Abuse Standards 6.1

Operation Manual For Live Content Moderators

Major changes since AS 6.0

- Added sexual language and solicitation policy
- 3-criteria rule (one person in photo, one person tagged, and tagged person reporting):
  - *NOT* applicable to photos
  - Only applicable to PhotoComments
- Sex & Nudity issues clarified as follows:
  - Content uploaded for the purpose of sexual arousal is no longer a violation; only sex and nudity standards apply
  - Any kissing, groping, fondling – heterosexual and homosexual – is allowed; only sex and nudity standards apply
  - Cartoon bestiality/necrophilia/pedophilia can be confirmed – no need to escalate unless the act is being actively promoted or encouraged. Real life violations are still escalated.
  - Child nudity can be escalated if unsure of sexual context or age; includes older teenagers who may still be minors

- Graphic violence policies updated as follows:
  - No exceptions for news or awareness-related context for graphic image depiction; confirm all such content
  - Human/animal abuse subject to clear involvement/enjoyment/approval/encouragement *by the poster*
  - Even fake/digital images of graphic content should be confirmed, but hand-drawn/cartoon/art images are ok
- Hate speech: Humor and cartoon humor is an exception for hate speech *unless* –
  - Slur words are being used, or
  - Humor is not evident in the post/photo
- Attacks: Humor overrules racist and other attacks *unless* –
  - Slur words are being used, or
  - Humor is not evident in the post/photo
- Photoshopped images are always confirmed even if there's no obvious attack on the person
- International Compliance:
  - Anti-PKK and anti-Ocalan content should be UNCONFIRMED
  - ONLY confirm depiction if no other context or if context shows support
- PhotoComments:
  - ONLY review the comment that has been reported
  - The reported comment can be seen flagged in the comment trail on the details page
  - All other comments are irrelevant to the moderation process
  - All Abuse Standards policies are applicable

Sexually explicit language and sexual solicitation policy

A fresh policy has been formulated around content containing sexually explicit/descriptive language, as well as users soliciting (requesting or offering) sexual activity. The content **must be descriptive** in nature (not just the mere mention of sexual acts, genitals and other related words) in order to confirm the content.

Below is a full definition of the policy, relevant instructions to moderators and some examples of confirmed and unconfirmed content.

New definition:

Users may not describe sexual activity in writing, except when an attempt at humor or insult. This includes –

- Describing a state of sexual arousal, defined as hard nipples, wetness or erections
- Describing an act of sexual intercourse, defined as sexual penetration, self pleasuring, or exercising fetish scenarios
- Description is defined as adding detail and going beyond mere naming or mentioning

Implications on decision-making (CR Tool):

- Delete only if titles of objects match the new definition (Meaning, relevant content such as caption for photos, comments for posts, etc.)
- Ignore messages/posts that use sexual language to insult or make a joke
- Delete if user is soliciting (requesting or offering) sexual services

Examples:

- [Unconfirmed] I want to fuck you [fucking = sex. No details given]
- [Unconfirmed] Yeah I'd like to poke that bitch in the pussy [poke in the pussy = sex. No details given]
- [Unconfirmed] I've got a hard-on for you girl [hard-on = sexual arousal. No details given.]
- [Unconfirmed] I'm gonna eat that pussy all night [eat pussy = oral sex. No details given]
- [Unconfirmed] Hello ladies, wanna suck my cock? [suck cock = oral sex. No details given]
- [Unconfirmed] How about I come over and fuck you in the ass, girl [fuck in the ass = anal sex. No details given]
- [Unconfirmed] I have a big penis and I love girls touching it [I have a big penis is not sexual activity. Touching a penis = hand job. No details given]
- [Unconfirmed] Kelly loves to suck cock (not reported by Kelly) [suck cock = oral sex. Need Kelly to report for bullying]
- [Unconfirmed] Jones likes to take it up the ass (not reported by Jones) [take up the ass = anal sex. Need Jones to report for bullying]
- [Confirmed] Looking for girls who want to have fun. Inbox me for a good time. [sexual solicitation]
- [Confirmed] Ladies and girls, I need some pussy. Call me on 555 143 5746 [sexual solicitation]
- [Confirmed] For 9 inches of pure pleasure call 1800 DIK PUSS [sexual solicitation]
- [Confirmed] I love drinking fresh hot cum [details of oral sex: drinking cum]

UNCONFIRMED

Sex and Nudity:

Graphic Violence:

Attacks against protected categories:

Anti-Apo/PKK – International Compliance:

Caption: This was the original flag of the PKK, designed and formulated in 1978.

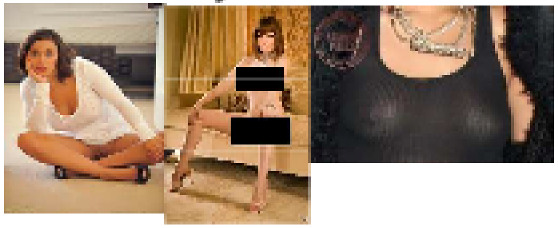
Animal (and human) abuse – no context or not showing pleasure/encouragement/approval:

Caption: Somebody has to report these assholes so they can get arrested.

