Is This The Real Life?

Data/Science/Society
Student ID: ge38rak

Winter Term 2020/2021

Adrian David Castro Tenemaya
adrian.castro@tum.de
Technische Universität München

ABSTRACT

Even before internet became popular, fake news and people disguising themselves as someone else existed all over the world. From gossip spreading, to the use of fake passports, going "off the grid" and living in the woods. People always tried to get away from something they were, or something they have done, or tried to change the course of history by telling lies.

The effect however, has been ever so slow as the fastest mean of communication available at the time.

Communication at the speed of light is a game changer. In seconds, people across the globe know about fake cults, conspiracy theories, and much, much more. Today, a *tweet* alone can alter the course of history, save lives, or even take them away.

Today, people can be whoever they want online. A waitress, a business owner, a monk, or a dog. On the internet, no one knows you are a dog.

Today, people can change their faces on the internet, or create entirely new ones, never before seen.

Today, more than ever, pen is mightier, and faster, than the sword.

1 INTRODUCTION

Perhaps you killed someone, perhaps you were the most horrible criminal in the country for stealing cows and selling them over coast, or perhaps you were a peasant who found out that a prince just went missing, whose appearances were strikingly similar to yours. In old times, all these were valid motives to disappear from the community, and reappear sometime later with another name, another haircut, another identity.

It was also quiet easy to tell lies, and gossip. The only way you could verify someone's story, was to ask around, and believe your most trustworthy person, like your sister or brother. However, as the saying goes, "lies have short legs". Stories could only travel as fast as the fastest camel, or horse, and as far as what your kingdom allowed you to. Because of this, news spreading was slow, and mostly unreliable, as the person carrying the news would most likely alter it, forget it or change it completely.

In fact, before modern identification, the only way you could prove who you were and if what you were saying was true, was your word.

Today, disappearing is not as simple as the old times. But lies just got longer legs. So long, that they can reach not only your hateful neighbor or the ears of your handsome Queen, but overseas, on a small island in the Pacific Ocean with an internet antenna, onto the screen of your Twitter follower "Leila31XoXo".

In the following chapters we will explore what "lies" beneath fake profiles, generated profile pictures and their impact on modern society.

2 ALL YOUR DATA ARE BELONG TO US

When the internet started to become mainstream, most people didn't really understand how vulnerable their online presence was, and the same was true the other way around. Webmasters and people who owned the very first servers and windows to the internet, didn't really understand how impactful their user data could be.

The first restaurant to accept online orders was none other than Pizza Hut, in 1994, and it was one of the first 10.000 websites to hit the World Wide Web, and it used an online form to gather their customer's email and phone number information to deliver pizza. Very simple and effective.

Thirty years later, you can create an account, look at a map of all the Pizza Huts in the world, look for a job, download the app, sign up for a fantastic pizza prize. And ah, yes, also order a pizza. In a relatively short amount of time, the number of additional services that website offer skyrocketed, and so has the means of tracking users.

This data is not only used to help users serve a hot, steamy pizza to their doorstep, but also to understand which kind of pizza they want the most, at specific times of the year. This data is used to track and predict where the next Pizza Hut store should be, to maximize the number of pizzas to be delivered.

Some people however, began to be highly concerned with the fact that their presence online was being used not only to satiate their appetite, but also for other means. Websites like *Google* and *Yahoo!* saw the potential of using user's searches through the web to show them advertisements. So suddenly, after you visit *PizzaHut.com* for your midnight cravings, the next time you visit *Bloomberg.com* a small box pops out, with *Domino's* menu.

People grew concerned and, as it always was, wanted to regain their freedom of choosing what their information was being used for.

3 THE RISE OF FAKE IDENTITIES

Her name is Josie O. Campbell. She lives in Tigard, Oregon, is 45 years old and has been working as a veterinary assistant and laboratory animal caretaker for 10 years. Her car is a 1993 Mazda Lantis, and her favorite color is purple.

The previous paragraph is a description of a fake person, generated by the website FakeNameGenerator.com. It is a rather simple, yet effective free tool that creates a fictional person given a set of basic parameters, such as the gender, name set (American in

this case, but it could have been Arabic, Hispanic or even Chinese), and country. In the past years, the need for protecting one's privacy online went up dramatically, aided by the fact that people are increasingly more conscious about their online persona, and the digital trail they might leave behind.

Internet users are using these sorts of fake accounts to sort of "anonymize" their presence online, by just hitting the button "generate" every time they visit a new website that they don't really care about, or entirely trust.

This situation has caused a lot of trouble amongst social media websites and online forums, as any suspicious or malicious activity is a lot more difficult to tackle with fake information. Because of this, platforms have become more and more active in the pursuit of the deletion of fake accounts. Facebook alone deleted 2.2 *billion* [13] of fake accounts just in the first quarter of 2019, not including the ones that didn't even get activated in the first place.

