HOW TO IMPROVE THE PROCEDURES ADOPTED BY THE CODE OF PRACTICE SIGNATORIES AS REGARDS CONSUMERS COMPLAINTS AND FLAGGING

(ERGA REPORT)

ERGA

European Regulators Group for Audiovisual Media Services

Table of Content

1. EXECUTIVE SUMMARY

The *"Empowering consumers"* pillar of the Code of Practice on Disinformation (hereinafter: the Code) puts platforms undertake to construct easily accessible tools for users to be able to report material which they define as disinformation. Besides the empowerment of the consumers, the Code under the *"Scrutiny of placements"* pillar-requires the closure of fake accounts and marking systems for automated bots, which aim is connected to flagging and complaint handling as these accounts pose a general problem and where automated monitoring is not able to eradicate such accounts, consumer complaints and reporting can also help to identify disinformation disseminating networks. In respect of these two pillars the platforms undertake to develop transparent, effective, understandable, and fully comprehensive procedures for consumer reporting.

The final 2019 Report of the European Regulators Group for Audiovisual Media Services (ERGA)[1] highlighted issues and shortcomings regarding the platforms' complaint handling procedures. Following from the Code and the Report, the main task of Focus Group 2 was to collect and analyse information about the processes the platforms use when they handle complaints (flagging) about disinformation. Seeking information, Focus Group 2 identified two sources which can give relevant information about the mentioned process. (1) the *written information from the website of the platforms* (Facebook, Google, Twitter) and (2) the *information from practice* - reports from stakeholders (complainants and complained).

In addition to focusing on generally available information across the Member States, we also conducted a detailed case study of one smaller Member State, Hungary. Since the Code must be complied with in all Member States, and the complaint handling procedures are conducted in a uniformed manner the shortcomings identified regarding Hungary may be similar to other Member States as well.

The observation is based on four different attributes of the processes. These attributes are created by inspecting the focal point of the complaint handling processes which are:

(I) *Transparency* (how transparent are the platform's complaint-handling system and procedure);

(II) *Clarity and accessibility* (the availability and comprehensibility of the complaint and reporting facilities, categories and related information to an average user);

(III) *Feedback* (the practice of platform feedback from receipt of the complaint to the final decision); and

(IV) *Reparation* (what remedies the platform provides against its decision).

[1] ERGA Report on disinformation: Assessment of the implementation of the Code of Practice https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf

After examining the three main platforms (Facebook, Google and YouTube, Twitter) it was possible to determine in which areas further improvements would be needed. According to the feedbacks collected and the results of the empirical study on the platforms the following shortcomings and recommendations can be identified:

(I) Transparency

Transparency cannot be achieved without knowledge of the platform's procedure and decision-making mechanism but this lacks in almost all cases.

Platforms should provide transparent description of their complaint-handling process in the terms of condition, provide adequate general information for the users on the available notification methods and their process, make the operation of the various automated content control systems transparent, and the principles applied should be made public, as well and the consistency of the procedures should be guaranteed so that all users and all content are judged according to the same principles by the platforms.

(II) Clarity and Accessibility

From the user's point of view it is crucial that the notification form and the relevant policies are easily accessible and clear this is usually done up to a certain point.

In order to reach clarity and accessibility, all information and flagging interfaces and messages must be in the notifiers' own language. Platforms should provide easily accessible and easy-to-understand information for all users and the notification interfaces must be user-friendly, easily accessible and easy to understand on all user-platforms (desktop site, mobile site, application, etc.).

(III) Feedback

Feedback is most often lacking, as is proper information on how and what remedies are available against the decision taken at the end of the proceedings, if there is any.

Platforms should provide feedback and adequate information to notifiers regarding the content they report at all stages of the flagging process as well as to those whose content has been complained about or flagged, at all stages of the investigation procedure. Platforms must make a serious effort to respond to all complaints without exception and measures taken in error should be easily remedied.

