



● 作者/Justin Hauffe ● 譯者/李育慈 ● 審者/馬浩翔

眼見不盡為憑

Don't Believe Your Eyes

取材/2021年8月美海軍學會月刊(*Proceedings*, August/2021)

透過人工智慧所發展的深偽技術，如今甚至已經到了以假亂真的地步。
在有心人士利用下，成為散布假消息與混淆視聽的工具，
當這種技術應用在情報或心理戰層面時，更具奇效。



(Source: USMC)



深偽(Deepfake)是看似真實正確，實則是詐欺作為的超寫實影音資料。深偽技術利用人工智慧(AI)與機器學習應用程式，編輯並複製真人的臉孔、身體與聲音，藉以製作任何人、做任何事或說任何話的仿真影片。¹ 深偽技術往往在社會媒體上搭配社會支持與相關網路以取得可信度。結合兩者可能澈底破壞影音證據可信度，最終改變社會建構事實認知的方式。

對海上軍種及美軍其他軍種而言，此種危險更形迫切。敵人可能利用深偽技術破壞指管、動搖合法情報來源之影音檔可信度，並且損害外界對美軍的觀感——這可能造成駐外美軍以身犯險。

無害的娛樂

深偽技術源自娛樂產業，其中部分大型作品需要雄厚的資金與人力。早在1994年的電影《阿甘正傳》(Forrest Gump)中，漢克斯(Tom Hanks)飾演的主角即已依數位化加入歷史鏡頭中，與已故許久的總統及名人互動。

近年來，深偽技術變得更為普及、廉價並且易於取得。2017年，俄羅斯「無線實驗室」(Wireless Lab)公司開發出智慧型手機應用程式——Face-App，成為首批最容易取得的深偽技術之一，其利用人工智慧使用戶的照片看起來更年輕、更老，甚至轉換成另一種性別。² 2019年，一家中國大陸的公司研發出技術更驚人的應用程式——Zao，用戶能直接以自己的臉孔取代電影中的演員。³

喜劇演員皮爾(Jordan Peele)在對美國提出的戲謔(儘管也相當嚴肅)示警中，製作了一段深偽影片，片中歐巴馬總統說了一段他從未說過的粗

鄙言論。⁴ 最近，麻省理工學院先進虛擬技術中心(Center for Advanced Virtuality)則製作一則短片，讓尼克森總統宣讀1969年阿波羅登月計畫失敗的演講稿。⁵

這兩則與美國總統有關的深偽影片皆是善意，旨在警告公眾深偽技術的危險，但深偽工具亦可用來為惡。2012年，俄羅斯針對出言批評總統普丁的美國駐俄大使麥弗爾(Michael McFaul)製作一段深偽影片。⁶ 該影片謊稱麥弗爾大使有戀童癖，藉以迫使美國替換另一位對普丁較為友善的大使。

社群媒體的資訊本質

社群媒體扮演的角色對深偽而言相當重要。社群媒體不只傳遞媒介，其分享資訊的「方式」亦強化深偽的攻擊性。社群媒體讓網紅得以將目標瞄準特定團體(如社會、人口、意識形態/政黨傾向等)，這些用戶往往認同特定傾向、性格與意見。因此，找到網路，並修改訊息以回應網路之需，實輕而易舉。

社群媒體大部分資訊傳播係透過「社會認同」——即透過個人周邊社群團體來評判資訊是否為事實——而非取決既有機構之可信度。社會認同極具效力；心理學研究指出，人類較可能聽從其夥伴支持的建議或要求。⁷ 因此，在社群媒體網路上傳播的資訊，無論真偽，其被接收者相信的可能性因而提高。

研究指出，社會認同會影響用戶觀感，即便用戶明白社群媒體上許多消息通常誇張不實。皮尤研究中心(Pew Research Center)研究發現：

大約有三分之二(68%)的美國成人表示，至少偶爾從社群媒體取得資訊……然而，這些人之中有許多懷疑他們在那裡見到的資訊真偽：57%的人聲稱，他們預期在社群媒體上得到的消息大多是錯誤的。但多數社群媒體新聞用戶則表示，以此管道取得的消息，幾乎不會影響他們對當前事件的瞭解，甚至有更多人宣稱，這可以幫助他們瞭解梗概，而非有如霧裡看花

(36%比15%)。⁸

基本上，社群媒體已進一步社會化人們對事實的建構。

深偽與證據本質

深偽技術特別危險，因為它利用先前已被接受的確切證據——影音紀錄——來傳播不實資訊。⁹ 例如，錄音被用來做為尼克森總統試圖阻撓水門案醜聞的國會調查證據。¹⁰ 近來國際社會接受沙烏地阿拉伯記者卡

