Digital Society Final Paper
TikTok

How TikTok Helps and Marginalises The Queer Community

Bianca Caissotti di Chiusano (i6245461)
28-10-2022

"Computers are Binary, People are not: How AI systems undermine LGBTQ Identity" is the title of an article published on "Access Now", a non-profit organization fighting for digital human rights (Leufer, 2021). This paper will analyze how this concept applies to TikTok and its algorithm, through the eye lens of Data Colonialism and the Ethics of AI.

TikTok, also known as Douyin, is owned by ByteDance and was launched first in 2016 in the Chinese market and then in 2017 worldwide. It gained popularity fast, surpassing two billion users in October 2020. The platform is not only a short-form video hosting service but also a place that revolutionised the music and fashion industries through the use of trends and by being an overall place for self-expression. It is important to note that Douyin and TikTok are in a way two different platforms. ByteDance prefaced that "TikTok is not a service offered in China" (Ryan et al., 2020).

For the purpose of this paper, it should be taken into account that the collected user data is not physically processed in China. If a user is located within the European Economic Area (EEA) or Switzerland they are contracting with TikTok Technology Limited which resides in Dublin, Ireland (TikTok Terms of Use, 2020).

Queer people have always sought a feeling of community, especially on online platforms. TikTok may feel like a safer space to express oneself as it does not necessarily connect you to people that you know in real life. The platform declares itself as a medium that "helps people find community through shared interests, giving users a canvas to express their creativity and discover the world around them". Please remember that the word *community,* along with the word connection, is used extensively within their Terms of Use, Privacy Policy and Community Guidelines, as this Paper is going to come back to this concept throughout its course.

Within the Legal Terms, TikTok attempts to provide safeguards against any type of discrimination. For example they "do not permit content that contains hate speech or involves hateful behaviour… content that attacks, threatens, incites violence against or otherwise dehumanizes an individual or a group on the basis of …. sexual orientation, sex, gender and gender identity". How TikTok monitors content and looks out for this behaviour is also disclosed under the headings: "What Information we collect" and "How We Use Your Infomation". Other than collecting profile information, they also store user content (even if it is not uploaded to the platform), audio recordings, comments, direct

messages, and all interactions. Moreover, they automatically collect any technical information, usage information and more. The mechanism behind this data collection is considered a new form of colonialism called Data Colonialism, where the major colonial powers are the Platform itself and their collaborating third parties, that appropriate data primarily for profitable exploitation (Couldry et al, 2019). In the case of TikTok, a user with an account enters a contract with the platform as soon as the Terms Of Use are accepted. However, what is unknown by a user is whether these terms apply even if they are not accepted or if the user doesn't have an account but is using the Platform. With Data Colonialism comes a spread of ideologies, which reframe this type of monitoring through a false narrative of how and why value is extracted, into more acceptable terms. This is very evident within the Privacy Policy and Terms of Use of TikTok.

The first ideology that this paper is going to tackle, is referred to as the marketing ideology of personalisation. Each user's For You Page (FYP) is unique to them and personalised based on their interactions within the platform. To make this possible, TikTok discloses that they "infer your characteristics (such as age and gender)" based on the data that they collect, to "provide you with personalised advertisements based on your interests". Moreover, "we will collect information from your device's phone book or social media contacts" to "help you connect on the platform". These are all examples of "language of personal freedom melded with the logic of surveillance", which spread the idea that personally targeted content can only come from surveillance, and its justification makes such tracking seem almost attractive and inevitable (Couldry et al, 2019).

In the book Weapons of Math Destruction (WMD), author Cathy O'Neil introduces the concept of a model to explain how systems (such as recommender systems) work. A model is described as "an abstract representation of some process… the model takes what we know and uses it to predict responses in various situations"(O'Neil, 2016). During this chapter, O'Neil makes an intuitive example of how a model can simply be the mechanism of what she chooses to cook her family for dinner. This model takes as input information she has about her family, ingredients, preferences, allergies and more, still taking into consideration the possibilities of unavoidable uncertainties. Lastly, as O'Neil stresses, the "key component of every model, formal or informal, is its definition of success" (O'Neil, 2016). To connect this example model to TikTok, the dinner would be Personalised content, the family would be the users, and the definition of success is the amount of engagement with the provided content. Personalised content is often content with which the user can self-identify, and what keeps them scrolling through their feed. And as the cycle continues, the more time spent on the app, the more data can be collected. These measurable features will construct an individual's digital identity, or "algorithmic identity" as described by John Cheney-Lippold, author of "We Are Data" (Simpson, E. et al, 2022). On one side, targeted content tends to prevent online conflict as it connects a user to a like-minded community. On the other hand, however, it negatively affects the social utility of the platform (Milano et al, 2020).

