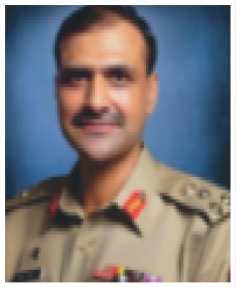
IMPACT OF AI GENERATED DEEPFAKES ON NATIONAL SECURITY



Brigadier
Dr Abdul Rauf
Serving Army Officer

Major
Ehsan Ullah Tarar
Serving Army Officer

Abstract

Spreading falsehoods is as old as mankind itself. It has been used by individuals to conduct frauds and by nation state actors to win wars. But "deepfake" technology represents an exponential increase in the ability to distort reality which is further fuelled by the social media algorithms. Deepfakes use Artificial Intelligence (AI) to generate fabricated audio or video of real people appearing to say or do things they did not. As machine learning methods advance, deepfakes are becoming more realistic and harder to detect. This technology can spread rapidly to many users with varying motives. While deepfakes have some upsides, they also enable significant harms. Our current online "marketplace of ideas" already suffers from truth decay and cognitive biases, which deepfakes will worsen. Individuals and organisations face new forms of exploitation, intimidation and sabotage through deepfakes. There are also profound risks to democracy and national security.

This paper describes 3 pillars that help generate and spread deepfakes including its technological foundations to the role to cognitive bias and search engines' filter bubbles. It provides a thorough assessment of the causes and impacts of deepfakes as a disruptive technological change affecting society. The analysis examines the potential role of deepfakes in education, evaluating possibilities for enhancing learning as well as harms. Additionally, the paper explores the use of deepfakes for fraud, particularly in the banking sector. Deepfakes' ability to shape military narratives and influence operations, especially by nation-state actors, receives attention regarding national security implications.

Towards the end it presents potential responses to mitigate the risks of deepfakes, spanning technological, educational and military domains. Proposed solutions discussed include technical detection tools, penalties for deepfake creation and distribution, military countermeasures, economic policy levers and market-driven adaptations. Based on this analysis, improvements to laws and policies to address deepfakes while noting potential pitfalls have been recommended. Overall, a multilayered toolkit combining technological detective methods, calibrated regulations, economic incentives and legal deterrence emerges as an effective approach to counter malicious uses of deepfakes while nurturing innovation.

Keywords

Deepfakes, National Security, Technical Controls, Next Big Threat, Disinformation, Artificial Intelligence (AI), Algorithms, Profiling, Micro-Targeting, Automated Fact-Checking, Social Media.

Introduction

Deepfakes[1] is a term that was first coined in 2017 to describe realistic photo, audio, video and other forgeries that are generated using Artificial Intelligence (AI) technologies.[2] These technologies make it possible to show events that never happened. These forgeries could pose a variety of national security challenges in the future. As these technologies continue to mature, they could have significant implications for national legislative oversight, defence authorisations and appropriations and the regulation of social media platforms.

While varying definitions exist, deepfakes predominantly describe synthetic media generated through machine learning (ML) techniques, chiefly generative adversarial networks (GANs)-a subfield of AI. GAN frameworks entail two ML models, neural networks, trained in opposition.[3] The first network, the generator, produces fabricated data (e.g., images, audio, video) emulating authentic attributes. The second network, the discriminator, attempts to identify the counterfeit data. Through successive iterations, often numbering in the thousands or millions, the generator refines its outputs seeking to increasingly elude discrimination. Ultimately, the GAN reaches equilibrium where synthetic data attains sufficient verisimilitude to evade detection, thus producing deepfakes. This competitive, self-learning process underpins deepfake creation. Further research into GAN architectures and training methodologies will likely yield critical insights into both deepfake synthesis and detection.

Propaganda has always been a crucial tool in warfare, used to manipulate public opinion, boost morale and demonise the enemy.[4] It is a psychological weapon that can influence the perception and behaviour of individuals, groups and societies. Propaganda can shape the narrative of a conflict, making one side appear as the righteous and the other as the villain. It can also be used to spread fear, uncertainty and doubt among the enemy ranks, thereby weakening their resolve. During World War II, for instance, both the Allies and the Axis powers extensively used propaganda to rally their people and demoralise the enemy.

The concept of propaganda has gained more importance with the advent of 5th generation warfare,[5] which is characterised by the use of information and communication technologies. In this new form of warfare, the battle is not only fought on the physical battlefield but also in the minds of people. The internet and social media platforms have become the new battlegrounds[6] where information is weaponised to influence public opinion and destabilise societies. Fake news, misinformation and disinformation are spread at an unprecedented scale and speed, making it difficult for people to distinguish between truth and falsehood.

