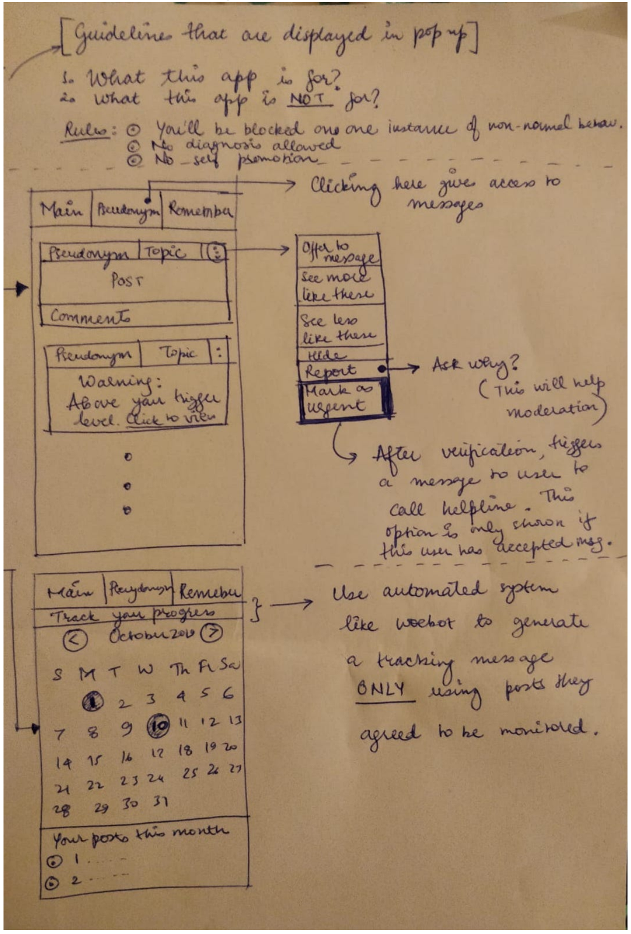
Assignment II - CS 6474 Social Computing

Sagarika Srishti (ssrishti3)

October 25, 2019

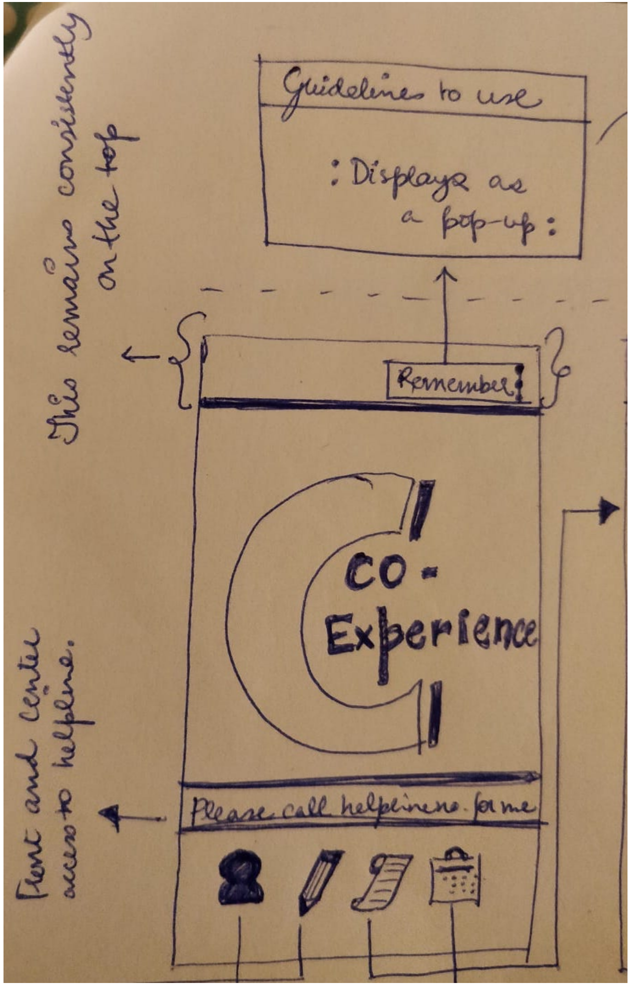
1. (a) I propose a design for an app that a user can access through their Facebook account, like apps PageModo, SurveyMonkey, and Woobox. This app will allow Facebook users who either are survivors of sexual abuse and are seeking support, or are people who want to provide help and support to others, or both, by giving them a platform to talk about their feelings that they're unable to share otherwise. For an app that is built for handling such a sensitive issue, some features are absolutely critical to it. First and foremost, the user should have an option of whether they want their real name, gender, location and other such personal details in the post or not. As seen in the class reading on social media disclosure of sexual abuse, people tend to use throwaway accounts, or accounts with pseudonyms for such purposes. Hence, the app will have an option for the user to select a pseudonym that they want to use in this platform. Secondly, there should be a clear set of rules and guidelines for the group that any user will have to read and understand as