A significant problem with the marketing ideology of personalisation is that it tends to create a “filter bubble”. This, not only limits a user’s experience but will also reinforce personal biases and contribute to the further polarization of society (Milano et al, 2020). Consider the following case, of a user identifying as a member of the LGBTQ+ community being recommended queer content on TikTok based on the information that the platform has collected about them. It is fair to assume, then, that the same process applies to a user that is homophobic or transphobic. The latter was actually tested and written about in an article by the nonprofit organisation “Media Matters”: “we reviewed and tracked which videos TikTok recommended after we liked an anti-LGBTQ post that was fed to our For You page … TikTok almost instantly began recommending more” (Little, 2021). As Safiya Noble, author of Algorithms of Oppression, states in the introduction of her book: “Oppression operates in the same formats, runs the same scripts over and over. It is tweaked to be context-specific, but its all the same source code”(Noble, 2018).

This challenge is discussed under the “Transparency” section on TikTok.com, which reads that “by optimizing for personalization and relevance, there is a risk of presenting an increasingly homogenous stream of videos”, and that their goal is “to find a balance between suggesting content that’s relevant to you, while also helping you find content and creators that encourage you to explore” (TikTok, 2020).

When a user opens the TikTok app for the first time, they will be initially asked to select a range of current interests. A possible alternative, as suggested by Paraschakis, would be to use a user-based approach to the configuration of recommendations (Paraschakis,2018). It is fair to argue, however, that the latter is not done partly because of Data Colonialism, along with the myth that these “infrastructures of connection and data extraction are inevitable” if users want to be fulfilled, find community and see personalised content (Couldry et al, 2019). For example, TikTok’s “Transparency” section reads “TikTok helps people find community through shared interests”. This stems from another ideology that came along with Data Colonialism, referred to as the Ideology of Connection. See this ideology, as having several branches. By community, is meant groups, of users, that not only share interests but also, for example, should be protected. A lot of the justifications for data collection also include safety and protection. This paper interprets the latter as part of the ideology of connection. For example, TikTok collects user information within direct messages to “block spam, detect crime and protect our users”. Other examples include; “We monitor user content and other information to protect the safety and well-being of our community”, and lastly “We partner with our Affiliated and use a combination of technology (including automated means), human monitoring, and reports from our users to identify and enforce violations of our Community Guidelines and Terms to protect you and all of our users.”

While the truth is that by connection, it is not meant the basic human value as its known, but more as a connection to an unequal infrastructure, to which users will be submissive to a more powerful platform through terms and conditions, like the ones presented throughout this paper

(Couldry et al, 2019). This in itself is a digital embodiment of "The Stange Case of Dr Jekyll and Mr Hyde". Where Dr Jekyll represents the ideologies and myths that make Data Colonialism possible, while Mr Hyde, lurking in the background, is data appropriation and compliance to an opaque model.

Sometimes, it may be difficult to distinguish between Dr Jekyll and Mr Hyde, identifying what is good and what is bad in the way an algorithm works and is justified. Furthermore, it can be argued that this is a challenge because it is usually subjective. What is considered just for someone, may not be for someone else. This also varies depending, on culture, political stance etc.

An important factor to remember is that human decision-making is biased, thus autonomous decision-making, which may appear neutral, also contains the same biases, as they are created by human beings. This often leads to negative effects such as misrepresentation, behavioural manipulation and overall discrimination and oppression. It is crucial to address these issues and continuously improve the design process of algorithms within platforms.

Algorithmically driven decision-making tends to have missing social and human context, and as Safiya Noble states, this is of particular concern to marginalised groups who are problematically misrepresented in media, such as the LGBTQ+ community. Needless to say, a closer inspection of what values are prioritized in automated decision-making systems is needed (Noble, 2018).

