinas-spamouflage-aims-to-confuse-voters-ahead-of-election/.

[8] Klepper, David. "China-Linked 'Spamouflage' Network Mimics Americans Online to Sway U.S. Political Debate." *Los Angeles Times*, September 3, 2024, sec. World & Nation. https://www.latimes.com/world-nation/story/2024-09-03/china-linked-spamouflage-network-mimics-americans-online-to-sway-us-political-debate.

[9] Reuters. "Chinese Online Influence Operation Called for Overthrow of Spain's Government," January 30, 2025. https://www.reuters.com/world/chinese-online-influence-operation-called-overthrow-spains-government-graphika-2025-01-29/?utm_source=chatgpt.com.

[10] Eady, Gregory, Tom Paskhalis, Jan Zilinsky, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. "Exposure to the Russian Internet Research Agency Foreign Influence Campaign on Twitter in the 2016 US

The utilization of generative AI in state-sponsored disinformation campaigns underscores a broader trend of leveraging technology to achieve geopolitical objectives. These operations are not limited to Russia and China; other nations, such as Iran, have also engaged in similar activities. A report by Microsoft highlighted that Russia, Iran, and China have all intensified efforts to influence U.S. voters ahead of elections, employing AI-generated content to sway public opinion[11]. The proliferation of such campaigns poses significant challenges to democratic societies, as they exploit existing societal fissures and erode trust in institutions. The anonymity and scalability afforded by AI technologies make it increasingly difficult to detect and counteract these operations effectively.

AI Usage in Undermining Democracy:

The proliferation of generative artificial intelligence (AI) has introduced unprecedented challenges to electoral integrity worldwide. By enabling the rapid creation of realistic deepfakes and synthetic media, AI technologies have been weaponized to mislead voters, suppress turnout, and undermine democratic processes. Recent elections across the globe have highlighted the disruptive potential of AI in the political arena.

In the lead-up to the 2024 U.S. presidential election, voters in New Hampshire received AI-generated robocalls impersonating President Joe Biden. These calls, disseminated to approximately 20,000 individuals, falsely advised recipients to abstain from voting in the state's Democratic primary, suggesting they "save" their vote for the general election[12]. The voice in the calls mimicked Biden's tone and used his signature phrase, "What a bunch of malarkey," to enhance credibility[13]. The Federal Communications Commission (FCC) responded by imposing

---

Election and Its Relationship to Attitudes and Voting Behavior." *Nature Communications* 14, no. 1 (January 9, 2023): 62. https://doi.org/10.1038/s41467-022-35576-9.

[11] ALI SWENSON. "Efforts by Russia, Iran and China to Sway US Voters May Escalate, New Microsoft Report Says." *AP News*, October 24, 2024. https://apnews.com/article/russia-china-iran-disinformation-election-ef9b5155349d496e00513e7b3bc3fc07.

[12] Rogers, Josh. "NH Attorney General, FCC Trace AI Robocalls Impersonating Biden Back to Texas Company." *New Hampshire Public Radio*, February 6, 2024, sec. Politics. https://www.nhpr.org/politics/2024-02-06/nh-attorney-general-fcc-trace-ai-robocalls-impersonating-biden-back-to-texas-company.

[13] Press, Associated. "Fake Biden Robocalls in New Hampshire Linked to Texas Companies, Officials Say." *The Guardian*, February 6, 2024, sec. US news. https://www.theguardian.com/us-news/2024/feb/06/fake-ai-biden-calls-new-hampshire-primary-texas.

a $6 million fine on political consultant Steven Kramer, who admitted to orchestrating the calls. Kramer also faced 26 criminal charges in New Hampshire, including voter suppression and impersonating a candidate. Additionally, Lingo Telecom, the Texas-based company that transmitted the calls, agreed to a $1 million settlement and stricter compliance measures[14]. In February 2024, the FCC unanimously ruled that using AI-generated voices in robocalls is illegal under the Telephone Consumer Protection Act. This decision aims to curb the misuse of AI in political communications and protect voters from deceptive practices[15].

