**Reporting system:**

**1. Introduction to the reporting system:**

Reporting is particularly important for MyDiary Notebook. **It is the tool by which the platform keeps inappropriate/harmful content from doing any harm to our users or put the platform in a legal jeopardy.**

When the user reports a post/account/comments on Public wall/friend’s diary and he/she thinks it is inappropriate or against our community standard and terms of service use, he/she has the right to report this post/account/comments. When he/she report it, a window will appear that include multiple issues, such as: misleading, inaccurate post, fake account … He/she should choose the issue that express his/her problem and report it. The team of reporting system is responsible for managing the system and deal with these reports according to certain criteria and guidelines, which are included in this document. In some cases, posts and accounts may end being suspend, or even shut down in some extreme cases.

**2. Report design and data:**

Reports are the main and the only tool by which we collect information about inappropriate/harmful content, and it is almost the only tool for users to tell us so. They are the main blocks for all cases to follow by MyDiary Notebook team. Accordingly, the information that reports provide are very essential to capture harmful or inappropriate content.

The following are the data needed for each report that help the team to know the

Details of any inappropriate content, which include:

1. Report Main information:

* Time of Reporting.
* Type of Report:

1. Report Post.
2. Report Account.
3. Report Hacked Account.

* Report issue.
* Report sub issue (the category).
* Number of reports on post/comment until date.
* Notes: when the user writes in "the tell us more" box at reporting window.

1. Post:

* From which page the reported post/account:

1. My Diary.
2. Friend’s Diary/Circle name.
3. Public Wall.
4. The Story of the Anonymous.

* Time of posting/commenting.
* The post/comment, the post/comment will be viewed as it is.

Comment

Post

1. Reporter:

* Reporter username.
* Date of creating account.
* The connection between the reporter and the reported (ex: if he/she friends with you & in which circle or if he/she not of your friends).

1. Reported:

* Reported username.
* Date of creating account.

1. **Issues of reporting and classifying issues:**

To help a user report a post/account that includes harmful or inappropriate content and to help the team follow on that report, he/she can report the inappropriate content by choosing issues like misleading or inaccurate content that appear on the reporting window, these issues are classified by Xs which stand for the level of harm. So, when the number of Xs is high it means that the level of harm is high. The number of Xs goes from 1X to 5Xs.

Classifying issues by the following:

|  |  |
| --- | --- |
| **Number of X’s** | **Issues** |
| X | Misleading  Inaccurate post |
| XX | Offensive content  Harmful language |
| XXX | Bullying and harassment |
| XXXX | Hate speech  Nudity or sexually content |
| XXXXX | Violent or graphic content  Suicide, self-injury  Abuse  Unlawful Activity  Copyright violation  I think it shouldn’t’t be on MyDiary Notebook  Violent or dangerous organizations  Fake account & identity theft  Hacked account  Others |

Table 1 (Classifying issues by X’s(

1. **Define a case, and when reports collectively become a case:**

When reports on a post/account reach a certain level (certain number of reports) a case will be established.

A case is a special situation that the team should deal with to resolve or be

alerted about. For each case, an analysis of the situation should be performed

to take the appropriate decision for the case. Cases are classified to:

1. Cases on a post: Alerts & Decisions.
2. Cases on accounts to be alerted about.
3. Cases on accounts for a decision to be made.
4. Cases on comments.
5. **Cases on a post: Alerts & Decisions:**

Cases on posts are divided to areas: The area of alerts and the area for decision

making. When a post reaches a certain level of reports number, the team

should be alerted, so to follow on that post before it does any harm. When a

post reach certain level. The team needs to review posts on P.W when they

reach specific reporting percentages and becomes a case. While there is no

alert for posts on the friend’s diary.

The decisions for cases on a post are divided to: Automated Decision on cases

with one issue and Automated Decision on cases with multi issues.

Automated decision on cases with one issue follow a special rule for each issue

and algorithm includes percentage or a specific number of reports.

