Your response

Question 1: Do you have any views on our audit-based assessment, including our proposed principles, objectives, and the scoring system? Please provide evidence to support your response	We generally support the audit based approach, and specifically the flexibility and future-proofing that it provides to ensure that technologies submitted for accreditation can be reviewed against and appropriate and holistic set of requirements, including performance, scalability and privacy protection. **Technology vs. Data** The term "Technology" is used in this material in quite a general way, and I believe it would be useful to provide a clear definition as I believe the way it is used here differs from common usage. "Technology" is defined by Oxford Languages as "the application of scientific knowledge for practical purposes, especially in industry." or "machinery and equipment developed from the application of scientific knowledge." We would then expect "Technology" to be packaged up into an "Implementation" or "Software" or "Tool", wrapping the core concept or algorithm from the "Technology" in a practical form that makes it usable in the real world. When the "Tool" based on the "Technology" is made available along with all the other things that might be needed (such as data), then it becomes a "Product" or "Solution" or Service" for consumption. In the digital world it has historically been the case that there is some degree of separation between these concepts. For example: • "matching the hash of a file against a database of hashes of known CSAM" is an example of what we usually think of as a technology. • "a C++ software library that takes a file and compares it with a CSV file database of hashes of known CSAM using MD5 hashes" is what we think of as a tool or implementation. • "SaaS using that C++ library that returns a value reflecting whether a provided file matches the

	Internet Watch Foundation database" is a solution. We believe that the conceptual difference is important, and that in describing what is to be accredited, the consultation material envisages complete solutions or services, rather than what some might expect to be described as "technology". At very least highlighting exactly what the scope of accreditation is meant to be in the documents seems really important and is missing. We believe that a definition of "technology" that was consistent with the use of the term in this consultation would be significantly different from the common usage of the term. This draws into question whether the interpretation of the legislation is correct or appropriate. If section 121 is read with a common usage view of what the word "technology" means, one might reasonably imagine that it was possible to require (under a technology notice) and accredit a technology such as "matching the hash of a file against a database of hashes of known CSAM" If such an accreditation were given, this would apply to ANY solution using such technology (whether developed in-house or otherwise). Since most large platforms have their own implementations of hash-matching, the approach taken by Ofcom appears to requires EACH to be independently accredited in testable form. This places considerable additional burden on platforms and technology providers. **Proposals:** • The definition of "technology" for the purposes of technology notices and accreditation used by Ofcom in preparing the consultation materials should be explicitly provided. • Ofcom should justify why it has used this definition of technology by reference to the Online Safety Act or other evidence • Ofcom should review whether this audit process is in fact targeted at "technology" as intended in the Act, or whether its interpretation is distorting or misinterpreting the intention of the Act. **Technology vs. Deployment Characteristics.**

	Some of the questions in the example questions for audit raise a further questions about what is being accredited. Imagine a provider offering a service somewhat similar to Microsoft's current PhotoDNA offering (but better in some key way). The provider currently has a handful of small customers, but there is little market demand from larger players. Imagine Ofcom sees the potential for this technology to address concerns it has about harms on a very large platform. Ofcom might be pleased to receive an application to accredit the technology. As the technology currently serves only small customers, it is currently only running on a single server within a cloud service provider. It is clear that this solution as deployed today would not have the capacity, throughput or redundancy needed for deployment by a large platform. However, the solution has been developed in such a way that it would be straightforward to deploy it across many servers behind load balancers with a high degree of geographic spread resulting in the required capacity throughput and redundancy (and it is obvious that it is so) that a large platform would need. Although it is obvious that it can be done, it has not been done yet (as there is no demand and it would be costly) so the provider cannot demonstrate or provide detailed metrics for the scaled version of the solution. It is the intention of the accreditation process that such a solution would fail accreditation (as it has not been demonstrated at the required scale), or is it the intention that responses to the audit questions can describe approaches to scaling and if sufficient information is given could achieve accreditation? The answer to this question is very important in determining the barriers to new technologies and new providers achieving accreditation. Cyacomb is a relatively small company with an inherently scalable technology. It is our view that the questionnaire contains many questions which are about deployment of a technology rather than the characteristics of the technology itself, and that these should be separated out. It is our view that in terms of metrics relating to core characteristics such as accuracy, bias or privacy protection audit should look at the current demonstrated capabilities of a solution, but that when considering deployment