Figure 1: Guidelines


[Guidelines that are displayed in pop up]
1. What this app is for?
2. What this app is NOT for?
Rules: ⊙ You'll be blocked on one instance of non-normal behav.
⊙ No diagnosis allowed
⊙ No self promotion
Clicking here gives access to messages
Main
Pseudonym
Remember
Pseudonym
Topic
Post
Comments
Pseudonym
Topic
Warning: Above your trigger level. Click to view
Offer to message
See more like these
See less like these
Hide
Report
Mark as urgent
Ask why? (This will help moderation)
After verification, triggers a message to user to call helpline. This option is only shown if this user has accepted msg.
Main
Pseudonym
Remember
Track your progress
October 2019
S M T W Th Fr Sa
1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 31
Your posts this month
1. ....
2. ....
Use automated system like woebot to generate a tracking message ONLY using posts they agreed to be monitored.

part of signing up for the platform. These guidelines should clearly state the norms of behavior, which is very important given the sensitivity of the topic of discussion, and also have additional links to helpful resources like helpline numbers that a user can be directed to if they are showing signs of extreme distress. The guidelines can also be accessed through a tab on the top of the app. Thirdly, as seen in the reading (Andalibi, Haimson, De Choudhury, & Forte, 2016) that males find it particularly harder to self-disclose when it comes to sexual abuse, the app will allow the user to form their preferred listener and speaker profile while signing up. For example, while signing up or through account settings, a person can select the categories about which they are likely to post on the platform, like 'sexual abuse of males', 'rape', 'child molestation', etc. Similarly, a user can select the categories whose posts they would be okay with seeing, which issues, what kind of people, etc. For example, if a person is not comfortable reading posts about rape, they can exclude the category from their listener profile. This feature can help male survivors to select an appropriate audience for opening up, so that they feel comfortable knowing that the audience that will see their self-disclosure is likely to help them.

