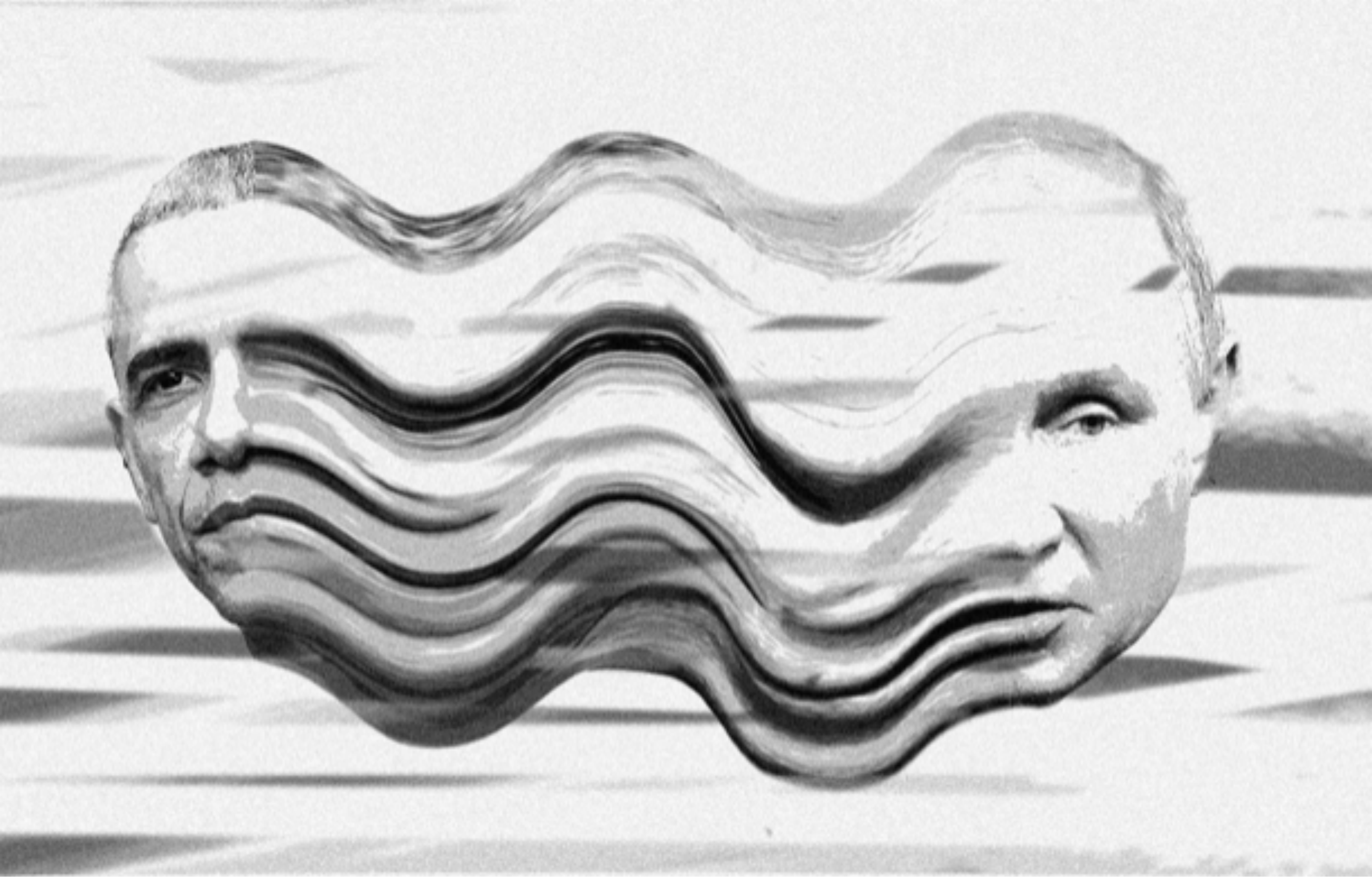

Brief No. 12.4

REPLICATING REALITY

Advantages and Limitations of Weaponized Deepfake Technology

Megan Hogan

P | I | P | S

Replicating Reality:
Advantages and Limitations of Weaponized Deepfake Technology

APRIL 2020

Megan Hogan

P | I | P | S
The Project on International Peace and Security
Global Research Institute
College of William & Mary

Replicating Reality:
Advantages and Limitations of Weaponized Deepfake Technology

Deepfakes are a form of synthetic media that use artificial intelligence to produce highly realistic, fake videos. Deepfakes are extremely effective weapons of disinformation capable of both undermining trust in institutions and elections and inciting political violence. By the end of 2020, virtually undetectable deepfakes will be a reality.

The United States faces a choice. The Department of Defense can either continue to restrict its research to developing video authentication algorithms or expand its effort to include deepfake weaponization for coercive diplomacy and warfighting. Each option has benefits and costs. Ultimately, the United States should develop weaponized deepfake technology as a capability to deny, defeat, or defend against any adversary that seeks to harm U.S. national interests—even if this capability is never used.

Introduction

Deepfakes are highly realistic fake video and audio recordings generated by an artificial intelligence technique called "deep learning"—hence the name "deepfake."[1] Deepfakes first appeared on the messaging board site Reddit in 2017, when an anonymous user transposed the faces of female celebrities onto the bodies of adult film stars.[2] Though Reddit deleted the videos and suspended the user who produced them, deepfake technology remains publicly available online today.

While it has been possible to manipulate video footage for decades, such manipulation has been easily detectable, time consuming, and expensive. The power and peril of current deepfake technologies are that they lower the cost of disinformation. Previously, actors needed a high degree of technical expertise to produce a realistic fake video of someone. Today, all an actor needs are a handful of photographs and an internet connection.[3]

The United States has an interest in expanding its current deepfake detection efforts to include deepfake weaponization for defense and warfighting. As a military tool, deepfakes can provide novel ways to intimidate and sabotage adversaries, disrupt enemy lines of communication, and influence adversary decision-making. As a weapon of disinformation, deepfakes can compel adversaries to comply with U.S. demands, manipulate public opinion to shape political events, and destabilize political regimes. Though the development and use of weaponized deepfake technology are not without risk, the United States can take measures to mitigate the dangers associated with weaponization.

Given the inherent risks associated with deepfakes, the United States should use this technology only during combat operations against state and non-state adversaries. Absent war, the United States should weaponize deepfakes to improve its existing deepfake detection capabilities. If the

United States foregoes weaponization, it will deny itself use of a potent weapon and both its offensive and defensive capabilities will fall behind.

The Rise of Deepfakes

What we see depends mainly on what we look for.
– Sir John Lubbock, 1892[4]

Humans rely on their sense of sight and hearing to identify threats and form their view of reality. The advent of deepfakes and synthetic media, however, may soon make it impossible to use these senses to distinguish fact from fiction.

