TEN DIGITAL TRENDS FOR 2020

DEEPFAKES: THE GOOD, THE BAD, AND THE MURKY

JAYR SOTELO:
Motion Graphics/Group Head

Hopefully by now all of us know what a "deepfake" is; if you don't, I will give you a quick recap. Deepfake is a technique for video manipulation common in academic research institutions – and by amateurs in online communities – using machine learning (generative adversarial neural networks or GANNs) to augment and distort a base video in differing degrees. Deepfakes commonly involve replacing or augmenting an actor's face with another to change what they say or their appearance while saying it. The phrase "deepfake" was coined in 2017, and it's a portmanteau of the term "deep learning" and "fake."

This technique has been used for entertainment purposes and academic research for now. The ability to put somebody's face into an existing video creates a lot of different reactions, from laughter and awe to paranoia and fear. Just type the word "deepfake" into YouTube or Google, and you will get explainer videos and profiles that specialize in taking movies or TV shows and swapping actors' faces with other actors' faces.

It's like that '90s movie *Face/Off*, but without getting Nicholas Cage to act like John Travolta for an extra cringe factor. If you haven't had the privilege of watching *Face/Off*, then you can search YouTube profiles like Ctrl Shift Face (226,000 subscribers) and The Corridor Crew (2.4 million subscribers) that have shown how these are made and even specialize in making deepfake videos for the fun of it.

A good reference for how much reach this new technique has in current popular culture is the deepfake that BuzzFeed did a while ago featuring President Obama saying things that Jordan Peele is mouthing, and it even shows a side-by-side comparison of Obama and Peele so you can see that it is fake. Ellen DeGeneres did something very similar as well, although that one was made just using video and compositing tools, no deep learning needed.

THE BAD

Much has already been written (Wired, MIT Technology Review, HelpNetSecurity) about what this means to our future media habits and our culture of Internet and content freedom. At the most basic level, how do we know if the videos we are seeing are real? There is already problematic content that misinforms people – you know, actual fake news – imagine the day when this technique gets so sophisticated that "we" are not going to be able to discern fake from real. Another way this is getting used is with voice, being able to synthetically fake the voices of popular world leaders and celebrities regularly in the media.

Well, I have really bad news for you. Top artificial intelligence researchers across the country say there is no reliable method for detecting deepfake videos programmatically, even by analyzing videos for telltale indicators of a fake: assessing light, shadows, blinking patterns, and even how a person's facial movements – such as the angle they tilt their head when they smile – relate to one another.

And just this September, a new Chinese app called "ZAO" has gone viral. It uses only a picture of you to insert yourself into movies as Leonardo DiCaprio or one of the cast members from *Game of Thrones*. This app can only be accessed using a Chinese phone number, but a number of hackers have been able to use it in the U.S. The technology, sophistication, and ease to create deepfake videos – already outpacing our ability to detect and police this content – is accelerating faster than techniques designed to prevent bad actors' use of deepfakes.

THE MURKY

New academic and commercial research into deepfake detection is constantly underway, with new articles published almost every month. Yet with each update, it seems more and more difficult to overcome the challenges that the deepfake technology creates – especially when paired with the scale and impact videos can achieve through the distributed media landscape.

Delip Rao, Vice President of Research at the AI Foundation, agrees that the challenge is far greater than simple detection and says that these papers need to be put in perspective. One deepfake detection algorithm unveiled in May 2019 boasted 97 percent accuracy, for example, but as Rao notes, that 3 percent could still be damaging when thinking at the scale of Internet platforms. "Say Facebook deploys that [algorithm] and assuming Facebook gets around 350 million images a day, that's a LOT of misidentified images," says Rao. "With every false positive from the model, you are compromising the trust of the users."

It's incredibly important we develop technology that can spot fakes, says Rao, but the bigger challenge is making these methods useful. Social platforms still haven't clearly defined their policies on deepfakes, as Facebook's tussle with a fake Mark Zuckerberg video recently showed, and an outright ban would be unwise.

If there is no way to discern the manipulation by assessing the visual part of the video, is there another way of analyzing the video to figure out if it is augmented? This opens a very interesting security and authenticity concept that is currently being explored by tech and media companies: Is it possible to encode metadata to any video file that verifies post-capture handling? This concept is akin to creating a new file extension – an edit record that is inseparable from the data of the file itself that would allow users to certify a video is genuine. An ideal form of this concept also gives you a history of how, when, and if the video has been modified and with what end goal. If this concept were adopted, the onus is then placed on file-sharing platforms, social media sites, and media companies to authenticate files prior to publishing in any public setting. Despite the promise of this approach, innumerable questions remain to be answered before it could be implemented. Some are technical – does this system require a distributed ledger (such as blockchain) to independently validate the metadata attached to various files? And some are logistical – how would this system adapt for files created prior to its existence?

Furthermore, let's say the verification mechanism works, but there are still videos of people doing things that could damage their public perception or otherwise threaten their livelihood. Not only in the court of public opinion or psychologically – some images simply cannot be unseen – but also legally as a video holds evidentiary value in the judicial system. In that scenario, who holds the responsibility for verifying the authenticity of the footage? Is the individual the sole arbiter of what they said and not what they were filmed doing?

