INSIDE DEEPFAKES: UNDERSTANDING THE A.I. NIGHTMARE CREATING A NEW WAVE OF VIOLENCE AND LIABILITY

Presented by:

Jill Ostrove, Esq.
VP of Sexual Misconduct Risk Management

Jill Ostrove is the VP of SML (Sexual Misconduct and Molestation Liability) Risk Management at ePlace Solutions, Inc. She is a licensed attorney in California, where she received her JD with honors in both criminal law and public service from Western State University College of Law and her Bachelor of Arts in Political Science from University of California, Riverside.

Prior to joining ePlace, Jill gained significant experience in various areas of law, including civil, insurance defense, and as in-house corporate counsel. However, from a young age, her passion was firmly rooted in criminal prosecution.

As a prosecutor, Jill handled thousands of cases from investigation through filing charges, hearings, jury trials, and sentencing. Those cases encompassed, in part, domestic violence, child abuse, animal cruelty, sex offenses, and gang crimes.

In her current position at ePlace, Jill develops thorough policies, procedures, and extensive training materials to prevent tragedies to all age groups and high-risk industries. Jill additionally trains audiences all over the country on sexual abuse prevention and routinely speaks on topics such as A.I.; social media dangers; child sexual abuse; recruitment strategies; general sexual and workplace violence prevention; HR issues for public entities; domestic violence and sexual assault; mandatory reporting; and many others.

Since she was a child, Jill has always focused her attention on helping vulnerable and at-risk populations, whether that is volunteering on the Board of Directors for local nonprofits, working with victims of crime, helping the homeless population, or running her own nonprofit animal rescue.

HOUSEKEEPING

TRIGGER WARNING

We will be discussing topics and real cases that mention or depict child sexual abuse, sexual assault, and other serious and violent crimes.

We will also be listening to deepfake audio that incudes explicit language and racist statements.

Please take a break when needed and take care while listening.

Agenda

- All About Deepfakes
- Deepfakes and Sextortion
- ID Techniques, Examples, and the Future of Deepfakes
- Prevention and Considerations

PRO TIP

Child Sexual Abuse Material

NOT

Child Pornography.

Do you know why?

ALL ABOUT DEEPFAKES

Deepfakes are Designed to Gaslight the Viewer

Seeing is No Longer Believing

What Are Deepfakes?

Deepfakes (AKA "synthetic media," or "synthetic sexually explicit material") are photos, videos, or audio where a person's face, body, or voice has been digitally altered through A.I. They are the result of using sophisticated technology that allows for the overlay of faces and voices in videos or audio. This creates fake, but highly realistic content.

Deepfake attempts were up 3000% in 2023. The rise of AI-generated explicit content has surged by over 290% on the internet's top 10 platforms. Three seconds of audio is sometimes all that's needed to produce an 85% voice match from the original to a clone. 60% of consumers have encountered a deepfake video within the last year. Only 15% state that they have never encountered a deepfake video.

How are Deepfakes Made?

One photo is all it takes to create a deepfake. Anyone with a single photo of themselves ***anywhere*** is susceptible to a deepfake getting created.

Less than one minute. It takes less than one minute to create a face swap or voiceover deepfake.

Deepfakes rely on artificial neural networks, a form of machine learning that attempts to mirror the way neurons interact in the human brain. More recent deepfakes have come from algorithms that encode the features of a face from their training data.

The Technical Answer

Deepfake technology is a combination of two algorithms—generators and discriminators. Generators take the initial data set to create (or generate) new images based on the data gathered from the initial set. Then, the discriminator evaluates the content for realism and feeds its findings back into the generator for further refinement.

The combination of these algorithms is called Generative Adversarial Networks (“GANs”), basically, advanced machine learning. They learn from one another by refining inputs and outputs so that the discriminator can’t tell the difference between a real image, sound clip, or video and one created by a generator.

Image-Based Sexual Abuse

What is Image-Based Sexual Abuse?

According to the NCOSE, image-based sexual abuse ("IBSA") is a type of gender-based violence. It is the sexual violation of a person committed through the abuse or weaponization of any image of the person depicted.

IBSA is an umbrella term that includes the creation, theft, extortion, threatened or actual distribution, or any use of sexually explicit or sexualized materials without the meaningful consent of the person/s depicted and/or for purposes of sexual exploitation. It also includes sexual violence or harassment committed towards a person's representation (e.g., a person's avatar) in virtual reality or online gaming.

An "image" is any visual depiction or representation of a person—including but not limited to materials such as photographs, videos, edited/altered images, or personal representations in virtual reality or online gaming.

Types of IBSA

- Deepfakes
- Sextortion
- “Revenge porn”
- Video voyeurism or “spycamming” (i.e. down blousing or upskirting)
- Non-consensual distribution of sexually explicit material
- Recording sexual violence

Common Types of Deepfakes

- Photos
  - Text-to-photo

- Voice alteration
  - Tone/voice/gender/age change
  - Language change
  - Text-to-speech

- Videos
  - Video creation from photos
  - Text-to-video

- Face-swapping
- Body-swapping
- Dubbing or lip syncing

How Did We Get Here?

- **1990s**: Beginning of CGI attempts to create human-looking images.

- **1997:** Academic paper described creation of the, “Video Rewrite Program,” which basically automated what movie studios were already doing. The program could synthesize new facial animations from an audio output. It built upon older work that interpreted faces, but was the first to put this all together and animate it convincingly.

- **2010s**: Significant advancements in the computer science field and developments in machine learning.

- **2014**: Creation of the Generative Adversarial Network

- **2015**: Google published a blogpost on what it called “inceptionism”, but which rapidly came to be known as “DeepDream”. In it, engineers from the company’s photo team asked: What happens if you take the AI systems that Google had developed to label images and ask them to create images instead?

How Did We Get Here?

- **2017**: Reddit moderator named “Deepfakes” created a subreddit for users to exchange deepfake sexual imagery they had created using photos of celebrities and open-source face-swapping technology. Although the forum has since been deleted, the word deepfake has persisted.

- **Early 2018:** A video starring Jordan Peele and President Obama. In the video, it’s revealed that Jordan Peele is doing some sort of voice over, putting the words into President Obama’s mouth, yet the movement of President Obama’s mouth and head seem natural.

- **2021**: OpenAI released Dall-E, and face-swapping became old news. The first major image generator, Dall-E offered the science-fiction promise of typing a phrase in, and getting a picture out. It was the 1st time images of this type existed; They were not simply remixed versions of previous pics, but wholly new things. The first version of Dall-E wasn’t great at photorealism, but it showed great promise.

Deepfakes Trigger Mandated Reporting

A.I.-generated images of CSAM are illegal if they contain real children **or** if images of actual children are used to train data.

Reminder: It's not your duty to investigate what the images are based on or if they're real before reporting. It's just your duty to report. Let law enforcement do their job. Internal investigations are 100% separate 100% of the time.

You won't get in trouble for reporting something that may be incorrect.

The Big Takeaway

- Deepfakes of minors are child sexual abuse material.
- Child sexual abuse material is child abuse.
- Child abuse triggers mandated reporter duty.

A Brand New Issue

The deepfake voiceover kidnapping ransom scam.

DEEPFAKES AND SEXTORTION

What is Sextortion?

According to the FBI, sextortion is a crime that involves adults or minors coercing anyone into sending explicit images online.

Sextortion can start on any site, app, messaging platform, or game where people meet and communicate. The person may coerce or trick the victim into sending nude or revealing videos or pictures of themselves to the perpetrator.

After the criminals have one or more videos or pictures, they threaten to post them to get the victim to send more pictures. The shame, fear, and confusion minors feel when they are caught in this cycle often prevents them from asking for help or reporting the abuse.

The FBI also has recently seen an increase in financial sextortion targeting minors. In these cases, the perpetrator receives sexually explicit material from the child and then threatens to release the compromising material unless the victim sends money and/or gift cards. This increasing threat has resulted in an alarming number of deaths by suicide.

Pro Tip

Sextortion scams and related suicides are happening over the course of HOURS, not days, weeks, or months.

They are disproportionately affecting teens, specifically teen boys aged 14-17.

A Brand New Issue with Sextortion

As a result of deepfakes, sextortion now:

- No longer needs the victim to send an initial picture or video; and
- No longer ends when the victim refuses to comply with demands to send additional material.

The same applies to the idea of "revenge porn." The initial material is no longer needed to be supplied to the perpetrator. The perpetrator can create their own deepfake with the victim's face or body and use it for blackmail.

Sextortion Triggers Mandated Reporting

Federal law prohibits the production, advertisement, transportation, distribution, receipt, sale, access with intent to view, and possession of CSAM.

Reminder: It's not your duty to investigate what the images are based on or if they're real before reporting. It's just your duty to report. Let law enforcement do their job. Internal investigations are 100% separate 100% of the time.