Deepfake Technology

Deepfake video production begins by feeding a series of photographs into software called a "neural network."[5] The neural network then makes statistical connections between the visual appearance of the photographs and whatever output a forger desires. If, for example, a forger wants to produce a video of former U.S. president Barack Obama delivering a fabricated speech, she will direct a neural network to learn the associations between particular words and the shape of Obama's mouth as he says them.[6] If a forger wants to affix Obama's face onto another person's body, she will direct a neural network to learn the associations between Obama's face and the other person's body.[7] However, if a forger wants to produce a unique image, rather than combine existing photos together, she can use a generative adversarial network (GAN)—a deep-learning technique composed of two competing algorithms. The first algorithm, the "generator," creates a deepfake from the statistics it was given in its training dataset.[8] The second algorithm, the "discriminator," evaluates whether the deepfake is real by comparing it to the training dataset and returning a probability of authenticity. If the discriminator yields "1," it predicts the deepfake is a real video. If the discriminator yields "0," it predicts the deepfake is fake. Researchers can then use the discriminator's prediction to modify the original deepfake, eliminating any signs of forgery.

Forgers also can use GANs to generate deepfake audio recordings. WaveGAN—the application of GANs in unsupervised raw-waveform audio synthesis—uses a training dataset of an individual's audio recordings to synthesize new speech.[9] Like visual GANs, WaveGANs are composed of two competing algorithms: a generator algorithm, which produces a convincing audio deepfake, and a discriminator algorithm, which evaluates the audio deepfake's authenticity. With only 40 minutes of audio recording, WaveGAN can generate a nearly perfect copy an individual's voice.[10] Even so, audio deepfakes are still largely limited by phrase composition and intonation; WaveGAN is only truly capable of synthesizing "one second slices of audio waveforms" with global coherence.[11]

Creating a deepfake requires the manipulation of aural and visual data. Consequently, GANs and other deep-learning techniques often leave evidence of tampering, such as oddly placed shadows, chopped speech, or resolution inconsistencies.[12] While these errors may not be discernible to the human eye or ear, sophisticated deepfake detection algorithms can be trained to identify them, allowing us to distinguish between legitimate media and deepfakes. However, because deepfakes are self-correcting, deepfake detection algorithms are not reliable. Detection algorithms that are relatively accurate are unlikely to be so for long.[13]

Weaponized Deepfake Technology

Disinformation, widely defined as "the purposeful dissemination of false information intended to mislead or harm," has been a weapon in the political arsenal for millennia, and for good reason.[14] People are not "rational consumers of information"; we seek "swift, reassuring answers and messages that give [us] a sense of identity and belonging."[15] Confirmation bias—the tendency to believe that new information is real or true if it confirms preconceived beliefs—and cognitive closure—the tendency to seek quick clarity during times of heightened uncertainty—make people vulnerable to black-and-white messaging.[16] Psychological warfare exploits these fundamental biases and behaviors, manipulating popular opinion "to sway policy or inhibit action by creating division and blurring the truth within a target population."[17]

Deepfakes are an emerging tactic in an age-old practice, but the harm they can inflict is significant and more damaging than traditional propaganda campaigns. For over a century, audio and video recordings have functioned as evidence of truth. Films of liberated Nazi concentration camps, the American moon landing, and the September 11th terrorist attacks have all informed and shaped our view of reality.[18] Synthetic media operates in the same manner. Doctored images and videos, including deepfakes, can cause "people to believe in and remember experiences that never occurred" and even influence their decision making.[19] Seeing and hearing are truly believing; the aural and visual elements of deepfakes make them extraordinarily compelling weapons of disinformation, more so than traditional forms of propaganda.

U.S. Deepfake Detection Efforts

The Defense Advanced Research Projects Agency's (DARPA's) "Media Forensics" (MediFor) program currently "brings together world-class researchers to attempt to level the digital imagery playing field by developing technologies for the automated assessment of the integrity of an image or video and integrating these in an end-to-end media forensics platform."[20] The process begins at the University of Colorado in Denver, where MediFor researchers manipulate media to create convincing deepfakes.[21] At SRI International, analysts then use these videos to train computers to detect deepfakes, using artificial intelligence to develop video authentication algorithms.[22] If successful, DARPA's MediFor platform will be able to automatically detect media manipulations, explain how these manipulations were done, and determine the overall integrity of media to guide decisions regarding their use.[23]

MediFor uses three essential elements of media integrity to detect media manipulations: digital integrity, physical integrity, and semantic integrity.[24] Digital integrity refers to consistent pixels, representations, and metadata. A lack of digital integrity is the most common indication of media manipulation. Consequently, MediFor researchers first look for color shifts and replicated pixels to identify deepfakes. When digital integrity indicators break down, such as under high levels of compression, physical and semantic integrity indicators are used. [25] Physical integrity refers to images consistent with the laws of physics. Fluctuating shadows and lighting, elongation and compression, and multiple vanishing points all suggest a lack of physical integrity. Semantic integrity refers to contradicting evidence in images. MediFor researchers ensure that all dates, times, and locations are verifiable, that media assets were not repurposed, and that associated images depict content not in the original image (e.g., two photos of a building taken at the same time showing a different number of doors).[26]

Social media platforms and technology companies are also making strides to detect and combat weaponized deepfake attacks. In September 2019, Google created and released a dataset of over 3,000 deepfakes to accelerate the global development of deepfake detection tools.[27] In early January 2020, Facebook banned from its platform all videos "edited or synthesized by technologies like artificial intelligence in a way that average users would not easily spot," excluding videos manipulated for parody or satire.[28] In February 2020, YouTube banned all media content that has been "technically manipulated or doctored in a way that misleads users (beyond clips taken out of context)," while Twitter vowed to add labels to or delete deepfaked tweets.[29] In addition, independent researchers and artificial intelligence companies such as Dessa are working to build systems that can automatically identify and remove deepfakes.[30]

Though Big Tech and DARPA's MediFor team have made impressive strides in recent years, deepfake detection will almost always lag behind the innovation found in deepfake synthesis. A deepfake detector built by Dessa in late 2019, for example, could identify the deepfakes from the Google dataset with almost perfect accuracy, but failed to identify deepfakes found on the internet 40 percent of the time.[31] Dessa's deepfake detector suggests that the fight against deepfakes will require "nearly constant reinvention" in order to be successful.[32] It is therefore unlikely that any detection technology produced by MediFor or any independent researcher will be sufficient to defend the United States from deepfake attacks in the near future.

Applications of Weaponized Deepfakes

> *With careful editing, an indecisive firefight could be recast as a heroic battlefield victory. A few countering voices might claim otherwise, but how [can] they prove it? ... videos and images [move] faster than the truth.*
>
> – P.W. Singer and Emerson T. Brooking, 2016[33]

Deepfakes have a wide array of applications both on and off the battlefield. As a weapon of disinformation, deepfakes can shape political conversation by diverting attention from an issue, obscuring the truth, or directing audiences towards a certain course of action.[34] As a military tool, deepfakes can be used for deception or act as a cost-effective, non-escalatory attack—either alone or in conjunction with conventional military operations.

Influence Adversary Crisis Decision-Making

Weaponized deepfake attacks can influence adversary decision-making in one of two ways: by instigating a crisis (forcing an adversary to engage in quick, uncertain decision-making that benefits the United States), or by altering the information available during a crisis (modifying an adversary's strategic calculus in a way that benefits the United States).