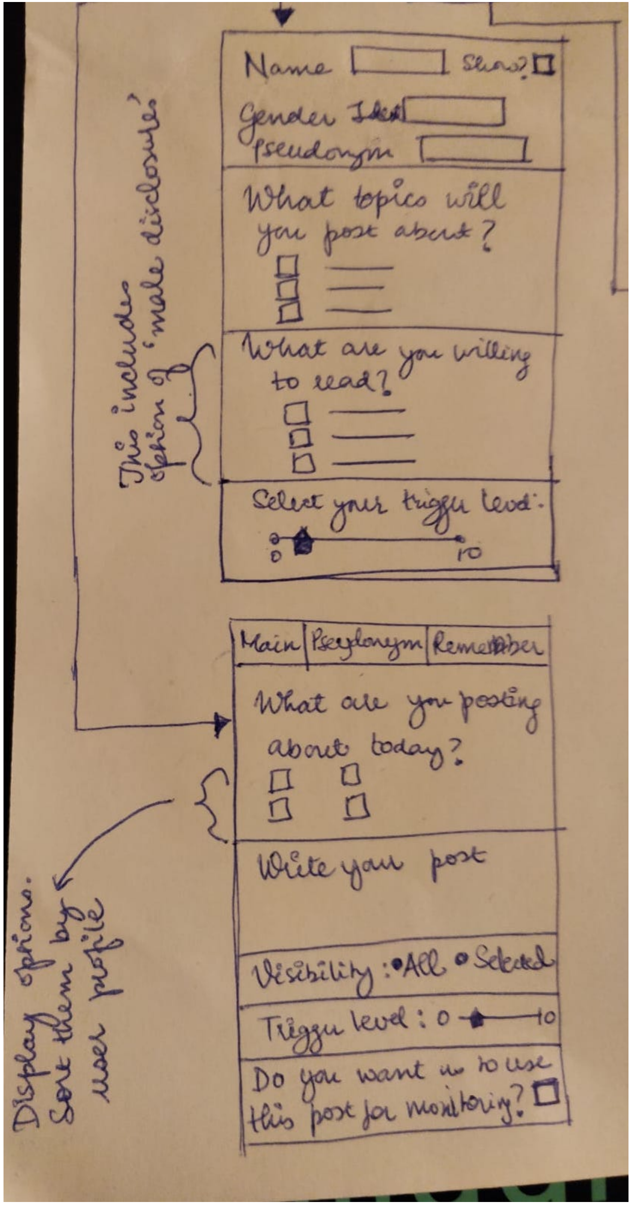
Figure 2: App interface


This remains consistently on the top
Guidelines to use
: Displays as a pop-up :
Remember!
CO-Experience
Please call helpline no. for me
Front and center access to helpline.

Removals and bans would be necessary in this app. A user's access to the community would be blocked even on one occurrence of bullying, spamming, or making fun. Also, the app would have automatic or human-annotated warning tags, depending on the content of the post. For example, a post about rape self-disclosure should have a warning tag of 'Graphic narration: rape content'.

Given the gravity of the issue, there are some features that absolutely shouldn't be present in the app. In this app, reposting a post, to Facebook or to any other platform would not be allowed.

Figure 3: Signing up for the platform


Name
Show? □
Gender Ident
Pseudonym
What topics will you post about?
What are you willing to read?
Select your trigger level:
0
10
This includes option of 'male disclosures'
Main
Pseudonym
Remember
What are you posting about today?
Write your post
Visibility: ●All ○ Selected
Trigger level: 0
10
Do you want us to use this post for monitoring? □
Display options.
Sort them by user profile

Also, I think that people who come to the platform for self-disclosing want to get something off their chest, and are seeking support and comfort. They're not seeking diagnosis, for example 'this post shows signs of bipolar disorder.' Hence, any comments on health or diagnoses should be marked as spam. Moreover, given that this is a support forum, and not a doctor forum, self-promotion shouldn't be supported by the app. No counsellor or doctor should have the chance to promote their practice on the platform. Another feature that the app wouldn't allow, unless specifically allowed by a user, is personal messaging, i.e., all comments should be public. This feature if allowed can be like Quora's messaging feature, where a user only allows specific users to be able to message them.

(b) To assess the effectiveness of the above app, I propose a study design in an academic setting. To recruit participants for the study, a survey can be integrated into the university health service of a university, for example, when a student signs in to the Patient Portal at Stamps Health Services of Georgia Tech, they will be asked to participate in a survey if they're interested, with the front page stating what the survey is about. Additionally, students who want to disclose can be recruited by sending out an email to the entire student body of the university. In the mail, the purpose and salient features of the app will be explained. Interested students will have the advantage of having a platform for self-disclosing as an incentive, and a small in kind incentive such as food coupons to a university restaurant can be further added.

Since we are external to Facebook and don't have server data, data for the study will be collected through self-assessment forms that the users will fill at the beginning and end of their duration in the study. A set of questions would be asked to participants at the beginning of the survey, when they have expressed an intent and need to self-disclose about sexual abuse, and at the end, when they have self-disclosed and can give a feedback about their experience with the app.

To measure the effectiveness of the app, half of the participants would be asked to use this app for self-disclosure, and the other half would be directed to another platform like Reddit or Rainn. The surveys conducted for both sets of users at the beginning and at the end would give an idea of how well the self-disclosure needs of a participant are being met through this app, as compared to other platforms.

