CODE OF CONDUCT ON COUNTERING ILLEGAL HATE SPEECH ONLINE +

a) The first Code of Conduct on countering illegal hate speech online was adopted on 31 May 2016 by the Commission and Google (YouTube), Facebook, X (formerly Twitter) and Microsoft-hosted consumer services, as relevant. Between 2018 and 2022, Instagram, Dailymotion, Snap Inc., Jeuxvideo.com, TikTok, LinkedIn, Rakuten Viber and Twitch subscribed to the commitments of the Code (hereafter all the companies subscribing to the Code will be referred to as the "the Signatories").

b) Over these years, the Code facilitated key progress, including on the swift review and removal of illegal hate speech content and increased trust and cooperation between the platforms, civil society organisations and Member States authorities in the form of a structured process of mutual learning and exchange of knowledge.

c) Conscious of the significant opportunities and challenges represented by fast technological development and the evolutions in the policy and legal frameworks in the European Union, the Signatories continue sharing a collective responsibility and pride in promoting the respect of fundamental rights, as enshrined in the EU Charter of Fundamental Rights, in the online environment, as well as the Commission's and EU Member States' commitment to tackle illegal hate speech online.

d) The Signatories recognise the importance of transparency, monitoring and accountability on their systems and processes as regards the way they deal with alleged illegal hate speech content available on their services, and in particular in relation to their notice-and-action mechanisms. Also, as shown by the success of the collaborative approach under the Code of Conduct since 2016, the Signatories recognise the importance of a multi-stakeholder approach involving representatives of civil society organisations, experts as well as public authorities in the joint efforts to improve content moderation policies, with the aim to assess and mitigate the risk of the public dissemination of illegal hate speech through their services.

e) The Signatories note that the regulatory framework has advanced with Regulation (EU) 2022/2065 (Digital Services Act, hereafter referred to also as "DSA")[1] to increase transparency and accountability for intermediary services, informed also by the important achievements of the 2016 Code of Conduct. They recognise the need to further strengthen the Code, building on the DSA, in order to increase its efficiency in fighting illegal hate speech online.

f) This Code aims to become a voluntary code of conduct under Article 45 of the DSA. It is in this spirit that the Signatories have agreed on this **Code of Conduct+** (hereafter referred to also as "the Code"), identifying voluntary commitments aimed at creating a framework that facilitates the compliance with and the effective enforcement of the DSA in the specific area of illegal hate speech content, including new measures to address the most recent challenges and threats.

g) This Code of Conduct is without prejudice to the obligations imposed on Signatories under the DSA, where applicable. The DSA will always take legal precedence for Signatories falling under its scope.

h) Where commitments overlap with legal obligations that apply to some or all Signatories under the DSA, this Code aims to gather specific and proportionate information on how obligations are implemented with respect to illegal hate speech, in particular through the transparency provisions in section 3 below.

[1] Regulation(EU) 2022/2065 of 19 October 2022 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services

i) Adherence to and compliance with commitments and measures agreed under voluntary codes of conduct by a very large online platform or a very large online search engine where relevant may be considered as appropriate risk mitigation measures under Article 35(1)(h) of the DSA.

j) To facilitate participation in the Code, Signatories subscribe to implement the relevant commitments through measures that are reasonable, proportionate, and effective in light of the identified risks, the size and nature of their subscribed services, the risk of illegal hate speech on their platforms, the resources available to them and other relevant factors.

k) The European Commission welcomes the efforts made by the Signatories and will support and monitor the implementation of the **Code of Conduct +** in the coming years.

***

The Signatories of the **Code of Conduct+**[2] have agreed on the following public commitments:

1. Terms and conditions for addressing illegal hate speech[3]

1.1 The Signatories are to have in place terms and conditions informing users that they prohibit illegal hate speech on their services. The Signatories are to include in such terms and conditions clear information on the policies around illegal hate speech[4] on their services, and the measures taken in case of breach of these policies.

1.2 The Signatories will inform their users about significant changes made to their terms and conditions that are relevant to how they tackle illegal hate speech, in accordance with Article 14(2) of the DSA.