The advent of AI and deepfake technology has taken propaganda to a whole new level. Deepfakes are hyper-realistic fake videos or audios created using AI, which can make it appear as if a person is saying or doing something they never did. This technology can be used to create convincing propaganda materials that can further fuel conflicts and tensions. For instance, a deepfake video of a world leader declaring war could spark real-world violence and chaos. As such, deepfakes pose a significant threat to national security and the integrity of information, making it even more challenging to combat propaganda in the era of 5th generation warfare.

Technical Foundations

Impersonating someone digitally keeps getting more realistic and believable. Deepfake technology is at the forefront of this. It uses machine learning to put faces and voices into real videos and audio of people. This lets them make very convincing fakes that look like someone said or did something. Even though deepfakes can be made with permission, they often would not be. This section talks about the technology behind deepfakes and why their spread and impact keeps growing.

The Technology Arms Race for Realistic Media Manipulation

New and advancing technology is making it possible to create very convincing fake videos, images and audio recordings of people, known as deepfakes.[7]

In the past, editing images or videos using tools like Photoshop[8] was detectable with close inspection. Digital forensics experts could often identify changes by searching for inconsistencies. However, the emergence of generative artificial intelligence technology has changed the game. It allows for the creation of highly realistic altered or fabricated media that is difficult to prove as fake.

This technology often employs neural networks[9] for machine learning. Neural networks start out like a blank slate, with connections between nodes that are set randomly. Like the brain strengthening neural pathways through experience, neural networks are trained by processing many examples. With enough diverse training data, the network can build accurate models for creating convincing media. The researchers at the University of Washington developed a neural network tool that alters videos to make it look like someone said something they did not actually say. They demonstrated it with a video of former USA President Barack Obama.[10]

But it goes beyond just neural networks. Generative adversarial networks, or GANs, take things to the next level. GANs use two neural networks - one generator network creates simulated media and the other discriminator network tries to detect if it's fake. The generator keeps improving to fool the discriminator. This iterative adversarial process results in highly realistic fabricated images, videos and audio.

Technology Access

The ability to create realistic deepfake videos, images and audio will not remain solely in the hands of experts and responsible groups. Deepfake technology is guaranteed to spread rapidly beyond its original circles of specialised researchers and companies. As past experience shows, technologies tend to proliferate over time, even risky or dangerous ones. For example, firearms meant only for state militaries are now available to regular citizen.

New technologies mainly stay contained if they require rare physical resources that are restricted. Nuclear weapons need scarce radioactive materials like enriched uranium or plutonium that are difficult to acquire. But deepfakes do not rely on rare physical materials has they prefer intangible knowledge and data. And knowledge spreads quickly in the interconnected digital age unless deliberately kept secret and restricted. However, the creators of deepfake technology often have incentives to publish their methods or commercialise them. The academic culture rewards sharing innovations openly. Companies want to profit from apps and services. These motivations run counter to keeping the technology contained. The fundamentals of deepfakes like AI techniques are already public knowledge. Tools are emerging that make it easy for anyone to generate deepfakes.

The Recipe for Disaster

The rise of deepfake technology comes at a dangerous time. Individuals can now reach huge audiences directly, bypassing traditional media gatekeepers. This section explores how psychological tendencies and technical dynamics accelerate the spread of harmful deepfakes.

Few years ago, distributing fake images, videos or audio was difficult for most people or groups. A few major media organisations controlled national and global broadcasts and publications. They acted as gatekeepers, though governments and advertisers exerted influence on content.[11] Today, countless platforms allow anyone to publish content that can go viral globally. Social networks blend one-to-many and many-to-many communication, democratising access to audiences. Overall gatekeeping is reduced, though companies moderate some illegal or egregious content based on terms of service, legal mandates, or public pressure. But generally, there is far less oversight on accuracy or suppression of legal-but-harmful material compared to traditional media. This means dubious content can easily reach and spread among online audiences.

Two dynamics explain the viral reach: information cascades[12] and filter bubbles.[13] Information cascades happen when people focus too much on what others appear to know rather than their own information. Enough individuals pass along doubtful claims without verification, lending them false credibility. The cycle reinforces itself on social networks. Plus, we instinctively share novel, negative information-making falsehoods temping clickbait. Studies show fabricated news spreads faster and wider than truth on social media. Once lodged in people's minds, lies stubbornly persist. Psychology shows we remember negative claims more than positive ones.

Figure-1: FIlter Bubble: Serves you what you like

Source: https://www.sciencespo.fr/public/chaire-numerique/en/2023/06/06/student-essay-warnings-from-the-us-the-relevance-of-filter-bubbles-in-polarized-countries/

Search engines and social platforms intensify our fascination with emotional content such as gossip, violence and sex. Filter bubbles reinforce our worldviews, while algorithms personalise feeds with agreeable information. These biases and technical dynamics create ideal conditions for deepfakes to spread virally, flourishing in the internet's fertile environment for fabricated and manipulated content.

