POLICY BRIEF | CIVIL SOCIETY

YouTube's Violation of Palestinian Digital Rights: What Needs to be Done

By: Amal Nazzal · December, 2020

In recent years, an increasing number of Palestinians have reported having their right of expression suppressed by dominant social media platforms, such as Facebook, WhatsApp, Twitter, and YouTube. In 2019, Sada Social, a Palestinian digital rights organization, documented as many as 1,000 violations including the removal of public pages, accounts, posts, publications, and restriction of access. However, we know little about YouTube's policies and practices towards Palestinian digital content, whether these policies are in line with human rights law in general, and how they are applied to Palestinian content.[1]

This brief draws on research conducted by 7amleh – The Arab Center for the Advancement of Social Media, to examine YouTube's policy and community guidelines, its social media content moderation, and its violation of Palestinian digital rights through hyper-surveillance. The analysis builds on interviews conducted with Palestinian journalists, human rights defenders, activists, and international human rights organizations, such as WITNESS and Article 19. The brief discusses key findings from the interviews, including the vague and problematic definition of "violent content," which is used to restrict content posted by Palestinian YouTubers, as well as YouTube's violative practices, such as language and locative discrimination. Finally, it offers recommendations to YouTube, civil society organizations, policymakers, human rights activists and

defenders, as well as social media users, for protecting Palestinian digital rights.

"Your Video Violates Our Community Guidelines"

YouTube, the American video-sharing platform, was established in February 2005 and was bought by Google in November 2006 for $1.65 billion. In 2019, the company's revenue amounted to $136.819 billion, making it one of the companies with the largest revenues in the world. Indeed, around 500 hours of video are uploaded to YouTube every minute. In 2019, the number of YouTube channels grew by 27% to well over 37 million channels. As the popularity and influence of video content grows, the Google-owned company has become one of the most important channels for digital content distribution worldwide. In the Middle East, the YouTube user rate has increased by 160% from 2017-2019, with more than 200 YouTube channels in the region that include over one million subscribers. Despite YouTube's popularity, many have criticized its policies and practices regarding how it manages and moderates its content. Among other controversial policies and practices, YouTube has faced criticism over its algorithms which help popularize videos that promote conspiracy theories and falsehoods, as well as videos that ostensibly target children but also contain violent and/or sexually suggestive content.On a regional level, particularly in Syria, many YouTube violations of digital rights have been recorded and documented. Over several years during the Syrian uprising, Syrian journalist and photographer, Hadi al-Khatib, collected 1.5 million videos documenting human rights violations in the country, including footage of hundreds of chemical attacks carried out by the Syrian regime, evidence that would be critical to prosecute perpetrators of war crimes. However, al-Khatib reported that in 2018, over 200,000 videos, amounting to 10% of the videos he had uploaded, were taken down and disappeared from YouTube. Deleting videos during crisis and war time has serious implications, and can result in erasing crucial evidence in later trials.

Although YouTube's mission statement promises "to give everyone a voice and show them the world," this is not always the case with its ongoing and unjustified declaration that some videos violate its community guidelines. YouTube has four main guidelines and policies that are listed under its community guidelines: spam and deceptive practices, sensitive topics, violent or dangerous content, and regulated goods. However, many users have accused YouTube of removing videos without clearly falling afoul of any of these guidelines.

This raises significant problems when it comes to users' digital rights. Firstly, it indicates that YouTube is not held accountable regarding the clarity and equity of its four guidelines; it can maneuver between them interchangeably to justify content removal. Senior campaigner at Article 19, Barbara Dockalova, explained that any of YouTube's obscure policies could be used to remove content: "If your video hasn't been deleted under 'violent or graphic content,' it could be deleted under 'harmful or dangerous content.'"

Another member at Article 19, Gabrielle Guillemin, added that YouTube is vague in defining certain terms: "Terrorism, for example, is defined differently according to YouTube, as some of the human rights defenders whom we worked with were considered terrorists." Indeed, in September 2018, Article 19 recommended that YouTube align its definition of terrorism with the one proposed by the UN Special Rapporteur on counter-terrorism and human rights. Article 19 also confirmed that YouTube's policies and community guidelines fall below international legal standards on freedom of expression. In its 2018 statement, Article 19 urged that YouTube be transparent about how it applies its guidelines in practice by providing detailed examples, case studies, and thorough explanations of what it considers to be "violent," "offensive" and "abusive" content, including "hate speech" and "malicious" attacks.