Effective crisis management involves making prudent decisions under severe time pressure and uncertainty. However, the very factors that define crises adversely impact the quality of decision-making. Time pressure, for example, "reduces the quality of human judgement."[35] When the human body initiates an adrenaline response to a stressor, such as time pressure, the amygdala shuts down neural pathways within the brain to the prefrontal cortex—the "logical" part of the brain. Complex decision-making disappears, attention narrows, and memory fails.[36] The ability to make good decisions in the midst of a crisis is therefore both "extraordinarily important and extraordinarily difficult."[37] By launching a deepfake attack, the United States can instigate a crisis within the political and military leadership of an adversary, exploiting innate cognitive limitations to induce poor decision-making.

Within an existing crisis, weaponized deepfake attacks can influence adversary decision-making by exercising reflexive control, downplaying costs and risks, and exploiting selection bias.[38] Reflexive control is defined as "a means of conveying to…an opponent specially prepared information to incline him to voluntarily make the predetermined decision desired by the initiator of the action."[39] As an adversary attempts to resolve a crisis, a strategic deepfake attack can make it appear as though there are only a few viable policy options (when in reality, there are many). In this way, the United States can covertly persuade adversaries to choose a policy option in line with U.S. interests.

A weaponized deepfake attack can also make the costs and risks of a certain course of action appear to be lower than they actually are. If an adversary "intercepts" an audio deepfake of a time-bound order to evacuate an area, for example, the adversary can be tricked into believing that the area is safe to take (when in reality, American troops are waiting to ambush them). In this way, a deepfake attack can be used to convince an adversary to pursue a detrimental course of action.

Finally, weaponized deepfake attacks can influence adversary decision-making by exploiting selection bias. Selection bias is the tendency to "accept new information from experts, the mass media, and outside critics only when it supports [a] preferred alternative."[40] Weaponized deepfakes of experts and mass media figures supporting a policy choice the United States favors could be used to create a false sense of support for that choice and drown out any opposition. In this way, a deepfake attack can convince an adversary that a detrimental policy choice it wants to pursue will, in fact, lead to a preferred outcome.

Coercive Diplomacy

In coercive diplomacy, weaponized deepfakes can replace conventional attacks as a limited use of force, compelling adversaries to comply with U.S. demands, while simultaneously preventing unwanted military escalation. [41] Unlike conventional attacks, which can be costly to execute and can cause demonstrable physical destruction, deepfake attacks cost-effectively undermine trust in institutions and leadership, increasing domestic political constraints on foreign policy behavior. The United States can, for example, threaten to destabilize an adversary if it continues to pursue an objectionable course of action. If the adversary ignores the threat, the United States can use weaponized deepfakes in a media campaign that targets the adversary's political and military leadership, encouraging elite infighting, fomenting unrest within the general populace, and undermining public trust in institutions and authority.

Hybrid Warfare

Hybrid warfare refers to the use of conventional military force supported by subversive instruments, including cyber warfare and information operations.[42] To achieve a policy goal or to bolster an offensive military action, the United States can deploy weaponized deepfake attacks in conjunction with conventional warfare. Not only would wartime deepfake attacks impede the adversaries' ability to react, respond, and communicate, but they also would allow the United States to shape the international narrative of a military operation in the critical early days of a conflict.

Both state and non-state actors have successfully combined information operations with military action to achieve specific policy objectives.[43]

- *Second Lebanon War.* The 2006 Second Lebanon War was a 34-day hybrid military conflict between the Israeli Defense Forces (IDF) and Hezbollah.[44] For the first time in its military's history, Israel deployed psychological operations in conjunction with conventional force.[45] Over the course of the conflict, the IDF dropped approximately 17 million propaganda leaflets in 47 different leaflet missions over Lebanon.[46] Hezbollah also engaged in information operations, using its TV network "Al-Manar," radio station "Al-Nur," and numerous websites to "exploit collateral damage in order to portray Israel's attacks as egregious and inhumane."[47]

- *Russia-Georgia War.* The 2008 Russia-Georgia War is notable for "its inclusion of a series of large-scale, overt cyberspace attacks that were relatively well-synchronized with conventional military operations."[48] In total, "fifty-four news, government, and financial websites were defaced or denied," and "thirty-five percent of Georgia's internet networks suffered decreased functionality during the attacks, with the highest levels of online activity coinciding with the Russian invasion of South Ossetia on August 8, 9, and 10."[49] These cyberattacks supported a broader disinformation campaign, which promoted a false narrative that Georgia assaulted South Ossetia before Russian troops invaded and vilified Georgian President Mikheil Saakashvili as a war criminal.[50]

- *Gezi Park Protests.* Following widespread unrest during the summer of 2013, then-prime minister of Turkey Recep Tayyip Erdogan coupled police brutality with a relentless disinformation campaign to restore order.[51] While police repeatedly used excessive force to prevent and disperse peaceful demonstrations, the Justice and Development Party (AKP) "purported to reveal the 'real reasons' behind the Gezi protests," including "traitors, coupplotters, the CIA, Mossad, MI6, Europeans who envied Turkey's economic success," and "the Jewish lobby."[52] The AKP even went so far as to form its own "social media army," hiring approximately 6,000 social media experts to coordinate a "response plan against online activists critical of Turkish officials."[53]

Hearts and Minds

Deepfakes are psychologically compelling. They exploit inherent human biases and tendencies to shape public attitudes and influence political events. Similar to conventional propaganda (e.g. news articles, TV programs, and radio shows), weaponized deepfake attacks can be used to weaken existing domestic support for ruling parties in adversarial states or regions.

The United States has used information operations in the past to alter public opinion.

- *Iran.* In the summer of 1953, the Central Intelligence Agency and the United Kingdom's Secret Intelligence Service arranged a coup in Tehran to overthrow Iranian Prime Minister Mohammad Mosaddegh.[54] Led by senior officer Kermit Roosevelt Jr., "Operation Ajax" (or "TPAJAX," as it was called in official documents) was the product of a months-long disinformation campaign planned and funded by the CIA against Mosaddegh, who was the most popular figure in modern Iranian history at the time.[55] Iranians working for the CIA "staged anti-Mossadegh protests, marching through the streets carrying portraits of the Shah and chanting royalist slogans"; foreign agents bribed parliament members; and CIA propagandists launched increasingly virulent press attacks on Mossadegh, accusing him of communist leanings, designs on the throne, Jewish patronage, and even "secret sympathy for the British."[56] Despite the mission's initial failure, Mossadegh was eventually overthrown, and Iran remained a staunch Cold War ally of the West for more than 20 years.