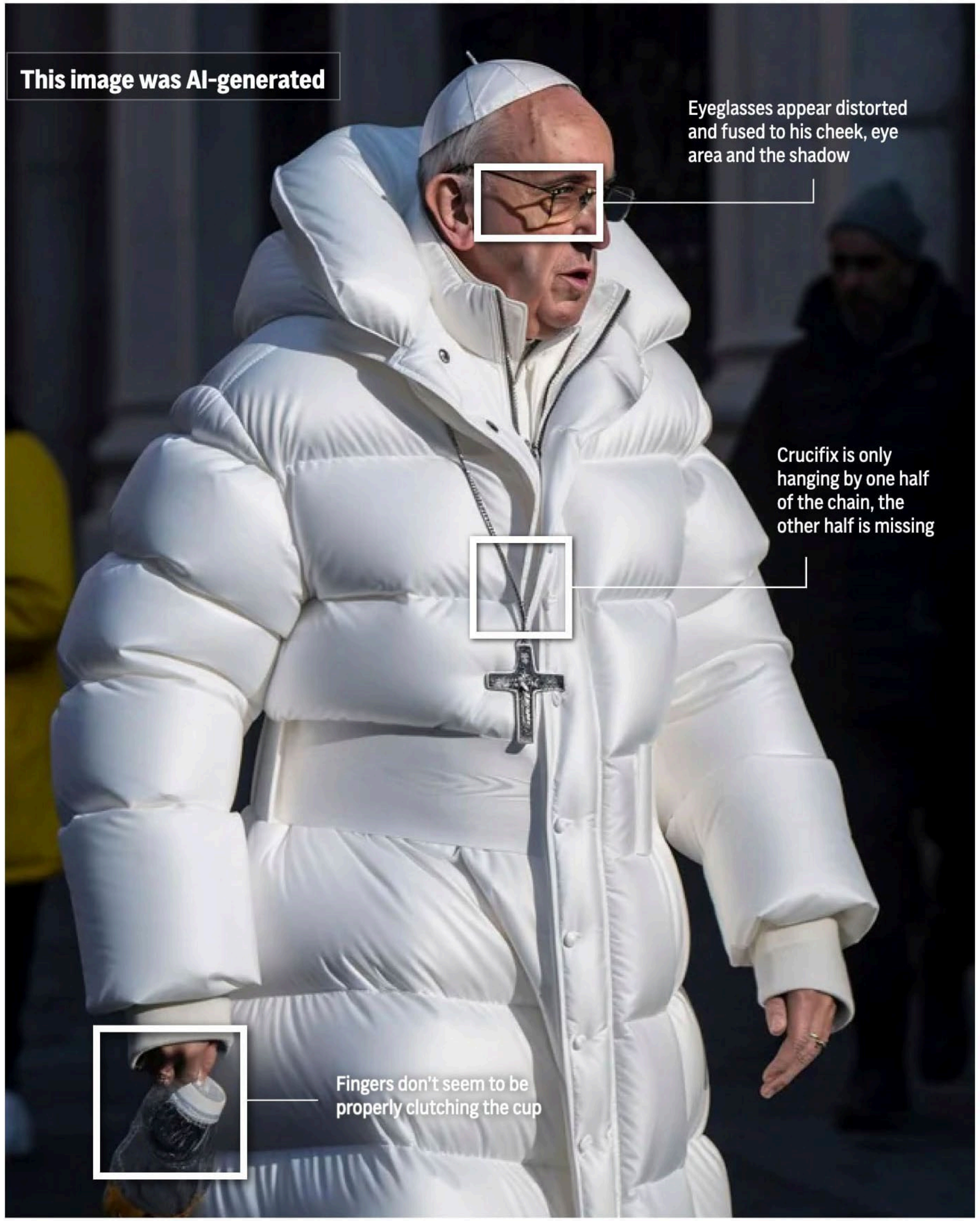
You won't get in trouble for reporting something that may be incorrect.

The Big Takeaway

- Sextortion of minors is child sexual abuse.
- Child sexual abuse material is child abuse.
- Child abuse triggers mandated reporter duty.

EXAMPLES AND ID TECHNIQUES

Ways to Identify a Deepfake



This image was AI-generated

Eyeglasses appear distorted and fused to his cheek, eye area and the shadow

Crucifix is only hanging by one half of the chain, the other half is missing

Fingers don't seem to be properly clutching the cup

Deepfake Catcher Apps

- FakeCatcher, Detect Fakes Website by MIT, Deeptrace

Search Engines

- Search engines can give details like how far an image has traveled, unearth counterfeits, or show you where it originated from.

- Use search engines to do a reverse image check:

  - **Take a screenshot** from the video you think may be a deepfake.
  - **Upload this image** onto a search engine like Google.
  - **Authenticate the image's history**: Where it's been used, other sources who have posted or used this, or if there is another version that exists somewhere.

- **Examine the face** - Experts advise looking closely at the edges of the face. Does the facial skin tone match the rest of the head or the body? Are the edges of the face sharp or blurry? Are there some awkward proportions? Have any facial moles or marks changed?

- **Look at the mouth** - Do their lip movements match the audio perfectly?

- **Look at the teeth -** Are they clear, or are they blurry and somehow inconsistent with how they look in real life?

- **Keep an eye on the eyes** - Deepfakes don’t blink as we do, and their eyebrows might not move like ours either. Look at glasses.

- **Unnatural body movements** – Is the person moving erratically? Do the movements from one image to another line up. Are the positions of the head and body uncoordinated?

- **Odd coloration** - Keep an eye on the video color, and on whether the lighting fluctuates from one frame to the next. Is the person’s skin color unusual? Are there shadows where there shouldn’t be?

- **Awkward facial expressions/emotions** - One of the first obvious things to take note of is their facial expressions. See if their face or nose is pointing another direction or if something doesn’t look right. Does their face not match the emotions according to the conversation?

- **Unnatural hair** - Do they appear to have perfect hair without any flyaways? Is there no single stray hair in sight? What about the facial hair?

- **Irregular reflections or shadowing -** Deepfakes often do a poor job of recreating shadows and reflections. Look closely at reflections or shadows on surrounding surfaces, in the backgrounds or even within participants' eyes.

- **Pupil dilation -** Dilation is more challenging to identify unless the video is presented in high res. In most cases, AI doesn’t alter the diameter of pupils, leading to eyes that appear off. This is true if the subject's eyes are focusing on objects either close or far away.

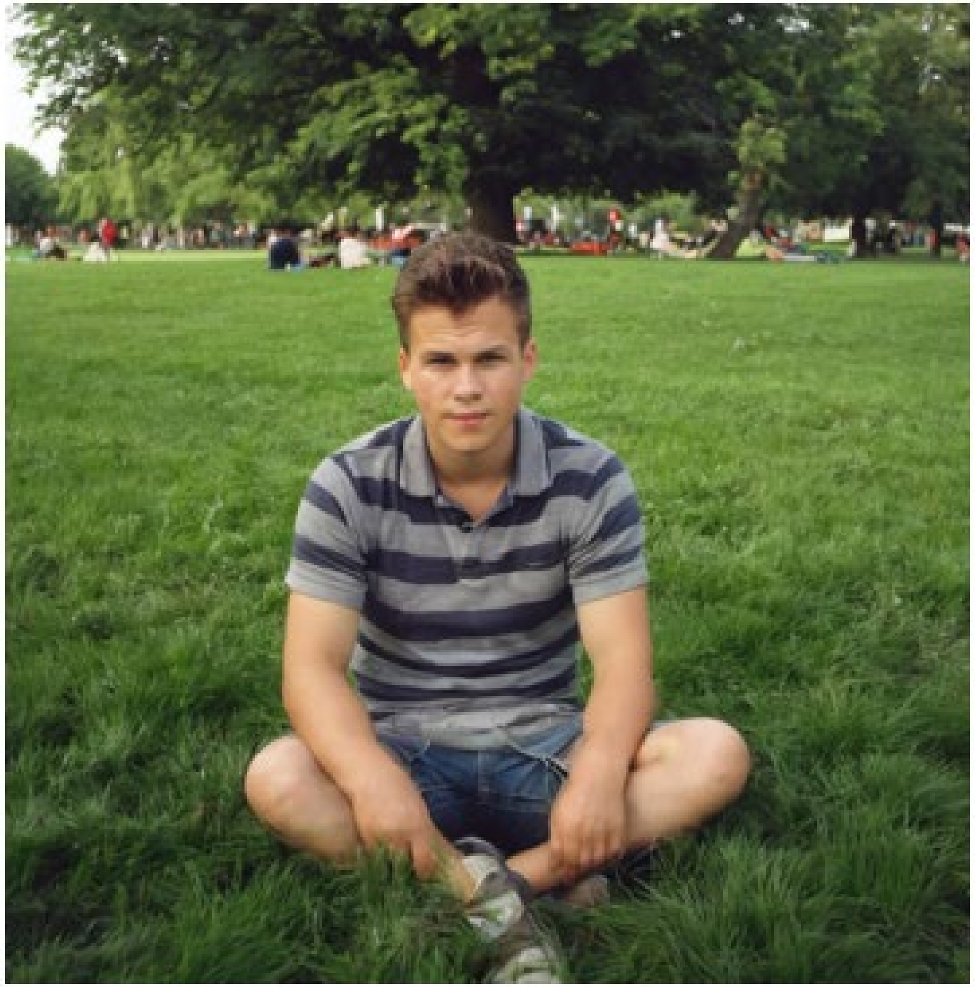
- **Artificial audio noise -** Deepfakes often add artificial noise, or artifacting, to audio files to mask changes in audio.