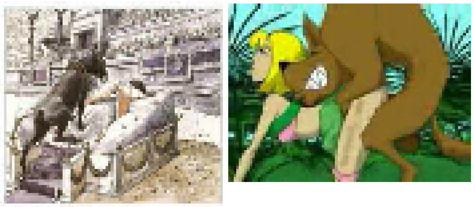
No Caption

CONFIRMED

Sex and Nudity:



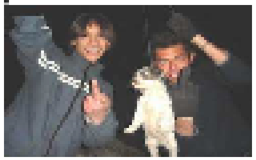
Cartoon Bestiality:



Graphic Violence – including fake images:

Any kind of Photoshop image:

Animal (and human) abuse – showing pleasure/encouragement/approval:



Caption: Wow, what a cool way to torture a puppy!

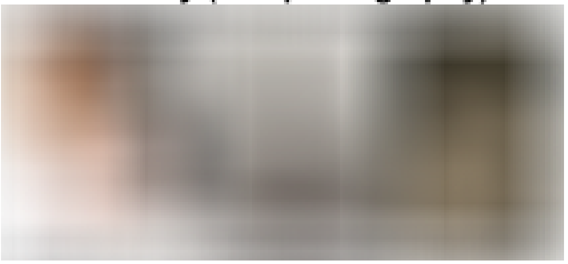
Attacks against groups – slur words or humor not obvious:

PKK (and Apo Ocalan) – either showing support, or no context:

Caption: Kurds all over the world must stand together against Turkish tyranny

ESCALATED

Child nudity (and pornography):



Bestiality – promoting/encouraging:

Caption: I love the ancient people; we should start fucking animals too!

Types of reported content - 1

Photos

Photographic or other visual content with accompanying caption

- **Relevant content**: Must check…
  - □ Photo (for visual violations),
  - □ text on photo (for name match and other violations),
  - □ caption (for name match and other violations) and
  - □ 'reported by' (as basis for name match)
  - □ Note: Mentions, or people whose names are mentioned and tagged within the caption, are considered valid name matches with the reporter because the text will still be visible even after the tag is removed. Please confirm if such mentions are seen.
- **Irrelevant content**: Must ignore…
  - □ Comments – No need to check for name matches in comments thread, and no need to look for other violations
  - □ Photo tags (these tags can be removed by the reporter so no need to check here for name matches). However, see note above on 'mentions'.
  - □ 'Posted by', also called 'poster'

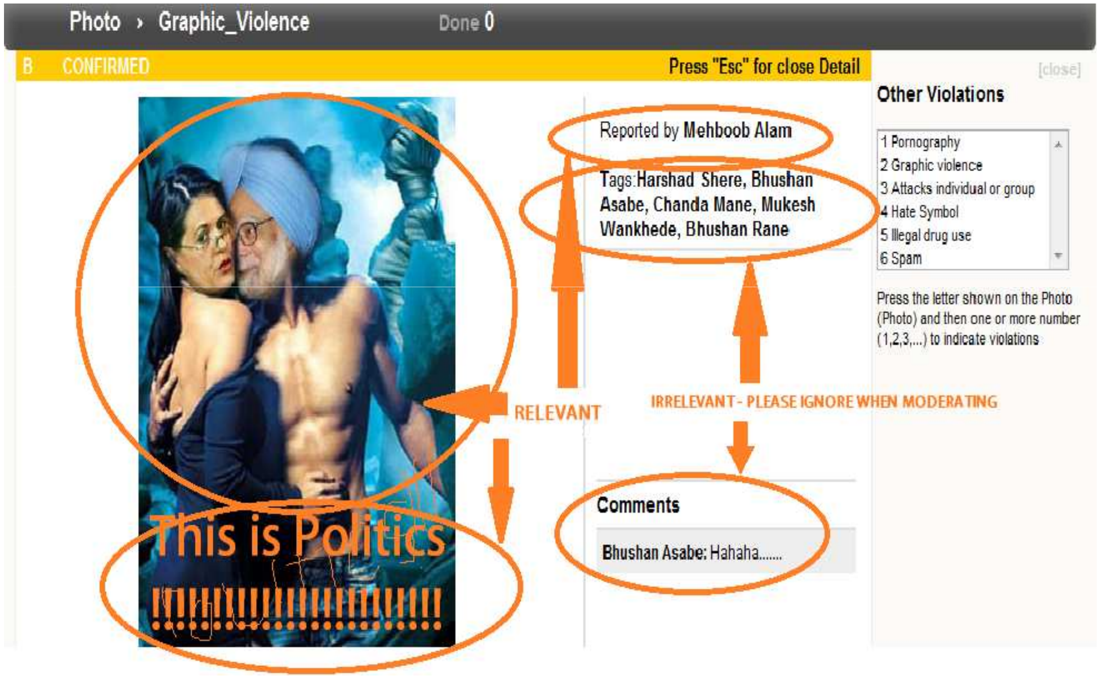


Photo › Graphic_Violence Done 0

B CONFIRMED Press "Esc" for close Detail [close]

Reported by Mehboob Alam

Tags:Harshad Shere, Bhushan Asabe, Chanda Mane, Mukesh Wankhede, Bhushan Rane

RELEVANT

IRRELEVANT - PLEASE IGNORE WHEN MODERATING

Comments

Bhushan Asabe: Hahaha.......

