



Sources	US Army Research Laboratory, IUL		
Date	March 2021		
Potential scale of impact	Certainty of outcome	Impact horizon	
★★★★★	★★★★★	H1	H2 H3

[US army scientists](#) have developed a novel tool that can help soldiers detect deepfakes - hyper-realistic video content made using artificial intelligence tools that falsely depicts individuals saying or doing something - that pose threat to national security.

The advance could lead to a mobile software that warns people when fake videos are played on the phone.

[The number of deepfake videos is growing.](#) There were close to 8,000 deepfake video clips online at the beginning of 2019, which doubled to about 15,000 over the course of nine months. The recent emergence of [sophisticated deepfakes of Tom Cruise on Tik-Tok](#) has created something of a stir in communities worried about exploitation and misinformation.

Their circulation can be harmful to society – from the creation of [non-consensual explicit content](#) to doctored media by foreign adversaries that are used in disinformation campaigns. UCL’s Dawes Centre for Future Crime suggest that [deepfakes are the most worrying criminal use of AI](#) for the next decade.

Most of the currently used state-of-the-art deepfake video detection methods are based upon complex machine learning tools that lack robustness, scalability and portability – features soldiers would need to process video messages while on the move. The researchers describe a new lightweight face-biometric tool called [DefakeHop](#), which meets the size, weight, and power requirements soldiers have. The new framework has been developed by combining principles from machine learning, signal analysis, and computer vision. One of the main advantages is that the AI tool needs fewer training samples, learns faster from the sample data, and can offer high detection performance of fake videos.

The scientists hope that their new tool will lead to a “software solution that runs easily on mobile devices to provide automatic warning messages to people when fake videos are played.”

