

Social Media **vs.** Cyberbullying

Comparative Study of YouTube, Facebook,
and Twitter's Responses towards Cyberbullying



**Authors**

Janitra Haryanto

Benyamin Imanuel Silalahi

Editor

Treviliana Eka Putri

Designer and Layouter

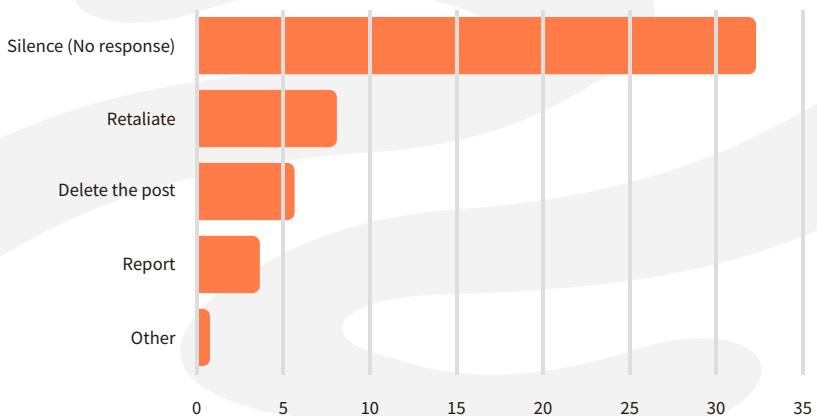
Riawan Hanif Alifadecya

Introduction ✓



The role of social media has significantly developed within the life of Indonesia's society. According to research by Wearesocial Hootsuite in 2019 (as quoted by Katadata,) the total number of active social media users in Indonesia has now reached 150 million users¹. Unfortunately, social media is currently evolving into a platform for cyberbullying. According to a survey from the Indonesian Internet Service Provider Association (APJII) (summarized by Katadata) in 2018, as much as 49% of social media users in Indonesia have experienced cyberbullying². Besides this fact, social media users in Indonesia have not yet developed the awareness to report the cyberbullying they experienced. According to the same survey by APJII (summarized by Katadata), only 3.6% of Indonesian social media users have reported their experience of cyberbullying to the authorities. The majority of Indonesian social media users (37.5%) choose to let the action go unreported³.

How People Respond the act of Cyberbullying in Social Media (%)



Picture 1. Social Media Users' Responses towards Cyberbullying based on a survey conducted by APJII, summarized by Katadata⁴

Social media has given the space for bullying act to shift from the real world into the online one. The benefits of anonymity and catering to virtual expression provided by social media (such as like or dislike expression) has intensified the complexity of cyberbullying. However, the condition of cyberbullying cannot be considered as ideal. This condition is being illustrated by the numerous accounts of cyberbullying experienced by Indonesian social media users, and also by the passive response they have given to their experience. By weighing this fact, social media corporations have the responsibility to ensure the reduction of cyberbullying cases on their platform.

The authors tried to answer two main points: (1) how social media platforms respond to cyberbullying and (2) comparison of these social media's responses. Knowledge and comparison of how these social media responses to cyberbullying act in their platform is essential to understand how safe we are from being the next victim of bullying in social media. The authors had limited the type of social media as the research objects into three social media: YouTube, Facebook, and Twitter. The decision to choose these platforms is based on the research by Wearesocial Hootsuite (as cited by Kompas), where these three social media ranked among the top ten social media platforms frequently used in Indonesia⁵. The authors used the desk research method by cultivating secondary data which is accessible online.



Cyberbullying: What, Why, and How



• What is Cyberbullying?

By definition, cyberbullying is an act of bullying, which is understood as an act of harassing or harming other people within cyberspace. This act is done continuously through electronic media, such as computers or mobile phones⁶. Cyberbullying utilized the information and communication technology to consciously or intentionally to conduct an act of bullying targeted towards specific someone by exploiting the unequal power relations⁷.

Different from conventional bullying, the practice of cyberbullying can be done anonymously. This condition increased the complexity of disturbance felt by the victim of cyberbullying. Besides, cyberbullying can happen anywhere, whenever, and to whomever. Whereas conventional bullying made the victim feel insecure when they go to a specific place during a particular time, cyberbullying, to some extent, can make the victim feel insecure almost all the time⁸.

During the first invention of the term, cyberbullying is assumed to manifest in four platforms: the chat room, short messages, e-mail, or text messages⁹. However, the existence of social media as this extended version of discussion or chatting space and accessible to a wider audience has increased the chance of cyberbullying cases to happen. Cyberbullying done through social media have a larger impact since social media, besides providing people room for discussion and increased the opportunity for more people to join in the act, is also complementing the perpetrator with other tools to conduct the cyberbullying act such as images, video, text in the timelines or the comment section.

One of the most severe impacts of cyberbullying is that the harassing contents reach a wider audience and may stay within cyberspace for a long period¹⁰. This led to many people witnessing the cyberbullying act itself, causing the negative consequences felt by the victim to expand. Moreover, cyberbullying brings negative impacts on the physical, psychological, and social health of both the victim and the perpetrator¹¹. Cyberbullying also has psychological consequences for both the victim and the perpetrator¹². The perpetrator will suffer the risk of depression, anxiety, eating disorder, and substance abuse. At the same time, the victim may feel degrading of confidence, a higher level of depression, and emotional disorder that is often related to a behavioral disorder or suicide attempt¹³. Thus, the psychological impact of cyberbullying is harsher compared to conventional bullying¹⁴.

● Why Does Cyberbullying Happen?

One of the methods to explain why cyberbullying happens is to trace the motivations of the perpetrators. A mental health research titled “High School Students' Perceptions of Motivations for Cyberbullying: An Exploratory Study”¹⁵ ” tried to investigate and map the motivations of high school students who committed the act of cyberbullying. This motivation is divided into two: internal and external motivations. The internal motivation is affected by the perpetrators' emotional condition, driven by feelings of detest, want for revenge, to create an air of superiority, boredom, pressures, protectiveness, jealousy, seeking for justification, trying out new personality, and anonymity. While the external motivations are usually affected by the perpetrators' characters or the circumstances itself, such as the lower level of consequences of cyberbullying, since they can escape the face-to-face situation and the feeling that the victim is different. Other motivations that may not occur in the general cases are homophobia and racism.