The Multiverse of Deepfakes: Applications across Sectors

Much like a coin has two sides, deepfake technology has positive applications as well as more concerning misuses. On the one side, deepfakes could allow for more seamless educational experiences, creative filmmaking and empowering people to better relate and tell their stories. The same algorithms that synthesise fictitious media also have immense promise for enhancing and personalising digital interactions. However, the other side highlights deepfakes' capacity to unfairly defame people, manipulate evidence and erode public trust on a mass scale. Just as a computer expert's skills can be channeled into strengthening cybersecurity or exploiting vulnerabilities unethically, the proliferation of deepfakes comes with risks despite the innovation. Ultimately, with sound policies and public awareness, societies can reap the benefits of deepfakes while mitigating misinformation that threatens institutions and values. Like any powerful technology, the impact depends not on the techniques alone, but how humanity chooses to wield them.

While deepfake creation has gained popularity for entertainment, such as inserting actor Nicholas Cage into films[14,15] where he did not originally appear or generating interactive exhibits with Salvador Dalí, more constructive applications also exist. Within healthcare, early research indicates promise using generative adversarial networks (GANs) to synthesise fictitious medical images.[16] These synthetic inputs allow training machine learning models to detect rare diseases, while also preserving patient privacy. Such benign use cases suggest while deepfakes present risks, their capabilities may also enable valuable innovations if guided ethically. Further interdisciplinary study between computer scientists, ethicists and domain experts will likely yield critical insights on responsibly harnessing deepfake technologies for social benefit. The ability of deepfakes to infiltrate multiple sectors and deceive individuals on a large scale threatens to undermine public trust, destabilise governments and compromise the very foundation of national security. Here are some direct and indirect implications of deepfakes on National Security.

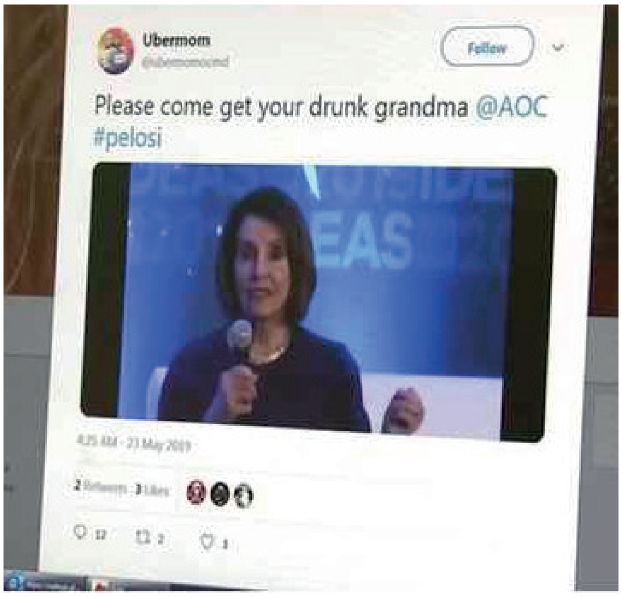
Social Fabric

Deepfakes pose a serious challenge for social cohesion, as they can erode trust, credibility and accountability.[17] Deepfakes can create a reality distortion field that can manipulate the perceptions, beliefs and behaviours of the target audience and can undermine the effectiveness, legitimacy and integrity of the actors involved. These can manipulate, polarise, or divide the social fabric of a society by creating false or misleading information or narratives that could affect the perceptions, beliefs and behaviours of the public. For example, deepfakes could be used to:-

- Create fake news, such as by using deepfake videos of leaders, officials, or experts, that could misinform, mislead, or persuade the public about important issues, events, or policies. For instance, in 2019, a deepfake video of US House Speaker Nancy Pelosi was circulated online, showing her slurring her words and appearing drunk.[18] The video was intended to discredit her and influence the public opinion about her political agenda.

- Spread rumors, such as by using deepfake images, videos, or audio, that could create or amplify scandals, controversies, or conspiracies about individuals, groups, or organisations. For example, in 2019, a deepfake video of Facebook CEO Mark Zuckerberg was posted on Instagram, showing him boasting about his power and control over billions of people's data.[19] The video was created by artists as a protest against Facebook's policies on deepfakes and misinformation.

