THREAT ANALYSIS

Recorded Future®

By Insikt Group®

September 24, 2024

Targets, Objectives, and Emerging Tactics of Political Deepfakes

Insikt identified 82 deepfakes targeting public figures in the last year, including 30 countries that held elections during this timeframe or have upcoming elections in 2024.

Deepfakes have had tangible impacts on elections and negative effects on the reputations of impersonated figures, and provoke a broader erosion of trust in democratic processes.

Deepfakes primarily impersonated heads of state and elected officials, but also candidates and journalists. Common attack vectors include scams, manipulating false statements, and electioneering.

Executive Summary

Public figures with high visibility during elections are increasingly impersonated by deepfakes, resulting in tangible impacts on election outcomes, negative effects on the reputations of impersonated figures, and a broader erosion of trust in elections. Deepfakes also exacerbate existing online harms, such as by taking advantage of media attention on elections to promote cybercriminal scams and to spread non-consensual pornography.

This report uses a dataset of 82 deepfakes identified by Insikt Group between July 2023 and July 2024 impersonating highly visible public figures, which includes heads of state, elected officials, electoral candidates, and journalists. The dataset includes examples from 38 countries, including 30 that held elections during this timeframe or have upcoming elections in 2024. The most common objectives observed were promoting financially motivated scams (26.8% of cases), followed by fabricating inauthentic statements to mislead the public (25.6%), electioneering (15.8%), character assassination (10.9%), and spreading non-consensual pornography to intimidate and cause reputational harm (10.9%). Emerging tactics identified include using fake whistleblowers, using deepfake audio masquerading as leaked recordings, spoofing assets from legitimate media organizations, using deepfakes of foreign leaders to endorse local political parties, and impersonating public figures' family members.

Deepfakes have already been operationalized and deployed successfully to target elections and have almost certainly had a more tangible effect on elections than other applications of generative artificial intelligence (AI) thus far. Research suggests that deepfakes need not be credible to durably affect public figures' reputations, as long as amplification of the deepfakes maximizes the number of exposed users.

AI companies, electoral campaigns, governments, and media organizations must develop countermeasures against deepfakes impersonating public figures. Communication strategies should focus on acting quickly and exposing audiences to impersonated individuals' real likenesses to discredit deepfakes, as lack of familiarity often feeds deepfakes' credibility and plausibility and is often exploited by influence actors. Organizations should also take inventory of copyrighted materials, as many deepfakes source material from copyrighted content such as older audio and video clips, providing organizations with leverage against such use via intellectual property lawsuits or other AI-specific courses of action.

Finally, organizations should also consider existing online harms exacerbated by deepfakes, such as the use of public figures' likenesses to promote cybercriminal scams or produce deepfake pornography. Impersonated organizations and individuals should prioritize cooperating with public authorities to identify cybercriminal campaigns and prevent financial losses in addition to taking down platforms publishing non-consensual deepfake pornography.

Key Findings

- Deepfakes are a global problem, with over 38 countries experiencing deepfakes impersonating public figures in the last year. 30 of these countries have conducted elections during the dataset timeframe or will conduct elections in 2024.
- Frequently impersonated public figures include current heads of state, elected officials, and well-known candidates. However, deepfakes also often exploit the public's lack of familiarity with recently elected officials, new candidates, and journalists.
- Deepfakes have been used to erode public confidence in governments, news media, and political parties, as demonstrated by claims that the Reform UK political party had fielded fake candidates following its publishing of campaign materials suspected to be AI-generated.
- 26.8% of deepfakes impersonating public figures in the last year likely capitalized on heightened media attention during electoral periods to promote scams, highlighting how cybercriminal objectives can overlap with influence tactics by promoting deepfakes on social media platforms.
- Deepfakes impersonating public figures also attempted to either produce false or extreme statements (such as false policy endorsements) or portray victims as engaging in unethical or illegal behavior, seeking to weaken public trust and inflict reputational damages.
- Deepfakes are also seeing increased adoption by electoral campaigns for both consensual use, such as broadcasting speeches and statements in different languages, and non-consensual use, such as impersonating and discrediting rival candidates.
- Non-consensual deepfake pornography disproportionately targets women in politics and media and is likely to inflict psychological and reputational harm on victims and deter women from running from office, and historically has even led to the withdrawal of a male candidate in Türkiye's 2023 elections.
- Emerging tactics include making deepfakes of anonymous third parties (including paid actors) posing as whistleblowers and journalists, in addition to impersonating news media organizations, manipulating video or images of foreign heads of state to endorse domestic political parties, and impersonating elected officials' family members.

Table of Contents

Methodology

This report uses a dataset of 82 deepfakes identified by Insikt Group between July 2023 and July 2024 impersonating highly visible public figures, which includes heads of state, elected officials, electoral candidates, and journalists. The dataset includes examples from 38 countries, including 30 that held elections during this timeframe or have upcoming elections in 2024. This report defines deepfakes as AI audio and video material generated with the intent of altering or impersonating a person's voice, face, or body. A dataset of deepfakes was assembled and manually labeled according to discrete categories, reflecting the distinct origins, targets, and objectives of each deepfake.

Targets

Insikt Group identified deepfakes impersonating public figures in over 38 countries. Public figures from the European Union (EU) were the most frequently impersonated, with thirteen EU countries having public officials impersonated by deepfakes in the last year. In addition, deepfakes were heavily present during elections. 30 (78.9%) of the countries whose public officials were impersonated by deepfakes have held elections during this timeframe or have upcoming elections in 2024.

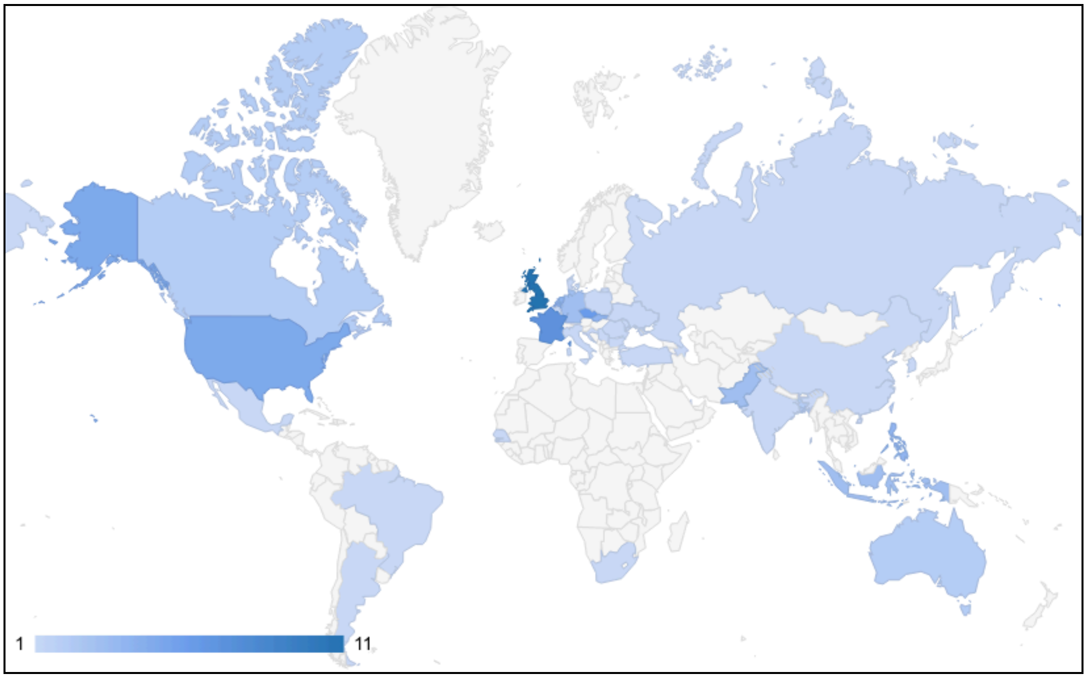


Figure 1: *Map of public officials impersonated by deepfakes per country (Source: Recorded Future)*