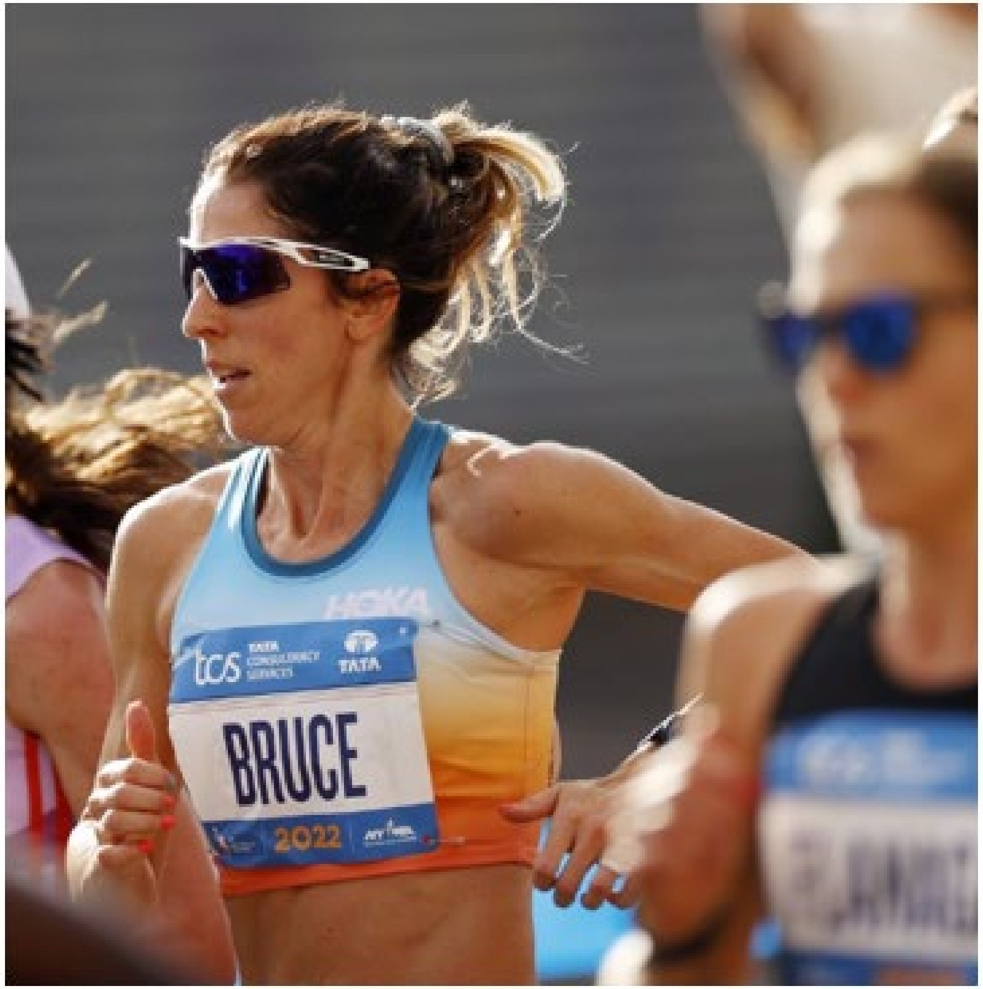
Deepfakes on Social Media

TikTok

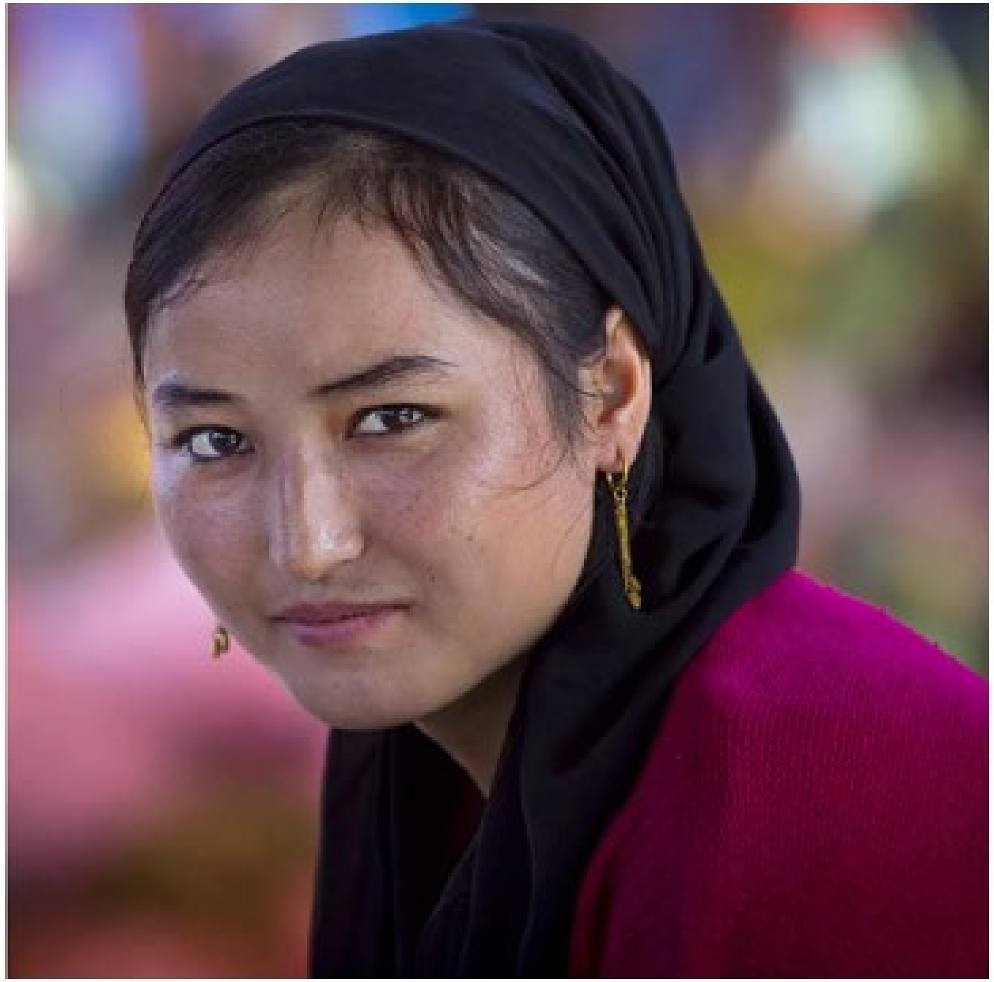
Northwestern University’s Deepfakes Study: “DeepFakes, Can You Spot Them?”

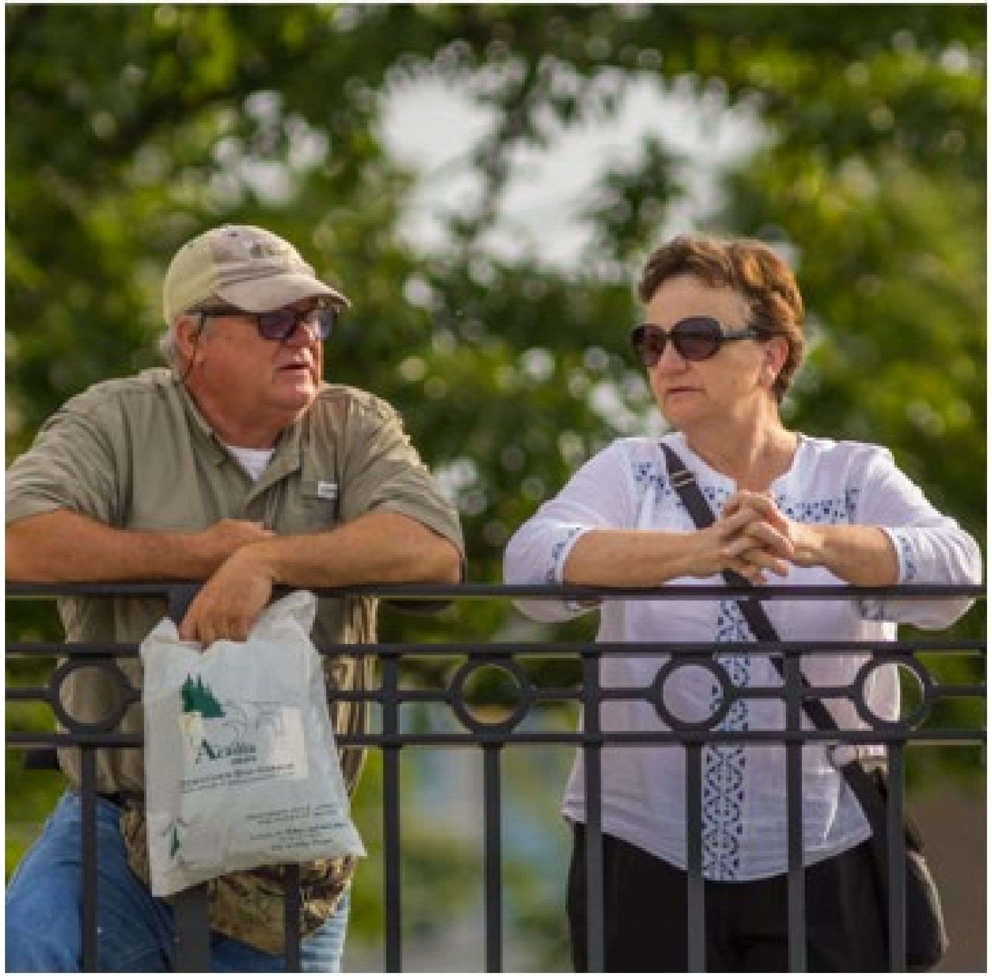
Are They Real? How Can You Tell?