Other Violations

1 Pornography
2 Graphic violence
3 Attacks individual or group
4 Hate Symbol
5 Illegal drug use
6 Spam

Press the letter shown on the Photo (Photo) and then one or more number (1,2,3,...) to indicate violations

Summary: When moderating this type of content, only review the relevant content for policy violations (Abuse standards, name matches, hate-related violations, etc.)

Types of reported content - 2

Videos
Uploaded videos with accompanying caption

- **Relevant content**: Must check…
  - □ Video (for visual AND audio violations),
  - □ Audio (for name match and other violations) ,
  - □ text within video (for name match and other violations),
  - □ caption (for name match and other violations) and
  - □ 'reported by' (as basis for name match)
  - □ Note: Mentions, or people whose names are mentioned and tagged within the caption, are considered valid name matches with the reporter because the text will still be visible even after the tag is removed. Please confirm if such mentions are seen.

- **Irrelevant content**: Must ignore…
  - □ Comments – No need to check for name matches in comments, and no need to look for other violations
  - □ Video Tags, if any (these tags can be removed by the reporter so no need to check here for name matches). However, see note above on 'mentions'.
  - □ 'Posted by', also called 'poster'

- **Summary**: When moderating this type of content, only review the relevant content for policy violations (Abuse Standards policies, name matches, hate-related violations, etc.)

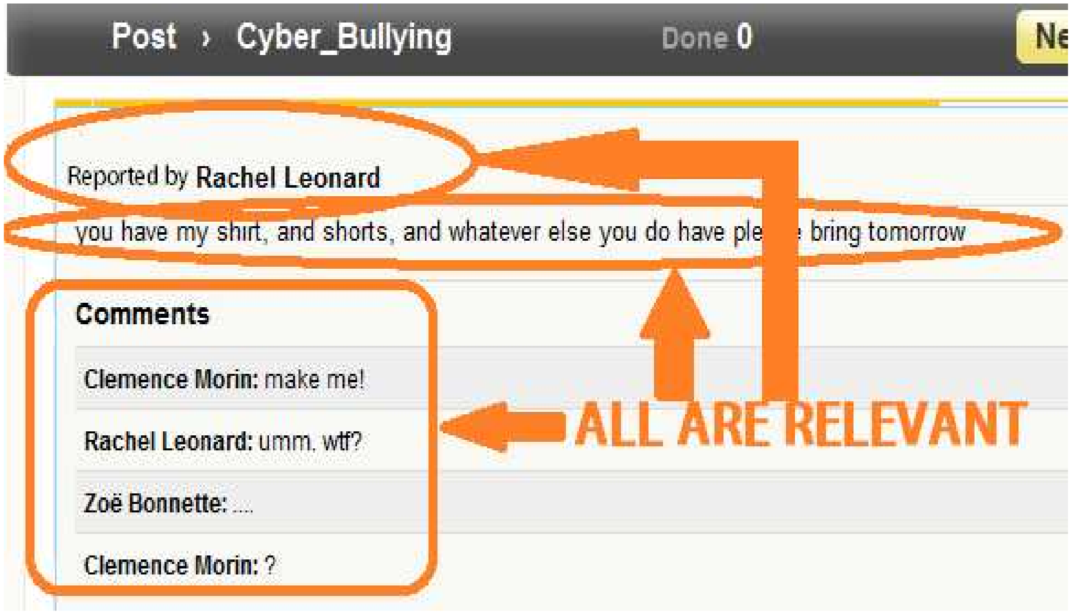
Types of reported content - 3

Posts

All types of posts that have been reported by the user (the reporter)

Types of Posts

- □ Group Posts
- □ Status messages
- □ Topic Posts
- □ Wall Posts
- □ Comments



Post › Cyber_Bullying Done 0 Ne

Reported by Rachel Leonard

you have my shirt, and shorts, and whatever else you do have ple bring tomorrow

Comments

Clemence Morin: make me!

Rachel Leonard: umm, wtf?

Zoë Bonnette: ....

Clemence Morin: ?

- <u>Relevant content</u>: Must check…
  - □ Actual Post/Reported Comment (for name match and other violations),
  - □ 'reported by' (as basis for name match)
- <u>Irrelevant content</u>: Must ignore…
  - □ 'Posted by', also called 'poster'
  - □ For comments: All comments other than the reported one

Summary: When moderating this type of content, only review the relevant content for policy violations (Abuse Standards policies, name matches, hate-related violations, etc.)

Types of reported content - 4

PhotoComments

Comments posted against another user's photo content

Note: In this category, one of the comments about the photo has been reported by the user, NOT the photo itself.