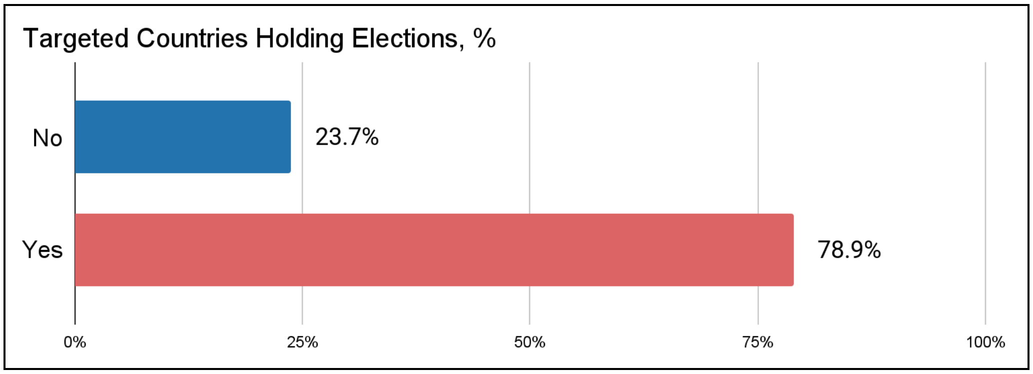


***Figure 2**: 78.9% of individuals impersonated in our dataset were in countries holding an election in 2024 (Source: Recorded Future)*

Insikt Group classified impersonated individuals under four categories:

- Heads of state and heads of government
- Electoral candidates
- Elected officials (including parliamentarians, ministers, and other members of government, excluding election candidates)
- Members of the media (including journalists)

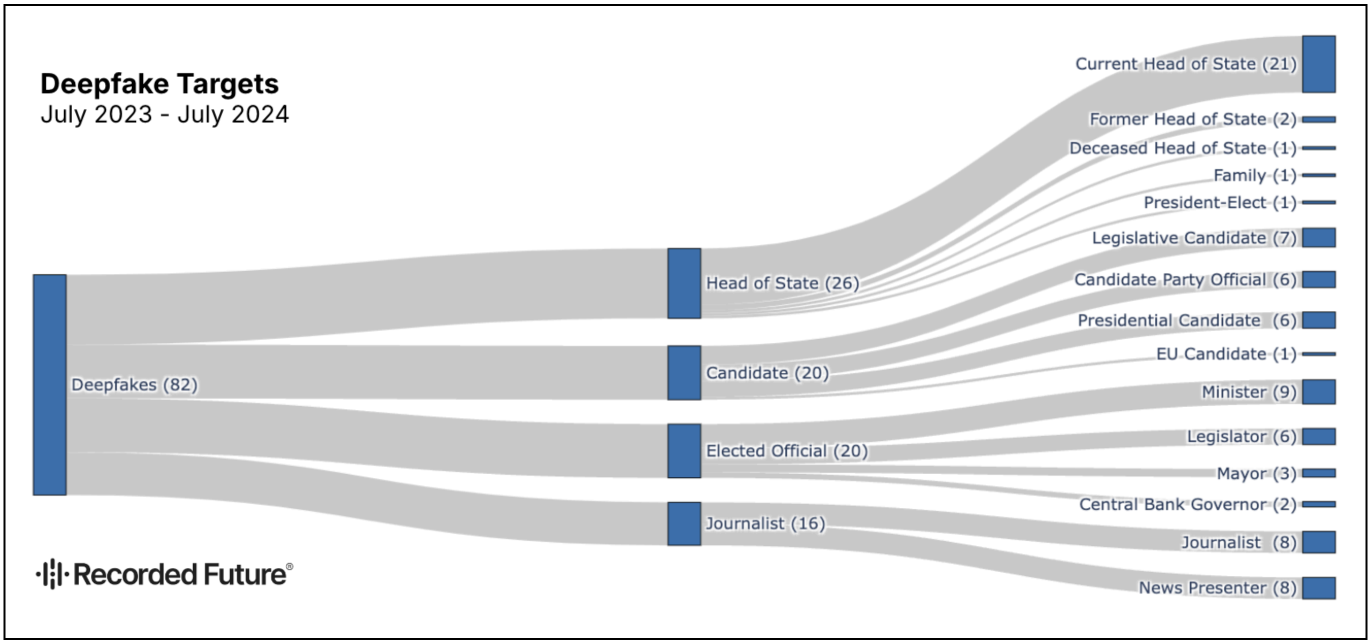


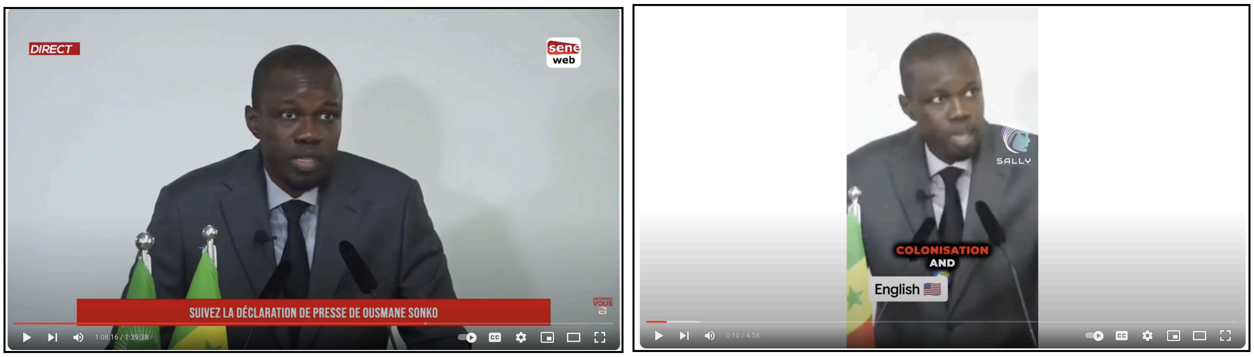
***Figure 3:** Breakdown of deepfakes by the category of impersonated individuals (Source: Recorded Future)*

Recorded Future data suggests that the most common targets of deepfake impersonations are the most visible public figures, including heads of state, electoral candidates, and elected officials. However, deepfakes impersonating well-known leaders are unlikely to be convincing. Research shows that the likelihood of a social media user correctly identifying a video as a deepfake is likely related to their familiarity with the impersonated figure, with only 43.9% of survey respondents being able to correctly identify a deepfake of an unfamiliar person, versus 72.7% for deepfakes of familiar people. Emerging tactics to overcome familiarity bias include using deepfakes of foreign political figures to endorse local political parties and impersonating new candidates or newly elected officials.

Heads of State

Current and former heads of state or government were our dataset's most commonly impersonated category of public figures, with heads of state from 23 countries impersonated by deepfakes in the last year. Regionally, heads of state in Eastern European countries were popular targets, including Polish Prime Minister Mateusz Morawiecki, Bulgarian President Rumen Radev, Croatian Prime Minister Andrej Plenković, Moldovan President Maia Sandu, Serbian Prime Minister Milos Vucevic, Slovakian Prime Minister Robert Fico and President Zuzana Čaputová, Czech President Petr Pavel, and former Czech President Václav Klaus.