(IV) Reparation

According to reparation, measures taken in error should be easily remedied. If it is proven that a content has been removed incorrectly, it should be restored fully as soon as possible. Also, if a page or user is found to have been incorrectly blocked or banned, it should be reinstated as soon as possible. In rare, justified cases, if a platform has caused a demonstrable, material financial disadvantage, an obligation of compensation via service could also be considerable.

2. COMPLAINT HANDLING AND FLAGGING IN THE CODE OF PRACTICE

According to the purposes of the Code one of the most important part is to decrease the spreading of disinformation, which contains both the fight against disinformation material and also accounts (and bots) which are used to disseminate disinformation on such platforms. Under the Code the handling of disinformation is from one perspective the task of the signatories through algorithms and monitoring activity, however the Code stresses that the reporting of such material is also a central part of its purpose. This empowerment of consumers means the obligation to create user friendly ways of reporting, that can help people who face disinformation to participate in the process of inspecting and erasing such material.

As the "Empowering consumers" pillar of the Code (section II. D.) states, platforms undertake to construct easily accessible tools for users to be able to report material which they define as disinformation.

As the Code says *"The Signatories of this Code recognise the importance of diluting the visibility of Disinformation by improving the findability of trustworthy content and consider that users should be empowered with tools enabling a customized and interactive online experience so as to facilitate content discovery and access to different news sources representing alternative viewpoints, and should be provided with easily-accessible tools to report Disinformation, as referred to in the Communication".*[2] (II. D. section of the Code)

Both the Code and the previously formulated European Commission's Communication document titled as "Tackling online disinformation: a European approach[3]" states that the final intention of the Commission is to create a more transparent, trustworthy and accountable online ecosystem. This purpose contains the need to create the most efficient ways of user reporting systems, as the Code also highlights the need to widen the autonomy of users. This can only be fully achieved if consumers have the possibility to complain in cases where they encounter materials that spoil their right to trustworthy content.

Besides the empowerment of the consumers, the Code under the *"Scrutiny of placements"* section (II. A. section of the Code) requires the closure of fake accounts and marking systems for automated bots, which aim is connected to flagging and complaint handling as these accounts pose a general problem. Where automated monitoring is not able to eradicate such accounts consumer complaints and reporting can also help to identify disinformation disseminating networks.

As the Code says: *"In line with the European Commission Communication, the Signatories recognize "the importance of intensifying and demonstrating the effectiveness of efforts to close fake accounts" as well as the importance of establishing "clear marking systems and rules for bots to ensure their activities cannot be confused with human interactions"".* (II. C. section of the Code)

In respect of these two pillars (Scrutiny of placements; Empowering consumers) the platforms undertake to develop transparent, effective, understandable, and fully comprehensive procedures for consumer reporting.

[2] Code of Practice on Disinformation https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.

[3] Tackling online disinformation: a European Approach COM/2018/236 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236

3. ERGA REPORT ON COMPLAINT HANDLING AND FLAGGING - SUMMARY OF THE TASK OF THE WORKSTREAM 2 FOCUS GROUP 2

Since the adoption of the Code of Practice two ERGA reports have been completed. The intermediate report[4] consisted the monitoring of the "transparency of political advertising", where the issues of user-friendly content flagging appeared in the research regarding such materials. However, the monitoring excluded the inspection of the whole flagging process. The understandability of advertising identification, the identification processes, removing and re-labelling procedures showed several discrepancies. These defects were phrased more detailed in the final Report.

The final Report of the ERGA monitoring[5], published in May 2020, highlighted issues and shortcomings regarding the platforms' complaint handling procedures. However, no specific changes were observed in connection with the *"Scrutiny of placements"* since the interim report. Some findings emphasised that the removing and flagging of political advertisements were not specified by the platforms, which indicates, that the transparency and the requirement of feedback through the flagging process was incomplete at the time of the monitoring.[6]

In case of the *"Empowering consumers"* pillar the report provided further information. According to its findings the discrepancies of the flagging process raise difficulties in the reporting of problematic content, which includes that the decisions of the platforms are not always clear, satisfactory or transparent. The conclusions of the report also stress that the platforms' processes, communication and principles have severe differences which make the reporting of disinformation hard and impenetrable for the consumer. The report also raised the attention to the problem of the lack of uniformity which is necessary in order to fulfil the commitments of the Code on a wider scale.