Relevant content: Must check…

- Reported Comment (for name match and other violations) and
- 'reported by' (as basis for name match – example given)
- Photo and Tags – as per new policy

Irrelevant content: Must ignore…

- 'Posted by', also called 'poster'

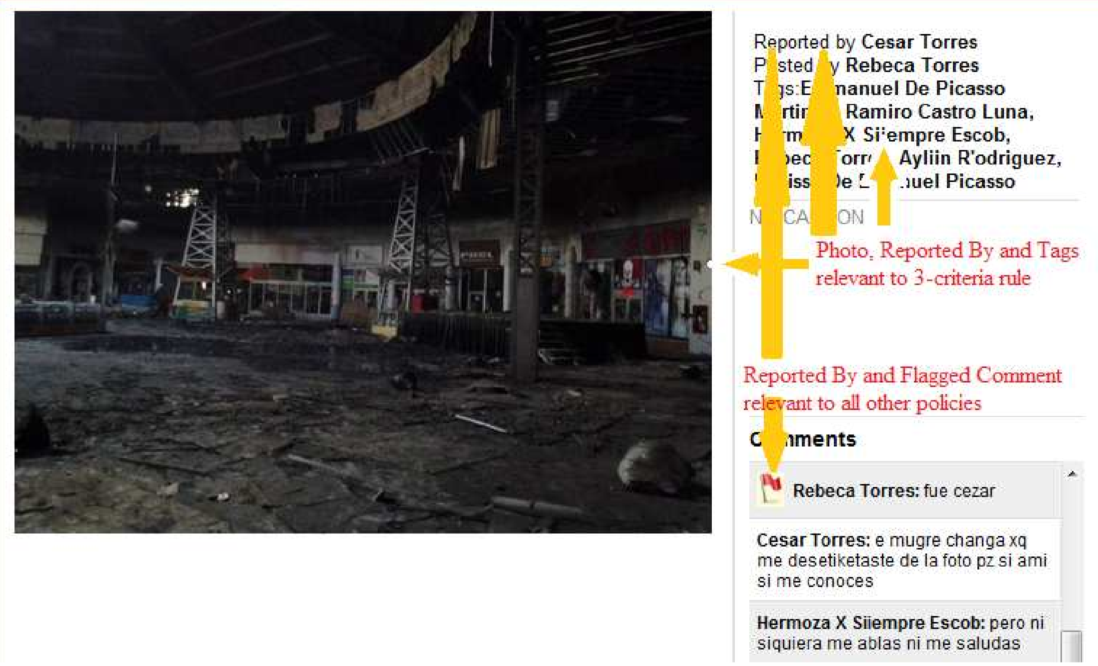
Summary: When moderating this type of content, only review the relevant content for policy violations (Abuse Standards policies, name matches, hate-related violations, etc.)

New

PhotoComments 3-criteria rule:

Confirm the photo comment if all 3 criteria (as below) about the parent content (the photo itself) are met:

1. Only one person is displayed in the photo
2. Only one user is tagged
3. The tagged user is reporting the photocomment



Reported by **Cesar Torres**
P sted y **Rebeca Torres**
T gs:E **nanuel De Picasso**
M rtir Ramiro Castro Luna,
H m X Si'empre Escob,
I bec orr Ayliin R'odriguez,
iss De uel Picasso

N CA ON

Photo, Reported By and Tags relevant to 3-criteria rule

Reported By and Flagged Comment relevant to all other policies

C mments

Rebeca Torres: fue cezar

Cesar Torres: e mugre changa xq me desetiketaste de la foto pz si ami si me conoces

Hermoza X Siiempre Escob: pero ni siquiera me ablas ni me saludas

Hate Content

	ORDINARY PERSON	PUBLIC FIGURE	LAW ENFORCEMENT OFFICER (LEO)	HEAD OF STATE (HOS)
EMPTY THREATS	Hate Content policies apply	UNCONFIRMED	CONFIRMED	CONFIRMED
CREDIBLE THREATS	<<============ESCALATE ALL===========>>			
REFERENCED NEGATIVELY	Hate Content policies apply	UNCONFIRMED	CONFIRMED	UNCONFIRMED
CYBERBULLYING	Hate Content policies apply	UNCONFIRMED	CONFIRMED	UNCONFIRMED
ATTACKED WITH HATE SYMBOLS	CONFIRMED	UNCONFIRMED	CONFIRMED	UNCONFIRMED
ATTACKED BASED ON THEIR BEING A SEXUAL ASSAULT VICTIM	CONFIRMED	CONFIRMED	CONFIRMED	CONFIRMED

Protected Categories (see table below)

Users may NOT create content that degrades individuals based on the below protected categories. Note: Protected category status should override the person's status as Public Figure or HOS. Example, "I hate Obama" is unconfirmed, while "Can't stand that nigger Obama" is confirmed.