Several researches pointed to anonymity as the main reason why cyberbullying becomes rampant. Anonymity makes the perpetrators lose control of themselves, in the sense that they are suddenly feeling capable of doing things that previously are unthinkable, especially actions that they would have never done directly in real life¹⁶. With this being said, The physical and emotional detachment from their victims have made the perpetrators feel that they can easily escape from the consequences of their action¹⁷.



Besides the motivational factors, the advancement of information and communication technology and the presence of social media itself can also be accounted as the reason why cyberbullying happens. The internet has endowed us with access, facility, and the opportunity to allow many shapes of cyberbullying to manifest themselves. The exposure of social media users of these cyberbullying contents has made them more aware of the various methods of cyberbullying and then imitating those methods.

• How Does Cyberbullying Being Regularly Carried Out

There are diverse methods to practice cyberbullying. Research about cyberbullying titled "Cyberbullying Experiences Survey (CES)"¹⁸ interviewed cyberbullying victims about the types and methods of cyberbullying they experienced. These types and methods are divided into four general classifications, i.e:

Category	Action
1. Public Humiliation	<ul style="list-style-type: none">• Spreading false information on behalf of the victims• Negatively manipulating the victims' photograph and uploading it to social media• Writing and sharing scornful messages to the public• Illegally accessing the victims' social media and altering private information• Uploading pornographic content related to the victims• Capturing and sharing digital conversations to other people• Sharing private electronic surveys filled by the victim• Illegally accessing the victims' social media and pretending to be the victim• Uploading disgraceful photographs of the victim
2. Malice	<ul style="list-style-type: none">• Calling rude names to the victim• Harassing the victim• Cursing the victim• Humiliating the victim• Mocking the victim

Category	Action
3. Unwanted Contact	<ul style="list-style-type: none">• Sending pornographic content to the victim• Sending sexually-related messages to the victim• Sending pictures that insulted the victim's race, religion, or ethnicity
4. Deception	<ul style="list-style-type: none">• Pretending to be someone else when talking to the victim• Requesting the victim to share private information when pretending to be someone else

Table 1. Acts of Cyberbullying according to Doanne et al. (2009)

Each social media platform has its own distinguished features. By considering this, the various types and methods to conduct cyberbullying depend on the features provided within the platform. For instance, cyberbullying on YouTube can be done through videos or the comment section. In the case of Twitter, it can be done by tweets or direct messages (DM). On the other hand, the various features provided by Facebook for its users also correspond to the various methods of cyberbullying in the platform, ranging from uploading videos/photos/writings on status, through comments or even short messages.

Social Media Platform Responses towards Cyberbullying

In conducting the comparative analysis of the policy initiatives made by the social media platform, the author wished to analyse three important aspects: the policy of the social media platform in regards to cyberbullying, the attempt at technological interventions and other non-virtual means. These aspects are differentiated based on the working mechanism of social media platform that normally has a standardized global practice, in which they build the framework enterprises, both technology-based or non-technology based, in correspondence to the prevailing standards

• Bullying in YouTube Community Guidelines

YouTube regulates the policy of cyberbullying in its community guidelines by incorporating both harassments with cyberbullying into the same mechanism. This regulation applies to videos, video descriptions, comments, live streamings, and other YouTube features and products. YouTube defines harassment and cyberbullying as the contents threatening individuals. This is being elaborately explained as contents such as:

Acts of Harassment and Bullying

- 1 It is containing mockery or inappropriate exclamations insulting an individual/group's essential attributes. Included in these essential attributes are age, caste, disability, ethnicity, gender identity and its expressions, nationality, race, immigrant status, religion, sex/gender, sexual orientation, victims of physical or sexual abuse, and veteran status³⁰.
- 2 Has the intention to humiliate or insult a minor. Minor here is being defined as someone who is under the age of 18 (noting here that the age range considered as a minor may differ between countries).
- 3 Spreading private information/someone's personal data, such as address, e-mail, account sign-in information, telephone number, ID number, or bank account information. Notes: publicly accessible numbers such as official government office numbers or business
- 4 Persuading someone to harass an individual both outside or inside YouTube
- 5 Supporting fans' offensive behavior such as doxing (spreading someone's private information), dogpiling (swarming negative comments towards individuals/groups), brigading (simultaneous planned attack towards individuals/groups) or off-platform targeting (attacking individuals/groups in real world).
- 6 Making threats of physical abuse or property destruction both implicitly or explicitly. Notes: implicit here means that the threat contains no specific date, time, or intention and can also be the act of showing weapons, imitation of violent acts, etc.
- 7 Depicting a group of authority (police or military) capturing or attacking an identifiable individual.
- 8 Containing a contents portraying serious violent acts (executions, torture, beating, etc.) towards an individual/groups.
- 9 Containing unwanted sexual contents (non-consensual) or any sexually degrading contents.
- 10 Exhibiting the methods to spread unwanted sexual images (non-consensual).

Table 2. Categories of actions considered as harassment and bullying by YouTube (compiled by authors).

There are several exceptions to these guidelines, such as when the initial purposes of the contents are education, documentary, scientific or natural aesthetic. A few examples are: public figures/public officials/world leaders' debate, orchestrated performance within artistic contexts (stand up comedy, diss track songs), and educational contents to raise awareness about harassment or cyberbullying. However, YouTube has emphasized that these exceptions are no justifications for harassing someone.

When there is a violation to the guidelines, YouTube will delete the contents and notify the owner through email. If this is the first violation, there will be no sanctions given. However, if the violation was not happening for the first time, YouTube will dispatch gradual sanctions, which is called 'strike.' There are 3 levels of strike, which are:

1 First Strike

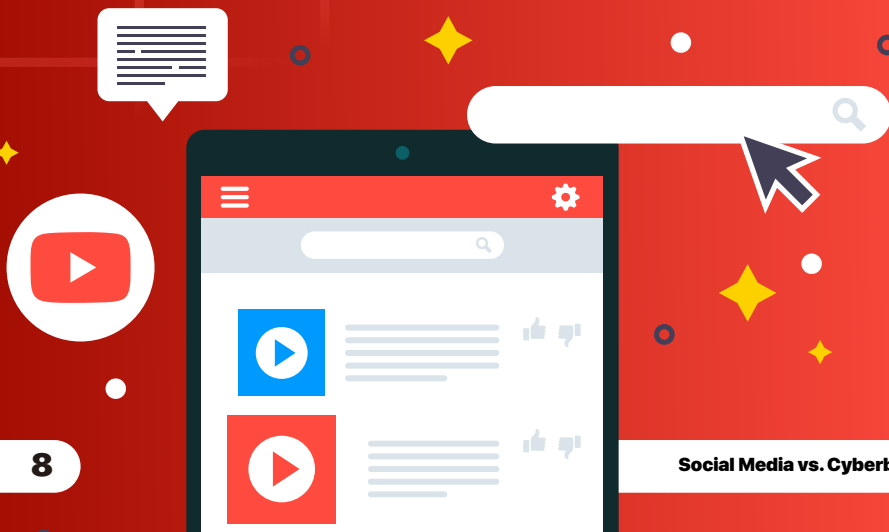
Users cannot upload videos, do live streamings, upload stories, create thumbnails, community uploads, collaborating on a playlist, adding or removing playlist using the save button for one week. This strike will last for 90 days in the user's account.

