The Information Commissioner's response to Ofcom's consultation on protecting people from illegal harms online

About the Information Commissioner

The Information Commissioner has responsibility for promoting and enforcing data protection and information rights. This includes responsibilities under the UK General Data Protection Regulation (UK GDPR), the Data Protection Act 2018 (DPA 2018), the Freedom of Information Act 2000 (FOIA), the Network and Information Systems Regulations 2018 (NIS), the Environmental Information Regulations 2004 (EIR) and the Privacy and Electronic Communications Regulations 2003 (PECR).

The Information Commissioner is independent from government and upholds information rights in the public interest, promoting openness by public bodies and data privacy for individuals. The Commissioner provides guidance and support to individuals and organisations, aimed at helping organisations to comply, and takes appropriate action where the law is broken.

Our approach to this consultation response

In this consultation response we have limited our comments to areas which interact with our data protection remit. We have not commented on topics such as online safety harms or the efficacy of the measures proposed, except where it is relevant to data protection.

The Data Protection and Digital Information Bill was reintroduced in the Houses of Parliament on 8 March 2023. When the Bill becomes law, it will amend elements of data protection legislation relevant to this response. This response was written in line with the current applicable law at the time of writing.

Executive summary

- It is essential that users of online services have confidence that their privacy will be protected. We expect services to comply fully with data protection law when meeting their online safety obligations.
- We are pleased that Ofcom has referred to compliance with data protection law throughout the documents under consultation. We share Ofcom's commitment to promoting compliance across both of our regimes and welcome the opportunity to respond to this consultation.

Automated content moderation measures in the illegal content codes of practice

- We are not opposed to the recommended content moderation measures in principle, but we raise important points of alignment with data protection law.
- Content moderation involves the processing of people's personal data. There are therefore impacts on people's information rights that should be taken into account in the design of the final measures. We do not agree that the privacy impact of automated scanning is minimal.
- The privacy safeguards in the automated content moderation measures should be expanded to include reference to data protection requirements. Relevant areas to consider include transparency, purpose limitation, data minimisation, accuracy and, where relevant, compatibility with the requirements in UK GDPR Article 22.
- Services should be required to take into account the importance of minimising incorrect reports of child sexual exploitation and abuse material to the National Crime Agency when configuring technical accuracy settings, and deciding on the proportion of material that is appropriate for human review.

Guidance on content communicated "publicly" and "privately"

- We consider that the guidance does not currently provide sufficient regulatory certainty to enable services to make a confident assessment about whether content is communicated "publicly" or "privately".
- This lack of clarity may incentivise services to inappropriately assess content as being communicated publicly. This risks diluting an important privacy safeguard in the Online Safety Act (OSA). We

consider that the guidance should be more definitive for services and include worked examples.

- Where a service has made a genuine attempt to make the public/private assessment and cannot make a decision with certainty, the default should be that they assess content as being communicated privately. This would be in line with the wider duty to have particular regard to avoiding breaches of privacy law in the OSA.

Risk assessment guidance

- We do not challenge Ofcom's evidence base for concluding that factors such as encrypted messaging and anonymity/pseudonymity functionality are risks for illegal harm.
- However, we are concerned that the guidance could in practice deter services from deploying functionalities such as end-to-end encryption because they are deemed too risky under online safety law.
- We therefore suggest that the guidance should make it clear that the online safety regime does not restrict or prohibit the use of these functionalities and that the emphasis is on requiring safeguards to allow users to enjoy the benefits while managing risks appropriately.

Data minimisation

- Our response flags areas where there may be a lack of clarity about what personal data is needed to comply with Ofcom's guidance and measures. It highlights the importance of taking account of data minimisation when Ofcom finalises its guidance and measures to ensure that services are not incentivised to process more personal data than is needed.

General comments

The ICO welcomes the online safety regime and its mission to make the UK the safest place in the world to be online. It is essential that users of online services have confidence that their privacy will be protected, and we expect services to comply fully with data protection law when meeting their online safety obligations.

The OSA has been designed to work alongside data protection law, for which the ICO remains the statutory regulator. The regime supports effective cooperation by requiring Ofcom to consult with the ICO on codes of practice and formal guidance with an impact on privacy.