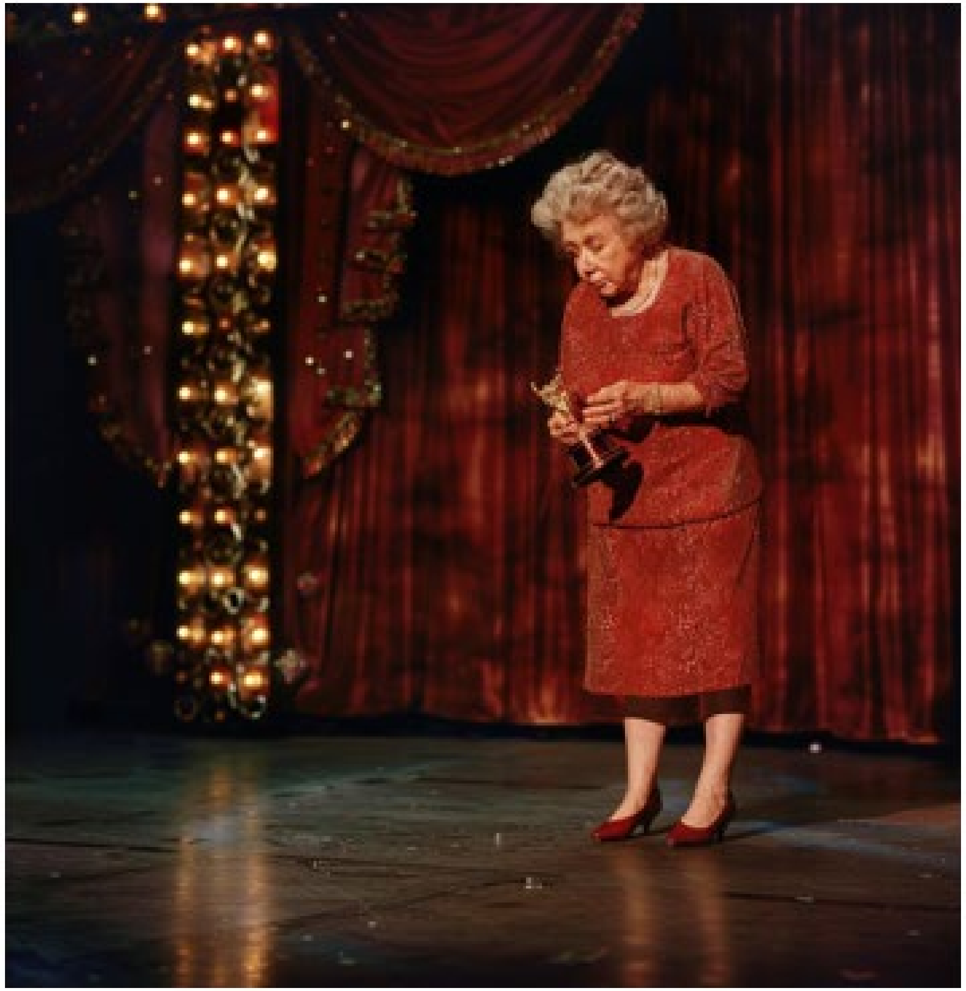


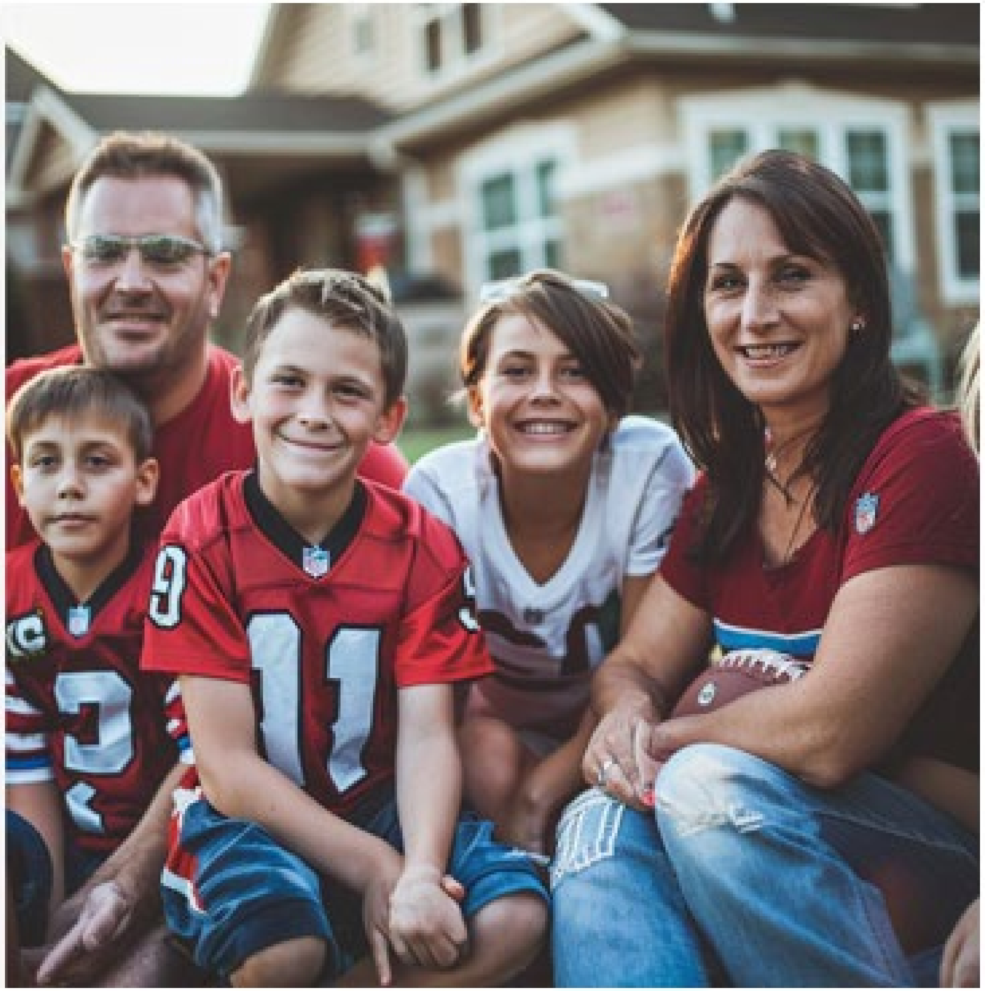
Are They Real? How Can You Tell?