	characteristics (throughput, latency, scalability) the audit should consider whether there is a clear path to achieving appropriate outcomes (no major blockers, standard engineering approaches clearly apply) rather than just the current state. Some technologies can be deployed in many different ways. For example, Cyacomb offers the same technology packaged as SaaS (Software as a Service) for small platforms, but no large platform is likely to want such external dependencies. Cyacomb offers technology for direct integration into large platforms, which enables them to ensure scalability and redundancy/availability in line with their own needs. Separating out core technology capabilities from deployment related matters could also have the potential to ensure that a technology didn't require separate accreditation just because it was being deployed in different ways with different deployment characteristics. **Proposals:** • The definition of "technology" for the purposes of technology notices and accreditation used by Ofcom in preparing the consultation materials should be explicitly provided. • The audit questions should be separate core capabilities of a technology from matters relating to its deployment. • Audit guidance should allow that for many possible technologies scaling and performance is a fairly trivial challenge, and not unduly penalise services that have not yet scaled if there is a very clear path for them to do so. **Continuous Improvement** Most software today operates on a basis of continuous improvement. This results in software being frequently updated. Some updates have no intended changes to functionality, but are carried out to ensure the latest security patches are applied to libraries, and the latest operating systems and devices are supported. Other updates are intended to fix bugs – changing functionality only for the better. More major updates may aim to improve performance (accuracy or speed), or to add significant new functionality. Some software goes through

	very significant restructuring or rewrites to reduce technical debt or provide for future expansion. Any of these updates has the potential to affect the performance of software, and therefore the performance of the technology it implements. The consultation documents appear to describe accreditation as applying to (software) solutions, rather than to technologies. They do not provide any information as to whether accreditation applies ONLY to the tested version of software (completely unworkable) or to ANY version of software that the provider chooses to claim is the same (with the risk that supposed improvements could change the characteristics such that it would not have been accredited). In other domains it is common for there to be guidance allowing accreditation to be retained by updated versions of software based on some set of risk criteria and manufacturers own testing. For example, software used in Digital Forensics (under the application of ISO17025 and as required by the Forensic Science Regulator) Police Forces can use updated software where it can be validated and shown to be compatible with the existing accredited process. **Proposals:** • Ofcom should provide guidance over what sort of changes to technologies allow them to remain accredited, and what would take them outside the original accreditation. • Ofcom could consider allowing provider testing and self-certification, or a light-weight re-accreditation process, to ensure technology can continue to improve (and maintain compatibility and security) without undue burden. **Technology and Data** The apparent broad definition of technology being used and the nature of the audit questions (and parameters for Independent Performance Testing if required) suggests that the data used to drive technology is considered an integral part of it. A data-driven solution cannot be tested without data. Using an example related to hashing again, is it the intention that a technology accredited using e.g. the IWF

