



HM Government

The Government Report on Transparency Reporting in relation to Online Harms

December 2020

Table of contents

1. Ministerial foreword	2
2. Executive summary	3
3. Introduction	4
4. Background	5
5. The multi-stakeholder Transparency Working Group	8
6. Key findings and recommendations by theme	9
7. Conclusion and next steps	24
8. Annex A - List of all recommendations	25
9. Annex B - Company commitments	29

1. Ministerial foreword



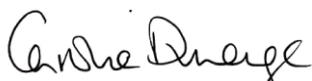
To harness the opportunities the internet offers, we must ensure that companies take greater responsibility for protecting their users from the harms that exist online. That is why we are introducing a new regulatory framework establishing a duty of care on companies to improve the safety of their users online, overseen and enforced by Ofcom. We have set out further details in the Full Government Response to the Online Harms White Paper.

At the heart of the new regulatory framework is the need to deliver greater transparency, trust and accountability. In the complex and constantly evolving online world it is crucial that the regulator is well informed, that users are empowered and that companies are held to account for keeping their users safe online.

The Online Harms White Paper set out the government's commitment to introduce mandatory transparency reporting for relevant companies. We remain steadfast in that commitment. Effective transparency reporting will be a cornerstone of the new regulatory regime, increasing visibility about how companies take decisions which affect their users and shining a light on the steps they are taking to fulfil the duty of care. This will help build the regulator's understanding of the online harms landscape and empower users of online services to make informed choices about the services they use. Transparency reporting will help ensure that users' rights online are safeguarded, including freedom of expression, and will help drive industry accountability.

The response to the Online Harms White Paper consultation highlighted a growing consensus about the importance of improving transparency about online services, which we have sought to build on. In developing this report we established a multi-stakeholder working group (the Transparency Working Group) which included representatives from industry and civil society. We are grateful for the insights and recommendations from the working group participants, which have helped to inform the future transparency framework. We expect Ofcom will continue to seek the views of a wide range of stakeholders in developing its future approach.

We also welcome the fact that a number of companies are already producing transparency reports on a voluntary basis. This is an encouraging step, and we hope that companies will continue to develop and improve their reporting efforts before the regulatory regime comes into effect.

A handwritten signature in black ink that reads "Caroline Dinenage".

Caroline Dinenage MP
Minister of State for Digital and Culture

2. Executive summary

2.1 Developing a culture of transparency, trust and accountability will be a critical element of the new regulatory framework. Ofcom will have the power to require annual transparency reports from companies as part of the new regulatory framework.

2.2 The objective of this report is to advance our shared understanding of the role transparency reporting can play as part of the future regulatory framework. The report is not an attempt to establish the specific information that companies should include in their transparency reports, as ultimately this will be for Ofcom to determine.

2.3 There is growing consensus about the value of transparency reporting. Transparency can help deliver against a number of shared objectives - empowering users, improving accountability and building our shared understanding of, and response to, online harms.

2.4 Currently, a number of companies produce transparency reports on a voluntary basis. This is a positive step and it is encouraging to see how the reports that companies are producing have developed over time. However, there are several key limitations associated with the voluntary approach.

2.5 In October 2019, the government established a new multi-stakeholder Transparency Working Group. The objective of the working group was to bring together a wide range of stakeholders to discuss transparency, to build consensus and to agree recommendations on what transparency reporting should look like, both as part of the future regulatory framework but also in the interim period.

2.6 This report presents the Transparency Working Group's recommendations. It includes a discussion of how the transparency framework could work in practice and provides insights about what types of information are likely to be most valuable. It also sets out areas that would benefit from further discussion in the future.

3. Introduction

3.1 As outlined in the Online Harms White Paper (“the White Paper”), developing a culture of transparency, trust and accountability will be a critical element of the new regulatory framework. The regulator will have the power to require annual transparency reports from companies in scope of the duty of care (although not all companies in scope of the duty of care will be required to produce transparency reports), outlining information about how they are addressing harmful content and activity on their services. These reports will be published online so that users can make informed decisions about their internet use.

3.2 In the White Paper, the government also committed to publishing the first Government Transparency Report (now referred to as The Government Report on Transparency Reporting in relation to Online Harms). This report is an interim step before the legislation is in place and the online harms regulatory framework is fully operational. The objective of this report is not to establish the specific information that companies should include in their transparency reports, as ultimately this will be for Ofcom to determine.

3.3 Instead, the aim of this report is to advance our shared understanding of the role transparency reporting can play as part of the future regulatory framework. The report includes a discussion of the objectives of transparency reporting, recommendations about how the transparency framework could work in practice, and insights about what types of information are likely to be most valuable. It also sets out areas that would benefit from further discussion in the future.

3.4 In developing this report we established a multi-stakeholder working group (the Transparency Working Group) which included representatives from the technology sector and civil society, to build consensus about the future transparency framework. The group produced a number of recommendations which are set out in this report. We also used the opportunity to discuss the steps that companies will be taking to develop their transparency reporting in the interim period.

3.5 Transparency reporting can help empower users, build our shared understanding of online harms, improve company accountability and help to protect users’ rights online, including freedom of expression. Further details on the future transparency requirements, and the wider future online harms regulatory framework, are set out in the Full Government Response to the Online Harms White Paper consultation.

4. Background

4.1 There is growing consensus about the value of transparency reporting. Governments, users, civil society organisations, academia and companies alike have recognised that transparency can help deliver against a number of shared objectives: empowering users; improving accountability; and building our shared understanding of, and response to, online harms.

4.2 There is no single definition of what constitutes a transparency report. However, the majority of transparency reports share a number of key features: they are publicly available; they contain information about the activity of online service providers; and they are issued periodically.

4.3 In the last decade there has been an increase in the number of companies producing transparency reports about the operation of their online services on a voluntary basis. This is a positive step and it is encouraging to see how the reports that companies are producing have developed over time. Many of the existing transparency reports contain really valuable information about company activity to combat online harms and explanations about the processes and decision-making procedures that are in place.

4.4 Initially, companies' reports tended to focus on requests from government and law enforcement agencies (this included requests for user information and requests for removal of content). However, the focus of transparency reporting has since expanded. Many companies now use their transparency reports as a way to demonstrate what they are doing to protect their users from harmful content and activity. Companies are also using transparency reports as a way to provide users with information about how key decisions, for instance content moderation and removal decisions, are made.

4.5 However, there are several key limitations associated with the voluntary approach:

- not all companies who could produce reports choose to;
- those who do report decide what to include in their reports and may not be incentivised to publish certain information which might be useful to users, civil society and government;
- there is a lack of independent verification of the information provided, which may reduce confidence in the accuracy and value of the data; and
- there is significant variation between the reports that different companies currently produce

4.6 Some variation in the transparency data that companies publish is inevitable, given the diversity of online companies that exist. However there is considerable scope to improve our shared understanding of what meaningful transparency data looks like.

4.7 We will encourage companies to improve their transparency reporting efforts in advance of the introduction of the new regulatory framework, which will give the regulator the power to require certain companies to publish reports.