2 Second Strike

Users cannot create any content or access any features as mentioned in the first strike category for two weeks.

3 Third Strike

The User YouTube account permanently deleted.



There is a technological intervention used by YouTube to ensure that its users comply with these community guidelines about cyberbullying and harassment. For instance, comments that are containing potential keywords that might violate the guidelines will not be directly shown in the video's comment section; instead, it will go through a filtering folder to be rated by users. There is also a button to report disturbing or harassing comments or videos that are available to users.

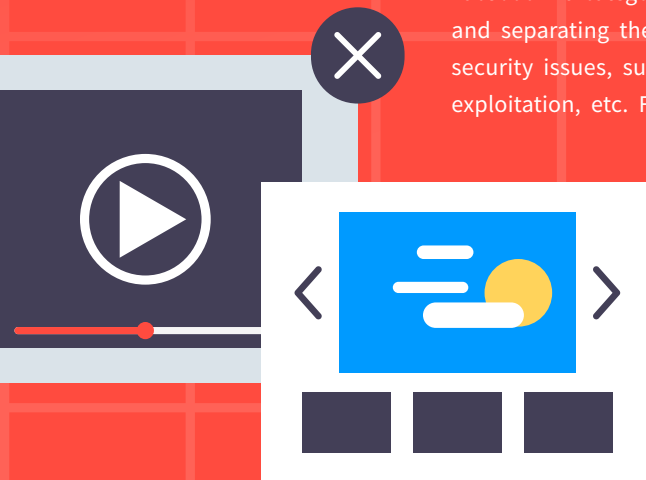
Youtube is also using the technology of Artificial Intelligence (AI) and Machine Learning to automatically detect comments that may contain cyberbullying elements. This technology utilized the reviews or feedback of YouTube users towards comments that are considered as toxic, and then detecting similar comments²¹. When it is already detected, YouTube moderators will decide whether the comments will be deleted or not.

Unfortunately, as of now, there is no information in regards to YouTube's possession of technological intervention against videos containing cyberbullying. On the other hand, YouTube has the advanced technology to detect copyright infringement called Content ID. The detection and mitigation of cyberbullying-related content are still manual.

◦ Bullying in Facebook Community Standards

In responding to acts of bullying in their platform, Facebook refers to the regulations named Facebook Community Standard. The Facebook Community Standard is applicable worldwide, in the sense that problems which arise in the Facebook platform in Indonesia will be handled through the same global standards. Within the Facebook Community Standard, Facebook is categorizing bullying along with harassment and separating them from other cases related to other security issues, such as suicide, adult and child sexual exploitation, etc. Facebook also has the prerequisite in

managing the problems of bullying and harassment. There are two categories provided by Facebook, the status and age of the user:²²



1

Public Figure (adults and minors)

According to Facebook, the handling of cases involving public figures must prioritize the assurance of open space for discussion in the comment section. This is in accordance with Facebook's principle to guarantee their users' freedom of speech²³. Therefore, Facebook differentiated the handling of cases between public figures and individuals.

2

Individuals (adults and minors)

For individually targeted bullying and harassment, Facebook takes up a greater measure in comparison to public figures cases. For individual adults, they must report privately by themselves if they are being harassed or bullied.

Within the Security section, sub-section Bullying and Harassment, Facebook identified actions that can be considered as harassment and bullying, such as:²⁴

Bullying Target(s)	Bullying Action
All User (Public Figures and Individuals)	<ul style="list-style-type: none"> ▶ Continuously contacting someone in the manner of: <ul style="list-style-type: none"> • non-consensual, as in the person do not wish to be contacted • complemented by sexual harassment, and exhibition of this unwanted contact to the public ▶ Taking any action against someone, in the manner of: <ul style="list-style-type: none"> • Attacking someone due to their status as victims of sexual abuse, sexual exploitation, or domestic violence. • Persuading individuals or groups to inflict self-harm or to commit suicide. • Attacking someone by calling them names related to sexual activities • Uploading contents about sadistic affairs, or the victim of sadistic affairs by asserting claims which undermines the sadistic nature of the events itself • Uploading contents about the victim or survivor of a sadistic events by attaching names or images claiming that it is fake or the person is being paid to pretend as the victim/survivor