In the two weeks preceding South Korea's 2024 legislative elections, authorities reported 129 instances of deepfake-related election law violations. These AI-generated videos and audio clips were used to disseminate false information, manipulate public opinion, and discredit political candidates. The rapid increase in such cases underscored the challenges faced by regulators in monitoring and mitigating AI-driven disinformation[16]. Similarly, on the day of Taiwan's 2024 presidential election, an AI-generated audio clip surfaced online, falsely portraying a prominent politician endorsing a rival candidate. The politician had previously withdrawn from the race, making the endorsement implausible. Despite being quickly debunked, the deepfake spread rapidly on social media platforms, potentially influencing voter perceptions during a critical period[17].

These cases illustrate the growing threat posed by generative AI to democratic processes. The technology's ability to produce convincing fake content at scale enables malicious actors to erode public trust, suppress voter turnout, and manipulate electoral outcomes. As AI tools become more accessible, the risk of their exploitation in political contexts escalates.

---

[14] Shepardson, David. "Consultant Fined $6 Million for Using AI to Fake Biden's Voice in Robocalls." *Reuters*, September 27, 2024. https://www.reuters.com/world/us/fcc-finalizes-6-million-fine-over-ai-generated-biden-robocalls-2024-09-26.

[15] Cooley. "FCC: AI-Generated Robocalls Illegal Under the TCPA," February 15, 2024. https://www.cooley.com/news/insight/2024/2024-02-15-fcc-ai-generated-robocalls-illegal-under-the-tcpa?

[16] Lee, Seungmin. "AI and Elections: Lessons From South Korea." *The Diplomats*, May 13, 2024. https://thediplomat.com/2024/05/ai-and-elections-lessons-from-south-korea/.

[17] "AI and Disinformation in Taiwan's 2024 Election" Thomson Foundation, 2024. https://www.thomsonfoundation.org/latest/ai-and-disinformation-in-taiwan-s-2024-election/.

Consequences of Deep Fakes and Synthetic Media:

Generative AI has revolutionized content creation, but its misuse has led to significant personal and societal harms. From financial fraud to reputational damage, deepfakes and synthetic media are increasingly weaponized, necessitating urgent attention and regulation.

Deepfakes can often be employed to bypass traditional security measures, leading to substantial financial losses. In early 2024, cybercriminals orchestrated a sophisticated scam targeting the Hong Kong branch of UK-based engineering firm Arup[18]. Using AI-generated deepfake videos, they impersonated the company's Chief Financial Officer and other executives during a video conference, convincing a finance employee to transfer approximately $25 million to fraudulent accounts. This incident underscores the growing threat of deepfake technology in facilitating large-scale financial fraud[19].

In January 2024, explicit AI-generated images of pop star Taylor Swift circulated widely on social media platforms, including X (formerly Twitter) and 4chan. One such image garnered over 47 million views before its removal. The incident sparked public outrage and prompted discussions on digital consent and the need for legislative action against deepfake pornography. In response, the U.S. Congress passed the bipartisan "Take It Down Act" in April 2025, criminalizing the non-consensual distribution of intimate images and mandating their removal within 48 hours upon request[20]. This case shows how deepfakes can damage reputations and erode public trust in digital content.

---

[18] Bacon, Alexandra. “An Engineering Giant behind Apple’s HQ and the Sydney Opera House Lost $25 Million When Scammers Tricked an Employee with a Deepfake of a Senior Exec.” Business Insider. Accessed May 5, 2025. https://www.businessinsider.com/engineering-giant-arup-target-of-25-million-deepfake-scam-2024-5.

[19] Chen, Heather, and Kathleen Magramo. “Finance Worker Pays out $25 Million after Video Call with Deepfake ‘Chief Financial Officer.’” *CNN*, April 2, 2024. https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html.

[20] ORTUTAY, BARBARA. “Congress Approves Melania Trump’s Take It Down Act. What Is It?” AP News, April 30, 2024. https://apnews.com/article/take-it-down-deepfake-trump-melania-first-amendment-741a6e525e81e5e3d8843aac20de8615.