Figure-2: Videos posted to social media call into question Speaker Nancy Pelosi's sobriety. Her slurred speech and slow movement are the results of a manipulated video



Ubermom

Please come get your drunk grandma @AOC #pelosi

4:25 AM - 23 May 2019

2 Retweets 3 Likes

12 2 3

- Foment hatred, such as by using deepfake videos of celebrities, influencers, or activists, that could incite or provoke violence, discrimination, or extremism against certain communities, cultures, or ideologies. For instance, in 2018, a deepfake video of former US President Barack Obama was created by comedian Jordan Peele, showing him making outrageous statements and warning the public about the dangers of deepfakes.[20] The video was created as a public service announcement to raise awareness and educate the public about the threat of deepfakes.

- Erode trust, such as by using deepfake audio or text, that could undermine or damage the credibility, reputation, or relationship of individuals, groups, or organisations. For example, in 2014, a deepfake audio of Turkish President Recep Tayyip Erdogan[21] was leaked, showing him ordering his son to hide millions of dollars in cash. The audio was later revealed to be a fake, but it sparked protests and accusations of corruption against President Erdogan.

Electoral Process

The emergence of deepfake technology presents novel and profound challenges for the integrity of democratic elections. As synthetic media achieves greater sophistication, the ability to generate falsified yet credible imagery, audio and video of political figures could be weaponised to manipulate public opinion and erode trust in electoral processes.

The core vulnerability arises from deepfakes enabling the widespread dissemination of fabricated misinformation to the voting populace. Deceptive fakes impersonating candidates or public officials could be used to promote false narratives, scandals, or controversies intended to damage their electability and credibility. Without rapid validation methods, the uncertainty surrounding the authenticity of questionable media itself seeds distrust, even if fakery is never conclusively proven. This distorts the information landscape and undermines voters' ability to make informed decisions grounded in facts.

The implications of unchecked deepfakes being leveraged to influence election outcomes extend far beyond the political realm. The integrity of the democratic process itself risks being compromised through technological means without appropriate safeguards. Further, deceptive synthetic media attribution could also be used to fabricate provocative statements or actions by elected officials. This raises the specter of manufactured pretenses for domestic unrest or severed international relations if deepfakes are mistaken as genuine.

Deepfakes in Military

Deepfakes are a new tool for deception in warfare and intelligence operations. They may be especially powerful in shaping public narratives but can also directly enable attacks or strategic policy shifts.

The emergence of deepfakes, with their unprecedented capacity to synthesise realistic audiovisual media, represents a watershed moment for the future of military engagement.[22] On one hand, deepfakes could usher in a new era of highly effective training simulations, strategic deception operations and augmented visualisation to give militaries an edge. The same generative AI that can conjure deepfakes may also boost situational awareness and human-machine teaming. However, deepfakes also open the door to unbridled psychological operations, communications disruption and falsified evidence of atrocities on scales yet unseen. Security experts have sounded alarms about "reality hacking" through weaponised deepfakes which could destabilise geopolitics and undermine democratic values. While deepfakes may enhance some military tactics, their unchecked use by state and non-state actors could inflict deep wounds difficult to repair on the world's social fabric. More than any technological revolution, it will be choices shaped by ethics and wisdom that determine whether deepfakes propel progress or chaos between nations.Some potential uses of deepfakes in military operations are described in proceeding paragraphs:-

- **Positive Applications**
  - **Training Simulations.** Deepfakes could create highly realistic training environments to prepare soldiers for diverse scenarios.
  - **Strategic Deception.** Safe uses of deepfakes could mislead adversaries and gain tactical advantages, akin to traditional camouflage and misdirection.
  - **Enhanced Visualisation.** Generative models could create 3D battlespace visualisations from limited sensor data to improve situational awareness.
  - **Remote Presence.** Deepfake avatars could allow commanders to interact remotely with a human-like presence for meetings, speeches, etc.

- **Negative Applications**
  - **Psychological Operations.** Deepfakes could broadcast demoralising messages seemingly from enemy leadership or spread misinformation among opposing forces.
  - **Impersonation.** Non state actors may impersonate military leaders with deepfakes to cause confusion in command structures.
  - **Blackmail/ Coercion.** Realistic deepfake media could be used to extort or influence key decision makers.
  - **Evidence Tampering.** Deepfakes could falsify images or videos of military operations, massacres, weapons systems, etc.
  - **Infrastructure Disruption.** Critical systems like communications could be compromised using deepfaked biometrics, surveillance footage, etc.
  - **Propagandised Media.** State media may leverage deepfakes to exaggerate victories, deny battlefield losses, or justify aggression.

Education Sector

Deepfakes can have both positive and negative impacts on the education sector, depending on how they are used and regulated. On the one hand, deepfakes could be used to enhance education standards by enabling almost real simulations that could enrich the learning experience, engage the students and foster creativity and innovation. For example, deepfakes could be used to:-

- Create interactive content that allows historical figures to speak directly to students and deepen the connection between the past and the present.