International initiatives

4.8 There are ongoing efforts to develop common standards globally to support company transparency reports on tackling terrorist content, including through the Global Internet Forum to Counter Terrorism (GIFCT) and the Organisation for Economic Co-operation and Development's Voluntary Transparency Reporting Protocol. The government will continue to work alongside its international counterparts and representatives from industry to support this work.

Transparency and the Online Harms White Paper consultation

4.9 The Online Harms White Paper, published in April 2019, set out that the regulator will have the power to require annual transparency reports from companies in scope of the regulatory framework. The regulator will publish companies' transparency reports on its website, to support users and parents in making informed decisions about internet use. It will also produce its own annual report which will highlight key insights from the reports that companies have produced. The White Paper set out three key objectives for mandatory transparency reporting. These were:

- to help the regulator gain an understanding of the level of harms on online services and the mitigating action being taken by companies;
- to help users gain a greater understanding and awareness of whether and to what extent companies are taking positive steps to keep their users safe, and the processes different companies have in place to prevent harms; and
- to help ensure that companies take responsibility for the impacts of their services on their users and to incentivise accountability within the industry.

Key findings from the Online Harms White Paper consultation

The consultation on the Online Harms White Paper ran from 8 April 2019 to 1 July 2019. The consultation asked respondents "Beyond the measures set out in the White Paper, should the government do more to build a culture of transparency, trust and accountability across industry and, if so, what?"

- Suggestions for additional measures (beyond the White Paper proposals) to increase transparency included requiring increased clarity and detail of reporting and additional engagement with international partners.
- Civil society groups showed considerable support for the proposals advanced in the White Paper around transparency, which they saw as a crucial mechanism to increase companies' accountability and foster positive relationships with the regulator.
- Rights groups were supportive of greater transparency and accountability but were keen to emphasise that transparency reporting should promote users' rights and should contain information about how companies uphold users' right to freedom of

expression online by building their understanding of the processes which affect them and empowering them to make informed decisions about the services they use.

- While the technology sector was also overall broadly supportive of transparency, there was less consensus about the format the reporting should take. Some small and medium-sized enterprises (SMEs) highlighted resource and capability challenges associated with collecting or reporting certain types of information. Other respondents, including dating sites and retailers, echoed this concern, stating that transparency reporting might be overly onerous on them should it require significant re-engineering of their given service if it had not been designed to gather certain types of data.
- Many responses also emphasised that reporting should be qualitative, not just quantitative, avoiding a one-size-fits-all approach, and that the data reported should be clear and meaningful. Respondents also asked that transparency reports be written in plain English and made accessible to the public.
- Respondents suggested that alongside transparency reports, there should be other avenues for companies to share information with the regulator as public transparency reports may not always be the most appropriate vehicle for sharing information with the regulator.

4.10 Since the publication of the [Initial Government Response](#)¹, we have continued the conversation about transparency and our policy development in this area with the Transparency Working Group (further details are set out in Section 5).

4.11 We have emphasised that, in line with the wider regulatory framework, we will take a risk-based and proportionate approach to developing the transparency reporting requirements. The Initial Government Response set out that there will be a minimum threshold that a company would need to meet before reporting requirements would apply.

4.12 Transparency reporting has been highlighted, by both government and civil society, as a priority area for companies to take action in the interim period before regulation is introduced. The Covid-19 pandemic has also shone a light on the importance of improving transparency about the actions companies are taking to keep users safe on their services. It is vital that the government, the regulator and users are able to understand what steps companies are taking to keep their users safe online.

¹ <https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response>

5. The multi-stakeholder Transparency Working Group

5.1 In October 2019, the government established a new multi-stakeholder Transparency Working Group. Noting the widespread interest in this area, the working group was designed to build on existing broad support for greater transparency. We also designed the group to ensure a consultative and collaborative approach to policy development. The working group is chaired by the Minister for Digital and Culture.

5.2 The objective of the working group was to bring together a wide range of stakeholders to discuss transparency, to build consensus, and to agree recommendations on what transparency reporting should look like, both as part of the future regulatory framework but also in the interim period. Stakeholders contributed to the working group on the understanding that their findings and recommendations would be included in this report.

5.3 The group comprised stakeholders from the tech industry (a mixture of representatives from larger companies and small and medium-sized enterprises), alongside rights organisations and organisations which focus on the safety of children online. The composition of the group reflects the diversity of organisations with an interest in this space and included representatives with a range of views about what transparency reporting should look like.

5.4 The group met three times between October 2019 and March 2020. Following the announcement that we were minded to appoint Ofcom as the regulator for online harms, they were invited to observe the second and third sessions². Topics for discussion included:

- the indicative list of areas that transparency reporting would cover which was set out in the White Paper;
- how to ensure that future transparency reporting requirements are proportionate for different companies;
- what meaningful information looks like in the areas set out in the White Paper;
- what more can be done by companies in the interim period before the regulator is operational; and
- what the biggest challenges are for the transparency reporting framework and how government, working with industry and the regulator, will overcome them.

5.5 Throughout this process we have sought to understand stakeholders' views on the transparency reporting framework, in order to build a future regulatory regime which is flexible and proportionate, and also one which delivers against the key objectives of the White Paper. We have presented a summary of the key findings and recommendations from the Transparency Working Group in the next section. These have been grouped thematically.

5.6 Alongside this, we have included boxes setting out the government position on what the future transparency framework will look like, which has been informed by the insights and recommendations of the group.

² The organisations involved in the working group are: Childnet, The Coalition for a Digital Economy, Facebook, Global Partners Digital, Google, Internet Watch Foundation, Match Group, Microsoft, National Society for the Prevention of Cruelty to Children, Open Rights Group, Snap, Stonewall, Twitter, The Association for UK Interactive Entertainment, South West Grid for Learning, the Department for Digital, Culture, Media and Sport, the Home Office, and Ofcom, as an observer.

6. Key findings and recommendations by theme

Theme 1: What are the objectives of transparency reporting?

Key insights

6.1 The group was supportive of the three objectives set out in the White Paper:

- to help the regulator gain an understanding of the level of harms on online services and the mitigating action being taken by companies;
- to help users gain a greater understanding and awareness of whether and to what extent companies are taking positive steps to keep their users safe, and the processes different companies have in place to prevent harms; and
- to help ensure that companies take responsibility for the impacts of their services on their users and incentivise accountability within the industry.

6.2 A number of other points were highlighted by the group as crucial in ensuring that transparency reporting delivers against these objectives, namely:

- It is crucial that the information included in transparency reports is reliable. Verification of the data and quality assurance of the processes used to gather the data will help build confidence in the reliability of the data.
- The information included in transparency reports must be meaningful, with a focus on the information that will be of most use to users, civil society and the regulator.
- It is also important that the analysis and interpretation of the information included in transparency reports is rigorous, to ensure that the right conclusions are drawn.

Recommendations

Recommendation 1: To ensure that the information included in companies' transparency reports is reliable, the regulator will need to be able to verify and quality assure the data and processes used to gather the data.

Recommendation 2: Transparency reporting should promote and support users' rights (including freedom of expression) and enable users and civil society to understand the response of companies in scope. This should be seen as a key objective of the reporting framework.