	hash dataset would be accredited only for use with that dataset? Would it only be accredited for use with the snapshot of the IWF data it was tested with? The IWF database keeps growing – can the technology use the updated database is it improved without falling out of accreditation? What limits might need to be placed on use of updated data? While the IWF has established strict governance and consistent standards over many years, there are other databases (e.g. commercial databases of terrorism related content) that operate much more opaquely. Is there a difference between how these should be viewed in terms of the risk/benefit of allowing updates? **Proposals:** • Clarify the role of data in data-driven solutions. Is specific data part of the "technology" for accreditation, or can it be considered separately. • How would updates to data affect accreditation?
Question 2: Do you have any views on our proposals for independent performance testing, including the two mechanisms for setting thresholds; the approach to testing technologies in categories against particular metrics; and data considerations? Please provide evidence to support your response.	The responses we have given to Question 1 are all relevant to this question also. Defining what is being tested and accredited is crucial to defining appropriate methodology and datasets. With respect to the mechanisms for setting thresholds, we recognise the risks identified at 4.78. Necessarily the setting of thresholds creates these risks. Appropriate granularity of testing categories will probably have at least as great an impact on these risks as the setting of thresholds. If the categories are too broad, then technologies which appear to perform "poorly" may in fact just have specialised differently. For example, it is reasonable to expect that a CSAM detection model designed to work in real-time on a user device (mobile phone) to prevent livestreaming on smaller platforms may perform "poorly" in absolute terms compared with a model running in the cloud to protect a major social media site streaming video. We suspect these should be separate categories as so many characteristics would be different. However, taken to the opposite extreme most providers of technology try to differentiate by application (not just performance) and if categories are very

	narrowly defined, there could only be one or two solutions in each. The table at 4.66 seems to suggest categories which we believe are far too broad to represent the diversity of applications of technology, and therefore create a huge risk that some categories of technology (especially for high privacy environments and operating client side) may be un-accreditable by the design of the process. Technologies that are best-in-class and could deliver immense good could be unavailable as tools for Ofcom in issuing technology notices because the categories are too broad. We are far more concerned by the direction on categories than by the nature of threshold mechanisms. It may also be relevant, in considering categories, to consider whether the intended application of technology is prevention, mitigation, or response. Different thresholds will be relevant depending on the consequences of false positives, which are very different if blocking content or warning users when compared with reporting to law enforcement. In considering both mechansims, it is worth looking at how the risks noted at 4.78 may be otherwise mitigated. It appears that a technology notice may specify an accredited technology or more than one alternative. It does not appear to specify a category. It therefore appears to be the case the Ofcom is under no obligation to put any specific technology into a notice. If "poorly performing" technologies are accredited, Ofcom has the option not to specify them when issuing notices. The reverse is not true. If for some need there is no accredited technology, Ofcom has no option to issue a notice requiring the use of accredited technology, and must instead use the "best endeavours" approach which seems to us to be a weaker option from a regulatory perspective. Given that Ofcom always has the option NOT to include accredited technology in a technology notice, and the diversity of technologies, applications, business models and delivery models, we believe that excluding too much technology from accreditation is a greater risk than accrediting too much technology. We therefore believe Mechanism A is a more appropriate approach to thresholds.

Proposals

- Ofcom should revisit the question of categories and consider differentiating based on the intended application of technology as well as its implementation class. In particular, it is not reasonable to compare client-side technologies with cloud-based ones as resource availability is so different, yet each serves vital roles mitigating serious harms. The proposed categories are not fit for purpose.
- Ofcom should adopt Mechanism A for setting thresholds. Over-accrediting can be corrected in the framing of specific technology notices by selecting the best and most appropriate technologies. Under accrediting does not appear to have a similar mitigation.

Data Considerations

Solutions matching against databases (using hashing) can have the benefit of extraordinarily low false positive rates. However, true positive rate will depend as much on the composition of the test set as it does on performance. For example, if a data matching solution works using the IWF database, the test set…

- Could contain entirely images IN the IWF database - in which case true positive rate would be 100%
- Could contain entirely images NOT IN the IWF database - in which case the true positive rate would be 0%

The true positive rate will be determined by the composition of the test set as well as the capability of the technology.

The composition of the test set will be, to some degree, arbitrary. Our understanding is that there is no reliable data at platform scale on the prevalence of CSAM material overall, let alone the proportion of that material that is represented in a particular database, or the proportion of content that has been transformed or modified in different ways. Absent a large scale human-powered review of content that data cannot exist.