- *Guatemala.* In May and June of 1954, the CIA and U.S. Ambassador John Peurifoy supported and directed Guatemalan military leaders to remove President Jacobo Arbenz from power.[57] "PBSUCCESS, authorized by President Eisenhower in August 1953, carried a $2.7 million budget for 'psychological warfare and political action' and 'subversion.'"[58] The subsequent disinformation campaign's efforts "to persuade Guatemalan citizens and political/military leaders that a major invasion force was steadily moving toward the nation's capital so unnerved Arbenz and others that the government fell without much of a battle."[59] Though the regimes that followed the 1954 coup were "far more repressive than Arbenz's elective government," the overthrow remains one of the CIA's most well-known regime change successes.[60]

- *Brazil.* Fearing that "the government of Brazilian President Joao Goulart would, in the words of U.S. Ambassador Lincoln Gordon, 'make Brazil the China of the 1960s,'" the United States backed a coup led by Humberto Castello Branco on April 1, 1964.[61] To ensure the coup's success, the CIA engaged in covert measures to undermine Goulart's popularity and "help strengthen resistance forces in Brazil."[62] Such measures included "covert support for pro-democracy street rallies…and encouragement [of] democratic

> and anti-communist sentiment in Congress, armed forces, friendly labor and student groups, church, and business."[63] Following Goulart's ousting, U.S.-Brazilian relations strengthened considerably, resulting in both ideological convergence and Brazilian acceptance of American leadership.[64]

Enhanced Detection Capabilities

Weaponizing deepfakes can enhance existing U.S. deepfake detection capabilities by improving our understanding of how these weapons are made and used. Deepfakes are composed of two competing algorithms: one which creates a deepfake (the "generator") and one which detects that deepfake (the "discriminator"). Because the algorithms that generate deepfakes continuously learn to replicate the appearance of reality more effectively, the best defense against a weaponized deepfake attack is a cutting-edge offense. Detection algorithms derived from past deepfake attacks will fail to defend against sophisticated future attacks; they are designed to "target the disinformation of yesterday rather than that of tomorrow."[65]

"Weaponization for defense" is a common strategy in cybersecurity. "White hat hackers"—also referred to as "good hackers" or "ethical hackers"—are cybersecurity experts who exploit computer systems or networks to identify security flaws and make recommendations.[66] To better defend against cyberattacks, these "good hackers" create cyberweapons, modifying small portions of known malicious code to circumvent a network's defense.[67]

Recent DARPA research suggests that weaponization for defense may be an effective strategy for deepfake detection. At University of Albany, Professor Siwei Lyu developed a successful deepfake detection algorithm by studying eye movement in convincing "puppet master" deepfakes.[68] Together with his research team, Lyu created the very things he aimed to detect—cutting-edge deepfakes of notable political figures—in order to enhance his own existing detection algorithms.[69] By producing the fakes himself, Lyu was able to understand more deeply the thought process behind the making of a weaponized deepfake, ultimately leading to a higher quality detection algorithm.

Arguments Against Weaponization

> *The marketplace of ideas already suffers from truth decay as our networked information environment interacts in toxic ways with our cognitive biases. Deepfakes will exacerbate this problem significantly.*
>
> – Danielle K. Citron & Robert Chesney, 2019[70]

The development and use of weaponized deepfake technology are not without risk. If the United States launches a deepfake attack, the attack can lead to unintended political violence. If targets or the international community attribute a deepfake attack to the United States, the attack can backfire. Additionally, if the United States publicly reveals that it is using deepfakes as a weapon, it may experience video verification difficulties, as well as spur the global proliferation of deepfake attacks.

- *Unintended Political Violence*. Weaponized deepfake attacks are scandalous and inflammatory by design. Consequently, they are likely to leak out of targeted audiences and into public discourse. In unstable states or regions, weaponized deepfake attacks can spark unintended political violence contrary to U.S. interests or against U.S. allies or personnel. Political violence, in turn, runs the risk of regional destabilization and civilian casualties, as well as has negative implications for U.S. trade.

- *Potential Backfire*. If targets attribute a weaponized deepfake attack to the United States, they can use the attack as effective anti-American propaganda. Adversarial non-state actors can circulate the weaponized deepfake to mobilize or recruit local citizens in a scenario comparable to the 2012 Benghazi consulate attack, in which citizens "protested a video satirizing the Prophet Muhammad, overwhelming a small U.S. consulate."[71] Adversarial state actors, in contrast, can argue that all messages and videos from the United States are fabricated. Such a message would undermine U.S. soft power, public diplomacy, and claims of fact.

- *Verification Difficulties*. If a third party finds the United States responsible for weaponizing deepfakes or launching a deepfake attack, it will be challenging for the United States to prove that any future videos it releases are legitimate. This difficulty would also extend to U.S.-led or partnered organizations. Even if the United States were to run its videos through a deepfake detection algorithm, because all current U.S. government-made deepfake detection algorithms are the products of military research, U.S. videos could still be reasonably dismissed as "fakes."

- *Nonproliferation Difficulties.* If the United States publicly reveals that it is weaponizing deepfake technology or if a weaponized deepfake attack is attributed to the United States, the United States will set a dangerous global precedent—that weaponized deepfake attacks are acceptable. Given this precedent, it will be difficult for the United States to justify any future actions it may take to discourage adversaries from using or developing weaponized deepfake technology, likely leading to deepfake proliferation.

- *Threat to Democratic and Financial Institutions.* Deepfake proliferation has the potential to corrode public trust in democratic and financial institutions by making election influence operations and disinformation campaigns more sophisticated. If the United States weaponizes deepfake technology or launches a weaponized deepfake attack and proliferation ensues, global capitalism and democracy will likely suffer from increased market volatility and truth decay, respectively.

Argument for Weaponization

Media weapons [can] actually be more potent than atomic bombs

– Propaganda Handbook of the Islamic State[72]

Weaponized deepfakes are a powerful tool for the United States to incorporate into its arsenal. In addition to their ability to exploit cognitive biases, weaponized deepfakes are difficult to defend against, fast-acting, and have no clear escalation thresholds. Given these qualities, it is likely that weaponized deepfakes will proliferate irrespective of U.S. policy. If the United States foregoes weaponization, it will deny itself use of a potent weapon and its offensive capabilities will fall behind.

- *Offense Dominant.* Weaponized deepfake attacks are offense dominant, meaning they are easier to use than defend against. By virtue of AI-driven self-correction, deepfakes are not easily detectable. Deepfake detection algorithms consistently lag behind offensive deepfake creation. Attackers benefit from this perpetual game of catch-up; there will always be a period within which a deepfake attack is undetectable. Additionally, the United States possesses the most advanced deepfake detection algorithms available. These algorithms are not publicly available and require a high degree of technical ability to access and use. Consequently, U.S. adversaries likely are ill-prepared to defend against a U.S. weaponized deepfake attack.

- *Fast Acting.* Weaponized deepfake attacks are fast-acting, making them an attractive option for time-sensitive operations. Disinformation, including deepfakes, tends to travel at higher speeds than mainstream news. According to an 11-year study on Twitter, researchers found that fake news travels six times faster than real news, because it is retweeted more frequently by Twitter users.[73] Outside social media, disinformation spreads through the forwarding function of messaging apps, including WhatsApp and Telegram. Given the tendency to share disinformation rapidly, the effects of a weaponized deepfake attack will likely spread too quickly to be countered effectively.

- *No Clear Escalation Thresholds.* As weaponized deepfakes are an emerging technology, they have no clear escalation thresholds. At present, their use does not intensify conflict the way conventional military operations, such as troop mobilization or aerial strikes, do. As a form of grey zone warfare, information warfare falls in a similar category. Under existing international law, "it is unclear whether nonlethal attacks that are neither physically intrusive nor physically destructive would constitute acts of 'war,' 'force,' or 'aggression.'"[74] Conventional military retaliation to a deepfake attack may be viewed as an disproportionate response.[75] The United States can take advantage of this ambiguity to launch weaponized deepfake attacks against its adversaries without prompting overt, militarized responses. However, adversaries can also take advantage of this ambiguity to launch deepfake attacks against the United States. To counter this threat, the United States should develop the capability to respond in kind to deter adversaries at all rungs on the escalation ladder.

