P | I | P | S

# Replicating Reality

## Advantages and Limitations of Weaponized Deepfake Technology

PIPS White Paper 12.4: *Executive Summary*

Megan Hogan, Research Fellow
Tom Plant, Research Intern

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.

### *Deepfake Technology*

Creating a deepfake uses artificial intelligence algorithms to manipulate audio and visual data. Generative Adversarial Networks and other deep-learning techniques leave evidence of tampering, such as oddly placed shadows, chopped speech, or resolution inconsistencies. While these errors may not be discernible to the human eye or ear, sophisticated deepfake detection algorithms can be trained to identify irregularities, allowing us to distinguish between legitimate media and deepfakes. However, because deepfake algorithms are self-correcting, deepfake detection can be unreliable. Relatively accurate detection algorithms are unlikely to remain so as programmers use detection algorithms to improve the performance of generating algorithms.

### *Weaponized Deepfake Technology*

Deepfake disinformation is an emerging tactic in an age-old practice, but the harm it 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 by informing and shaping our view of reality. Doctored images and photos, including deepfakes, can cause people to believe in and remember experiences that never occurred—influencing their decision making.

*Applications of Weaponized Deepfakes*

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 altering public opinions. As a military tool, the United States can weaponize deepfakes for deception, advancing detection capabilties, and as a cost-effective, non-escalatory means of attack—either alone or in conjunction with conventional military operations.

- *Influence Adversary Crisis Decision-Making.* Weaponized deepfake attacks can influence adversary decision-making by instigating a crisis or by altering the information available during a crisis.

- *Coercive Diplomacy.* Weaponized deepfakes can replace conventional attacks as a limited use of force, compelling adversaries while simultaneously preventing unwanted military escalation.

- *Warfare*. To achieve a policy goal or to bolster an offensive maneuver, weaponized deepfake attacks can function alongside conventional warfare. Deepfakes can impede the adversaries' ability to react on the battlefield and politically.

- *Hearts and Minds.* Deepfakes exploit inherent human biases and tendencies to shape public understanding and influence political events. These capabilities allow deepfakes to weaken support for ruling parties in adversarial states.

- *Enhanced Detection Capabilities*. Weaponizing deepfakes can enhance existing U.S. deepfake detection capabilities. Because deepfake algorithms continuously learn to replicate reality, the best defense against a deepfake attack is a cutting-edge offense.

*Arguments Against Weaponization*

The use of weaponized deepfake technology is not without risk. A U.S. deepfake attack can lead to unintended consequences. Weaponized deepfakes can erode trust in democratic and financial institutions, lead to unintended violence, undermine U.S. claims of video and audio authentication, and lead to the global proliferation of weaponized deepfakes.

*Arguments for Weaponization*

Given that deepfakes are difficult to defend against, fast-acting, and have no clear escalation thresholds, 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.

*Recommendation for Use*

Deepfake weaponization is strategically advantageous, but the effects and success of such attacks are context-specific and cannot be easily generalized. If the United States chooses to weaponize deepfakes, it should use deepfakes within military campaigns against foreign adversaries. The United States should restrict weaponization outside of war to developing detection algorithms and anticipating adversarial weaponized deepfake attacks.

- *Deception and Influence.* On the battlefield, weaponized deepfakes deceive enemies by clouding judgement, obscuring the location and intention of U.S. forces, or degrading or generating distrust of communications. In respect to influence operations, targeted weaponized deepfakes can sow popular discontent and increase elite infighting.

- *Building Defenses.* Given the strategic advantages of deepfake technology, it is very likely that adversaries will launch weaponized deepfake attacks against the United States in the near future. Combatting these attacks requires advanced detection capabilities. The United States should be at the forefront of weaponized deepfake technology because a sophisticated generation algorithm gives insight into what methods adversaries will use to deceive.