2 Review and possible removal of or disabling access to illegal hate speech content

2.1 Pursuant to Articles 16 and 22(1) of the DSA, the Signatories will have in place notice and action mechanisms to allow any user in the EU, including Trusted Flaggers, to notify them of the presence on their service of specific content that the user considers to be illegal hate speech content.

2.2 After receiving a valid notice, the Signatories will review it in a timely, diligent, non-arbitrary and objective manner and act expeditiously to remove or to disable access to the reported content if it is in violation of the Signatories policies, and/or on the basis of the applicable law in accordance with the relevant jurisdiction and in accordance with Article 16 of the DSA.

2.3 The Signatories commit to review the majority (at least 50%) of notices received under Articles 16 and 22 of the DSA from Monitoring Reporters within 24 hours. Signatories will apply their best efforts to go beyond this target and aim for two-third (at least 67%) of those notices subject to Annex 1. Signatories are free to set up the processes they see fit to reach this target, such as designated reporting processes. This acknowledges that the time it takes to action a notice may depend on contextual factors, such as the volume, severity, virality, and complexity of the content flagged.

3 Transparency, accountability and monitoring.

---

[2] The Signatories of the Code of Conduct+ are: Facebook, Instagram, YouTube, X (formerly Twitter), TikTok, LinkedIn, Twitch, Viber, Dailymotion, jeuxvideo.com, Snap Inc. and Microsoft-hosted consumer services, as relevant.

[3] "Terms and conditions" hereafter refers to terms of service as well as any applicable policies (such as Community Standards, Guidelines, etc). The Signatories acknowledge that there are a variety of valid approaches to crafting Terms and Conditions, e.g. distinctions in how they address illegal content vs. lawful content that violates services' policies, and the Signatories accordingly have discretion in how they comply with commitment 1.1.

[4] Illegal hate speech as defined by applicable laws, including the Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, as transposed in national jurisdictions, as well as possible forthcoming updates to this Framework Decision, where relevant.

3.1 The Signatories agree to assess the adherence to the specific commitments on the review of notices under 2.3 and to monitor trends over time, following a methodology agreed between the Signatories and the European Commission (See Annex 1).

3.2 Beyond the monitoring on the review of notices under 2.3, the Signatories commit to provide additional information accompanying their results of each monitoring exercise, as referred to in Annex 2.

4 Intra-industry and multi-stakeholder cooperation to address and prevent risks of the spread of illegal hate speech

4.1 The Signatories will continue to implement and strengthen, where relevant, their partnerships with Monitoring Reporters, as defined in Annex 1.

4.2 The Signatories will participate in regular exchange fora aimed at facilitating an effective flow of best practices, expertise, and tools on measures to address the dissemination of illegal hate speech content through their services. These fora will be made up of:

4.2.1 An annual convening of Signatories and Monitoring Reporters organised by the European Commission with input from the Signatories.

4.2.2 An exchange of best practices among the Signatories, materialised in a separate session during the annual convening in 4.2.1.

4.2.3 A structured dialogue among Signatories and Monitoring Reporters to exchange information on trends and developments related to illegal hate speech online. This may be aided by a set of standard questions to answer when sharing a trend with the Signatories.

4.2.4 A shared online knowledge hub facilitated by the European Commission and with adequate safeguards to protect information, gathering jurisprudence and legal resources useful to assess the legality of alleged illegal hate speech, and/or past reports under the Code. Signatories or the Monitoring Reporters may contribute through best practice resources, including relevant information on the challenges and opportunities presented by emerging technologies in both the generation and detection of illegal hate speech.

4.3 In the exceptional event of a material risk of a significant increase in the dissemination of illegal hate speech on the Signatories' platforms in the Union, the Signatories and the European Commission might agree together to initiate a meeting under the umbrella of the exchange fora per 4.2 (or more, should it be necessary and proportionate) to discuss trends, exchange information, and explore avenues of cooperation. Such a meeting can only be initiated by the Commission and the relevant Signatories together.

