SnapShield:
Evaluating a Framework for Early Detection of Inappropriate Live Video Streaming

By Gabriel Ruiz, Massey University

19 October 2019

Introduction:

It was the 9th of October of this year when the sound of explosions disturbed the peaceful small Geman town of Halle. At the synagogue site of the explosion the Cantor, the person in charge of reading of the Torah, watched as a good Samaritan was gunned down by a passer-by outside their gated wall before hurling improvised explosives over the wall. (Fogel, 2019) But the cantor was not the only one watching. Somewhere else in the world 5 viewers watched the 35 minute live "event" on Twitch, a live streaming service used mostly by gamers to live-stream their video gaming exploits. The video was watched by 2,200 users in the time between when the video was automatically uploaded after the conclusion of the live stream and the time it got tagged and removed by the service. (Gonzales, 2019)

Earlier this year, in Christchurch, New Zeland, an unnamed Australian man used Facebook's live-streaming option to exhibit his massacre of 51 Muslim worshipers at the local mosque. Facebook has this to say about its live-streaming service: "Live is the best way to interact with viewers in real time. Field their burning questions, hear what's on their mind and check out their Live Reactions to gauge how your broadcast is going." (Facebook Live) According to this, the service is meant to bring broadcasters closer to their audience. It is a product meant to unite TV celebrities and fans, grandkids and grandparents, even churches and parishioners. It was not meant to be a medium to propagate hate and terror.

In a world of unintended side effects, what is the responsibility of real time content providers to assure the free-speech of the individual content creators, provide privacy for the content creators and their target audiences, and still provide a medium "that promotes actions that maximize happiness and well-being for the majority of a population"? (Lazari-Radek & Singer, 2017) In March of this year, the Australian prime minister called on social media platforms to do what is needed to assure that they "ensure their technology products are not exploited by murderous terrorists." (Tough new laws… 2019) What are the ethical bounds of electronic content aggregators and social media providers?

It is the purpose of this paper to first evaluate academic literature on the ethical responsibilities of technology solution creators who leverage user-generated content to communicate, entertain, and ultimately profit. Secondly, this paper proposes a technological and social method for assisting said creators in safeguarding their products from being hijacked with unintended uses. Finally, the paper gauges the possible ethical challenges offered by the proposed method in the light of being a signatory to the Code of Ethics of the organization IT Professionals New Zealand.

Literature Review:

Since the Facebook Live service started in 2015 there have been numerous instances of violence, including child abuse. (Warzel, 2019) In New Zealand, the Harmful Digital Communication Act of 2015 lays some responsibility on the shoulders of content providers. (Online Safety Laws and Rules) Additionally, a company can suffer the loss of advertising income when it is unable or reticent to protect its users, as was the case in New Zealand after the 2019 mosque massacre. (Peacock, 2019) But even if there were no governing laws or financial consequences, as is often the case, what guiding ethical principles can live content aggregators use to guide the development of software solutions and services?

In spite of its limitations, utilitarianism can be a good starting point when deciding how to craft a software feature with the potential of affecting a large group of people. (Utilitarianism, 2019) In a society with a strong belief in personal freedom and responsibility, it can be easy to let those personal ethical considerations overshadow the need for some greater good. In the case of live video streaming, it can be tempting to let the end-user decide if the content is appropriate or not to them. It is always difficult to balance the good of the few and the good of the many. In his 2019 commentary entitled “Beware Geeks Bearing Gifts: Assessing the Regulatory Response to the Christchurch Call,” Victoria University’s Peter A. Thomson says Facebook founder Mark Zuckerberg acknowledges “the tension between protecting legitimate dissidents and affording privacy to ‘bad actors’” (2019)