Several newcomers and presidents-elect were impersonated shortly after electoral victories. Malicious actors likely sought to exploit the period between election results and inauguration to capitalize on the lack of familiarity before the electee gained significant international visibility. For example, Senegalese President Bassirou Diomaye Faye was impersonated immediately after being elected on March 24, 2024, and before taking office on April 2. A deepfake video inauthentically labeled as depicting Diomaye used a 2021 speech by party leader and current Senegalese Prime Minister Ousmane Sonko to attempt to use Sonko's likeness and AI-generated audio to depict Diomaye as disparaging France. Nigerian fact-checkers found that the deepfake had been first uploaded to YouTube in September 2023 but was amplified on social media following Diomaye's electoral victory. Influence actors likely sought to exploit Diomaye's limited international recognizability as a newly elected president to portray him negatively to an international audience via a deepfake using the likeness of another politician. Additionally, the deepfake was produced in English despite the original clip being in French, suggesting a deliberate targeting of international audiences who were less likely to be familiar with Diomaye's likeness before his electoral victory.



***Figures 4 and 5**: (Left) Original clip and (Right) resulting deepfake of Ousmane Sonko, presented as Diomaye (Source: YouTube, YouTube)*

Elected Officials

Elected officials, government officials, and members of national legislatures were the second most commonly impersonated category of public figures in our dataset, representing 24.4% of deepfake impersonations. Among these, the most common sub-categories by position were:

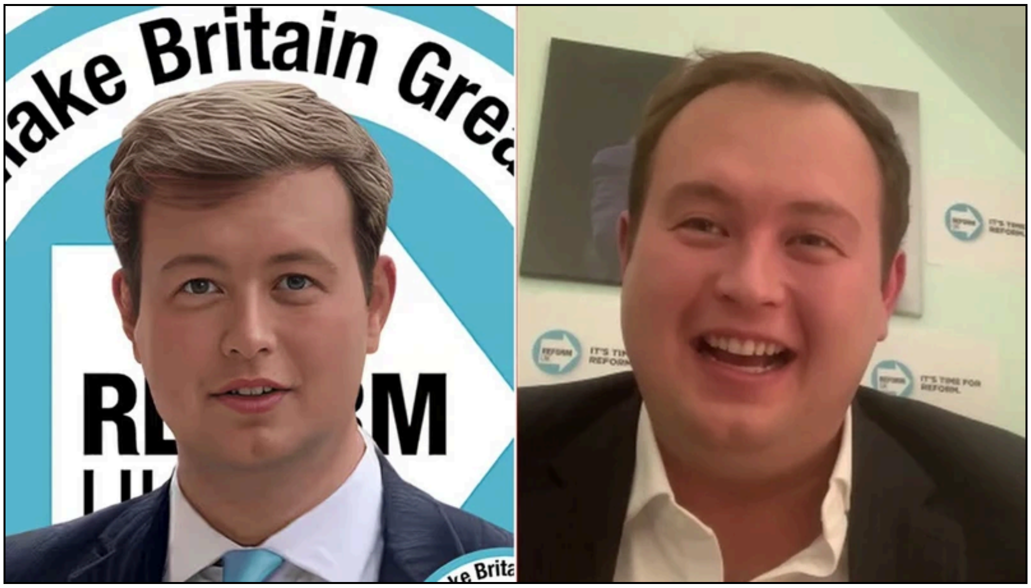
- **Government ministers,** including Canadian Finance Minister Chrystia Freeland, Singaporean Deputy Prime Minister Lawrence Wong, Australian Finance Minister Katy Gallagher, and UK Education Secretary of State Gillian Keegan, were impersonated by deepfakes used to promote scams.
- **Legislators,** including US Representative Alexandria Ocasio-Cortez, UK MPs Penny Mordaunt and Stella Creasy, and Dutch MP Caroline van der Plas, were all impersonated in deepfake pornography videos or images.
- **Mayors**, including London Mayor Sadiq Khan, then-Mexico City Mayor Claudia Sheinbaum, and Costa Rica-MS (Brazil) Mayor Cleverson Alves dos Santos, were depicted as endorsing a pro-Palestinian march in London, promoting an investment scam, and insulting constituents, respectively.
- **Heads of central banks** were also impersonated by deepfakes promoting scams, including Gaston Reinesch in Luxembourg and Mugur Isărescu in Romania.

Electoral Candidates

Deepfakes have impersonated candidates in presidential, legislative, and municipal elections that have taken place in the last year or are scheduled to take place in 2024. Deployment of deepfakes impersonating candidates has resulted in tangible impacts on elections, including the withdrawal of candidates, and significant reputational damages, including disputing the existence of real candidates.

In some cases, deepfakes impersonating candidates were strategically amplified shortly before polls opened, showing clear intent to manipulate voters' decisions and discredit candidates. This strategy is also likely to disproportionately affect candidates running in countries that impose media blackouts in the days ahead of elections. Slovakia, for example, imposes a media blackout four days prior to elections, with a reported 13% of Slovakian voters deciding their vote during this period for the 2019 EU elections. Two days prior to Slovakia's parliamentary election on September 29, 2023, a deepfake audio clip alleged to have been a recording of Slovakian journalist Monika Tódová and Member of the European Parliament (MEP) and chairman of the Progressive Slovakia party Michal Šimečka discussing electoral fraud surfaced on social media. In some countries, such as the Netherlands, up to 42% of eligible citizens reportedly decided who to vote for in the days leading up to the 2019 EU elections.

Deepfakes can also undermine trust in new candidates with whom voters are less familiar relative to heads of state or opposition leaders. On July 4, 2024, during the UK's general election, conspiracy theories reportedly led some individuals to believe that Reform UK candidate Mark Matlock was fake and that his likeness had been generated using AI based on election materials. The party was reportedly asked to clarify the existence of its candidates, and the candidate was forced to state that he was “not AI”. While these allegations came out following the vote and were unlikely to have durably impacted voters’ decisions, this incident likely demonstrates deepfakes’ systemic erosion of the broader information environment, with the credibility of campaign materials modifying candidates’ appearances being questioned.



***Figures 6 and 7**: (Left) Modified picture of Mark Matlock; (Right) Real picture of Mark Matlock (Source: GB News)*

Journalists

While not as common as heads of state, elected officials, or candidates, deepfakes have also impersonated journalists and news presenters, often coupled with imitations of legitimate news organizations’ overlays and branding. Deepfakes impersonating journalists are also becoming increasingly common: non-governmental organization The Coalition for Women in Journalism reported that it detected as many deepfakes in Q1 2024 as it had throughout the previous year in 2023.

Deepfakes impersonating journalists and media organizations typically fulfill one of two objectives: to discredit the journalists and erode trust in media organizations more broadly or to hijack the legitimacy of these organizations to spread narratives. For example, deepfake audio was used to fabricate false apologies from journalists discrediting German news show Tagesschau and public-service broadcaster network ARD. Journalist Susanne Daubner could be heard in the deepfake audio apologizing for the show’s “manipulation” of topics such as the war in Ukraine, COVID-19, and anti-Alternative for Germany (AfD) marches in Dresden in February 2024.