As long as database technologies are not tested against a dataset with 0% match then a generic test dataset could

	give measures of relative performance (assuming a sufficiently large test set) but the absolute metrics (false negative in particular) will have little real world meaning. We believe that database driven solutions, including hashing, have a vital contribution to make. The extraordinarily low false positive rate is essential in applications where high confidence is needed on data at large scale. They are also capable of matching images for victims where it is very hard to discern if their age is over or under the legal threshold (where AI approaches can struggle since the necessary age information cannot be derived from the pixels of an image). Ensuring appropriate test methodologies is therefore important, and test sets that allow meaningful testing may need different consideration than test sets for AI based methods. One useful methodology is to separate out the performance of the technology from the completeness of the database. For example, a technology may be able to match highly accurately against a database, and may be able to do so with little bias. This is separate to how much coverage the database achieves, how many false positives the database contains and how much bias there is in the database (e.g. over or under representation of certain groups). Generally hash matching technologies can target three different types of accuracy: • Cryptographic matching (exact file) • Similarity matching (targeting same image but allowing for resizing or recoding in transit) • Similarity matching (targeting visually similar but deliberately altered images e.g. cropping, mirroring etc) This suggests that in evaluating technology having corresponding test data sets would be useful. Otherwise the relative prevalence in the test dataset of original/resized/visually similar will have a bigger impact on results than performance. Table A3 lists data transformation examples. The re-sizing and re-compressing of images is carried out automatically by many apps. While re-sizing is mentioned in the table, re-compression is not. This is an important real-world test. We would suggest that transformations include realistic analogs for the processes of major tools

	and platforms (or ideally, data transformed using those platforms, although for CSAM this may be impractical).
	
	Proposals:
	
	• Ofcom should differentiate between the data requirements and interpretation of results relating to:
	  ○ Cryptographic matching
	  ○ Similarity matching (resize/recode)
	  ○ Similarity matching (more generally altered)
	  ○ AI/Machine Vision/Machine Learning solutions
	• Data transformations should include the most common operational re-sizing and re-compression types in addition to those already included. Ideally Ofcom would take an evidence based approach (and encourage research into) the prevalence of different transformations to ensure the dataset represents real world behaviour and activity.

Question 3: Do you have any comments on what Ofcom might consider in terms of how long technologies should be accredited for and how often technologies should be given the opportunity to apply for accreditation? Is there any further evidence we should consider?	The length of time for which accreditation would be appropriate should be considered in terms of the scope of accreditation and the environment of operation. **Scope of Accreditation** Under Question 1, "Continuous Improvement", we questioned how much change would be allowed without invalidating accreditation. This question is intrinsically linked with the question of duration. If only a single software version is accredited, then the accreditation should be short. Software generally needs at least regular security updates. Conversely if the accreditation allows for considerable continuous improvement, there is no reason why it cannot be for a much longer period. For example, hash based matching software has been in use for decades, and there is no reason why it could not be used for many more decades with little change from a technical perspective. Technology notices can apply for 36 months, and there is no reason to believe the timing of issue will align with when technologies are accredited. To permit Ofcom to

	issue a notice covering the maximum term, it seems desirable for the scope of the accreditation to be greater e.g. 4 years would allow 1 year between accreditation and issue of notice (recognising these are unrelated events) and then issue of a 36 month notice. 5 years would allow a 2 year period during which a maximum length notice could be issued etc. **External Environment** The appropriate duration for accreditation may also depend on the environment of operation. The technical environment may change. Some technologies are largely file format agnostic (e.g. cryptographic hashing) while others may need to support new image types (or variants) as they emerge. Generally these are slow moving changes (e.g. the emergence of HEIC over JPG). New operating systems that require technical changes are also relatively infrequent on PC, although the mobile space is faster moving. Software that isn't attended to for a year or so tends to cease functioning. The nature of CSAM changes over time, but the fundamentals remain the same. A machine learning based detector of CSAM is probably as capable of recognising a 20 year old image as a new AI generated one. Database driven solutions must be updated regularly to remain relevant however – to keep up with the constant flow of new material. As in our response to previous questions, allowing updates of data without re-accreditation (albeit with appropriate controls) will be necessary if the accreditation period is to be long. The language and means of concealing messages (emojis etc) changes much more rapidly, so language based tools may be harder to accredit for long periods unless continuous retraining is permitted (and required). In counter terror we understand from data providers we have spoken to that offender behaviour, especially language again, can evolve very rapidly. **Proposals:** • Technology notices can apply for up to 36 months so accreditation should where possible cover this whole period – e.g. 4 or 5 year validity for accreditation.