- Generate realistic scenarios that enable students to practice their skills and knowledge in various fields and disciplines, such as medicine, law, or engineering.

- Produce personalised feedback that adapts to the needs and preferences of each student and motivates them to improve their performance.

On the other hand, deepfakes could also distort, mislead, or corrupt the education sector by creating false or misleading information or narratives that could affect the quality, integrity and credibility of education. For example, deepfakes could be used to:-

- Produce fake lectures spreading false, biased, or harmful information to mislead students on key issues.

- Pretend as teachers, risking security, privacy, or reputation, exploiting students for personal gain.

- Cheat in examinations, undermining fairness, validity, or ethics, violating academic integrity.

Banking Frauds

Deepfakes pose a serious threat to digital integrity, distorting reality to manipulate audience perceptions and undermine trust. They can facilitate fraud by creating fake documents, impersonating victims, or forging signatures, impacting both organisations and individuals. Examples include payment fraud, email hacking, identity theft and biometric spoofing. Deepfakes amplify traditional fraud schemes and enable new forms of deception. Some notable banking frauds are:-

- In 2019, a UK energy company sent €220,000 to a scammer who used deepfake voice technology to sound like the CEO of its German[23] parent company. The scammer phoned a UK executive, demanding an urgent payment to a Hungarian supplier. Believing it was the German CEO, the UK executive complied. When asked for more money later, the exec got suspicious, confirming with the real CEO. This marked the first AI voice deepfake scam.

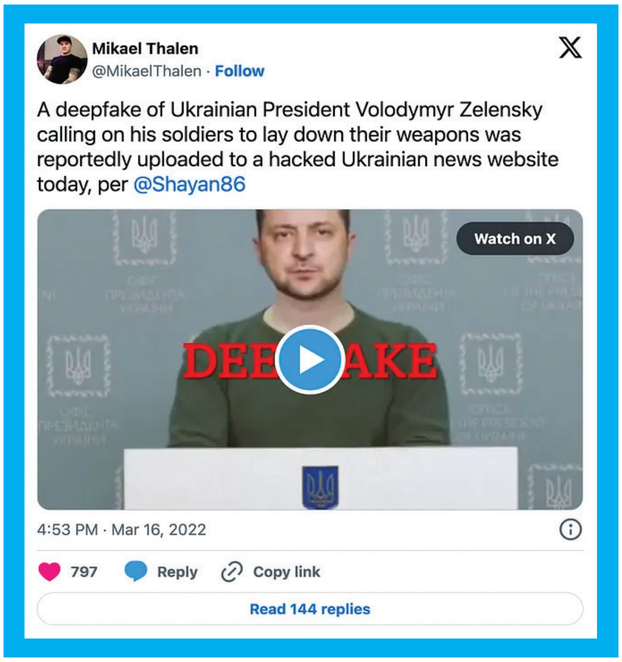
- In 2020, a fake video of a Canadian entrepreneur endorsing a cryptocurrency scam surfaced on YouTube. It aimed to deceive investors into a bogus website, seeking personal and financial details. The scam used altered visuals and voice from a public video. Though YouTube removed it later, thousands had already seen it.

- In 2021, a scammer used a deepfake voice to impersonate a US bank executive, bypassing voice biometric security to enter a customer's account. Pretending to be the executive, he called customer service, successfully resetting the account's password and security questions. With access granted, he transferred funds elsewhere.

Figure-3: Anonymous hacker programmer uses a laptop to hack the system in the dark. Creation and infection of malicious virus

National Security

The potential for exploiting deepfakes to enable disinformation campaigns and influence operations bears consideration. State actors or individuals could spread synthesised footage of public figures, eroding trust and sparking societal discord. U.S. intelligence found Russian operations during the 2016 election aimed to undermine democracy and hinder Clinton’s candidacy. Similarly, in March 2022, Ukraine’s President Zelensky confronted deepfake footage undermining trust. Developing strategies to detect deepfakes and build public resilience against disinformation is crucial.



Mikael Thalen
@MikaelThalen · Follow

A deepfake of Ukrainian President Volodymyr Zelensky calling on his soldiers to lay down their weapons was reportedly uploaded to a hacked Ukrainian news website today, per @Shayan86

4:53 PM · Mar 16, 2022

Exploiting deepfakes for extortion or compromising public figures is concerning. Foreign intelligence has used synthetic media to create fake social media accounts for recruitment. Similar techniques could produce convincing yet inflammatory footage, like military misconduct, to incite violence. In USA, Section 589F of the FY2021 National Defence Authorisation Act mandated an intelligence assessment on deepfake risks, highlighting the need to protect officials, personnel and classified information from coercion. Research into detection methods and resilience against synthetic media threats is vital for informed policies and programmes.

