Society Against Violence and Abuse

A set of proposed ideas for making Instagram
a safer platform

Critical Research

Orlikowski and Baroudi classify research as critical where a critical stance is taken toward taken-for-granted assumptions about organizations and information systems, and where the aim is to critique to status quo “through the exposure of what are believed to be deep-seated, structural contradictions within social systems”

Klien, Hienz, and Michael Myers. A SET OF PRINCIPLES FOR CONDUCTING CRITICAL RESEARCH IN INFORMATION SYSTEMS. MIS Quarterly

Globally 1 out of 3 women have experienced gender based violence or abuse in their lifetime.

This form of gender based hate crime has a very long history, and has taken numerous forms with varying intensity over the years.

Yet even today, less than 40% women seek help of any sort, less than 10% report to the police and choose to disregard or deal with the issue themselves.

We have come to a point where there are multiple organisations and more help around us than ever before, but VAWG (Violence and Abuse against women and girls) is still rampant.

Hence it becomes necessary to understand how and where we are failing women and girls, and what are the challenges that abide.

We conducted critical research on the following two subjects, and understood what works and what doesn't in both approaches towards the issue

MAVA

Men Against Violence and Abuse (MAVA), is an organisation that engages men and boys instead of women and girls, in conversations about gender inequality and male dominance in our society, and also how it affects them at a personal level. They view men as equal partners and stakeholders, not just as perpetrators in this issue of VAWG.

TOXIC TWITTER

Amnesty International is a group of human rights activists, that took a different turn on twitter's toxicity. They used data science and crowd sourcing to conduct research and study on abusive tweets that women recieve every 30 seconds. They found patterns and fatual data from the masses of tweets, that twitter needed to address and hasn't. They held the platform responsible for turning it into a battlefeild for women.

Next →

MAVA

Some of our insights and critique on MAVA are :

Archaeology and Geneology of Knowledge

Traditional efforts to tackle gender-based violence against women have concentrated on empowering women to assert themselves.

Cultural Capital

Effects of mainstream media (films, ads and internet) and also stories of gods in religion like hinduism, glorifies male dominance, violence and gender norms. This has a direct impact to the way men are brought up

Discipline

Using the existing power structures as an advantage to enable power sharing among the oppressed

MAVA primarily focuses on adolescent men and not men in power that hold a bigger percentage for social reproduction

Habitus

Socially conditioned by the ideals of masculinity also makes men significantly blind towards the contributions of women

MAVA wants oppressors to be a part of conversations and movements where the oppressed are usually the only ones fighting at the forefront, the majority often does not take onus of protecting the minority.

Economic Capital

Women's vulnerability to male dominance has been long associated and impacted by their economic disempowerment.

This has a direct impact on women being unable to accept positive changes in men

TOXIC TWITTER

Some of our insights and critique on Toxic Twitter campaign are :

Archaeology and Geneology of Knowledge

V and A against women on Twitter is not new, its simply any extension of existing and systematic discrimination against women that has found its way into the digital sphere.

Cultural Capital

Instead of women using their voices to 'impact the world', many are instead being pushed backwards to a culture of silence

India's linguistic diversity, local slang makes it difficult for Twitter's reporting system to analyse and block reported content

Discipline

Amnesty, as a Human Rights group calls out powerful groups by investigating and exposing facts. It uses media as a tool to unite people from everywhere and seek for justice.

Habitus

Lack of real world consequences tend to bring out the worst in people

Amnesty focused on crowdsourcing information, primary and secondary research and data analysis compared to individual emancipation/strategic action.

Economic Capital

Twitter also has workplace effects for many women. Due to constant abuse, their work is also unfairly hindered thus affecting their economic capital

For female journalists, being active on Twitter or any other social media platform is quite necessary, yet they're forced to prepare themselves for any kind of backlash not just to do their job but to fight for their space as women on on the internet.

Objective

Our main objective with the project was to make the internet a safe space for all genders by attempting to curb online abuse. But we narrowed it down to :

“Making social media platforms safer safer by attempting to curb online hate and abuse”

Social Media Reporting Systems

Facts and Figures :

From the 40 responses we recieved on our survey,
- 92.5% have reported gender based abuse online
- 85% have reported on Instagram, as compared to other social media platforms
- 55% recieved an uncertain response to their report, and 22% recieved no action at all
- 50% people didn’t choose to report because they didn’t believe it would make an impact

Link to the form : https://forms.gle/d1byqw5i23eR55nW6

Instagram was chosen as the working space for the project

Ideation

These are some of our ideas from our SCAMPER brainstorming session

Substitute the action of blocking the offender with a learning process or a tedious game that involves a learning process to get back into instagram

S

Combine online reporting and police reporting systems to ensure real world consequences

C

Verify each account for Instagram at the time of creating one. Ask users to update and verify their account periodically. (like tinder/bumble)

A

Minify the engagement that abuser gets on their account when people encourage others to report them.

M

Using reporting system in emergency situations (SOS) that could perhaps help save someone in a real world scenario

P

Eliminate the process of a user having to choose to view sensitive content without enough context

E

Using reverse pyschology on offenders to reconsider their comment

R

View all our research and ideation on this miro board :
https://miro.com/welcomeonboard/Qtf9rBIDVDNPsjJOhP2Nk222q3qMnBY2Q20QaT4uErAsvT94LRkeG8PuaxJlza48

Final Outcome →

Idea 1

Description

Modifying the account creation process, which would make users take the policies and code of conduct seriously

Verifying account history and linking accounts to a person's biometrics helps Instagram background check their users and debars them from creating multiple fake accounts. Customization during account creation process is a new add-on which allows users to choose from a given set of hashtags. Selecting them under the customization section, Instagram would make sure to not show any posts related to those hashtags. Thus, providing users a comfortable user experience. A clear explanation of the policies and code of conduct at the beginning (during the account creation process) is necessary. We understood from our research that regional slurs and abuse is not detectable by Instagram. Which is why mentioning the region you're in, would modify the experience accordingly and make the reporting procedure customized to the region.