Hijacking the legitimacy of news organizations and journalists often helps bolster deepfakes’ credibility. In June 2024, both French journalist Elise Lucet’s likeness and branding from public broadcaster France 5 were used in a deepfake attempting to legitimize a gambling scam originally impersonating French football player Kylian Mbappé, who was also impersonated by a deepfake. Lucet’s likeness had also previously been used in deepfakes promoting a gambling scam in March 2024.

Objectives

Deepfakes identified by Insikt Group in the last year attempted to accomplish the following objectives:

- **Financial Gain**: Deepfakes are frequently used to impersonate public figures to endorse fraudulent schemes, especially during periods of heightened media attention, such as elections. These scams, often involving investment opportunities or cryptocurrencies, can cause significant financial losses for unsuspecting individuals.
- **Misleading the Public by Creating False Statements**: Deepfakes are used to create false statements by public figures that can mislead the public. The impersonated figure is often portrayed as criticizing their own party, endorsing policies they do not support, or making extreme statements.
- **Character Assassination**: By portraying public figures engaging in unethical, illegal, or embarrassing behavior, deepfakes aim to damage their reputation. This includes showing them involved in scandals, corruption, or other forms of discrediting content, weakening public trust in these individuals.
- **Electioneering**: Deepfakes are being increasingly deployed by political campaigns to promote voter participation, discredit opposition candidates, or depict candidates in misleading ways, such as speaking another language or from prison.
- **Spreading Non-Consensual Deepfake Pornography to Intimidate and Cause Reputational Harm**: Female politicians and journalists are disproportionately targeted by deepfake pornography. These explicit videos can cause reputational harm, intimidate the impersonated figures, and discourage participation in public life.

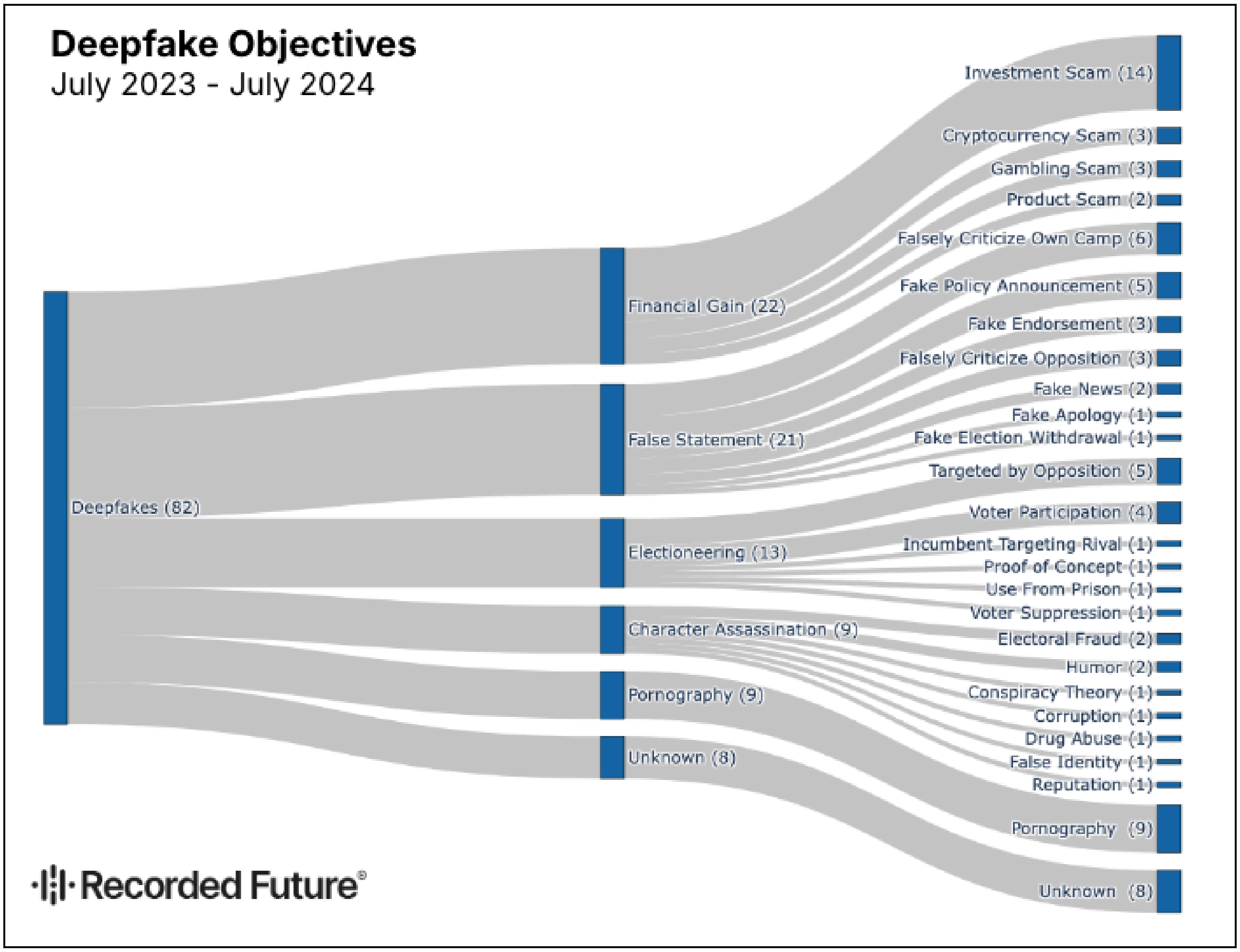
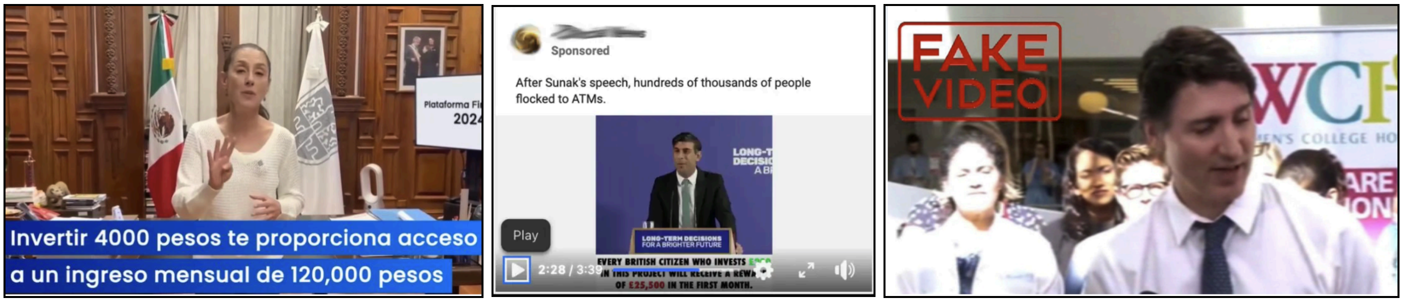


***Figure 8**: Breakdown of influence objectives used by deepfakes impersonating public figures (Source: Recorded Future)*

Financial Gain

26.8% of deepfakes targeting public figures in the last year likely capitalized on heightened media attention during electoral periods to promote scams, highlighting how cybercriminal objectives can sometimes overlap with influence tactics by promoting deepfakes on social media platforms. Deepfakes of well-known public figures promoting scams include Mexico's President-Elect Claudia Sheinbaum and former UK Prime Minister Rishi Sunak promoting investment scams, and Canadian Prime Minister Justin Trudeau endorsing a cryptocurrency scam. Investment scams were the single most frequent sub-category of objectives (fourteen) observed in our dataset.