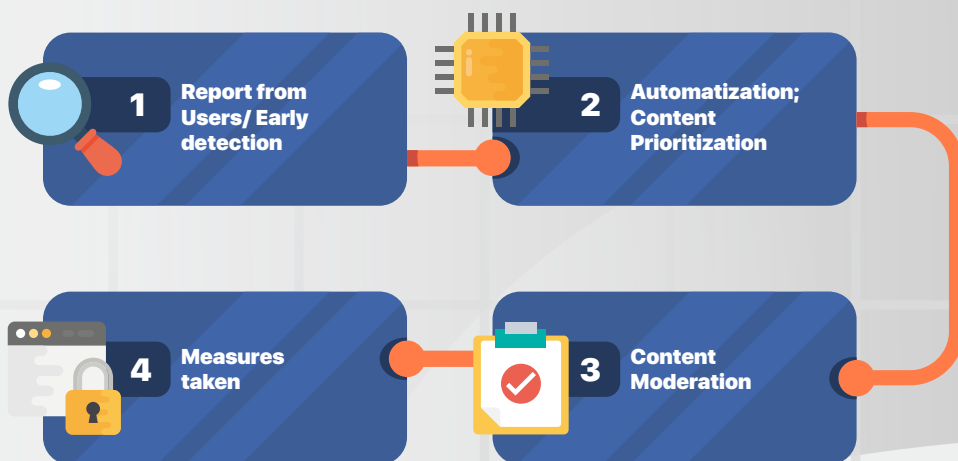
Bullying Target(s)	Bullying Action
	<ul style="list-style-type: none"> • Threatening to spread someone's private telephone number, address and email. • Creating or uploading contents to Page or Group purposefully established to attack individuals/groups by wishing these individuals/groups to die, suffering from life-threatening illness, or become disabled, declaring to be involved in sexual activity or claiming that these individuals/groups have sexually transmitted disease. • Sending a message to individuals/groups within the messaging thread wishing someone to die, suffering from a life-threatening illness, become disabled, or suffering from physical injuries.
Adult Public Figures	<p>► Intentionally targeting adult public figures by exposing them to contents comprised of:</p> <ul style="list-style-type: none"> • death wish/life-threatening illness/disability, • statement to be involved in sexual activity or support to be involved in sexual activity, • claims about sexually transmitted disease, worship, celebration, or mock death of the pertaining adult public figures.
Minor Public Figures	<p>► Intentionally targeting minor public figures by exposing them to contents comprised of:</p> <ul style="list-style-type: none"> • Comparison with animals/insects which can be considered as culturally degrading in terms of intellectuality or physical conditions, or with inanimate objects ("cow", "monkey", "potato") • Manipulation in the purpose of highlighting, circling, or negatively attracting attention to physical attributes (nose, ears, etc.)
Minor Public Figures/Individuals	<p>► Targeting individuals or minor public figures through actions, such as:</p> <ul style="list-style-type: none"> • wishing someone to die/suffering from life-threatening illness/disability, • claiming about sexually transmitted diseases/sexual activity, • creating Page or Group intended to attack someone with curse words, • horrendous physical description, • claim about someone's identity or religious blasphemy, using expression of derogation or disdain.

Bullying Target(s)	Bullying Action
Individuals	<p>► Targeting individuals by exposing them to contents comprised of:</p> <ul style="list-style-type: none"> • Comparison with animals/insects which can be considered as culturally degrading in terms of intellectuality or physical conditions, or with inanimate objects (“cow”, “monkey”, “potato”) • Manipulation in the purpose of highlighting, circling, or negatively attracting attention to physical attributes (nose, ears, etc.) • Attack through negative physical description • Ranking (comparison) of physical appearance or personality of individuals • Depiction of someone else in the process of, or just finished doing, menstruation, urination, regurgitation, or excretions within the context of degrading an individual or contains expression of disgusts • Physical bullying in the purpose of degrading an individual • Character claims or negative capabilities or self-bullying, only if objects are targeting more than one individual (on Page or Groups toward specific individual) • Worship, celebration, or mock death of an individual • Self-bullying • Unwanted images manipulation • Comparison with public figures, fictional character, or other individuals based on physical appearances • Phrases about identity or religious blasphemy, except in high-risk countries where the policy of Violence and Provocation must be enacted • Comparison with animal or insects which are not culturally degrading in terms of intellectuality or physical conditions (“tiger”, “lion”) • Neutral and positive physical description, • Non-negative and capability claims. • Violation of bullying and harassment which is being framed in an elegant context • Attacks through derogatory terms indicating the lack of sexual activity
Adult individuals/minor individuals	<p>► Targeting adult individuals or minor individuals by exposing them to contents comprised of:</p> <ul style="list-style-type: none"> • Swear words • Sentences about romantic relationships, sexual orientation, and sex/gender identity • Question about hygiene • Coordination or support to excommunicate • Character or negative capability claims, except in the context of alleged criminal acts against adults • Expression of disdain or disgust, except in the context of alleged criminal acts against adults

Bullying Target(s)	Bullying Action
Minor individuals	<p>► Targeting minor individuals by exposing them to contents comprised of:</p> <ul style="list-style-type: none"> • Allegations about criminal acts or other actions against the law • Videos of physical bullying or abuse towards minor in the context of a quarrel which is being shared in a non-condemning manner.

Table 3. Facebook's bullying and harassment provision based on the target of action (compiled by authors)

Based on Facebook Community Standard, Facebook is responding to bullying through the following process:



Picture 2. The Follow-up Process of Harmful Contents (adapted from Video in Facebook Source of Information)²⁵

Facebook detected contents accommodating harassment and bullying action based on autonomous detection and reports. Facebook's autonomous detection is done by employing artificial intelligence technology and reports received from users. Feedbacks generated from these two methods are then being processed into two steps: automated selection and content moderator²⁶. During the automated selection process, Facebook is harnessing the automation process to filter whether the suspicious contents bear dangerous potential. After passing the automated selection stage, these contents will be sent to the content moderator to be filtered and confirmed of its violations. If the contents are proved to be violating Facebook Community Standards, Facebook might deactivate the user's account²⁷.

Facebook detailed the consequences of Community Standard violation in their Newsroom section. The consequence is given to user and Page (through Page and Page admin) who violated the Facebook Community Standard. If violation does indeed happen, the detail of the consequences given by Facebook are as follows:

- 1 Facebook did not indicate the number of violations which led to permanent or temporary account restrictions to avoid the risk of pre-arranged violation.
- 2 If a Page has exceeded certain limits of violation, the Page's publications will be discontinued, and if the Page repeats the violation, the Page will be deleted.
- 3 Temporary account restrictions only enable the user to read and see timeline updates in Facebook²⁸.

Beside of these measures, Facebook also provides other methods such as giving cautions to video and photo contents which may harbor the act of bullying or harassment, the Bullying Prevention Hub²⁹ (information site for youths, parents and educator,) is cooperating with various organizations in Indonesia³⁰ (such as YLBH Apik, YCAB Foundation and SudahDong,) while also providing a set of Facebook Security Tools. Facebook Security Tools enables bullying victim to do preventive measures and responding secretly without being known by the harasser, such as³¹:

- 1 Blocking or ending Facebook friendship without the harasser's knowledge that their account has been blocked by the victim,
- 2 They may stay as 'friends' on Facebook, although the victim did not follow the harasser anymore. If the victim decided to stop following the harasser, the harasser would not get any notifications regarding this. This feature enables the victim to avoid any contact with the harasser, without prior knowledge to the harasser.
- 3 Reporting the bullying or harassment to Facebook for the reports to be carried out without the harasser knowing the identity of the victim when they are being contacted.