With the findings of the *"Empowering consumers"* pillar monitoring, the report informed the Commission about the inefficiency of the platforms' practices and necessitate further monitoring of such tools. In this respect, recommendations where phrased by the ERGA Report, that aims the improvement of the monitoring regarding the existing Code's commitments[7]. Under this section, ERGA recommends that some sets of guidelines should be drafted with the aim to *"improving and harmonizing the platforms' reactions to consumer's complaints and flagging"* [8] . The Report emphasised here the need for standardised principles and that improved processes of flagging and consumer complaint handling would be essential to reach the maximum level of consumer empowerment. In order to make recommendations connected to the Code, further and more detailed inspections became necessary. As the recommendations formulated, further monitoring requires also the cooperation of the ERGA members, therefore, the creation of a subgroup was necessary in order to measure whether the platforms properly implemented the provisions of the Code or additional recommendations are essential.

[4] Report of the activities carried out to assist the European Commission in the intermediate monitoring of the Code of practice on disinformation (ERGA Report), June 2019. https://erga-online.eu/wp-content/uploads/2019/06/ERGA-2019-06_Report-intermediate-monitoring-Code-of-Practice-on-disinformation.pdf.

[5] ERGA Report on disinformation: Assessment of the implementation of the Code of Practice https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf.

[6] p. 17-19.

[7] p. 50-51.

[8] p. 50.

4. METHODOLOGY AND ACTIVITIES IMPLEMENTED

The main task of Focus Group 2 was to collect and analyse information about the processes the platforms use when they handle complaints (flagging) about disinformation. Seeking information, we identified two sources which can give relevant information about the mentioned process.
The two sources are:
- Written information from the website of the platforms (Facebook, Google, Twitter)
- Information from practice - reports from stakeholders (complainants and complained)
Since we could not examine every single major platforms' detailed mechanisms in every single Member State, we performed a general, comprehensive study, in addition, we conducted a case study focusing on a detailed analysis of their practices in one Member State, Hungary. In the course of our research, regarding all the available information of major platforms, we have come to the conclusion that the complaint handling procedures are conducted in a uniformed manner in Europe, so that, apart from any minor – mainly linguistic – differences, an in-depth examination of one Member State could lead to general conclusions.

4.1. ANALYSIS OF THE WEBSITE OF THE PLATFORMS

In the first round we collected and analysed all the information which was available at the website of a certain platform (flagging surface, community standards, terms and conditions, community guidelines etc.) and was deemed to be relevant concerning complaint handling.

As far as the methodology is concerned, we assumed that these information are common (every EU citizen meets the same rules in its own language), therefore, we analysed these websites from one member state perspective (Hungary). If the information was not available in Hungarian, then we used the English original version. This method also helped us to discover the discrepancies, namely what information is available in a certain member state and what are the shortcomings.

For the analysis we used the following main information categories:

- Is there an interface to flag complaints on disinformation or not?
- Is this interface available for a common user directly or indirectly?
- Is the description of the flagging/complaint process easily accessible, transparent and appropriately detailed?
- Is fact-checking part of complaint handling process?

These main categories were only the starting point of our analysis which means we go into a more detailed analysis in a certain category when we had the relevant information.

4.2. REPORTS FORM STAKEHOLDERS (COMPLAINANTS AND COMPLAINED)

In the second round we gathered information from the practice. We collected information from those who have experienced complaint handling processes of the platforms. This means that we not only asked stakeholders representing users/consumers and making complaints (consumer

associations, NGOs, government agencies etc.) but also those whose activity was flagged for disinformation. The collection involved the stakeholders of those countries (Greece, Hungary, Italy, Poland) which took part in the work of the focus group.