Recommendation 3: The transparency reporting framework should be future-proof and should incentivise continuous innovation, encouraging companies to take action from the start.

Recommendation 4: The framework should also incentivise collaboration and coordination between companies, for instance in relation to the sharing of best practice or specific tools and technologies, in recognition of the fact that the perpetrators of online harms may operate between platforms.

Recommendation 5: Further discussion is needed on how the online harms regulator will analyse and compare transparency information between different companies.

How will transparency reporting under the online harms regulatory framework deliver against these objectives?

- Transparency reports will help shape the regulator's understanding of what companies are doing to keep their users safe. These reports will be used alongside a variety of other sources of information to help the regulator understand how companies are fulfilling the duty of care and help determine the regulator's priorities.
- Transparency reports will provide users themselves with important information about the steps that companies are taking to tackle online harms. Alongside the transparency reports that companies produce, the regulator will also produce a report which will include key findings and analysis from the company reports. This will make it easier for users to make informed decisions about which services they use.
- Better informed users will help drive industry accountability and encourage action from companies. By improving the shared understanding of how companies are tackling online harms, users, civil society, government and the regulator will be able to assess whether companies are taking sufficient action to tackle online harms.
- Transparency reports will also provide insights into how companies are protecting freedom of expression, helping users, civil society and the regulator to hold companies to account for doing so.

Theme 2: What are the major challenges associated with transparency reporting/what are the barriers to achieving these objectives?

Over the course of the three sessions, a number of potential challenges associated with transparency reporting were raised, and the group discussed how they might be tackled.

Key insights

6.3 The transparency landscape is complex, nuanced and will develop over time. It is therefore essential that the transparency framework is future proof.

6.4 The diversity of the companies in scope was also highlighted as a key challenge. Companies, especially small and medium-sized enterprises, need time in order to build their transparency reporting systems and capabilities.

6.5 Furthermore, it is important to consider what reporting should look like for companies who operate a range of services.

6.6 Participants agreed that without contextual information, statistics can be misleading.

6.7 Similarly, the issue of perverse incentives was raised. It was emphasised that statistics about the actions companies are taking to tackle online harms (e.g. removal of users) could skew perceptions about the prevalence of harm.

6.8 It was also emphasised that high profile events might lead to sudden calls for data that companies might not collect.

Recommendations

Recommendation 6: Given the potential for rapid change in this space, it is critical that the transparency framework incentivises innovation to ensure it meets the objectives set out in the White Paper.

Recommendation 7: It will be important to ensure that transparency reporting supports and safeguards freedom of expression.

Recommendation 8: The transparency reporting framework must take into account that gathering data for transparency reporting takes time and requirements for new data could require companies to pivot their existing design.

Theme 3: Who should report and how should reporting differ between companies?

Key insights

6.9 A wide range of service types will be in scope of the future online harms framework. There are significant differences in the size, capability and risk-profile of in scope companies. It would not be proportionate to require all companies in scope to produce transparency reports.

6.10 The information that will be meaningful (to users and the regulator) may also vary between different types of company.

6.11 As companies can scale up very quickly it is important to incentivise early action and continuous improvement, whilst acknowledging that companies may be starting from very different baselines.

Recommendations

Recommendation 9: The transparency reporting requirements should reflect the diversity of services in scope.

Recommendation 10: To ensure proportionality, the regulator should use a threshold in determining who needs to report. This should be based on factors such as company revenue/capability, the functionality of the service and the reach/audience of the service. However, the regulator may need some flexibility to place transparency reporting requirements on services falling below the threshold if they are sufficiently high risk.

Recommendation 11: Expectations of small and medium-sized enterprises should be different to the expectations of the largest companies.

Recommendation 12: The regulator should have a role in encouraging best practice and continuous improvement when companies are in the early stages of their business cycle.

Who will be required to publish transparency reports under the online harms regulatory framework?

- Companies providing Category 1 services will be required to publish reports containing information about the steps they are taking to tackle online harms on these services.
- Category 1 services will be determined through a three-step process. First, the primary legislation will set out high level factors which lead to significant risk of harm occurring to adults through legal but harmful content. These factors will be: the size of a service's audience (because harm is more likely to occur on services with larger user bases, for example due to rapid spread of content and 'pile-on' abuse); and the functionalities it offers (because certain functionalities, such as the ability to share content widely or contact users anonymously, are more likely to give rise to harm). Second, the government will determine and publish thresholds for each of the factors. Ofcom will be required to provide non-binding advice to the government on where these thresholds should be set. The final decision on thresholds will lie with the government, to ensure democratic oversight of the scope of the regulatory framework. Ofcom will then be required to assess services against these thresholds and publish a register of all those which meet both thresholds. These services will be designated as Category 1 services.
- The Secretary of State for Digital, Culture, Media and Sport will also have the power to extend the scope of companies who will be required to publish transparency reports, beyond Category 1 companies, by setting additional thresholds based on factors such as the functionalities and the audience of the service.
- To ensure that the transparency reporting framework is agile and future-proof, the regulator will need flexibility in determining the specific information companies will need to provide. The legislation will set out a list of the types of information that the regulator may require companies to report on, relating to a number of areas.

Theme 4: The wider transparency landscape

Key insights

6.12 Transparency reporting is only one part of the wider transparency, trust and accountability framework.

- The regulator should also have additional information gathering powers in order to better understand whether companies are taking sufficient action to fulfil the duty of care. In order to understand how well companies' internal processes are working, powers for the regulator to audit these systems would be really useful.
- It would not be appropriate to include certain sensitive information in public transparency reports.

- In addition to transparency reporting, there should be additional avenues for companies to share information privately with the regulator.

6.13 Civil society, academics and other experts have an important role to play in the wider transparency landscape:

- The regulator should harness the insights of civil society and academics in order to maximise the value of the transparency information that companies publish.
- Valuable insights could be gained by providing academics with access to certain company information, for research into online harms, including looking at, and the impacts of mitigations on, free expression. There are significant commercial sensitivities here and so safeguards would be needed to ensure that such research is conducted in a safe and secure way which is consistent with data protection requirements.

6.14 It is also important that the transparency reporting framework is aligned with other relevant initiatives.

- Transparency reporting should complement initiatives which promote user safety online. There is an important role that parents, schools and companies have to play in building user awareness about how to keep safe online.
- The regulator should consider how to align the transparency reporting requirements with existing international initiatives and influence future ones. This will avoid placing disproportionate burdens on business or creating confusion for users.

Recommendations

Recommendation 13: The regulator should be equipped with other information gathering and investigation powers so it can understand whether companies are fulfilling the duty of care and hold them to account.

Recommendation 14: Companies should have appropriate ways to share and disclose sensitive information to the regulator directly. Further discussion is needed on what information should be disclosed directly to the regulator as opposed to included in transparency reports.

Recommendation 15: Transparency is an ecosystem and it is important to take a holistic approach. As well as working with companies to ensure the framework is proportionate and incentivises innovation, the regulator should work with users, civil society and academia in developing the future framework.

Recommendation 16: Academics, civil society and external experts should play an important role in the transparency framework. For instance, by analysing and explaining key trends such as how and why the amount of harmful content on a platform changes over time, or may vary between different audiences.