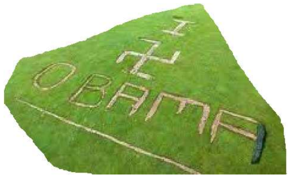
The table on the left shows what types of visual and verbal attacks can and cannot be used against various types of people:

Definitions:

- **Ordinary person** – non-public figures who aren't 'famous'
- **Public figure** – any person featured in any mass medium (internet, news, etc.)
- **LEO** – Any person belonging to a law enforcement agency such as the police, drug enforcement agencies, etc. Does not apply to military personnel
- **HOS** – Any person who is currently the head of a ruling political entity in a country

Examples of types of attacks

- **Empty threat** – Not Applicable; refer to Hate Content p11
- **Credible threat** – "Let's cut Jason's [target] throat [method] when he comes back home tonight [time]. 2 out of 3 sufficient to escalate.
- **Referencing negatively** – Not Applicable; refer to Hate Content p11
- **Cyber bullying** – Not applicable; refer to Hate Content p11
- **Attacking with hate symbols**



- **Attacking based on being a sexual assault victim** – "Janie says she was raped when she was 12. I think she deserved it…she must have been asking for it."

Race	Ethnicity	National Origin	Religion	Sex	Gender Identity	Sexual Orientation	Disability	Serious Disease
White, Black, Hispanic, Asian	American Indians, Aborigines	Americans, British, French, Chinese	Christians, Muslims, Buddhists, Hindu, Wiccans	Male, Female	Heterosexual, Bisexual, Homosexual, Asexual	Lesbian, Gay, Bisexual, Transgender	Physical, Sensory, Intellectual, Mental, Developmental	Any life threatening disease

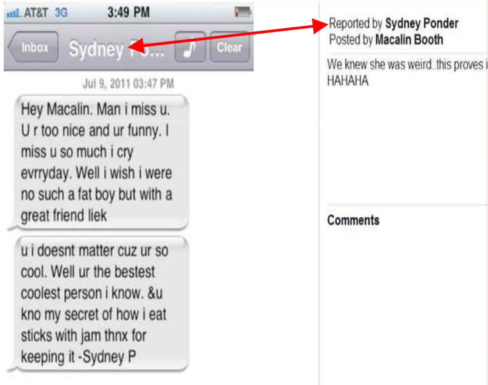
Name Match Policy

- The Name Match Policy is intended to protect the identity and privacy of individuals who are named in photos, videos and posts.
- The name match should be between the reporter (‘reported by’) and the actual content that was reported (photo+caption, video+caption or post+comments).
- Ignore all irrelevant content when checking for name matches
- Possible name matches: First name only, last name only, full name and fuzzy name match
- Fuzzy name match could be in the form of initials, nicknames, etc.
- Criterion – **IF THERE’S A POSSIBILITY THAT THE REFERENCED PERSON IS THE REPORTER, THEN CONFIRM**

Some examples

Confirmed

Name match valid:
Decision = Confirmed



Jul 9, 2011 03:47 PM

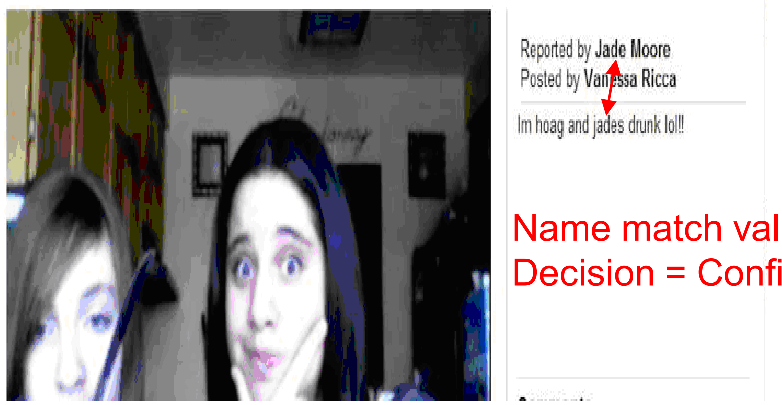
Hey Macalin. Man i miss u. U r too nice and ur funny. I miss u so much i cry evrryday. Well i wish i were no such a fat boy but with a great friend liek

u i doesnt matter cuz ur so cool. Well ur the bestest coolest person i know. &u kno my secret of how i eat sticks with jam thnx for keeping it -Sydney P

Reported by **Sydney Ponder**
Posted by **Macalin Booth**

We knew she was weird..this proves i
HAHAHA

Comments



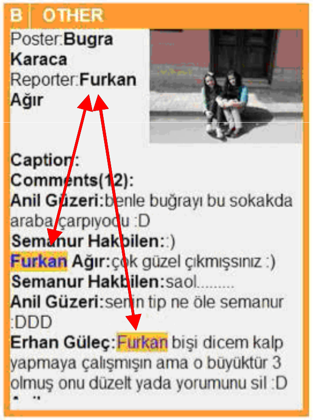
Reported by Jade Moore
Posted by Vanessa Ricca

Im hoag and jades drunk lol!!