4.3.1 Signatories' commitment to participation is contingent upon their level of exposure to the imminent and material risk of a significant increase in the dissemination of illegal hate speech on their platform. The Signatories and the European Commission can decide together whether or not to involve Monitoring Reporters in those discussions.

4.3.2 Meetings as described in 4.3 will not be initiated if the Signatories are already engaged in meetings on the same subject matter with the Commission under Art. 36 DSA.

4.4 Signatories commit, including through existing initiatives, to exchange between their relevant teams, where appropriate, about cross-platform operations and relevant trends that emerge on their respective services, with the aim of preventing the dissemination of illegal hate speech on other services, in full compliance with privacy legislation and with due consideration of security threats and fundamental rights threats.

5 Awareness raising, online civility and counter narrative initiatives

5.1 Recognising the value of raising awareness against hateful rhetoric and prejudice and to foster online safety and civility, the Signatories will continue supporting tools and approaches on counter and alternative narratives, new ideas and initiatives as well as educational programmes that encourage civility online and critical thinking. These will complement initiatives funded and facilitated by the European Commission, such as in the context of the current Citizens, Equality, Rights and Values Programme[5] and the High Level Group on combating hate speech and hate crime.

5.2 Signatories will make best efforts to provide information about training, resources and support available to Monitoring Reporters and any relevant civil society organisations regarding best practices and how to use online platforms for delivering educational and awareness raising campaigns.

[5] And relevant future EU funding programmes.

Annex 1: Methodology for the Monitoring Exercises.

Introduction

a) This document codifies the monitoring exercise and ensures clarity of process, consistency with the DSA, and strengthened transparency.

b) The results of each Signatory's monitoring exercise have to be seen in combination with the additional information on the measures taken to address illegal hate speech based on the Signatory's content moderation policies, proactive efforts, and enforcement actions provided according to the guiding questions set out in Annex 2, where relevant.

Definitions

- **Monitoring Exercise**: The Monitoring Exercise aims to evaluate the performance of the Signatories in respect to the commitments set out in paragraph 2.3 of the Code. It starts with the beginning of the Monitoring Period and ends with the report publication.

- **Monitoring Period**: Annual period of maximum six weeks (30 working days) during which the Monitoring Reporters report alleged illegal hate speech content to the Signatories as part of the Monitoring Exercise.

- **Trusted Flaggers**: Entities whose status of trusted flagger has been awarded according to the provisions set out in Article 22(2) of the DSA for their expertise and competence in notifying illegal content and whose information have been published in the publicly available database referred to in Article 22(5) of the DSA.

- **Monitoring Reporters**: Not for profit or public entities with expertise on illegal hate speech in at least one EU Member State and approved by the Commission and the Signatories to participate in the Monitoring Exercise. Monitoring Reporters can also be Trusted Flaggers as set out in Article 22 as relevant. They may also take part in any relevant voluntary Signatory program, operated at the discretion of said Signatories.

- **Monitoring Reporters List**: Approved entities shall be added to the Monitoring Reporters List according to the following confidential procedure[6]: the Commission will share with the Signatories a suggested Monitoring Reporters List at least 10 weeks prior to the start of each Monitoring Period. The Signatories will then have 2 weeks to provide any comments either verbally or in writing. The absence of comments or objections from the Signatories to the List within the 2 weeks shall be deemed as approval of the List. In case of disagreements between the Signatories and the Commission, the parties have one week to come to a compromise. If no compromise can be found, the Signatories shall have the final word on the list of Monitoring Reporters, but must provide reasons for any exclusion[7].

- **Third party coordinator**: A third party entity selected by the Commission which may take an administrative role in helping the Commission to identify potential new Monitoring Reporters and providing the Commission with clerical support during the Monitoring Exercise. Any third party coordinator should be approved in writing by the Signatories after notification by the Commission. Failure from the Signatories to react within 2 weeks after notification by the Commission shall be deemed as approval of the proposed third party coordinator. Any fees

[6] The procedure for agreeing to the Monitoring Reporters list shall remain confidential between the Commission and the Signatories to preserve the integrity and impartiality of the process and ensure that the Signatories are able to share sensitive information with the Commission. All communications during this procedure shall remain confidential as between the Signatories and the Commission and shall be subject to the rules on confidentiality pursuant to Article 84 of the DSA.

