Zenarosa, Jason

Prof. Korey Bell

Writing 60

4 February 2023

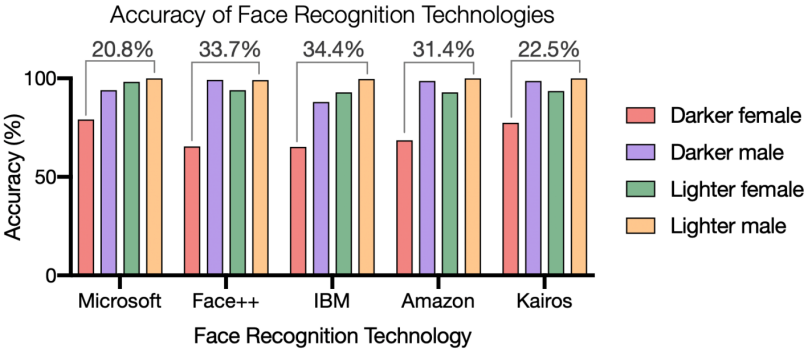
The Unforeseen Consequences of Emerging Artificial Intelligence Technologies

Artificial intelligence (AI) is everywhere these days. It has a wide variety of applications, from self-driving cars to election fraud detection to virtual personal assistants. This bleeding edge technology is constantly changing and evolving, all while being implemented in almost every industry to improve efficiency and decision-making. The potential for AI to revolutionize how we live and work is enormous, and it's exciting to see how it will continue to shape our world in the years to come. However, as with any emerging technology, artificial intelligence comes with its own set of unforeseen consequences. These can range from ethical concerns to practical issues, and it's important to consider them as we continue to develop and implement AI. The purpose of this essay is to argue that the lack of regulation for AI technology perpetuates systemic biases against marginalized groups, allows disinformation to disseminate on a large scale, and makes possible new kinds of impersonation and identity theft.

***

One consequence of the lack of regulation over AI systems is how they can perpetuate and even amplify societal biases if they are not properly designed and tested. Biases can be introduced into AI systems in a number of ways, such as through the data sets used to train them or the algorithms used to make decisions. For example, if an AI system is trained on a data set that contains a disproportionate number of images of one particular race or gender, it may be more likely to make mistakes when recognizing people from other races or genders. This is

exemplified in the figure shown to the left, which was sourced from the Gender Shades Project,



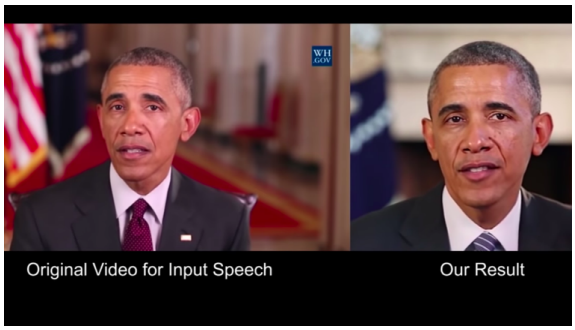
the goal of which is to evaluate "the accuracy of AI powered gender classification products" (Buolamwini, Gebru). The bar graph reveals how facial recognition technologies have lower accuracy on certain groups, especially women and people of color.

Nani Jansen Reventlow, a human rights lawyer and founding director of Digital Freedom Fund, discusses how the use of AI by different public bodies disproportionately impacts marginalized groups, in her article, "How Artificial Intelligence Impacts Marginalised Groups." According to Reventlow, one area in which this is particularly evident is "punishment and policing" (Reventlow). The perpetuation of racial bias and discrimination remains a significant problem in the U.S criminal justice system, and the deployment of facial recognition and predictive policing technologies by law enforcement only exacerbates this issue, particularly for marginalized communities. These communities, especially those with darker skin tones, are disproportionately likely to be misidentified and targeted by law enforcement due to AI technology. According to Alex Najibi, a bioengineering PhD candidate from Harvard, this discrepancy in AI accuracy is caused by "Black people [being] overrepresented in mugshot data, which face recognition uses to make predictions" as a result of "Black Americans [being] more likely to be arrested and incarcerated for minor crimes than White Americans" (Najibi). This leads to a higher likelihood of over-policing in marginalized communities and a perpetuation of systemic discrimination.