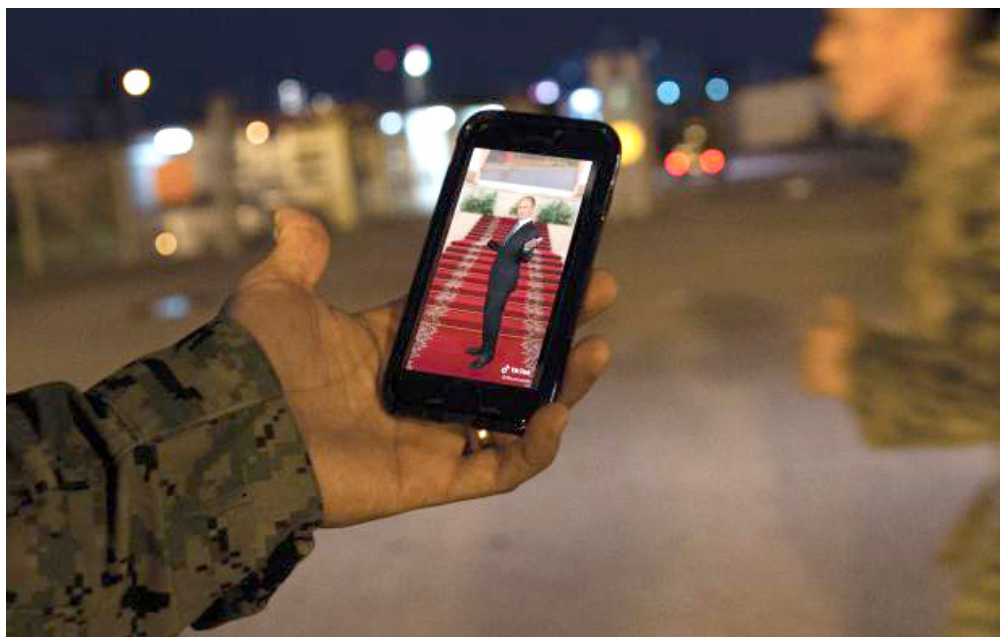
沙吉(Jamal Khashoggi)走進沙國駐土耳其領事館的影片，證實沙國政府在記者謀殺事件中所扮演的角色。¹¹

然而，精密複雜的深偽影片可能改變人們對證據的認知。想像一下審訊一名搶劫銀行嫌疑犯。原告可能製作一段影片，其中顯示被告闖入銀行，而被告亦可能製作一則反深偽影片，其中「顯示」被告在搶案發生時，身處在一個完全不同的地點。若無目擊者或其他證據，法庭該如何判斷孰真孰假？軟體或能偵測深偽技術，卻亦能遭矇騙。當社會認同取代既有權威機構來驗證資訊的真確性，陪審員可能步入一個已預先傾向從影片二擇一的法庭。

軍事威脅分析

以下數個假設性想定有助闡釋深偽技術對軍事的威脅：

- 海軍軍醫署長宣稱美國製造的特定新冠肺炎疫苗會致癌，這則深偽影片正在社群媒體流傳。
- 情報人員提供美軍一則深偽影片，片中描述一名伊朗



俄羅斯總統普丁並未擁有抖音(TikTok)帳號，但用戶卻可能誤以為他有。有數個不同的偽普丁帳號存在，其中用戶@1facerrussia製作這名俄羅斯領導人表演滑稽舞步的深偽流行影片。該帳號擁有300萬名追蹤者，大多數似乎都相信此則插科打諢，但仍有不少網路貼文解釋這些是深偽影片，這說明此媒介即便用在諷刺上，仍極具影響力與誤導性。(Source: USMC & Tiktok User @1facerrussia)



想像海軍軍醫署長宣稱美國製造之特定新冠肺炎疫苗會致癌的深偽影片正在社群媒體流傳。若軍方正準備進行重大部署或系列部署，該影片可能充其量浪費時間分散注意力。但最糟的情況是，它可能在關鍵時刻造成疑惑，並降低遵守疫苗接種規定的配合度。(Source: US Army & USN on YouTube)

將領在簡報中以美國為目標的核子攻擊計畫。

- 美海軍在穆斯林國家進行港口訪問期間，一部描述美海軍士兵汗巖清真寺的深偽影片正在穆斯林社群流傳。

為因應每個想定，軍方必須攻擊此工具(即深偽影音)並保護目標(即深偽技術意圖影響的人心)。基本上，軍方應首先確認該影片不實，繼而說服目標閱聽眾勿相信影片內容。

攻擊工具

打擊深偽技術有兩個主要步驟。第一步是偵測。約翰傑刑事司法學院(John Jay College of Criminal Justice)教授馬拉絲(Marie-Helen Maras)和亞歷山德魯(Alex Alexandrou)表示，此偵測技術稱為數位影像鑑識，著重在「偵測低階影像竄改，諸如刪除或複製畫面及/或某部位，剪接與複製貼上部分原始影像，將其置於其他地方，如拷貝移動竄改」。¹²