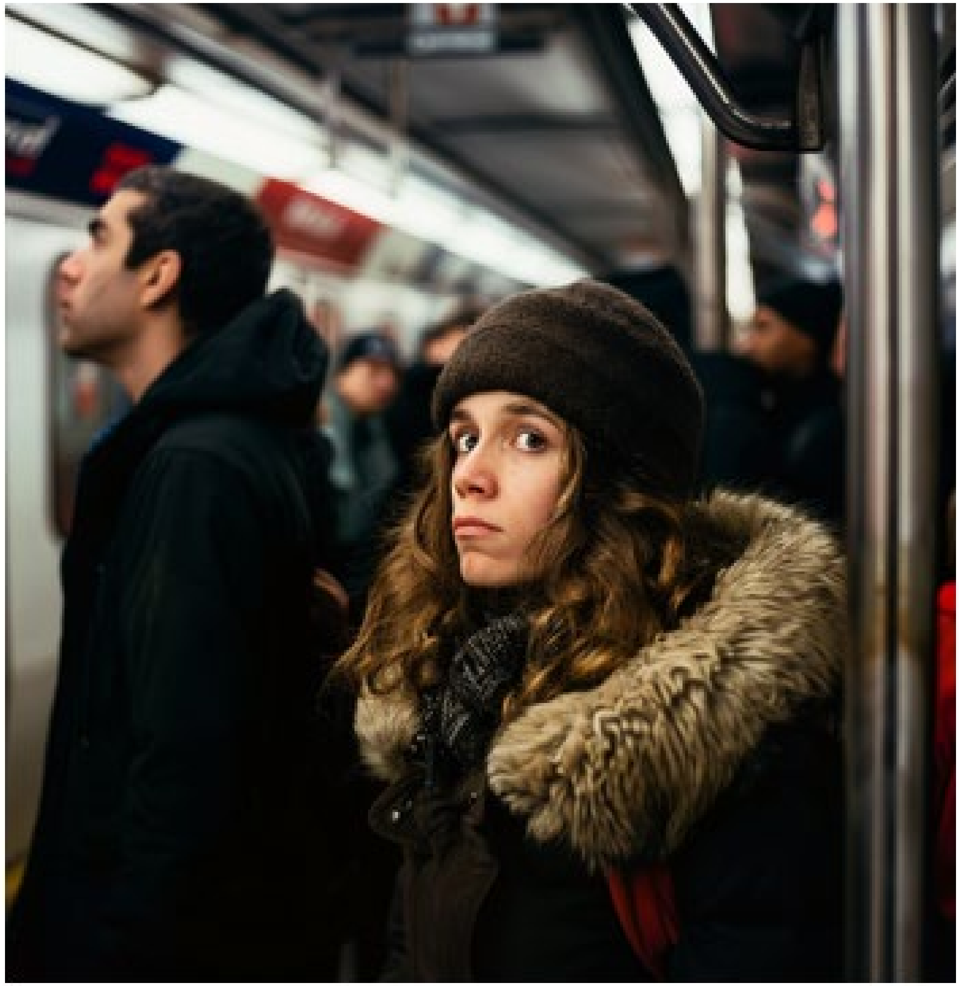


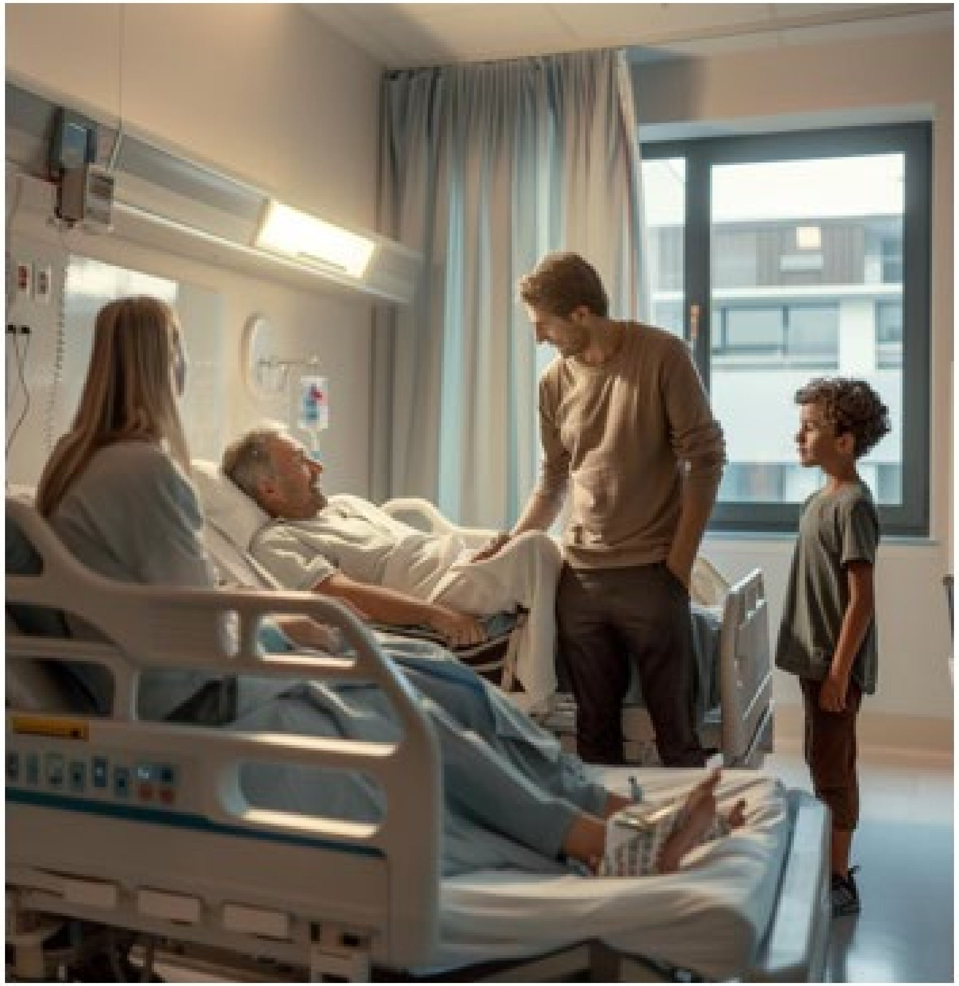
Are They Real? How Can You Tell?



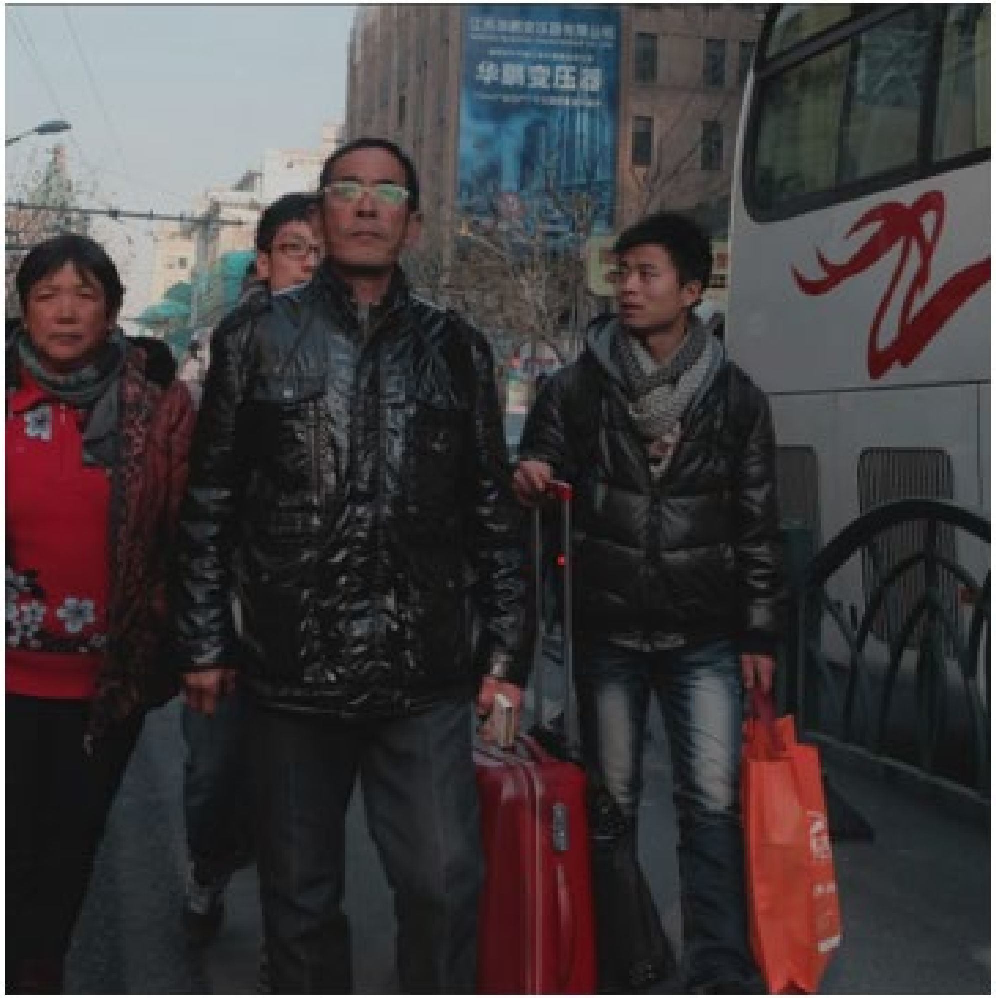


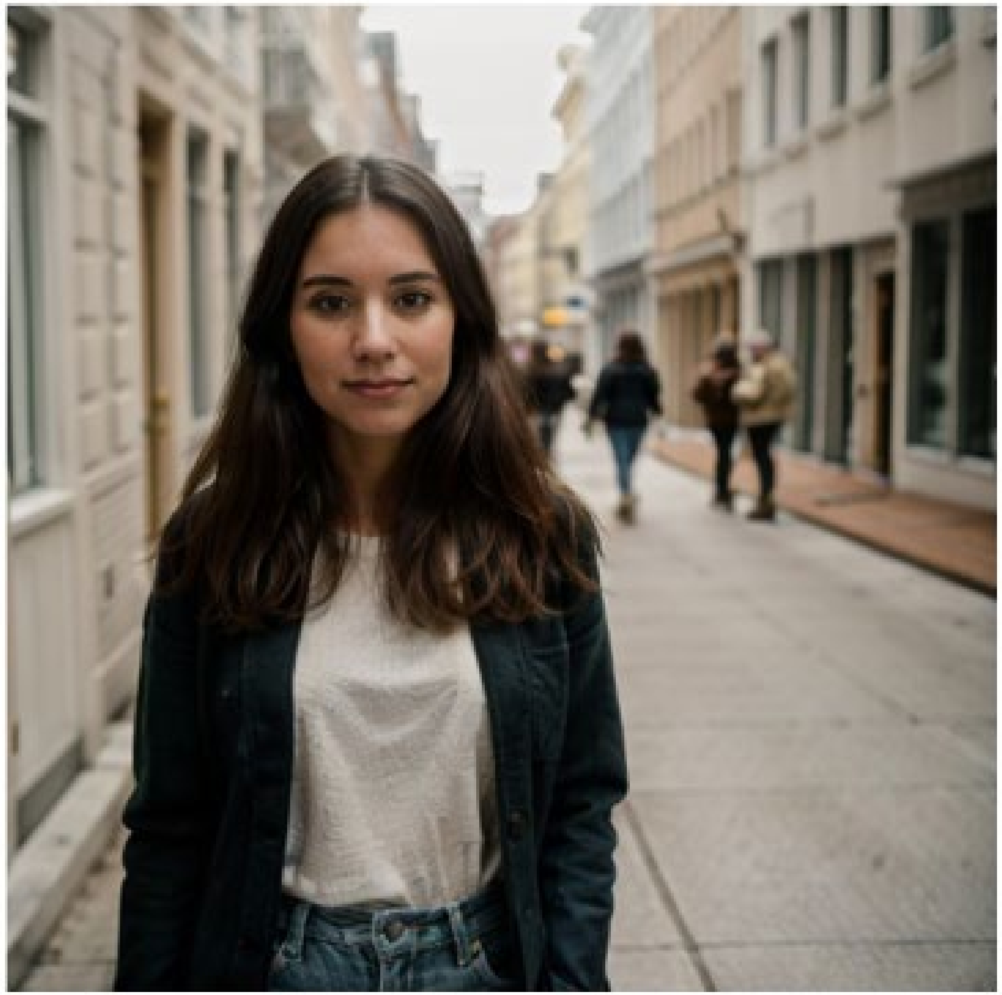
Are They Real? How Can You Tell?