Name match valid:
Decision = Confirmed

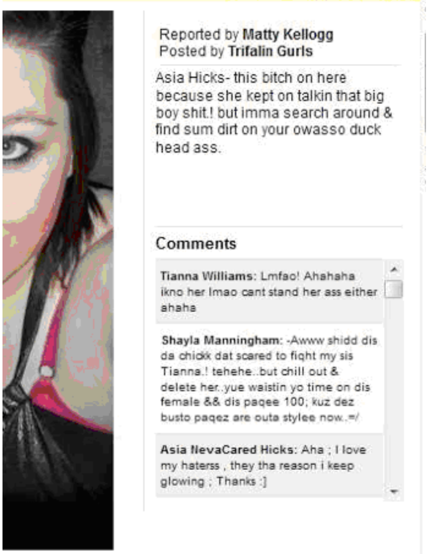
Unconfirmed

Name match not valid:
Decision = Unconfirmed



Poster:**Bugra Karaca**
Reporter:**Furkan Ağır**

Caption:
Comments(12):
Anil Güzeri:benle buğrayı bu sokakda araba çarpıyodu :D
Semanur Hakbilen::)
Furkan Ağır:çok güzel çıkmışsınız :)
Semanur Hakbilen:saol........
Anil Güzeri:senin tip ne öle semanur :DDD
Erhan Güleç: Furkan bişi dicem kalp yapmaya çalışmışın ama o büyüktür 3 olmuş onu düzelt yada yorumunu sil :D



Reported by **Matty Kellogg**
Posted by **Trifalin Gurls**

Asia Hicks- this bitch on here because she kept on talkin that big boy shit.! but imma search around & find sum dirt on your owasso duck head ass.

Comments

Tianna Williams: Lmfao! Ahahaha ikno her lmao cant stand her ass either ahaha

Shayla Manningham: -Awww shidd dis da chickk dat scared to fiqht my sis Tianna.! tehehe..but chill out & delete her..yue waistin yo time on dis female && dis paqee 100; kuz dez busto paqez are outa stylee now..=/

Asia NevaCared Hicks: Aha ; I love my haterss , they tha reason i keep glowing ; Thanks :]

Clear target but no name match:
Decision = Unconfirmed

See exception for photocomments in slide 5

How to assess credibility

Credibility Test: Consequence, Specificity and Practicability

Question 1 (Consequence): Is the proposed violence

(a) targeted at a head of state/law enforcement officer,
(b) terrorism or
(c) organized crime?

Answer "Yes" to any of the above - Escalate as Egregious
Answer "No" - Move to Question 2.

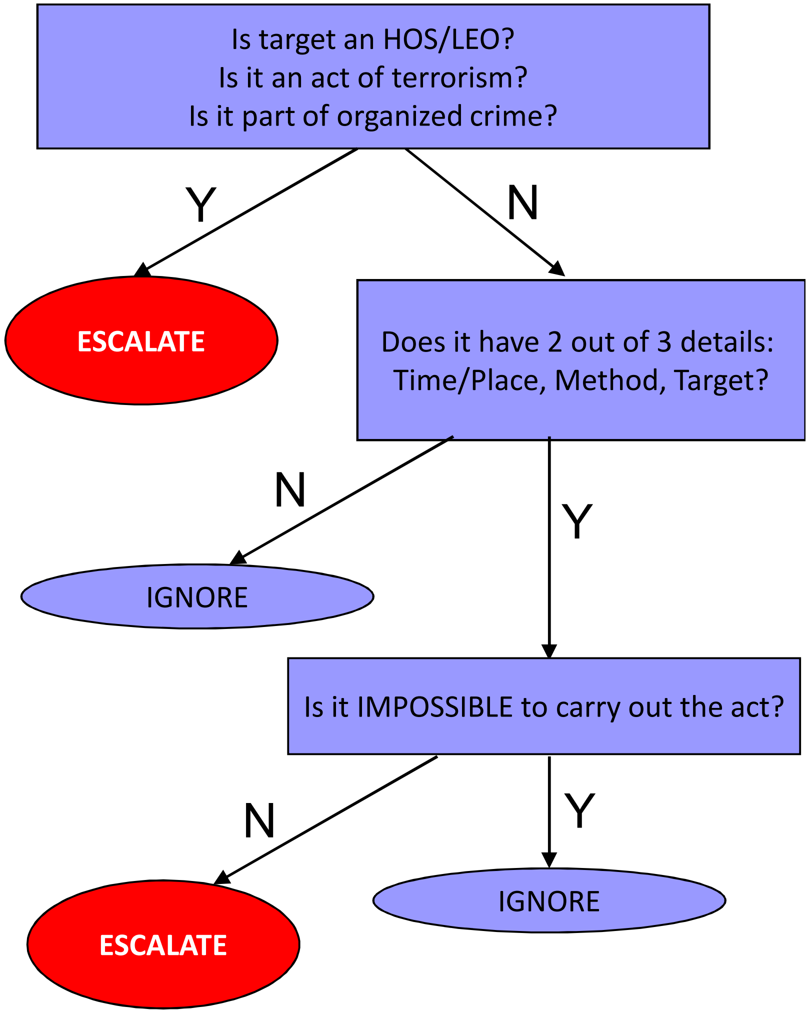
Question 2 (Specificity): Does the content specify 2 out of these 3 details: Time/Place, Method or Target?

Answer "Yes" - Move to Question 3
Answer "No" - Ignore content.

Question 3 (Practicability): Is it it clearly impossible for the individuals proposing violence to carry it out?