第二步是發展並追蹤數位

指紋，用戶能追溯影像來源。¹³

2021年，臉書與密西根州立大學(Michigan State University)合作研發人工智慧軟體，據稱可偵測深偽影音，並追溯其源頭。¹⁴

此種進步眾所樂見，惟即便軟體完美運作，仍有兩項主要挑戰會迅速削弱其效力。首先，深偽技術會持續發展，而偵測軟體必須同步發展以維持效力。其次，偵測軟體能否成功告知公眾哪些是事實、哪些是深偽，取決於假設目標閱聽眾信任偵測軟體及使用該軟體的機構(如臉書)。最終，「信任」是對抗深偽技術最有價值的利器。

取決於無知

偵測與追蹤只是初步階段。美軍打擊深偽的目的，是防止目標閱聽眾將深偽誤以為真而回應之。欲成功達此目的，有賴教育並獲取目標受眾的信賴，使其瞭解深偽存在，以及這些質疑深偽的機構可受信賴。

軍方應從教育部隊開始著手。內部教育課程無須大費周章，可以簡化成在國防部內傳

播一至兩分鐘的影片，介紹深偽技術的仿真製造能力，最終強調僅信賴官方資訊來源的重要性。

說服外國民眾勿回應敵方深偽影片，將是一項挑戰。這些民眾較可能偏好國內媒體，並傾向懷疑美軍。因此，與私部門合作不可或缺。好消息是，許多控制深偽技術傳播平臺的科技公司已採取措施，遏止深偽傳播。儘管如此，由於深偽技術會持續演進以避免遭到偵測，而這絕非打擊深偽的終章。

十九世紀法國詩人波特萊爾(Charles Baudelaire)說道，「惡魔的最佳詭計是說服你，祂並不存在。」目前全球許多社會尚未注意到深偽存在。

為制敵機先，美軍應儘可能教育民眾深偽技術的威脅。國防部應從教育軍人開始，繼而向外推展。官兵們則應妥善準備因應突如其來發生的深偽威脅。在這場打擊深偽的行動中，無知是最大弱點，教育則是最大防禦。

作者簡介

Justin Hauffe中校現擔任美空軍KC-135R空中加油機飛行員暨外事軍官。他以優秀成績畢業於加州蒙特瑞(Monterey)海軍研究生院(Naval Postgraduate School)，並自2006年從美國空軍官校畢業後即投身軍旅。

Reprint from *Proceeding* with permission.

註釋

1. Marie-Helen Maras and Alex Alexandrou, "Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos," *The International Journal of Evidence & Proof* 23, no. 3 (1 July 2019): 255-62.
2. Kate O'Flaherty, "The FBI Investigated FaceApp. Here's What It Found," *Forbes*, 3 December 2019.
3. Marie C. Baca, "Viral Chinese App Zao Puts Your Face in Place of Leonardo DiCaprio's in 'Deepfake' Videos" *The Washington Post*, 3 September 2019.
4. Ian Hislop, "How the Obama/Jordan Peele Deep-face Actually Works," BBC, www.youtube.com/watch?v=g5wLaJYBAm4, 2019.
5. Suzanne Day, "MIT Art Installation Aims to Empower a More Discerning Public," *MIT News*, 25 November 2019.
6. Deb Reichmann, "I Never Said That! High-Tech Deception of 'Deepfake' Videos," *The Seattle Times*, 1 July 2018.
7. Robert B. Cialdini, *Influence: The Psychology of Persuasion*, rev. ed. (New York: Collins, 2006), 20.
8. Elisa Sharer and Katerina Eva Matsa, "News Use Across Social Media Platforms 2018," *Pew Research Center's Journalism Project* (blog), 10 September 2018.
9. Maras and Alexandrou, "Determining Authenticity of Video Evidence."
10. Marisa Iati, "Inside the Supreme Court Ruling That Made Nixon Turn over His Watergate Tapes," *The Washington Post*, 3 October 2019.
11. Julian E. Barnes, Eric Schmitt, and David D. Kirkpatrick, "'Tell Your Boss': Recording Is Seen to Link Saudi Crown Prince More Strongly to Khashoggi Killing," *The New York Times*, 12 November 2018.
12. Maras and Alexandrou, "Determining Authenticity of Video Evidence."
13. Haya R. Hasan and Khaled Salah, "Combating Deepfake Videos Using Block-chain and Smart Contract," *IEEE Access* 7 (2019): 41596-41606.
14. Jaclyn Diaz, "Facebook Researchers Say They Can Detect Deepfakes and Where They Came From," NPR, 17 June 2021, www.npr.org/2021/06/17/1007472092/facebook-researchers-say-they-can-detect-deepfakes-and-where-they-came-from.