The relationship between transparency reporting and the regulator's information gathering powers and powers to support investigations

- The regulator's information gathering powers will also play a crucial role in supporting its various regulatory functions. These powers will help the regulator build an in-depth understanding of the online harms landscape, prioritise its activity and oversee companies' compliance with the regulatory framework.
- The regulator will have a broad power to require the information that it needs to carry out its functions. This will give the regulator the flexibility to determine the specific information it requires.
- The regulator will use information from a wide range of sources to help prioritise its investigation and enforcement activity. Alongside the information which companies have provided (in their transparency reports and in response to information requests), the regulator will also use user complaints data and publicly available information to help determine whether an investigation might be warranted.
- The regulator will also have a number of additional powers to support its oversight and enforcement activity. Where there are reasonable grounds to suggest that a company may be non-compliant, Ofcom will have the power to enter companies' premises and access documentation, data and equipment in order to understand whether companies are taking sufficient measures to fulfil the duty of care.
- Ofcom will also have the power to interview employees, which will allow it to develop further understanding of how the company is complying with the duty of care.
- Finally, Ofcom will have the power to require a company to undertake, and pay for, a skilled person report on specific issues of concern. This power will be particularly useful on issues where external technical expertise is needed, for instance to validate the effectiveness of automated moderation systems. As with all its powers, Ofcom will be required to take a proportionate approach to the use of these powers.
- Further detail on the regulator's information gathering powers can be found in Part 4 of the Full Government Response to the Online Harms White Paper consultation.

Theme 5: What type of information should transparency reporting cover?

Key findings

6.15 The group was broadly supportive of the six high level categories of information for transparency reporting that was set out as an indicative list in the White Paper. This included:

- Evidence of effective enforcement of the company's own relevant terms and conditions, which should be consistent with the codes of practice issued by the regulator. Processes that the company has in place for reporting illegal and harmful

content and behaviour, the number of reports received and how many of those reports led to action.

- Proactive use of technological tools, where appropriate, to identify, flag, block or remove illegal or harmful content.
- Measures and safeguards in place to uphold and protect fundamental rights ensuring decisions to remove content, block and/or delete accounts are well founded, especially when automated tools are used and that users have an effective route of appeal.
- Where relevant, evidence of cooperation with UK law enforcement and other relevant government agencies, regulatory bodies and public agencies.
- Details of investment to support user education and awareness of online harms, including through collaboration with civil society, small and medium-sized enterprises and other companies.

6.16 Over the course of the sessions, a number of other potential categories not captured in the initial six were discussed, including:

- Information about tools for users to help them manage potentially harmful content and activity.
- Information about the process and steps an organisation has in place to assess risk of harm at the design, development and update stage of the online service.

6.17 The group agreed that including both qualitative and quantitative information in the reporting requirements is vital to ensuring the reports are as meaningful as possible. Contextual information which helps to explain quantitative data is necessary.

Recommendations

Recommendation 17: In addition to the six high level categories of information outlined in the White Paper, the regulator should be able to require additional information on tools for users to help them manage potentially harmful content and activity, and about the process and steps an organisation has in place to assess risk of harm at the design, development and update stage of the online service.

Recommendation 18: Transparency reporting should cover both qualitative and quantitative information.

Recommendation 19: The transparency reporting framework should reflect the focus on systems and processes which underpins the duty of care.

What types of information will transparency reports cover?

An indicative list of the high level categories of information that companies might need to include in their transparency reports is set out below. This list builds upon the initial list proposed in the Online Harms White Paper in April 2019, and has been informed by the recommendations of the working group:

- Information about the enforcement of the company's own relevant terms and conditions, which should reflect the regulator's codes of practice.
- Information about the processes that the company has in place for reporting harmful content and activity (including in relation to illegal harms), the number of reports received and the action taken as a result.
- Information about the processes and tools in place to address illegal and harmful content and activity, including, where appropriate, tools to identify, flag, block or remove illegal and harmful content and the processes that companies have in place for directing users to support and information.
- Information about the measures and safeguards in place to uphold and protect fundamental rights, ensuring decisions to remove content, block and/or delete accounts are well founded, especially when automated tools are used, and that users have an effective route of appeal.
- Where relevant, information about evidence of cooperation with UK law enforcement and other relevant government agencies, regulatory bodies and public agencies.
- Information about measures to support user education and awareness of online harms and strengthen users' media literacy, including through collaboration with civil society, small and medium-sized enterprises and other companies.
- Information about tools for users to help them manage harmful content and activity.
- Information about the process and steps an organisation has in place to assess risk of harm at the design, development and update stage of the online service.
- Information about other steps that companies are taking to tackle online harms and fulfil their obligations under the online harms framework, including to deliver a higher level of protection to children where a platform is likely to be accessed by children.

Theme 6: What does meaningful information look like in the areas discussed?

6.18 Having discussed the types of information that transparency reporting should cover, at a high level, the group discussed what information would be most meaningful within these categories.

6.19 The objective of this particular discussion was to develop our understanding of what meaningful information could look like and to help identify areas of consensus, potential challenges and additional questions for future discussion.

6.20 It will be for the future regulator to determine the specific metrics relevant companies will need to provide.

6.21 Furthermore, some of the information which the regulator may need from companies may need to be shared directly with the regulator as opposed to published in transparency reports,

given the commercial sensitivity and/or the potential for the information to be used by bad actors to cause harm.

6.22 The regulator will work closely with industry and other stakeholders to determine the specific information that companies will need to include in their transparency reports, and how this will differ between companies. The regulator will also work closely with law enforcement and companies to build its understanding of potentially sensitive information which should not be included in transparency reports, but should be shared directly with the regulator.

6.23 Below is a summary of the discussions about what kind of information would be meaningful in each of the six categories. It should be emphasised that not all recommendations (discussed below) would be applicable to all companies, and some of the recommendations might apply differently, depending on the specific characteristics of the service(s) in question. Furthermore, additional discussion will be needed to explore how the regulator should use and analyse the information it requires from companies.

Category 1: Evidence of effective enforcement of the company's own relevant terms and conditions

Summary: Companies will be required to set out their policies on what behaviour is considered unacceptable in their terms and conditions/acceptable use policies. These policies set out different types of behaviour and activity that will not be tolerated by the service as well as information about how such activity will be dealt with.

Key insights

6.24 It is important for users, civil society and the regulator to understand how companies' terms and conditions are enforced.

6.25 It is useful to understand the amount of content on a platform which violates a company's acceptable use policy. Information about how often users experience and interact with this content or activity might be more useful than the total number of pieces of harmful content found on a platform, as a lot of content is taken down by companies automatically. Statistics about how often users experience and interact with content/activity are likely to be approximations.

6.26 Focusing solely on the quantity of harmful content which is removed might create perverse incentives which could negatively impact users' freedom of expression. There is a risk that transparency reporting could incentivise the over-removal of legal content if companies feel that the quantity of content removed is what they are being measured against. The inclusion of contextual information alongside quantitative data can help mitigate this risk. Furthermore, effective user appeal and redress processes will also play an important role in ensuring users' freedom of expression is protected.

