Berlin, May 2024
Analysis and Consulting

KONRAD ADENAUER STIFTUNG

MONITOR

SECURITY

The Influence of Deep Fakes on Elections

Legitimate Concern or Mere Alarmism?

Ferdinand Gehringer, Dr. Christopher Nehring and Mateusz Łabuz

- The number of deep fakes of a political nature, including those directly targeting elections, has increased significantly.
- Currently, there is a clear tendency to implement deep fakes mainly to discredit political opponents, but also for consensual political advertising.
- Contrary to popular expectations, what seems to be the most worrisome consequence of deep fakes is their cumulative psychological and social effects.
- So far, in Germany, as in many other countries, deep-fake technology is primarily used for fraud and non-consensual pornography, and not yet in electoral processes or during election campaigns, but the cases of political deep-fakes are increasing.
- Building safeguards against the negative effects of deep fakes should include measures designed to protect and enhance information and election integrity as well as resilience.

Table of contents

2024 can be symbolically called a super-election year: around half of the world's population will be called upon to go to the polls. Presidential elections will be held in around 30 countries and parliamentary elections in around 20 others. Elections have already been held in Taiwan or Indonesia, and will also be held in some of the world's biggest democratic countries, including India, Mexico, and the USA. In addition to the German federal state elections in Thuringia, Saxony and Brandenburg in Germany, the European Parliament will also be elected in 2024.

Elections are an essential element of social co-determination and the cornerstone of participatory democracy. They determine the future distribution of power and enable citizens to have their say on this crucial issue. The development of modern technologies, including generative artificial intelligence, creates new challenges related to the integrity of elections. This includes the manipulative potential of deep fakes[1], which are considered one of the threats to democracy. The observed intensification of the use of deep fakes to intervene in the will-forming process in open societies or to influence election results is of particular importance in the super-election year. Especially in the face of the increasing erosion of trust in information, political institutions and processes worldwide.

A Rising Number of Deep Fakes Incidents

The number of deep fakes of a political nature, including those directly targeting elections, has increased significantly. In 2023 and early 2024, there were numerous attempts to influence the course of the election campaigns with the use of deep fakes. This is a significant change compared to previous years when such cases were isolated and rather poorly coordinated[2]. Currently, there is a clear tendency to implement deep fakes mainly to discredit political opponents, but also for consensual political advertising. Importantly, the characteristics of information space and consumption patterns are also changing[3], and the key element in the dissemination of deep fakes is algorithmic amplification, which allows harmful content to quickly reach recipients and reduces the feasibility of moderation, debunking and tackling illicit content.

Deep fakes are widely used in the campaign before the 2024 US presidential election. In addition to classic cases of discrediting political opponents, there was a very disturbing attempt to reduce voter turnout in the primaries in New Hampshire in January 2024 by a political consultant. Fake robocalls impersonating Joe Biden's voice could reach 5,000-25,000 potential voters, though the outcome and the range of the campaign is still under investigation[4]. Even if the influence on the outcome was probably manageable, this process shows that the potential behind the use of AI is easily accessible to everyone and can reach many people in a short time with a specific message.[5]

Also, the attacks in Slovakia (September 2023) and Bangladesh (January 2024) should be seen as particularly significant, because of the strategy used. In both cases, deep fake disinformation materials targeting politicians running in the elections were disseminated just before the vote. In the case of Slovakia, one of the candidates allegedly discussed election fixing (audio deep fake)[6]. In Bangladesh, one candidate allegedly announced her withdrawal (video deep fake)[7]. Regardless of the actual effects, targeting the so-called decisional checkpoints, defined by Chesney & Citron[8] as short periods before the vote, during which disruption of the decision-making processes may result in irreversible consequences due to the inability to debunk the false information effectively, could have been an attempt to test the ground for similar campaigns in the future.

In Turkiye, one of the candidates in the presidential elections decided to withdraw after being targeted by discrediting deep fake pornographic material, so-called deep porn (May 2023)[9].

In Argentina (October 2023), both leading candidates in the presidential elections used deep fakes, producing, among others, their election posters, but also materials ridiculing the opponent.[10] In this case, we can recognize the implementation of basic political advertising activities but also memetic warfare (using memes for disinformation purposes) on a mass scale.

Deep fake recordings of imprisoned Imran Khan were generated in Pakistan to mobilize his supporters. His opponents, in turn, tried to demobilize voters by distributing recordings of Khan discouraging citizens from participating in the elections – these, again, appeared just before the vote (January 2024)[11].