Again in the book Weapons of Math Destruction, Cathy O'Neil writes about the use of a predictive program called PredPol in police departments around the U.S. The model seeks to optimize resources and is said to be blind to race and ethnicity. Throughout the chapter, O'Neil makes an interesting observation; "most of us, including the police, view crime as a pyramid", where at the top there is homicide and at the bottom, there are shoplifting and parking violations which occur more frequently. She then proceeds stating "Prioritizing the crimes at the top of the pyramid makes sense. Minimizing violent crime, most would agree, is and should be a central part of a police force's mission"(O'Neil, 2016). In the case of TikTok, and for the purpose of this paper, think of crimes as violations of the Community Guidelines. In this pyramid, their primary concerns are listed in their "table of contents", which starts with Minor Safety, and continues to include Dangerous acts, Suicide, Adult Nudity, Bullying, Hateful Behaviour, and more. As can be observed, there is no specific section in the Community Guidelines tackling LGBTQ+ discrimination specifically, and this paper does not criticise that, as these discriminations are still being included in the guidelines, just in a more broad or general way. This observation is made to point out that, content which goes against their "prioritised" guidelines will be taken down more often, maybe even processed with human monitoring. As O'Neil also states: "The privileged are processed more by people, masses by machines". Not everything can be taken into account within TikTok's guidelines or choices. Models are just simplifications, that do not include all of human complexity (O'Neil, 2016). There are choices to be made regarding what should be included, for example in TikTok's Community Guidelines, but also when to have blind spots. As O'Neil explains, blind spots in a model "reflect the judgments and priorities of its creators … Models are opinions embedded in mathematics" (O'Neil, 2016). Hence, the Community

Guidelines and the primary concerns of the platform, including how points are written, what information is given and most importantly, how much transparency there is, do reflect upon the ideologies of the creators of TikTok.

Transparency is an ethical principle in automated decision-making within a digital society. It is essentially what connects the platform and its algorithms with its users. Moreover, it is one of the guiding principles to improve the design process of a platform such as TikTok, along with Nonmeleficence, Fairness, Accountability and Privacy. As previously discussed, treating the topic of LGBTQ+ discrimination is not one of the priorities of TikTok's Community Guidelines, and this leaves several loopholes. Consider the following examples that acknowledge why being transparent about certain choices of implementation is important.

The first example is about Fairness when handling Minor Safety and Adult Nudity. The Community Guidelines under the Minor Safety section include "the promotion of cosmetic surgery to minors may lead to higher rates of body dissatisfaction … including before-and-after videos, videos of surgical procedures" (Community Guidelines, 2022). What happens to content dealing with gender-affirming surgery, such as what is known as "top surgery"? What if the user sharing their experience on the platform, perhaps to even educate other users, is a minor? Should all users under the age of 18 be protected from accessing such information? On the other hand, is it fair for users of all ages to be treated equally, and exposed to such content even at the age of 13?

Another point in the Community Guidelines regarding Minor Safety tackles depictions of self-harm, including disordered eating. It reads "we also remove content that depicts attempted suicide or behaviour likely to lead to self-inflicted death" (Community Guidelines, 2022). Research from WHO claims, however, that a number of vulnerable groups, such as members that identify in the LGBTQ+ community, have been identified as having a higher risk of suicide (Preventing Suicide: A Global Imperative, 2014). Multiple research, along with that also affirms that unfortunately, self-harm (and eating disorders) are very prevalent within the queer community. Thus a lot of content created by queer users may be taken down because it depicts "attempted suicide". Again, is this ethical? Should there be extended use of, perhaps, trigger warnings? Or is content taken down for other reasons, and not because of a depiction of self-harm?

Still on the topic of fairness, in 2020, reports showed that TikTok had been censoring LGBTQ+ content by, for example, shadow-banning hashtags regarding the community in several countries such as Russia, Bosnia and Jordan (Ellis, 2021). Further reports indicate that queer content was suppressed in at least 8 languages, meaning that this censorship not only affected users within the particular countries but also any other users around the world that speak those languages (Ellis, 2021). This banning is due to some countries' local laws, and because of the pornographic content that would be shown when searching for LGBTQ+-related content. Thus, to what extent should local laws be respected? If shadow-banning is the consequence of respecting these laws, then are all users being treated the same?

Looking at the guiding principle of “Nonmaleficence”, a platform should not cause any harm, unequal opportunities or emotional distress. Among other guidelines, TikTok will remove any “expressions of abuse, including threats or degrading statements intended to mock, humiliate, embarrass, intimidate, or hurt an individual”. Again, however, there are some loopholes as in those same guidelines they state that “to enable expression about matters of public interest, critical comments of public figures may be allowed; however, serious abusive behaviour against public figures is prohibited”. There is a lack of transparency regarding what the platform considers critical comments or abusive behaviour, this leaves space for inflicting harm.