Secondly, the interviews conducted in this study indicate that, while YouTube continues to indiscriminately remove content it deems unfit, it has a tendency to

keep, and even defend, highly problematic content. In June 2019, YouTube decided that several hateful homophobic messages against the American producer, Carlos Maza, did not violate its terms. Then, in August 2019, LGBTQ+ YouTube content creators in California filed a lawsuit against YouTube claiming that YouTube's algorithm discriminates against LGBTQ+ content, whereby machine learning moderation tools and human reviewers unfairly target channels that include words such as "gay," "bisexual," or "transgender" in the title.

In May 2020, YouTube Artificial Intelligence (AI) was accused of automatically deleting comments containing certain Chinese-language phrases considered critical of the country's ruling Communist Party (CCP) within 15 seconds of a YouTube user leaving a comment under videos or in livestreams. The CCP case demonstrated how YouTube designs automated decision-making (AI programming) to mass flag and remove certain content for political ends, regardless of the context.

One of YouTube's justifications for using AI programming is that it could be much more efficient in detecting inappropriate content, as massive amounts of videos are uploaded every minute, making it impossible to manually review. However, Dia Kayyali at WITNESS explains that the problem with using AI to review content is that, while "computers might be efficient in detecting violence, they are not as nuanced as humans. They are not good at figuring out if a video is ISIS propaganda or vital evidence of human rights violations." As a result, she continues, "YouTube's policy of content moderation has become a double-edged sword, where essential human rights content is getting caught in the net."

Promoting Israeli Content, Silencing Palestinians

Collaboration between Israeli security units and social media platforms such as Facebook, WhatsApp, and Twitter, is becoming increasingly well documented. Indeed, the office of the Attorney General of Israel has been illegally running a

"Cyber Unit" to censor Palestinian content and to monitor Palestinians' social media accounts. What is more, Israeli and Palestinian Authority (PA) security forces have arrested 800 Palestinians using AI programming because of their posts on social media, particularly on Facebook. However, we know little about YouTube's pro-Israel leanings. In a range of videos, researchers found that many YouTube videos about the Israeli army and military remain on YouTube, regardless of the explicit celebration of militarization and violence. This unfettered development of Israeli content which unequivocally promotes and celebrates the lethal use of force has become so normal that it is now marketable on YouTube. This, on a platform whose community standards restrict images of violence, and which even has a "Firearms Content Policy." Under this policy, YouTube states that: "Content that sells firearms, instructs viewers on how to make firearms, ammunition and certain accessories, or how to install those accessories is not allowed on YouTube." Despite this policy, YouTube promotes "Alpha Gun Models," a business run by Israeli gun model, Orin Julie. Julie has more than 440,000 followers on Instagram, and a famous YouTube channel titled, "I am the Queen of Guns," with more than 2,870 subscribers. On her Instagram account, she defines herself as an influencer and promoter of gun companies, a clear violation of YouTube's Firearms Content Policy. Nonetheless, YouTube continues to promote her content.Another incident of hate speech and harassment which reveals YouTube's double standard is the video promoted by Israeli clothing brand, Hoodies. In the video, Israeli model Bar Refaeli removes a niqab before sporting a range of different clothes. The Islamophobic video ends with the slogan "Freedom is basic." Refaeli shared the 30-second video on her Facebook page, an account with nearly three million followers, and despite activists' critiques, the video remains on YouTube. YouTube has been a particularly important platform for Palestinian human rights defenders and activists documenting Israeli violations and sharing them on YouTube in the hopes of raising public awareness and holding the Israeli regime accountable. However, several human rights defenders

had their content taken down, harming efforts to archive and document Israeli violations against children, people with disabilities, and Palestinians as a whole.

Interviews with Palestinian human rights defenders showed that YouTube removes their content under the pretext that it is "violent." According to Palestinian journalist Bilal Tamimi, YouTube violated his right to post a video showing Israeli soldiers abusing a twelve year-old boy in the village of Nabi Saleh. With sorrow, he said: "YouTube understands what violates its own terms and conditions, not what violates us as Palestinians." He continued: "Why do I need YouTube if I cannot report the violations which the Palestinians, and Palestinian kids in particular, are daily facing?" In the end, Tamimi embedded the deleted video into a longer video which passed YouTube's AI screening, a tactic used to circumvent the platform's content removals.

Similarly, the chief editor of Al-Quds News, Iyad Al-Refaie, explained in his interview: "Violence varies according to YouTube. For example, when YouTube sees a video of a Palestinian toddler killed by the Israeli army, it has issues with publishing it, but it is fine with promoting Israeli militarization and videos of Israeli kids trained to shoot guns." One of the videos which Al-Refai referred to shows Israeli children learning how to carry a weapon, along with other Israeli YouTube channels celebrating Israeli militarized violence.