The rise of deepfake technology may give way to what Professors Danielle Keats Citron and Robert Chesney call the “Liar’s Dividend.”[24] This idea suggests that as synthetic media becomes more prevalent, individuals might dispute genuine footage, especially if it depicts unethical behaviour, by claiming it’s a deepfake. As awareness of deepfakes grows, so does the potential for this tactic. Citron and Chesney warn this could encourage deepfake use for deception and avoiding accountability. Further research on the impacts of deepfakes on truth-seeking processes could inform detection methods and institutional reforms to counter the Liar’s Dividend.

Evidence suggests deepfake tactics may already be deployed for political ends. For instance, opponents of Gabon’s President Ali Bongo alleged video footage demonstrating his health was synthetically generated.[25] They later cited this as justification for an attempted coup, despite outside experts being unable to conclusively determine the video’s authenticity. As one expert observed, the uncertainty itself proved sufficient to “undermine credibility and cast doubt,” irrespective of veracity. This exemplifies how merely the potential for deepfakes, whether utilised or not, can enable the Liar’s Dividend by fostering distrust in genuine information.

Some Notable Misinformation Operations

A University of Essex report warns of deepfakes’ threats to democracy, human rights activists and journalists and potential voter manipulation. It notes risks like generating inflammatory content to radicalise populations or incite violence. Deepfakes may also embarrass officials or blackmail individuals with access to classified information. There is already evidence that, foreign intelligence has used deepfakes to create fake social media accounts for source recruitment.

Some notable incidents involving deepfake misinformation operations include:-

- **Gabon’s Failed Coup-2018**

  In 2018, a deepfake video of Gabon’s President Ali Bongo was circulated to suggest that he was in poor health after not being seen in public for months.[26] This fueled uncertainty during an attempted coup. Released by the government, it stirred doubt among citizens. Critics remained skeptical. The military’s failed coup, one week later, cited the video as evidence of the president's instability. Despite speculation, experts deem it a deepfake.

Figure-4: Army officers appeared on national television today to say they have seized power in Gabon

- **Myanmar Political Unrest-2020**

  In 2020, a deepfake video of Myanmar's leader Aung San Suu Kyi circulated on social media, purportedly announcing the country's constitution's revocation, raising concerns of unrest. Shown during a junta spokesman's press conference in Naypyitaw, it was swiftly debunked but caused alarm. This incident underscores deepfake dangers, spreading misinformation and emphasises government action necessity for protection.

Figure-5: Regardless whether video is a forced confession or a deepfake; results are most likely the same

- **US Presidential Election-2020**

  In the 2020 US Presidential Election, deepfake videos of candidate Joe Biden surfaced, depicting him as confused or frail, aiming to discredit his campaign. These incidents highlighted deepfake's political impact, stressing the necessity for verification tools and digital literacy to counter misinformation and manipulation of public opinion.

- **War Crimes Depiction**

  Few state media is suspected of using deepfake technology to spread disinformation about conflict, sharing videos showing fake war crimes by soldiers. This further escalates tensions and misleads public. It highlights deepfake's dangerous role in information warfare, emphasising the need for verification tools and digital literacy to discern truth.

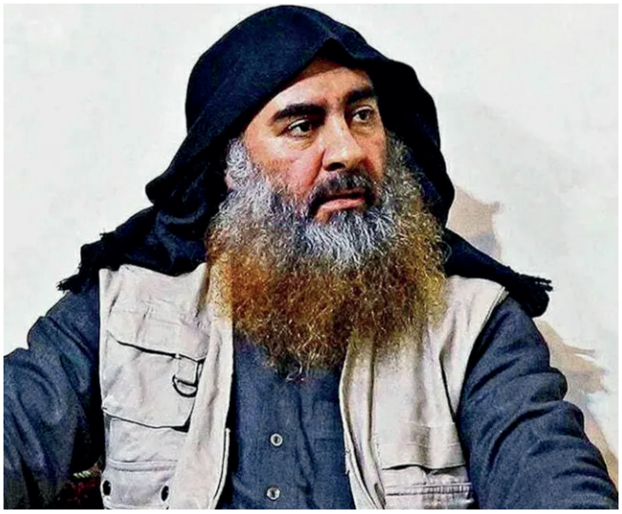
- **Masses Confusion and Dissent in North Korea**

  North Korean state media reportedly used deepfake technology to fabricate videos of South Korean politicians praising North Korea. This aims to sow confusion and distrust, undermining South Korean leaders and distorting public perception. The incident highlights deepfake's political misuse, urging for detection tools and public awareness to counter digital manipulation's potential harm.