Invertir 4000 pesos te proporciona acceso
a un ingreso mensual de 120,000 pesos
Plataforma Fi
2024
Sponsored
Play
2:28 / 3:39
LONG-TERM DECISIONS FOR A BRIGHTER FUTURE
EVERY BRITISH CITIZEN WHO INVESTS
IN THIS PROJECT WILL RECEIVE A REWA
OF £25,500 IN THE FIRST MONTH.
FAKE VIDEO

After Sunak's speech, hundreds of thousands of people flocked to ATMs.

Figures 9, 10, and 11: *Deepfake videos using Claudia Sheinbaum (left), Rishi Sunak (center), and Justin Trudeau (right) to promote scams (Source: Mexico News Daily, TBIJ, CTV News Toronto)*

In addition to the reputational risks to impersonated individuals, such as association with cybercriminal activity, such deepfakes can cause significant harm to social media users who are duped by the promoted scams. In one example, a Canadian man reportedly lost over $11,000 after falling for a deepfake video of Prime Minister Trudeau promoting a cryptocurrency exchange. Deepfakes promoting scams can also reach significant audiences; a deepfake of Slovakian Prime Minister Robert Fico posted to Facebook in February 2024 was reportedly seen by 22% of Slovakians prior to the country's presidential elections in March and April 2024.

Misleading the Public by Creating False Statements

The second most common objective of deepfakes impersonating public figures is to fabricate false statements from impersonated figures, as observed in 25.6% of incidents in our dataset. In practice, influence actors deploying deepfakes during elections have used various tactics to publish fabricated statements that may be considered extreme or a departure from the corresponding impersonated individual's established political policies or ideologies. Common tactics used in conjunction with false statement deepfakes depicted public figures as doing the following:

- **Falsely criticizing their own party**, such as fake audio of UK Prime Minister Keir Starmer allegedly calling Labour supporters "thick" or UK MP Wes Streeting allegedly calling fellow Labour MP Diane Abbott a "silly woman"
- **Falsely criticizing opposition**, such as likely fake audio of Taiwan People's Party (TPP) candidate Ko Wen-Je accusing opposition candidate Lai Ching-te of visiting the US for a "job interview"
- **Making fake policy announcements**, such as a fake video of Indian Home Minister Amit Shah allegedly announcing the end of reservation rights for Scheduled Castes (SCs), Scheduled Tribes (STs), and Other Backward Classes (OBCs)
- **Making fake endorsements,** such as fake audio of London Mayor Sadiq Khan allegedly dismissing Remembrance Day and endorsing a London pro-Palestine march in November 2023 or Chancellor of Germany Olaf Scholz falsely endorsing a ban on the Alternative for Germany (AfD) political party
- **Other false statements**, such as a Pakistan Tehreek-e-Insaf (PTI) candidate, Muhammad Basharat Raja, allegedly announcing his withdrawal from the 2024 Pakistan general elections

According to researchers from the University of Amsterdam, although deepfakes making implausible false statements and amplifying extreme viewpoints are less likely to be believed as being authentic, they have a stronger effect in delegitimizing impersonated politicians. Influence actors looking to target elections or impersonate specific political figures do not necessarily need to invest in undetectable deepfakes or generate deepfakes of individuals unfamiliar to the general public if the messaging is sufficiently extreme. Even if impersonated figures are well-known or deepfakes are badly made, painting public figures as professing political positions that are significant departures from their known or past political positions can still successfully discredit victims.

Electioneering

Deepfakes are also increasingly becoming a tool used by political parties and candidates for electioneering and attacking the opposition, with the technology's role in elections not being specifically regulated by electoral laws in many cases. Electioneering efforts included encouraging voter participation (for example, by reviving Indonesian leader Suharto) and voter suppression (by using fake audio of President Joe Biden calling on voters to skip a primary election) or depicting candidates as speaking in another language (such as former French presidential candidate Marine Le Pen speaking Russian) or from prison.

In some countries, political parties are using deepfakes to attack opposition candidates overtly. This includes use by incumbents against challenging parties, such as Turkish President Recep Tayyip Erdoğan playing a video deepfake at a campaign rally accusing an opposition rival of being linked to the outlawed Kurdistan Workers' Party (PKK). During Argentina's presidential elections, Sergio Massa's campaign made and amplified a deepfake video of then-candidate Javier Milei discussing organ markets.

Character Assassination

Character assassination remained among the most common influence objectives when impersonating public figures in broader influence operations, and has also been common in deepfakes. Some attempts appealed to social media users' humor (such as showing Ukraine President Volodymyr Zelensky or French President Emmanuel Macron dancing) or portraying victims as engaging in unethical or illegal behavior (such as showing Philippine President Ferdinand Marcos Jr. doing drugs or resurfacing alleged corruption claims targeting Croatian Prime Minister Plenković). Other deepfakes promoted more conspiratorial narratives, such as showing opposition leaders and journalists as being involved in planning electoral fraud in Slovakia or asking Russian President Vladimir Putin about his alleged body doubles in a live interview.

Spreading Non-Consensual Deepfake Pornography to Intimidate and Cause Reputational Harm

In addition to shaping public perceptions and negatively affecting victims' reputations, the spread of non-consensual deepfake pornography impersonating public figures can also inflict significant psychological harm on victims, and deepfakes, in general, can deter women's political participation. Both consequences are likely to have long-term impacts on electoral processes.

Deepfakes can be a gendered security issue; for example, women are disproportionately targeted by non-consensual deepfake pornography. Insikt Group's dataset of deepfakes found that 37.8% of impersonated public figures were women, jumping to 62.5% among impersonated journalists. Women in politics account for 100% of instances of deepfake pornography in our dataset. Insikt Group was unable to identify any examples of journalists or male political leaders being impersonated by pornographic deepfakes in the dataset timeframe, though there was one earlier than this time period: Muharrem İnce in May 2023 (see below). Political leaders and representatives such as Italian Prime Minister Giorgia Meloni, US House Representative Alexandria Ocasio-Cortez, and Dutch Representative Caroline van der Plas have all been impersonated in pornographic deepfakes, in addition to a host of female UK politicians, including Deputy Prime Minister Angela Rayner, former Commons Leader Penny Mordaunt, and former Home Secretary Priti Patel.