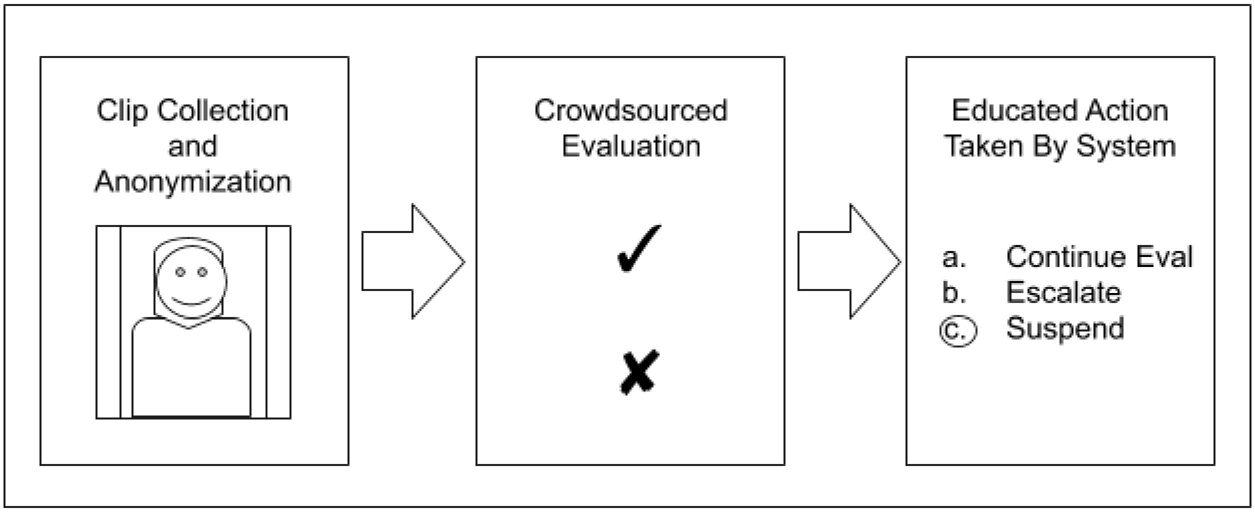
As the short University of Texas Ethics Unwrapped article points out, pure utilitarianism can lead to solutions that are not an “acceptable course of action, let alone the most ethical one.”(2019) In the case where harm is unavoidable because of individual rights, the principle of nonmaleficence and the corollary principle of harm reduction can serve as an additional guide in development. (Four fundamental ethical principles, 2012) In both the Twitch synagogue shooter and the Christchurch mosque attack, streams were available long enough to propagate throughout the internet. Catching potential problems early is key to minimizing the unavoidable harm caused by some software design decisions.

The author of this paper was disappointed to not find an abundance of scholarly articles that evaluated the ethical considerations relevant to the live-streaming of aggregated content. In spite of the limitations, the author proposes the following two principles in terms a software developer can understand, namely to apply the concepts of the Minimax algorithm when considering how to create strong ethical software features. These principles are to consider how to maximize the good to society in general while minimizing the possible damage to individual rights and privileges. The method proposed in the next section attempts to find a balance between the two.

Method:

Having ascertained the ethical responsibility of content aggregators to safeguard the utilitarian needs of the community from the literature the author of the paper proposes SnapShield. In the face of the failed Facebook algorithms in the 2019 mosque massacre, this software method leverages crowdsourcing and democratization to monitor live streams for inappropriate content. (Cheng, 2019) SnapShield is composed of three major sections. (see Fig. 1) First, software running on the streaming server randomly selects 3-5s clips of the video feed and partially anonymizes the clips. Those clips are then sent to random users of the SnapSheild app to evaluate. Finally, the software uses the user evaluations to determine the appropriate action to take with the aid of AI, preset parameters, and/or human operators.