6.27 In certain areas, for instance when dealing with terrorist content, information about the time taken to remove illegal content could also be valuable, given the speed at which this content is shared and the risks associated with this. At the same time it is important to consider

the potential to create perverse incentives, so further discussion of this topic is needed. The regulator should work closely with industry and other stakeholders on this in future.

6.28 There is real value in involving academics and external experts in the discussion and analysis of transparency data.

6.29 Worked examples and discussion of the types of decisions companies are required to make when moderating content could provide valuable insights to users.

Recommendations

Recommendation 20: Transparency reporting should include information about the processes which companies use to enforce their terms and conditions.

Recommendation 21: Transparency reporting should help users and the regulator to understand how well these processes are working and whether, at a systematic level, the moderation decisions that companies are making are accurate.

Recommendation 22: Transparency reports should include information about how often content which violates terms and conditions, or which has been identified as illegal by an appropriate body, is seen or shared.

Recommendation 23: Transparency reports should not focus solely on the quantity of harmful content which is removed as this might create perverse incentives which could negatively impact users' freedom of expression.

Recommendation 24: The regulator should work with companies, civil society and academia to help understand and explain the data included in transparency reports.

Category 2: Processes that the company has in place for reporting illegal and harmful content and behaviour, the number of reports received and how many of those reports led to action.

Summary: Many companies have reporting functionality on their online services where users are able to report illegal or harmful content. There are usually categories which specify the type of rule violation which may have occurred, such as bullying or hate speech. These reporting categories and the thresholds for what is considered to be unacceptable behaviour (in relation to legal but harmful content and activity) may vary between companies.

Key insights

6.30 Understanding the reporting processes that a company has in place will be important for the regulator in understanding how companies are fulfilling their duty of care.

6.31 Information about the number of reports received, and the steps companies are taking as a result, will also be useful for the regulator and for users.

6.32 Transparency reporting could also provide insights into the experience of particular groups of users, such as children and young people and those with protected characteristics.

6.33 Transparency reporting could help build our understanding of whether/how certain groups are disproportionately affected by certain harmful content and activity. However, this also needs to be balanced with ensuring that the privacy of those users is protected. Further discussion and exploration of this topic with companies and representatives from these groups would be beneficial.

Recommendations

Recommendation 25: Transparency reports should include information about what companies' reporting processes look like and how well they are working.

Recommendation 26: Transparency reports should include additional breakdown about user reports. Additional information about the experience of particular groups of users in relation to different categories of content/behaviour which violates terms and conditions, would be useful. Further discussions and exploration of this topic with companies and representatives for these groups would be beneficial.

Category 3: Proactive use of technological tools, where appropriate, to identify, flag, block or remove illegal or harmful content.

Summary: Many companies have tools and processes in place to proactively identify potentially illegal content on their services.

Key insights

6.34 It is important to understand what tools and processes companies are using to proactively identify, flag, block and remove illegal content and how well they are working.

6.35 Information about the use of proactive use of technological tools will help develop our shared understanding of why users see, or do not see, certain content on online services.

6.36 Information about the provision, use and awareness of the safety tools, and the information and support available to users, was also highlighted as important.

6.37 Some of the information about companies' processes that will be most useful to users (e.g. information about where they can get help and support) is integrated into the services. It is important that the transparency reporting compliments this.

6.38 Transparency around the use of algorithms is also an important part of the equation. It is important that the regulator and users understand the impact that use of algorithms may have, but the information that will be valuable is likely to differ between these audiences.

6.39 Certain information about the operation of companies algorithms is commercially sensitive or could pose issues to user safety if publicly disclosed. Transparency reports may not be the appropriate vehicle for certain information about the use of algorithms.

6.40 There is value in the regulator having the power to conduct audits in relation to companies' internal systems and processes.

Recommendations

Recommendation 27: Transparency reporting should include information about the tools which are being used to identify, flag, block or remove illegal or harmful content and how well they are working.

Recommendation 28: Transparency reporting should also include relevant information about user awareness and use of companies' tools.

Recommendation 29: Transparency reporting should also contain some information about internal quality assurance processes for moderation decisions and 'safety by design' processes, but the regulator should also be mindful that companies' processes are constantly evolving.

Recommendation 30: The regulator should also undertake further discussion on the approach to information which may be commercially sensitive or could pose issues to user safety, in considering what should be disclosed privately rather than in transparency reports.

Category 4: Measures and safeguards in place to uphold and protect fundamental rights, ensuring decisions to remove content, block and/or delete accounts are well founded, especially when automated tools are used and that users have an effective route of appeal.

Summary: Companies take action to remove content and delete or block accounts where they believe behaviour or content is not compliant with their terms and conditions. Many companies have processes in place to help ensure that these decisions are made accurately and consistently.

Key insights

6.41 The group recognised the importance of ensuring the framework protects freedom of expression online.

6.42 Where companies have mechanisms for users to appeal moderation decisions, information on the number of user appeals, and the processes in place for dealing with those appeals, will be crucial.

6.43 Statistics about the number of appeals, and the number of decisions which are subsequently overturned, can be really useful. This applies to both appeals to reinstate an account/content, and appeals when content has been requested to be taken down and has stayed up. However, there are a number of challenges associated with this information. There may be obstacles which reduce the number of users who appeal decisions (e.g. if the appeals process is overly complex). Users may also report content, or appeal decisions on content, which does not contravene terms and conditions simply because it is content they do not like or want to see.

6.44 Not all reports meet the thresholds for removal and in many cases the fact that reported content remains on a platform is because it did not breach terms and conditions. Similarly, just because decisions are not appealed by users does not necessarily mean they were accurate.

6.45 Furthermore, focussing on the numbers alone may create perverse incentives and might negatively impact freedom of expression (e.g. by incentivising over-removal of content). As with other categories for transparency, statistical data (e.g. information about the amount of content being appealed) should be combined with contextual information and information on processes.

6.46 More can be done to communicate to users what happens after an appeal has been made. It would also be useful if transparency reporting could shed light on the factors which are discouraging users from reporting things to companies.

6.47 Algorithms are playing an increasingly prominent role in content moderation. Both the regulator and users would benefit from additional information about how algorithms determine what content is seen. This applies both to content that is removed by algorithms as well as content which is de-prioritised.

6.48 It is also important to consider the potential impact that measures which incentivise additional content moderation can have on the rights and wellbeing of the moderators themselves.

Recommendations

Recommendation 31: Transparency reports should include information about what measures and safeguards companies have in place to uphold and protect users' rights, such as freedom of expression, and how effective they are.

Recommendation 32: Transparency reports should include information about the quality assurance processes for moderation decisions, and how well they are working.

Recommendation 33: Transparency reports should contain information about user awareness of appeals processes and the 'user journey' process once reports have been made.

Recommendation 34: Transparency reporting should also provide information about the error rates of moderation processes. For example, analysis of random samples of moderation decisions might provide valuable insights.

Recommendation 35: Transparency reports should include, where appropriate, information on the use of algorithms and automated processes in content moderation. However, given the commercial sensitivities and safety implications associated with publishing certain information, further discussion on this topic is needed.