In a world where video has played a pivotal role in shaping modern history, solving this challenge is vital. Without a viable method to consistently spot the deepfakes, the broader consequences if the authority of video slips away are enormous.

When it comes to voice, there is a lot of research and content being generated by deep machine learning or artificial intelligence (AI). Consider *The Joe Rogan Experience* AI experiment. Nearly 1,300 episodes of *The Joe Rogan Experience* provided the training data for a system to create fake ramblings with not just Joe's voice but some of his beliefs. This could be labeled under murky, for taking people's words and using them in a different context, but it could also be labeled under good, as the idea behind the system might be applied to keeping some of the great thinkers, writers, and podcasters alive in an AI form. Imagine George R.R. Martin passes away without finishing the last books in the *Game of Throne*s series – well, maybe we could use AI and deep learning to finish his grand opus.

THE GOOD

There are clearly perils brought on by the existence and unregulated use of this technology. But what about the positives? When it comes to advertising and entertainment, this could be a huge disruptor. The potential for personalization and magical experiences is boundless. For a very simple example, imagine "Elf Yourself" revitalized for 2020, taken to a whole new level. With just a single photograph, people can be dropped into fully realized narratives rather than simple animations.

Pick any movie or sports footage, and put whomever you want at the center of it with no editing, coding, or technical expertise needed. Talk about catering to the individual. Or another way of using this technology would be taking a new movie and being able to experience it as a home casting director – replaying it with your favorite actors or actresses in whatever roles you should choose.

Deepfakes could be a new platform paralleling the trend of having advertising be closer to the audience, either through user-generated content or by playing within the context of the media. Despite the risks, I remain hopeful this would be a good opportunity for brands to mine for creative potential, making every one of us the star of their ads. And when talking about imagination and creativity, this is where we should be thinking about how we can use these types of technologies to enhance the future and how to caution the young and old alike on the landscape of media that we will be seeing in the future.

Three decades in, the Internet keeps proving that what we think is impossible can be made possible. But also what we believe to be true can be challenged by technology and the people seeking to manipulate it.

Deepfakes are a rich topic for discussion, involving social media mechanics, democratized content creation, celebrities, politicians, media responsibility, and ethics. But at the end of the day, what we consume in our digital diets needs to follow the same basic rules as our food consumption: We have to understand what things are indulgent and what are good. If we want to keep a healthy mind, we need to feed it content that is good for our minds. And that sometimes means limiting our click-bait, if not completely eliminating it.

Despite the constant change, with the continual revolution from impossible to possible and back again, I feel the statement "don't believe anything you see on the Internet" remains as true today as when it was first coined in the '90s. Be careful out there and beware of the deepfakes.

REFERENCES

1. Deepfake. YouTube.
2. Deepfake. Google.
3. Ctrl Shift Face. YouTube.
4. The Corridor Crew. YouTube.
5. Knight, Will. "Even the AI Behind Deepfakes Can't Save Us From Being Duped." Wired, October 2, 2019.
6. Chen, Angela. "Three Threats Posed by Deepfakes That Technology Won't Solve." MIT Technology Review, October 2, 2019.
7. Balasubramaniyan, Vijay. "Deepfakes and Voice as the Next Data Breach." HelpNetSecurity, October 21, 2019.
8. Sabir, Ekraam; Cheng, Jiaxin; Jaiswal, Ayush; AbdAlmageed, Wael; Masi, Iacopo; and Natarajan, Prem. "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos." USC Information Sciences Institute, May 16, 2019.
9. Kelly, Makena. "Instagram Will Leave Up Deepfake Video of Mark Zuckerberg." The Verge, June 11, 2019.
10. Elf Yourself.
11. RealTalk: We Recreated Joe Rogan's Voice Using Artificial Intelligence. YouTube.
12. Vincent, James. "This AI-Generated Joe Rogan Fake Has to Be Heard to Be Believed." The Verge, May 17, 2019.

Jayr Sotelo

Growing up in Mexico, when Jayr wasn't playing soccer, he was watching *The Simpsons*. Homer taught Jayr three things: how to speak English, that animation was his life's calling, and that donuts really are man's best friend.

Jayr knew that textbook doodles alone weren't enough to launch a career, so he went the old-fashioned route: college. At Texas Christian University, he majored in graphic design with a minor in computer science. He attended the art academy of Trier in Germany. Then he backpacked and doodled his way through Budapest, attending the Hungarian University of Fine Arts. He'd come a long way since Homer Simpson.

Jayr began his career at Janimation Studios working on McDonald's, MTV, and VH1. He joined The Richards Group in 2010 and founded our motion graphics group, where he now creates award-winning doodles for Orkin, Clamato, The Home Depot, and Ram Trucks.

Jayr lives in Dallas with his wife, Gaby, and the two characters he is most proud of creating–his kids, Jimena and Melquiades.

JAYR SOTELO:
Motion Graphics/Group Head

THE
RICHARDS
GROUP

Have a question or would like to debate a particular trend? Please feel free to contact us. We love this stuff.

Go to richards.com/trends to read the rest of the Ten Digital Trends for 2020.

2801 North Central Expressway
Suite 100
Dallas, TX 75204-3663
214-891-5700