4

Sharing contents to only trusted persons by using the Filter Audience tool.

5

Deleting or hiding comments from the harasser that might cause inconvenience without being known by the harasser.

6

Neglecting the harasser's messages without being known.

• Bullying in Twitter Rules

Twitter based its policy measures on Twitter Rules. Twitter Rules are separated into three sections: Safety, Privacy, and Authenticity.)³² The guidelines related to bullying in Twitter Rules are not being categorized within a specific section dedicated to bullying; rather it is being separated under prohibited behaviors within the Security aspect. In Twitter Rules, Twitter also explained that each action has different consequences. Referring to Doanne et al. categorization of bullying act, prohibited actions and its consequences related to bullying outlined by Twitter are as follows:

Provisions	Consequences
Hateful Conduct	
<p>1 Expressing hatred which led to violence, threats, or harassment toward others on the basis of race, ethnicity, citizenship, caste, sexual orientation, gender, gender identity, religious affiliation, age, and special needs. This is being defined in detail with Twitter as:</p> <ul style="list-style-type: none"> • Set up a profile picture, username and display name which contain targeted hatred towards individuals, groups, or protected categories • Uploading a statement containing violence towards the target, which inflicted a serious injury and physical violence for a long period of time, until it causes death or severe injury, for example "I will kill you." 	<p>The consequences are depending on several factors, however it encompasses the significance of the damages being inflicted, users' track records and the repetition of violation. The level of consequences consisted of:³⁵</p> <ul style="list-style-type: none"> • Request to delete content • Limit user to read-only mode • Extension of read-only mode period (if repetition occurs) • Permanent account restrictions

Provisions	Consequences
Hateful Conduct	
<ul style="list-style-type: none"> • Uploading a content that bears a wish, promotion, and expression of desire for someone or a group for protected individuals to experience death, severe and prolonged injury, or acute illnesses, for example "I hope you got cancer and die." • Targeting individuals with contents depicting massacres, violence, or other particular acts targeting vulnerable groups, for example: media which portrays the Holocaust victims. • Targeting individuals with content purposefully to made to ignite fears or spreading stereotypes about vulnerable groups, for example "all religious groups are terrorist." • Continuously targeting individuals, sending contents of mockery, designation, or other contents with the purpose of ruining, disgracing and sending negative or dangerous stereotypes towards vulnerable groups³⁴. • Uploading contents of logo, symbols or pictures with the purpose of promoting fears and hatred towards someone on the basis of race, religion, special needs, sexual orientation, gender identity or ethnicity/citizenship. For example, the symbol of Nazi's swastika 	
Abuse/harassment	
<p>2 Targeting someone for harassment or persuading others to do harassment. One of the actions which can be found within this category is wishing someone to experience physical abuse³⁶. This is being defined in detail by Twitter as follows:</p> <ul style="list-style-type: none"> • Uploading contents that bear a wish, promotions, or an expression of desire for someone to experience death, severe and prolonged injury, or severe illnesses, for example, "I hope you got cancer and die." • Conducting both an act and sexual contents which objectifying individuals without consent. SAs for example discussing someone else's body without their consent. • Hurling taunts at individuals with the purpose of harassing or intimidating. • Persuading others to harass an individual or groups, both through cyber or non-cyber harassment. 	<p>The consequences are depending on several factors, however it encompasses the significance of the damages being inflicted, users' track records and the repetition of violation. The level of consequences consisted of³⁷:</p> <ul style="list-style-type: none"> • Request to delete content • Limit user to read-only mode • Extension of read-only mode period (if repetition occurs) • Permanent account restrictions

Provisions	Consequences
Suicide or self-harm	
<p>3 Promoting or persuading someone to do suicide or an act of self-harm, for example, suicide strategy³⁸.</p>	<p>The consequences are based on the load of the content and users' track record³⁹.</p> <ul style="list-style-type: none"> • First time violation, the user will be obliged to delete the content and his/her access will be temporarily restricted. • When violations occur continuously, a user's account can be restricted. • Besides these two consequences, Twitter will also contact the reported individuals and informing someone they recognized to let them know that he/she is in a potential danger,




Table 4. Provisions of bullying according to Twitter (Source: Authors)

Other than the aforementioned provisions, Twitter also has another provision to prohibit the spread of sensitive content⁴⁰, the spread of private data⁴¹ and impersonation of an individual or groups.⁴² However, these provisions were not directly applied to solve the issues of bullying in Twitter or if this are being applied, the majority are just more of an emphasis of the aforementioned provisions. According to Twitter Rules, Twitter response to bullying is adopting the similar arrangement process with other social media platforms: it receives feedback from the contents containing acts of bullying (whether from automated detections or user reports), reviewing the contents, and then continues with the follow-up actions. Although the consequence as the result of follow-up actions are being published, Twitter does not publish the process of the follow-up actions itself. Twitter is also responding to bullying by flagging down contents which have the potential to harbor violations. It also engage with various related organizations in Indonesia, such as Ending the Sexual Exploitation of Children in Indonesia (ECPAT Indonesia), ICT Watch, YCAB Foundation, Lembaga Studi dan Advokasi Masyarakat Indonesia (ELSAM) and Wahid Foundation.⁴³



Comparison of Social Media Responses towards Cyberbullying



			
Provisions	<ul style="list-style-type: none"> • Including cyberbullying along within the category of harassment. • The provisions were generally laid out on 10 points explaining the types of the pertaining harassment/bullying. 	<ul style="list-style-type: none"> • Including cyberbullying along within the category of harassment • Provisions were laid out in detail based on two categories: users' popularity (Public Figures/ Individuals) and age (Adults/Minors.) • Considering variables such as local culture/contextualization in determining contents violation. 	<ul style="list-style-type: none"> • Including cyberbullying as a type of violations • Provisions are written in detail in 3 types of violations: Hateful Conduct, Abuse/Harassment & Suicide or Self-Harm.
Consequences & Regulation Enforcement	<ul style="list-style-type: none"> • Deleting contents • Give warning to the first violation after the content removal • Giving strike with the consequence of features limitation within a particular period. • Permanently removing contents on the third strike. 	<ul style="list-style-type: none"> • Follow-ups of a report/violation findings are still considering the availability of public discussion space. • Consequences are being given after contents are received from Facebook search results or reports, filtered by artificial intelligence, and checked by content moderator. • Consequences are being carried out towards Use and Page (including Page and Page admin). • Indicators on the implementation of consequences were not being explained in detail by Facebook to avoid pre-arranged violations. 	<ul style="list-style-type: none"> • Founded its consequences on various factors, including these three factors: inflicted damages, the perpetrators' track records, and the repetition of violation. • Consequences are being gradually carried out from content removal request, limitation on read-only mode, the addition of read-only mode periods, until permanent account deletion.