Most of us who have an account on social medias like Twitter or Instagram, have quite an experience with this kind of fake accounts. For example, you might have received a shady Instagram follow request from a young, prosperous lady with a shadier description, usually with the emoji "Adults only", some cucumbers and hearts. Or perhaps you might have heard that your coworker wants to see if his fiancé is loyal, and has created a fake account to go and take a shot to his significant other.

Indeed, there are legitimate uses for fake accounts. To preserve one's identity, to innocently (or not) spying on other people's lives, or to simply please your mother with an Instagram where you just post your best panorama pictures.

However, the problem comes with the fake, automatically created accounts. Their use varies widely, as we previously pointed out, with most of them being used for financial profit.

4 GRIND AND CLICK FARMS

When I was in high school in Italy, I used to play a free online game called "Dead Frontier", a survival MMORPG (Massive Multiplayer Online Role-Playing Games, for you not-nerds in the audience). The goal of the game is simple: shoot at zombies, loot their undead corpses for money or weapons, and level up your character to destroy stronger zombies, get better weapons, and slowly but surely get bad grades at school.

This process is called "grind", which basically means repeat the same thing over and over for hundreds of hours. It is a long process, not even so much fun if you ask me now. But indeed, people did it, and still do it today in more modern games. *Dead Frontier*, and other similar games that involve grinding, do offer a simple and expensive way out to get a better account: microtransactions.

Often abbreviated as MTX, microtransactions are a business model where users can purchase virtual goods with micropayments [4], and in modern games it's the most widely adopted way to get money out of user's parents credit cards.

In *Dead Frontier*, the maximum achievable level is 220, and according to some of my friends, it takes thousands of hours to get there, even with microtransaction-powered enhancements. Some people then, just go and buy a fully-leveled up account, for hundreds or even thousands of euros, which still won't give your wasted months back.

Ever since Facebook introduced the infamous "Thumbs Up" or "Like" button on his platform, people immediately saw an opportunity to make money, as it always is. The more likes, views and followers you have on your social media page, the more likely you are to gain trust amongst your users. However, this takes time, effort and a lot of money to pull off. It's a grind.

Exactly like in games, businesses don't always have the time to grow millions of people out of thin air just so they are old enough to hold a smartphone and click "Like" on their Facebook page, so they pay what are called "Click Farms" to mimic this process, in a shorter amount of time.

Click farms work by buying thousands of smartphones and hiring hundreds of people to do simple tasks, such as to click on their on websites, videos, or following social media accounts. Other than that, they have to do other tasks that don't necessarily have anything to do with their customers, such as googling for some keywords, or scrolling a webpage. This is in fact a crucial point in their work, as these actions prevent their activity to be flagged as potentially malicious or fraudulent. As of today, you can buy roughly ten thousands views on a post on Instagram with as little as 25 dollars. Less if you have a discount code.

As you can imagine, this is a huge problem for both advertisers and the advertising companies. If you are a business trying to sell something online, no matter how much you pay, you will never be able to convert fake likes and clicks into revenue. The data that these fake clicks generate is hugely misleading, and hurtful, as marketing decisions are made through analysis of user's behavior.

However, there are categories of people that base their business entirely on those numbers alone, like influencers or politicians. If you are an average internet user, then the number of likes and followers of an account matters a lot, regardless of what they actually say or do. It's the first impression that matters, and having millions of followers and tens of thousands of likes on content helps that.

5 FAKE NEWS

You might have subscribed to online news magazines from time to time, and read through their articles. They might not look different at all from the ones that we see in real life, except from the fact that they are written on pixels on screen instead of ink on paper. Even the titles sound the same: "Godzilla Invaded New York!", "President Shaking Hands With Pope", "Winter Is Coming: Buy BrandName".

However, titles, *recommendations* that you see and the articles that you read, evolved along with the algorithms that show them to you. This became overtime one of the most debated arguments of how the internet works, and how companies make money. On YouTube for example, content creators have to keep track of every algorithm change, leading to the fact that everything they create, is not actually done to appease the users, but to appease the algorithm. If you don't show in people's recommendations, you don't exist.

Titles, thumbnails, video length, sponsors, advertisements. Everything has to be done in such a way that the algorithm favors your videos amongst a million other. Of course, this also applies everywhere a content creator posts their content, such as Google, Instagram, LinkedIn or Twitter. This is why it is more and more common to see titles such as: "HOW TO BECOME A BIRD IN 10 STEPS", "JOKE: BIG BALL HITS PEOPLE, PRANK!" and thumbnails

with bright, saturated pictures of people with their mouths open in awe, and some dollar bill symbols, because why not.