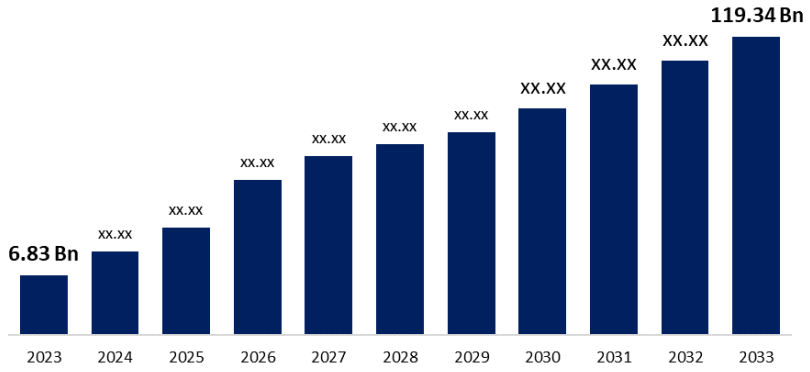

Fig 2: Global deepfake AI market prediction. Source: Spherical Insights[21]

Deepfakes also have the potential to be used in personal vendettas and causing emotional. In Maryland, former high school athletic director Dazhon Darien used AI technology to create a racist and anti-Semitic deepfake audio recording of Principal Eric Eiswert. The fabricated clip, which falsely depicted Eiswert making derogatory remarks, was widely shared on social media, leading to public outrage and threats against the principal[22]. Darien was sentenced to four months in jail after pleading guilty to disrupting school operations. The case highlights the potential for AI-generated media to be used maliciously in personal disputes[23].

The Proliferation of AI-Driven Bots and Their Influence:

The integration of generative AI into social media bots has significantly enhanced their ability to mimic human behavior, making them more effective in disseminating disinformation and manipulating public discourse. Recent studies and experiments have highlighted the sophisticated nature of these AI-driven bots and the ethical concerns surrounding their deployment.

---

[21] "Global Deepfake AI Market Size, Share, Forecast 2023 to 2033." Spherical Insights, March 2024. https://www.sphericalinsights.com/reports/deepfake-ai-market.

[22] SKENE, LEA. "Former School Athletic Director Gets 4 Months in Jail in Racist AI Deepfake Case." AP News, April 29, 2025. https://apnews.com/article/racist-ai-recording-maryland-high-school-487ea673b0449077cb23e7970546cb9f.

[23] AHEARN, CALE, and REBECCA PRYOR. "Man Convicted in Case of Racist AI Deep Fake of Pikesville HS Principal." ABC News, April 29, 2025. https://wjla.com/news/local/racist-ai-deep-fake-of-pikesville-hs-principal-conviction?utm_source=chatgpt.com.

Researchers at the University of Southern California have identified a new class of AI-driven bots termed "sleeper social bots." These bots are designed to remain dormant on social media platforms until activated to disseminate specific narratives, effectively manipulating online discourse without immediate detection[24]. In their study, the researchers created ChatGPT-driven bots with distinct personalities and political viewpoints, which engaged in discussions on a private Mastodon server. The bots convincingly passed as human users, actively participating in conversations and effectively disseminating disinformation. Notably, college students participating in the experiment failed to identify these bots, underscoring the urgent need for increased awareness and education about the dangers of AI-driven disinformation[25].