			
Technology-based enterprises	<ul style="list-style-type: none"> • Filtering comments which potentially violate the rules, before being published. (Determined by users/video owners.) • Involving artificial intelligence to detect toxic comments. • Providing the service of content reports which considered by user has the element of bullying. • Initiating YouTube Creator Academy to avoid creator do bullyings. 	<ul style="list-style-type: none"> • Involving artificial intelligence to filter content feedbacks related to bullying. • Providing Facebook Security Tools, including reporting mechanism and victim protection in secret. • Providing the Bullying Prevention Hub. 	<ul style="list-style-type: none"> • Providing reporting mechanism and victim protection in secret.
Non-technological based enterprises	<p>Non-technological based enterprises to tackle bullying in YouTube is not found.</p>	<ul style="list-style-type: none"> • Holding events which have the purpose to socialize the provisions, such as: Safe in Social Media. • Cooperating with NGOs and related communities in Indonesia such as: YLBH Apik, YCAB Foundation & SudahDong. 	<ul style="list-style-type: none"> • Cooperating with NGOs and related communities in Indonesia such as: ECPAT Indonesia, ICT Watch, YCAB Foundation, ELSAM & Wahid Foundation.

Table 5. Comparison on the Findings of YouTube, Facebook & Twitter Responses towards Cyberbullying (Source: Authors)

○ The categorization of bullying act and their variety of consequences

To summarize, actions that are considered bullying have been regulated in the social media guidelines. However, each social media platform has different approaches to prevent the occurrence of cyberbullying. Within Facebook Community Standards, bullying fell under the same category with harassment and explained through prohibited actions within their platform, based on the target of the bullying. This categorization, based on the target of the bullying shows that Facebook understands each individual has varying degrees of protection needs. The classification that highlights the status of the individual being also concerned implicitly emphasizes that someone with great popularity also has a great responsibility to carry themselves in public.

Although it is quite detailed in indicating the actions considered as bullying, Facebook did not publish the indicators of reasoning to the appropriation of consequences with the same detail as the former. And although this policy averted users to outsmart the Facebook Community Standard, this policy also reduces user's transparency in regards to content restrictions that are still permitted on Facebook. The categorization and appropriation of consequences are also different on Twitter. In Twitter Rules, bullying actions can be found among other prohibited actions by Twitter within the Security section. Twitter has a more transparent indicator of reasoning in the appropriation of consequences as compared to Facebook. However, the response of Twitter in handling bullying cases, especially in the case of cyber harassment, is still deemed insufficient by organizations for humanity, such as Amnesty International.⁴⁴ One of the critiques being launched against Twitter response is the slowness of the follow-up actions taken by Twitter towards reports of violation concerning cyber harassment, especially towards women. Specific provisions concerning the regulations of bullying can also be found on YouTube. However, it has not yet addressed the bullying which might occur in the YouTube public timeline, such as the comment section. On YouTube, there are clear and specific rules in regards to the video contents uploaded by users to avoid cyberbullying. Still, it is not the case for the comment section, where bullying can also happen. The guidelines stipulated by YouTube are still focusing on videos only so that discussion spaces such as the comment section and chats only follow the rule, which is general in nature.



• Responses are being carried while still maintaining public space for discussion

The provision of public space for discussion is one of the principles endorsed by social media platforms. In the case of Facebook, for example, while limiting prohibited actions and responding to those actions, social media platforms are also prioritizing the maintenance of public discussion, as not to put the freedom of speech principle in jeopardy. In the case of bullying happening on Facebook, this is being shown in how Facebook handles bullying cases towards public figures. For Public Figures (the term addressed by Facebook towards celebrities, public figures, influencers, etc.), which experienced bullying, Facebook treatments are limited by the principle of non-interference with public space for discussion. In this regard, Public Figures have more lax protection compared to Individuals (regular users). The bullying classification which differentiates between Public Figures and Individuals cannot be found on other social media platforms. However, Twitter is also

prioritizing the maintenance of public space for discussion, although it did not specifically concern users who are considered as Public Figures only. Whereas YouTube in fact did not have any elaborate or specific guidelines concerning the comment section, which can also be considered as a public discussion space where bullying might occur. Diverging from Facebook and Twitter, YouTube is a social media platform where its feature revolves around videos, and they do not have timelines like Facebook and Twitter and therefore operate on an entirely different logic compared to those two. One of the distinct features is that YouTube enables their user to deactivate the video's comment section or discussion space on a live video, thereby negating the existence of public discussion space itself.

◉ The response is taken by considering local cultural context and intentions

One of the study findings shows that in tackling bullying issues, social media platforms are also observing the local cultural context and also the intentions of the content being uploaded in their platforms. The consideration of local cultural context in handling the cases of bullying can be found on Facebook Community Standard. In some provisions, Facebook wrote the term "culturally" and "within the context." This can be argued that both terms are emphasizing the local cultural context in judging whether content can be categorized as bullying or harassing content. This consideration becomes an important point since each community has different value standards in determining whether content can be categorized as an act of harassment or bullying. One term considered as dehumanizing in one society does not always mean the same in another society. The consideration of local cultural context can be found in Facebook Community Standards, but not in Twitter Rules. Twitter Rules tend to rule out contents based on its intentions, for example, by allowing abusive contents if the purpose was for education. This approach of intention is also being utilized by YouTube. YouTube did not elaborately explain about the local cultural context in their cyberbullying guidelines. The only local contextualization on YouTube is regarding the guidelines concerning minors, whereas the age consideration of a minor might differ from one country to another. However, in regard to the matter of intention, YouTube is clearly opening the space for contents such as comedy satire, stand-up comedy, diss track songs, public figures debates, and other contents that may share some similarities with cyberbullying. However, YouTube also emphasizes that these contents should not serve as the justification for cyberbullying.