It was a general experience during the collection of information on the procedures concerning disinformation that they merge/part of the general complaint handling processes of the platforms.

5. FINDINGS

The observation was carried out by using four different attributes of the processes. The attributes were created by inspecting the focal point of the complaint handling processes. In the following both the shortcomings and the recommendations are explained from the perspective of these four attributes. Under these attributes the three inspected platforms (Facebook; Google and YouTube; Twitter) are discussed separately. These attributes are following:

I. *Transparency* (how transparent is the platform's complaint system and procedure)
II. *Clarity and accessibility* (the availability and comprehensibility of the complaint and reporting facilities, categories and related information to an average user)
III. *Feedback* (the practice of platform feedback from receipt of the complaint to the final decision)
IV. *Reparation* (what remedies the platform provides against its decision)

I. Transparency

When examining the operation of a social media platform, transparency is one of the most important requirements that we need to examine. With regard to the complaint handling process, the platforms mostly do not fully meet this requirement. This is because neither the detailed rules of their complaint procedure, nor the deliberation process, nor the standards considered are public. This experience is illustrated by the following examples:

Facebook:

Facebooks' community policies include the types of offensive contents which can (or should) be reported by its users, however, the process following the complaint is mostly unknown. It is unclear, what happens after the report, what are (if there are any) the deadlines for examining complaints, what are the exact principles used to weight offensive content, and whether there is a possibility of appeal if a user does not agree with the platform's final decision.

Facebook recently published its' strategy for stopping false news[9]. Despite the description of this strategy, there is no general description of the complaint handling process in any Facebook document in an exact way. The Community Principles contain the types of offensive content and what Facebook can do with them, a description of how to report is also easily accessible, but the platform does not provide information on the rest.

The *Hungarian Civil Liberties Union (HCLU)*[10] is a human rights NGO, which monitors legislation, pursues strategic litigation, provides free legal aid assistance, provides trainings and launches

[9] https://about.fb.com/news/2018/05/hard-questions-false-news/
[10] https://hclu.hu/en

awareness raising media campaigns in order to mobilize the public. In the opinion of the Hungarian HCLU organisation stresses that, the algorithm of Facebook presumably detects when a content is reported by several people over a period of time, and removes it without merely examining the content, which restricts the given user's freedom of speech on the platform. Removing content from Facebook is quick, but happens without justification, and any recovery is slow and contingent.

Google and YouTube:
Google provides a separate site for reporting infringing ads, which can be used to report issues with each of Google's platforms. In this site, in addition to shopping ads, the "fake news" option always appears under the "misleading content or scam" option. However, under Google's advertising policies, there is no information on what exactly the platform means by fake news. Among the readable descriptions, fake news is not named, it also appears only namely on the COVID-19 information interface.

Google ads also do not provide a detailed step-by-step description of how complaints are handled. There is a description of the information required, however, the process description and response time cannot be found in the Google Ads information site.

We can say that Google is very open about the way to delete a user. Google says: "If your content violates this policy, we'll remove the content and send you an email to let you know. If this is your first time violating our Community Guidelines, you'll get a warning with no penalty to your channel. If it's not, we'll issue a strike against your channel. If you get 3 strikes, your channel will be terminated." Users can read detailed information about this "strike method"[11]. However, no information is available on the follow-up to the notification, the deadlines and the feedback process. Thus, the background processes of complaint handling remain unknown.

Finally, users can also report expected terms to appear in YouTube Search: they request removal here if violation of the "AutoComplete Policy" appears. On the other hand, these guidelines do not include guidelines for misinformation.

There are a number of differences between Google ads and the YouTube reporting system, and it is difficult in practice to have YouTube ads (including pseudo news) on the google support page, but this is not directly accessible from YouTube. In the reporting ads on YouTube, fake news is not available and will not redirect the user to the Google ads.