Instagram benefits

1. Conduct a background check on its users to understand their history of locations to get a better sense of their communicative attributes.

2. Opens the door to communicating advertisements in regional language to its users.

3. Better filtration system is equally proportional to easier recognition of the user's attributes.

4. Provide similar sponsorship based on the interests of the user, on all their accounts. Easier to recognize its the same person on multiple accounts.

5. Removing the person from all their accounts after being reported on one. Makes sure Instagram is a safe space.

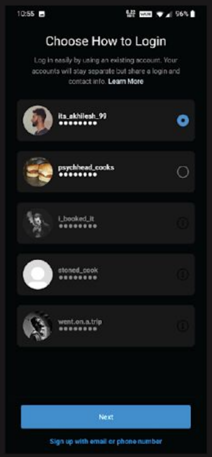
CR concepts

Panopticon :
Self Surveillance

Discourse

Instagram

Account creation process update



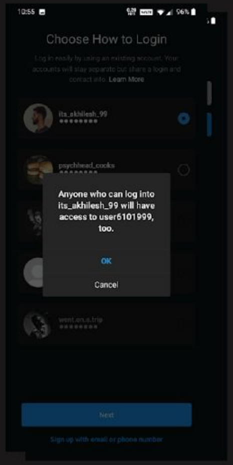
Choose How to Login

Log in easily by using an existing account. Your accounts will stay separate but share a login and contact info. **Learn More**

- its_akhilesh_99 ●●●●●●●● ●
- psychhead_cooks ●●●●●●●● ○
- i_booked_it ●●●●●●●●
- stoned_cook ●●●●●●●●
- went.on.a.trip ●●●●●●●●

Next

Sign up with email or phone number



Choose How to Login

Log in easily by using an existing account. Your accounts will stay separate but share a login and contact info. Learn More

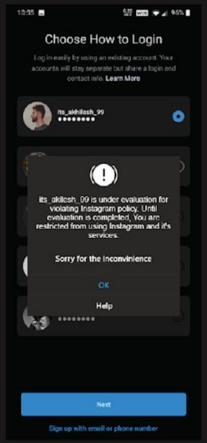
Anyone who can log into its_akhilesh_99 will have access to user6101999, too.

OK

Cancel

Next

Sign up with email or phone number



Choose How to Login

Log in easily by using an existing account. Your accounts will stay separate but share a login and contact info. **Learn More**

its_akhilesh_99 ●●●●●●●●

its_akilesh_99 is under evaluation for violating Instagram policy. Until evaluation is completed, You are restricted from using Instagram and it's services.

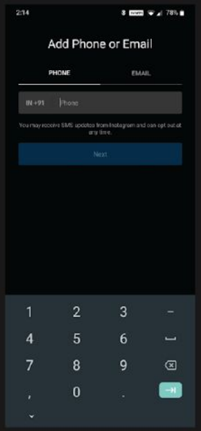
Sorry for the inconvinience

OK

Help

Next

Sign up with email or phone number



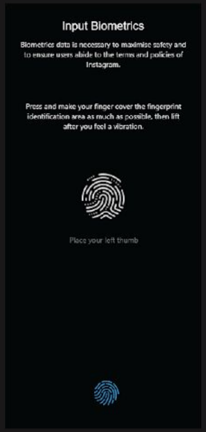
Add Phone or Email

PHONE | EMAIL

IN +91 | Phone

You may receive SMS updates from Instagram and can opt out at any time.

Next

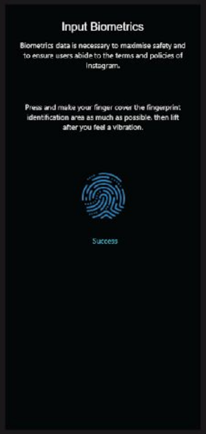


Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.

Place your left thumb

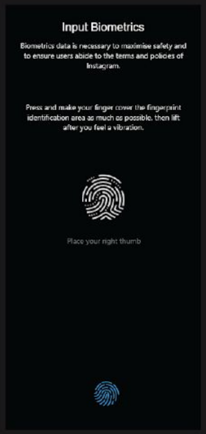


Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.

Success

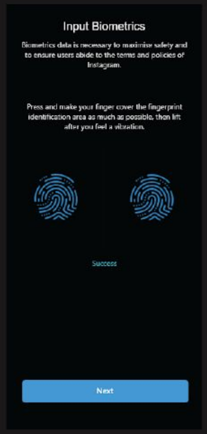


Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.

Place your right thumb



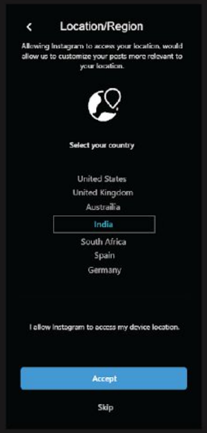
Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.

Success

Next



Location/Region

Allowing Instagram to access your location, would allow us to customize your posts more relevant to your location.

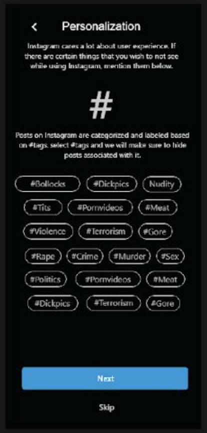
Select your country

United States
United Kingdom
Austrailia
India
South Africa
Spain
Germany

I allow Instagram to access my device location.

Accept

Skip



Personalization

Instagram cares a lot about user experience. If there are certain things that you wish to not see while using Instagram, mention them below.

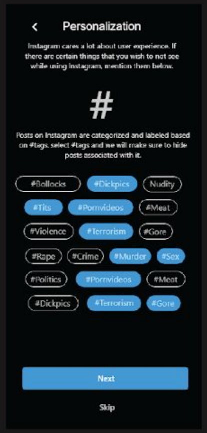
#

Posts on Instagram are categorized and labeled based on #tags. select #tags and we will make sure to hide posts associated with it.