◦ **Technology and non-technological based enterprises**

From the perspective of preventive and upkeeping measurements of the aforementioned provisions, YouTube, Facebook, and Twitter had the advanced methods in the form of devices, reporting mechanisms, and consequences for the perpetrators. However, Facebook has had more efforts compared to Twitter and YouTube. By the existence of the Bullying Prevention Hub, Facebook offers educational actors towards users and those who are concerned with the users' safety (parents and teachers, for example) to act wisely. This effort has not been visible in Twitter and YouTube. The efforts of both platforms are still limited to cyberbullying content detection through users' report which then being developed into various features such as the screening of sensitive comments by users before being displayed, and also providing flexibility for users to do the follow-ups in clandestine.

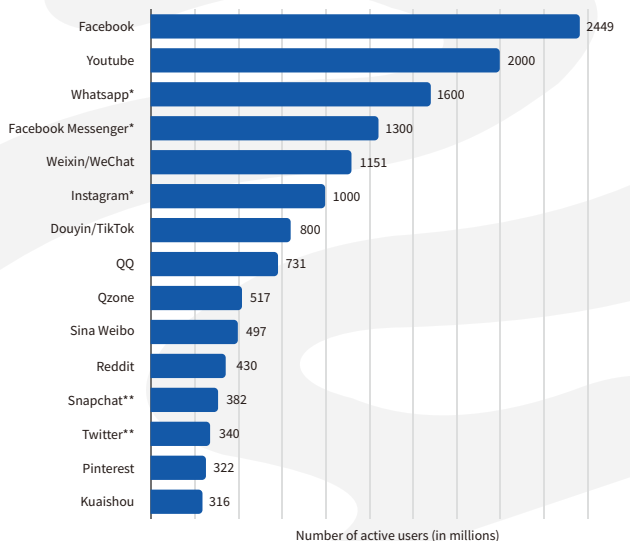
Comprehensive and Thorough: Facebook Standard of Handling Cyberbullying



The three social media platforms mentioned already have adequate regulations concerning cyberbullying. Between YouTube, Facebook, and Twitter, Facebook has the most specific regulations and measures to prevent cyberbullying. YouTube and Twitter have not yet been on par with Facebook concerning such measures. Moreover, it can be assumed that Facebook's advances compared to the two other social media can be attributed to reasons such as (1) the position of Facebook as social media with the most number of users worldwide and (2) the variety of Facebook's features. According to Statista, Facebook is a social media with the most number of users, with 2.4 million users worldwide.⁴⁵ This position has motivated Facebook to invest more to develop its platforms and also to increase their user's experience. Facebook's sizable efforts can also be attributed to the numerous features that they possessed. Facebook has the features to upload images, text, and videos which can be used evenly. Although Twitter has similar features with Facebook, Twitter is a social media which is more inclined towards microblog, so that the text feature is being used more often. Differing from both social media, YouTube only has the video upload and text features, whereas the text upload feature is only a minor feature (text in the form of comments within the comment section). These two hypotheses can be further examined in future research.



Most popular social networks worldwide as January 2020, ranked by number of active users(in millions)



Sumber
We Are Social;
various sources (company data),
Hootsuite, Data Reports

Additional Information:
worldwide; various sources
(company data);
Data Report as January 25, 2020,
social networks
and messenger chat

Picture 3. Number of Global Social Media Users in January 2020 (Source: Statista)⁴⁶

Regarding the second hypothesis about Facebook, the other two platforms, Twitter and YouTube may adopt Facebook's policy and responses. Twitter can adapt Facebook's policy and responses, especially to have a more specific categorization regarding cyberbullying. Twitter can also consider the local cultural context in determining whether a content violates the Twitter Rules or not. YouTube can still adapt Facebook methods in making rigorous guidelines, especially in regulating cyberbullying that happens in the form of texts. This is due to the large potential of bullying that might happen in the video's comment section or YouTube video live-chat. With this being said, rigorous guidelines about the type of bullying, harassment, towards whom the act was directed, and also the consideration of local cultural context are indeed necessary. The specification of guidelines and adequate preventive measures will make social media geared up to face the advance of cyberbullying. By making the definition of cyberbullying more specific and elaborate and also sufficient technological intervention, social media can guarantee the availability of secured discussion space from bullying without jeopardizing the quality of the discussion space itself.

