Your response

Question (Volume 2)	Your response
Question 6.1: Do you have any comments on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.	*[Is this answer confidential? **Some**]* The Digital Trust & Safety Partnership is a content- and product-agnostic approach to best practices for trust and safety. DTSP offers best practices and assessment methodologies that can be applied by any organization that offers a public-facing digital product or service for which it conducts trust and safety operations, or otherwise implements controls to manage content- or conduct-related risk. The DTSP Best Practices and the Safe Framework have been deployed by some partner companies as part of their systemic risk assessments under the EU Digital Services Act, and have been recognized as an example of best practice for risk assessment in industry codes and standards in Australia.[9] Given our content-neutral approach, in which it is up to each organization to determine the content- or conduct-related risks applicable to its products and/or services, DTSP will refrain from commenting on Ofcom's assessment of the causes and impacts of specific sources of potential harm. However, the Safe Framework offers a rigorous methodology by which organizations can identify risks, which is broadly compatible with the approach and guidance that Ofcom has published. [✂]

[9] For example, see https://www.esafety.gov.au/sites/default/files/2023-06/Schedule-1%E2%80%93Social-Media-Services-Online-Safety-Code-%28Class-1A-and-Class-1B-Material%29.pdf.

Question (Volume 2)	Your response
Question 6.2: Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer.	*[Is this answer confidential?* ***No****]* The Safe Framework, while broader in application than any one specific source of harm, does incorporate into its Tailoring Framework a Risk Profile Questionnaire, used to measure the extent to which a product or service's features and policies implicate content- or conduct-related risks. See our response to Question 9.1 for more detail on this matter.

Question (Volume 3)	Your response
Question 8.1: Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice? Please provide underlying arguments and evidence of efficacy or risks to support your view.	*[Is this answer confidential?* ***No*** *]* The Ofcom recommendation regarding Governance and Accountability includes an annual review of risk management activities: "The provider's most senior governance body in relation to the service should carry out and record an annual review of risk management activities in relation to online safety, and how developing risks are being monitored and managed." The Code recommends the provider to have an internal monitoring and assurance function to provide independent assurance and report to: the body that is responsible for overall governance and strategic direction of a service or an audit committee. The independent assurance may be provided by an existing internal audit function. As DTSP has a risk-based approach, its members already undertake review of risk management externally and internally. The other recommendations such as written statements of responsibilities are not set out at a granular level of detail in the DTSP Best Practices Framework. However, under PE1.1, we do include "Constitute roles and/or teams within the company accountable for policy creation, evaluation, implementation, and operations." Most of the best practices under product development such as risk assessment and pre-launch and post-launch evaluations also align with an internal monitoring and assurance function.

Question (Volume 3)	Your response
Question 8.2: Do you agree with the types of services that we propose the governance and accountability measures should apply to?	*[Is this answer confidential? **No**]* All participating DTSP partners agree on five fundamental commitments that a digital service should make to promote a safer and more trustworthy internet. Also known as the DTSP Best Practices Framework, we regard the following to be industry best practices: **Product Development:** Identify, evaluate, and adjust for content- and conduct-related risks in product development. **Product Governance:** Adopt explainable processes for product governance, including which team is responsible for creating rules, and how rules are evolved. **Product Enforcement:** Conduct enforcement operations to implement product governance. **Product Improvement:** Assess and improve processes associated with content- and conduct-related risks. **Product Transparency:** Ensure that relevant trust and safety policies are published to the public, and report periodically to the public and other stakeholders regarding actions taken. These five commitments are underpinned by 35 specific (but non-comprehensive) best practices, which provide concrete though non-exclusive examples of the variety of activities and processes that organizations may have in place to mitigate risks from harmful content and conduct, depending on their particular product and risk model. There is some overlap between the Product Governance commitment in the DTSP Best Practices Framework and the proposed governance and accountability measures. While DTSP does not require its partners to have a senior governance body to carry out the annual review of risk management, the practices include designating a team or manager accountable for integrating trust and safety feedback. DTSP does not prescribe which practices different types of services should employ. Instead DTSP takes a tiered approach to the level of assessment of implementation of these practices by companies.