Twitter:
The platform states in principle that it strives for transparency in its process, but does not state how it intends to do so. No specific deadline, procedure or method is indicated.

On the plus side, it publishes aspects of your consideration somewhere, but this can only be achieved in English and after several clicks. The platform does not write specific rules, but only defines in principle the basic objectives of its activities and the aspects taken into account during the consideration.

[11] https://support.google.com/youtube/answer/2802032

II. Clarity and accessibility

Clarity and accessibility are essential requirements for the user-friendliness of the platform. In this category, we examined whether the complaint options and rules provided by the platform are clear to the average user, and how easily the fake news policies (if any) are accessible and interpretable. Unfortunately, not all platforms provide the category of disinformation or fake news among the reporting options, so often the user has to decide which other category to classify a given offensive post.

Facebook:

At Facebook, reporting is done through online forms with different content, depending on the type of complaint or comment. Complaints about ads and reports and other content are distinguished. The blank is easily accessible from the top right corner of the content (by clicking on the "..." sign). Here many options appear, and on the bottom there is the "Find support or report post" choice. Users can report any kind of problems and through the main options "False news" appear. However, a disadvantage is that an explanation for each reportable category is not available here.

Among the platforms examined, Facebook had the most detailed and easiest-to-interpret description of fake news, which is available in the users' own language as well. In the "COVID-19: Community Standards Updates and Protections"[12]. Facebook gives information about fake news in Part IV. About False News Facebook says: "There is also a fine line between false news and satire or opinion. For these reasons, we don't remove false news from Facebook but instead, significantly reduce its distribution by showing it lower in the News Feed."[13]

Google and YouTube:

Google's policies regarding Google ads activity may include Prohibited Content and Prohibited Activity. These include counterfeit products and, for example, misleading. However, there is no point that is directly related to disinformation. Here is the category "deception" a separate point.

YouTube handles ads containing disinformation through a given blank, given by the Google AdSense site. However, YouTube does not include the same ad notification form, nor does it have a direct link to the site. On YouTube's own page, it is possible to report audiovisual content directly, by clicking on the "feedback" tab.

It is possible to report problematic content through this reporting page, on the side there is the clear "Send feedback" option. In addition to taking a screenshot of the problematic content, the legal help page also provides assistance and allows the user to report content, however, no "fake news" option is provided. The report can be submitted by clicking on the "There is a problem other than the above" tab. Information about making a report is available under the Policies and Security tab.

Opening the YouTube Help page, policies and information about disinformation are misleadingly found under the category of "Spam, deceptive practices & scams policies"[14]. All of YouTube's policies, information and help are available in the user's native language.

[12] https://www.facebook.com/communitystandards/
[13] https://www.facebook.com/communitystandards/false_news/
[14] https://support.google.com/youtube/answer/2801973?hl=en

Twitter:
The notification interface is available for each post separately with one click. On the other hand, there is no “disinformation” or “fake news” in the list of specific types of violations, so the content of the other categories must be explained. The basic rules for reporting violations are also available in Hungarian, but the other pages containing detailed rules are unfortunately not. The detailed rules of Twitter are available in a total of 17 languages, but many European countries have been left out. In addition to Hungary, this includes e.g. Poland, Slovakia, Slovenia, the Czech Republic and Croatia. Despite the basic “Twitter Rules”, the platform policies are limitedly available in languages other than English, making it difficult for users to learn and interpret the detailed rules.

Disinformation or fake news is not a separate category, but it is possible to classify the type of violation on the basis of the content of the communication: incitement to violence, incitement to hatred, manipulated content or content that violates the purity of elections. However, this must be decided by the user at the time of notification, but the platform does not provide sufficient information for consideration.