- ### Deefakes Use by ISIS

  ISIS have utilised deepfake technology to create videos of politicians claiming conversion to Islam, aiming to demoralise foes and rally supporters.[27] If these videos are believed, they could significantly impact public sentiment and political stability. This misuse of deepfakes can sway public opinion and destabilise politics, emphasising the technology's broader societal and geopolitical risks.

Figure-6: Abu Bakr al-Baghdadi was killed, according to US President Donald Trump



- ### Pornography Generation

  Deepfake technology, though innovative, is often misused, especially against women, through non-consensual pornography where faces, often of celebrities, are imposed onto explicit content. This digital violation harms victims' privacy, safety and mental health, leading to distress and even extortion threats. Predominantly targeting female celebrities, this underscores gender inequality and the urgent need for legal and technical solutions. In 2019, the Deeptrace report revealed that 96% of online deepfake videos are pornographic, further emphasising the severity of the issue.[28]

Combating Deepfakes

To tackle deepfake risks, a multifaceted strategy is needed. Firstly, invest in technology for detecting and debunking fake media. Secondly, adjust legal liability to deter malicious deepfake creation. Thirdly, enforce regulations by governments and platforms to address systemic risks. Fourthly, cautiously deploy state countermeasures against harmful deepfakes. Lastly, market-driven solutions can offer protection, but privacy concerns must be considered. These combined efforts can mitigate threats while fostering beneficial innovation. Government measures to combat deepfakes include:-[29,30]

- ### Detecting Deepfakes

  As technology progresses, so do the methods for identifying fabricated explicit videos online. Governments, researchers and technology firms invest in methods like reverse engineering to authenticate content. Social media, deepfake and explicit content platforms deploy strategies to remove non-consensual content, especially involving public figures. This is done to curb the dissemination of such harmful explicit content. For instance, tools on the DeepSwap website and regulated policies prevent users from creating explicit content involving unsuspecting individuals without their prior consent. Here are some individual methods for identifying deepfake explicit videos:-[31,32]

  - **Unusual Eye Movement**. These videos often feature unnatural eye movements, particularly in terms of blinking and the positioning of the character's eyes, which can appear odd and unrealistic.

  - **Inconsistent Body Movement**. Some deepfake videos depict the character's body distorting when they move, indicating that the video may be fabricated and potentially non-consensual.

  - **Other Indicators**. Other signs could include unnatural hair and awkward positioning of facial features, which suggest that a video or photo is synthetic explicit content:-

    - Faces making expressions that do not look right.
    - Not blinking at all.
    - Bodies shaped in a way that does not look natural.
    - Faces changing shape, like one picture is just put on top of another.
    - Heads not staying in the same place.

- Skin colors that do not look right.
- Hair that does not look natural.
- Heads and bodies positioned awkwardly.
- Voices that sound like robots.
- Lighting or colours that look strange.
- Mouths not moving in synchronisation with the words being said.
- Pictures that are blurry or do not line up right.
- Noise in the background of the picture.

- **Media Literacy for Consumers**

  Media literacy is crucial in fighting disinformation and deepfakes. Public awareness campaigns and education programs can enhance critical thinking and media literacy skills, teaching people to recognise altered media, spot mismatched speech patterns, cross-reference sources and use reliable fact-checking resources. An informed citizenry is the most effective defence against malicious deepfakes.

- **Meaningful Regulations**

  Governments, along with technology, civil society and policymakers, can create regulations to discourage harmful deepfakes. Laws holding creators accountable, supported by technology ethics, deter misuse while preserving innovation. Yet, regulation may stifle beneficial uses or push activities underground. Open discussions among stakeholders can shape policies to target bad actors without overreach.

- **Easy-to-Use and Accessible Technology Solutions**

  Governments can fund user-friendly technology to detect deepfakes, authenticate media and promote authoritative sources. Simple apps and browser plugins can warn users of misleading content before sharing. Official validation from reputable sources can curb the spread of falsified materials. Prioritising trusted institutions and fact-checkers' visibility is crucial.

- **Social Media Policies**

  Social media platforms address deepfakes with policies and terms of use. Governments can collaborate with them to curb deepfake spread and enforce accountability. Guidelines banning malicious deepfakes are a start, but platforms may need proactive measures like detection tools and promoting credible information during high-risk events. Social media firms, as major distribution channels, can lead in deepfake mitigation.

Recommendations

As deepfake technology grows more advanced, it is imperative that we develop comprehensive strategies to counteract malicious uses. A multifaceted approach is required, combining technological tools, legal deterrents, market innovations and policy changes. Specific recommendations include:-

- Invest in robust detection technologies like artificial intelligence systems to authenticate media and expose manipulations. This will enable platforms and watchdog groups to identify deepfakes.

