

Think About This...

DEEPFAKES

WHAT YOU NEED
TO KNOW

Shelly Palmer

DEEPPAKES: THE DAWN OF THE POST-TRUTH ERA

For about 200,000 years, modern humans have relied on our eyes and ears to separate truth from lies and fact from fiction. Even if we ignore the rise of fake news (**and how difficult it is to do anything about it**), technology (like deep learning) is on the verge of making it impossible to know if what you are seeing and hearing is real or fake.

WHAT IS DEEP LEARNING?

Deep learning, a subset of **machine learning**, is loosely based on the way engineers and scientists believe biological brains work. The “deep” in deep learning refers to the number of layers (depth) of data transformation. If you want to go deep on the subject (pardon the pun), search “**credit assignment path**” and pick a scholarly article that meets your needs. Also known as hierarchical learning or neural networks, deep learning systems learn by trial and error. Said differently, as opposed to being explicitly programmed to recognize a particular object, a deep learning system will learn how to recognize an object after a number of attempts. Sort of the way a newborn baby picks up a rattle and tries to eat it, then bangs it on the floor, then tries to break it, then bangs it on its head, then shakes it. Each time, the baby is learning what the rattle is, and what it isn't.

(You can start learning how to build deep learning tools right now by checking out **Google's** AI Education site. It's just one of thousands of sites filled with free exercises and information.)

DRIVING INNOVATION AND PRODUCTIVITY – THE LIGHT SIDE OF DEEP LEARNING

Deep learning can be used to identify nearly anything, like **determining cats from dogs in images** or identifying whether a photo of a kitchen features a Kohler or Moen faucet. Companies like Instagram use deep learning to **detect (and prevent) bullying** in photos and captions. Stanford University used deep learning to parse 2 million medical records to **predict when terminally ill patients will die** (which it does with 90 percent accuracy).

A neural network at MIT studied more than 60 hours of musicians playing various instruments and **learned to identify more than 20 instruments** that can be **isolated with a mouse click**. Nvidia used deep learning to **“fake” slo-mo videos** and create the 210 necessary frames to slow a standard 30 fps video to a 240 fps slo-mo video.

There are an unlimited number of good uses for deep learning systems. They have the power to truly change our world, but maybe not the way well-intentioned engineers intend.



DEEPFAKES – THE DARK SIDE OF DEEP LEARNING



Considering that one of the first artificial life forms created by human beings was a computer virus, it should not surprise you to learn that there are some seriously bad actors using deep learning.

In December 2018, someone used a deep learning algorithm to take Gal Gadot's face and **paste it onto an adult actress's body** in a porn video. The resemblance is close, although not close enough to fool anyone looking, ahem, too closely. The video came from a redditor named "deepfakes," who used open-source machine learning tools (like Google's TensorFlow) to create the video.

It's far from the first time we've seen this sort of thing. We've seen deep learning used creatively in Hollywood when the movie **Rogue One** de-aged Carrie Fisher and the movie **Captain America: Civil War** de-aged Robert Downey, Jr.

There's an inherent difference, however, between de-aging an actor to play a younger version of a character they're known for portraying and pasting a celebrity's face on an adult actress's body – but is there really a difference? To the untrained eye, both types of images look real. What does "real" mean?

To be fair, deepfakes has also done some less offensive deep learning fakery, like **putting Nicolas Cage in Raiders of the Lost Ark**. A while back, someone **inserted Harrison Ford into the movie Solo**, as though swapping the lead actor in that movie would have made it any less of an insult to Star Wars fans. (I digress.) That video came from YouTuber "**derpfakes**," who has been face swapping actors into movies for most of 2018.

WHAT'S NEXT?

We're rapidly approaching a point where actors will have a new source of revenue. Sooner than later, instead of offering Robert Downey, Jr. \$50 million to play Iron Man, the movie studio will offer him a nominal licensing fee to use his likeness applied to an anonymous actor. Sound crazy? It was the plot of the film [The Congress](#), which came out in 2013.

As technology improves (and it is improving at an exponential pace), it will become harder for us to discern what is real and what is fake. There's a difference, both in technical difficulty and in ethical responsibility, between making a public figure say whatever you want to them to say (remember [Talk Obama To Me?](#)) and swapping a celebrity's face with an adult film actress's face.

Deep learning tools are becoming easier to use and more readily available every day. It's only a matter of time before this technology is as easy to apply as an Instagram filter. When everyone is one tap away from creating an alternate reality, we will be firmly living in the post-truth era.

Was the video of that murder real or a deepfake? Was the person captured on that security camera actually the person? What if someone were to deepfake you into a revenge porn video? Sadly, there are an unlimited number of ways to use deepfake technology to change the arc of a story (and to alter the truth).

Remember, deepfake technology works with both audio and video. No voice is safe. No image is safe. Incorporate AR (artificial reality) techniques into this "end of the truth" scenario, and we simply will not be able to tell reality from fiction. If you can't believe your eyes and your ears, what can you believe?