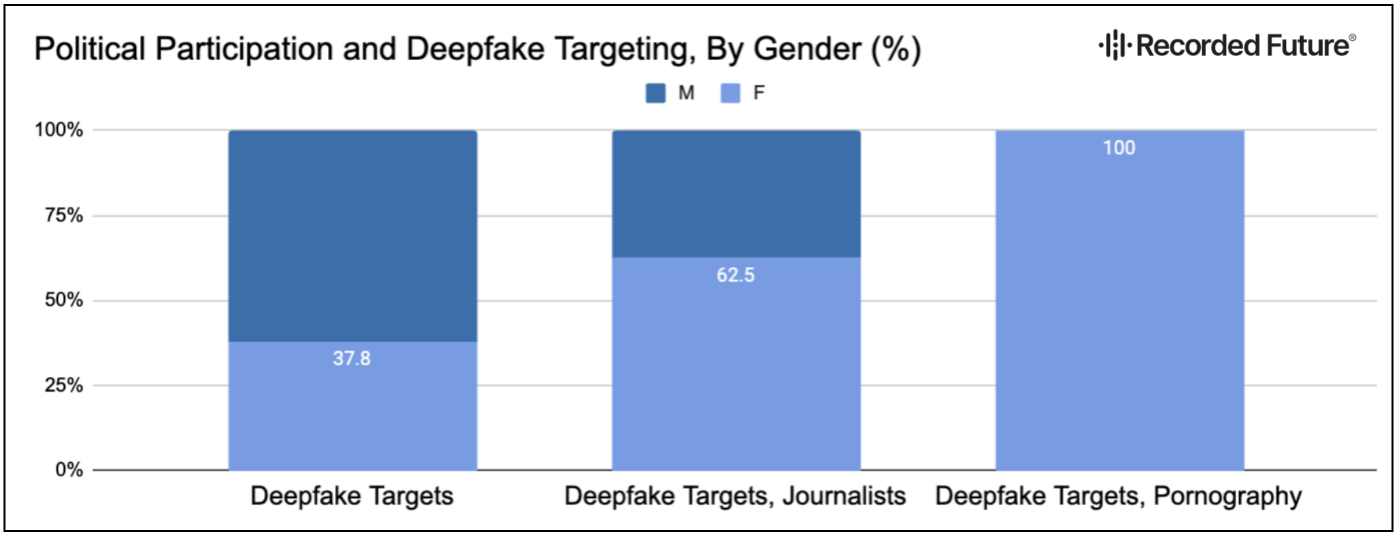


***Figure 12**: Percentage of female public figures targeted by deepfakes, female journalists targeted by deepfakes, and women in politics targeted by deepfake pornography (Source: Recorded Future)*

The strategic dissemination of non-consensual deepfake pornography can also have a direct impact on elections. In May 2023, candidate Muharrem İnce withdrew from the Turkish presidential election following the release of an alleged sex tape, which the candidate stated was likely a deepfake. The dissemination of the deepfake reportedly coincided with "a steep drop in his popularity as polls indicated his garnering less than 2 percent of the vote".

Emerging Tactics

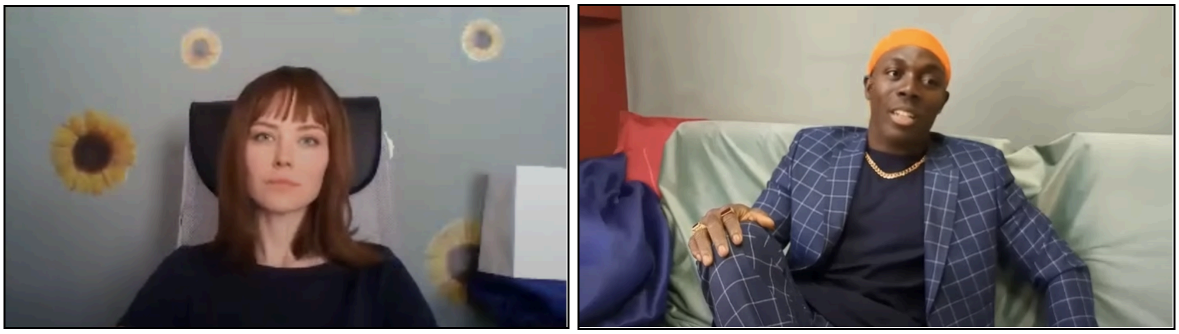
Influence actors, especially well-resourced actors from Russia and China, have begun coupling deepfake videos with new tactics to boost the plausibility and credibility of deepfakes. For example:

- Using anonymous third parties as the subjects of deepfakes allows influence actors to obfuscate their activities from detection by AI tools and social media platforms
- Audio deepfakes are likely enabling actors to target vulnerable users outside of social media, including via phone calls
- Some deepfakes are using media overlays from media outlets as if they originated from authentic sources
- Deepfakes of foreign leaders are being used to promote political parties in a domestic context
- Deepfakes impersonating public figures' family members may also be used to promote or discredit political leaders

Fake Whistleblowers

Rather than producing deepfakes of public figures, influence actors have begun using deepfakes of anonymous third parties or suspected paid actors posing as whistleblowers to make claims of personal or political scandals involving targeted public figures. This tactic allows influence actors to avoid detection efforts seeking to identify deepfakes of known individuals and directly bypass protections for production (on commercial AI platforms) and dissemination (on social media or video-hosting platforms). Using deepfakes of paid actors is also likely to help avoid attribution. Researchers have previously been able to identify actors' real identities and attribute content to known influence actors, such as RT and the Internet Research Agency (IRA).

Storm-1516 is a Russian influence actor that has perfected this technique and has consistently been able to produce and upload deepfakes, which are only detected once disseminated to target audiences rather than being detected by AI production tools or by the video-hosting platforms themselves. Some examples of anonymous third parties, suspected to be paid actors, being used to portray fake whistleblowers or journalists attempting to smear political leaders include an anonymous Bugatti employee used to target First Lady of Ukraine Olena Zelenska, the brother of a fictitious victim of French law enforcement, an anonymous Hamas member threatening the 2024 Paris Olympics, a Nigerian sex worker used to target German Minister of Foreign Affairs Annalena Baerbock, and a fictitious German journalist used to target EU Commission President Ursula von der Leyen.



Figures 13 and 14*: Two Storm-1516 deepfakes using likenesses of anonymous third parties to target European leaders (Source: Recorded Future)*

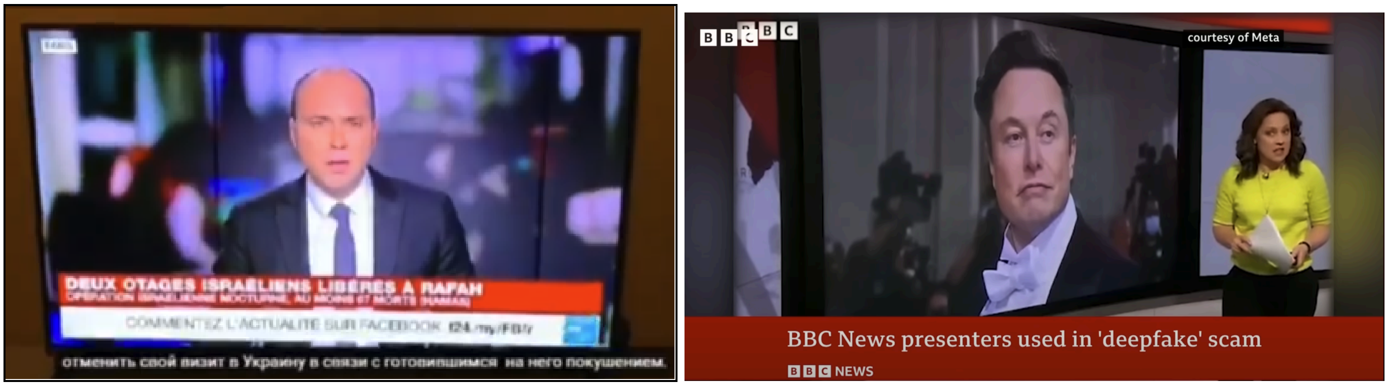
Inauthentic Audio

Audio deepfakes have also been operationalized to impersonate public figures in the last year. While deepfake audio represented only 17% of our overall dataset, deepfake audio clips accounted for 40% of the deepfakes used to generate false statements from impersonated public figures, such as UK Prime Minister Keir Starmer and US Vice President Kamala Harris, both engaged in elections in 2024.