The costs and risks of weaponizing deepfake technology are not inevitable. The United States can actively mitigate the effects of unintended political violence, backfire, and verification difficulties. Weaponized deepfake proliferation, however, is inevitable, because deepfake technology is increasingly undetectable and freely available online.

- *Target Hardening.* The United States can take measures prior to and following a deepfake attack to lessen the effects of unintended political violence on allies and U.S. assets. Such measures may include financially and militarily supporting allies prior to an attack or increasing the security of U.S. personnel in the region. To prevent a deepfake attack from immediately going viral on social media (and therefore lower the probability of backfire), the United States can suppress a targeted state's or region's internet connection in the immediate aftermath of a deepfake attack. Internet "throttling"—a variant of national-level internet cutoffs—slows down internet access without blocking it altogether, allowing vital online functions to continue, while rendering mass coordination difficult.[76] The United States also can micro-target the delivery of self-destructing deepfakes to key individuals in adversarial states and non-state groups. Specific messaging app settings, such as the status feature in WhatsApp and the timer feature in Telegram and Facebook Messenger, cause videos to disappear permanently upon viewing. If used in a weaponized deepfake attack, these features could postpone the deepfake's eventual leak into public discourse by prohibiting mass forwarding.

- *Independent Media Verification.* To circumvent video verification difficulties, the United States can encourage third parties, such as the International Fact-Checking Network (IFCN), to develop robust video authentication capabilities. The IFCN is an umbrella organization of independent fact-checking agencies that promotes best practices and exchanges in the media verification field.[77] In addition to combatting fake news and monitoring misinformation, IFCN agencies identify manipulated and out-of-context photos and videos.[78] The European Commission has encouraged the IFCN to incorporate deepfakes into its media verification network.[79] In partnering with such an independent media verification agency, the United States would allow a place for both targeted deepfake attacks and legitimate video content in its military strategy.

- *Deepfake Proliferation.* Deepfake nonproliferation difficulties are unavoidable. If the United States ever publicly admits to developing weaponized deepfake technology or to launching a deepfake attack, it will inevitably be difficult to justify preventing other actors from also doing so. However, weaponized deepfake attacks are inexpensive, effective, and increasingly within the technical capability of actors. For these reasons, weaponized deepfakes will likely proliferate irrespective of whether the United States offensively develops or uses this technology. Though the United States would, hypothetically, struggle to justify its nonproliferation efforts, it would not, in reality, attempt to limit the spread of a universally available technology. Given the threat that deepfake proliferation poses to global democracy, the United States should weaponize deepfakes to improve its existing deepfake detection capabilities.

Recommendations for Use

> *Around the world at this hour and every hour of the 24, there is a constant battle on the ether waves for the possession of man's thoughts, emotions, and attitudes—influencing his will to fight, to stop fighting, to work hard, to stop working, to resist and sabotage, to doubt, to grumble, to stand fast in faith in loyalty …*
>
> – Robert D. Leigh, director of the Foreign Broadcast Intelligence Service, 1944[80]

Deepfakes' ability to distort reality convincingly and undetectably make them exceptional weapons of disinformation and deception. As a military tool, deepfakes can provide novel ways to intimidate and sabotage adversaries, disrupt enemy lines of communication, and influence adversary decision-making. As a weapon of disinformation, deepfakes can compel adversaries to comply with U.S. demands, manipulate public opinion to shape political events, and destabilize political regimes. The U.S. military and intelligence community therefore has an interest in developing weaponized deepfake technology.

Given the inherent dangers associated with weaponized deepfakes, the United States should use this technology only during: (1) combat operations against adversary military forces and leadership and (2) combat operations against adversarial non-state actors and leadership. Against state actors, weaponized deepfakes can be used to achieve tactical objectives or within influence operations. Against non-state actors, deepfakes can be used within counterinsurgency or counterterrorism operations. Absent war, weaponized deepfakes should be used to refine and enhance existing deepfake detection algorithms. Though the limited use of weaponized deepfakes can occur during periods of non-war, the effects and success of such attacks are uncertain. Recommendations for use therefore cannot be easily generalized in non-war contexts.

War: Deception and Influence

Weaponized deepfake attacks promise to be highly sophisticated and virtually undetectable. Consequently, they will prove challenging for states, non-state actors, and civil societies to counter effectively.

- *State targets*: On the battlefield, weaponized deepfakes can be used against state actors to achieve a number of tactical objectives. To reduce the quality of an adversary's judgement, for example, the United States can launch a weaponized deepfake to instigate a crisis. To deceive an enemy about the location and intention of U.S. forces, the United States can employ deepfakes in a military deception campaign, creating or amplifying an artificial fog of war. To degrade an enemy's line of communication, the United States can insert false video and audio into an adversary's communication networks, generating confusion and intensifying mistrust of messages received.

  In additional to achieving tactical objectives, weaponized deepfakes can be used within wartime influence operations. Deepfakes could be used to sow popular discontent against a regime, increase elite infighting, or destabilize adversarial states or regions. Similarly, the United States can use weaponized deepfakes to trigger desired foreign regime transitions, manipulating public support away from an adversarial leader already in power and towards a leader the United States supports.

- *Non-state targets*: Against non-state actors, the United States can use weaponized deepfakes to enhance existing U.S. counterinsurgency and counterterrorism operations. Weaponized deepfake attacks can support counterinsurgency operations by shaping public perceptions of the insurgent group.[81] The United States could, for example, launch a series of weaponized deepfake attacks against the insurgent group, depicting the insurgents as inherently violent and irredeemably corrupt and the government as peaceful, understanding, and willing to appease rebelling insurgents.

  Within counterterrorism operations, weaponized deepfakes can be used to counter two strategies of terrorism: intimidation and outbidding. Terrorists using an intimidation strategy seek to convince a population that the terrorist group is strong enough to punish disobedience and the government is too weak to stop them.[82] The United States can counter an intimidation strategy by spreading deepfakes of terrorists being captured or punished by local authorities within the local populace. Terrorists using an outbidding strategy seek to convince a population that they are worthier of support than rival groups, often using violence to demonstrate their resolve.[83] The United States can counter outbidding strategies by advancing deepfakes of terrorist leaders quickly and easily surrendering to local governments and rival groups.

  The United States has launched successful influence operations against terrorist organizations in the past. In 2011, under the Obama administration, the State Department's Center for Strategic Counterterrorism Communications (CSCC) was created, a hub through which U.S. officials disseminated anti-Islamic State messaging and videos over Facebook and Twitter.[84] Dissolved in 2016, the CSCC was replaced by the Global Engagement Center (GEC), an organization which continues to act as a "secretive counter-propaganda center" to fight Russian, Chinese, and Iranian disinformation online.[85] Currently, the GEC's efforts are not limited by any U.S. propaganda standards or policies. In addition to the Department of Defense, the GEC could utilize weaponized deepfake attacks to support its mission.