All of the unanswered questions come from a lack of transparency. TikTok should provide greater transparency on how and why decisions are made and what happens when content is flagged. Especially when the users are issuing reports. Moreover, the community guidelines should be more precise and accessible (including on the app) to the users. Lastly, there should be more human moderators, that are qualified to distinguish and interpret whether the use of LGBTQ+ terms and content is legitimate or comes from trolling (Ellis, 2021). For this to be possible, the company should continue to diversify hiring and gather further resources regarding the current state of LGBTQ+ self-expression. After all, as O’Neil mentions in Weapons of Mass Destruction, models are always changing, like people, and the updates or adjustments are what make it a “dynamic model”.

Overall, TikTok helped create a safer space where users from marginalised groups, such as members of the LGBTQ+ community can connect with each other and discover themselves. On the other hand, however, the homogenous stream of content created by their collected data enhances the creation of a filter bubble. This can create a false perspective of how accepting the outside world is, or even TikTok itself. In fact, as discussed, TikTok has a problematic history of policies and practices with regard to LGBTQ+ users and content. In conclusion, much more transparency is needed. “If you were told by an usher, upon entering an open-air concert, that you couldn’t sit in the first ten rows of the seats, you might find it unreasonable. But if it were explained to you that the first ten rows were being reserved for people in wheelchairs, then it might well make a difference”(O’Neil, 2016). The latter is a great example given by Cathy O’Neil that explains exactly why transparency matters. If it seems like such a simple concept, then why do most companies, including TikTok, hide the existence or results of their models? Most probably because of how much their algorithms are worth, and the success that they’ve already had. Even if TikTok were to be much more transparent its model would still remain unclear. This raises one last question: who do they value more, their user or their platform?

Word Count: 3077

References (APA 7):

Community Guidelines. (2022, February). TikTok. https://www.tiktok.com/community-guidelines?lang=en#38

Couldry, N., Mejias U.A. (2019) The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism. Stanford University Press. (Chapter 1: The Capitalization of Life without Limit)

Ellis, S. (2021). Social Media Safety Index. GLAAD. https://www.glaad.org/sites/default/files/images/2021-05/GLAAD%20SOCIAL%20MEDIA%20SAFETY%20INDEX_0.pdf#page=23

Little, O. (2021, May 18). *TikTok's recommendation algorithm is promoting homophobia and anti-trans violence | Media Matters for America*. Media Matters for America. https://www.mediamatters.org/tiktok/tiktoks-recommendation-algorithm-promoting-homophobia-and-anti-trans-violence

Leufer, D. (2021, April 30). How AI systems undermine LGBTQ identity. Access Now. Retrieved October 18, 2022, from https://www.accessnow.org/how-ai-systems-undermine-lgbtq-identity/

Milano, S., Taddeo, M. & Floridi, L. Recommender systems and their ethical challenges. *AI & Soc* 35, 957–967 (2020). https://doi.org/10.1007/s00146-020-00950-y

Noble, S. (2018). Algorithms of Oppression. NYU Press. (Introduction: The Power of Algorithms) https://safiyaunoble.com/wp-content/uploads/2020/09/Algorithms_Oppression_Introduction_Intro.pdf

O'Neil, C. (2016). Weapons Of Math Destruction. Penguin Random House. (Chapter 1: Bomb Parts: What is a Model?)

O'Neil, C. (2016). Weapons Of Math Destruction. Penguin Random House. (Chapter 5: Civilian Casualties: Justice in the Age of Big Data)

Paraschakis D (2018) Algorithmic and ethical aspects of recommender systems in e-commerce. Malmö. http://muep.mau.se/bitstream/ handle/2043/24268/2043_24268%20Paraschakis.pdf?sequence=3&isAllowed=y

Preventing suicide: a global imperative. (2014). World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/131056/97892?sequence=1

Privacy Policy. (2021, October 5). TikTok. https://www.tiktok.com/legal/page/eea/privacy-policy/NL

Ryan, F., Fritz, A., & Impiombato, D. (2020). TikTok privacy concerns and data collection. In *TikTok and WeChat: Curating and controlling global information flows* (pp. 36–42). Australian Strategic Policy Institute. http://www.jstor.org/stable/resrep26120.7

Simpson, E., Hamann, A., & Semaan, B.C. (2022, January 14). How to Tame “Your” Algorithm: LGBTQ+ Users’ Domestication of TikTok: Proceedings of the ACM on Human-Computer Interaction,Volume 6, Issue GROUP, Article No.: 22pp 1–27 https://doi.org/10.1145/3492841

TikTok Terms of Use | TikTok. (2020). Tiktok.com. https://www.tiktok.com/legal/terms-of-service-eea?lang=nl

TikTok.(2020,June 18). How TikTok recommends videos #ForYou | TikTok Newsroom. Newsroom | TikTok; TikTok. https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you