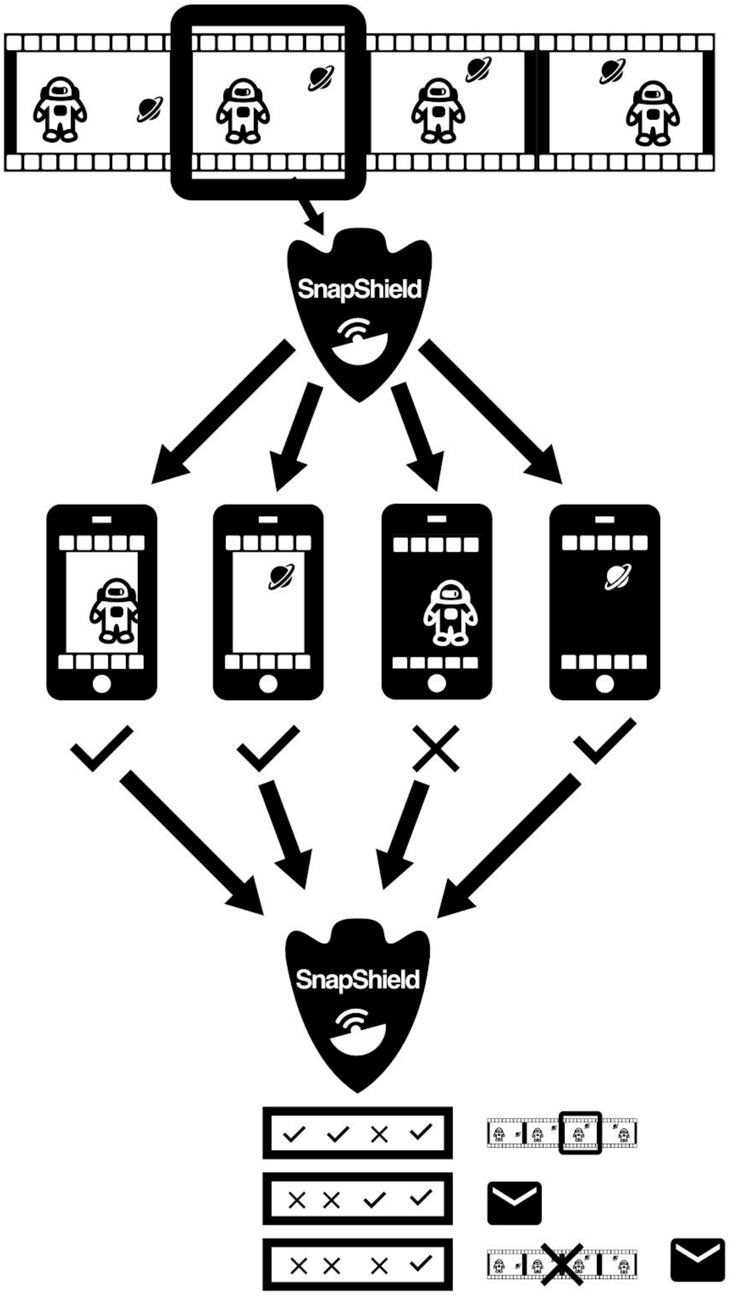
Figure 1


Clip Collection and Anonymization
Crowdsourced Evaluation
✓
✘
Educated Action Taken By System
a. Continue Eval
b. Escalate
c. Suspend

When a live stream is first started SnapShield will start to collect random sample clips of the video. The first step in preserving some anonymity is to not select too long of a clip. Then the clip experiences additional changes to further obfuscate its origin without making them unusable. For example, the audio of the clip may be slowed down or the color on part of the video image inverted. The result is multiple copies of the random clips each modified in some small way. This will help provide some degree of anonymity for the next step in the process. Finally SnapShield stores these modified clips on a server for consumption. The server will repeat the process at user-defined intervals.

SnapShield takes advantage of the same “voyeuristic principles” that make social media giants like Instagram so popular to help evaluate small sections of live streams. (Team, 2015) Users of the SnapShield mobile app can browse the modified clips. The user can swipe left or right to determine their opinion on the appropriateness of the clip. No single user is presented with more than one clip from the same live feed and the clip disappears forever once played. Appropriateness can be left to the user judgment or the app can prompt the user with a quick overview of the community standards.

Finally, the SnapShiled server takes the user responses from each batch of clips and compares it with a predetermined procedural list. For example, a single inappropriate evaluation may trigger the system to upload additional clips for evaluation. If fifty percent of the feedback is negative SnapShield may trigger a notification to the content aggregator to verify that the contents of the feed are appropriate while seventy-five percent negative feedback may trigger an automatic shutdown of the video pending further evaluation.



No one method could completely prevent the misuse of a live video feed feature. However, SnapShield uses popular technical solutions to protect the content aggregator and the public in turn by stopping such abuse as soon as possible. Like the technology it employs, there is a bevy of possible challenges to this method. The next section of the paper evaluates some of the ethical challenges that can arise from this innovative solution to live-video feed security.

Ethical Implications of Method:

In this section, the paper will first describe how the ethical principle of Minimax applies to the SnapShield method as delineated above. Then the method will be evaluated in terms of the IT Professionals New Zealand Code of Ethics. Finally, the author will describe some of the potential abuses a method like SnapShield could face.

This method maximizes the benefit to society. First, it incorporates a democratized method for evaluating the appropriateness of videos. It also allows for early intervention, reducing the amount of damage that could be inflicted on the public. However, this maximum benefit to the community comes at the cost of loss of privacy for the individual. This is why the method minimizes the damage to the individual by taking measures that obfuscate the identity and full content of the live stream. This is accomplished by providing only a small part of the stream, assuring that no single user gets more than one section of the stream, and manipulating the clip.