Answer "Yes" - Ignore content.
Answer "No" - Escalate as Egregious
Answer "I don't know" - Escalate as Egregious

Assessment Flowchart


Is target an HOS/LEO?
Is it an act of terrorism?
Is it part of organized crime?
Y
N
ESCALATE
Does it have 2 out of 3 details: Time/Place, Method, Target?
N
Y
IGNORE
Is it IMPOSSIBLE to carry out the act?
N
Y
ESCALATE
IGNORE

The target needs to be clearly identifiable; examples below -

"I’m going to stab Lisa H." (Escalate as Egregious)

“I’m going to stab the idiots in France" (Do not escalate)

Abuse Standards Violations

All the items below should be confirmed; anything not on this list can be unconfirmed

Sex and Nudity

1. Any OBVIOUS sexual activity, even if naked parts are hidden from view by hands, clothes or other objects. Cartoons/art included. Foreplay allowed (Kissing, groping, etc.). even for same sex (man-man/woman-woman
2. Naked 'private parts' including female nipple bulges and naked butt cracks; male nipples are ok.
3. Pixelated or black-barred content showing nudity or sexual activity as above.
4. Naked children, including cartoon versions (able to stand on their own) and older minors - Escalate if unsure of sexual context (child porn)
5. Depiction of sexual assault or rape in any form.
6. Mothers breastfeeding without clothes on.
7. Escalate bestiality, necrophilia, and pedophilia. Confirm cartoon/digital versions BUT escalate if content is promoting.
8. Digital/cartoon nudity. Art nudity ok.
9. People "using the bathroom".
10. Blatant (obvious) depiction of camel toes and moose knuckles.
11. Sex toys or other objects, but only in the context of sexual activity.
12. Depicting sexual fetishes in any form.

Illegal Drug Use

1. Unconfirm all marijuana unless context is clear that the poster is selling.buying/growing.
2. Illegal drugs shown NOT in the context of medical, academic or scientific study.

- **Note**: Any depiction of marijuana alone (any amount) or implements used for smoking marijuana are ok (unconfirm)

Theft Vandalism and Fraud

1. Praising or displaying crimes that they or their friends committed
2. Organizing criminal activity or soliciting illegal services.
3. Encouraging others to engage in criminal activity.
4. Escalate based on credibility assessment

Hate Content
(Valid Name Match not required)

1. Slurs or racial comments of any kind
2. Attacking based on protected category
3. Hate symbols, either out of context or in the context of hate phrases or support of hate groups.
4. Showing support for organizations and people primarily known for violence.
5. Depicting symbols primarily known for hate and violence, unless comments are clearly against them.
6. "Versus photos" or "Vs photos": photos comparing two people side by side.
7. Any photoshopped images of people, whether negative, positive or neutral
8. Images of drunk and unconscious people, or sleeping people with things drawn on their faces.
9. Videos: Street/bar/schoolyard fights even if no valid name match is found. School fight videos are only confirmed if the video has been posted to continue tormenting the person targeted in the video.

Notes:

- Hate symbols are confirmed if there's no context OR if hate phrases are used
- Humor overrules hate speech UNLESS slur words are present or the humor is not evident.

Graphic Content

1. Content showing Poster's delight in/involvement in/promoting of/encouraging of violence against humans or animals for sadistic purposes (e.g. torture, staged animal fights, animal starvation, obvious neglect, etc.)
2. Depicting the mutilation of people or animals, or decapitated, dismembered, charred, or burning humans.
3. Poaching of animals should be confirmed. Poaching of endangered animals should be escalated
4. Urine, feces, vomit, semen, pus, and ear wax. (Cartoon feces, urine and spit are OK; real and cartoon snot is OK)
5. Violent speech (Example: "I love hearing skulls crack")
6. Photos and digital images showing internal organs, bone, muscle, tendons, etc. Deep flesh wounds are ok to show; excessive blood is ok to show.
7. Crushed heads, limbs, etc are ok as long as no insides are showing

- Note: No exception for news or awareness related content.

IP Blocks and International Compliance

Escalated:

1. Holocaust denial which focuses on hate speech
2. All attacks on Ataturk (visual and text)
3. Maps of Kurdistan (Turkey)
4. Burning Turkish flag(s)

Confirmed (unless clearly against PKK and/or Ocalan):

1. PKK support and depiction
2. Abdullah "Apo" Ocalan-related content

Self Harm

- **Note**: All self harm content should be escalated.

1. Threat and serious promotion of suicide.
2. Supporting people, groups, and symbols that advocates and promoting eating disorders as a lifestyle choice.
3. Depicting self-mutilation and groups and people that promote and support it (ex: cutting groups)

Bullying and Harassment

1. Valid name matches no matter what the content is (negative, positive or neutral)
2. Contacting other users persistently without prior solicitation or continue to do so when the other party has said that they want no other further contact with the sender.
3. Attacking anyone based on their status as a sexual assault or rape victim.

Credible Threats
(Escalate as per credibility assessment)

1. Credible threats or incitement of physical harm against anyone
2. Credible indications of organizing acts of present or future violence
3. Any threats of violence against Heads of State (HOS) or Law Enforcement Officers (LEO) should always be escalated even if not credible
4. Any credible indication of terrorist activity or organized past/future crime.