Artificial Intelligence Designed to Discriminate

In interviews, Palestinian human rights defenders reported experiencing language and locative discrimination against their content on YouTube. That is, they reported that their Arabic-language videos, which included Arabic titles or subtitles, are under higher surveillance from YouTube in comparison to videos posted in other languages. In addition to language discrimination, Palestinian activist and journalist Muath Hamed explained that YouTube's surveillance AI machines are designed and operationalized for a higher level of scrutiny of

content emerging from the West Bank and Gaza – what is known as locative discrimination.Hamed, who is an active YouTuber, elaborated on the language discrimination he experienced with YouTube. His personal YouTube channel, "Palestine 27K," has 10,500 subscribers, with around 897 videos. He explained that many of his videos were flagged and deleted, and the only explanation was that they included Arabic-language content. As he put it: "One way or another, our videos that have Arabic titles or subtitles are under more observation from YouTube."He added that he was able to understand exactly how YouTube's lexicon discriminates against Arabic content, and consequently, how to evade detection:

I've become more experienced with YouTube policies and tactics of surveillance […] *Over time, I developed a YouTube lexicon* […] *For example, Hamas, Islamic Jihad, Hezbollah, etc. are keywords which will be highly flagged. […] So I started to blur some items within the video, such as political parties' flags.*

Hamed, who at the time was residing in the West Bank, revealed a shocking reality about YouTube's locative discrimination against Palestinians in a simple online experiment: "I sent the same video which was deleted from my YouTube account to my friend's YouTube account in Europe […] and YouTube was fine with the video published in a European country."Muath Hamed and Samer Nazzal, a Palestinian journalist and YouTuber based in Ramallah, argue that YouTube AI hyper-surveys Palestinian content, especially when it has a high number of views and starts becoming influential. They noticed that after their videos received a certain number of views (one about a Palestinian boy throwing stones at an Israeli soldier, and the other about Ahed Tamimi in Nabi Saleh demanding the Israeli soldiers release her brother from custody), YouTube started tracking their videos, even old ones, where they got more strikes, blocks, and even demonetization as punishment. In other words, as punishment, YouTube users lose the ability to earn advertising income on their content. It is unclear how YouTube AI moderates its

online content, how it hires its staff, or how it applies digital surveillance within specific contexts. Article 19 member Barbara Dockalova explains:

We know little about how the data is treated, how the algorithms are developed […] YouTube won't share such information […] We also know very little about YouTube staff. How are they trained? Do they know anything about the context? How does the revision process go?

To better understand how AI could be designed in various discriminatory ways, such as language and locative discrimination, it is worth explaining how AI could be programmed to interfere in the online moderation process. The figure below shows the workflow of AI to increase a company's effectiveness with online content moderation. The three ways, shown in green, demonstrate where AI algorithms could be designed and trained to perform in a discriminatory way. But how, exactly does this discrimination occur?