Recommendation 36: Transparency reporting should also include information about the human resources behind the content moderation decisions, including what training they have had and what support they are offered.

Category 5: Understanding how companies are engaging and cooperating with UK law enforcement and other relevant government agencies, regulatory bodies and public agencies.

Summary: In certain circumstances, companies will need to engage with law enforcement agencies in order to effectively tackle certain types of illegal content on online services. For example, the Counter Terrorism Referral Unit in the Metropolitan Police identifies terrorist content and activity online that violates UK terrorism legislation, and works with companies by referring this content for online services to review and potentially remove under their community guidelines or terms of service.

Key insights

6.49 Many companies have existing processes in place for cooperating and engaging with law enforcement. Further work can be done to ensure effective cooperation between services, especially in tackling illegal activity. This might involve sharing information between services about online criminal activity, including when and how it is occurring as well as how to tackle it.

6.50 Government and law enforcement agencies could look for ways to be more transparent about requests to companies to remove illegal content.

Recommendations

Recommendation 37: Transparency reporting should include information about how companies are engaging and cooperating with UK law enforcement and other relevant government agencies, regulatory bodies, public agencies, and Non Governmental Organisations. Further discussion is needed in this area.

Category 6: Details of investment to support user education and awareness of online harms, including through collaboration with civil society, small and medium-sized enterprises and other companies.

Summary: A number of companies invest in initiatives to educate users about online harms and to promote online safety.

Key insights

6.51 Both users and the regulator will benefit from understanding the work companies are doing to support user education and awareness of online harms. Civil society and academia also have an important role in this space.

Recommendations

Recommendation 38: Transparency reporting should give companies the opportunity to provide information on what they are doing to support user education and awareness of online harms, how effective this work is, and incentivise best practice in tackling online harms and

keeping users safe. This information should cover what companies are doing outside of their service and also on the service itself.

Theme 7: Reporting in the interim period

6.52 Transparency reporting is an area where companies can take action now. The group as a whole discussed priorities in the interim before the regulator is set up, detailed in this section. In addition, we asked companies to set out their own plans for how they will develop their transparency reports in the interim (included in Annex B).

6.53 The group also made a series of broader recommendations relating to the interim period before the regulatory regime comes into effect as follows:

6.54 Companies could improve their UK-level data, especially around prevalence (e.g. around Child Sexual Exploitation and Abuse (CSEA) data). This could be a really useful area to prioritise in the interim.

6.55 Companies could provide additional breakdown (within reporting categories/categories of rule violation). For instance, additional detail about the proportion of children and young people who are reporting illegal and harmful content would provide meaningful insights. Furthermore, as previously discussed, additional work to understand the experiences of people with protected characteristics (as established by the Equality Act 2010³) would be useful, however this must be balanced against protecting the privacy of these users.

6.56 In order to engage young people better, additional work is needed to improve young people's confidence and trust in reporting. This is an area of real opportunity for the interim period.

6.57 There is still work to be done on what questions transparency reports can help answer and what data would help to answer those questions.

³ Section 4 of the Equality Act 2010.

7. Conclusion and next steps

7.1 The multi-stakeholder Transparency Working Group has ensured that development of the future transparency framework has been informed by a range of perspectives. Over the course of the discussions thus far, a number of topics have been highlighted which merit further discussion. It is our intention to continue to convene the Transparency Working Group during the interim period to further develop our shared understanding of the role transparency reporting can play.

7.2 Following the announcement that the government was minded to appoint Ofcom as the regulator for online harms, Ofcom were invited to observe the working group. As we move towards regulation, we intend for Ofcom to play a more central role in convening the working group in order to ensure the group helps deliver against our shared aims.

7.3 Following publication of this report, participants will be given the opportunity to discuss the future operation of the working group, including on the priority topics for future discussion.

7.4 Ultimately it will be for the regulator to determine the specific information that companies should include in their transparency reports and how this differs between different companies.

7.5 We hope that this report has advanced our shared understanding of the role transparency reporting can play as part of the future regulatory framework. This report is one step on a journey towards delivering the transparency, trust and accountability framework that was set out in the Online Harms White Paper.

7.6 We look forward to working with the regulator and representatives from industry, civil society and academia in the coming months and years, to build on these discussions.

Annex A: List of all recommendations

Recommendation 1: To ensure that the information included in companies' transparency reports is reliable, the regulator will need to be able to verify and quality assure the data and processes used to gather the data.

Recommendation 2: Transparency reporting should promote and support users' rights (including freedom of expression) and enable users and civil society to understand the response of companies in scope. This should be seen as a key objective of the reporting framework.

Recommendation 3: The transparency reporting framework should be future-proof and should incentivise continuous innovation, encouraging companies to take action from the start.

Recommendation 4: The framework should also incentivise collaboration and coordination between companies, for instance in relation to the sharing of best practice or specific tools and technologies, in recognition of the fact that the perpetrators of online harms may operate between platforms.

Recommendation 5: Further discussion is needed on how the online harms regulator will analyse and compare transparency information between different companies.

Recommendation 6: Given the potential for rapid change in this space, it is critical that the transparency framework incentivises innovation to ensure it meets the objectives set out in the White Paper.

Recommendation 7: It will be important to ensure that transparency reporting supports and safeguards freedom of expression.

Recommendation 8: The transparency reporting framework must take into account that gathering data for transparency reporting takes time and requirements for new data could require companies to pivot their existing design.

Recommendation 9: The transparency reporting requirements should reflect the diversity of services in scope.

Recommendation 10: To ensure proportionality, the regulator should use a threshold in determining who needs to report. This should be based on factors such as company revenue/capability, the functionality of the service and the reach/audience of the service. However, the regulator may need some flexibility to place transparency reporting requirements on services falling below the threshold if they are sufficiently high risk.

Recommendation 11: Expectations of small and medium-sized enterprises should be different to the expectations of the largest companies.

Recommendation 12: The regulator should have a role in encouraging best practice and continuous improvement when companies are in the early stages of their business cycle.

Recommendation 13: The regulator should be equipped with other information gathering and investigation powers so it can understand whether companies are fulfilling the duty of care and hold them to account.

Recommendation 14: Companies should have appropriate ways to share and disclose sensitive information to the regulator directly. Further discussion is needed on what information should be disclosed directly to the regulator as opposed to included in transparency reports.

Recommendation 15: Transparency is an ecosystem and it is important to take a holistic approach. As well as working with companies to ensure the framework is proportionate and incentivises innovation, the regulator should work with users, civil society and academia in developing the future framework.

Recommendation 16: Academics, civil society and external experts should play an important role in the transparency framework. For instance, by analysing and explaining key trends such as how and why the amount of harmful content on a platform changes over time, or may vary between different audiences.

Recommendation 17: In addition to the six high level categories of information outlined in the White Paper, the regulator should be able to require additional information on tools for users to help them manage potentially harmful content and activity, and about the process and steps an organisation has in place to assess risk of harm at the design, development and update stage of the online service.

Recommendation 18: Transparency reporting should cover both qualitative and quantitative information.

Recommendation 19: The transparency reporting framework should reflect the focus on systems and processes which underpins the duty of care.