Non-War: Building Defenses

Weaponized deepfake proliferation has the potential to undermine global democracy and capitalism by corroding public trust in democratic and financial institutions. Deepfake technology also has characteristics that enable rapid and widespread diffusion. As a result, weaponized deepfake attacks outside of wartime can be disastrous. Acceptable non-war uses of weaponized deepfake attacks therefore cannot be easily generalized: the state or region-specific probability of unintended political violence, potential backfire, verification difficulties, and nonproliferation difficulties must be weighed against the expected benefits of the attack. Additionally, the United States must be sufficiently prepared to address both international and domestic blowback for conducting information warfare in peacetime.

Absent war, the United States should weaponize deepfakes to improve its existing deepfake detection capabilities. Weaponizing deepfakes can enhance existing U.S. deepfake detection capabilities by improving our understanding of how these weapons are made and used. Because the algorithms that generate deepfakes continuously learn to replicate more effectively the appearance of reality, the best defense against a weaponized deepfake attack is a cutting-edge offense. Detection algorithms derived from past deepfake attacks will fail to defend against sophisticated future attacks. Given that weaponized deepfake technology is inexpensive, publicly available, and increasingly undetectable, it is very likely that adversaries will launch weaponized deepfake attacks against the United States in the near future. To counter this threat and strengthen trust in global democratic and financial institutions, the United States must develop the capability to both detect deepfakes and respond in kind to weaponized deepfake attacks.

Conclusion

Weaponized deepfake technology promises to revolutionize disinformation through highly realistic, virtually undetectable fake audio and video recordings. In addition to their ability to exploit cognitive biases and fundamental human tendencies, weaponized deepfakes are difficult to defend against, fast-acting, and have no clear escalation thresholds. As a military tactic, deepfakes can provide novel ways to exploit, intimidate, and sabotage adversaries. As a disinformation tactic, deepfakes can manipulate public opinion to shape political events or destabilize political regimes. Though the development and use of weaponized deepfake technology are not without risk, the United States can take active measures to mitigate the costs and risks associated with weaponization.

The United States ultimately has an interest in expanding its current deepfake detection efforts to include deepfake weaponization for defense and warfighting. If the United States chooses to weaponize deepfakes, it should use them within sanctioned military campaigns against foreign adversaries. Though weaponized deepfake attacks can be used against adversarial actors during peacetime, the effects and success of such attacks are context-specific and cannot be easily generalized.

Acknowledgements

The author thanks P.W. Singer, Christopher Meserole, Joe Littell, JD Maddox, and Jeffrey Kaplow for questions and comments that dramatically improved the paper, as well as LTC Casey Dean, USA, CPT T.S. Allen, USA, and Thomas Plant for their extensive research and editorial assistance. The author also thanks Major General Mark Matthews, USAF (Ret.), the U.S. Army Mad Scientist Team, the Joint Forces Staff College, and the Global Research Institute at William & Mary for supporting and promoting the PIPS program. Finally, the author would like to thank Professor Dennis Smith, whose meticulous edits were instrumental to this paper's success, and Professor Amy Oakes, who believed in this project from its inception and without whom this paper would not have been possible.

[1] "What is a deepfake?," *The Economist*, August 7, 2019, https://www.economist.com/the-economist-explains/2019/08/07/what-is-a-deepfake.

[2] "Deepfakes: What are they and why would I make one?," *BBC*, accessed February 25, 2020, https://www.bbc.co.uk/bitesize/articles/zfkwcqt.

[3] "What is a deepfake?," *The Economist.*

[4] Sir John Lubbock, *The Beauties of Nature and the Wonders of the World We Live In (*New York: Macmillan, 1892), 4.

[5] "What are deepfakes – and how can you spot them?," *The Guardian*, January 13, 2020. https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

[6] "What is a deepfake?," *The Economist.*

[7] Ibid.

[8] A Beginner's Guide to Generative Adversarial Networks (GANs)," *pathmind*, accessed February 25, 2020. https://pathmind.com/wiki/generative-adversarial-network-gan.

[9] Chris Donahue, Julian McAuley, and Miller Puckette, "Adversarial Audio Synthesis," in *Proceedings of the Seventh International Conference of Learning Representations,* (New Orleans, LA: ICLR, 2019): 5, https://arxiv.org/pdf/1802.04208.pdf.

[10] Joe Littell, "Don't Believe Your Eyes (Or Ears): The Weaponization of Artificial Intelligence, Machine Learning, and Deepfakes," *War on the Rocks,* October 7, 2019, https://warontherocks.com/2019/10/dont-believe-your-eyes-or-ears-the-weaponization-of-artificial-intelligence-machine-learning-and-deepfakes/.

[11] Donahue, McAuley, and Puckette, "Adversarial Audio Synthesis," 1.

[12] Drew Harwell, "Top researchers race to detect 'deepfake' videos: 'We are outgunned,'" *The Washington Post*, June 12, 2019, https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/.

[13] Chris Meserole and Alina Polyakova, "The West is ill-prepared for the wave of "deep fakes" that artificial intelligence could unleash," *Brookings*, May 25, 2018. https://www.brookings.edu/blog/order-from-chaos/2018/05/25/the-west-is-ill-prepared-for-the-wave-of-deep-fakes-that-artificial-intelligence-could-unleash/.

[14] Though disinformation is generally considered to be a modern political art, the strategic use of false information can be traced back to the ancient Romans. Rome's leaders used sensationalist propaganda to communicate and justify their power and policies to an immense empire. Enemy societies were often depicted as primitive and impoverished. In his famous account of the Gallic Wars of 58-50 BC, for example, Julius Caesar described enemy Germanic tribes as barbaric, lawless, and dangerous, reinforcing the idea that Rome represented peace, good governance, and rule of law. In his description of the Caledonian tribes of ancient Scotland in the early third century AD, Dio Cassius labeled Caledonians as uncivilized, unclothed, and unshod nomads, implying that such peoples could benefit from Roman rule. In addition, enemy rulers were often the subjects of smear campaigns. The historical depiction of Cleopatra as a wily seductress who entrapped Marc Antony is one such example. The smear campaign was spearheaded by Octavian, the future Augustus Caesar, under the guise of protecting Rome from moral decay. For more see: Jacquelyn Williamson, "Ancient Egypt's A List: Power, Empire, and Propaganda," *Smithsonian Associates*, October 27, 2018. https://smithsonianassociates.org/ticketing/tickets/ancient-egypts-a-list-power-empire-and-propaganda. and Neil Faulkner, "The Official Truth: Propaganda in the Roman Empire," *BBC*, February 17, 2011. http://www.bbc.co.uk/history/ancient/romans/romanpropaganda_article_01.shtml.
Quote from: Christina Nemr and William Gangware. *Weapons of Mass Distraction: Foreign State-Sponsored Disinformation in the Digital Age* (Washington, DC: Park Advisors, 2019), https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distraction-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf.

[15] How disinformation impacts politics and publics," *National Endowment for Democracy*, May 29, 2018. https://www.ned.org/issue-brief-how-disinformation-impacts-politics-and-publics/.
Nemr and Gangware, *Weapons of Mass Distraction.*

[16] Mona Kasra, "Can you spot a fake photo online? Your level of experience online matters a lot more than contextual clues," *NiemenLab,* June 24, 2019. https://www.niemanlab.org/2019/06/can-you-spot-a-fake-photo-online-your-level-of-experience-online-matters-a-lot-more-than-contextual-clues/. Daniele Anastasion, "The Price of Certainty," *The New York Times,* November 1, 2016. https://www.nytimes.com/2016/11/01/opinion/the-price-of-certainty.html.