As the bodies responsible for regulating data protection and online safety in the UK, the ICO and Ofcom demonstrated their shared commitment to protecting people online by publishing a joint statement in November 2022. The statement set out our overall vision of ensuring coherence across online safety and data protection requirements and promoting compliance with both regimes. We are pleased that Ofcom has engaged with us during the development of the documents under consultation, and we welcome the opportunity to respond to the consultation. We stand ready to continue our engagement as Ofcom finalises the measures and guidance.

Compliance across the data protection and online safety regimes

We expect services to comply fully with data protection law when following the guidance and implementing the measures set out by Ofcom in this consultation. Service providers should therefore familiarise themselves with the data protection legislation and relevant ICO guidance to understand how to comply with the data protection regime. We expect services to take a data protection by design and default approach when implementing online safety systems and processes, as required by the UK GDPR.

The privacy duties set out at sections 22 and 33 of the OSA confirm the importance of data protection compliance by requiring services to have particular regard to the importance of protecting users from a breach of any statutory provision or rule of law concerning privacy when deciding on and implementing online safety measures. We are therefore pleased to see that Ofcom has referred to compliance with data protection law in the documents under consultation. We encourage Ofcom to continue to reinforce the importance of data protection compliance and to refer services to relevant ICO guidance resources where appropriate.

Compliance with the online safety duties will inevitably involve the processing of personal data. For example, services may be required to collect new types of personal data or to use personal data that they already hold for a different purpose. Some of the personal data could be special category or criminal offence data which involve additional protection under data protection law.

The processing of children's personal data may be necessary, and the UK GDPR also requires this to be given specific protection. The ICO's Children's code will be relevant where children's data is processed for online safety purposes by an information society service that is likely to

be accessed by children[1]. The code sets out specific safeguards for children's personal data to ensure that online services are appropriate for use by children. Services should refer to our Children's code guidance for further details.

We consider specific requirements of data protection law where they are relevant to Ofcom's proposals.

Response to consultation recommendations

The Illegal Harms Risk Register and Risk Profiles (Volumes 2 and 3 and Annex 5)

Risk and service functionalities

In Volume 2 Ofcom sets out its understanding of the causes and impact of online harm. It concludes that certain functionalities stand out as posing particular risks. These include:

- End-to-end encryption (E2EE), and
- Pseudonymity and anonymity.

We are pleased that the introduction to Volume 2 (page 3) notes that:

> "the functionalities are not inherently bad and that they have important benefits. End-to-end encryption plays an important role in safeguarding privacy benefits online. Pseudonymity and anonymity can allow people to express themselves and engage freely online…."
>
> and
>
> "The role of the new online safety regulations is not to restrict or prohibit the use of such functionalities, but rather to get services to put in place safeguards which allow users to enjoy the benefits they bring while managing risks appropriately"[2].

We agree that the functionalities have these benefits. They also confer wider benefits that keep users safe online. For example, we outlined our views on the safety benefits of E2EE in our November 2021 document *A*

[1] For the avoidance of doubt, the test as to whether a service is likely to be accessed by children under the Children's code is separate from the test set out in s37 OSA. The ICO has published guidance on when services are likely to be accessed by children for the purposes of the Children's code.

[2] See also paragraphs 6.11 and 6.12 of Volume 2

Framework for analysing End-to-End Encryption in an online safety context. Whilst we do not challenge the evidence base for concluding that these factors are risks for illegal harm, it is important that the regulatory approach to these functionalities seeks to reconcile addressing the immediate content-related harms with longer term privacy and safety impacts.

We are concerned that the benefits of these functionalities are not given enough emphasis in the risk assessment guidance and risk profiles (Annex 5). These are the documents that U2U services are most likely to consult on a regular basis. We consider that there is a risk that the risk assessment process may be interpreted by some services to mean that functionalities such as E2EE and anonymity/pseudonymity are so problematic from an online safety perspective that they should be minimised or avoided. If so, the risk assessment process could create a chilling effect on the deployment of functionalities that have important benefits, including keeping users safe online. This could be addressed if the risk assessment guidance itself transposed the parts of Volume 2 that make it clear that there is no intention to restrict or prohibit the functionalities in question and that the emphasis is on providing safeguards for user safety. This is an important message, and it should be more prominent in the risk assessment guidance itself.