III. Feedback

Platform feedback is key to the user knowing their report was not in vain. There is an information obligation on the platform for both our notified user and the owner of the reported post. Unfortunately, it can be said, that reporters rarely receive a response to a reported complaint, while if an authority makes a report to a dedicated contact email address, platforms do respond. This is shown by the responses and experiences received, detailed below:

Facebook:
Facebook provides feedback on reported complaints and user violations under Support Inbox. It is indicated whether a report or complaint has been submitted. Despite Support Inbox and the fact that the reported complaint can be seen, no further information about the process or result can be found on the interface. Support box is a great interface for following your own reports, however, the user is not informed about the possibilities of appealing against the decisions here either.

Internet Hotline is a legal advisory service in Hungary (also an active member of the INHOPE organisation) aims to quickly remove illegal content found around the Web based on their reports. The operation of the Internet Hotline is primarily aimed at content that is harmful to minors. The hotline service was launched in 2005 in Hungary and has been operated by the Hungarian Media Authority since 2011.[15] The Hotline will first ask the user to report the harmful content to the social platform themselves, and if the user does not receive feedback within a few days, Hotline will contact the platform directly. Hotlines’ experiences are basically positive, they usually get feedback from the social platforms within a few days. In several cases, the platform will request more accurate information and then typically remove the content or resolve the issue indicated by the user. On the other hand, the Hotline also reported, that users have repeatedly indicated that even if they themselves report on social media sites, they do not respond, or find that the content is not infringing. There was also an example, where the reporting user received a feedback that the

[15] https://english.nmhh.hu/article/190105/What_is_the_Internet_Hotline

infringing content did not violate Community policies, however, after the Hotline reported it, the platform later deleted the content.

The Hungarian Internet Hotline reported that it had recently turned to Facebook several times for content suspected of child pornography that removed them within about a day, but the feedback came in an email only a few days later. Their basic experience is that the action is followed by feedback a few days later. It is a rarer case, but if there is no response to their report within a day or two and the matter is not resolved (e. g. the infringing content is still available), the alert will be repeated. The platform then typically responds and resolves the matter. Instagram communication is also similar to Facebook, in most cases collaborative. An example from Internet Hotline report was that they have reported several profiles on Instagram that have posted videos where children abuse, humiliate, shame each other. Instagram did not respond, but the profiles were removed two days later.

A special flagging system works in the case of Facebook and Instagram, where direct communication with the staff of the platforms is possible by a dedicated e-mail address, available only for specified authorities and organisations. The e-mail address is a contact created exclusively for services such as the Internet Hotline. This option makes their job much easier, as they can provide a detailed description of a problem in form of an e-mail, unlike an average user. It is important to note that this channel is specifically reserved for cases where the user has already used the reporting options available on the platform and has not received any feedback from the service provider, and if this is a more serious matter (such as child pornography content). This is also a request from Facebook, which has confirmed several times before: the e-mail address is used to report more serious, urgent matters. The Media Authority of NRW and the Media Authority of Hamburg/Schleswig-Holstein also reported their position as 'Trusted Flagger' on Facebook.

Google and YouTube:

According to the *Hungarian Internet Hotline*, YouTube usually removes the content that is reported, but its response comes later. In one of their cases, for example, they turned to YouTube and asked for removal of an unauthorized recording of a minor. Youtube deleted the content within 3 days.

The Hungarian Internet Hotline reported that since 2017, YouTube has provided an opportunity to participate in the *Trusted Flagger program*, under which the Hotline can report the infringed video directly on the YouTube interface, flagging even more content at once. In addition, they communicate with YouTube staff via e-mail, which e-mail address reserved for legal complaints and for civilians.

Twitter:

Among the rules of Twitter, it does not write specific rules, it only defines in principle the basic goals of its activity and the aspects taken into account during the consideration.[16] Therefore, no specific rules can be found for informing the reporting user and the owner of the reported content. At the same time, it can be said that Twitter will inform the affected user about the sanctions imposed on him and the remedies available to him/her in due time.