Mostly such rules will result in moving the post from a public wall/friend’s

diary.

All automated decisions should be reviewed by the team within one to two

days. To ensure that the automated rules operate according to their intended

aims.

The below table shows for each issue the specific percentage of reporting on

post for a case of Alert to be developed, and for the post to be flagged as in

the zone of Alerts.

Also, it shows the percentage of automated decisions. When the post reaches

the demonstrated reports level a decision will be made automatically with the

action is shown.

|  |  |  |
| --- | --- | --- |
| **Issues** | **Alert** | **Automated Decisions on cases per issue** |
| **Misleading** | 5% - 9% (alert  for the team to review the post on P.W when it reaches these reporting percentages and become a case ). | Decision: when the post on P.W and has 10% of reporting & ≥ 500 views, the post will be only me.  Decision: when the post on friend’s diary and has 5 reports, the post will be only me. |
| **Inaccurate post** | 5% - 9% | Decision: when the post on P.W and has 10% of reporting & ≥ 500 views, the post will be only me.  Decision: when the post on friend’s diary and has 5 reports, the post will be only me. |
| **Offensive content** | 2% - 3% | Decision: when the post on P.W and has 4% of reporting & ≥ 500 views, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 3 reports, the post will be only me. |
| **Harmful language** | 2% - 3% | Decision: when the post on P.W and has 4% of reporting & ≥ 500 views, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 3 reports, the post will be only me. |
| **Bullying and harassment** | 10 – 19 | Decision: when the post on P.W and has 20 reports, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 2 reports, the post will be only me. |
| **Hate speech** | 2% - 3% | Decision: when the post on P.W and has 4% of reporting & ≥ 500 views, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 2 reports, the post will be only me. |
| **Nudity or sexually content** | 3% - 4% | Decision: when the post on P.W and has 5% of reporting & ≥ 500 views, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 2 reports, the post will be only me. |
| **Violent or graphic content** | 2% - 3% | Decision: when the post on P.W and has 4% of reporting & ≥ 200 views, the post will be on only me.  Decision: when the post on friend’s diary & and has 2 reports, the post will be only me. |
| **Suicide, self-injury** |  | Decision: when the post on P.W and has 2 reports, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 1 report, the post will be only me. |
| **Abuse** | 2% - 3% | Decision: when the post on P.W and has 4% of reporting & ≥ 200 views, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 2 reports, the post will be only me. |
| **Unlawful activity** | 1% - 2% | Decision: when the post on P.W and has 2% of reporting & ≥ 100 views, the post will be on friend’s diary.  Decision: when the post become on friend’s diary and has 2 reports, the post will be only me. |

Table 2 (Cases on a post: Alerts & Decisions)

**Copyright violation:**

Copyright violations are serious violations. Each case should be followed

instantly by the team, even if automated decisions exist. If one report is made

by the copy right holder, the report immediately become a case that needs

very quick follow up by the report officer, the case should be resolved within

less than 48 hrs.

This might include taking the post down without enough certain

evidence, if good amount of doubt has been established that the post creator

is not the creator of the content. The following window will appear when the

user clicks on the copyright violation:

You are not the copyright holder.



**Form of report of Copyright violation:**

Are you the copyright holder? .

If yes:

Note: Your name and identity will be use only and solely for the purpose of reporting and following this specific case.

Solely own  with other  held by your company

Name:

Copyright holder name:

Describe the copyright violation:

**A close up of a logo

Description automatically generated**

**Submit**

Form 1 (Report of copyright violation)

If users who are not copyright holders report a copyright violation, this the automated decisions:

Decision: when the post on P.W and has 10 reports, the post will be moved to

friends’ diary.

The system will send notification to the user that:

“Due to copyright violation claims, please note that your post titled: ……… has been moved from the public wall to friend’s diary, until we make sure that there is no copyright violation”.

Decision: when the post on friend’s diary and has 13 reports, the post will be

only me.