[7] Entities which have been awarded the status of Trusted Flagger under Article 22 of the DSA shall not be excluded from the Monitoring Reporters list.

incurred will be undertaken by the Commission. The Signatories and the Commission will agree on terms of reference should such a coordinator be onboarded.

- **Disputed Cases**: Disputed Cases can occur when there is a disagreement between a Signatory and a Monitoring Reporter on the deemed illegality of a reported case or a disagreement between the Signatory and the Monitoring Reporter as to whether the content was removed or made inaccessible on the service during the Monitoring Period.

- **Error Cases**: Error Cases occur (a) when the same piece of content is reported several times (duplicates), (b) when the wrong account/page is flagged, (c) when there is an error in the URL or related screenshot provided, (d) where there is no evidence that a notice was received by the Signatory, (e) where the notice was not valid as per article 16 or 22 of the DSA, as applicable.

Monitoring exercise methodology

1. Frequency of the Monitoring Exercises and the Signatories' participation

1.1. The Monitoring Exercise is to take place annually, following a standardised and cycle-based approach. The regularity is justified by the need to provide for enough time for the Signatories to consider and potentially strengthen their approach to mitigating the risk of illegal hate speech on their platforms which may be informed by the results of the Monitoring Exercise.

1.2. All Signatories shall accept to be subject to the monitoring exercises and no exemptions can be granted. Any new Signatory joining the Code will benefit from one "trial monitoring exercise" (usually the one following the date of joining), during which the procedure and the results only serve as a test. It will be at the discretion of the new Signatory to make publicly available the results of the trial monitoring exercise.

2. The role and responsibilities of Monitoring Reporters

2.1. Monitoring Reporters are responsible for recording results and in doing so will provide, for each piece of content they report during the monitoring period, a unique URL and a screenshot, and specify what part of the content they allege is illegal hate speech (e.g. video, post, hashtag, etc.). Error Cases should not be accounted for as illegal hate speech in current and future publicly reported categories and as such will not be captured in final results. Where a Monitoring Reporter has submitted a significant number of Error Cases (more than 33% of their submitted notices to any of the Signatories in a single Monitoring Period) the Commission, after hearing the Monitoring Reporter and consulting the concerned Signatory/ies, shall suspend the Monitoring Reporter's participation in forthcoming exercises until further notice, unless reasonable explanations for and tangible commitments to address these Error Cases are provided.

3. Reporting illegal hate speech using designated reporting processes

3.1. During the Monitoring Period, Monitoring Reporters are to report content that, to their best knowledge, is illegal hate speech and found on Signatories' services, using designated reporting processes when and where applicable. Without prejudice to Article 22(1), this is to ensure that DSA reporting processes are genuinely being tested and the integrity of the results presented. The Code imposes no obligation on the Signatories to provide a designated reporting process for Monitoring Reporters (and if no such designated reporting process is provided, the Monitoring Reporter shall use the process available for ordinary users). Designated processes will be communicated by the Signatories to the Commission, who in turn, will provide information about these processes to the Monitoring Reporters.

4. Ordinary user reporting

4.1. Monitoring Reporters can, in addition to the designated reporting processes, or where such processes are not available, choose to report content that according to the best of their knowledge is illegal hate speech in their respective countries using reporting processes available for ordinary users (i.e. illegal content reporting processes pursuant to Article 16 of the DSA or combined processes for illegal and policy-violating content where such combined processes are made available by the relevant Signatory). These cases will appear as a separate category in the reporting about the Monitoring Exercise. Monitoring Reporters are strongly encouraged to use reporting through designated reporting processes, where they are available. The Monitoring Reporters will receive information on the Signatories' reporting processes in the trainings mentioned in section 6 below.