Another area where Reventlow claims that AI has a significant impact on marginalized groups is in the context of "movement and border control" (Reventlow). For example, in the United States, the use of biometric data such as fingerprints, retinal scans, gait recognition, and facial recognition technology has been used to arbitrarily deny individuals entry or even be used as justification for their forced removal (Reventlow). In other words, the lives of immigrants, a majority of whom happen to be people of color, are being subject to indiscriminate judgment from AI systems which have been displayed to have lower accuracy on people of color. Many refugees have even resorted to "burning off their own fingerprints" to avoid being tracked and targeted by AI systems used by border control (Reventlow). These systems are inhumane and an example of how the lack of regulation with AI technology can lead to discrimination against marginalized groups.

***

Another one of the consequences of the progression of AI technology, particularly deepfake technology, is its potential to be utilized for spreading false information and impersonating individuals. Sunny Valley Networks, a network security company based in Silicon Valley, defines deepfaking as the use of AI "...to fabricate information by replacing or synthesizing faces, audio, and emotions" (Sunny Valley Networks). Deepfake technology can be used to cast a person's face onto someone else or perfectly mimic their voice, effectively making it a highly advanced form of digital identity theft.

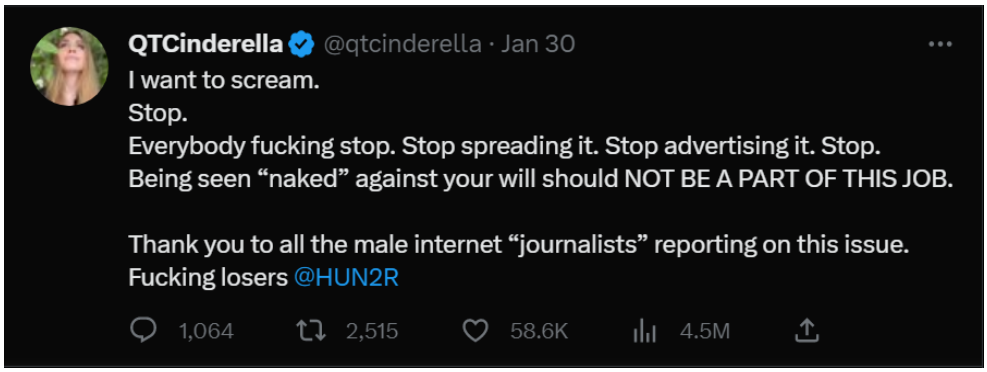


This image is taken from a YouTube video titled "Synthesizing Obama: Learning Lip Sync from Audio" by three AI researchers from the University of Washington: Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. The researchers used dozens of videos from Barack Obama's press releases to train an AI that generates convincing videos of Obama speaking.

The use of deepfake technology has the potential to cause mass disinformation. For instance, watch this video that American comedian/actor/filmmaker Jordan Peele made using a deepfake of former President Barack Obama. Although this video quite obviously and intentionally reveals that it is a fake, it exemplifies the way this technology can be used to impersonate politicians to "make it look like [they are] saying anything… Even if they would never say those things" (Peele). Additionally, both the video by Jordan Peele and the video by the University of Washington researchers were produced over four years ago, and deepfake technology has only improved since then. As deepfaking technology becomes more sophisticated, it will become increasingly difficult to determine what is real and what is fake. The ability to generate convincing false content using deepfake technology raises concerns about the authenticity of the information that we see and hear online, and the potential for it to be used for malicious purposes, such as spreading propaganda or manipulating public opinion.

The use of deepfake technology to produce pornographic material without an individual's permission is a concerning issue as it invades an individual's privacy and security. The victims of this type of deepfake pornography may suffer severe harm, such as damage to their reputation and emotional distress. For instance, Brandon "Atrioc" Ewing, a livestreamer on the platform Twitch, recently came under fire for purchasing deepfake pornography of several other female streamers. Among those impersonated was "QTCinderella" (real name unknown), a streamer

with over 800,000 followers on Twitch. QTCinderella has been very vocal on social media about how being portrayed in deepfake pornography has affected her mental wellbeing ever since the incident has come to light.



– *QTCinderella (via Twitter)*

As you can see, the creation and dissemination of deepfake pornography without consent is a blatant violation of the individual's privacy and autonomy. Not only is it a violation of privacy, but it also strips the individual of their agency and control over their own image. This is particularly damaging given the potential for deepfake porn to be spread rapidly and widely on the internet, leaving a permanent and ineradicable mark on the victim's online presence. This not only perpetuates harm but also deters people from expressing themselves freely online. In many cases, the victims are left feeling violated and helpless. The psychological toll on the victim can be severe and long-lasting; it can cause body dysmorphia (QTCinderella), which in turn can lead

to trauma and depression. The non-consensual creation and dissemination of deepfake porn is a gross invasion of privacy and autonomy, and it has serious consequences for the victim's well-being.

