



Personalising fakes: towards the disinformation apocalypse?

CHRISTOPH NEHRING

This article explores the effects and implications of AI-generated disinformation. It examines its forms, in particular deepfakes, and its impact on recent and upcoming elections. It also offers practical ideas for identifying and combating AI-driven disinformation, with particular attention to the role of influencers, journalists and other media professionals, and the unique challenges they face.

— —

Introduction

Artificial intelligence (AI) is rapidly transforming the global information landscape, creating new opportunities and unprecedented risks. Generating, spreading and boosting dis- and misinformation are prominent examples. Despite widespread fear and confusion, empirical knowledge on AI disinformation, its forms, impact and effects remain scarce, which in turn contributes to uncertainty, fear, mistrust and the demand for balanced, high-quality information.

Disinformation, deepfakes and manipulation

As early as 2023 disinformation experts spoke about the potential of generative AI as a “weapon of mass deception”, boosting and supercharging dis- and misinformation. Even though such doomsday scenarios have not yet materialized, AI possesses several qualities that have a strong effect on the production and distribution of dis- and misinformation. AI can make disinformation:

— *faster* (both in generating content as

well as automatically distributing such content)

— *cheaper* (e.g. by automating production and distribution, reducing human and financial resources needed)

— *more persuasive* (e.g. by using super real deepfakes)

— *more customized* (e.g. by using AI software for data analysis, identifying more effective messages and challenges to reach certain target audiences)

— *more far reaching* (e.g. by using AI bots and automation for distribution of disinformation or simply because AI tools are available for all normal social media users).

Experiments conducted by hackers and journalists have shown, for example, that the costs of using ChatGPT to power a fully automated fake news web site run entirely and only by ChatGPT dropped from 400 US\$ in 2023 to 105 US\$ in 2024.

Forms of AI disinformation

The *generative artificial intelligence* (genAI) revolution affects software designed to create every kind of content, text, images, videos and audio. Thus, known forms of AI disinformation include:

a. *Fake news web sites*: Even though difficult to detect, several thousand such sites whose contents (text, images, videos) are entirely created by ChatGPT or other chatbots have already been identified. Some, such as “electionwatch” or “TheDCWeekly” focus on organized disinformation about US politics and the 2024 US Presidential election, while others are commercial web sites simply rewriting and republishing old news for profit.

b. *AI-Images*: AI created images have started to flood social media platforms, messenger services and web portals. Some show either persons (most often politicians) in situations that never happens (e.g. Donald Trump dancing with underage girls) or depict events that never happened (e.g. a terrorist attack on the US Pentagon). While professional fake news web sites and outlets most likely backed by state actors also use AI images to publish along with fake articles, the vast majority of such AI generated images, however, are created and spread by “normal” social media and forum users. Such images are especially wide spread during the conflict in the Middle East, mostly emphasizing war damage and victims in Gaza. In some instances, such AI images that were spread on social media found their way to large stock databases (e.g. Adobe Stock), where they were sold for commercial use. Ukrainian users, on the other hand, are increasingly using AI generated images to portray support for the Ukrainian Army in their struggle against the Russian war of aggression. This trend demonstrates the effects of the “democratization” of genAI tools and their misuse.

c. *Deepfakes*: So-called “deepfakes” (deriving from “deep learning” and “fakes”) are AI produced or manipulated video and audio content. There are various different types of deepfakes, differing in either their

application (e.g. face swapping for deep porn or fraud and scamming calls) or the intention behind them. Deepfakes that are produced for the purpose of political disinformation have appeared in many different contexts, e.g. the Russian war against Ukraine and particularly during election campaigns all around the world (see below). Most often, they are used to produce fake discrediting evidence for scandalous statements or positions, participation in illegal or otherwise discrediting events or pornography. Their victims are most often publicly exposed person, e.g. celebrities, politicians, CEOs, influencers and journalists. Deepfakes have various qualities that have led to a high level of public fear and confusion: a) The impressive quality of such fakes; b) Their ability to convince and persuade audiences; c) The lack of reliable detection software and methods and d) the insecurity and inability of audiences to recognize and deal with deepfakes. During the following parts of this essay, its focus will be on deepfakes as one of the most imminent and dangerous examples of AI disinformation.

Experts and state investigators have found empirical proof for the existence of all these forms of AI disinformation. Yet, due to the so-called “detection challenge” of AI-content, it remains difficult to correctly assess and determine the quantity and actual amount of AI disinformation. Up to date, there is no 100% accurate detection method for AI generated content, no automatic upload filters or take down service etc. This means, while we can observe how the quality and the quantity of AI generated disinformation is rapidly increasing, its true quantity remains difficult to assess.