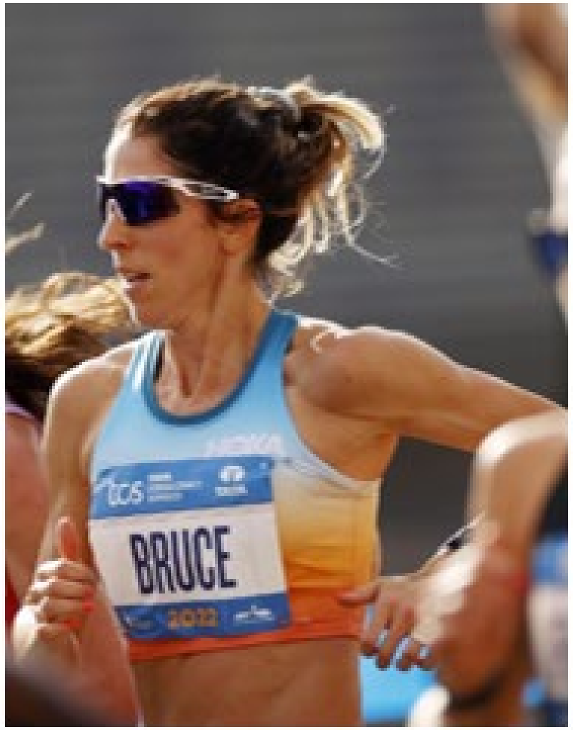




Are They Real? How Can You Tell?