[16] https://help.twitter.com/en/rules-and-policies/enforcement-philosophy

IV. Reparation

The issue of reparation is closely related to the feedback and information obligation, both for the reported and the complainant user. In this category, we looked at the remedies available if the reported post is deleted or the account holder's account is restricted. When and where does the platform inform the user about these devices and to what extent are these options available at all? In our experience, most of the information is lacking here as well, and not all platforms provide a clear remedy for their decisions.

Facebook:

The *Hungarian* HCLU assumed, that an event they organised and publicly posted on Facebook, had been deleted from the platform, and the reasons for the deletion had not been communicated to the organisers. HCLU complained about this, but did not receive an answer to the question of how or by whom the event was cancelled.

The *Italian AGCOM* reported, that they received different complaints from consumers that denounced that Facebook and Google had removed their accounts with no reason. Therefore, AGCOM asked platforms for the reasons for the removal, but for different cases the removal was a mistake due to a wrong machine content control. AGCOM reported three specific cases of erroneous removals solved by its intervention. In one case AGCOM received a complaint from an Italian journalist, correspondent from Turkey, because his Facebook account was closed. He asked information from Facebook but after certain time without any change or answer, he finally wrote to AGCOM. AGCOM wrote formally to Facebook on the legal Italian mail of the platform, as they received a formal complaint from a journalist, who complained, that probably the closure of his account could be due to the contents he wrote form Turkey. After some days AGCOM received a letter from Facebook Ireland in which they said that the account had been closed for a mistake and that they had reopened it.

One other case reported by AGCOM was about a complaint from a citizen that criticised that his Facebook account was closed with no reason. In his case AGCOM asked informally to Facebook, and after some days the platform admitted, that the account had been closed for a mistake and that it had been reopened.

In Poland, several organisations fighting for consumers' rights, but the *Ministry of Digital Affairs* is specially created as a "contact point" for dealing with complaints against Facebook. The Ministry of Digital Affairs has signed at the end of 2018 a *Memorandum of Understanding* with the Facebook representative. Regarding to this document everybody whose account has been banned can fill in a special form on the Ministry's website[17] after an unsuccessful attempt or an unsatisfying result of a Facebook complaint. During the first year the Ministry of Digital Affairs has received 750 reports on deleted content or Facebook accounts, Facebook positively considered about 24 percent of all appeals lodged (representing almost half of the applications processed) and over 200 rejected. Polish consumer protection organisations (such as Office of Competition and Consumer Protection, European Consumers' Centre and Consumers Federation) said on request, that they receive rather

[17] https://www.gov.pl/web/gov/odwolaj-sie-od-decyzji-portalu.

small number of complaints on this topic. In the last year, they received 2 inquiries regarding the Facebook not allowing the opening of an account. One inquiry related to a consumer who allegedly provided false registration data, the other one concerned a situation where Facebook claimed that the consumer had already had an account.

Google and YouTube:
One case which the Italian AGCOM reported was about the YouTube channel of an Italian local radio, where the platform banned the radios' YouTube channel because of contents in violations of rules for the protection of minors monitored by the machine control. After a human control, Google verified that it was a mistake and the radios' channel was reopened.

Twitter:
A disabled or restricted profile can appeal the platform's decision.[18] Here the blocked user should describe the nature of the appeal and provide a contact information (e-mail address or phone number) so that later can be notified of the outcome of the appeal.

6. RECOMMENDATIONS

Along the issues identified based on our research, we formulated some suggestions that can help correct the shortcomings that have arisen. These were divided into four main groups based on the main shortcomings. These are general suggestions, as they could be applied universally to all platforms.

