Question	Your response
We welcome input from industry on the areas listed below. We encourage stakeholders to respond with feedback so that we can ensure that the guidance helps providers and other stakeholders understand:	
A) Ofcom's powers and providers' duties for transparency reporting, as well as Ofcom's approach to implementing the transparency regime. B) Ofcom's approach for determining what information service providers should produce in their transparency reports. C) Ofcom's plans to engage with providers prior to issuing transparency notices, and on what matters, and whether the proposed engagement plan will be sufficient for helping services to comply with their duties. D) Ofcom's plans to use the information in providers' transparency reports in Ofcom's own transparency reports.	Confidential? – N B) We suggest incorporating aspects of the Integrity Institute's Transparency of Risks framework. (See How the Risk or Problem Manifests on the Platform Section) It is essential that Ofcom request metrics that comprehensively cover the scale, cause, and nature of harms that occur on platforms for Ofcom to be able to understand platforms and meet their goals such as: Strengthening safety governance in online systems Ensuring platforms are designed and operated with safety in mind Promoting trust in services' safety measures. The Risk Assessments and Mitigations Report includes metrics such as: Scale: How many users are exposed to known violating content or content involved in the risk in a reasonable time window like daily, weekly, or monthly? How many total exposures to known violating or risky content in a reasonable time window? The average views and reach of harmful content before it is moderated? The average time it takes to moderate? (If some platform moderates 100% of hateful content, but it is viewed by 100% of users

	before it is moderated, it is still extremely harmful) Cause: What fraction of exposures were due to platform recommendations? What % are from creators the user follows? What % are from DM's? What % are from pages or groups? What % are reels or shorts? Are there other surfaces/features the are high risk? (It is good to know if there is a disproportionately risky feature or risk for certain groups) Nature: How are exposures to harms distributed among users? (Evenly among all, or concentrated on a subset?) Is there a particular, vulnerable demographic group that experiences outsized harm? (Specificity on segmenting users is important to compare across general demographics like age, gender, languages or countries, provides a much more accurate assessment of what harms exist on the platform) (For additional metrics and how to use the framework, see the Assessing the Current Scale, Cause, and Nature of the Risk from the report, the Metrics and Transparency Deck, or the best practices in Ranking Design Transparency) The good news is that most of these metrics are already collected by the companies and are a single query away. Some include these metrics are already reported by the companies in existing transparency reports and public data sets, so for larger platforms, this shouldn't be a capacity issue to provide.

	These metric frameworks highlight the role that platform design, operation, and governance play in exposures to risky content. This should be made clear as a goal for Ofcom of transparency. For instance, one specific recommendation on the goals of transparency is to change the language of the Ofcom goals from “(iv) promotes trust in services safety measures” to “promotes trust that the services are designed, operated, and governed safely” Some additional specific metrics to include: Moderation precision and recall across harm areas as well as policy verticals, content moderation outcomes across countries and regions, languages, user groups, etc. in which the service is offered, and speed of response broken down by the factors listed above. It is our recommendation to ask companies to provide this info for all of the harms they share in their transparency reports. Reports of the total number of pieces of content removed, without sharing the accuracy of their classifiers or models, lack substance if measurement errors, like precision and recall, are not also included. Prevalence is a tempting metric to watch across platforms, but a misleading one. It does not account for the real prominence of content. Such figures are highly variable between platforms, fairly easy to manipulate to the platforms' preferred messaging, and incomplete. Citing the fraction of content that gets 'served' to users does not equal prominence within a timeline, likelihood that the user actually saw that content, time spent with the content, etc. It also flattens the differences between the impact of violative content. For example, there is a big difference in the impact on people (particularly children) between severe gore and mild violence, between light nudity and pornography, or between a mention of an extremist group and effective recruitment by an extremist group.

	Yet all kinds of content under a policy appear as one marker for 'prevalence'. More nuance and qualitative understanding is required. These nuances could be tracked in the public through releases of public data sets of random samples of public content, weighted by impressions, and by making data sets of violating content available to researchers with relevant expertise. When thinking about what information service providers should produce in their transparency reports, it is important to find a balance between comparable and bespoke statistics. Achieving comparable metrics across platforms is challenging as only a small number of truly comparable statistics exist and a more critical factor is the comprehensiveness of the metrics rather than the uniformity. Platforms face unique challenges and use different variables. It is especially difficult to find comparable statistics considering the unique context facing all the major platforms as even similar numbers or variables can have different meanings across platform designs, user contexts and policy details. Additionally, platforms including or not including certain features massively contribute to the types of harms most likely on the platform.(Ex. Child sexual exploitation on platforms with direct messages vs platforms without, platforms with end-to-end encryption compared to platforms without) These design choices create complexity in understanding the nature of the harms on the platform. It is our recommendation that while it is a desirable goal to track some metrics that enable platforms to be compared to each other, and that should be done when possible, that shouldn't be done at the expense of platforms providing a comprehensive view into their safety. C) It would be helpful for Ofcom to provide examples and a range of alternative

	metrics/info that platforms could provide and still reach compliance. The overarching engagement strategy sounds sufficient, but Ofcom may want to consider a structured opportunity for platforms to ask clarification questions in response to the notices D) We recommend that if the providers' transparency reports are kept private, then Ofcom should strive to include as much (nonsensitive) data as possible from the providers' reports in their reports. It is essential that a comprehensive review of platform safety be possible with publicly available data.
Are there any aspects in the draft guidance where it would be helpful for additional detail or clarity to be provided?	Confidential? – N
Are the suggested engagement activities set out in the draft guidance sufficient for providers to understand their duties and Ofcom's expectations?	Confidential? – N