Burden of Clarity

The concept of ‘burden of clarity on user’:

There are certain types of content where the mere depiction of an image is considered a violation, unless the caption (or other relevant content) suggests that the user is not promoting, encouraging or glorifying the act.

For such content, as below, the user must make it clear that the posted photo is not a violation, either through the caption or through text on the photo.

- **Hate Symbols** – includes Swastika and other acknowledged hate symbols; also includes pictures of Hitler, Bin Laden and others associated primarily with hate and violence.

- **Graphic Violence** – includes highly graphic images of violence towards humans and animals

- **Self Harm** – includes suicide, eating disorders and self-mutilation (exception is body art, piercings, tattoos, etc.).

- **Illegal Drugs** – Depiction of drugs (other than marijuana or its derivatives – hash, hash oil, etc.) which are typically abused, and implements depicting drug abuse

Content That Should Be Escalated

International Compliance/IP Blocks:

- Photos AND/OR text making fun of/attacking/depicting negatively/criticizing, Ataturk.
- Burning the Turkish flag [other flags are ok to be shown burning]
- Maps of Kurdistan [as of now, only maps are escalated; other references are merely confirmed]
- Holocaust denial [any discussion of holocaust denial should be escalated]

Egregious:

- Child pornography/Pedophilia [ESCALATE IF UNSURE OF SEXUAL CONTEXT OR AGE]
- Threats of school violence, credible or otherwise
- Necrophilia and bestiality
- Credible threats and indications -
    - Credible Threats against ordinary people or public figures that qualify (2 out of 3 – Time/Place, Target, Method)
    - Credible threats against Law Enforcement Officers (LEO)
    - Any threat (credible or not) against Heads of State (HOS)
    - Credible indications of past/future crime and organized crime
    - Any indication of terrorist activity
- Poaching of endangered species [always check http://www.iucnredlist.org/apps/redlist/search before escalating]

Sensitive:

- Credible depictions of suicide and/or promotion thereof.
- Self Harm photos
- Eating disorders (ana – anorexia, mia – bulimia; generically known as ana-mia)

Note: There is currently only one escalation option in the tool; it can be used for all 3 above types

DEALING WITH SPAM

- At the moment, we ignore all spam posts and photos unless they violate other Abuse Standards like nudity, graphic violence, etc. Only clear and obvious visual and verbal violations visible in the post should be confirmed/escalated as appropriate
- Ignore porn links unless the url is sexually descriptive or pornographic thumbnails can be seen.

NOTE:

NO ITEM SHOULD BE MARKED AS CONFIRMED IF IT COMES IN THROUGH THE SPAM QUEUE/FILTER. If it violates other standards, confirm as appropriate under “Confirmed - Other”.

Glossary of important terms

- **Reported Content (or simply, Content)**: A user-generated post on the social media website that we moderate for. Can be a photo+caption, a post+caption or video+caption. Refers to content that has been reported as having a particular violation
- **Caption**: The text accompanying a photo or video post. For posts, this is the original post on which comments are later made.
- **Comments**: Text uploaded by users in response to a posted piece of content. Usually found as a conversation thread with one or more users commenting on the original post or any other subject.
- **Photo tags**: These are people who have been tagged to a particular piece of content. Considered as irrelevant content when moderating because users can remove tags of themselves.
- **Mentions**: These are user's names being mentioned and possibly tagged within the text, such as caption text. These are considered valid name matches (relevant content) when the name has a match with that of the reporter.
- **Poster**: The user who originally posted the content. Represented by 'posted by' as seen on the moderation tool. Considered as irrelevant content when moderating.
- **Reporter**: The user who reported a piece of content. Represented by 'reported by' as seen on the moderation tool. Considered as relevant content when moderating because it is the basis for the name match policy.
- **Owner**: In PhotoComments, the user who owns the original photo that is being commented on.

- **Violation**: A piece of content that violates the terms of use of the user's social media account.
- **Non-Violation**: A piece of content that has been reported for review but does not violate any of the terms of use of the user's social media account.
- **Relevant Content**: Elements of a piece of content that should be reviewed when making a decision.
- **Irrelevant Content**: Elements of a piece of content that should be ignored when making a decision.
- **Confirmed**: With respect to the moderation tool, a decision which implies that there is a violation on a piece of content, as reported by the user.
- **Unconfirmed**: With respect to the moderation tool, a decision which implies that there is no violation on a piece of content, as reported by the user.
- **Confirmed – Other**: With respect to the moderation tool, a confirmed decision which violates the terms of use, but does not violate the abuse standard for which it has been reported. Example, nudity in a piece of photo content that has been reported for hate speech.
- **Escalate** – A decision taken on a piece of content that sends it to the social media client's internal review team for further action.

Thanks for reviewing this document carefully. If you have any questions or need to clarify any points, please take it up with your team leader in an email (using the oDesk messaging system to ensure security of the information) with a copy to the Certification and Training Manager:

shudeepc@odesk.com

Note to Team Leads:

Please copy the above ID in your clarification response to your team member(s). This is to ensure integrity of information and uniformity of process knowledge.