5. Disputed Cases

5.1. Without prejudice to Art. 20 and 21 DSA, should there be Disputed Cases between Monitoring Reporters and Signatories in the context of the monitoring exercise, the Signatories will determine the best route to facilitate their resolution. That could include additional legal review(s) in order to determine the illegality of the piece of content and/or individual outreach to Monitoring Reporters on the specific case(s), expecting Monitoring Reporters to actively cooperate in the process and respond as promptly as possible, and not excluding the possibility that the Commission may be called to facilitate the resolution of these disputes. In no circumstance can the Commission (or any third party coordinator) be asked to have a role in qualifying the possible illegality of a disputed content. As per the parameters of this Code, content that is not deemed illegal will be out of scope of Monitoring Exercises.

5.2. If disagreements remain, Disputed Cases will appear as a separate category of the report. This is to ensure the report accurately reflects the results of the Monitoring Exercise and does not conflate Disputed Cases with alleged illegal hate speech.

5.3. Disputed Cases might be further discussed in the annual meeting foreseen under the Code (see Section 4 of the Code). Where a Monitoring Reporter has submitted a significant number of unresolved Disputed Cases (higher than 70%) for two Monitoring Periods, the Commission, after hearing the Monitoring Reporter, and consulting the concerned Signatories, may decide to suspend the Monitoring Reporter's participation in forthcoming exercises until further notice.

6. Data consolidation

6.1. Ahead of the Monitoring Exercise, the Commission (with the possible administrative support from a third party coordinator) will provide the Monitoring Reporters with training on quality notices and data collection, with input from the Signatories.

6.2. The Signatories will be notified of the end of a monitoring period, which will be followed by a data consolidation period (to be defined each year) coordinated by the Commission (with the possible administrative support from a third party coordinator) in order for Signatories to analyse and validate the provisional results presented (including identifying any Error Cases or Disputed Cases), engage with Monitoring Reporters on points of discrepancy and allow sufficient time to discuss Disputed Cases.

6.3. If, during a given Monitoring Period, the combined number of Error Cases and Disputed Cases of a single Signatory is higher than 75% of the total notices sent, the Commission shall communicate accordingly in the publication of the results. This may preclude the publication of the Signatory's results as referred to in point 7 in view of the fact that the results may be considered as not reliable.

7. Report publication

7.1. After the data consolidation, the Commission will publish the final results of the Monitoring Exercise within three months following the end of the monitoring period. The Commission will

share an advance copy of the final results with the Signatories at least two weeks ahead of the publication of the report. The final results will be computed by the European Commission and include:

7.1.1. Overall Figures :

7.1.1.1. Number of notices from Monitoring Reporters received;

7.1.1.2. Number of notices from Monitoring Reporters who are not Trusted Flaggers;

7.1.1.3. Number of Monitoring Reporters involved, including details on languages and Member States covered;

7.1.1.4. Average percentage number of notices across Signatories removed by all platforms involved in the Monitoring Exercise;

7.1.1.5. Average percentage number of notices across Signatories reviewed within 24h and within 48h;

7.1.1.6. Average distribution of grounds of illegal hate speech, based on data provided by Monitoring Reporters.

7.1.2. Figures per Signatory:

7.1.2.1. Number of notices from Monitoring Reporters received (split between designated and ordinary reporting processes as set out above, where relevant);

7.1.2.2. Number of notices per type of content, if and as applicable (video, image, text, other);

Review time

7.1.2.3. Percentage of notices (disregarding Error Cases and Disputed Cases) reported by Monitoring Reporters through designated reporting processes reviewed within 24h and within 48h;

7.1.2.4. Percentage of notices (disregarding Error Cases and Disputed Cases) reported by Monitoring Reporters through the ordinary user reporting processes reviewed within 24h and within 48h;

7.1.2.5. Percentage of Disputed Cases reported both through the designated reporting processes and through ordinary users reporting processes reviewed within 24h and within 48h, per reporting channel.