#Bollocks #Dickpics Nudity
#Tits #Pornvideos #Meat
#Violence #Terrorism #Gore
#Rape #Crime #Murder #Sex
#Politics #Pornvideos #Meat
#Dickpics #Terrorism #Gore

Next

Skip



Personalization

Instagram cares a lot about user experience. If there are certain things that you wish to not see while using Instagram, mention them below.

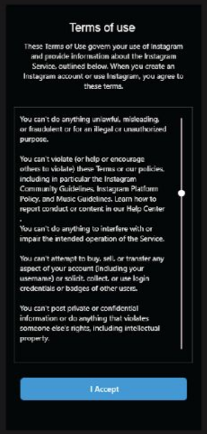
#

Posts on Instagram are categorized and labeled based on #tags. select #tags and we will make sure to hide posts associated with it.

#Bollocks #Dickpics Nudity
#Tits #Pornvideos #Meat
#Violence #Terrorism #Gore
#Rape #Crime #Murder #Sex
#Politics #Pornvideos #Meat
#Dickpics #Terrorism #Gore

Next

Skip



Terms of use

These Terms of Use govern your use of Instagram and provide information about the Instagram Service, outlined below. When you create an Instagram account or use Instagram, you agree to these terms.

You can't do anything unlawful, misleading, or fraudulent or for an illegal or unauthorized purpose.

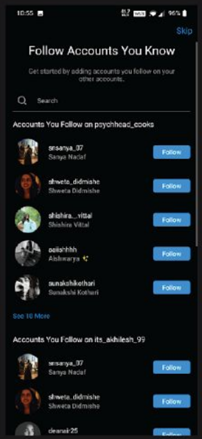
You can't violate (or help or encourage others to violate) these Terms or our policies, including in particular the Instagram Community Guidelines, Instagram Platform Policy, and Music Guidelines. Learn how to report conduct or content in our Help Center .

You can't do anything to interfere with or impair the intended operation of the Service.

You can't attempt to buy, sell, or transfer any aspect of your account (including your username) or solicit, collect, or use login credentials or badges of other users.

You can't post private or confidential information or do anything that violates someone else's rights, including intellectual property.

I Accept



Skip

Follow Accounts You Know

Get started by adding accounts you follow on your other accounts.

Search

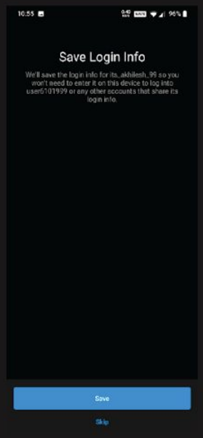
Accounts You Follow on psychhead_cooks

snsanya_07 Sanya Nadaf	Follow
shweta_didmishe Shweta Didmishe	Follow
shishira._vittal Shishira Vittal	Follow
aaiishhhh Aishwarya ✨	Follow
sunakshikothari Sunakshi Kothari	Follow

See 10 More

Accounts You Follow on its_akhilesh_99

snsanya_07 Sanya Nadaf	Follow
shweta_didmishe Shweta Didmishe	Follow
deanair25	Follow

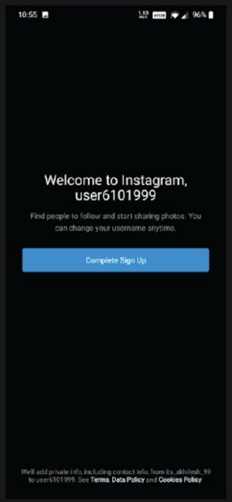


Save Login Info

We'll save the login info for its_akhilesh_99 so you won't need to enter it on this device to log into user6101999 or any other accounts that share its login info.

Save

Skip



Welcome to Instagram, user6101999

Find people to follow and start sharing photos. You can change your username anytime.

Complete Sign Up

We'll add private info, including contact info, from its_akhilesh_99 to user6101999. See **Terms**, **Data Policy** and **Cookies Policy**.

Idea 2

Description

Sensitive content warning with descriptors and keywords, to give more context to the same and refining feed by giving users more control

This idea looks at posts labeled as 'sensitive content' on IG, where posts that can trigger users with explicit imagery are blurred. However, this becomes counter-productive since the blurriness of the image only make users more curious to tap on it and end up getting scandalized with the graphic images/video, sometimes to not be able to report it too. In this case, having more context through keywords about what the 'sensitive content' is really about, gives users more information, making them aware and allowing them to make an informed decision. Taking a step further, we're proposing a "Hide" feature under each post, where users can opt to not see posts of a certain kind. We're also extending one of the ideas from the account creation process here, where users decide what kind of content they don't wish to see by customizing hashtags.

Instagram benefits

1. Keywords can be used to improve image recognition, for it to describe the content without the need of a person.

2. Descriptors can gain traction over those specific issues with relevant audiences

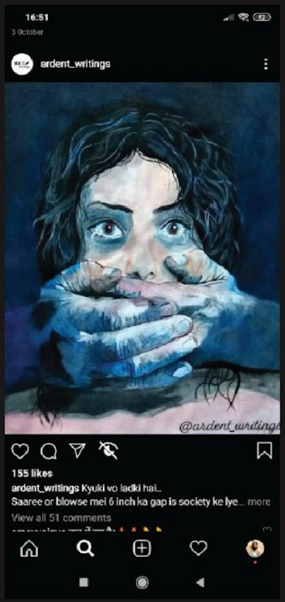
3. Better filtration system increases instagram's archive and narrows down on the range of ads that will grab the user

CR concepts

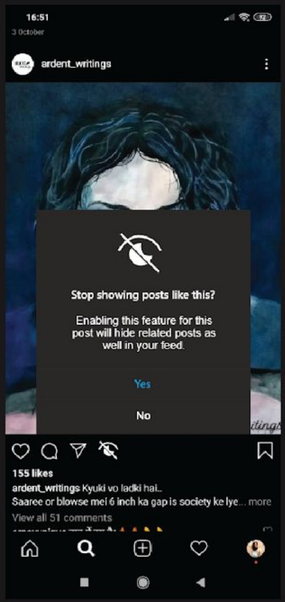
Panopticon :
Self Surveillance

Cognitive Interests

Option for users to choose the kind of posts they wish to see



ardent_writings Kyuki vo ladki hai..
Saaree or blowse mei 6 inch ka gap is society ke lye... more
View all 51 comments



Stop showing posts like this?