One problem with gathering information from university students is that they are not representative of the overall demographic. A possible way to mitigate this problem is to have research collaborators from other walks of life, like working professionals from the industry, a blue-collar workers union, etc. Another problem with the study is that, how will I as the researcher know when a user has posted on the Facebook app, so that a reminder can be sent to them to fill out the second part of the survey once they've used the app for self-disclosure and received support. Some form of linking to Facebook so that we could know when a user has posted could help mitigate this problem.

2. (a) Twitch is an online service used for watching or broadcasting live or prerecorded videos, where the streamer can add audio commentary along with the video, and the viewers can participate via live comments on the videos. While the primary purpose of Twitch when it was started in 2011 was the streaming of video games being played (Investopedia, 2015), the content on Twitch has since expanded to include categories like music and art. There can be various cases of hate speeches and harassing attitudes on Twitch, such as comments on a person's race/ethnicity, offensive remarks about women players, hate speech about a person's sexual orientation, accent etc. Such behavior can discourage streamers and cause them to leave the platform, and can even harm their mental well-being. Cyberbullying has even lead to suicides in the past (Foundation, 2019), and hence is a very serious problem which should be handled. Out of the design choices discussed by Kiesler et al. in the chapter Regulating Behavior in Online Communities (Kraut & Resnick, 2012), I think the most appropriate design choice for solving the issue of hate speech and pro-harassment attitudes on Twitch will be the implementation of Roles, Rules, Policies, and Procedures.

There are various reasons because of which I think this is the right design category. For starters, in a platform as large as Twitch (over 100 million unique users in a month), trolls are bound to be present. And as we have seen in the class reading 'Identity and Deception in the Virtual Community', the trolls who spread hate speech on purpose feed on people's reactions. Hence, a DNFTT (Do Not Feed The Troll) policy should be adopted as a ground rule in Twitch. Secondly, when the user goes to a particular channel, there can be an additional tab alongside the preexisiting Live Channels, Videos and Clips tabs, which can have guidelines decided upon by the followers of that channels. The members and viewership of Twitch are predominantly male and white (Jørgensen & Karlsen, 2019), and may not have the awareness about the normative behavior, for example, can unintentionally engage in a harassment attitude, hence the clear guidelines will help. Thirdly, a subset of people who are regular visitors, have a minimum set account age (age of account not the account holder's age), and no history of non-normative behavior can be selected at random, for a given period of time, to act as moderators, and decide on appeals, if any. This maintains fairness, and increases the legitimacy of the procedure. This also ensures that the moderators have limited and rotating power. Meta-moderation system like Slashdot.com can also be implemented. Fourthly, if a person notices some bad behavior still exists, like casual sexist comments wrapped in humor, but there's not a rule against it, they can ask for the creation of a rule, have a petition, and if the number of signs reach a certain percentage, the moderators will have to take a decision about it.

For a platform like Twitch, naive approaches like removals and bans won't work as registration for being able to participate in chats is free, and a troll once blocked will just form a new account. Other moderation techniques like pre-screening, disemvoweling, and degrading posts can't be applied since this is a live stream and there's no time for screening the posts, and there's a surety that whatever the user posts will appear on the screen.

In terms of robustness, bringing the millions of viewers to a consensus about the guidelines of a channel can be problematic. I think the process of randomly choosing moderators with rotation is hard to game as no individual has greater control than others.

(b) There are various subreddits on Reddit whose discussion is around the topics of health and well-being, like r/health, r/mentalhealth, r/nutrition, etc. On these subreddits, people mostly ask for or provide recommendations/advices. Hence, this is a perfect breeding ground for the spread of misinformation and low-credibility information, with posts about pseudoscience, fake medicines claimed to be working, photoshopped pictures of results, etc. present. Out of the design categories discussed by Kiesler et al. (Kraut & Resnick, 2012), the most appropriate one according to me to deal with this problem is 'Selection, Sorting, and Highlighting'.