Removal (or equivalent enforcement action taken – to be specified if possible):

7.1.2.6. Percentage of notices (disregarding Error Cases and Disputed Cases) reported by Monitoring Reporters through designated reporting processes removed;

7.1.2.7. Percentage of notices (disregarding Error Cases and Disputed Cases) reported by Monitoring Reporters through the ordinary user reporting processes removed;

7.1.2.8. Percentage of notices from Monitoring Reporters received through the designated reporting processes (where available) and the ordinary user reporting processes which are Disputed Cases, per reported channel;

7.1.2.9. Percentage of cases initially qualified as Disputed that reached a resolution in terms of illegality/legality in the dialogue between the Monitoring Reporters and the Signatories.

Feedback

7.1.2.10. Percentage of notices (disregarding Error Cases and Disputed Cases) reported by Monitoring Reporters that received a response from the relevant Signatory, including a breakdown for notices reported through ordinary users reporting processes and through dedicated reporting processes where available.

8. Target

8.1. The targets are referred to in paragraph 2.3 of the Code. Signatories are free to set up the processes they see fit to reach this target, such as designated reporting processes. This acknowledges that the time it takes to action a notice may depend on contextual factors, such as the volume, severity, virality, and complexity of the content flagged.

8.2. The Signatories and the Commission agree to consider Monitoring Exercise results as part of a wider picture, in the context of past results, the sample size of notices received, where relevant the contextual information provided in response to the questions in Annex 2, and the engagement and cooperation between Signatories and Monitoring Reporters, when considering the implementation of the commitments in the Code.

9. Review

9.1. The Signatories, by majority, by their own initiative or following a prompt from the Commission, can review the methodology described in this Annex 1 when it would benefit from changes as Signatories' practices and approaches evolve and in view of technological, societal, market and legislative developments. Without prejudice to Art. 45 DSA, this review will be conducted jointly by the Signatories and the Commission, with input from Monitoring Reporters as relevant and should, to the extent possible, occur no sooner than 12 months following the results of the first monitoring exercise under this revised version of the Code.

Annex 2 – Key performance indicators and additional qualitative and quantitative information on hate speech content moderation

a) The Signatories will undertake to provide, where relevant, pertinent, and legally possible additional meaningful and comprehensible information on the measures taken to address illegal hate speech as part of their content moderation policies. This may include measures taken to provide training and assistance to persons in charge of content moderation, the measures taken that affect the availability, visibility and accessibility of information and the trends detected.

b) This information is to be provided in a short summary drafted and provided by the Signatory, accompanying their results of each monitoring exercise (hereafter: Summary Document), and will be based on the structured set of reference points set out in this Annex. The Signatories commit to providing information to these points where relevant and pertinent to their platforms and to the best possible knowledge and information available.

c) The points below refer to 'hate speech'. Signatories can provide this information in relation to either illegal hate speech and/or content in violation of their relevant policies (including based on reports received), provided that they clearly state what is the scope of the information provided.

d) With due regard to the differences between each of the Signatory's platforms, in providing this information Signatories should only provide information to the selection of points in this Annex that are relevant to their platform[s], via a written statement of no more than four pages in length accompanying the results of their monitoring exercise. Signatories are advised that the relevant reporting period for the points in this Annex shall correspond to the previous calendar year or previous six month period, and be aligned with the most recent reporting periods used in the context of Articles 15, 24 and 42 of the DSA, as applicable, and acknowledging that each Signatory can refer to information provided in this context.

1. Terms and conditions for addressing illegal hate speech

1. [Code: 1.1] Link(s) to the Signatory terms and conditions addressing illegal hate speech;

2. [Code: 1.2.] Information on any significant changes to Signatory terms and conditions relevant to how they define and tackle illegal hate speech since the last monitoring exercise, including how these changes were communicated to the users.