Recommendation 20: Transparency reporting should include information about the processes which companies use to enforce their terms and conditions.

Recommendation 21: Transparency reporting should help users and the regulator to understand how well these processes are working and whether, at a systematic level, the moderation decisions that companies are making are accurate.

Recommendation 22: Transparency reports should include information about how often content which violates terms and conditions, or which has been identified as illegal by an appropriate body, is seen or shared.

Recommendation 23: Transparency reports should not focus solely on the quantity of harmful content which is removed as this might create perverse incentives which could negatively impact users' freedom of expression.

Recommendation 24: The regulator should work with companies, civil society and academia to help understand and explain the data included in transparency reports.

Recommendation 25: Transparency reports should include information about what companies' reporting processes look like and how well they are working.

Recommendation 26: Transparency reports should include additional breakdown about user reports. Additional information about the experience of particular groups of users in relation to different categories of content/behaviour which violates terms and conditions, would be useful. Further discussions and exploration of this topic with companies and representatives for these groups would be beneficial.

Recommendation 27: Transparency reporting should include information about the tools which are being used to identify, flag, block or remove illegal or harmful content and how well they are working.

Recommendation 28: Transparency reporting should also include relevant information about user awareness and use of companies' tools.

Recommendation 29: Transparency reporting should also contain some information about internal quality assurance processes for moderation decisions and 'safety by design' processes, but the regulator should also be mindful that companies' processes are constantly evolving.

Recommendation 30: The regulator should also undertake further discussion on the approach to information which may be commercially sensitive or could pose issues to user safety, in considering what should be disclosed privately rather than in transparency reports.

Recommendation 31: Transparency reports should include information about what measures and safeguards companies have in place to uphold and protect users' rights, such as freedom of expression, and how effective they are.

Recommendation 32: Transparency reports should include information about the quality assurance processes for moderation decisions, and how well they are working.

Recommendation 33: Transparency reports should contain information about user awareness of appeals processes and the 'user journey' process once reports have been made.

Recommendation 34: Transparency reporting should also provide information about the error rates of moderation processes. For example, analysis of random samples of moderation decisions might provide valuable insights.

Recommendation 35: Transparency reports should include, where appropriate, information on the use of algorithms and automated processes in content moderation. However, given the commercial sensitivities and safety implications associated with publishing certain information, further discussion on this topic is needed.

Recommendation 36: Transparency reporting should also include information about the human resources behind the content moderation decisions, including what training they have had and what support they are offered.

Recommendation 37: Transparency reporting should include information about how companies are engaging and cooperating with UK law enforcement and other relevant government agencies, regulatory bodies, public agencies, and Non Governmental Organisations. Further discussion is needed in this area.

Recommendation 38: Transparency reporting should give companies the opportunity to provide information on what they are doing to support user education and awareness of online harms, how effective this work is, and incentivise best practice in tackling online harms and keeping users safe. This information should cover what companies are doing outside of their service and also on the service itself.

Annex B: Company commitments

In the working group sessions, companies were invited to outline what progress they had already made with their transparency reports, and how they saw their transparency reports evolving in the period before the regulatory regime comes into effect. The following statements were provided by the companies.

Microsoft

- We are actively engaged on this issue through participation in the Organisation for Economic Co-operation and Development and Global Internet Forum to Counter Terrorism workstreams on transparency reporting.
- We also note that Electronic Frontier Foundation and other groups have launched a consultative period on the [Santa Clara Principles](https://santaclaraprinciples.org/)⁴, on transparency more broadly.
- We encourage consistency among national, regional, and international efforts to support transparency reporting both by technology companies and governments around terrorist and violent extremist content.
- Moreover, we support transparency reporting guidelines that are “future-proofed” and take into account the need for continuous innovation.
- As a company, Microsoft has increased the resources dedicated to transparency reporting around terrorist and violent extremist content, in order to make good on the public commitments we’ve made as part of the [Christchurch Call](https://www.christchurchcall.com/call.html)⁵ and [nine steps](https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2019/05/Christchurch-Call-and-Nine-Steps.pdf)⁶.
- We stand ready to engage further, including through consultative processes with the UK government and other stakeholders, including civil society organizations.

Snap Inc.

- Snap has voluntarily produced a bi-annual Transparency Report since November 2014.
- These reports provide important insight into the volume and nature of governmental and law enforcement requests for Snapchatters' account information and metadata, and other legal notifications received.
- Our decision to report aligns with our values and ideology, and we always look to collaborate with law enforcement while making sure we are open with our community, helping to shape as safe an environment as possible.
- From 2020, our Transparency Reports have provided insights into the volume and nature of accounts reported on Snapchat for violations of our Terms of Service or Community Guidelines. These disclosures provide our community with useful information on key categories of content reported and enforced on Snapchat.

⁴ <https://santaclaraprinciples.org/>

⁵ <https://www.christchurchcall.com/call.html>

⁶ <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2019/05/Christchurch-Call-and-Nine-Steps.pdf>

- Our next Transparency Report will include overviews of the enforcement of our rules in individual countries, as well as information and resources on our approach to safety and privacy at Snap.

Match Group

- Match Group operates a portfolio company composed of several brands, such as Tinder®, Match®, Meetic®, OkCupid®, Hinge®, Pairs™, PlentyOfFish®, and OurTime®. Safety and transparency are key priorities for Match Group and for all of the brands we operate. Match applauds the valuable work the Transparency Working Group has undertaken in its efforts to develop an online culture of transparency and trust.
- Each of our brands have developed and operate separate platforms to offer their unique services tailored to the various demographics and geographies they serve, and also to maintain the separation of privacy of user data. Each brand is in the process of implementing tools and systems that will allow all brands to report on a unified set of metrics. These tools will allow Match Group to provide clear and accurate reports at the Group level. Match Group is currently working on this process and is looking forward to publishing its first transparency report.

Google

- Google was founded on the belief that everything we do should always respect the user. Transparency is integral to user trust and core to Google's commitment to respect human rights. Our quantitative and qualitative transparency efforts shed light on how the policies and actions of governments -- and our policies and actions across Google services -- affect privacy, security, and access to information. As the Internet evolves, this means continuously advancing how we share this information with our users and the public. Google first launched our Transparency Report in 2010 with three reports: government requests for content removal, global traffic patterns, and government requests for user data. Today, the Transparency Report is home to 12 different reports that disclose information across a broad range of areas, including security, privacy, and access to information. For the purpose of the government's inquiry, we wish to highlight three of these:
 - Requests for user information: A variety of laws allow government agencies around the world to request user information for civil, administrative, criminal, and national security purposes. In this report, we share information about the number and type of requests we receive from government agencies. Earlier this year, we expanded this report to include requests for Cloud enterprise customer data. The data is reported in a 6-month timeframe.
 - Government requests for content removal: Courts and government agencies around the world regularly request that we remove information from Google products. We review these requests closely to determine if content should be

removed because it violates a law or our product policies. In this report, we disclose the number of requests we receive in six-month periods, with a country/region view. We expanded the report earlier this year to show the percentage of content that was removed for policy reasons vs. local law in each country or region.