Audio deepfakes are convincing in approximately 25% of cases and are also more likely to be effective on non-native speakers and older listeners. Malign actors are also very likely experimenting with more direct dissemination methods, such as targeted phone calls, which could allow for more granular targeting of older, non-native-speaking audiences. In January 2024, a political consultant used automated calls with deepfake audio of President Biden and spoofed caller IDs to target Democratic voters in a primary election, urging them not to vote. Malign influence actors will almost certainly consider audio deepfakes via phone calls or conferencing software to conduct election interference, from broad voter suppression campaigns to targeted social engineering and intelligence collection efforts.

Spoofing Media Assets

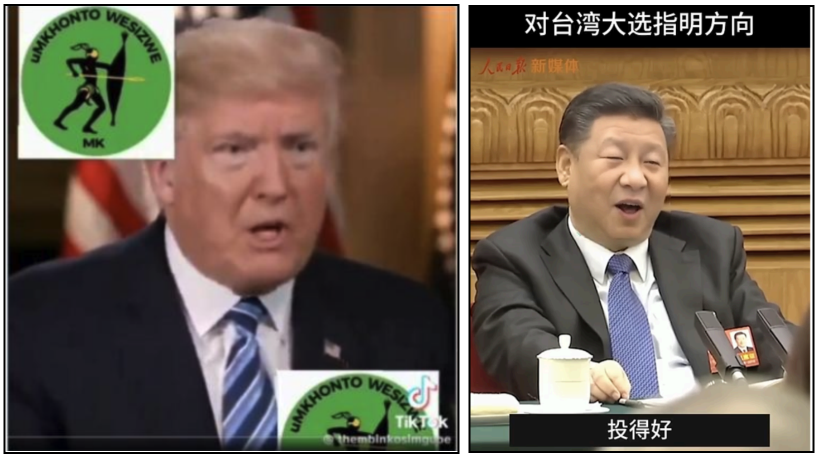
In addition to impersonating journalists and news presenters, influence actors have also started using deepfakes overlaid with media assets (such as logos from credible news sources, news tickers, and jingles), attempting to hijack the credibility of impersonated media organizations' brand identities. In the last year, both France24 and the BBC saw news presenters being impersonated by deepfakes using legitimate media assets.


DEUX OTAGES ISRAÉLIENS LIBÉRÉS À RAFAH
COMMENTEZ L'ACTUALITÉ SUR FACEBOOK f24.my/FB
отменить свой визит в Украину в связи с готовившимся на него покушением.
BBC
courtesy of Meta
BBC News presenters used in 'deepfake' scam
BBC NEWS

***Figures 15 and 16**: Deepfakes impersonating journalists, overlaid with France24 and BBC media assets (Source: CheckYourFact, BBC)*

Impersonating Foreign Leaders

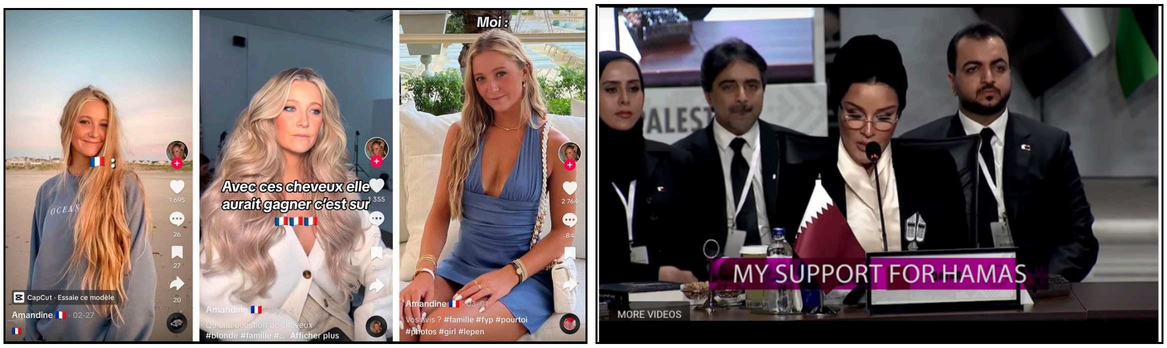
Deepfakes of foreign heads of state were also used to influence domestic elections and endorse specific political parties. In December 2023, a deepfake video of Chinese President Xi Jinping began circulating on TikTok, showing President Xi encouraging Taiwanese citizens to vote. Despite the video initially being meant to parody Xi, Taiwanese users repurposed the video with different captions in support of both mainstream political parties, the Kuomintang (KMT) and Taiwan People's Party (TPP). Former US President Donald Trump's likeness was also used in March 2024 to support uMkhonto weSizwe (MK), a new political party in South Africa, before the country's May 2024 general elections.



***Figures 17 and 18**: Deepfakes of foreign heads of state: (Left) former US President Donald Trump and (Right) China President Xi Jinping (Source: AFP, Taiwan FactCheck Center)*

Impersonating Family Members

In two cases, influence operations involving public figures have also impersonated female family members of political figures as a means to garner support or discredit targets. In April 2024, TikTok videos surfaced deepfakes targeting Marine Le Pen and Marion Maréchal, elected officials in France. The videos purportedly showed fictitious nieces of the Le Pen family using the two officials' faces grafted onto social media influencers, depicting the fake family members as promoting France's Rassemblement National (RN) party and inviting their followers to follow "Lucia Meloni", allegedly another influencer deepfake based on Italian Prime Minister Giorgia Meloni. While these videos were not attributed to any specific malign actors, they likely contributed to the targeting of young voters ahead of the 2024 EU elections. In Qatar, Sheika Moza bint Nasser, chair of the Qatar Foundation and mother to the current Emir of Qatar, Sheikh Tamim bin Hamad Al Thani, was also digitally altered as part of a broader malign influence campaign targeting the Qatari government.


Moi :
Avec ces cheveux elle aurait gagner c'est sur
CapCut · Essaie ce modèle
Amandine
MY SUPPORT FOR HAMAS
MORE VIDEOS

***Figures 19 and 20**: Deepfakes of Marine Le Pen and Marion Maréchal (left) and Sheika Moza (right) (Source: RTS, Jerusalem Post)*

Detection and Response Strategies

Deepfake analysis and mitigation is an emerging field with little consensus or examples of effective mitigation strategies. Most of the mitigation strategies associated with deepfakes impersonating public figures are likely to be centered around effective response strategies post-dissemination, as publicizing their likeness is an integral part of public figures' roles and cannot easily be mitigated.

Assessing Plausibility and Credibility

The first step in responding to deepfakes impersonating a public figure is to assess the plausibility and credibility of deepfake materials, in addition to the overall social, cultural, and political context in which they are deployed. As seen in previous sections of this report, even unrealistic deepfakes that make sufficiently implausible claims can effectively damage a public figure's reputation.