Transparency

- Platforms should provide transparent description of their complaint-handling process in the **terms of condition**, which includes in detail:
  - The structure of each type of complaints and flagging procedures
  - The conditions and rules for each type of consumer complaints and flagging procedures
    - For example: which points of the terms and conditions could be infringed by disinformation, what may be considered disinformation, what content cannot be reported in a similar way, etc.
  - The process and deadlines for the assessment of each complaint
    - Based on our research, and the nature of disinformation, we suggest that these deadlines should be around the maximum of 72 hours. This timeframe gives enough time for the platforms to thoroughly asses the complaint, but short enough to provide quick remedy for the complainant against the harms of disinformation.
  - The exact possible consequences of these procedures

- The platforms should provide adequate **general information** for the users on the available notification methods and their process.

[18] https://help.twitter.com/forms/general?subtopic=suspended.

- The platforms should make the operation of the various automated content control systems transparent.
- The principles applied in the different content reporting and control procedures should be made public and the consistency of **the procedures should be guaranteed** so that all users and all content are judged according to the same principles by the platforms.

Clarity and accessibility

- All information and flagging interfaces and messages must be in the **notifier's own language**
- Platforms should provide **easily accessible and easy-to-understand** information for all users:
    - About each flagging method and procedure
    - The nature and distinction of each flagging category

- **Notification interfaces** must be user-friendly, easily accessible and easy to understand:
    - The reporting interface should be available **directly next to the objected content** so that the user can report the content as soon as he encounters it
    - The reporting interface should be easy to use, with clear indications and options
    - The possible reasons for the flagging should be **easily identifiable** on the reporting interface, where the user may select more than one category (this may be limited to a maximum of 2 or 3 categories to avoid possible abuse)

- Notification and information interfaces should be easily accessible and understandable **on all user-platforms** (desktop site, mobile site, application, etc.).

Feedback

- Platforms should provide adequate information to notifiers regarding the content they report at all stages of the flagging process, which means:
    - Providing appropriate feedback to the notifier **at the time of notification**, including:
        - description of the further stages of the procedure, in particular the time frames
        - the way in which the notification is processed, in particular, whether the notification is evaluated by an automated system or by human intervention
        - the possible consequences of the notification
        - any interim measures relating to the content in question, in particular when temporary removal may take place
    - **After evaluating the reported** content, sending appropriate feedback, which includes:
        - the result of the investigation, in particular about the possible consequences of the flagging
        - the evaluating methods used in the procedure

    - The information may be provided by automatic messages or notifications, but it may even be considered to provide each complaint with a unique identifier code or number, which can be used to find out about the status of a given complaint procedure on a central interface.

- The platforms should provide adequate information *to those whose content has been complained about or flagged*, at all stages of the investigation procedure, which means:
    - Sending appropriate information **at the beginning** of the procedure, including:
        - Exactly what content was reported
        - Which clause of the terms of use is examined in connection with the given content
        - What are the possible consequences of the procedure
        - What are the remedies available at this stage of the procedure
        - Description of further stages of the procedure, in particular about the time frame and deadlines
        - The way in which the notification is processed, in particular, whether the notification is checked by an automated system or by human intervention, or whether the platform may contact a third party (e.g. fact-checker)
        - Any temporary measures and the reasons why they are applied by the platform

    - Sending appropriate information **at the end of the procedure**, including:
        - The result of the procedure
        - Any specific measures taken
        - The exact justification for the connection between the content and violated terms of use. That is, why the content was classified as infringing in a given case.
        - Available remedies, in particular the possibilities of appeal after a possible ban from a given platform, with direct link to the available remedies

Reparation

- Platforms must make a serious effort to respond to **all complaints without exception**.
- Measures taken in error should be **easily remedied**:
    - If it is proven that a content has been removed incorrectly, it should be restored fully as soon as possible
    - If a page or user is found to have been incorrectly blocked or banned, it should be reinstated as soon as possible, especially with regard to previous friends and followers, previously uploaded content, and other profile interactions.

In rare, justified cases, if a platform has caused a demonstrable, material financial disadvantage (e. g. loss of revenue due to an erroneously cancelled event, costs incurred in connection with an erroneously blocked advertisement etc.), an obligation of compensation via service (e. g. free advertisement, bigger audience reach etc.) could also be considerable.