[17] Nemr and Gangware, *Weapons of Mass Distraction.*

[18] "Inside the Pentagon's race against deepfake videos," *CNN,* accessed February 25, 2020. https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/
[19] Robert Nash, Kimberly Wade, and Rebecca Brewer. "Why do doctored images distort memory?," *Consciousness and Cognition* 18, no. 3 (2009): 773-780, http://www.sciencedirect.com/science/article/pii/S1053810009000798. Jeremy Bailenson et al. "Facial Similarity between Voters and Candidates Causes Influence," *Public Opinion Quarterly* 72, no. 5 (2008): 935–961, https://doi.org/10.1093/poq/nfn064.
[20] Matt Turek, "Media Forensics (MediFor)" *Defense Advanced Research Projects Agency*, accessed February 26, 2020, https://www.darpa.mil/program/media-forensics.
[21] Catalin Grigoras and Cole Whitecotton, "Seeing isn't believing: CU Denver masters the science of truth in audio/video," February 13, 2019, in *CU on the Air*, produced by Blubrry Podcasting, MP3 audio, 35:39. http://cuontheair.blubrry.net/2019/02/13/seeing-isnt-believing-cu-denver-center-masters-the-science-of-truth-in-audio-video/.
[22] "Inside the Pentagon's race against deepfake videos," *CNN.*
[23] Turek, "Media Forensics (MediFor)," *Defense Advanced Research Projects Agency*.
[24] Linda Casola and Dionna Ali, "Adversarial Attacks," in *Robust Machine Learning Algorithms and Systems for Detection and Mitigation of Adversarial Attacks and Anomalies: Proceedings of a Workshop*, (Washington, DC: The National Academies Press, 2019), https://www.nap.edu/read/25534/chapter/4.
[25] Ibid.
[26] Ibid.
[27] Karen Hao, "Google has released a giant database of deepfakes to help fight deepfakes," *MIT Technology Review,* September 25, 2019, https://www.technologyreview.com/f/614426/google-has-released-a-giant-database-of-deepfakes-to-help-fight-deepfakes/.
[28] Tony Romm, Drew Harwell, and Isaac Stanley-Becker, "Facebook bans deepfakes, but new policy may not cover controversial Pelosi video," *The Washington Post,* January 7, 2020, https://www.washingtonpost.com/technology/2020/01/06/facebook-ban-deepfakes-sources-say-new-policy-may-not-cover-controversial-pelosi-video/.
[29] Casey Newton, "How big tech companies could team up to stop deepfakes," *The Verge*, Februrary 6, 2020, https://www.theverge.com/interface/2020/2/6/21124934/misinformation-policies-youtube-twitter-deepfakes-synthetic-media. Yoel Roth and Ashita Achuthan (@TwitterSafety), "Building rules in public: Our approach to synthetic & manipulated media," *Twitter Blog,* February 4, 2020, https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html.
[30] Cade Metz, "Internet Companies Prepare to Fight 'Deepfake' Future," *The New York Times,* November 24, 2019, https://www.nytimes.com/2019/11/24/technology/tech-companies-deepfakes.html.
[31] Ibid.
[32] Ibid.
[33] P.W. Singer and Emerson T. Brooking, *LikeWar: The Weaponization of Social Media* (New York: Houghton Mifflin Harcourt, 2018), 8.
[34] "How disinformation impacts politics and publics," *National Endowment for Democracy*.
[35] Janet Sniezek et al., "Training for Crisis Decision-Making: Psychological Issues and Computer-Based Solutions," *Journal of Management Information Systems* 18, no. 4 (2002): 147-168, https://www.jstor.org/stable/40398546
[36] Diane Musho Hamilton, "Calming Your Brain During Conflict," *Harvard Business Review*, December 22, 2015, https://hbr.org/2015/12/calming-your-brain-during-conflict.
[37] Janet Sniezek et al., "Training for Crisis Decision-Making: Psychological Issues and Computer-Based Solutions."
[38] David Welch, "Crisis Decision Making Reconsidered," *The Journal of Conflict Resolution* 33, no. 3 (1989): 430-445, https://www.jstor.org/stable/174123.
[39] Timothy Thomas, "Russia's Reflexive Control Theory and the Military," *Journal of Slavic Military Studies* 17, no. 2 (2010): 237-256, https://doi.org/10.1080/13518040490450529.
[40] Ibid.
[41] "Coercive diplomacy" is a political-diplomatic strategy that combines "threats of force, and, if necessary, the limited and selective use of force … to induce an adversary to comply with one's demands." A subset of coercive warfare and compellence, coercive diplomacy encompasses "defensive" compellent actions only, demanding that targets stop or reverse actions already taken rather than offensively forcing targets to do something. For more see: Jack Levy, "Deterrence and Coercive Diplomacy: The Contributions of Alexander George," Political Psychology 29, no. 4 (2008): 537-552, https://fas-polisci.rutgers.edu/levy/articles/Levy%20%20Deterrence%20&%20Coercive%20Diplomacy.pdf and Lisa A.

Nemeth, The Use of Pauses in Coercion: An Examination in Theory, (Fort Leavenworth, KS: Publisher, 2012), 6, https://apps.dtic.mil/dtic/tr/fulltext/u2/a506197.pdf.

[42] *Understanding Russian "Hyrbid Warfare" And What Can Be Done About It: Testimony before the House Armed Services Committee,* 115th Cong. (2017) (statement of Christopher S. Chivvis, senior political scientist at the RAND Corporation).

[43] Joshua Ball, "What is Hybrid Warfare?," *Global Security Review,* August 1, 2018. https://globalsecurityreview.com/hybrid-and-non-linear-warfare-systematically-erases-the-divide-between-war-peace/

[44] "Second Lebanon War," The Israel Defense Forces, accessed February 25, 2020, https://www.idf.il/en/minisites/wars-and-operations/second-lebanon-war-2006/.

[45] Ron Schleifer, "Psyoping Hezbollah: The Israeli Psychological Warfare Campaign During the 2006 Lebanon War," *Terrorism and Political Violence* 21, no. 2 (2009): 221-238, https://doi.org/10.1080/09546550802544847.

[46] Herbert Friedman, "Psychological Operations During the Israel-Lebanon War 2006," *PsyWar*, August 14, 2006, https://www.psywar.org/israellebanon.php.

[47] Lisa Brennen, "Hezbollah: Psychological Warfare Against Israel" (Phd diss., Naval Postgraduate School, 2009), 64, https://apps.dtic.mil/dtic/tr/fulltext/u2/a496916.pdf.

[48] Sarah White, "Understanding Cyberwarfare: Lessons from the Russia-Georgia War," *Modern War Institute,* March 20, 2018, https://mwi.usma.edu/understanding-cyberwarfare-lessons-russia-georgia-war/.

[49] Ibid.

