



## **Deepfakes/Real Harms Toolkit**

Deepfakes are a type of media generated using artificial intelligence, whereby images are combined to create new footage that depicts events, statements or actions that may never have actually happened.

As this technology becomes easier to use, there have been more and more cases of deepfakes being used to create fake AI-generated pornographic materials of women. This is the predominant use case of deepfake technology, with estimates suggesting that more than 95% of all deepfakes on the internet feature non-consensual abuse imagery.

In recent years, a number of people who have been targeted by this technology have spoken out.

Watch this brief video to learn more:

<https://www.youtube.com/watch?v=qidzRZ74f9c>

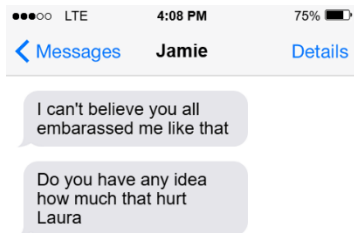


In the next section of this study, you will see three examples of how people interact with deepfake videos. These have been based on real events. After each section, we will ask you to briefly think (for 10-15 seconds) about the content or to write out a reflective response.

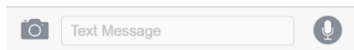
## Deepfakes/Real Harms Toolkit

### 1: Are deepfakes just jokes?

You have two friends Laura and Jamie, who you've known for years. They have been getting a lot closer in the past few months, and a few of your friends think they may be dating. As a joke, one of your friends puts together a fake video of the two of them deepfaked into an intimate scene from a romcom and posts it in a group chat where Laura and Jamie can see it. The content is not real, and it is clearly labelled as coming from a deepfake website. Reflect for a moment. **Why do you think your friend made this false video?**



You like the video but notice that Laura and Jamie have left the group chat. Jamie has cut off contact with a lot of your friend group. You text Jamie and he tells you that "I can't believe you all embarrassed me like that". He sounds very upset.



Imagine the type of apology you would write to him and write it below (15 words min)

---

---

---

---

Researchers from the University of Washington found that a majority of people believe that making or sharing entertainment deepfakes of strangers is unacceptable. **Just because deepfakes aren't real and just because a deepfake may have been created as a joke, doesn't mean it isn't harmful.** Impersonating someone else without their consent can be upsetting and humiliating.

*Do you agree that non-consensual manipulated images and videos are harmful?*

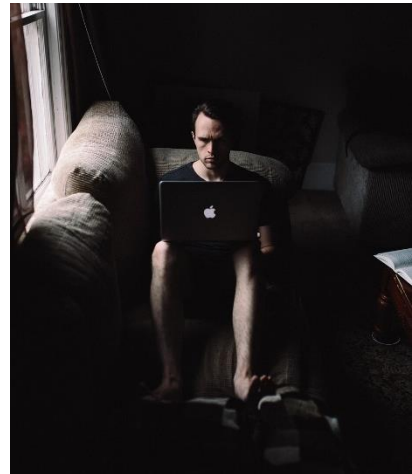
## Deepfakes/Real Harms Toolkit

*2: If deepfakes aren't posted online, are they harmless?*

You have an old family friend named Doug. He has had a crush on your cousin Zoe for years. They used to be very good friends, but they've grown apart. One night when Doug has been drinking, he starts talking about Zoe, he tells you that he's been editing adult videos and putting images of her into them. He says this is okay to do and amounts to the same as a sexual fantasy because he isn't sharing them with anyone else. He says that If they're only for his viewing, it's not hurting anyone. Reflect for a moment. *Why do you think Doug believes this?*

You are not the only person he has talked about this to while drinking. It gets back to Zoe, and she writes a long post on social media detailing how upset she is.

She asks you to write a message to Doug explaining why this is making her so upset. What would you write? (15 words min)