2. Review and possible removal of or disabling access to illegal hate speech content

3. [Code 2.1] Information on whether and how the Signatory enables reporting of hate speech content, specifying which languages and in which Member States;

4. [Code 2.1] Information on the path for users to submit a hate speech notice, distinguishing by access point (mobile app/desktop app/ browser) as well as grounds of notice (the law (Article 16 of the DSA or terms and conditions), if different;

5. [Code 2.2] Information about the moderation path for hate speech notices, distinguished by grounds of notice (the law (Art. 16)/ T&C);

6. [Code 2.1, 2.2 and 2.3] Additional details on the variety of measures, including proactive measures, that a Signatory takes with the aim of detecting and reviewing hate speech on its platform[s]. This can include, to the extent it is available, information on content detection and reach, review flows, and appeals, such as for example:

    a. The proportion of hate speech content detected by automatic detection tools of the estimated total of hate speech content found on their platform in the EU – in the previous calendar year or previous six month period;

b. Additional details on the interaction between automation and human review in content moderation, such as information on hate speech content reviewed by humans and hate speech content reviewed solely by automated means in the EU – in the previous calendar year or previous six month period;

c. The proportion of actioned hate speech content which was appealed by the users in the EU and the extent to which appealed content was restored – in the previous calendar year or previous six month period;

d. Any additional details on review time of notices in the EU (such as for example % within 24h, 48h, 72h etc) – in the previous calendar year or previous six month period;.

e. Any additional details on relevant information or metrics related to hate speech on the Signatories' platforms, for example, where available, on the role of recommender systems and the % of illegal hate speech that was removed before accumulating any views.

7. [Code 2.2 and 3.2] Any additional information outlining how the Signatory moderates its platform[s] for content violating its policies on illegal hate speech. Additional information could also include more details on the training provided to content moderators, technologies used, **or** languages covered, for example:

   a. Further information on the structure and ways of working of the pertinent teams (such as trust and safety), including, for example, team structure, languages covered and relevant training;

   b. Further information on new technologies adopted during the reporting period to moderate content that violates its terms and conditions on illegal hate speech, where relevant.

3. Transparency, accountability, and monitoring

[Code 3.1 and 3.2]: See monitoring exercise results and responses to this Annex 2.

4. Intra-industry and multi-stakeholder cooperation to address and prevent risks of the spread of illegal hate speech

8. [Code 4.1 and 5.2] Information about the Signatory's work with any partners to inform its policies and practices around illegal hate speech (this may include descriptions and information on partnerships, actions taken as result of these partnerships, numbers/country of operation, and possible evolution over time of existing partnerships where relevant)[8];

9. [Code 4.2 and 4.4] Information on the Signatory's participation and contribution to the regular exchange fora, including the annual convening, the intra-industry exchange of best practices, the structured channel, and the online knowledge hub;

10. [Code 4.3] (If applicable) Signatory to confirm its participation in discussions held regarding a crisis as defined in article 36 of the DSA.

5. Awareness raising, online civility and counter narrative initiatives

11. [Code 5.1] Information on any additional measures, such as efforts aimed at improving media literacy and critical thinking, awareness raising campaigns fostering civility and safety online,

[8] Note that this provision (as with all provisions in this code) shall not be construed to require the disclosure of information that is subject to confidentiality restrictions or otherwise prohibited by law.

or measures taken to discourage the creation of both illegal and non-illegal hate speech content on their platforms (such as comment controls, kindness prompts, etc);

12. [Code 5.2]: Information on training, resources and support available to Monitoring Reporters, and relevant civil society organisations. This information can include the number of participants, results of surveys/evaluations of the training by participants, or any follow up to the training sessions.

Looking ahead

13. Any information on significant trends (e.g. AI generated content) it would like to highlight in the past calendar year relating to the spread of illegal hate speech on its platform[s] and, any information the Signatory would like to highlight on how it has adapted its policies and procedures as a result;

14. Additional learnings from the participation in this Code that the Signatory would like to highlight.

REPORTING TABLE LEGEND

Code paragraph	**Reporting**
1.1	Question 1
1.2	Question 2
2.1	Question 3, 4, 5 and 6 + Monitoring Exercise
2.2	Question 6 and 7 + Monitoring Exercise
2.3	Question 6 + Monitoring Exercise
3.1	Monitoring Exercise
3.2	Responses given to this Annex 2
4.1	Question 8
4.2	Question 9
4.3	Question 10
4.4	Question 8
5.1	Question 8 and 11
5.2	Question 8 and 12
As relevant	Questions 12 and 13