AI and Misinformation

Darian Fautz

If you've spent any time on the internet within the last year or two, then you've most likely encountered a deepfake. A lot of the time they can be pretty harmless and very obviously fake. I've encountered plenty of TikToks of Hillary Clinton playing Fortnite with the likes of Donald Trump and Vladimir Putin. The trouble with deepfakes comes when they are created and distributed more surreptitiously and with malicious intent.

On October 26, 2023, supermodel Bella Hadid released a statement on the current conflict in Palestine and Israel.[1] Bella herself is Palestinian and many of her fans were anxiously anticipating her thoughts on the situation.[2] She made an Instagram post of her thoughts and ended her statement with a call for peace.[3] Two days later, a video was posted to X, formerly Twitter, that showed Bella standing at a podium and saying, "I stand with Israel against terror."[4] As of this blog being written, the post has garnered over 30 million views.[5] That video is now flagged with a warning from X that it is AI.[6] Bella has yet to comment on the video, but many X users were concerned with the misinformation being spread due to the deepfaked video.[7]

The spread of misinformation and other harmful effects from deepfakes has been exacerbated by their ease of being created. Previously making a deepfake required the creator to use elaborate software, but now, deepfake tools are available through smartphone apps, giving everyday consumers access to the technology for free.[8] There are two ways of making a deepfaked video.[9] The first is called a replacement or a "faceswap."[10] Using this method, "the identity of a source subject is transferred onto a destination subject's face. The destination's facial expressions and head movements remain the same, but the identity takes on that of the source."[11] The second method is called a re-enactment.[12] In a re-enactment video, the deepfake maker uses the source person to drive "the facial expressions and head movements of a destination person, preserving the identity of the destination."[13] Audio for deepfakes can be

---

[1] Jennifer Zhan, *Bella Hadid Says Palestine 'Cannot Afford Our Silence'*, VULTURE (Oct. 26, 2023), https://www.vulture.com/2023/10/bella-hadid-palestine-israel-statement.html#:~:text=The%20people%20and%20children%20of,trauma%20of%20my%20Palestinian%20blood.%E2%80%9D.

[2] *Id.*

[3] Bella Hadid (@bellahadid), INSTAGRAM (Oct. 26, 2023), https://www.instagram.com/p/Cy4DlmjAZxA/?hl=en&img_index=2.

[4] Danel Ben Namer (@DanelBenNamer), X (Oct. 28, 2023, 3:55 PM), https://twitter.com/DanelBenNamer/status/1718355794297503881.

[5] *Id.*

[6] *Id.*

[7] *Id.*

[8] Stuart A. Thompson, *Making Deepfakes Gets Cheaper and Easier Thanks to A.I.*, N.Y. TIMES (Mar. 12, 2023), https://www.nytimes.com/2023/03/12/technology/deepfakes-cheapfakes-videos-ai.html

[9] Catherine Bernaciak & Dominic A. Ross, *How Easy Is It to Make and Detect a Deepfake?*, SOFTWARE ENG'G INST. (Mar. 14, 2022), https://insights.sei.cmu.edu/blog/how-easy-is-it-to-make-and-detect-a-deepfake/.

[10] *Id.*

[11] *Id.*

[12] *Id.*

[13] *Id.*

created using AI to clone celebrity voices and then altering mouth movements to match the audio to say whatever the creator desires.[14]

The deepfake of Bella Hadid mentioned above was made using footage of her appearance at a ceremony for the Global Lyme Alliance.[15] Since the source and destination were the same person, there was probably no faceswapping or reenactment involved; it's likely that an ordinary user simply utilized a free celebrity voice AI model to modify Bella's speech and then adjusted her mouth movements in the video to match.

One of the main concerns with deepfakes when they first became popular was their function in pornographic media. Deepfakes allowed individuals to exploit non-consenting celebrities or other individuals by using their faces and voices to create pornography.[16] This first malicious use of deepfakes concerned many legal scholars; many voiced their concerns about other potential harmful uses for deepfakes, such as manipulation of democratic discourse and erosion of trust.[17]

The fears of these scholars have most definitely come to fruition. The deepfake involving Bella Hadid is just one example of these ongoing concerns. Along with the spread of misinformation, experts today are concerned that the prolific use of deepfakes has made many internet users skeptical of things they are seeing online that are actually real.[18] It has even been suggested that bad actors will argue that real videos and pictures are actually AI generated in order to cause confusion and further an agenda, a concept called "the liar's dividend."[19] These concerns are especially grave in times of political conflict and war.

Unfortunately for those concerned, there is no current national legislation regarding deepfakes or other deceptive uses of AI[20] and fewer than 10 states have laws on the books governing AI-generated content. Although, this may change with Representative Yvette Clarke's DEEPFAKES Accountability Act.[21] The act proposes prosecution of internet users who fail to label "malicious deepfakes."[22] A "malicious deepfake" would include deepfakes related to sexual content, criminal conduct, incitement of violence and foreign interference in elections.[23] There would also be an avenue to sue for damages under civil claims.[24] Scholars are concerned that this act is too narrow and may call into question First Amendment issues, so it's unclear whether other member of the House will support the act.[25]

---

[14] Thompson, *supra* note 8.
[15] Shayan Sardarizadeh (@Shayan86), X (Oct. 29, 2023, (9;54 PM) https://twitter.com/Shayan86/status/1718808464850382931.
[16] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Cal. L. Rev. 1753, 1772 (2019).
[17] *See id.*
[18] Tiffany Hsu & Stuart A. Thompson, *Making Deepfakes Gets Cheaper and Easier Thanks to A.I.*, N.Y. TIMES (Oct. 28, 2023), https://www.nytimes.com/2023/10/28/business/media/ai-muddies-israel-hamas-war-in-unexpected-way.html?searchResultPosition=2.
[19] Chesney & Citron, *supra* note 16 at 1785.
[20] Emmanuelle Saliba, *Bill would criminalize 'extremely harmful' online 'deepfakes'*, ABC NEWS (Sept. 25, 2023, 2:20 PM), https://abcnews.go.com/Politics/bill-criminalize-extremely-harmful-online-deepfakes/story?id=103286802.
[21] *Id.*
[22] *Id.*
[23] *Id.*
[24] *Id.*
[25] *Id.*

In the meantime, President Biden has issued a new executive order trying to curb some of AI's dangerous uses.[26] The executive order allows mostly undisturbed AI development but imposes some restrictions and rules.[27] One of the rules regarding deepfakes directs the "Commerce Department to come up with guidance for watermarking A.I.-generated content, which could help crack down on the spread of A.I.-generated misinformation."[28] Although the rules imposed by this executive order are very modest, this is a step in the right direction and keeps the dangers of AI on the government's radar. As a consumer and a citizen, staying vigilant about AI-generated content while avoiding the pitfalls of the 'liar's dividend' will be challenging, but this is all one can do until sweeping regulations are presented.

[26] David E. Sanger & Cecilia Kang, *Biden to Issue First Regulations on Artificial Intelligence Systems*, N.Y. TIMES (Oct. 30, 2023), https://www.nytimes.com/2023/10/31/technology/executive-order-artificial-intelligence-regulation.html?searchResultPosition=1.
[27] *Id.*
[28] *Id.*