- YouTube Community Guidelines Enforcement Report: At YouTube, our Community Guidelines set the rules of the road for what we don't allow on YouTube. We rely on a combination of people and technology to flag inappropriate content and enforce these guidelines. Flags can come from our automated flagging systems, from members of the Trusted Flagger program (NGOs, government agencies, and individuals) or from users in the broader YouTube community. This report provides data on the flags YouTube receives, enforcement of our policies, and user appeals and reinstatements of content. We also include a few deep dives on specific policy issues, including hate speech and violent extremism. This report is updated quarterly.
- Outside of the Transparency Report infrastructure, we also make available a quarterly bulletin disclosing action taken across Google against coordinated influence operations. Qualitative transparency is equally as important to our users. We provide robust explanations of the policies that govern our services and our approach to moderating content, including how to provide flags or feedback and how to appeal. For example, we provide in-depth information about how we tackle specific issues, like disinformation, or about specific products, like YouTube, Play, and Search.

Our plans for next year transparency reporting:

- Since 2010, not only have the number of reports in the Transparency Report grown, as you can see from the explanations of our key reports, but the amount of data we share has expanded as well. We continue to look at ways to improve the user experience of our reports in order to carry out our mission to share data that sheds light on how the policies and actions of governments and corporations affect privacy, security, and access to information. This mission is consistent with the government report's recommendation that, "Transparency reporting should promote and support users' rights and freedom of Expression."
- Publishing this data takes an extraordinary amount of unseen effort in terms of preparing and validating data, importing it into the site infrastructure, and cross-checking accuracy. Over the next year, a major focus of our work will be back end engineering work—invisible to the public—but that will improve our infrastructure to make publishing and validating the data easier and, hopefully, quicker. For the YouTube report, we are looking closely at where we may be able to add more data about our processes and enforcement broadly, as well as how to refine our reporting processes and expand what data we make available, based on both what other companies are doing and feedback from external stakeholders, including governments, researchers, and human and civil rights organisations.
- It is important to note that even the addition of a single statistic to any of reports takes at least 6 months of planning, data science investment, and design and

engineering work to finalise and publish. It's also important to note that "transparency" reporting can take many forms, not just a site or PDF with statistics. For example, YouTube recently launched a site called How YouTube Works, which seeks to centralise answers to some of the questions we are most commonly asked about processes and policies on YouTube. In addition to the thorough info to users in our Help Center, we consider this a transparency effort. We will expand and refine the site over the coming year, and encourage the government to consider these types of efforts to simplify and centralise information as a form of qualitative transparency reporting.

Twitter

- Transparency is foundational to the kind of Internet that we all want to see. Twitter is the only major service to make public conversation data available via an API for the purposes of study, and this has resulted in a number of important benefits. In November 2020, we worked with Demos to highlight why greater data transparency should be at the centre of our societal response to online harms, like information operations (https://blog.twitter.com/en_us/topics/company/2020/nation-states-exerting-power-online-sharing-data-can-guard-again.html). In October 2020, we similarly worked with the government's Anti-Muslim Hatred Working Group to highlight how data transparency had enabled them to conduct critical research (https://blog.twitter.com/en_gb/topics/company/2020/twitteruk-amhwguk-working-partnership.html). Our work to increase transparency efforts across the company is tireless and constant, and we welcome opportunities to support publicly available data being used to advance research objectives in a safe, compliant way with the public's basic expectation of privacy.
- Part of our transparency efforts also include our biannual Twitter Transparency Report, which we've produced since July 2012 to share global trends across a number of areas of our enforcement on Twitter, including the Twitter Rules and legal requests we receive.
- The report is ever-evolving. In August 2020, for example, we launched the Twitter Transparency Centre. Our goal with this evolution is make our transparency reporting more easily understood and accessible to the general public. This includes:
 - Brand new website that includes all our disclosed data in one place
 - Data visualizations making it easier to compare trends over time
 - Country comparison module
 - Tooltips to help explain key terms and provide more insights on the terms we use
 - History of transparency milestones and updates
 - New metrics and methodology on the enforcement of the Twitter Rules (from July 2018 through December 2019)
 - New policy categories to better align with the Twitter Rules

- We also have started to include state-backed information operations datasets, which were released to the public to empower research and awareness of these campaigns. We now host over 35 different datasets that we believe are connected to state-backed information operations, and hosted our first research workshop on the data with the Carnegie Partnership for Countering Influence Operations in July 2020.
- We want to empower research that can advance public understanding of critical issues online. For example, to further support Twitter’s ongoing efforts to protect the public conversation and help people find authoritative health information around COVID-19, we released a new endpoint into Twitter Developer Labs in April 2020. This is to enable approved developers and researchers to study the public conversation about COVID-19 in real-time. The data can be used to research a range of topics related to the coronavirus pandemic, including areas like the spread of the disease, the spread of misinformation, crisis management within communities and more. Making this access available for free is one of the most unique and valuable things Twitter can do as the world comes together to protect our communities and seek answers to pressing challenges.
- We remain deeply committed to transparency at Twitter - it continues to be one of our key guiding principles. We have welcomed the opportunity to participate in the working group, and look forward to continuing to work with government, civil society and the wider community on these important issues.

Facebook

- Facebook supports the idea of a regulatory framework for online content that ensures companies are making decisions about online speech in a way that minimizes harm but also respects the fundamental right to free expression. Regulation should seek to balance these often conflicting issues and bring more procedural accountability for platforms. In order to do so, regulation should include requirements for companies to publish their content standards, create mechanisms to report violations of these standards, provide response to decisions, and notice when content is removed.
- Facebook is transparent about its Community Standards, the global set of policies that outlines what is and is not allowed on Facebook and is publicly available on our website. Our Community Standards apply to everyone, all around the world, and to all types of content. They are based on feedback from our community and the advice of experts in fields such as technology, public safety and human rights, collected in various forms, including via our Content Policy Forum, a meeting that we hold every two weeks to discuss new policies or amendments to existing policies. Minutes from these meetings are public.
- It is also important for service providers to provide transparency into how their systems are performing, however the type and nature of the measurements or metrics used should not be so fixed as to hinder efforts to accurately respond to the changing dynamics of reporting on online content. Different types of services may require different levels of transparency, depending on the size and nature of the provider.

- Facebook already does so by sharing regular transparency and enforcement reports, such as the Community Standards enforcement report detailing how much content we remove for violating certain of our policies, how much of that content was detected proactively by our automated tools, how much content was appealed when people believed we had made a mistake, and how many of those appeals were successful. Additionally we regularly publish another report that includes metrics on the number and nature of legal requests we receive from governments and other entities around the world – including requests for data and requests to restrict access to content which they believe violates local law.
- Given the dynamic nature and scale of online speech and the different expectations of users of their experience online, any system operating at scale and for a global user base will be imperfect. For this reason, in order to safeguard freedom of expression, it is essential for platforms to be transparent about its decisions and have appropriate redress mechanisms. Facebook provides feedback and updates to users that report content and informs users whose content has been removed. Additionally, we give users the possibility to appeal our decisions regarding certain content that we took action on and certain content that was reported but not acted on.



HM Government