In Indonesia and India, deep fakes of leading political actors were widely disseminated to improve their images or attractiveness[12]. Thus, some politicians were to sing popular songs, and in Indonesia, AI allowed for resurrecting gen. Suharto, whose deep fake avatar was to conduct political advertising activities (February 2024)[13].

These are only selected examples, as deep fakes directly related to electoral processes have appeared, among others, in Poland, Bulgaria, Taiwan, Zambia and France. They should show the potential of misuse and manipulation. At the same time, deep fakes were used for other political purposes, which may also have consequences for electoral processes in the future.

The Elusive Influence of Deep Fakes

The question about the possibility of undermining democratic elections should no longer be if, but when, how and to what extent. Deep fakes are already influencing the course of democratic elections, but this influence is largely indirect and strongly related to the disinformation landscape. Each subsequent election may be a potential target of deep fakes, but the method of their use will depend on the specificity of information space and the political context. Attacks might be based on exploring specific vulnerabilities, including strengthening cognitive bias (e.g., attempts to confirm unfavorable rumors)[14], or targeting topics that particularly arouse public opinion (e.g. presenting a candidate fighting corruption while accepting a bribe).

Assessing the real consequences of using deep fakes to influence elections is difficult, especially measuring the scale of the phenomenon and quantifying its impact. Deep fakes are part of disinformation activities and therefore contribute to information disorders. It is not only about direct disinformation attacks, but also about the long-lasting consequences of undermining citizens' trust, enhancing uncertainty, or disturbing the epistemic value of the media, information, and their carriers[15].

In this context, the effects of deep fakes on elections are mainly psychological and social.[16] This does not mean that deep fakes do not have the potential to directly attack and disrupt the electoral processes. The key in this regard seems to be targeting the above-mentioned decisional checkpoints[17], which might already be a part of a new disinformation playbook.

Understanding the consequences of deep fakes requires building a methodological framework for assessing their impact. The use of specific forms of materials will be important. There are already clear signs that audio files pose a greater threat due to difficulties in debunking, although their persuasiveness still seems to be understudied.

Elections organized on a smaller scale (e.g. local ones) may be similarly sensitive, as the risk of manipulation may be increased due to the ease of reaching a larger percentage of potential voters with highly personalized messages. Candidates will also have less capabilities to debunk deep fake content than large campaigners who benefit from the media attention.

Drawing the line between the consequences of deep fakes and disinformation would be extremely difficult. However, the lack of an appropriate methodological framework and difficulties in empirical verification make it difficult to measure the real impact of deep fakes. Especially since activities aimed directly at elections may be complemented by activities in other areas: Deep fakes might be part of disinformation, military deception, or other manipulation attempts[18]. Yet, as it seems from an analysis of 2023 and early 2024 election-related deep fakes, so far have not had the effect that many observers expected[19]. Only in two cases, during the election in Slovakia and Turkiye, did deep fakes have a measurable effect on the election itself. But even there they did not decide the election.

The Biggest Threat: Not a Single Deep Fake, but Cumulative Effects

Contrary to popular expectations, what seems to be the most worrisome consequence of deep fakes is their cumulative psychological and social effects. First of all, deep fakes do not so much aim to persuade voters of one candidate or another but to distract, distort, smear, and disrupt. Deep fakes are used for negative campaigning and do not necessarily encourage people to vote for one particular candidate or party, but either not to vote at all or not to vote for a candidate or party. To do so effectively, deep fakes do not need to be particularly convincing or life-like as long as they have a clear emotional message that sticks. Studies have also shown that in these cases, labels such as “this content was produced with AI” do not keep audiences from sharing and believing in them.

Secondly (2), the mere fact that deep fakes and AI-manipulation technology exist is enough to have powerful and disruptive real-world effects. In 2019, for example, the rumor that a video message from President Ali Bongo of Gabon was a deep fake and that he was dead, was enough to trigger a military coup in the country[20]. But neither was he dead nor was the video a deep fake. And in autumn 2023, the accusation that an image of dead children killed during the Hamas attack on Israel published by the Israeli government was a deep fake was enough to completely distort and disrupt online debates about the attack[21].

In research, this phenomenon is known as "liar's dividend", describing claims that a piece of information is a deep fake in order to distract, distort or defend[22]. It might also be used for court proceedings, taking a form of "deep fake defense"[23]. In summer 2023, a lawyer defending Tesla in a Los Angeles court against charges due to a deadly accident with their autonomous driving programme used this tactic to carefully cast doubts and mistrust against a statement allegedly made by Elon Musk[24]. If everything could be a deep fake, then there is always reason for doubt. This logic and its political misuse have proven to be a powerful weapon.