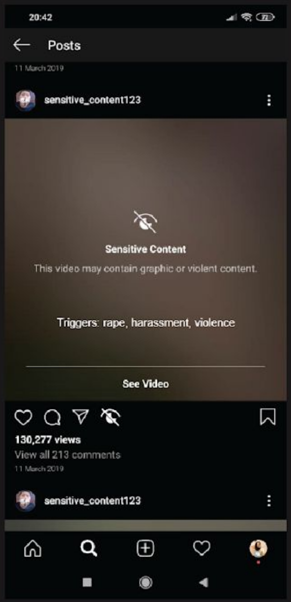
Enabling this feature for this post will hide related posts as well in your feed.

Yes

No

ardent_writings Kyuki vo ladki hai..
Saaree or blowse mei 6 inch ka gap is society ke lye... more
View all 51 comments

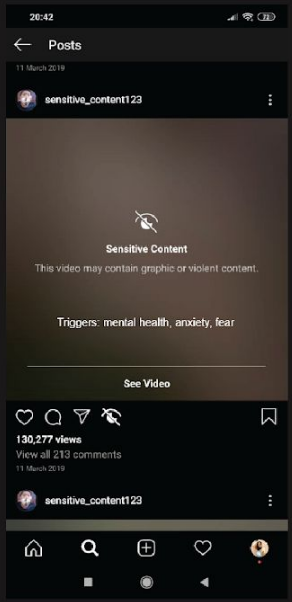
Keywords on sensitive content to give user context about the post



Sensitive Content
This video may contain graphic or violent content.

Triggers: rape, harassment, violence

See Video

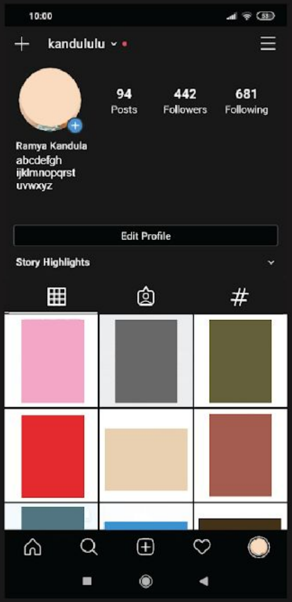


Sensitive Content
This video may contain graphic or violent content.

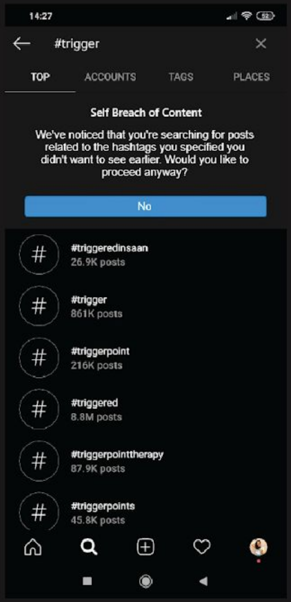
Triggers: mental health, anxiety, fear

See Video

Customize content through hashtags



Cautioning user when self breach of content occurs



Self Breach of Content

We've noticed that you're searching for posts related to the hashtags you specified you didn't want to see earlier. Would you like to proceed anyway?

No

#triggeredinsaan
26.9K posts

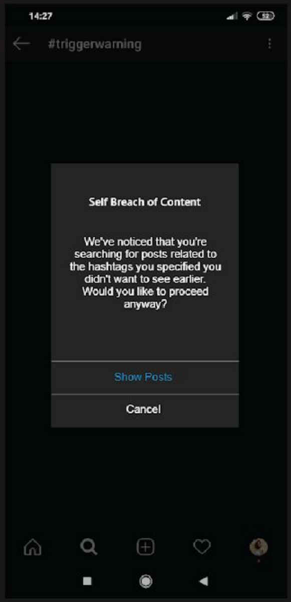
#trigger
861K posts

#triggerpoint
216K posts

#triggered
8.8M posts

#triggerpointtherapy
87.9K posts

#triggerpoints
45.8K posts

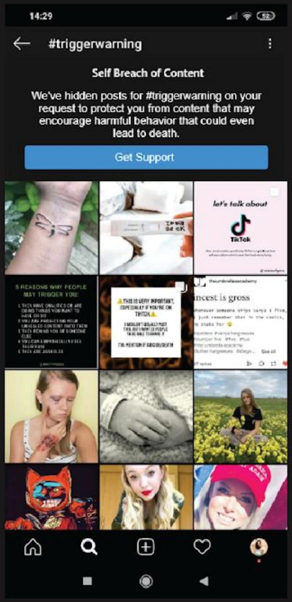


Self Breach of Content

We've noticed that you're searching for posts related to the hashtags you specified you didn't want to see earlier. Would you like to proceed anyway?

Show Posts

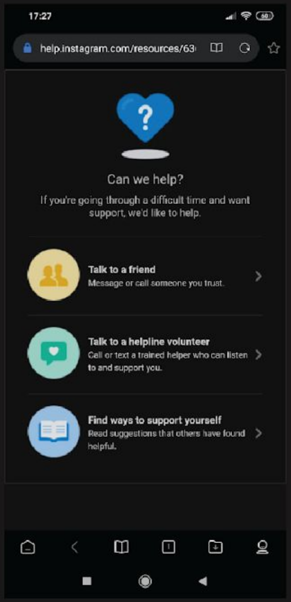
Cancel



Self Breach of Content

We've hidden posts for #triggerwarning on your request to protect you from content that may encourage harmful behavior that could even lead to death.

Get Support



Can we help?

If you're going through a difficult time and want support, we'd like to help.

Talk to a friend
Message or call someone you trust.