The code of ethics of the IT Professionals of New Zealand has eight basic tenents. Three are more pertinent to the method than the others. (itp.nz/Members/Code-of-Ethics, 2019) The development of SnapSheild already considers the first part of the concept of good faith in treating both the streaming users and the potential community views with dignity and equality, balancing the protection of individuals' rights with the protection of the community from bad apples. However special care will need to be taken to assure those cultural sensitivities are observed at all times. This is to observe the tenant of community focus. Finally, obtaining informed consent will allow facilitating the protection of the individuals.

Ultimately both the broadcaster and the evaluator need to know the risks involved. For the broadcaster they need to know that parts of their broadcast will be seen by strangers and that they need to moderate their content accordingly. For the evaluator, they need to be informed that they may be exposed to materials they may consider objectionable.

Some other challenges the method can face are the hijacking of the system, underrepresented peers, and human error introduced by internet trolls. Care will have to be taken to make sure that no one group can control the evaluation system. A government could, for example, use the system to censure streams in their countries by placing a majority of government-leaning users. Factors like the digital divide can also skew the results. In the US Caucasians are known for reporting African-American citizens to the police for unreasonable behavior. In this case, a majority of Caucasian users could mean fewer African-American live-streams go live. Finally, some kind of system will need to be put in place to catch users who flag user content negatively for the LOLs.

Conclusion:

Live-streaming is a practice that is becoming more popular year by year. (Clarke, 2019) It's not only beneficial for businesses but can also benefit society in general as Hong Kong protesters have proven in recent months. (Hundreds of Hong Kong, 2019) There is no doubt that social media needs to find a solution for better policing of live video streams. However, as the method proposed by this paper shows, that may mean sacrificing some personal freedoms for the benefit of the community.

Works Cited:

"22 Live Streaming Statistics for B2B Marketers." *ZoomInfo Blog*, 21 June 2019, blog.zoominfo.com/live-streaming-statistics/.

"Facebook Live." *Facebook Live | Facebook Media*, www.facebook.com/facebookmedia/solutions/facebook-live.

Fogel, Opinion by Yaffa. "Halle Survivor: Gun Control Saved My Life." *CNN*, Cable News Network, 11 Oct. 2019, edition.cnn.com/2019/10/11/opinions/halle-synagogue-attack-survivor-fogel/index.html.

Gonzalez, Oscar. "Twitch Video of Germany Shooting near Halle Synagogue Included Anti-Semitic Motives." *CNET*, CNET, 11 Oct. 2019, www.cnet.com/news/twitch-video-of-germany-shooting-near-halle-synagogue-included-anti-semitic-motives/.

"Hundreds of Hong Kong Protesters Storm Government Building over China Extradition Bill." *CNN*, Cable News Network, 1 July 2019, edition.cnn.com/asia/live-news/hong-kong-july-1-protests-intl-hnk/index.html.

"IT Professionals New Zealand: Te Pou Hangarau Ngaio." *Code of Ethics | IT Professionals*

New Zealand | Te Pou Hangarau Ngaio, itp.nz/Members/Code-of-Ethics.

Lazari-Radek, Katarzyna de, and Peter Singer. *Utilitarianism: a Very Short Introduction*. Oxford University Press, 2017.

"Live Streaming Video for Nonprofits." *Idealware*, 22 Sept. 2016, www.idealware.org/live-streaming-video-nonprofits/.

"Online Safety Laws and Rules." *Consumer Protection*, www.consumerprotection.govt.nz/general-help/consumer-laws/online-safety-laws-and-rules/.

Peacock, Colin. "The New Zealand Mosque Massacre: 2. 'End of Innocence' for Media and Nation." *Pacific Journalism Review : Te Koakoa*, vol. 25, no. 1and2, 2019, pp. 18–28., doi:10.24135/pjr.v25i1and2.490.

Team, The Labs. "Social Media and a Culture of Voyeurism." *Sutherland Labs*, www.sutherlandlabs.com/blog/social-media-and-a-culture-of-voyeurism/.

"Tough New Laws to Protect Australians from Live-Streaming of Violent Crimes." *Australian Government Crest*, 30 Mar. 2019, www.pm.gov.au/media/tough-new-laws-protect-australians-live-streaming-violent-crimes.

"Utilitarianism." *Ethics Unwrapped*, ethicsunwrapped.utexas.edu/glossary/utilitarianism.

Warzel, Charlie. "The New Zealand Massacre Was Made to Go Viral." *The New York Times*, The New York Times, 15 Mar. 2019, www.nytimes.com/2019/03/15/opinion/new-zealand-shooting.html.