Thirdly (3), influencing events and debates by claiming something is a deep fake is possible because this technology undermines the trust of any audience in the authenticity and truthfulness of any given piece of information. As studies indicate, the simple fear of deep fakes is enough to do so and to increase the level of uncertainty among recipients. Scholars have labeled this phenomenon "epistemic apocalypse", i.e. the blurring of the boundaries between real and fake[25]. Surveys in Germany, the UK and the US have shown that more than 70% of respondents expressed concerns about deep fakes and over 40% of US respondents stated that they had the feeling of being misled[26]. These fears result in more frequent questioning of the veracity of information.

Fear, mistrust, and insecurity thus might be the most important effects of deep fakes. And they have powerful, yet difficult-to-measure consequences for the political process, political communication and for elections. In a world where everything could be deep fake, election and information integrity are in danger notwithstanding the actual quantity or quality of deep fakes.

Implications for Germany

So far, in Germany, as in many other countries, deep-fake technology is primarily used for fraud and non-consensual pornography, and not yet in electoral processes or during election campaigns, but the cases of political deep-fakes are increasing.

The most prominent case of a political deep fake appeared in late 2023 when Chancellor Olaf Scholz featured in a video produced and published by left-wing activists, allegedly calling for an official ban of the far-right party AfD. Despite official debunking and the government's urge, popular social media platforms decided not to delete the video until ordered by court three months later.[27]

Russian pranksters (with alleged government ties) repeatedly targeted German politicians with deep fake phone calls of a political nature. In 2023, Minister of Economy Robert Habeck spoke for more than four minutes with Russian "comedians" posing as African politicians.[28]

Of all political actors, far-right politicians, activists and groups seem to employ AI-images and audio-deep fakes most often. Politicians of the far-right "AfD" have used AI to create symbolic and racist images of migrants[29] and anti-system protesters have used discrediting audio-deep fakes of Germany's most popular TV news ("Tagesschau") at demonstrations[30]. When confronted with massive national demonstrations against racism, its politicians and other supporters would also use the "deep fake defense", falsely claiming that pictures of the demonstrations were AI-generated to exaggerate the number of protesters.[31]

AI-generated images were also used during the wave of farmers' protests since late 2023. Some AI-images were merely symbolical, some were meant to exaggerate the extent of protests and in one particularly notable instant far-right politicians and other rightwing activists trying to capture the nature of the protests shared and commented on an AI-fake showing heyballs piled up in front of the Eiffel tower.[32]

As Germany prepares for the European elections in June 2024, regional elections in three federal states and national elections in 2025, it seems that most political actors are still exploring and testing the use of deep fake technology. However, repeating instances of successful deep fake-attacks have shown that traditional political actors still seem to be unprepared, not fully aware of the threat and lack appropriate response mechanisms.

Deep Fakes: A Systemic Threat and Necessary Measures

Instead of waiting for a "game changer-moment", a single deep fake that creates massive confusion and on its own decisively influences important political events, deep fakes should rather be seen as cumulative, systemic threats. Building safeguards against the negative effects of deep fakes should include measures designed to protect and enhance information and election integrity as well as resilience.

Political and public institutions should draft and introduce clear defense and response strategies to ensure election and information integrity. In addition, dedicated institutions should report regularly on the latest developments in the information space on various media and publish current narratives, disinformation campaigns and deep fakes used. Communicating with society about current threats and campaigns creates additional awareness.

Furthermore, political and media actors should outline, publish, and enforce common ethical standards on the use of AI in political communication, campaigning, advertising and reporting. A code of conduct should also set out clear limits for when AI-supported tools can be used in the election process and when they can no longer be used.

Media and AI-specific cyberliteracy measures need immediate bolstering. Programmes and educational measures aimed at increased awareness raising and resilience should become part of an election preparation curriculum for decision makers, candidates, but also officials, journalists, influencers and the general public. Therefore appropriate preparation should become an element of the school curriculum.

Rules and regulation, i.e. provisions of the EU AI Act or terms and community standards of social media platforms, need vigorous enforcement with "deep fake response units" and special resources for last minute election-related deep fakes. Lawmakers may also consider issuing additional (i.e. election-related) regulations. Cooperation and public-private partnerships between political actors, social media companies, traditional media and research can contribute to increased awareness, timely deletion, demobilization and debunking of content.