***

It is vital to be aware of how AI can reinforce and carry on present biases and discrimination and take action to guarantee that AI is utilized in a way that values and enhances the rights and respect of all people. To prevent these issues from arising, it is crucial to create regulations that ensure that AI systems are designed and tested with diversity and fairness as a priority. This can include utilizing diverse data sets, testing AI systems on a diverse group of users, and consistently monitoring and evaluating the performance of AI systems to identify and address any biases that may appear.

However, even with better quality data sets, it could be argued that certain use cases for AI, such as the ones exemplified above, are simply morally wrong. In the case of policing and border-control AI systems, no amount of ethical data sourcing will solve the core problem: these systems are, by design, meant to subject marginalized communities to subhuman conditions and treatment. In addition, under no circumstances should people have to witness themselves turned into pornography against their will. In cases like these, the use of AI should instead be heavily restricted.

There’s no doubt that AI technology will continue to be implemented in our society, as it has a wide range of positive and beneficial applications. However, it is imperative that the ethical implications of AI be taken into consideration. These issues must be addressed through regulations and restrictions that recognize AI’s potential for harm, in order to prevent AI from being used to discriminate, spread misinformation, or impersonate people against their will.

Works Cited

Buolamwini, Joy, and Timnit Gebru. *Gender Shades*, MIT Media Lab, 2018, http://gendershades.org/overview.html.

Ivascu, Mihai, and Alin Iftemi. “Using the Power of Blockchain to Combat Deepfake Videos.” *OpenDemocracy*, OpenDemocracy, 27 Nov. 2019, https://www.opendemocracy.net/en/democraciaabierta/c%C3%B3mo-usar-el-poder-de-blockchain-para-combatir-videos-deepfake-en/.

Mooney, Georgia. “'I'm Going to Sue You': QTCinerella and ATRIOC Deepfake Scandal Explained.” *The Tab*, The Tab, 2 Feb. 2023, https://thetab.com/uk/2023/02/02/qtcinderella-deepfake-ai-video-drama-explained-atrioc-pokimane-sweet-anita-292909.

Najibi, Alex. “Racial Discrimination in Face Recognition Technology.” *Special Edition: Science Policy and Social Justice*, Harvard University, 24 Oct. 2020, https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/.

Peele, Jordan. “You Won't Believe What Obama Says in This Video!” *YouTube*, BuzzFeedVideo, 17 Apr. 2018, https://www.youtube.com/watch?v=cQ54GDm1eL0.

QTCinderella. “I Want to Scream. Stop.everybody Fucking Stop. Stop Spreading It. Stop Advertising It. Stop.being Seen ‘Naked’ against Your Will Should Not Be a Part of This Job. Thank You to All the Male Internet ‘Journalists’ Reporting on This Issue. Fucking Losers @HUN2R.” *Twitter*, 30 Jan. 2023, https://twitter.com/qtcinderella/status/1620142227250094080?cxt=HHwWgIDRzbKx8_ssAAAA.

QTCinderella. “The Amount of Body Dysmorphia I've Experienced since Seeing Those Photos Has Ruined Me. It's Not as Simple as ‘Just’ Being Violated. It's so Much More than That.” *Twitter*, 31 Jan. 2023, https://twitter.com/qtcinderella/status/1620264657926885380?cxt=HHwWiMDQneaHq_wsAAAA.

Reventlow, Nani Jansen. “How Artificial Intelligence Impacts Marginalised Groups.” *Digital Freedom Fund*, Digital Freedom Fund, 29 May 2021, https://digitalfreedomfund.org/how-artificial-intelligence-impacts-marginalised-groups/.

Suwajanakorn, Supasorn, et al. “Synthesizing Obama: Learning Lip Sync from Audio.” *University of Washington*, 7 Sept. 2017, http://grail.cs.washington.edu/projects/AudioToObama/.

Suwajanakorn, Supasorn, et al. “Synthesizing Obama: Learning Lip Sync from Audio.” *YouTube*, University of Washington, 11 July 2017, https://www.youtube.com/watch?v=9Yq67CjDqvw.

“What Is Deepfake?” *Sunny Valley Networks*,

https://www.sunnyvalley.io/docs/network-security-tutorials/what-is-deepfake.