Of course, this is not entirely to blame on algorithms. We as users click on those articles and videos, basically telling the underlying system that yes, this system works, and we want to see that video and watch it to the end. This is how the internet uses algorithms to hack human nature.

This "hack" can in turn be used to manufacture articles in such a way that people are guaranteed to take a look at them, starting from the title. As an example, using "Pope Francis Meets Donald Trump" is much less effective in gaining clicks than "Pope Francis Shocks The World, Endorses Donald Trump for President". Although untrue, the latter is far more likely to be seen in social media, to get clicked on and to get shared amongst peers. In turn, this generates revenue for the article poster, via clicks on advertisement banners, which usually proliferate on their websites like shrooms in a forest.

You might have had the misfortune to find somebody in your life that shared some of these kind "clickbaity" of articles to you, or to their social media pages, like their *wall* on Facebook or as a *tweet* on Twitter. What I call misfortune is confronting them on the truthfulness of said article, and failing to making their posters to admit that they were wrong about believing that Elizabeth, Queen of England, is actually just a pawn of a secret group called the "World Order".

What happened there? Why did your friend start to believe in such stories, and why? Well, as it always has been, when enough people start to believe in something, that belief becomes reality. An untrue one, but an alternative reality nonetheless.

The same thing can be said for example about Galileo Galilei, the scientist who in 1609 defended his position of heliocentrism based on his astronomical observations in front of a jury of a very Christian jury. At that time, the occidental world lived within an alternative reality, where the earth is at the center of the universe, and the Earth was flat.

This happened because enough people believed that heliocentrism was a foolish idea, based on the Bible.

What if the same thing happened today? We are in an era when ideas and opinions can go far and wide, in the blink of an eye. News, fake or true, can be told by anyone anywhere, in any position of power, for whatever reason.

Election results can suddenly change with a few tweets, trends emerge with a few viral videos on TikTok, and vaccines and a global pandemic that kills hundreds of thousands of people suddenly becomes a hoax.

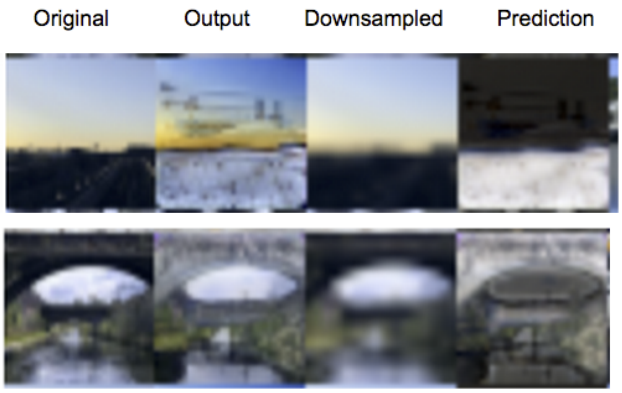

Figure 1: Skyline and bridges, generated by Eyescream (2015)

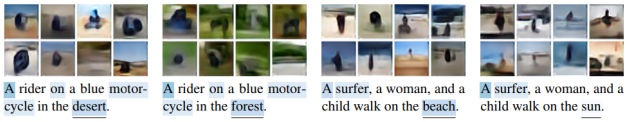

Figure 2: Example of most attended words while changing the background

Figure 3: Sequences of image edits performed using control discovered with GANSpace, applied to three different GANs.

6 BOTS AND AI

Around 2014, Generative Adversarial Networks, a type of neural network designed by Ian Goodfellow and his team [7], was introduced. The next year, Batch Normalization [9] and the definition of Residual Networks [8] was born. These, amongst many more discoveries up until today allowed neural networks to deeper, more efficiently, and with much better results.

One paper from 2015 [6] shows outstanding –for the time– results in terms of image generation using GANs. From a very low quality image, the network is able to generate a completely new image, either reconstructing the original, or coming up with its own interpretation of the available data (Figure 1). The same year, another team in Toronto developed a neural network that could generate images from a natural language text description (Figure 2). This was merely the start of the GAN revolution.

More recent studies allow GANs to be use parameters that can be tuned to generate specific variants of the same output image. An example is the image of the face of a person, that smiles or frowns, according to a given parameter, or turns its head left and right, or has glasses on or off, see Figure 3.

These advancements looked great, as they served to push the advancements in the field much forward. As the development in this

Figure 4: Josie O. Campbell: a fake person generated by an AI

field progressed, more discoveries were made, and higher quality images could be reproduced.

However, with more advanced and real-looking outputs, comes different problems. In Chapter 3 for example, we started by describing a woman in her forties using a generated output from a website. Of course, if you were browsing a Facebook page or an online forum with her profile, you would think twice about her being real. No pictures of her, and her connections are also without a profile picture. Suspicious.