We note that paragraph 6.14 of Volume 2 states that the fact that the identified risk factors can also be beneficial to users is a key part of the analysis underpinning the code measures. Whilst we welcome this approach, for the reasons set out above our preference would be for the risk register and risk profiles to clearly set out these considerations as matters to be taken into account as part of the risk assessment process itself. Not all regulated services will choose to conform to the measures set out in the codes of practice. Where services choose to take alternative measures to meet their online safety obligations, the risk assessment findings will be crucial in determining whether the measures they choose are necessary and proportionate to the risk of harm. However, it is not clear to us from the consultation documents how these services should take into account the wider benefits of a functionality such as anonymity or E2EE.

In summary, we therefore suggest that the guidance should make it clear that the online safety regime does not restrict or prohibit the use of these functionalities and that the emphasis is on requiring safeguards to allow users to enjoy the benefits while managing risks appropriately.

Risk assessment guidance (Volume 3 and Annex 5)

The risk assessment process

At step 2 of the risk assessment process services are required to assess the risk of harm based on relevant information and evidence. The OSA provides that a relevant factor in this assessment is how the design and operation of the service may reduce or increase the risks identified (s 9(5) OSA for U2U services and s 26(5) OSA for search services). Ofcom's draft guidance (Annex 5) reflects this requirement by setting out that services should consider whether there are any systems and processes already in place that reduce the risk of harm occurring on the service and demonstrate that these are effective in decreasing the risk of harm.

However, taken as a whole, we consider that the guidance lacks specificity about what kind of existing systems and processes a service could consider at step 2 and how services might demonstrate that the processes are effective in reducing the risk of harm.

More clarity may help providers to determine the accurate level of risk for their service. We suggest that Ofcom should consider the benefits of providing more detailed guidance to help services to understand the inputs and evidence that Ofcom expects to be relevant.

Core and enhanced inputs (Table 10)

The core and enhanced inputs set out in step 2 of the risk assessment process are likely to involve processing of personal data. For example, this could include data from user complaints and reports, and relevant user data including age.

In general, and subject to our specific comments below, we think that the guidance gives a clear explanation of the data that Ofcom expects services to consider as part of the risk assessment process. Not all of this will be personal data but where it is, the clarity in the guidance will help services to identify whether the personal data processing is relevant and necessary to the risk assessment process.

We are pleased that in Volume 3 Ofcom makes clear that any use of users' personal data will require services to comply with their obligations under UK data protection law (for example in Table 9.4 of Volume 3). We recommend that the guidance itself (Annex 5) also includes this reference so that services who only consult the guidance are clear about the need to comply with data protection law when compiling risk assessments.

A key data protection consideration when processing personal data for risk assessment is the data minimisation principle set out in Article 5(1)(c) of the UK GDPR. This requires the personal data that services process to be adequate, relevant and limited to what is necessary in relation to the purposes for which it is processed. This means that

services should identify the minimum amount of personal data they need to fulfil their purpose.

Certain types of personal data require particular care. The Children's code provides that services should collect only the minimum amount of children's personal data that they need to achieve their purpose. Where special category or criminal offence data is collected it is also particularly important to make sure that services collect and retain only the minimum amount of information that they require. Where possible, services should ensure that personal data is anonymised or pseudonymised to reduce the potential for it being linked to a particular person. This is mentioned on page 73 of Volume 3. We would recommend that it is also included in the guidance itself (Annex 5).

User complaints data

In relation to user complaints data being a core input for risk assessment, we note that this may conflict with the records retention analysis in the options assessment for the user complaints measure (Volume 4 paragraph 16.26). We comment on this in detail below in our response to the user complaints measure.

Relevant user data including user base demographics

The risk assessment guidance includes requirements for services to consider information relating to user characteristics. User base demographics are included as a general risk factor, and user data is a required core input for risk assessment. For some types of data, such as information relating to age, services may collect this data already, or may do so as part of other recommended measures within Ofcom's codes. However, for other user characteristics, such as gender or vulnerabilities relating to mental health, services may not routinely collect this data.

The requirement to use user data to inform risk assessment could lead to an assumption amongst some services that they need to collect this personal data if they do not already do so. This could have a significant impact on the privacy of users. We would therefore welcome further clarification within the risk assessment guidance that the need to consider user data does not require services to obtain personal data relating to user characteristics that they do not already hold.