[50] Robert Cutler, "Russia's Disinformation Campaign over South Ossetia," *The Central Asia-Caucasus Analyst,* August 20, 2008, https://www.cacianalyst.org/publications/analytical-articles/item/11678-analytical-articles-caci-analyst-2008-8-20-art-11678.html

[51] Berivan Orucoglu, "How President Erdogan Mastered the Media," *Foreign Policy,* August 12, 2015, https://foreignpolicy.com/2015/08/12/how-president-erdogan-mastered-the-media/.

[52] Amnesty International, *Gezi Park Protests: Brutal Denial of the Right to Peaceful Assembly in Turkey*, Report no. EUR 44/022/2013 (London: Amnesty International, 2013), https://www.amnestyusa.org/files/eur440222013en.pdf. Orucoglu, "How President Erdogan Mastered the Media." and ibid.

[53] Amnesty International, *Gezi Park Protests: Brutal Denial of the Right to Peaceful Assembly in Turkey*, Report no. EUR 44/022/2013 (London: Amnesty International, 2013), https://www.amnestyusa.org/files/eur440222013en.pdf. Orucoglu, "How President Erdogan Mastered the Media."

[54] The extent of the United States' involvement in the 1953 Iranian coup is highly debated today. In his 2003 book *All the Shah's Men*, Stephen Kinzer of the *New York Times* provided a detailed account of CIA covert action in Iran throughout the early 1950s. A decade later, the CIA itself released 21 declassified documents to the National Security Archive, confirming that the CIA devoted "extensive resources and high-level policy attention toward bringing about Mosaddeq's overthrow, and smoothing over the aftermath." For more see: Malcom Byrne, "CIA Confirms Role in 1953 Iran Coup," The National Security Archive, August 19, 2013. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB435/. In light of these documents, some American scholars believe the United States played only a minor role in Mossadegh's overthrow, and that Kinzer wildly exaggerated the United States' involvement. For more see: Ray Takeyh, "What Really Happened in Iran," *Foreign Affairs*, July/August 2014. https://www.foreignaffairs.com/articles/middle-east/2014-06-16/what-really-happened-iran. Lawrence Wu and Michelle Lanz, "How The CIA Overthrew Iran's Democracy in 4 Days," February 7, 2019, in *Throughline*, produced by NPR. Podcast, MP3 audio, 37:47. https://www.npr.org/2019/01/31/690363402/how-the-cia-overthrew-irans-democracy-in-four-days.

[55] Steven Kinzer, *All The Shah's Men* (Hoboken, NJ: Wiley, 2003), 7.

[56] Kinzer, *All The Shah's Men,* 6.

[57] David Barrett, "Congress, the CIA, and Guatemala, 1954", Stud. Intel. Winter/Spring 2001, No. 10:23-31. https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol44no5/html/v44i5a03p.htm.

[58] Nicholas Cullather, *Operation PBSUCCESS: The United States and Guatemala,* (Washington, DC: The Central Intelligence Agency, 1994), 89.

[59] Barrett, "Congress, the CIA, and Guatemala, 1954", Stud. Intel.

[60] Ibid.

[61] J Dana Suster, "Mapped: The Seven Governments the U.S. Has Overthrown," *Foreign Policy*, August 20, 2013. https://foreignpolicy.com/2013/08/20/mapped-the-7-governments-the-u-s-has-overthrown/.

[62] Lincoln Gordon to State Department, cable, 29 March 1964, *Brazil Marks 40th Anniversary of Military Coup*, ed. Peter Kornbluh (Washington, D.C.: The National Security Archive and Chadwyck-Healey, 2004). https://nsarchive2.gwu.edu/NSAEBB/NSAEBB118/index.htm

[63] Ibid.

[64] "Remembering Brazil's Military Coup 50 Years Later," the North American Congress on Latin America, April 1, 2014, https://nacla.org/news/2014/4/1/remembering-brazils-military-coup-50-years-later.

[65] Meserole and Polyakova, "The West is ill-prepared for the wave of "deep fakes" that artificial intelligence could unleash."

[66] Rob Sobers, "What Does it Take to Be an Ethical Hacker?," *Varonis,* June 19, 2018, https://www.varonis.com/blog/white-hat-hacker/.

[67] Daniel S. Hoadley, *Artificial Intelligence and National Security,* CRS Report No. R45178 (Washington, DC: U.S. Congressional Research Service, 2019), https://fas.org/sgp/crs/natsec/R45178.pdf.

[68] "Exposing Fake Videos," *University at Albany,* June 20, 2018, https://www.albany.edu/news/87379.php.

[69] Sarah O'Carroll and Siwei Lyu, "Detecting Deepfake Videos with Siwei Lyu," Episode 4, UAlbany News Podcast, University at Albany, podcast audio, September 21, 2018, https://ualbanynewspodcast.simplecast.com/episodes/detecting-deepfake-videos-with-siwei-lyu-1552184e/transcript.

[70] Danielle Citron and Robert Chesney, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," 107, no. 1743 (2019), https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=1640&context=faculty_scholarship.

[71] Joe Littell, "Don't Believe Your Eyes (Or Ears): The Weaponization of Artificial Intelligence, Machine Learning, and Deepfakes," *War on the Rocks,* October 7, 2019, https://warontherocks.com/2019/10/dont-believe-your-eyes-or-ears-the-weaponization-of-artificial-intelligence-machine-learning-and-deepfakes/.

[72] Singer and Brooking, *LikeWar*, 148.

[73] Chris Stokel-Walker, "Fake news travels six times faster than the truth on Twitter," *NewScientist*, March 8, 2018. https://www.newscientist.com/article/2163226-fake-news-travels-six-times-faster-than-the-truth-on-twitter/. Peter Dizikes, "Study: On Twitter, false news travels faster than true stories," *MIT News,* March 8, 2018. http://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308.

[74] Lawrence Greenberg, Seymour Goodman, and Kevin Soo Hoo, *Information Warfare and International Law,* (Washington, DC: National Defense University Press, 1998), 15, http://www.dodccrp.org/files/Greenberg_Law.pdf.

[75] Ibid.

[76] Ibid., 89.

[77] "The International Fact-Checking Network," *Poynter,* accessed February 25, 2020, https://www.poynter.org/ifcn/.

[78] Glenn Kessler, "Rapidly expanding fact-checking movement experiences growing pains," *The Washington Post,* June 25, 2018, https://www.washingtonpost.com/news/fact-checker/wp/2018/06/25/rapidly-expanding-fact-checking-movement-faces-growing-pains/.

[79] European Commission, "Tackling online disinformation: a European Approach," *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee and the Committee of the Regions* COM(2018) 236 final, April 26, 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236.

[80] Singer and Brooking, *LikeWar*, 33.

[81] Alan Vick et al., *Air Power in the New Counterinsurgency Era* (Santa Monica, CA: RAND Corporation, 2006), Chapter 4, https://www.jstor.org/stable/10.7249/mg509af.12.

[82] Andrew Kydd and Barbara Walter, "The Strategies of Terrorism," *International Security* 31, no. 1 (2006): 49-80, https://www.jstor.org/stable/4137539.

[83] Ibid.

[84] Guy Taylor, "State Department Global Engagement Center targets Russian propaganda, 'deep fakes,'" *Associated Press*, December 12, 2018. https://apnews.com/9f7892a163582b5fd0297e2a81124c35.

[85] Ibid.