With GANs however, we can give good ol' Josie Campbell a face, like the one in 4 generated using the output of a website called ThisPersonDoesNotExist.com. The site uses a GAN architecture called *StyleGAN2* [11], which greatly outperforms the previous architecture *StyleGAN* [10] in terms of quality and realness.

Now that Josie Campbell has a face of her own, even though it is fake, she suddenly became much more credible, and you wouldn't question right away that she isn't a real person.

Suddenly, we have created a person from thin air, and she can write and do whatever we please. She can be a woman that fights for human rights, or she can be an advocate for Donald Trump's latest fake news about the Democrats.

In Section 4 we talked about click farms, and the people who work in them. They are being paid incredibly little to operate up to two hundred phones each. Now, imagine the impact of pictures generated by GANs can have in the creation of new accounts. For every account you can have a totally different person, with a unique set of traits and pictures.

It would be incredibly hard for social media algorithms to filter and block such activities, as they look so real at a first glance, even for manual reviews.

If you throw fake news into the mix, you can possibly lead an army of online accounts that fight for your points of views and beliefs. You can possibly fool the popular opinion by generating thousands of people that will talk about your cause, whatever that might be.

What if you want to take a step forward, and generate a video that shows them doing some common activities, like walking or skiing? This can be done, with some limitations, with the use of "DeepFakes" [12, 14]. This is a recent technique where the face of a person in a video is replaced by another one, imitating their expressions and poses. Features that put this method under the spotlights, both for their incredible results, and their potential consequences.

Suddenly, you could make Obama the starring character of the movie "300", or make fun of a friend by pasting his face into some viral video. Or you could change the face of a porn video actress to someone else's face.

It started to become obvious that Deepfakes were not a tool for funny jokes and memes anymore, but they could become a way to bully and blackmail. In fact, this technique also applies to all kinds of objects, not only faces. In 2019, a group of researchers in Israel proved that you could also change the outcome of a CT scan [2], perhaps injecting a fake cancerous nodule into someone's lung's output image.

As the cherry on top, OpenAI, one of world's leading AI-focused companies, released in June 2020 a software called GPT-3 [5]. This AI model is able to do incredible things, from powering a text-based game called AI-Dungeon [1], to writing essays able to score a C [3]. The GPT-3 model is not easily available to the public, and for good reasons. Just imagine the level of detail and stories that this technology could create, from blog articles to populating a profile's *Facebook Wall* of realistic interactions.

7 FIGHT AGAINST THE MACHINES

Fake accounts, bots, face swaps. All these seem to be the beginning of a dystopian reality where common users like you and me can be outnumbered by a bunch of zeros and ones on someone's screen.

Who can you trust? Are your Facebook friends sharing true stories? Did that President really say something like that? Did your friend really meet Mark Zuckerberg?

What if someone creates tens of thousands of realistic profiles, and creates an alternate version of historic reality? Is this happening already?

Unfortunately, we don't know yet. It is fair to say that with the current technology, anyone is able with enough time and/or money to create the things that I have described above. Heck, they exist already.

REFERENCES

[1] [n.d.]. AI Dungeon - Wikipedia. https://en.wikipedia.org/wiki/AI_Dungeon#Dragon_model_release_(GPT-3) (Accessed on 03/20/2021).

[2] [n.d.]. Hospital viruses: Fake cancerous nodes in CT scans, created by malware, trick radiologists - The Washington Post. https://www.washingtonpost.com/technology/2019/04/03/hospital-viruses-fake-cancerous-nodes-ct-scans-created-by-malware-trick-radiologists (Accessed on 03/20/2021).

[3] [n.d.]. What Grades Can AI Get in College? https://www.eduref.net/features/what-grades-can-ai-get-in-college/ (Accessed on 03/20/2021).

[4] 2021. Microtransaction. https://en.wikipedia.org/w/index.php?title=Microtransaction&oldid=1008847640 Page Version ID: 1008847640.

[5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:cs.CL/2005.14165

[6] Emily Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus. 2015. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. arXiv:cs.CV/1506.05751

[7] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. arXiv:stat.ML/1406.2661

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. arXiv:cs.CV/1512.03385
[9] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv:cs.LG/1502.03167
[10] Tero Karras, Samuli Laine, and Timo Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv:cs.NE/1812.04948
[11] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. arXiv:cs.CV/1912.04958
[12] Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. 2017. Synthesizing Obama. *ACM Transactions on Graphics (TOG)* 36 (2017), 1 – 13.
[13] H. Tankovska. 2021. Facebook fake account deletion per quarter 2020. https://www.statista.com/statistics/1013474/facebook-fake-account-removal-quarter/
[14] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2020. Face2Face: Real-time Face Capture and Reenactment of RGB Videos. arXiv:cs.CV/2007.14808