Talk to a helpline volunteer
Call or text a trained helper who can listen to and support you.

Find ways to support yourself
Read suggestions that others have found helpful.

Idea 3

Description

Reverse reporting system by putting abuser through similar harrowing steps that usually abused person needs to to go through

The number of steps to post an abusive comment is three while the number to report one is six. In adopting survivor-centric preventative measures in response to GBV, it is unfair that the onus to seek justice for the abused individuals/communities lies only with themselves. It is possible that survivors of online abuse have already been traumatized and have had this affect their mental health negatively. This proposed new feature for posting abusive comments aims at eliminating and minimalizing the need to report by attempting to nip the abuse in the bud while also educating the abuser. By providing the same sense of uncertainty, hopelessness and discouragement that is otherwise felt by the abused person even after reporting the abuser, would eventually lead to self-surveillance in the abuser.

Instagram benefits

1. Paid tie ups with multiple news websites, organizations and ad revenue, when promoted during the process

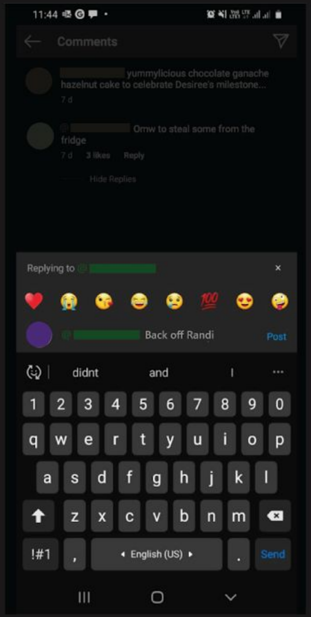
CR concepts

Panopticon :
Self Surveillance

Communicative Action

Habitus

Archaeology of Knowledge



11:44

Comments

[redacted] yummylicious chocolate ganache hazelnut cake to celebrate Desiree's milestone...
7 d

@[redacted] Omw to steal some from the fridge
7 d 3 likes Reply

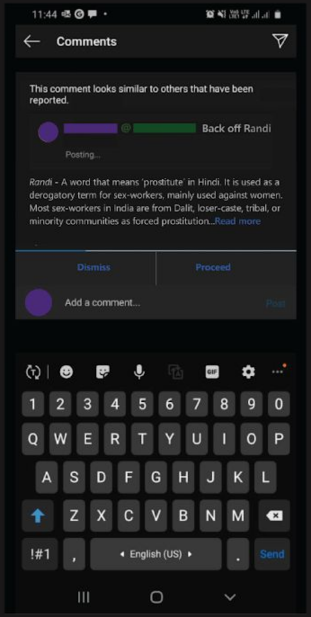
Hide Replies

Replying to @[redacted] ×

❤️ 😭 😘 😄 😢 💯 😍 🤪

@[redacted] Back off Randi Post

Abuser types in comment



11:44

Comments

This comment looks similar to others that have been reported.

[redacted] @[redacted] Back off Randi

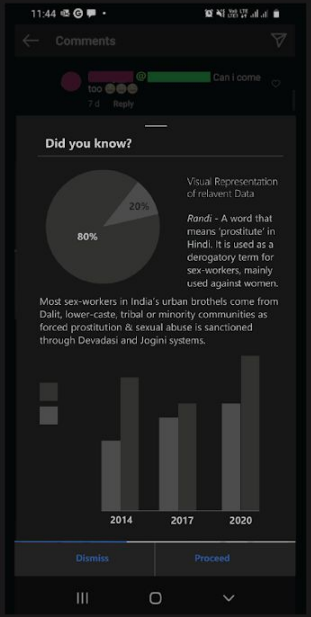
Posting...

Randi - A word that means 'prostitute' in Hindi. It is used as a derogatory term for sex-workers, mainly used against women. Most sex-workers in India are from Dalit, loser-caste, tribal, or minority communities as forced prostitution...Read more

Dismiss Proceed

Add a comment... Post

AI detects abusive words/phrases/ tones. User is asked if they want to still post comment



11:44

Comments

[redacted] @[redacted] Can i come too 😑😑😑
7 d Reply

Did you know?

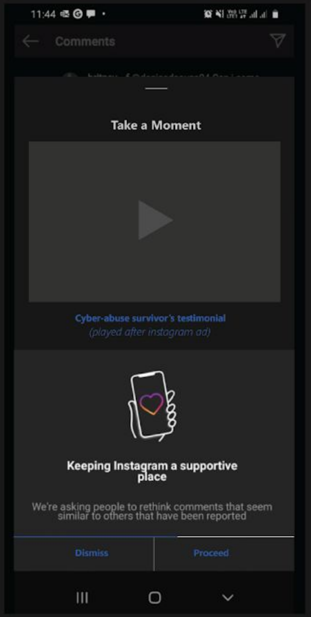
Visual Representation of relavent Data

Randi - A word that means 'prostitute' in Hindi. It is used as a derogatory term for sex-workers, mainly used against women.

Most sex-workers in India's urban brothels come from Dalit, lower-caste, tribal or minority communities as forced prostitution & sexual abuse is sanctioned through Devadasi and Jogini systems.

Dismiss Proceed

Suggests alternative use of words and shows info related to the abusive language



11:44

Comments

Take a Moment

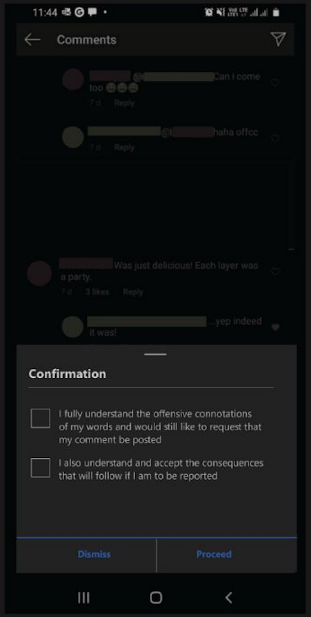
Cyber-abuse survivor's testimonial
(played after instagram ad)

Keeping Instagram a supportive place

We're asking people to rethink comments that seem similar to others that have been reported

Dismiss Proceed

If they wish to go ahead, they're presented with additional info tabs about the effect of the particular kind of abuse



11:44

Comments

[redacted] @[redacted] Can i come too 😑😑😑
7 d Reply

[redacted] @[redacted] haha offcc
7 d Reply

[redacted] Was just delicious! Each layer was a party.
7 d 3 likes Reply

[redacted] ...yep indeed it was!