- Carefully expand civil liability for creators and distributors of harmful deepfakes that cause reputational, emotional, or financial damage. This will deter malicious actors.

- Consider amendments to legal frameworks that encourage platforms to address deepfakes under a "reasonable" standard, while protecting free speech.

- Use sanctions, military options and covert actions to impose costs on foreign actors using deepfakes against national interests. A deterrence posture is needed.

- Encourage market-driven services like immutable logs and platform speech policies to combat deepfakes while ensuring privacy protections. Market solutions have promise but require oversight.

- Fund academic research and education programs to develop technical and societal resilience against deepfake harms. An informed public is empowered against manipulation.

- Launch public awareness campaigns by government and platforms to inoculate citizens against trusting fake content. Media literacy is critical.

- Have governments share information rapidly with platforms during elections and crises to help flag fake content. Collaboration can limit viral spread.

Conclusion

Despite the conventional wisdom that “sticks and stones may break my bones but words will never hurt me,” falsehoods in their various forms have long posed significant dangers to individuals, institutions and society as a whole.

In conclusion, deepfake technology presents novel capabilities for deception and manipulation with significant implications for national security. As demonstrated across the documents reviewed, deepfakes enable the realistic fabrication of audio, video and imagery to support propaganda, psychological operations and disinformation campaigns. State and non-state actors alike may exploit deepfakes to influence politics, undermine public trust and exacerbate social divisions.

Despite the profound risks, deepfakes also offer some constructive applications that should not be overlooked. Carefully regulated use of synthetic media for training, analysis and operations can benefit defence, technology and more. Balancing security with innovation will require a holistic response engaging technologists, lawmakers, platforms, academia and the public. Detecting and attributing deepfakes through technical and open-source means is an imperative. Legal and normative guardrails against harm must be erected while protecting free expression. As deepfake capabilities spread, democratise and evolve, sustained research, vigilance and cooperation across sectors is essential to ensure security, accountability and faith in information itself.

NOTES

1. Karnouskos, “Artificial Intelligence in Digital Media.”
2. Kelley and A. Harris, “Deep Fakes and National Security - CRS Reports.”
3. Donahue, “SD-GAN Tensorflow.”
4. “Psychological Warfare | Propaganda, Mind Games & Tactics | Britannica.”
5. “Fifth-Generation Warfare.”
6. Swift, “CloudTweaks | Social Media A New Military Battleground.”
7. Karnouskos, “Artificial Intelligence in Digital Media.”
8. “Official Adobe Photoshop - Leading AI Photo & Design Software.”
9. Donahue, “SD-GAN Tensorflow.”
10. “Obama Deep Fake.”
11. Rheingold, Smart Mobs.
12. “Networks, Crowds, and Markets: A Book by David Easley and Jon Kleinberg.”
13. Pariser, The Filter Bubble.
14. “The Power of DeepFake Technology in Today’s World.”
15. “Deepfake Technology in the Entertainment Industry.”
16. Skandarani, Jodoin, and Lalande, “GANs for Medical Image Synthesis.”
17. “How Deepfakes Erode Trust - Jason Thacker.”
18. “Doctored Nancy Pelosi Video Highlights Threat of ‘Deepfake’ Tech - CBS News.”
19. Cole, “This Deepfake of Mark Zuckerberg Tests Facebook’s Fake Video Policies.”
20. Vincent, “Watch Jordan Peele Use AI to Make Barack Obama Deliver a PSA about Fake News.”
21. “Turkish PM’s Office Says Erdogan Recordings Are Fake.”
22. Biddle, “U.S. Special Forces Want to Use Deepfakes for Psy-Ops.”
23. “Unusual CEO Fraud via Deepfake Audio Steals US$243,000 From UK Company - Noticias de Seguridad.”
24. Schiff, Schiff, and Bueno, “The Liar’s Dividend”; Sonnemaker, “‘Liar’s Dividend.’”
25. Cahlan, “Analysis | How Misinformation Helped Spark an Attempted Coup in Gabon.”
26. Lagos, “Gabon Coup.”
27. Rivers, “ISIS to Bring Al-Baghdadi ‘back from Dead’ Using Deepfake Tech to Fool US.”
28. Blair, “Deepfake Videos Online ‘almost All Porn’ as 15,000 AI-Generated Clips Found.”
29. “How Can We Combat the Worrying Rise in Deepfake Content?”
30. “The Emerging Threat of Deepfake Technology.”
31. “PM Modi Warns Deepfakes Could Wreak Havoc.”
32. “Deepfakes.”