Question (Volume 3)	Your response
	DTSP uses three levels of assessment that a company may undertake to examine trust and safety practices in support of a particular product, digital service, or function. The Level 3 assessment is designed as the most in-depth in terms of the breadth and depth of assessment procedures, while Level 1 is less detailed and provides for more summary-level analysis, with Level 2 falling in the middle. Using a risk-based approach, the depth of assessment is then determined by evaluating the size and scale of the organization, as well as the potential impact of its product or service.
Question 9.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential? **Some**]* There is significant overlap and complementarity between the proposed Service Risk Assessment Guidance and the Safe Framework. The Safe Framework assessment that DTSP members undertake aligns with Ofcom's recommended measures to undertake annual risk review. Ofcom lays out a four-step risk assessment process: 1)understand the harms that need to be assessed; 2)assess risk by considering the likelihood and potential impact of harms occurring on their service; 3)implement safety measures and record outcomes of the risk assessment; and 4)report, review and update the risk assessment. DTSP's approach is generally similar and strongly aligns with step three and four: implementation of safety measures and documenting risk assessment. Rather than agreeing or disagreeing with the proposals, we are providing evidence based on the Safe Framework and its implementation to help inform Ofcom's approach. DTSP appreciates the references to our work in Chapter 9 and how they have informed Ofcom's "scalable approach which allows services to differentiate based on their size, nature and likely levels of risk." [✂]

Question (Volume 3)	Your response
	[✂]

Question (Volume 4)	Your response
Question 11.1: Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?	*[Is this answer confidential? **No**]* There is once again substantial overlap between the Illegal Content Codes of Practice and the DTSP Best Practices Framework, even as these two documents serve specific and different purposes. As a globally applicable and voluntary framework, DTSP does not aim to provide granular guidance on illegal content in any particular jurisdiction.
Question 11.2: Do you agree that in general we should apply the most onerous measures in our Codes only to services which are large and/or medium or high risk?	*[Is this answer confidential? **No**]* Although DTSP agrees with the principle of proportionality, and that more onerous measures should be tiered in their application, in our view the Ofcom methodology risks overburdening smaller companies based on assumptions about the risks facing certain types of services. Attention to product design and other mitigating details could help redress this deficiency. DTSP takes a different approach to tiering obligations based on size and risk. As described in our previous responses and shown in the Safe Framework, companies

Question (Volume 4)	Your response
	may employ whatever combination of practices fulfill their overall commitments to trust and safety, but the level of assessment of those practices is based on a combination of the organization's size and scale as well as product impact, which is a combination of user volume and risk factors.
Question 11.3: Do you agree with our definition of large services?	*[Is this answer confidential?* ***No****]* The Safe Framework considers user volume to be an element of product impact and looks at annual revenue and number of employees for products/services in scope of assessment as determinants of organizational size and scale. Although user volume is an important factor that is included in our framework, viewing the size of a service only as a function of the number of users may impede the scalability of the Codes of Practice.
Question 11.4: Do you agree with our definition of multi-risk services?	*[Is this answer confidential?* ***No****]* The DTSP Best Practices Framework, as a content- and product-agnostic framework of practices, offers an inherently multi-risk approach. As indicated previously, the proportionality of our approach is derived from having companies use whatever combination of practices enables them to fulfill their commitments, and then evaluating the implementation of these practices in a proportionate manner using the Safe Framework.
Question 11.6: Do you have any comments on the draft Codes of Practice themselves?[10]	*[Is this answer confidential?* ***No****]* The draft Codes (Annex 7) provides interpretations and definitions. The similarities and differences of the definitions and interpretations between DTSP glossary and Ofcom codes can provide some clarity and more alignment.[11] It is important to have a common understanding of key terminology across organizations responsible for trust and safety online, including industry actors as well as partners in government and civil society. To this end, DTSP released an inaugural edition of its Trust & Safety Glossary of

[10] See Annexes 7 and 8.

[11] Codes, Section A 11, page 41, Definitions and interpretations https://www.ofcom.org.uk/__data/assets/pdf_file/0022/271165/annex-7-illegal-harms-consultation.pdf.