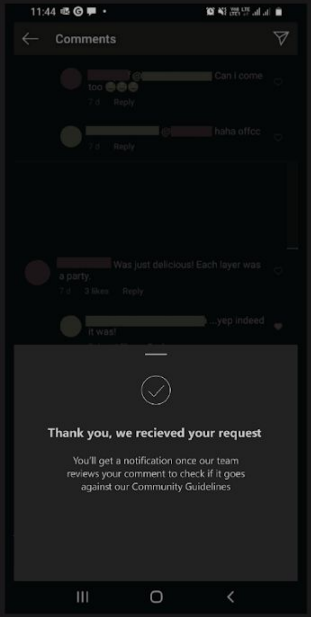
Confirmation

□ I fully understand the offensive connotations of my words and would still like to request that my comment be posted

□ I also understand and accept the consequences that will follow if I am to be reported

Dismiss Proceed

Users get an option to retract and end process anytime which when clicked will dismiss the comment



11:44

Comments

[redacted] @[redacted] Can i come too 😑😑😑
7 d Reply

[redacted] @[redacted] haha offcc
7 d Reply

[redacted] Was just delicious! Each layer was a party.
7 d 3 likes Reply

[redacted] ...yep indeed it was!

Thank you, we recieved your request

You'll get a notification once our team reviews your comment to check if it goes against our Community Guidelines

If user goes through all steps/tabs they receive a an uncertain response saying only that their comment will be reviewed and only then if it doesn't go against CG will it be posted

Idea 4

Description

Substituting reporting options with examples of abuse for the user to understand which category to choose

The purpose of this idea is to simplify and humanize the reporting process for the user by replacing overwhelming terminology with more accessible language as well as examples of abusive content to help the user categorize the post better. The list of types of abuse or harassment with examples would take into account the 'blocked tags' from the user's account customization and the type of content that they have chosen not to see or have reported in the past. The options could also be region specific and display categories and examples related to the user's geographic location and local language. At the end of the report, along with a message that notifies the user of their report being sent and an option to block the reported user, there would be a link to a feature that allows them to keep track of the progress of consequences on the accounts that they reported.

Instagram benefits

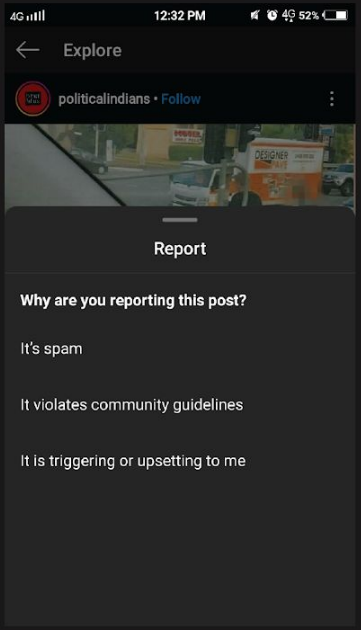
1. More engagement on the reporting system by making it more accessible

2. Lesser content that users dislike and therefore spend more time on the platform

3. Trust in the system with a better tone of voice

CR concepts

Social and Cultural Capital



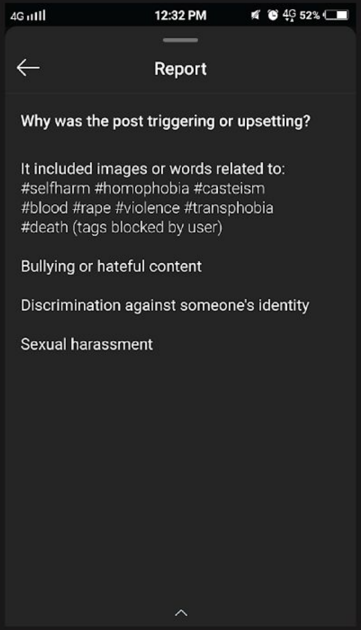
Report

Why are you reporting this post?

It's spam

It violates community guidelines

It is triggering or upsetting to me



Report

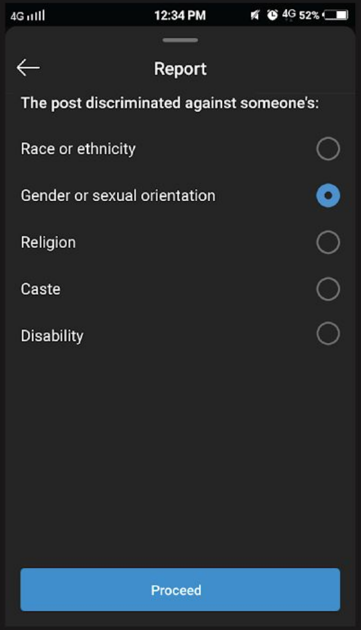
Why was the post triggering or upsetting?