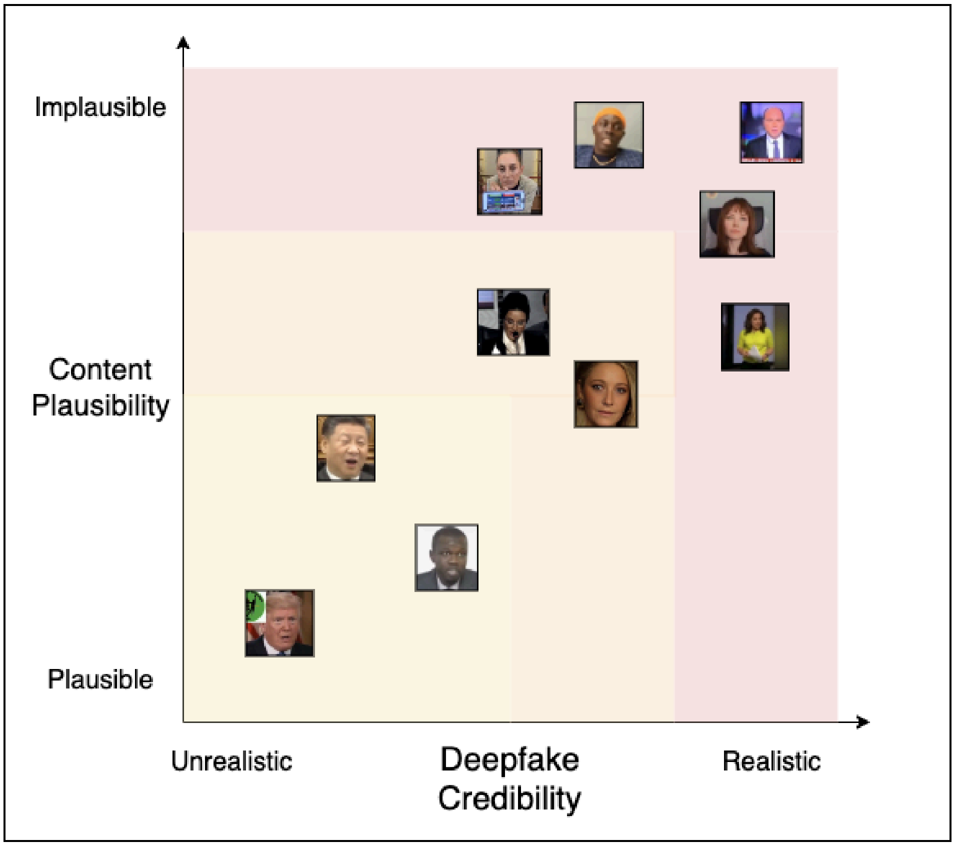


***Figure 21**: Assessed plausibility and credibility of deepfake videos impersonating public figures included in this report (Source: Recorded Future)*

Insikt Group recommends organizations use the following framework to assess the plausibility, credibility, and context of deepfake videos:

- **Plausibility: how plausible are the narratives advanced by the deepfake?**
  - Is the narrative in line with previous positions made public by the impersonated individual?
  - How do the victim's values align with those advanced by the deepfake?
  - What does the narrative promise?
  - How do the potential influencer's values align with the narratives?

- **Credibility: how realistic is the deepfake?**
  - Does the person exist?
  - If so, are users familiar with the person?
  - If not, do users readily believe their purported existence?
  - Is the deepfake speaking the right language?
  - Are there any glaring signs of visual or audio manipulation?
  - Are there any sudden cuts or jumps?
  - Are there any moving objects or hand gestures leading to visual discrepancies?
  - Does the audio match up to the visual?

    - Does the deepfake use any historical clips of the individual?

- **Context: what is the context for the deepfake?**
    - Has the public figure's likeness previously been used in deepfakes?
    - What channels and platforms were the deepfakes distributed on?
    - What entities are promoting the deepfakes? Are they known influencers or scammers?
    - What is the likely influence objective of the deepfake?

Debunking Strategies

Once deepfakes are uploaded and amplified on social media or other sources, organizations looking to protect the reputation of the impersonated individual need to rapidly and effectively counter the narratives promoted by the deepfake. Several debunking strategies have been effective in case studies covered in this report:

- **Act quickly.** Assessing and debunking deepfakes rapidly should be a priority for organizations, which should also cooperate with fact-checkers, social media platforms, and media outlets to take down deepfakes before copies are made.
- **Make users gain familiarity with the impersonated individual.** Making filmed messages of the victim juxtaposed with their deepfake is effective in helping users familiarize themselves with the candidate — this strategy was used by Pakistan Tehreek-e-Insaf (PTI) candidate Muhammad Basharat Raja, who posted a video while holding a smartphone in his hand displaying the deepfake.
- **Take inventory of copyrighted material.** If deepfakes use copyrighted source material, organizations can issue DMCA takedowns. Identify and coordinate with any news or media organizations whose assets were used.
- **Monitor follow-on narratives**. If reported to platforms rapidly, deepfakes can be taken down relatively quickly; however, ensuing narratives can survive on social media for much longer. Monitoring influence narratives post-amplification remains crucial for organizations to understand the impact of deepfakes on overall sentiment toward the impersonated individual.
- **Consider online harms**. If deepfakes are likely to pose severe harm (such as non-consensual pornography or scams), work with appropriate authorities to help identify and take down distribution channels and perpetrators.

Monitoring

- Clients can use Recorded Future's Brand Intelligence Module to track illicit use of organizations' branding and logos.
- Clients can use Recorded Future's Threat Intelligence Module to query and monitor for the availability of datasets of public figures' likenesses ("facesets").

Outlook

Deepfakes will almost certainly be used for illicit purposes and influence operations during the 2024 US elections and other global political and electoral processes. In addition to having a tangible impact on specific electoral outcomes, such as forcing candidates to withdraw from races and durably impacting public figures' reputations, deepfakes erode trust in elections more broadly. Candidates blaming authentic embarrassments and mistakes on deepfakes will likely further erode trust within the electoral process.

Regulating the use of deepfakes remains a difficult task for legislators. While progress has been made in criminalizing non-consensual deepfake pornography, regulations for the use of AI-generated images and deepfakes during campaigns have been applied unevenly across social media platforms and governments. Additionally, even such regulations will almost certainly fail to deter foreign interference and manipulation efforts, which will continue using deepfakes to discredit political leaders during elections.

Research suggests that, above a certain threshold, the quality of deepfakes has diminishing returns in terms of reputational damages inflicted on individuals. Hence, influence actors likely already have the appropriate capabilities to durably impact elections using deepfakes, unlike large language models, which still require a significant margin of improvement before becoming viable in influence operations beyond "back end" operations (such as researching target audiences and writing code) and writing ineffective influence content at scale.

Tracking drivers of development for emerging technologies such as deepfakes is, therefore, less consequential for organizations than developing effective response and mitigation strategies. A key factor to include in such strategies is rapidly exposing social media users to impersonated individuals' likenesses to reduce credibility and consistently fact-check narratives to reduce plausibility. Additionally, organizations should monitor follow-on influence narratives using deepfakes as the basis for broader narratives impersonating public figures.

Recorded Future reporting contains expressions of likelihood or probability consistent with US Intelligence Community Directive (ICD) 203: Analytic Standards (published January 2, 2015). Recorded Future reporting also uses confidence level standards employed by the US Intelligence Community to assess the quality and quantity of the source information supporting our analytic judgments.

About Insikt Group®

Recorded Future's Insikt Group, the company's threat research division, comprises analysts and security researchers with deep government, law enforcement, military, and intelligence agency experience. Their mission is to produce intelligence that reduces risk for clients, enables tangible outcomes, and prevents business disruption.

About Recorded Future®

Recorded Future is the world's largest threat intelligence company. Recorded Future's Intelligence Cloud provides end-to-end intelligence across adversaries, infrastructure, and targets. Indexing the internet across the open web, dark web, and technical sources, Recorded Future provides real-time visibility into an expanding attack surface and threat landscape, empowering clients to act with speed and confidence to reduce risk and securely drive business forward. Headquartered in Boston with offices and employees around the world, Recorded Future works with over 1,800 businesses and government organizations across more than 75 countries to provide real-time, unbiased, and actionable intelligence.

Learn more at recordedfuture.com