In 1984, George Orwell's prescient novel about a dystopian world, the main character works at the Ministry of Truth. His job is to rewrite the daily news and then adjust historical records to support the goals of the state. When I first read the book, I thought it was written to scare me into supporting the cold war or just a brilliantly written hyperbolic view of life under totalitarian or authoritarian rule. I never imagined any of it was possible to do. Rewriting history, changing the pictures and the text, swapping out the audio, and manipulating the daily news were science fiction in Orwell's day. Today, these capabilities are just over the horizon.



DeepNudes

ALREADY ILLEGAL: DEEPUDES

Revenge porn has been outlawed in many jurisdictions around the globe. In July, Virginia became one of the first states to **outlaw the sharing of computer-generated pornography** aka deepfakes. It did so by amending an existing law which criminalizes revenge porn, making clear that the category now includes “falsely-created” material (image above).



DNC Deepfaked Itself

SCARY: DNC DEEPPAKE

The Democratic Party deepfaked its own chairman to highlight 2020 concerns. In early August at DEF CON – one of the world’s biggest conventions for hackers – attendees were told DNC Chair Tom Perez couldn’t make it for the DNC’s presentation. In lieu of that, he “Skyped in” and chatted. Except he didn’t. Do you know what Perez’s voice sounds like? Neither did those in attendance.

Bob Lord, the DNC’s chief security officer, led the exercise (with the help of experts in the field of AI) as a way to demonstrate the art-of-the-possible. This is what’s coming leading up to the 2020 election. The video speaks for itself. Please **watch it**.



The “fixed” Lion King

FUN: “FIXING” THE LION KING

It’s important to remember that not all deepfakes are malicious. After Disney purists didn’t like the (stunning) computer graphics in this summer’s release of *The Lion King*, they “fixed” the trailer by melding the original (animated) version’s character faces onto the new version’s animals. The resulting footage is so well done that it takes a second to realize what you’re looking at.



Aged with FaceApp

SOMEWHERE IN BETWEEN: FACEAPP

In case you’ve been living under a rock this summer, or simply took a vacation from the Internet, FaceApp is an app on Android and iOS that uses AI to “generate highly realistic transformations of faces in photographs.” Basically, FaceApp puts the power of deepfakes into your hand, making it a snap to make a person in a photo look older, look younger, or change genders.

Aside from the mild controversy – **Fast Company** discovered that every photo uploaded to the app was sent to FaceApp’s cloud servers (which you agreed to when you accepted the app’s terms and conditions, but clearly did not read) – the tech here is remarkable. FaceApp has been offering its photo effects since early 2017, but the results have gotten much better in the years since it launched.

While FaceApp has limited features, it’s one of the easiest and most popular deepfake creators available to the masses. Its speed and impressive results make it fun ... and also dangerous.

WHAT DOES THIS ALL MEAN?

The technology behind deepfakes was already easy to find and use last fall, but it's even more so today. Imagine where it'll be next summer when the 2020 election is four months away.

Did Elizabeth Warren really say that? Was that footage of Bernie Sanders in that attack ad real, or was it a deepfake? As political advertisements already twist candidates' words and manipulate the truth for the perfect soundbite, can you believe anything you hear when it can all be manufactured on any laptop you can find at Best Buy?

IT GETS WORSE



An image from thispersondoesnotexist.com

While news cycle after news cycle covered allegations of possible social media manipulation prior to the 2016 election, is anyone ready for what comes next? Let's quickly imagine how an AI-assisted system will exacerbate the problem.

Go to ThisPersonDoesNotExist.com. See that face? That lifelike photo? That's not a real person. The site "showcases fully automated human image synthesis by endlessly generating images that look like facial portraits of human faces."

Fun! Right? Only until you think about the possible ramifications. How much work would it take to build an entire history and lifetime for this person? Use AI to generate the person's name and backstory (maybe seed the system with an obituary so it could find their living relatives on social media, alongside a cache of "real" pictures)?

Create an email address, Facebook account, and Twitter profile. Use FaceApp to age the photo up 30 years, 60 years. Photoshop the character's likeness into his or her (real) family photos. Do the same in other photos.

Use AI-assisted software to generate tweets and Facebook posts. Use **procedural generation** to make the character's posts seem real, and not scheduled. Have the character retweet things. Reply to things. Like things. Interact with the world. It's like the "Yahoo Boys" **military scam**, but automated and amplified to the nth degree, powered by technology and by machine learning.

Build a **generative adversarial network** (GAN) to create retinal scans and fingerprints (the same way fake faces are generated on thispersondoesnotexist.com). Snag some social security numbers off the dark web, and ... Poof! We've created a virtual person that is practically indistinguishable from a real person. It would take quite a bit of research to figure out that this fabrication was not a living, breathing person.

Use **headless chrome** to create a few thousand accounts per day (I'm not going to tell you how to mask your network connection, but you'd easily do that).

Bias the AI "red" or "blue." Or, if you choose, "anarchist" or "union organizer," or make your created person appear sympathetic to whatever cause serves your purpose.

Train the AI to follow "real" accounts (or fake ones that you've created). The AI-generated tweets and retweets or Facebook posts and shares are going to look (and be) real. They just won't be from humans.

Activate your influencer network of a half-million accounts. Wait for election day and see how you did.