References

- ¹Katadata, (2019). Berapa Pengguna Media Sosial Indonesia?. Katadata [online] available at <https://databoks.katadata.co.id/datapublish/2019/02/08/berapa-pengguna-media-sosial-indonesia>. [Accessed 31 Mar 2020]
- ²Jayani, D. H. (2019). Survei APJII: 49% Pengguna Internet Pernah Dirisak di Medsos. Katadata [online] available at <https://databoks.katadata.co.id/datapublish/2019/05/16/survei-apjii-49-pengguna-internet-pernah-dirisak-di-medsos>. [Accessed 6 Apr 2020]
- ³Ibid.
- ⁴Ibid.
- ⁵Pertiwi, W. K. (2019). Facebook Jadi Medsos Paling Digemari di Indonesia. Kompas [online] available at <https://tekno.kompas.com/read/2019/02/05/11080097/facebook-jadi-medsos-paling-digemari-di-indonesia>. [Accessed 31 march 2020].
- ⁶Hinduja, S., & Patchin, J. W. (2009). *Bullying Beyond the Schoolyard: Preventing and Responding to Cyberbullying*. California: Corwin Press.
- ⁷Smith, P. K., Del Barrio, C., & Tokunaga, R. S. (2013). Definitions of Bullying and Cyberbullying: How Useful are the Terms?. In Bauman, S., Cross, D., & Walker, J. (Eds). *Principles of Cyberbullying Research*. London: Routledge.
- ⁸Kowalski, R. M., Diumetti, G. W., Schoeder, A. N., & Lattanner, M. R. (2014). Bullying in the Digital Age: A Critical Review and Meta-analysis of Cyberbullying Research Among Youth. *Psychological Bulletin*, 140(4), 1073-1137.
- ⁹Diamanduros, T., Downs, E., & Jenkins. (2008). *The Role of School Psychologist in the Assessment, Prevention and Intervention of Cyberbullying*. In Zalaquett, C., & Chatters, S. (2014). *Cyberbullying in College: Frequency, Characteristics, and Practical Implications*. SAGE Open.
- ¹⁰Kowalski, R. M., Diumetti, G. W., Schoeder, A. N., & Lattanner, M. R. (2014). Bullying in the Digital Age: A Critical Review and Meta-analysis of Cyberbullying Research Among Youth. *Psychological Bulletin*, 140(4), 1073-1137.
- ¹¹Tjongjono, B., Gunardi, H., Pardede, S., & Wiguna, T. (2019). Perundungan Siber (Cyberbullying) serta Masalah Emosi dan Perilaku pada Pelajar Usia 12-15 di Jakarta Pusat. *Sari Pediatri*, 20(6), 342-348.
- ¹²Ibid.
- ¹³Ibid.
- ¹⁴Gilroy, M. (2013). Guns, Hazing and Cyberbullying Among Top Legal Issues on Campuses. *Education Digest*, 78, 45-50.
- ¹⁵Varjas, K., Talley, J., Meyers, J., Parris, L & Cutts, H. (2010). High School Students' Perceptions of Motivations for Cyberbullying: An Exploratory Study. *West J Emerg Med*, 11(3), 269-273.
- ¹⁶Mason, K. L. (2008). Cyberbullying: A Preliminary Assessment for School Personnel. *Psych in the Schools*. 45(4), 323-348.
- ¹⁷Razhaukas, J., & Stoltz, A. D. (2007). Involvement in Traditional and Electronic Bullying among Adolescents. *Develo Psych*, 43, 564-575.
- ¹⁸Doane, A. N., Kelley, M. L., & Cornell, A. M. (2009). Online Bullies: College Students' Reports of Internet Harassment and Cyberbullying. In Doane, A. N., Kelley, M. L., Chiang, E. S., & Padilla, M. A. (2013). *Development of the Cyberbullying Experiences Survey*. *Emerging Adulthood*, 1(3), 207-218.
- ¹⁹YouTube. (n.d.). Harassment and Cyberbullying Policy. [online] Google Support. available at https://support.google.com/youtube/answer/2802268?visit_id=1-636215053151010017-1930197662&rd=1&hl=en [Accessed 28 March 2020]
- ²⁰YouTube (n.d.). Hate Speech Policy. [online] Google Support. available at <https://support.google.com/youtube/answer/2801939> [Accessed 28 March 2020]
- ²¹Carson, E. (2017). Google Jigsaw Puzzle Perspective Toxic Comments Machine Learning AI. [online] Cnet. available at: (<https://www.cnet.com/news/google-jigsaw-puzzle-perspective-toxic-comments-machine-learning-ai/>) [Accessed 28 march 2020]
- ²²Facebook, (n.d.) 10. Perundungan (Bullying) dan Pelecehan. [online] available at <https://www.facebook.com/communitystandards/bullying>. [Accessed 29 Mar 2020.]
- ²³Ibid.
- ²⁴Ibid.
- ²⁵Facebook, (n.d.) Sumber Informasi. [online] available at <https://www.facebook.com/safety/resources>. [Accessed 29 Mar 2020.]

References

²⁶Ibid.

²⁷More on this issue, see Facebook, (n.d.) Akun Facebook pribadi saya dinonaktifkan. [online] available at <https://www.facebook.com/help/103873106370583> [Accessed 6 Apr 2020].

²⁸Facebook, (2018). Enforcing Our Community Standards. [online] available at <https://about.fb.com/news/2018/08/enforcing-our-community-standards/>. [Accessed 8 Apr 2020.]

²⁹More about the Bullying Prevention Hub, see Facebook, (n.d.) Hentikan Perundungan (Bullying). [online] available at <https://www.facebook.com/safety/bullying>.

³⁰More about Facebook Partners in handling bullying in Indonesia, see Facebook, (n.d.) Sumber informasi.

³¹More about Facebook Security Tools, see Facebook, (n.d.) Alat Bantu. [online] available at <https://www.facebook.com/safety/tools>.

³²More on Twitter Rules, see Twitter, (n.d.) The Twitter Rules. [online] available at <https://help.twitter.com/en/rules-and-policies/twitter-rules>.

³³Twitter, (n.d.) Hateful conduct policy. [online] available at <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>. Accessed 30 Mar 2020.]

³⁴More about this Provisions, see Ibid.

³⁵Ibid.

³⁶Twitter, (n.d.) Abusive Behavior. [online] available at <https://help.twitter.com/en/rules-and-policies/abusive-behavior>. [Accessed 30 Mar 2020]

³⁷Ibid.

³⁸Twitter, (n.d.) Glorifying Self Harm. [online] available at <https://help.twitter.com/en/rules-and-policies/glorifying-self-harm>. [Accessed 30 Mar 2020]

³⁹Ibid.

⁴⁰More on sensitive content, see Twitter, (n.d.) Sensitive Media Policy. [online] available at <https://help.twitter.com/en/rules-and-policies/media-policy>.

⁴¹More on the spread of private information, see Twitter, (n.d.) Private information Policy. [online] available at <https://help.twitter.com/en/rules-and-policies/personal-information?lang=browser>.

⁴²More on Identity Impersonation, see Twitter, (n.d.) Impersonation Policy. [online] available at <https://help.twitter.com/rules-and-policies/twitter-impersonation-policy>.

⁴³Twitter, (n.d.) Safety Partners. [online] available at https://about.twitter.com/en_us/safety/safety-partners.html#indonesia. [Accessed 31 Mar 2020]

⁴⁴More on the Amnesty International critics towards Twitter, see Amnesty International, (n.d.) Toxic Twitter - The Reporting Process. [online] Available at <https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-4/> [Accessed 30 Mar 2020.]

⁴⁵Clement, J., (2020). Most popular social networks worldwide as of January 2020, ranked by the number of active users. Statista [online] available at <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>. [Accessed 8 Apr. 2020.]

⁴⁶Ibid.







Center for Digital Society

Faculty of Social and Political Sciences
Universitas Gadjah Mada
Room BC 201-202, BC Building 2nd Floor,
Jalan Socio Yustisia 1
Bulaksumur, Yogyakarta, 55281, Indonesia

Phone : (0274) 563362, Ext. 116
Email : cfds.fisipol@ugm.ac.id
Website : cfds.fisipol.ugm.ac.id



facebook.com/cfdsugm



[Center for Digital Society \(CfDS\)](https://www.linkedin.com/company/Center-for-Digital-Society-(CfDS)-UGM)



[cfds_ugm](https://www.instagram.com/cfds_ugm)



[@cfds_ugm](https://wa.me/c/62/274563362)



[@cfds_ugm](https://twitter.com/cfds_ugm)



[CfDS UGM](https://www.youtube.com/channel/UCfDSUGM)