The team will send notification to the user that:

“ Due to copyright violation claims, please note that your post has been moved

from friend’s diary to only me, until we make sure that there is not a copyright violation.

* Anyone can report a copyright violation even if they are not a user.

**Automated decisions on cases with multi issues:**

This happen when more than one user report the same post in different issues,

ex: user 1 report that a post is misleading post, and user 2 report that the

post is inaccurate post, and user 3 report the post as hate speech post. As a

main rule for automated multi issues reporting the worst-case scenario will be

applied, accordingly: the rule of the lowest threshold (lowest number or

percentages rule among the multiple issues for developing alarm case) will

apply to the post under discussion. For example, in the previous example. The

rule of hate speech will be applied to that post.

The following table summarizes the rules for multi issues.

|  |  |  |
| --- | --- | --- |
| **Multi issues** | **No. of reports on Public Wall** | **No. of reports on Friend’s Diary** |
| Misleading + Inaccurate post | 10% of reporting & ≥ 500 views | 5 reports. |
| Offensive content+ Harmful language + Hate speech | 4% of reporting & ≥ 500 views. | 3 reports. |
| Nudity or sexually content | 5% of reporting & ≥ 500 views. | 3 reports. |
| Bullying and harassment | 20 reports. | 2 reports. |
| Suicide or self-injury | 2 report. | 1 report |
| Violent or graphic content + abuse | 4% of reporting & ≥ 200 views. | 2 reports. |
| Unlawful activity | 2% of reporting & ≥ 100 views. | 2 reports. |
| Copyright violation | 10 reports. | 13 reports. |

Table 3 (Rules for multi issues)

When reviewing any case with multiple issues more emphasis should be given

to reports associated with high Xs.

1. **Cases on accounts to be alerted about:**

Report account issues divided to fake account or identity theft, hacked

account, and Violent or dangerous organization.

When the user report accounts the team will review the history of the account.

Review past reports associated with that account. In the case of report account

the team might ask for permission to see and view the posts of user on

friend’s diary, and they must provide a written justification.

There is an alert for the team to review the account on the public wall and

friend’s diary when it reaches specific number of reports and becomes a case

and before it reaches the specific number of reports and become

unautomated decisions on cases. All decisions on accounts are made by teams’

members, and many times a committee decision is needed. However, for alerts

on accounts, which starts the follow up process by the team, automated rules

are applied. However, it should be noted that for hacked account there is no

alert as one report on an account will make the report a case for the team to

take a decision about.

**B.1 Alert for fake account or identity theft:**

For fake account or identity theft, if there were 3 - 4 reports on the public wall

and friend’s diary an alert will be establish for the team to review the case of

that account.

**B.2 Alert for Violent or dangerous organization:**

For Violent or dangerous organization account 5 - 9 reports on the public

wall and friend’s diary, it will be an alert for the team to review the case.

1. **Cases on accounts for a decision to be made:**

Cases on accounts should be established when reports reach a certain number

of reports. The team will review each case then they decide the decision. The

decision should be taken according to each case; no two cases are the same.

There are a group of guidelines for the team to help them in making the

decision. The following are the guidelines for each related issue:

**C.1 Fake account or identity theft: (time for decision)**

For 5 reports on fake account & identity theft issue a case will be established.

The following are guidelines for the team to follow while analyzing the case to

take the appropriate decision:

* The team must know if the account holder report that this account is fake or other users report that this fake account.
* The team must know when the fake account was created (Date).
* Impersonation (if the fake account represent true account that the username of fake account like the real username, or if the profile picture of fake account like the profile picture of real account, the post of user posted on public wall). So, the team must identify the true case/profile of real account and the fake account, and to review the accounts of real user on Facebook, Instagram, twitter, and they take into consideration:

1. To spot the difference in the way of posting, interacting, and the actions he/she is doing.
2. Any random unfamiliar actions that does not look alike with the true account.