Deepfakes and elections in 2023 and 2024

AI disinformation and particularly deepfakes have become a weapon used to influence political campaigns and elections during the past two years. In most instances, deepfake technology was used to produce video or audio content featuring politicians, candidates, but also journalists and other popular voices in negative, discrediting scenarios. Some are reputational attacks on individuals trying to undermine their credibility, image and public reputation, others are part of negative political campaigns, trying to discredit political opinions, decisions or events. All of them, however, try to influence voter behaviour by deliberately spreading artificially created false, untrue or decontextualized information.

In other cases, deepfakes are used for official political campaigning. Such deepfakes differ in as much as they are a) attributable to an “official” source (e.g. a candidate, party, institution or organization), b) (often) labelled as AI generated content and c) not necessarily contain false information. During the elections for the European Parliament in June 2024, several far-right and right-wing parties (e.g. in France and Italy) used deepfake-technology to advance their messages and narratives via memes, images or AI-generated songs. In Pakistan, former PM Imran Khan and his team used deepfake technology to make him appear in campaign videos despite being imprisoned; in *India*, *Indonesia* and *the Philippines*, parties and campaigning teams created deepfakes of dead politicians or popular public figures for election campaigning. During the presidential elections in *Argentina*, both candidates and their teams heavily used all forms of generative AI (images,

» AI disinformation and particularly deepfakes have become a weapon used to influence political campaigns and elections. «

videos, text) for their campaigns. This also included malicious deepfake videos of both candidates that crossed the line between campaigning and disinformation by deliberately spreading aggressive lies. In *Mexico*, then presidential candidate and former mayor of Mexico City, Claudia Sheinbaum, featured in a deepfake video allegedly promoting a crude financial scheme, thus undermining her political credibility. Every country and election in 2024 saw political deepfakes that were meant to discredit political candidates and/or promote certain (mostly aggressive) narratives. By far the most use of AI-fakes was apparent during the US Presidential Election: Both sides (official party channels as well as supporters) published AI-generated images to get their messages across; however, more dangerous forms of deepfakes, e.g. AI-generated “robocalls” using the voice of President Joe Biden calling upon voters not to participate, AI-fakes depicting Taylor Swift cheering for Donald Trump or AI-generated non-existent content allegedly featuring in JD Vance’s book, spread on social media.

While deepfakes were part of every election in 2024, all empirical evidence points towards the fact that — unlike apocalyptic expectations — they did not

have a significant impact on the outcome of elections. So far, only in two cases did deepfakes that occurred during the last 48 hours before election day show a decisive influence on the election: In Slovakia, an audio-deepfake of one of the candidates allegedly discussing how to buy minority votes on the phone, seems to have had a direct effect on the outcomes, even though it did not “swing” the final results in favour of another candidate. During presidential election in Turkey, on the other hand, a deep-porn video of one of the candidates led to the withdrawal of said candidate from the elections. This obviously influenced the outcome of the election, yet as all polls saw the running president clearly winning the election anyway, this deepfake may have influenced the result, but not “swing” the election.

Journalism and Influencer: the Global Information Space

Generative AI has the potential to completely change the global information space. This includes all forms of political communication, content creation and presentation (including journalism and influencer).

AI and Journalism

GenAI has a strong effect on content creation as well as content presentation when it comes to journalism. Yet, there is an apparent “AI-gap”: Whereas traditional, quality media struggles to come up with concrete answers, boundaries and regulation concerning the ethical use of AI in journalism, low-quality media, tabloids, state-sponsored propaganda

outlets and fraudsters are already using AI for their purposes. The Russian foreign propaganda outlet “RT”, for example, already uses “deepfake personas”, i.e. non-existent, completely AI-generated automated avatars (which they call “digital presenters”) for their Spanish language programme. Several Chinese and state other news channels have been known for quite some time to use AI for these purposes as well.

And while traditional quality media all over the world mostly refrain from using genAI for creation of “core news”, i.e. from



» Russia is considered to be one of the most active actors of Foreign Information Manipulation and Interference. «

creating information itself, other actors are less prone to do so. News organisations have identified thousands of websites that rely on AI (most often ChatGPT) to run fully automated “news websites”. These websites either republish and rewrite old content for

advertising revenues or to spread outright political disinformation. Channel 1, a new news station established in Los Angeles in 2024, on the other hand, is the first reported news outlet that claims to be a serious media actor, but runs its programmes entirely with genAI, i.e. both for content creation as well as content presentation. Another important issue of genAI in journalism is the question of social media platforms regulate, label and publish AI generated content. While it will be soon become mandatory for platforms to label and specify AI content in North America and the EU, there are no such

unified norms for other parts of the world. Most social platforms themselves claim in their community standards and terms of use that AI-content and AI-profiles need to be clearly marked and registered. Yet, just like in the past, the extent to which these rules are being enforced varies significantly.