Public and private sector must work together to strengthen deep fakes detection methods and their availability. Law enforcement agencies should also invest in appropriate measures in this regard and develop the digital competences necessary to detect harmful content and thus respond more quickly.

Understanding the consequences of deep fakes requires building a methodological framework for assessing their impact. Even if the manipulative use of deep fakes has not yet decided an election, it is only a matter of time before the influence of deep fakes on elections and the electoral process increases.

Especially in a super-election year like 2024, it is time to initiate and implement the necessary measures in order to be prepared for future elections.

[1] Pursuant to the Artificial Intelligence Act adopted by the EU deep fakes are defined as AI generated or manipulated image, audio or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful.

[2] Łabuz M. Nehring C. (2024). On the way to deep fake-democracy? Deep fakes in election campaigns in 2023. European Political Science. Accepted.

[3] Huijstee van M. et al. (2021). Tackling deepfakes in European policy. European Parliamentary Research Service. Brussels.

[4] Matza M. (2024). Fake Biden robocall tells voters to skip New Hampshire primary election. https://www.bbc.com/news/world-us-canada-68064247.

[5] H. Ramer (2024). Political consultant behind fake Biden robocalls says he was trying to highlight a need for AI rules. https://apnews.com/article/ai-robocall-biden-new-hampshire-primary-2024-f94aa2d7f835ccc3cc254a90cd481a99.

[6] Meaker M. (2023). Deepfake Audio Is a Political Nightmare. https://www.wired.co.uk/article/keir-starmer-deepfake-audio.

[7] Pieal J. N. (2024). AI in politics: How lines between reality and 'deepfake' are blurring. https://www.tbsnews.net/features/panorama/ai-politics-how-lines-between-reality-and-deepfake-are-blurring-779066.

[8] Chesney B., Citron D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review. Vol. 107(18). pp. 1753-1820.

[9] Michaelson R. (2023). Turkish presidential candidate quits race after release of alleged sex tape. https://www.theguardian.com/world/2023/may/11/muharrem-ince-turkish-presidential-candidate-withdraws-alleged-sex-tape.

[10] Nicas J. (2023). Is Argentina the First A.I. Election?. https://www.nytimes.com/2023/11/15/world/americas/argentina-election-ai-milei-massa.html.

[11] Shahzad A. (2024). Pakistan's jailed Imran Khan uses AI-crafted speech to lure votes. https://www.reuters.com/world/asia-pacific/pakistans-jailed-imran-khan-uses-ai-crafted-speech-call-votes-2023-12-18.

[12] Christopher N. (2023). AI Modi started as a joke, but it could win him votes. https://restofworld.org/2023/ai-voice-modi-singing-politics.

[13] Ware G. (2024). Deepfakes and disinformation swirl ahead of Indonesian election – podcast. https://theconversation.com/deepfakes-and-disinformation-swirl-ahead-of-indonesian-election-podcast-223119.

[14] Brown N. (2020). Deepfakes and the Weaponization of Disinformation. Virginia Journal of Law & Technology. Vol. 23(1).

[15] Esselink J. (2021). Deepfakes and extreme beliefs. An ethical assessment. Vrije Universiteit Amsterdam. Amsterdam; Fallis D. (2021). The Epistemic Threat of Deepfakes. Philosophy & Technology. Vol. 34(4). pp. 623-643; Farid H. (2022). Creating, Using, Misusing, and Detecting Deep Fakes. Journal of Online Trust and Safety. Vol. 1(4).

[16] Thus, an increasingly polarized society provides a fertile ground for fueling polarization through social media and algorithms. Algorithms tend to favor emotional and oversimplified content, hence

why greater responsibility of platform operators through regulation is opportune. Emotional outrage and social validation play a significant role in this context.

[17] Chesney B., Citron D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review. Vol. 107(18). pp. 1753-1820.

[18] Byman D. L., Gao C., Meserole C. (2023). Deepfakes and international conflict. The Brookings Institution. Washington.

[19] Łabuz M. Nehring C. (2024). On the way to deep fake-democracy? Deep fakes in election campaigns in 2023. European Political Science. Accepted.

[20] Delcker J. (2019). Welcome to the age of uncertainty. https://www.politico.eu/article/deepfake-videos-the-future-uncertainty.

[21] Maiberg E. (2023). AI Images Detectors Are Being Used to Discredit the Real Horrors of War. https://www.404media.co/ai-images-detectors-are-being-used-to-discredit-the-real-horrors-of-war.

[22] Chesney B., Citron D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review. Vol. 107(18). pp. 1753-1820.