Question (Volume 4)	Your response
	Terms.[12] This is the first industry effort by technology companies, representing various products, sizes, and business models, to develop a common trust and safety lexicon. The Trust & Safety Glossary of Terms consists of more than 100 terms across four categories: ● content concepts and policies; ● common types of abuse; ● enforcement practices; and ● trust and safety technology. The glossary has been updated to incorporate valuable input received from academic organizations, industry partners, regulators, and other global stakeholders during the public consultation held earlier this year, including input from Ofcom staff.
Question 12.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	[Is this answer confidential? **No**] As noted elsewhere, DTSP does not mandate which types or sizes of services should undertake particular practices. Instead, the Safe Framework provides a tiered approach to evaluation of whichever practices a partner company employs to manage content- and conduct-related risks. DTSP takes this approach because the specific practices a service employs to address risks will depend on the design of the product or service. Not every practice is suitable for every product or service. The rigid safe harbor system proposed by Ofcom, which increases the burden on organizations with alternatively designed or structured services increases their compliance risk and administrative burden by requiring them to justify the use of specific practices, which could discourage safety innovation and keep the codes of practice from being adequately future proofed. The proposed content moderation systems requirements generally overlap with components of the Product Governance, Product Enforcement, and Product Improvement commitments under the DTSP Best Practices Framework, specifically:

[12] See https://dtspartnership.org/wp-content/uploads/2023/07/DTSP_Trust-Safety-Glossary_July-2023.pdf.

Question (Volume 4)	Your response
	Having a content moderation function that allows for the swift take down of illegal content is similar to the Product Enforcement commitment, “Conduct enforcement operations to implement product governance” which has the aim of ensuring operations exist to implement the aims set forth in Product Governance to address Content- and Conduct-Related Risks. The process and systems measures in the proposed code align significantly with the DTSP Best Practices Framework: ***Setting internal content policies*** is similar to PG7, “Document for internal use the interpretation of policy rules and their application based on precedent or other forms of investigation, research, and analysis” ***Performance targets*** is similar to PE6.1, “Operationalize enforcement actions at scale where: Standards are set for timely response and prioritization based on factors including the context of the product, the nature, urgency, and scope of potential harm, likely efficacy of intervention, and source of report” ***Prioritisation*** is similar to PE1.2, “Develop and review operational infrastructure facilitating the sorting of reports of violations and escalation paths for more complex issues” ***Resourcing*** is similar to PI3, “Use risk assessments to determine allocation of resources for emerging Content- and Conduct-Related Risks” ***Provision of training and materials to moderators*** is similar to PE2, “Formalize training and awareness programs to keep pace with dynamic online content and related issues, to inform the design of associated solutions” The DTSP Best Practices Framework includes several other examples of best practices that are relevant to content moderation. These include: PG7, “Facilitate self-regulation by the user or community to occur where appropriate, for example by providing forums for community-led governance or tools for community moderation and find opportunities to educate users on policies, for example, when they violate the rules” PE3, “Invest in wellness and resilience of teams dealing with sensitive materials, such as tools and processes to

Question (Volume 4)	Your response
	reduce exposure, employee training, rotations on/off content review, and benefits like counseling”
Question 13.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	[Is this answer confidential? **No**] The DTSP Best Practices Framework and Safe Framework assessments are applicable to search, but we are not prescriptive about which practices search must use. The DTSP Best Practices are non-exclusive, and provide the option for organizations to employ other practices, dependent on their unique design and features, to fulfill the overarching commitments.
Question 14.1: Do you agree with our proposals? Do you have any views on our three proposals, i.e. CSAM hash matching, CSAM URL detection and fraud keyword detection? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential? **No**]* The DTSP Best Practices Framework is not prescriptive about particular technological approaches. Our relevant best practice is PE4, “Where feasible and appropriate, identify areas where advance detection, and potentially intervention, is warranted” It is through Safe Framework assessments that companies evaluate the specific risks and controls related to this practice.
Question 14.3: Do you have any relevant evidence on: ● The accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services; ● The ability of services in scope of the CSAM hash matching measure to access hash databases/services, with respect to access criteria or requirements set by database and/or hash matching service providers; ● The costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of	*[Is this answer confidential? **No**]* The DTSP Best Practices Framework is not prescriptive about particular technological approaches. Our relevant best practice is PE4, “Where feasible and appropriate, identify areas where advance detection, and potentially intervention, is warranted” It is through Safe Framework assessments that companies evaluate the specific risks and controls related to this practice.