The reasons why this approach works are manifold. First, sometimes newcomers, or people who are in general unaware that the grandma's tricks and tips that they've been using are not scientifically backed, end up spreading misinformation. This is a non-normative behavior that can be controlled by having guidelines, making the newcomers read them before allowing them to follow a subreddit. There can be a tab in the Community Details section on the right hand side of the screen in a subreddit (highlighting). Second, moderation systems that pre-screen, degrade, label, move, or remove inappropriate messages can work here as it will limit the impact of misinformation spread. For example, posts could be pre-screened for red flag phrases like 'anti-vax', 'Maybe this is all in your head', 'Get busy and distract yourself', 'Why can't you work?' (Tartakovsky, 2018). Third, like the option of 'Give Award' in Reddit posts, downvotes can be introduced for which a reason has to be provided, so that misinformation posts get marked as such and the ratings of the poster is reduced. Fourth, a cost can be introduced for posting, which can be recovered through upvotes, but this cost will discourage manipulators from posting. Fifth, a nofollow attribute can be attached to a post, which will only be removed once the poster has collected a minimum required number of karma through posts in that subreddit. This increases the credibility of information.

Naive approaches like gags and bans won't work here because on Reddit, people tend to use multiple throwaway accounts (Andalibi et al., 2016), which can be done by people spreading misinformation as well if they're blocked once.

As for the robustness of this approach, the downside of the downvoting feature is that people might not feel compelled towards giving a reason, and choose to just ignore it. Secondly, intelligent agents might learn over time what kind of posts are being downvoted and discounted, and can try to game the system by altering their method of posting.

(c) The design choices I suggested for parts (a) and (b) were dissimilar, and belonged to different categories. The design choices made for regulating non-normative behavior on different social computing platforms primarily differ because communities can differ greatly in what behaviors are normative and what are not. For example, the use of informal language is the norm on Facebook, but would be unacceptable for a platform like LinkedIn, even though the language might not contain anything offensive. Similarly, while Wikipedia expects editors to have a neutral point of view while editing any article, a media platform like Fox news would expect its guest editors to have a left or right leaning point of view. As the definition of normative behaviors changes from community to community, the methods applied to prevent or handle non-normative behaviors change as well.

Not just the setting of the platform, but the problems that it faces can also be very different. For example, while Twitch suffers from trolls and griefers, manipulators and misinformation spreaders are the main problems for recommendation sites like Rotten Tomatoes and subreddits where information about health and well being is shared. Again, different design choices are required when dealing with different kinds of non-normative behavior.

Although the two platforms are quite different in their functioning and norms, there are still some opportunities for adopting similar design choices for regulating non-normative behavior. For example, even though the types of bad behaviors on the two platforms are different, the design choice of having clearly stated out guidelines and policies, and strictly enforcing them via a graduated sanctions can be applied to both the platforms. Also, both of these platforms might suffer from the problem of bot activity. For example, bots might be bombarding a Twitch channel stream with hateful comments, while on a subreddit a bot might be responsible for spamming it with posts about fake health supplements with outlandish claims that don't work. In this case, a design choice to identify and remove or discount bot activity can be adopted for both Twitch and Reddit.

Since registration for participating in the live chats is free in Twitch, and there's no loss for a viewer who gets blocked and there's no cost for pseudonym switching. However, on Reddit, a user's profile matters as they keep collecting karma, which is reflective of their reputation and experience on the platform. Therefore, while rewards and penalties on profiles can work on Reddit for regulating non-normative behavior, it won't work on Twitch, and different design choice needs to be considered.

References

Andalibi, N., Haimson, O. L., De Choudhury, M., & Forte, A. (2016). Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 3906–3918). ACM.

Foundation, M. M. (2019). *Bullying, cyberbullying, suicide statistics.* `meganmeierfoundation.org/statistics`.

Investopedia. (2015). *How twitch.tv works and its business model.* www.investopedia.com/articles/how-twitchtv-works-and-its-business-model.asp.

Jørgensen, K., & Karlsen, F. (2019). Kaceytron and transgressive play on twitch.tv. In M. Consalvo (Ed.), *Transgression in games and play* (p. 83-120).

Kraut, R. E., & Resnick, P. (2012). Regulating behavior in online communities. In S. Kiesler, R. E.Kraut, P. Resnick, & A. Kittur (Eds.), *Building successful online communities: Evidence-based social design* (p. 125-162).

Tartakovsky, M. (2018). *9 things not to say to someone with mental illness.* psychcentral.com/blog/9-things-not-to-say-to-someone-with-mental-illness.