Question	Your response
We are also seeking input that will help us understand if there are other matters that Ofcom should consider in our approach to determining the notices, beyond those that we set out in the guidance. The questions below seek input about any additional factors Ofcom should take into account in various stages of the process, including: to inform the content of transparency notices; in determining the format of providers' transparency reports; and how the capacity of a provider can be best determined and evidenced.	
Are there any other factors that Ofcom might consider in our approach to determining the contents of notices that are not set out in the draft guidance?	Confidential? – N
Is there anything that Ofcom should have regard to (other than the factors discussed in the draft guidance) that may be relevant to the production of provider transparency reports? This might include factors that we should consider when deciding how much time to give providers to publish their transparency reports.	Confidential? – N Collecting statistics for internal use is normally easy; the difficulties from meeting these transparency requirements are primarily meeting legal requirements about due

	diligence, compliance, and ensuring numbers are as accurate as possible. Perfect accuracy could be challenging if not impossible in many situations. And the overall process of making sure the data meets accuracy and compliance guidelines will generally be more challenging than the making of the datasets themselves, from the datasets that companies normally make for their business decisions. The time it takes for companies to deliver reports could be a good metric to track. When it takes a company a long time to produce transparency reports, it could mean they don't already collect these statistics automatically and reliably, do not have confidence in their own data, or wish to reframe statistics to fit a narrative presentation more beneficial to them. Pushing for a relatively quick turnaround (between 1-2 months) also pushes them to automate good basic statistics, which not only increases the timeliness for regulators but is also useful for practitioners within the company. Certain metrics may require gathering data across different teams, systems, countries, etc. which could require significant time and coordination across teams. Also, accuracy, prevalence, etc. are continuously changing across policy areas, languages, timeframes, etc. so requests should be as specific as possible to yield the intended insights.
What are the anticipated dependencies for producing transparency reports including in relation to any internal administrative processes and governance which may affect the timelines for producing reports? What information would be most useful for Ofcom to consider when assessing a provider's "capacity", by which we mean, the financial resources of the provider, and the level of technical expertise which is available to the	Confidential? – N The primary factor should be the size of the platform, measured by the number of users. Ideally, capacity should be closely tied to the size of the platform, but that is not always the case. A platform could have very few employees and generate very little revenue, but could still have close to a billion users globally, and thus have little capacity for

service provider given its size and financial resources?	compliance. However, companies with extremely large user bases and very few employees should be seen as problematic, and they should not be able to use an irresponsibly low number of employees focused on platform safety as an excuse to avoid compliance. Companies should have an obligation to keep their users safe or face regulatory consequences. Total revenue or total profit should also be considered, because a highly profitable platform that has a lower number of users could still be considered risky. However, we recommend avoiding the use of the number of employees, because using employee counts as a measure of capacity for compliance purposes creates an incentive to keep the total number of employees low.
Are there any matters within Schedule 8, Parts 1 and 2 of Act that may pose risks relating to confidentiality or commercial sensitivity as regards service providers, services or service users if published?	It is our recommendation to consider that the majority of the data coming in is unlikely to be information that would be confidential or commercially sensitive. Aggregate statistics pose little risk to user privacy. Regarding the algorithmic systems, we already have examples of voluntary transparency that demonstrate there is little risk of commercial sensitivity. Twitter/X made the source code of their ranking systems open source without issue, and other companies have released the exact signals they collect from users for use in their ranking systems without any concerns. Meaningful transparency is needed to create more positive external incentives for the companies to build safe platforms. It is crucial that comprehensive transparency be made publicly available for users or civil society to understand the state of the platforms. For resources on the importance of meaningful transparency, see:

	Integrity Institute Best Practices for Platform Transparency Making Social Media Safer Requires Meaningful Transparency

Question	Your response
Finally, we are also seeking input into any matter that may be helpful for ensuring Ofcom's transparency reports are useful and accessible.	
Beyond the requirements of the Act, are there any forms of insight that it would be useful for Ofcom to include in our own transparency reports? Why would that information be useful and how could you or a third party use it?	Confidential? – N
Do you have any comment on the most useful format(s) of services' transparency reports or Ofcom's transparency reports? How can Ofcom ensure that its own transparency reports are accessible? Provide specific evidence, if possible, of which formats are particularly effective for which audiences.	Confidential? – N We recommend APIs or public data sets released in JSON or CSV formats, which should be sufficient at sharing the data. However, we stress that accessing them should not require extensive coding knowledge. Any reasonable user should be able to get the basic summaries and crosstabs. In order to create a truly accessible framework, access should not be limited to only those with technical skills.

Question	Your response
Please provide any other comments you may have.	
General comments	Confidential? – N

Please complete this form in full and return to OS-Transparency@Ofcom.org.uk