* The team must view the IP address of fake account.
* The team will review and observe if there are 2 accounts with the same identity are active from 2 devices.
* The team will review the friends of alleged fake account.
* The team will review the behavior of alleged fake account:

1. The clicks of the account.
2. His/her interactions with others.
3. His/her posts.
4. His/her actions.
5. The fake username should include a dot, number… because the username cannot be repeated.
6. Read the person’s profile.
7. Few friends.
8. Date of creating the account, (to put things in order).

**Decisions**:

* If it is fake account: termination or shut down of fake account. Remember the email and the main IP address.
* It is not fake account: keep it; (to have the reporter activity in the activity log).
* Not able to decide:

-A committee will convene in 24 hrs to follow the case.

- The Committee reviews the case and decide what will happen.

- The committee need more resources that include (review more

information/posts from the public wall or friend’s diary) of the user.

-To view the API address of user.

* Send notification to the user that he/she logged in from unusual device.

**C.2 Violent or dangerous organization:**

For a case ofViolent or dangerous organization account to be established,

there should be 10 reports on the public wall or friend’s diary, it will be

unautomated decisions on a case and the team should make

comprehensive research on the case, then the team should decide as soon

as possible.

**C.3 Hacked account:**

One report is enough to establish a case of hacked account, it will be

unautomated decisions on a case and the team should make

comprehensive research reviews the hacked account report case and make

a decision.

Thefollowing are guidelines for the team to help them analyze the case and

take the appropriate decision:

* Email will be received from a user by My Diary (email that he/she use it in sign up or previous or current email) that his/her account has been hacked.

This email will receive the emails from the user to (Reporting @mydiarynotebook.org).

* The team use two-factor authentication and it should be by default for all users, to send to the user verification code in this case by his/her email/phone number.
* The team should provide written justification, to review the activity log

of the user and review the document logs of the user as PDF file.

* The team will review the devices the user used, and the last 3 passwords of his/her account or the ones he/she remember.
* The team must be alerted and put good emphasis on the time of when the hacking occurred.
* The team might ask about the log-in history of the user (methods of logged in, devices, location, time, failed attempt to logged in).
* In the case that the hacker deletes posts of the user, these posts should be recovered. Deleted posts should be saved for 30 days *( Note for technical),* So the user can restore his/her deleted posts on his/her account.
* The team might suspend the account of user through processing and send email to the user that her/his account had been suspended, and ask the user to provide a proof of identity (official document: passport, driving license) within 72 hours.

Notes:

* Send notification to the user that “You have logged from unusual devices, if it is not you please…. ”. when he click he will go to change password.
* Send notification to the user that please check your activity log.

**Two cases of hacked account:**

1. **Silent hacker:**

In this case the hacker logs in and he/she just monitor the account without doing any strange activities.

In this case the user should:

* Change his/her password and log in again and log out from other devices.
* MyDiary notebook team will send question for the user as email by the following:

**Questions help you if you think that your account got hacked:**

* What the date of hacking?
* Do you think that there are unfamiliar activities?
* Yes. - No.
* Can access your account?
* Yes. - No.

If yes : change your password and log out from other devices.

If No: send MyDiary notebook team proof of identity.

Form 2 (your account hacked)

1. **Deniable account access: (*need revision*)**

Is the hacker who logs into account not for him/her and he/she make unusual

activities on the account.

In this case the user should:

* Change his/her password and log out from other devices.
* The team will review this case very carefully and take the appropriate decision.

**Decisions**:

* To reject the claim of real user.
* To accept the claim and give the user new password to his/her account.
* No able to decide (the committee will review the case and decide).

**D) Cases on Comments:**

Cases on comments are divided to: Abuse, harmful language, hate speech,

suicide, self-injury, others, so when the user report comment, it will be

established as case, and the team decide the decision, so the following table

summarize the rules for each issue when the privacy of comments Me and

him/her or Public wall:

|  |  |  |
| --- | --- | --- |
| **Issues** | **No of reports when it (Me & him/her)** | **No of reports on Public Wall** |
| Abuse, harmful language, hate speech | 1 report the post will be only me | If the user has 4% of reporting & ≥ 200 views. |
| Suicide, self-injury | There is no specific number, only follow up by the team. | There is no specific number, only follow up by the team. |
| others | There is no specific number, only follow up by the team. | There is no specific number, only follow up by the team. |

So, after the reviewing and follow up by the team, they decide the

appropriate decision in each case.