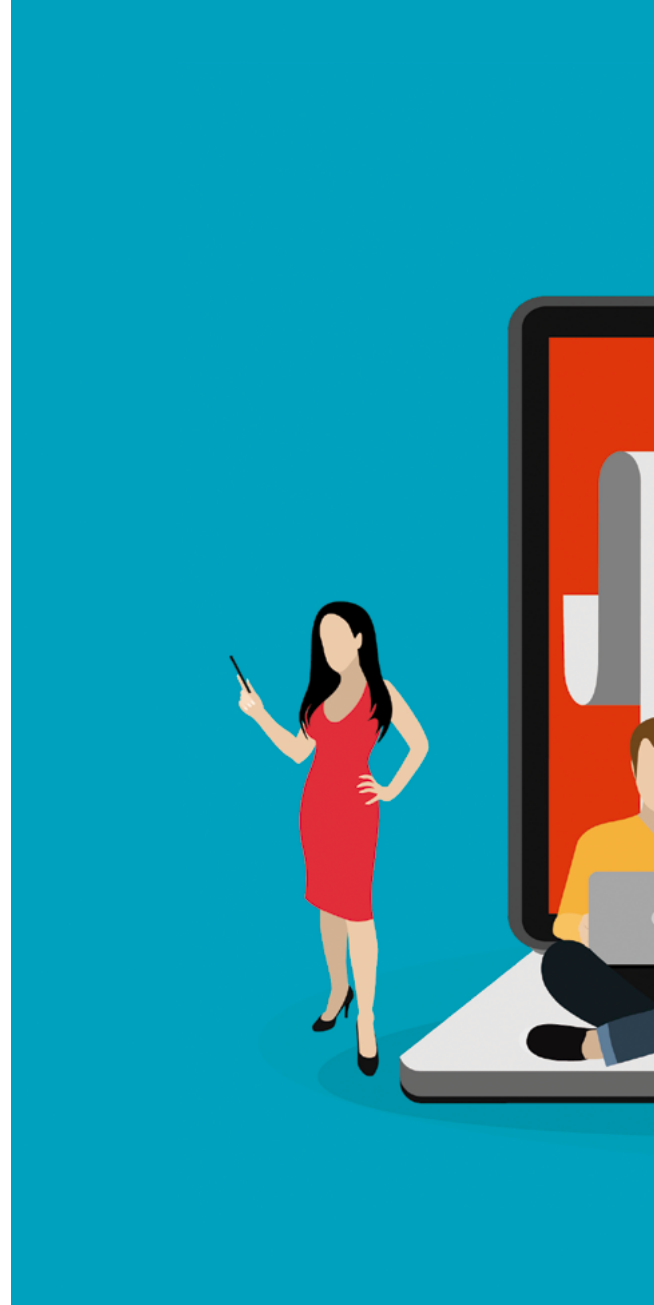
AI and Influencers

In the world of influencers, similar developments seem to be taking place and content generation and presentation are heavily affected by genAI. Virtual influencers, i.e. entirely AI-created and AI-driven avatars that pose as influencers on social media platforms have attracted millions of followers already (i.e. in China, Brazil, the USA or India). This is also true concerning the spread of mis- and disinformation, conspiracy theories etc. Throughout the world, influencers are gaining more and more significance as a target group and tool for professional disinformation actors, but also as professional creators and spreaders of disinformation.

Some TikTok-Influencers, for example, have turned AI-created videos about ever new conspiracy theories into their business model and are discussing in closed chat rooms about how to utilize genAI to increase revenues. In other cases, Russian embassies all over Africa have been proven to pay local influencers to spread disinformation. AI-influencer, on the other hand, have so far not been caught engaging in political campaigning and disinformation, yet they obviously bear a high risk potential in that regard.

AI manipulations and deepfakes also affect media professionals and influencers in other ways: Both groups (journalists and influencers) are regularly the victims of deepfake discreditation attacks using AI. Deepfake videos depicting journalists

promoting financial fraud schemes or dubious products without their consent have already become a regular occurrence in the US and all over Europe. Influencers, on the other hand, often face the risk of falling victim to deepfakes that target their reputation (and thus their business model). The most common scenario here is images





of female influencers being used for deep-porn. Such attacks may also be part of targeted reputation attacks for political purposes, as e.g. deepfakes of Taylor Swift following her mobilization during the US election in 2024 demonstrate.

» Artificial intelligence is fundamentally altering the landscape of disinformation, election interference, and information manipulation.«

“Over-Pollution” or: Drowning in a Sea of AI-Content

Yet another dimension of genAI in the global information space is the very real possibility of its “over pollution” with AI-content, automated AI-bots etc. Pessimistic scenarios propose the possibility of 90% of all online content being generated by AI in 2026, whereas automated online behaviour (e.g. bots and programmes) according to some studies already make up for the majority of all online activities. If genAI makes up for the majority of online content, presentation and activities, this will seriously affect political and other news, information and societies. Hence, “over pollution” might be one of the most serious long-term risks of genAI.