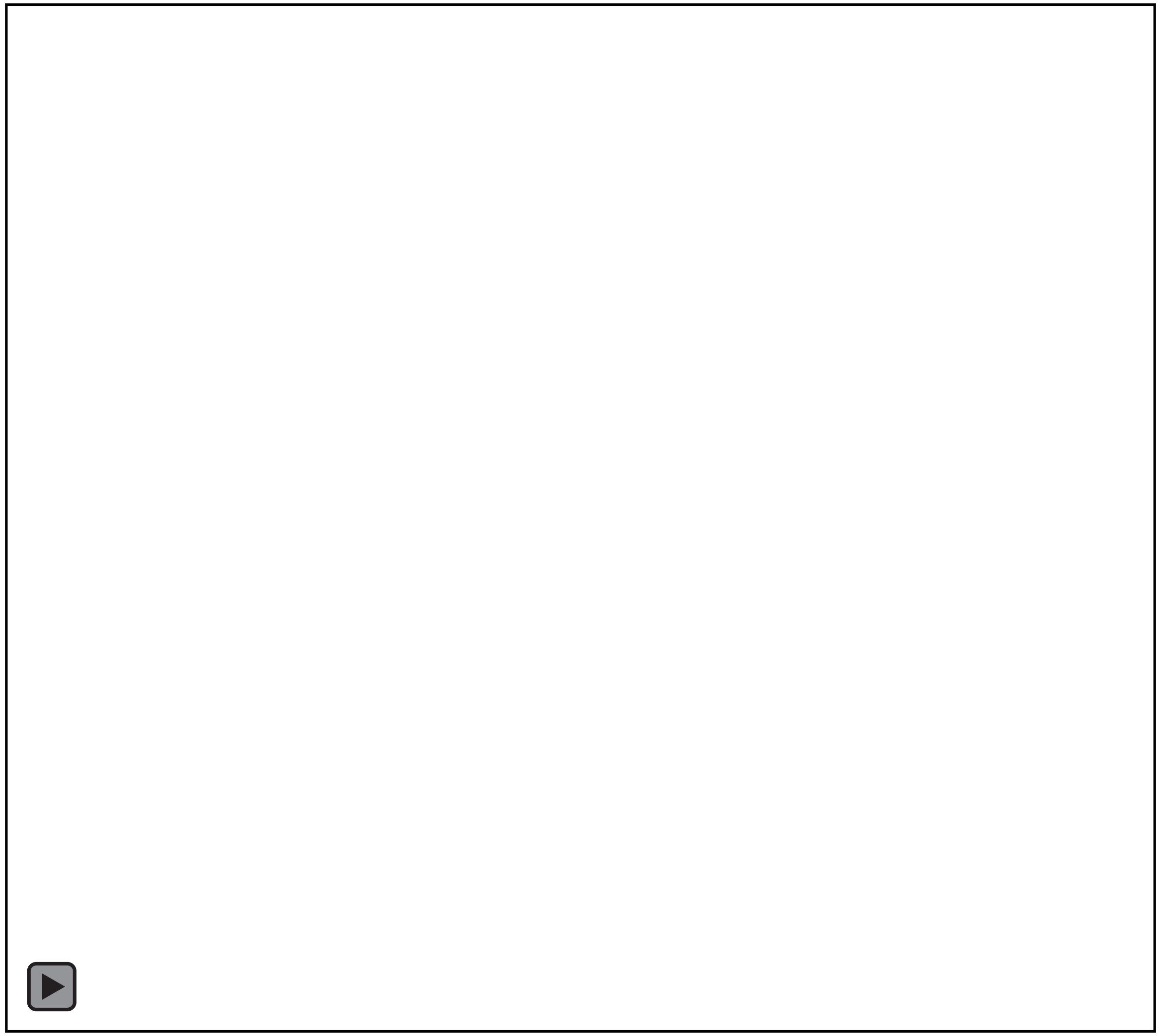


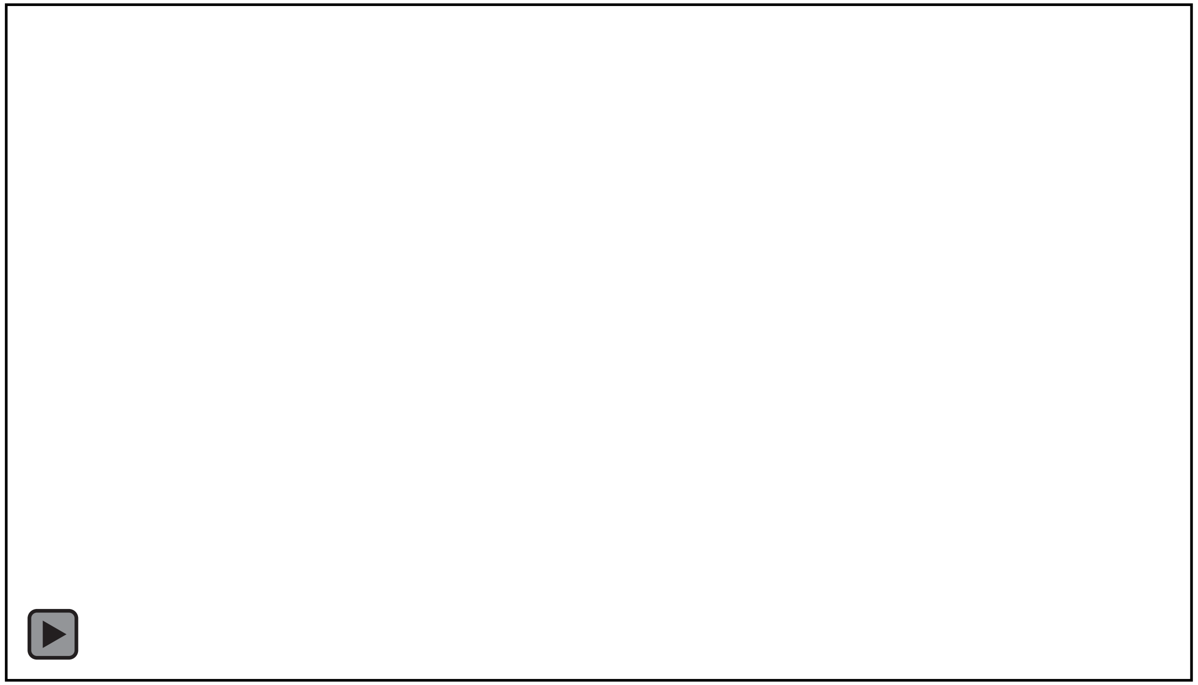


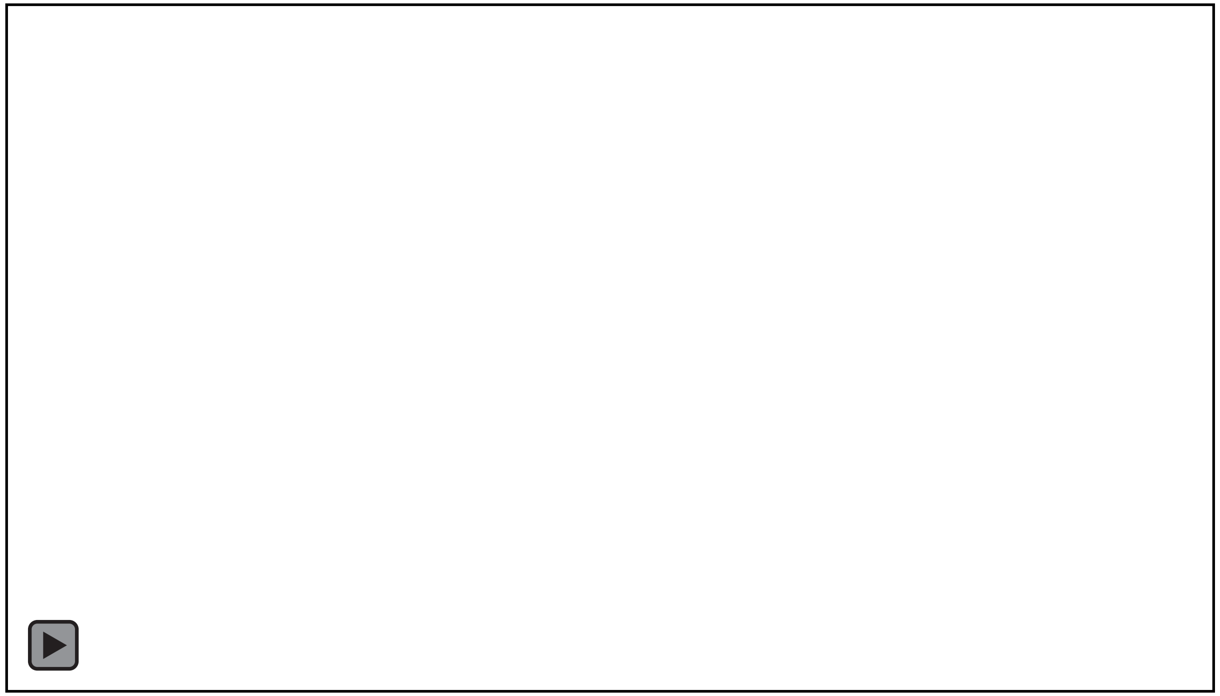
4 ARE REAL. THE REST ARE A.I.

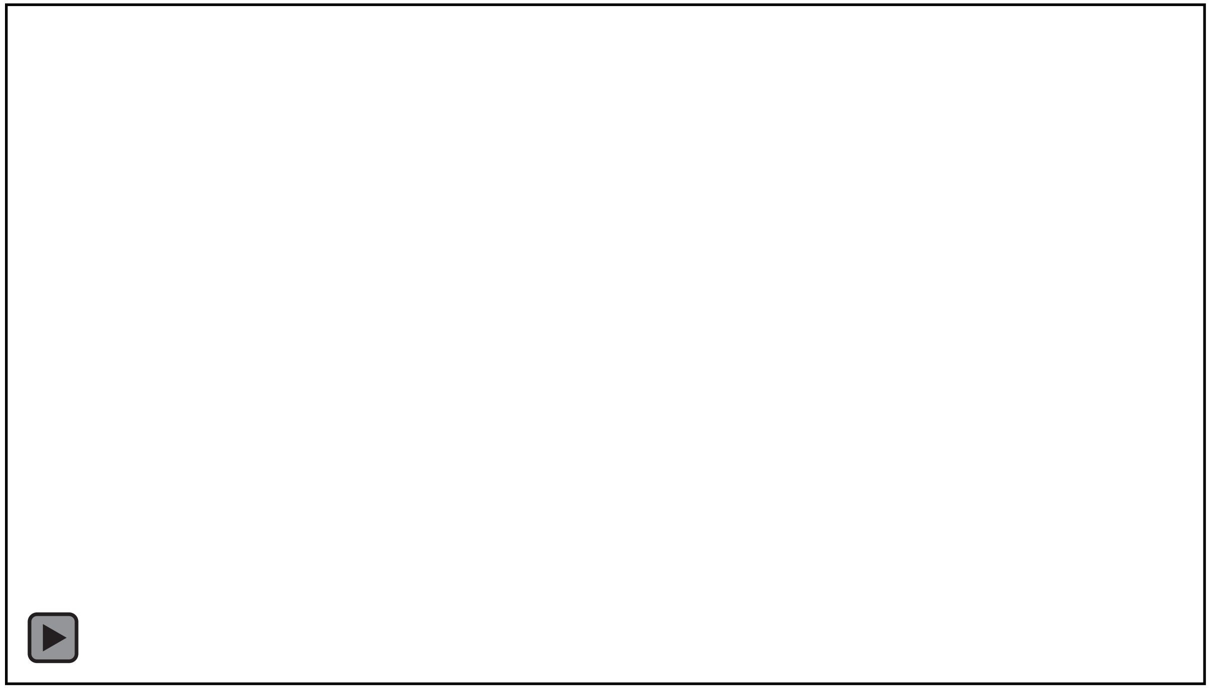
The Future of Deepfakes is Here

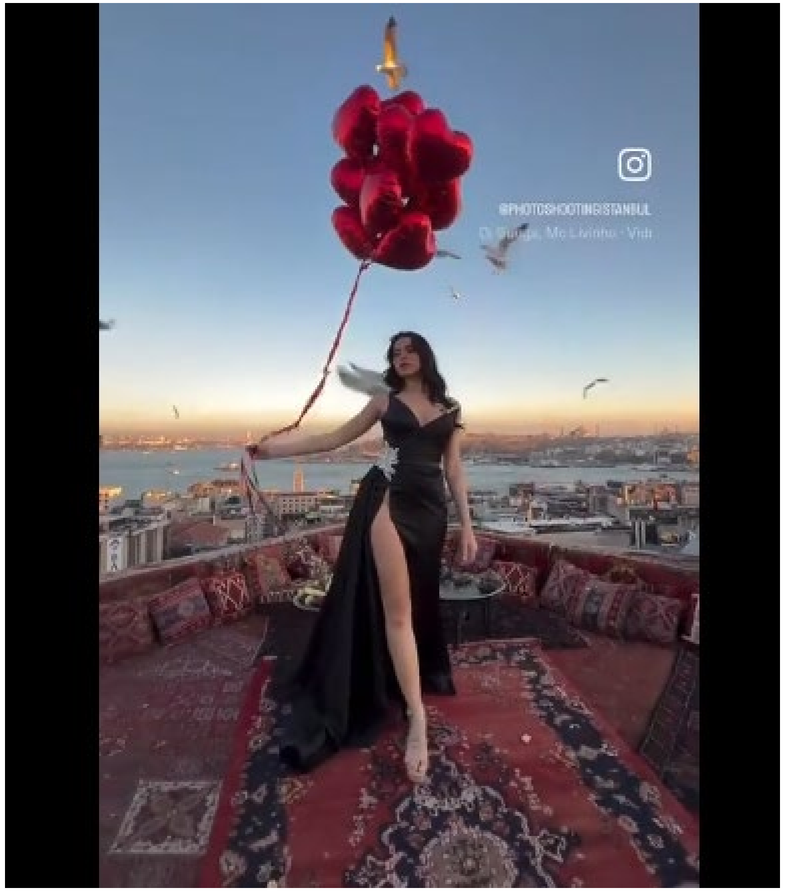
New York Times Article on Google Veo: “A.I. Videos Have Never Been Better. Can You Tell What’s Real?”