	• This implies there should be a flexible or light-weight approach to continuous improvement of software and updating of data. • In some areas the rate of change may be so high that to support accreditation beyond e.g. 1 year the accreditation must require improvements or updates to be regular.
Question 4: Do you have any views on how to turn these proposals into an operational accreditation scheme, including the practicalities of submitting technology for accreditation? Is there any additional evidence that you think we should consider? Please provide any information that may be relevant.	The consultation material appears to assume that technologies for accreditation will already exist in complete form, and that the accreditation process can therefore look at something (whether describing characteristics for audit or testing) complete and wholly representative. We would argue that this will not be the case in the many important scenarios. We further believe that the parliamentary process put in place the mechanism for technology notices precisely to allow the regulator to require the deployment of technology not yet in widespread use. We would like to see the accreditation process clearly positioned in terms of Technology Readiness Level. Clearly the lower TRLs are not accreditable – but at what point should it be possible to accredit technology? We believe that proof of the core capabilities of a technology can be achieved at TRL5 or TRL6. However, to get to TRL7 (demonstration in operational environment) requires at least one platform to have opened up and allowed testing of the technology. If we look at the current response of the E2EE messaging community to technologies to detect CSAM, none are willing. Technologies have been demonstrated at TRL6 (integrated into open-source E2EE messaging) but cannot reach TRL7 (operational environment). So if accreditation required TRL7 or higher as a baseline, technologies could only be accredited with the consent of at least one relevant platform! We do not believe it was parliament's intent (and there is nothing in the wording of the act to suggest otherwise) to place such a restriction on accreditation. Indeed, we believe that parliament requires an accreditation process specifically to decouple the regulators powers from the consent of platforms to test or deploy particular technologies. Our earlier comments regarding the difference between core characteristics and deployment characteristics

	(where they are today and were it is apparent they can readily scale to) apply here too. It would be helpful to provide specific examples. For example, imagine a provider who has developed a cutting edge new technology not available anywhere else. They have built and demonstrated the technology in an open source analog of real world deployment (TRL6) at small scale, and the technology has been appraised by security and online safety exports as offering an excellent solution. The technology is inherently scalable, but has not yet been scaled as there is no direct customer demand. Scaling the technology for production requires investment. The investment is not forthcoming in the absence of demand. In principle Ofcom would like to have this technology available to use in a Technology Notice. Assuming the cost of accreditation was within the reach of the company, would this technology be accreditable? Should it be? This consultation appears to suggest that it might be very difficult to accredit due to lack of evidence of operation at scale – is that the intention? While having scaled operation already in place is a great indicator scaling is possible, larger platforms will almost certainly want to do a deep technology integration rather than use external infrastructure (SaaS) anyway – so is it necessary? We believe the intent of the act in using the term Technology is to allow technology rather than scaled implementation to be accredited. **Proposal:** • Ofcom guidance should be explicit about the level of technology maturity necessary for accreditation. • We believe that technologies at TRL6 or above should be accreditable as long as there is a clear path to scale. • For technologies not at the required maturity for accreditation, Ofcom should consider support,

	guidance, or even a sandbox environment to help providers understand if they technology might be accreditable in future, and if so what they would need to do to achieve that. • Ofcom should consider arrangements to ensure that accreditation is accessible (cost, complexity) to small innovative companies.
Question 5: Do you have any comments on our draft Technology Notice Guidance?	We believe many of the points covered above are relevant to the guidance, especially in terms of robustly defining "technology" in the context of "accredited technology" – a key element of the guidance.

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