**5. Decision for suspending the user’s account:**

To keep our platform protected against any violations of speech or action,

a decision to suspend accounts with inappropriate content or against

our community standards and terms of service and use might be taken, which

means that the user is not able to use his/her account as usual, permanently

or temporarily, so that the team can review his/her account and make sure

that there is no inappropriate content, then the user will be permitted to use

his/her account.

Before suspending a user’s account, there is an alert to become a case for

the team to review the posts on the P.W and the story of the anonymous

when it reaches specific reporting percentages and becomes a case.

|  |  |  |
| --- | --- | --- |
| **Issues** | **Alert for posts on the P.W** | **Alert for posts on the story of the anonymous** |
| Misleading & inaccurate post | If the user has 40% reporting on his/her last 10 posts. | If the user has 40% reporting on his/her last 13 posts. |
| Offensive content & harmful language | If the user has 30% reporting on his/her last 10 posts. | If the user has 30% reporting on his/her last 13 posts. |
| Bullying & harassment | If the user has 20% reporting on his/her last 10 posts. | If the user has 20% reporting on his/her last 13 posts. |
| Hate speech, Nudity or sexually content | If the user has 20% reporting on his/her last 10 posts. | If the user has 20% reporting on his/her last 13 posts. |
| Violent or graphic content, Abuse, Unlawful activity | If the user has 2 report on posts out of 7 posts. | If the user has 2 report on posts out of 13 posts. |

Table 4 (Decision for suspending the user’s account)

**6**. **Notifications sends to the user:**

The following notifications will be sent to the post holder that other users

report his/her posts/accounts which are inappropriate or against community

standards and terms of service and use. It should serve as a warning to the

user to avoid posting such posts again.

1. **When its X ( Misleading, Inaccurate post):**

Your post privacy has been changed to lower privacy exposure settings

by MyDiary Notebook because it is against our terms of service and

use, privacy policy and community standards, please give more

attention to the accuracy of future posts.

1. **When its XX (Offensive content, Harmful language):**

Your post got down by MyDiary Notebook because it is against terms of

service and use, privacy policy and community standards, please be

more considerate in your future posts.

1. **When its XXX (Bullying and harassment):**

Your post got down by MyDiary Notebook because it is against terms of

service and use, privacy policy and community standards, please be

more considerate in your future posts.

1. **When its XXXX (Hate speech, Nudity or sexually content):**

Your post got down by MyDiary Notebook because it is against terms of service and use, privacy policy and community standards, please be more considerate in your future posts.

1. **When its XXXXX:**

* **Violent or graphic content:**

*Action*…Your content was in true violation of terms of service and use,

privacy policy and community standards, please note further such act may

lead to suspension to your account.

* **Abuse:**

*Action*…Your content was in true violation of terms of service and use,

privacy policy and community standards, please note further such act

may lead to the suspension of your account.

* **Unlawful Activity:**

*Action*…Your content was in true violation of terms of service and use,

privacy policy and community standards, please note further such act

may lead to suspension to your account.

* **Copyright violation:**

Your post got down by MyDiary Notebook as it is a copyright violation and is against terms of service and use, privacy policy and community standards. Please take this point in consideration in future posts.

* **Violent or dangerous organizations:**

An email to be sent .

* **Fake account & Identity theft:**

MyDiary Notebook will send an e-mail to the user in this case that (Your

account has been terminated as it might be:

Involved in fake account creation.

Hacked by others.

Involved in identity theft.

However, your account now in violation of our terms of service and use,

privacy policy and community standards).