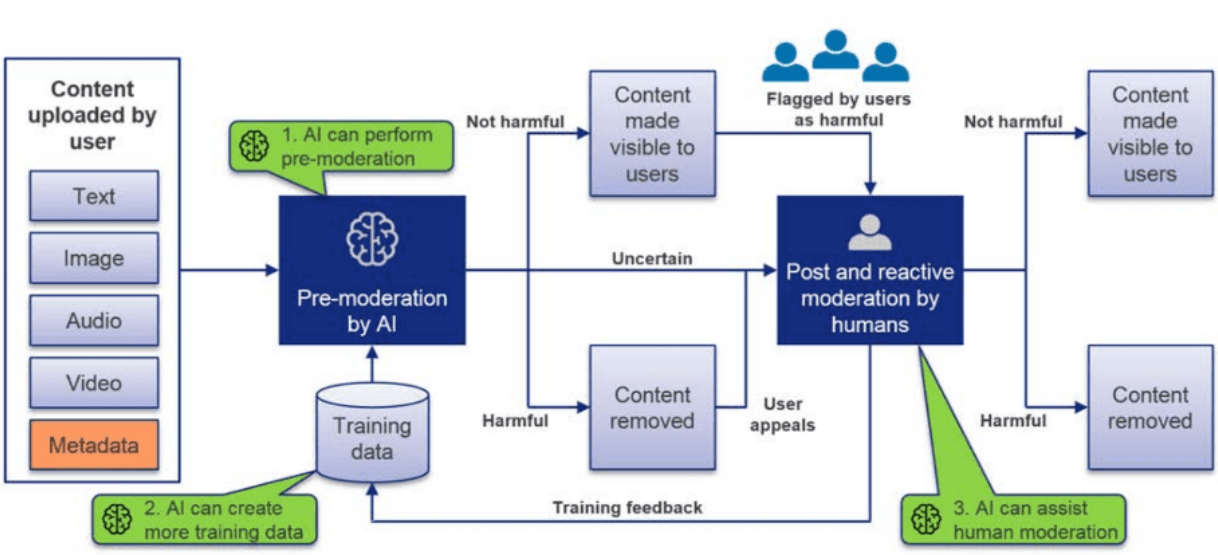

Figure 1 – There are three key ways in which AI can improve the effectiveness of the typical online content moderation workflow (**SOURCE:** Cambridge Consultants)

First, the workflow allows biased AI to evaluate user content, such as text, image, audio, or video, as harmful or not, which automatically produces a biased decision. This process is called "pre-moderation." Second, biased AI is programmed to automatically create more training data and assist human moderators, which

reproduces more biased data and discriminatory decisions. But how can AI algorithms be trained to perform discriminatorily? There are certain ways to train AI algorithms to do this. These include "hash matching," whereby a fingerprint of an image is compared with a database of known harmful images, or "keyword filtering," whereby words that indicate potentially harmful content are used to flag particular content. Which content is included in the database of content considered harmful could lead to clear cases of discrimination. For example, Israeli intelligence units use algorithms to censor tens of thousands of young Palestinians' Facebook accounts by flagging words such as *shaheed* (martyr), Zionist state, Al-Quds (Jerusalem), or Al-Aqsa, in addition to photos of Palestinians recently killed or jailed by Israel.

The activists, YouTubers, and journalists interviewed in this study expressed feeling discriminated against and excluded from a dominant digital power like YouTube, in addition to feeling angry, disappointed, and unmotivated to continue being active on YouTube. However, participants also proposed different tactics and techniques to evade and confront this discrimination, such as keeping back-up videos at all times, using alternative platforms such as Vimeo and Dailymotion, and trying to develop a YouTube lexicon to counter its biased policies.

Recommendations

YouTube's policies and community guidelines fall below international law and standards of freedom of expression, which YouTube should respect. Moreover, YouTube's definitions of key terms in its community guidelines, such as "violent content," are vague and problematic, and, in effect, violate users' digital rights. Indeed, YouTube's AI violates Palestinian digital rights by implementing algorithmic biases against the Arabic language and against content emerging from the West Bank and Gaza.To counter these discriminatory practices and protect Palestinian activists, journalists, and YouTube users, the following

recommendations should be implemented:

- YouTube community guidelines must respect human rights law and standards. YouTube should provide equal access to information and ensure that its community guidelines are available, with full clarifications, in official UN languages, including Arabic.
- A third party, civil society monitoring group, such as human and/or digital rights organizations, should ensure that AI is not hyper-surveying Palestinian content and discriminating against it. The third party must also serve as an adjudicating body, supporting users in appealing and protesting the removal of their content. YouTube users must have the right to challenge decisions when their content is removed.
- YouTube should publish transparency reports for deletions, blocks, or restrictions of content and profiles of Palestinian users; it must publish the number of requests to restrict content per actor, the number of requests approved, and the reasons for approval or rejection of requests; and it must clearly explain how a person can appeal the decision, as well as give a reasonable response timeline with contact details for more information.
- The Palestinian Authority must support Palestinians' lawsuits against social media platforms that violate their digital rights, and it should assign lawyers to assist them.
- Palestinian civil society organizations should raise awareness about digital rights through which they educate Palestinian users on their rights, on how to express themselves, and on how to resist these discriminatory practices.
- Activists, journalists, and human rights defenders engaged in online platforms should share strategies and tactics for evading language and locative discrimination, among other forms of violation. They should also work to develop technologies to reverse YouTube's biased AI.

1. *To read this piece in French, please click here. Al-Shabaka is grateful for the efforts by human rights advocates to translate its pieces, but is not responsible for any change in meaning.*

Al-Shabaka: The Palestinian Policy Network, is an independent, non-profit organization. Al-Shabaka convenes a multidisciplinary, global network of Palestinian analysts to produce critical policy analysis and collectively imagine a new policymaking paradigm for Palestine and Palestinians worldwide.

Al-Shabaka materials may be circulated with due attribution to Al-Shabaka: The Palestinian Policy Network. The opinion of individual members of Al-Shabaka's policy network do not necessarily reflect the views of the organization as a whole.