It included images or words related to:
#selfharm #homophobia #casteism
#blood #rape #violence #transphobia
#death (tags blocked by user)

Bullying or hateful content

Discrimination against someone's identity

Sexual harassment



Report

The post discriminated against someone's:

Race or ethnicity ○

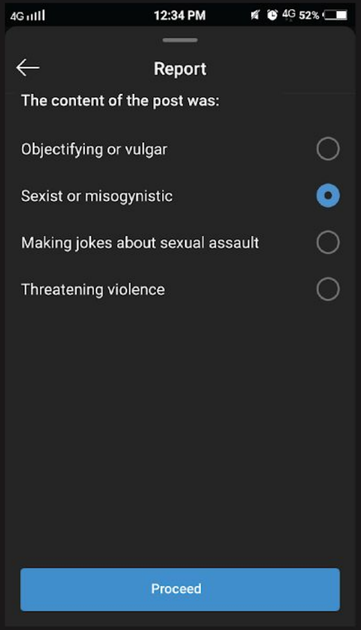
Gender or sexual orientation ●

Religion ○

Caste ○

Disability ○

Proceed



Report

The content of the post was:

Objectifying or vulgar ○

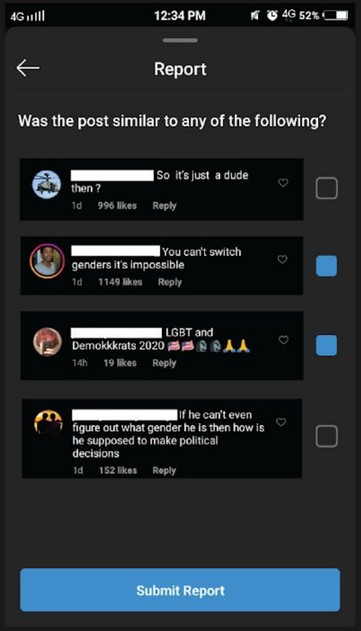
Sexist or misogynistic ●

Making jokes about sexual assault ○

Threatening violence ○

Proceed

An additional category to report beyond 'inappropriate content', to humanize the process.



Report

Was the post similar to any of the following?

[redacted] So it's just a dude then ?
1d 996 likes Reply ☐

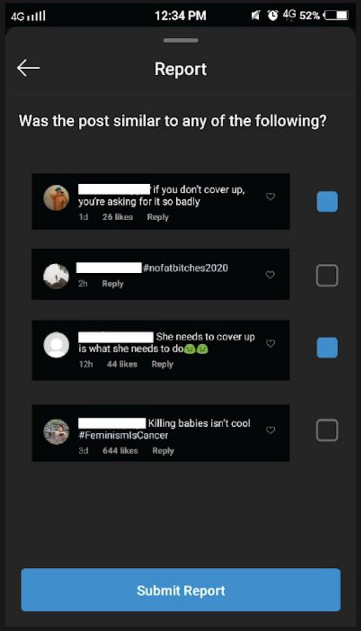
[redacted] You can't switch genders it's impossible
1d 1149 likes Reply ☑

[redacted] LGBT and Demokkkrats 2020 🇺🇸🇺🇸🗿🗿🙏🙏
14h 19 likes Reply ☑

[redacted] If he can't even figure out what gender he is then how is he supposed to make political decisions
1d 152 likes Reply ☐

Submit Report

Examples if selected 'Discrimination based on gender identity'



Report

Was the post similar to any of the following?

[redacted] if you don't cover up, you're asking for it so badly
1d 26 likes Reply ☑

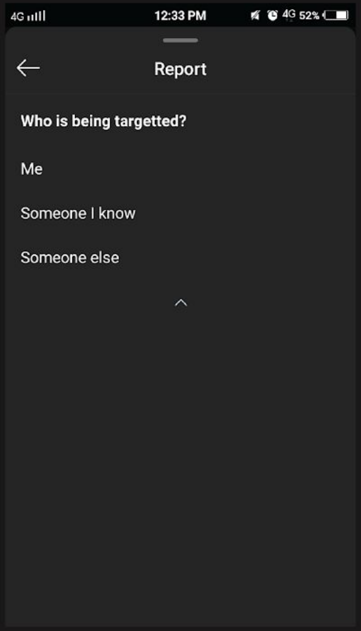
[redacted] #nofatbitches2020
2h Reply ☐

[redacted] She needs to cover up is what she needs to do🤢🤢
12h 44 likes Reply ☑

[redacted] Killing babies isn't cool #FeminismIsCancer
3d 644 likes Reply ☐

Submit Report

Examples if selected 'Sexist or misogynist' content



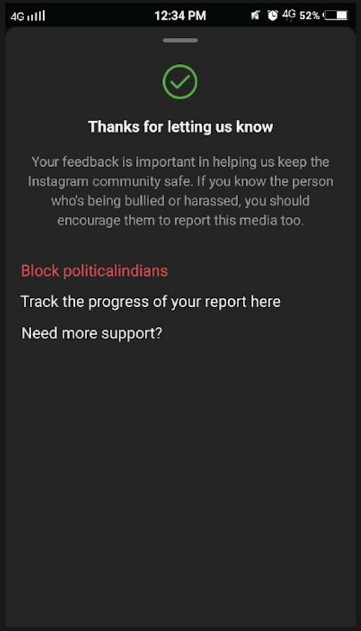
Report

Who is being targetted?

Me

Someone I know

Someone else



Thanks for letting us know

Your feedback is important in helping us keep the Instagram community safe. If you know the person who's being bullied or harassed, you should encourage them to report this media too.

Block politicalindians

Track the progress of your report here

Need more support?

Final screen.
Track progress will take the user to the reports history/activity feature on their profile.

Idea 5

Description

History of a user's abusive or offensive behavior can be transferred to other platforms or employers.

This feature came into being, to create an archive of reports. This allows the users to keep track of their reports and hold Instagram accountable in case the reported content doesn't check mark their guidelines, and have a constant scope for improvement. The other part of this archive is the history of the user's own reports. This space was created for an ethical background check for employees of a workspace. While this creates a set of real world consequences to their online behavior, it also tends to be a reality check. This archive is a space for self surveillance of the user's behavior online, to make them more conscious about their words and actions. Rewarding positive changes and conscious behavior online is equally important. Which is why curated highlights on user's profile is a small but meaningful addition - like being "report free". Instagram can create a joint database with other platforms as well, create guidelines for reporting content and have the power to remove hate online.