Question (Volume 4)	Your response
fuzzy matching[13] for CSAM URL detection; - The costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services; and - An effective application of hash matching and/or URL detection for terrorism content, including how such measures could address concerns around ‘context’ and freedom of expression, and any information you have on the costs and efficacy of applying hash matching and URL detection for terrorism content to a range of services.	
Question 15.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential?* ***No****]* The DTSP Best Practices Framework and Safe Framework assessments are applicable to search, but we are not prescriptive about which practices search must use. The DTSP Best Practices are non-exclusive, and provide the option for organizations to employ other practices, dependent on their unique design and features, to fulfill the overarching commitments.
Question 16.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential?* ***No****]* Several of the DTSP Best Practices overlap with the user complaints and reporting requirements. These include: PE5, “Implement method(s) by which content, conduct, or a user account can be easily reported as potentially violating policy (such as in-product reporting flow, easily findable forms, or designated email address)”

[13] Fuzzy matching can allow a match between U2U content and a URL list, despite the text not being exactly the same.

Question (Volume 4)	Your response
	PE6.2, "Operationalize enforcement actions at scale where: Appeals of decisions or other appropriate access to remedy are available" PE7, "Ensure relevant processes exist that enable users or others to "flag" or report content, conduct, or a user account as potentially violating policy, and enforcement options on that basis" PI5, "Establish appropriate remedy mechanisms for users that have been directly affected by moderation decisions such as content removal, account suspension or termination"
Question 17.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential?* ***No****]* Several of the DTSP Best Practices overlap with the proposed Terms of Service and Publicly Available Statements requirements. In particular, under Product Governance, DTSP partners commit to: "Adopt explainable processes for product governance, including which team is responsible for creating rules, and how rules are evolved." Specific example best practices in this area include: PG1, "Establish a team or function that develops, maintains, and updates the company's corpus of content, conduct, and/or acceptable use policies." PG2, " Institute processes for taking user considerations into account when drafting and updating relevant Product Governance" PG3, "Develop user-facing policy descriptions and explanations in easy-to-understand language" PG4, "Create mechanisms to incorporate user community input and user research into policy rules" PG5, "Work with recognized third-party civil society groups and experts for input on policies"
Question 18.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential? Yes /* ***No****]* The DTSP Best Practices Framework and Safe Framework assessments are not prescriptive about default settings and user support for child users.

Question (Volume 4)	Your response
Question 19.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential?* ***No****]* While not recommender specific, our best practices include effectiveness testing in product improvement. In product development, abuse pattern analysis, mitigation and control, and monitoring and evaluation might also be useful practices for testing recommender products and systems.
Question 20.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential?* ***No****]* Generally DTSP's practices align with enhanced user control proposals by Ofcom. However, the DTSP Best Practices Framework identifies holistic practices that do not compel granular or prescriptive outcomes. The practices that are aligned with some of Ofcom's suggestions are: PD8, "Iterate product in light of Trust & Safety considerations including based on user feedback or other observed effects, including ensuring that the perspectives of minority and underrepresented communities are represented" PD9, "Adopt appropriate technical measures that help users to control their own product experience where appropriate (such as blocking or muting)" PI4, "Foster communication pathways between the Practicing Company on the one hand, and users and other stakeholders (such as civil society and human rights groups) to update on developments, and gather feedback about the social impact of product and areas to improve"
Question 22.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*[Is this answer confidential?* ***No****]* DTSP practices are not service specific and not prescriptive. Some of the general practices that apply to Ofcom suggestions include: PE5, "Implement method(s) by which content, conduct, or a user account can be easily reported as potentially violating policy (such as in-product reporting flow, easily findable forms, or designated email address)"

Please complete this form in full and return to IHconsultation@ofcom.org.uk.