You may not like this game, but it is not going to be played by you or me. It's going to be played by nation-states with unlimited resources. Buckle your seat belt. This is going to be a rough ride.

CASE STUDY – ONLINE DATING IN A WORLD OF DEEPPFAKES



A BRIEF HISTORY OF ONLINE DATING

The first known “personals” appeared in a British agricultural journal circa 1685. Today, dozens of apps are eager to help you find true love. From Bumble to Tinder, from Farmers Only to Grindr, there’s an app for you, no matter your interests and proclivities. Facebook may be entering a crowded market, but it’s bringing a built-in user base that dwarfs the rest of the field.

CATFISH AND EVIL BOTS

Even if you’re looking for love in all the right places, online dating is a land of opportunity for fraudsters and scammers.

Catfishing – a scam where someone **creates a false identity** to deceive, harm, or prank another person – is a well-known dating scam. It’s so popular that it led to a movie in 2010, and a spinoff TV series that’s run since 2012. Catfishing is a low-tech predecessor to deepfakes.

When you use technology to automate fake (catfished) profiles, you get bot profiles. The bots interact at a very high level (in many cases they’re virtually indistinguishable from a real person txtng or emailing you). At a certain point in your “relationship,” the bots hit you with links to malware, do something unfortunate with online forms, or worse.

CATFISHING AND BOT HUNTING

Recognizing you're being catfished is a bit more difficult than recognizing you're chatting with an evil bot, but both scams use roughly the same tactics.

Telltale signs that the "person" you're chatting with is not who (or what) they say they are include, out of context messages instant replies, and the receipt of unprompted links.

DEEPPAKES WILL MAKE IT WORSE

Deepfakes will make it much harder to understand who is real and who is fake. Remember the lifelike photos on [ThisPersonDoesNotExist.com?](https://thispersondoesnotexist.com/) Whenever you refresh the homepage, a type of machine learning known as a Generative Adversarial Network (GAN) generates a new image that looks like a real person, but it's just a bunch of pixels arranged in a pattern the GAN has been "taught" to create.

Now, imagine these AI-generated people fully animated and able to chat with you in real time. A real person will exist behind the scene (at least in the beginning), but you won't know who it is. The exceptionally beautiful human being you're chatting with may be a sexual predator or a serious con artist – it will be very, very hard to tell.

Want an even deeper deepfake? Layer real-time translation software on your video (like the tech that [Wordly](#) is building), and... presto! You can have a real conversation with a fake person in dozens of languages. Wordly is great for stuff like speeches at trade shows and hotel concierge desks, but applied to a deepfake, it just expands the size of the scammer's universe.

YOU AIN'T SEEN NOTHIN' YET

In the very near future, deepfakes will become more common, more believable, and far easier to create. (Have a look at [FaceApp](#) or [Zao](#) for a preview of the insanity on the horizon.) Couple that with Facebook's problem with fake accounts – the company removed 2.2 billion fake accounts from its platform during Q1 2019 (Twitter also has a gigantic fake account issue) – and you have fertile ground for fraudsters, scammers, and hardened criminals.

Facebook is taking deepfakes very seriously. The company is to be commended for leading the DFDC. While it might be disappointing to learn that the would-be significant other of your dreams was a deepfake, it will be something else entirely when bad actors use the technology to do real harm.

Spend some time learning about GANs and Robotic Process Automation (RPA) this week. A primer on Headless Chrome would also be helpful. Then, search “ways to create fake accounts” on your favorite search engine. The results of that search query (which will surprise you) should make you want to read up on the work the **Deepfake Detection Challenge** (DFDC) is doing and get involved. It’s up to us to architect the future we want to live in.



An image from thispersondoesnotexist.com

POTENTIAL COUNTERMEASURES

Facebook has teamed up with the Partnership on AI, Microsoft, and academics from Cornell Tech, MIT, University of Oxford, UC Berkeley, University of Maryland, College Park, and University at Albany–SUNY to build the DFDC.

Deepfake detection is an enduring arms race that will never end. In case you are wondering... no, this technology will not protect the 2020 election from deepfakes. No science is up to that task.

Facebook's goal is to commission a realistic data set that will use paid actors, with the required consent obtained, to contribute to the challenge. This "benchmark data" will be used to help developers build better tools to detect deepfakes. Everyone should applaud this effort! As I've written about recently, deepfakes will be used extensively by both good and bad people.

THIS IS JUST THE BEGINNING

What will happen when we weaponize alternate reality? Will it redefine what it means to be a superpower? Will it be used to manipulate the stock market? Wage psychological war on a global scale? In a world where we can't tell the difference between reality and fantasy, what is truth? In the post-truth era, whom will we believe?



ABOUT THE AUTHOR

Shelly Palmer is CEO of The Palmer Group, a consulting practice that helps Fortune 500 companies with technology, media and marketing. Named LinkedIn's **Top Voice in Technology**, he is the co-host of "**Think About This**" on the Westwood One Podcast Network. He covers tech and business for **Good Day New York**, writes a weekly column for **Adweek**, and is a regular commentator on CNN and CNBC.

Follow @shellypalmer Visit: shellypalmer.com