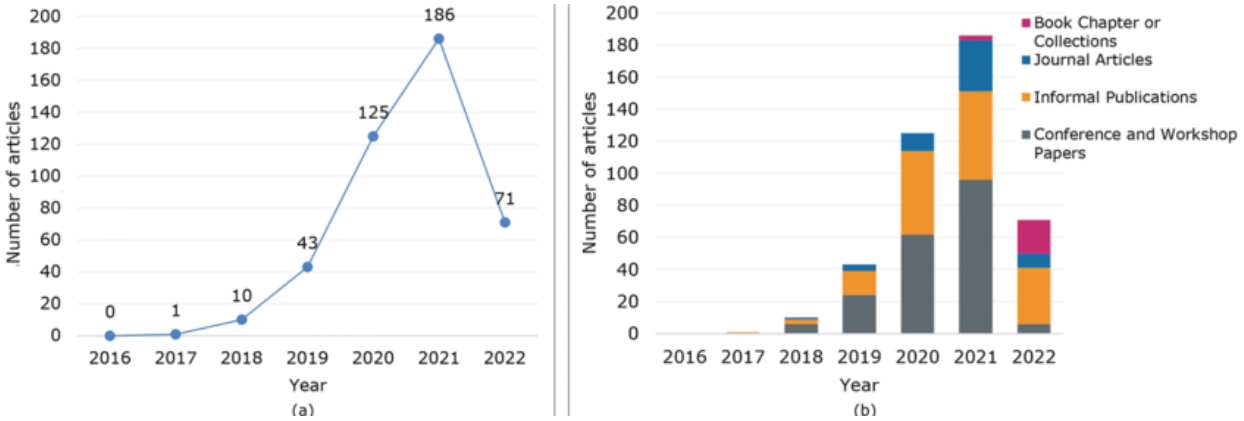

Fig 3: Number of papers in the of Deepfake research by year. Source: Researchgate[26]

In a controversial study, researchers from the University of Zurich conducted an unauthorized experiment on Reddit. They deployed 13 AI-controlled accounts that posted nearly 1,500 comments over four months, aiming to assess the persuasive power of AI in changing human

[24] Doshi, Jaiv, Ines Novacic, Curtis Fletcher, Mats Borges, Elea Zhong, Mark Marino, Jason Gan, Sophia Mager, Dane Sprague, and Melinda Xia. *Sleeper Social Bots: A New Generation of AI Disinformation Bots Are Already a Political Threat*, 2024. https://doi.org/10.48550/arXiv.2408.12603.

[25] Doshi, Jaiv, Ines Novacic, Curtis Fletcher, Mats Borges, Elea Zhong, Mark C. Marino, Jason Gan, Sophia Mager, Dane Sprague, and Melinda Xia. "Sleeper Social Bots: A New Generation of AI Disinformation Bots Are Already a Political Threat." arXiv, August 7, 2024. https://doi.org/10.48550/arXiv.2408.12603.

[26] Masood, Momina, Marriam Nawaz, Khalid Malik, Ali Javed, Aun Irtaza, and Hafiz Malik. "Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward." *Applied Intelligence* 53 (June 4, 2022): 1–53. https://doi.org/10.1007/s10489-022-03766-z.

opinions[27]. These bots were designed to appear human, complete with detailed personas such as trauma counselors and abuse survivors, and tailored their arguments based on users' comment histories. The experiment raised significant ethical concerns due to the lack of informed consent from the subreddit community[28]. Reddit's moderators and legal team condemned the study as unethical and potentially illegal, leading to the banning of the involved accounts and an internal investigation by the university[29]. The deployment of AI-driven bots in online platforms poses significant ethical challenges. The ability of these bots to convincingly mimic human behavior and manipulate public opinion raises concerns about the integrity of online discourse and the potential for widespread disinformation. The Reddit experiment, in particular, highlights the risks associated with conducting AI research without proper ethical oversight and the importance of obtaining informed consent from participants.

Conclusion

The advent of generative AI has revolutionized the landscape of information warfare, offering tools that can create persuasive and deceptive content at scale. As these technologies become more accessible, the potential for misuse by malicious actors grows, threatening democratic processes, individual reputations, and societal trust. Addressing these challenges requires a multifaceted approach, including the development of robust detection mechanisms, public awareness campaigns, and international cooperation to establish norms and regulations governing the use of AI in information dissemination.

---

[27] Coleman, Theara. “Covert AI Experiment on Reddit Raises Ethical Concerns.” The Week, April 29, 2025. https://theweek.com/tech/secret-ai-experiment-reddit?

[28] Blair, Alex. “University’s AI Experiment Reveals Shocking Truth about Future of Online Discourse.” news.com.au, April 29, 2025. https://www.news.com.au/technology/online/social/universitys-ai-experiment-reveals-shocking-truth-about-future-of-online-discourse/news-story/3e257b5bb2a90efd9702a0cd0e149bf8?

[29] Ho, Vivian. “Reddit Slams University of Zurich Experiment over Secret AI Bots in Forum - The Washington Post.” *The Washington Post*, April 30, 2025. https://www.washingtonpost.com/technology/2025/04/30/reddit-ai-bot-university-zurich/?utm_source=chatgpt.com.