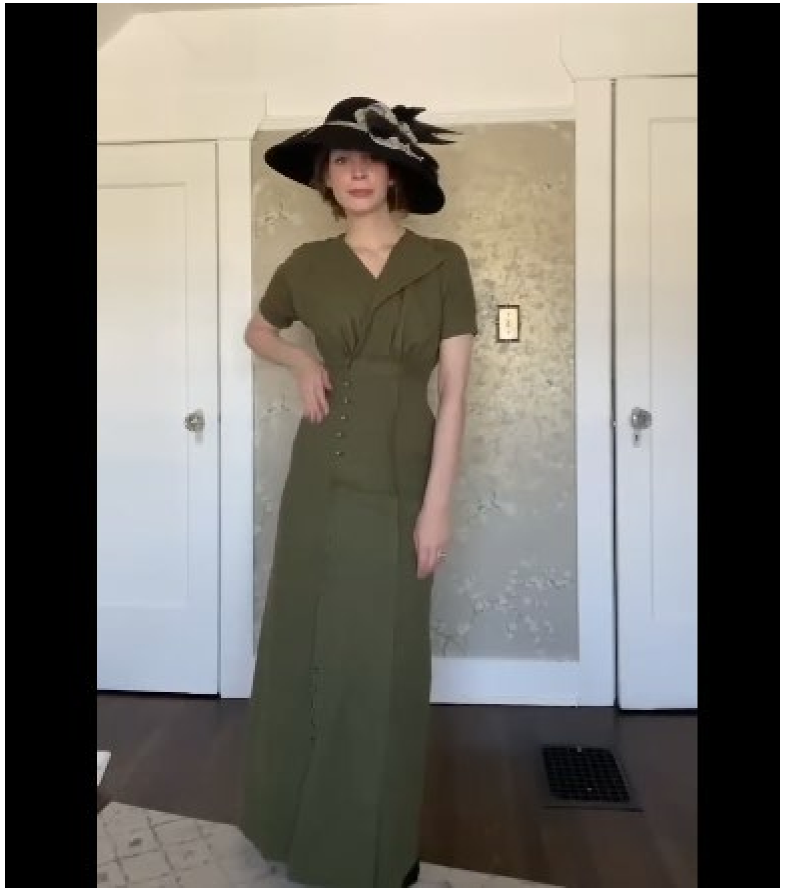












2 ARE REAL. THE REST ARE A.I.

Magic Hour AI
(www.magichour.ai)

Parrot AI
(www.tryparrotai.com)

PREVENTION AND CONSIDERATIONS

The Reality

The Bad News: You cannot stop A.I. and technological advancement.

The Reality: Kids today aren’t doing anything we weren’t doing

The Good News: At least there’s a playbook now!

“TAKE IT DOWN” Act

“Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks” Act.

In addition to making it to illegal to share online nonconsensual, explicit images — real or computer-generated — the law requires tech platforms to remove images within 48 hours of being notified about them.

The law boosts protections for victims of revenge porn and nonconsensual, AI-generated sexual images, increase accountability for the tech platforms where the content is shared and provide law enforcement with clarity about how to prosecute such activity. Previously, federal law prohibited creating or sharing realistic, AI-generated explicit images of children. But laws protecting adult victims varied by state and didn't exist nationwide.

YOU Have a Part in Prevention!

Report deepfakes!

- **FBI –** Email CyWatch@fbi.gov
- **Nat'l Center for Missing and Exploited Children ("NCMEC")** – "Take it Down" Initiative
- **Cyber Civil Rights Initiative** - An organization that provides a 24-hour crisis helpline, attorney referrals and guides for removing images from social media.

YOU Have a Part in Prevention!

Report deepfakes!

- **EndTAB** (Ending Tech-Enabled Abuse) - An organization that provides resources for education and reporting abuse.

- **Cybersmile** - A nonprofit anti-bullying organization that provides expert support for victims of cyberbullying and online hate campaigns.

- **UK-Based Childline:** Provides a 24/7 confidential helpline and support for children and young people under 19 via phone, online chat, or text.

**Report deepfakes.
It takes 1 minute.**

Be the person that stops the bystander effect.

YOU Have a Part in Prevention!

Open communication is only 1 part.

How YOU react, will determine if a child comes to you.

You may listen, hug them, tell them it's OK and it's not their fault.

But then you take away their phone.

Guess what will happen?

They'll never come to you again.

YOU Have a Part in Prevention!

Embrace the maturity of the child coming forward and asking for help.

Make sure they have additional avenues for help.

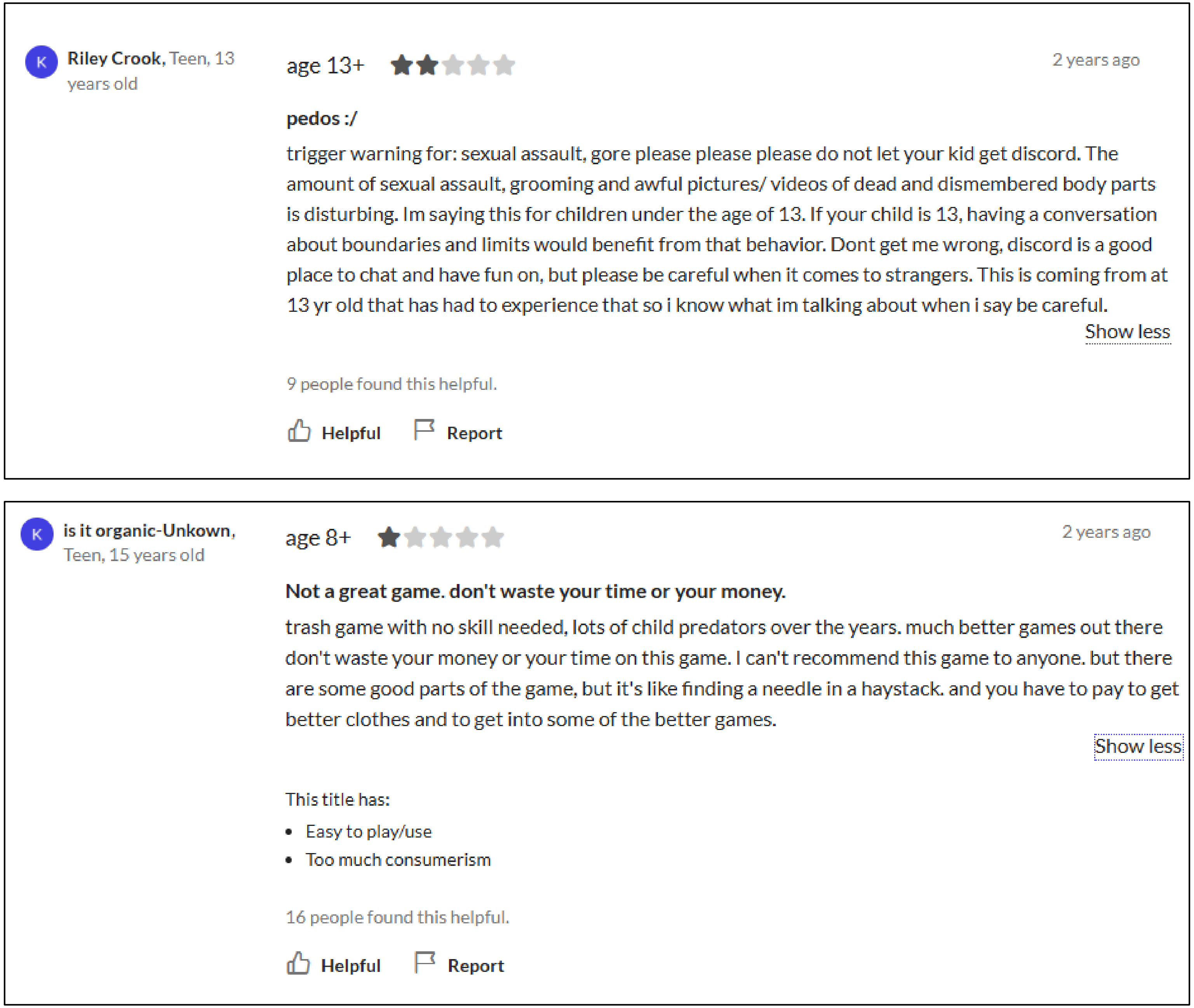
YOU Have a Part in Prevention!

NCMEC Netsmartz: Age-appropriate online safety education

Other reporting avenues:

- **Crisis Text Line (741741)**
- **988 Text Line** – 24/7 access in English and Spanish via phone, text, or online chat
- **Teen Line** (www.teenline.org): 24/7 access via phone, text, or email
- **CA Youth Crisis Line**

WWW.CommonSenseMedia.Org



K **Riley Crook,** Teen, 13 years old

age 13+ ★★☆☆☆

2 years ago

pedos :/

trigger warning for: sexual assault, gore please please please do not let your kid get discord. The amount of sexual assault, grooming and awful pictures/ videos of dead and dismembered body parts is disturbing. Im saying this for children under the age of 13. If your child is 13, having a conversation about boundaries and limits would benefit from that behavior. Dont get me wrong, discord is a good place to chat and have fun on, but please be careful when it comes to strangers. This is coming from at 13 yr old that has had to experience that so i know what im talking about when i say be careful.

Show less

9 people found this helpful.

Helpful **Report**

K **is it organic-Unkown,** Teen, 15 years old

age 8+ ★☆☆☆☆

2 years ago

Not a great game. don't waste your time or your money.

trash game with no skill needed, lots of child predators over the years. much better games out there don't waste your money or your time on this game. I can't recommend this game to anyone. but there are some good parts of the game, but it's like finding a needle in a haystack. and you have to pay to get better clothes and to get into some of the better games.

Show less

This title has:

- Easy to play/use
- Too much consumerism

16 people found this helpful.

Helpful **Report**

Reminder:

Security updates on your phone automatically changes settings.

You will need to review app settings after ***both*** phone and app updates.

THANK YOU!

QUESTIONS?

- Jill Ostrove, Esq
- JOstrove@eplaceinc.com
- (760) 681-3098
- www.eplacesolutions.com