Instagram benefits

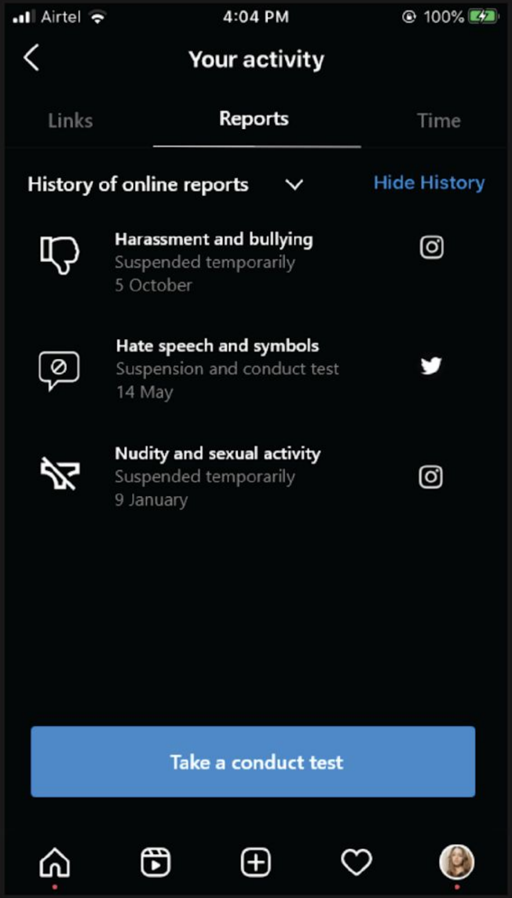
1. Joint database of reporting guidelines from all platforms can make the Ai better in recognizing hateful content

2. Increased screen time opportunity by adding a tracking status of the report

CR concepts

Cultural Capital

Panopticon :
Self Surveillance



Airtel 4:04 PM 100%

Your activity

Links | Reports | Time

History of online reports ⌄ Hide History

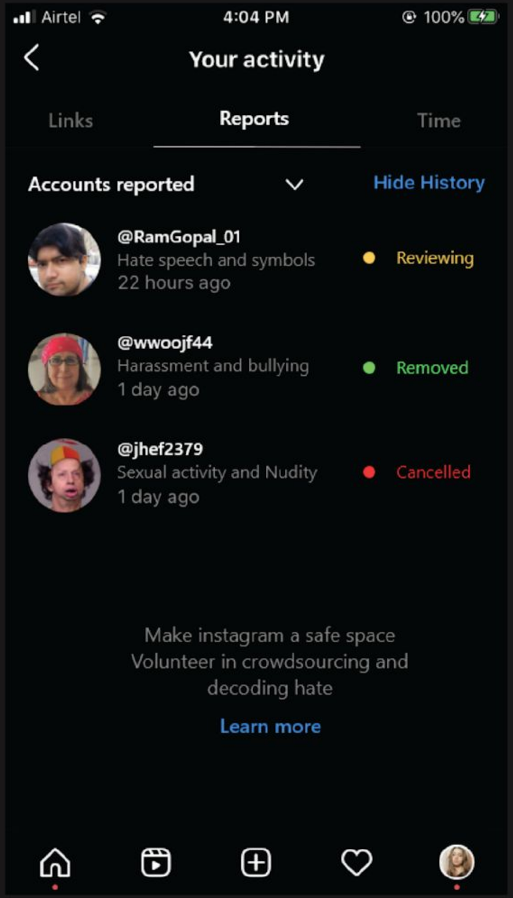
Harassment and bullying
Suspended temporarily
5 October

Hate speech and symbols
Suspension and conduct test
14 May

Nudity and sexual activity
Suspended temporarily
9 January

Take a conduct test

History of user's online reports
+ Option to redeem by taking
conduct test to qualify for it



Airtel 4:04 PM 100%

Your activity

Links | Reports | Time

Accounts reported ⌄ Hide History

@RamGopal_01
Hate speech and symbols
22 hours ago
● Reviewing

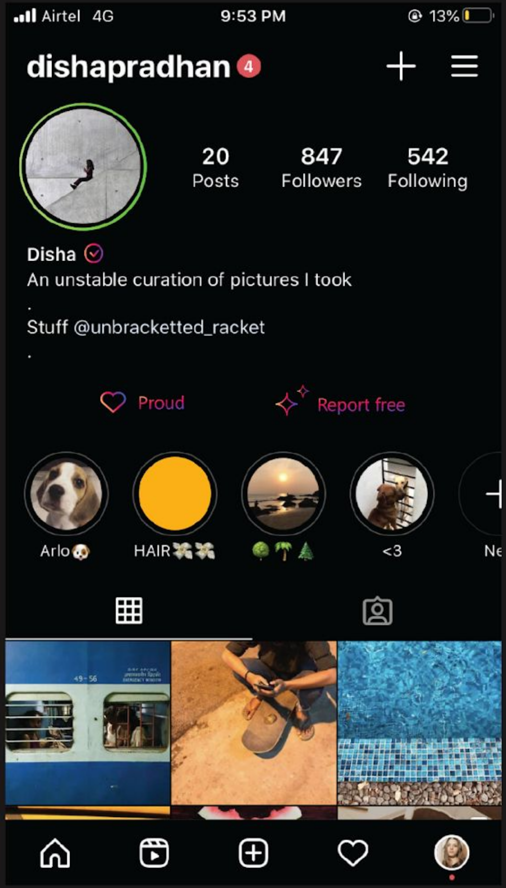
@wwoojf44
Harassment and bullying
1 day ago
● Removed

@jhef2379
Sexual activity and Nudity
1 day ago
● Cancelled

Make instagram a safe space
Volunteer in crowdsourcing and
decoding hate

Learn more

Tracking content reported



Airtel 4G 9:53 PM 13%

dishapradhan 4

20 Posts | 847 Followers | 542 Following

Disha
An unstable curation of pictures I took
.
Stuff @unbracketted_racket
.

Proud | Report free

Verify : Bumble approach to verify the
user is the same person they are
impersonating
+ Profile highlights

Adobe XD link :
https://xd.adobe.com/view/b681509e-4404-40e0-af77-0744ab9b7987-e3d2/?fullscreen

Thank You.