---

---

---

---

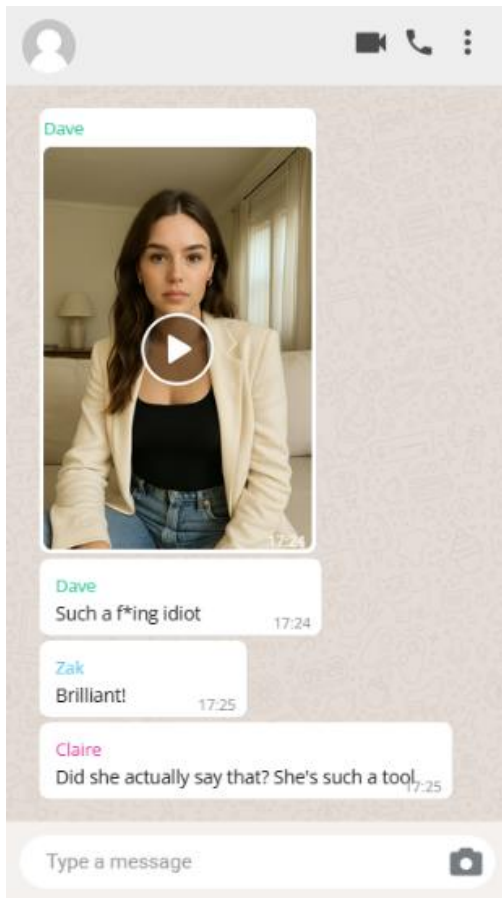
---

Research has shown that most people find creating non-consensual deepfake videos to be morally wrong, especially if they are of people you know. There is increasing legal focus on the use of deepfakes to create abuse imagery. **These acts are not without consequences: recently a man went to prison in the United Kingdom for using deepfakes to create abuse imagery of children for his own personal use.**

*Do you agree that it is wrong to make deepfake videos of people you know, even if you don't share them?*

## Deepfakes/Real Harms Toolkit

### 3: Do some people deserve to be targeted by deepfakes?



You get a notification on your phone one evening while relaxing at home. You open a group chat to see that your friends have shared a video of a popular celebrity who you dislike. In the video the celebrity talks negatively about herself, her intelligence and her career. The video is obviously fake, she is intentionally made to have a dumber sounding voice, and you can tell the image is fake due to how the face moves. However, you think it's funny and you like the video and send it to another friend. Reflect for a moment. *Why do you think a person made this video?*

Later that week, you see an interview with the same celebrity online, she is clearly distressed and crying and she talks about how hurtful it was to see so many people mock her and make her out to be someone she's not. She also talks about people sending her and her friends images of her edited into abuse imagery and pornographic films. How do you think it feels for a public figure to see fake videos of themselves online? (15 words min)

---

---

---

---

In interviews with celebrity victims of deepfake harassment and abuse imagery, people often report that they feel like the creators of deepfakes think they *deserve* to be targeted by deepfakes as they are “public figures”. Celebrities often highlight the emotional harm of seeing this content online and how it leads to increased harassment and could have real world impacts such as increased stalking and death threats. **Just because someone is in the public eye, it doesn't mean that they deserve to have to deal with non-consensual intimate imagery and defamatory videos.**

*Do you think deepfakes can be damaging and upsetting for public figures?*

## Deepfakes/Real Harms Toolkit

### 4. How do you feel about deepfakes?

|   | Strongly Disagree     | Disagree              | Neither Agree nor Disagree | Agree                 | Strongly Agree        |
|---|-----------------------|-----------------------|----------------------------|-----------------------|-----------------------|
| When deepfake videos are just jokes, they aren't harmful                    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> |
| Deepfake videos are only harmful when someone believes they're real         | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> |
| It is okay to make deepfakes of celebrities because they are public figures | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> |
| Some people are more deserving of being targeted by deepfakes than others   | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> |
| It's okay to make deepfakes when there are no legal consequences            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> |
| Deepfakes aren't harmful if they're not shared online                       | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> |

*Reflect on why people may believe in some or all of these commonly circulating myths about deepfakes.*

## Deepfakes/Real Harms Toolkit

### 5. Resources

Deepfake media is often used as part of bullying and harassment online. The public harassment of targets with deepfake abuse imagery is often accompanied by attempts to invalidate their experience using a number of “deepfake myths”. The victims of deepfake harms, particularly women who have been targeted with abuse imagery created using deepfakes show how often they hear variations of these myths online as a way to dismiss their experiences. This Toolkit aims to challenge the myths that underpin deepfake abuse.

If you have any questions or feedback about this intervention, you can email [deepfakes@ucc.ie](mailto:deepfakes@ucc.ie)

*If you have experienced any distress from the content of this study, please avail of the resources below:*

The Revenge Porn Helpline is a UK organization which offer supports for victims of intimate imagery including deepfakes: <https://revengepornhelpline.org.uk>

STOPNCII.ORG is a group which offers help to victims of non-consensual intimate imagery in removing real and manipulated content online: <https://stopncii.org/>

Find a Helpline is a global directory that connects people in crisis with mental health and support helplines in their country: <https://findahelpline.com>

Suicide Stop is a free online resource offering crisis helplines, chat support, and self-help tools for individuals struggling with suicidal thoughts: <https://www.suicidestop.com>

Lifeline International is a global network of crisis helplines dedicated to suicide prevention and mental health support: <https://lifeline-international.com>