AI in Russian Foreign Information Manipulation and Interference

Russia is considered to be one of the most active actors of “Foreign Information Manipulation and Interference – FIMI”. The coordinated and covert spread of false, misleading and manipulated information to influence societies, events and elections is one important tool for these activities. Russian disinformation targets nearly every election on the planet and uses a large variety of complex tools and instruments. Russian embassies and consulates, Russian media, PR companies, paid journalists and influencers, anonymous web portals and local proxies are the most important actors of Russian disinformation. Its tactics span from simple propaganda information, paying influencers and journalists to complex information operations that include fakeing traditional quality media outlets and publishing covert disinformation. Narratives and messages of Russian disinformation usually center

❖ **AI-driven strategies can be used to micro-target voters, tailor messages, and enhance campaign efficiency** ❖

around certain key topics (e.g. anti-West, anti-Ukraine, anti-LGBTQ) which are reworked into customized messages for local audiences all around the world. In the Global South, for instance, such narrative usually focus on discrediting the Global West (e.g. Colonialism, social tensions, economic and social injustice etc.).

Use of AI

In its disinformation operations Russian FIMI-actors have shown their willingness to exploit the full and unrestrained potential of genAI for their purposes. The Spanish language programme of the Russian foreign propaganda broadcaster “RT” now includes two “digital presenters”, i.e. AI-avatars; in the US, several fake news websites that published fully automated negative articles about the US presidential election written with genAI were traced back to Russia; during an elaborate global disinformation campaign called “Doppelganger”, which centers around faked websites of the world’s most famous traditional media, Russian actors were caught using ChatGPT to generate and translate social media posts and comments; and in the ongoing war against Ukraine, Russian actors have time and again deployed deepfake videos (e.g. a fake video of President Zelenskyy calling for surrender or a fake video of Ukrainian intelligence chiefs allegedly admitting

their hand in an Islamist terror attack in Moscow) for foreign and domestic disinformation.

Conclusion

Artificial intelligence is fundamentally altering the landscape of disinformation, election interference, and information manipulation. Already today, all of these malicious activities have seen the integration of AI and no election now occurs anymore without some level of AI-generated disinformation. AI enables the production and dissemination of such content at unprecedented speeds, lower costs, and with increasing ease, making disinformation campaigns not only more accessible but also more automated, customizable, persuasive and large scale.

Despite these advancements, the feared “information apocalypse” has yet to materialize. No single election has been decisively influenced by AI-driven disinformation, even though there have been notable cases where deepfakes played a role. For example, in recent elections in Turkey and Slovakia, while deepfakes gained attention and raised concerns, they did not ultimately “swing the results” altogether in favor of any candidate or party.

Meanwhile, AI is not only a tool for disinformation but also a growing force in political campaigning. AI-driven strategies can be used to micro-target voters, tailor messages, and enhance campaign efficiency. As this trend grows, so too does the range of risks associated with AI manipulations, particularly deepfakes. Beyond elections, deepfakes contribute to cybermobbing, fraud and scams, and cybersecurity breaches, with influencers especially vulnerable to falling victim to such malicious uses of AI technology (e.g. cyberbullying with deepfake pornography). ♦

Bibliography

- BONTCHEVA, K. (ed.) (2024). [Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities.](#)
- FERRARA, E. (2024). [GenAI Against Humanity: Nefarious Applications of Generative Artificial Intelligence and Large Language Models.](#)
- GEHRINGER, F. A., NEHRING, CH., & LABUZ, M. (2024, May 10). [The influence of Deep Fakes on Elections: Legitimate Concern or Mere Alarmism?](#) KAS Monitor 2024.
- HABGOOD-COOTE, J. (2023). [Deepfakes and the epistemic apocalypse.](#) *Synthese*, v. 201 103/2023.
- LABUZ, M. & NEHRING, CH. (2024, April 26). [On the way to deep fake democracy? Deep fakes in election campaigns in 2023.](#) *Eur Polit Sci.*
- MARCHAL, N., & XU, R. (2024, August 2). [Mapping the misuse of generative AI.](#) *GoogleDeepmind.*
- MUÑOZ, M. (2024). The AI Election Year: How to Counter the Impact of Artificial Intelligence. DGAP Memo, v. 1.
- SCHICK, N. (2020). *Deep Fakes and the Infocalypse.* Ottawa.



Christoph Nehring

Researcher, analyst and journalist. Guest lecturer and analyst at the media programme of the Konrad Adenauer Foundation, author for *Tagesspiegel*, *Deutsche Welle*, *NZZ*, *Spiegel* and many others. Passionate about AI and disinformation. He has been researching disinformation, manipulation and secret services for more than ten years.
LI: christopher-nehring-423b06257