[23] Delfino R. (2023). The Deepfake Defense - Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers. SSRN Electronic Journal.

[24] The Guardian (2023). Elon Musk's statements could be 'deepfakes', Tesla defence lawyers tell court. https://www.theguardian.com/technology/2023/apr/27/elon-musks-statements-could-be-deepfakes-tesla-defence-lawyers-tell-court.

[25] Schick N. (2020). Deep Fakes and the Infocalypse. Octopus Books. Ottawa; Fallis D. (2021). The Epistemic Threat of Deepfakes. Philosophy & Technology. Vol. 34(4), pp. 623-643; Habgood-Coote J. (2023). Deepfakes and the epistemic apocalypse. Synthese. Vol. 201.

[26] Home Security Heroes (2023). AI deepfakes in 2024 election. https://www.homesecurityheroes.com/ai-deepfakes-in-2024-election; Luminate (2023). Bots versus ballots: Europeans fear AI threat to elections and lack of control over personal data. https://www.luminategroup.com/posts/news/bots-versus-ballots-europeans-fear-ai-threat-to-elections-and-lack-of-control-over-personal-data.

[27] Hanfeld, M (2024): Bundesregierung lässt Deepfake-Video mit Olaf Scholz verbieten, in: FAZ, 22.2.2024 (https://www.faz.net/aktuell/feuilleton/medien/bundesregierung-laesst-fake-video-mit-olaf-scholz-verbieten-19538916.html).

[28] N.N.(2023): Fake-Anruf beim Wirtschaftsminister Russische Trolle legen Habeck rein, in: Spiegel, 6.12.2023 (https://www.spiegel.de/politik/deutschland/robert-habeck-russische-trolle-legen-ihn-mit-fake-anruf-rein-a-5bfc1066-a1ba-4bae-8f3b-7506d4b37c00).

[29] N.N (2024).: So nutzt die AfD KI-Fotos für Propaganda, in: Watson, 3.4.2024 (https://www.watson.ch/digital/afd/174644994-so-nutzt-die-afd-ki-fotos-fuer-propaganda).

[30] N.N. (2024): Demonstrationen in Dresden Justiz ermittelt wegen tagesschau-Fakes, in: Tagesschau, 27.2.2024 (https://www.tagesschau.de/inland/justiz-ermittlungen-tagesschau-audiodateien-100.html).

[31] Pascal Siggelkow (2024): Massenproteste gegen rechts Falsche Behauptungen über Demo-Bilder,

in: Tagesschau Faktenfinder, 22.1.2024 (https://www.tagesschau.de/faktenfinder/demonstrationen-rechtsextremismus-bilder-100.html).

[32] Thomas Laschyk (2024): AfD-Fans fallen auf KI-Fake über Bauernproteste in Paris herein, in: Volksverpetzer Faktenfinder, 5.2.2024 (https://www.volksverpetzer.de/faktencheck/afd-ki-fake-bauernproteste-paris/).

Imprint

Authors

Ferdinand Gehringer: Policy Advisor Internal and Cybersecurity of the Department International Politics and Security Affairs, in the Analysis and Consulting Division of the Konrad-Adenauer-Stiftung e.V.

Dr. Christopher Nehring: PhD in Intelligence Studies, expert on AI, Disinformation and Intelligence Services. Fellow & Guest Lecturer for the Media programme Southeast Europe of the Konrad-Adenauer-Stiftung e.V.. Writes regularly for Tagesspiegel, Deutsche Welle, TableMedia, NZZ and other media.

Mateusz Łabuz: Career diplomat working for the Polish Ministry of Foreign Affairs; PhD candidate at the Chemnitz University of Technology (Germany); Lecturer of Cybersecurity and Artificial Intelligence at the University of the National Education Commission in Kraków (Poland); Lecturer of Cybersecurity at the Pontifical University of John Paul II in Kraków (Poland).

Konrad-Adenauer-Stiftung e. V.

Ferdinand Gehringer
Homeland and Cyber Security Department
International Politics and Security Affairs
Analysis and Consulting Division

T/p +49 30 / 26 996-3460

ferdinand.gehringer@kas.de

This publication by the Konrad-Adenauer-Stiftung e.V. is for information purposes only. It may not be used by any political parties or election campaigners or contributors to the intention of election advertising. This applies to Bundestag, Landtag and local elections as well as elections to the European Parliament.

Publisher: Konrad-Adenauer-Stiftung e. V., 2023, Berlin
Design: yellow too, Pasiek Horntrich GbR
Produced with the financial support of the Federal Republic of Germany.