**7. How Case report appear for team:**

Case report give a display of the case characteristics, this includes:

* 1. Date of first report, last report.
  2. The status of the case (alert/Open case/Review Automated Decision).
  3. Issues related to the case.
  4. Number of reports.
  5. Percentages of reports per issue.
  6. Number of notes written per issue.

The table (5) shows how case report appears for the team when the user

reports a post which is Inappropriate, and he/she choose one of these issues

and the number of reports for each issue exceeds the specified alert

percentage to be a case. So, the table help the team in showing summary of

each case and its characteristics and follow on them.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **Alert** | **Number of reports** | **Percentage of reports** | **Number of users who wrote in (the tell us more box).** |
| Offensive content | 2% - 3% |  |  | when you tap on this button, you can view the (report) what the user wrote in this box. |
| Misleading | 5% - 9% |  |  |  |
| Inaccurate post | 5% - 9% |  |  |  |
|  |  | Total number of reports for all issues |  |  |

Table 5 (How case report appears)

**Percentage of reports = Number of reports for 1 issue**

**Total number of reports for all issue**

**Decision: Before making the decision, the following points should be taken into**

**consideration:**

* If reports represent the case.
* Review the claim of the users (verification).
* Based on claims of users, the team review the case then they decide the decision, however no claims are 100% accurate or 0% not accurate.

**8. Dealing with unexpected topics/issues/cases:**

These topic/issues/cases that we did not expect to happen and were not

covered in the reporting system and how to deal with and solve this unexpected

topics/issues/cases by the team of reporting. What you want to say…

**9. Design of the reporting system:**

The design of the reporting system about the pages and layout of the

reporting system, and what each page includes.

1. Page 1: sign in.
2. Page 2: sign up:
3. Page 3:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No. of case** | **Date & Time** | **Issue of the case** | **No of reports** | **Status: Need a decision/alert/Automated decision** | **Open/Closed case** | **Resolved/Reviewed by** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Table 6

When tap on the case, each case includes the following:

* **Number of cases:**
* **Date & Time:**
* **Time of first report:**
* **Time of last report:**
* **Number of reports:**
* **Report for account, post, comment:**
* **Issue/ multi issue:**
* **Issue and sub issue of the case:**
* **The reported:**
* **From which page the reported account/post/comment:**
* **Time of posting/commenting:**
* **Post/account/ comment as it:**
* **Notes what user write in (the tells us more box):**
* **Status: Need a decision/alert/Automated decision:**
* **Open/Closed case:**
* **Resolved/Reviewed by**
* **Decision:**
* **Justification:**
* **Future recommendation:**

**10.Reporting system team:**

A team should follow on the implementation of the reporting system.

**The human power associated with reporting and issues follow up through the**

**Reporting System (RS) should be arranged and established to carry and**

**achieve the following objective:**

**Objective for RS human capital:**

**Reports and cases should be followed 24 hours a day and 7 days a week, no**

**case should be neglected, delayed, or not be noticed an hour after it was**

**created.**

This team requires several employees who are responsible to review the

reports for posts or accounts.

The following is the working hours for the team, and there are 3 options, give

conditions to each one? :

1. **Have 3 shifts:**

* 8:30 AM – 5 PM (team).
* 5 PM- 1 AM ( 2 employee).
* 1 AM – 8:30 AM ( 2 employee).

1. **Have 2 shifts:**

* 8:30 AM – 5 PM ( team).
* 10 PM- 6 AM ( 2 employee).

1. **Have 1 shift:**

* Dana & Batool & Mohammad.

**11. Work plan:**

The Work plan of the reporting system is important for the team to know

their responsibilities and how tasks should be done with a specific time.

The work plan consists of number of tasks, task description, name of

employee who is responsible for the task, date of the task, if the task done or

in progress. To be developed one month after the launching.

1. **Guidelines for the team to follow the reporting system:**

Guidelines for the reporting team are rules and steps that are important for

the team to know how to deal with each report or cases that include

inappropriate/harmful content…, to keep our platform safe and make users

comfortable while using our platform.

1. **Documentation:**