Data from proactive technologies and age assurance

The outputs of behaviour identification technology, user profiling technology and age assurance or age verification processes are included within the scope of 'relevant user data', which is a core input for risk assessment (Table 10). This may include sensitive personal data, for example where technologies are used to profile and make inferences

about a user, particularly where the user is a child. It may not always be necessary or proportionate under data protection legislation for services to make use of this kind of data as a matter of routine. We are pleased that Ofcom has recognised this in its guidance and encouraged services to consult the ICO's guidance on UK GDPR requirements and the Children's Code.

The Illegal Content Codes of Practice for U2U services (Annex 7) and Search Services (Annex 8)

Governance and accountability (Annex 7 and 8, section A3)

Written statements of responsibilities (3C)

This measure requires providers of large or multi-risk U2U services and large general search or multi-risk search services to have written statements of responsibilities for senior members of staff who make decisions related to online safety risks.

Although the recommendation within the draft U2U and search service codes does not specify what types of decisions are related to online safety risks, Section 8.64 of Volume 3 states that these decisions include those related to:

> "the design of the parts of a product that users interact with (including how user behaviour / behavioural biases have been taken into account), how data related to user safety is collected and processed, and how humans and machines implement trust and safety policies." (page 17)

We support the statement of responsibility including these matters. It complements the accountability requirements under data protection law (UK GDPR Article 5(2)) and will help to ensure that services in scope of the measure comply with the OSA privacy duties by having regard to the importance of data protection law when making decisions related to the design and operation of online safety systems that process personal data.

Tracking evidence of new and increasing illegal harm (3E)

This measure requires services to track evidence of new and increasing illegal harm, including evidence derived from complaints processes, complaints moderation processes, referrals from law enforcement and information from trusted flaggers. This is likely to involve processing of personal data, and services will need to ensure they comply with data protection law when doing so.

As with the risk assessment process, where personal data is processed the data minimisation principle requires services to limit their use of personal data to what is relevant, adequate and necessary. Where

personal data can be anonymised, or pseudonymised, this will support data minimisation under data protection law.

U2U content moderation (Volume 4 sections 12 and 14, Annex 7 section A4)

Content moderation systems deployed by U2U services involve the processing of people's personal data.

In most cases, user-generated content is likely to be personal information in a service's moderation systems. This can be because:

- the information is about someone (for example, where the content contains information that is clearly about a particular user); or
- it is connected to other information, making someone identifiable (for example, the account profile of the user who uploaded it, which may include information like their name, online username and registration information).

Beyond the content itself, content moderation may also involve using personal information that is linked to the content or a user's account. For example, this can include a user's age, location, previous activity on the service, or a profile of their interests and interactions.

We have published guidance for U2U services setting out our data protection expectations for content moderation. Although we engaged with Ofcom as the guidance was being prepared, the finalised guidance was not available when Ofcom published the illegal harms consultation documents. We are committed to working with Ofcom to ensure that the online safety and data protection regimes are aligned and that organisations understand how data protection and online safety requirements interact in relation to content moderation.

We have the following observations about Ofcom's recommendations for U2U content moderation and automated content moderation (Annex 7 and the relevant sections of Volume 4).

Privacy impact assessments (Volume 4)

Volume 4 explains how Ofcom has analysed the impact on privacy of each of the measures it is proposing. In this part of our response, we consider the following recommended measures for U2U services:

- 4A-F – Requirements for a content moderation function designed to swiftly take down illegal content of which a service is aware
- 4G - Hash matching for child sexual abuse material (CSAM)
- 4H - Detection of CSAM URLs

- 4I - Keyword detection regarding articles used for fraud

There are privacy impact assessments for each of these measures in Volume 4. As drafted, these assessments do not take sufficient account of the impact of the proposed measures on information rights. We are keen to engage with Ofcom to ensure that the final measures take appropriate account of data protection considerations.

We note that the privacy impact assessments for (i) automated content moderation and (ii) the requirement to report UK-linked detected and unreported child sexual exploitation and abuse (CSEA) content to the National Crime Agency (NCA) under s66 of the OSA have not been comprehensively set out. Consequently, the assessment of the necessity and proportionality of the measures is incomplete.

This does not necessarily mean that we have concerns about the scope of the recommended measures. However, it will be important that they are supported by a fuller impact assessment which takes account of data protection impacts. A further consequence of the lack of inclusion of data protection in the impact assessment is that as drafted, the privacy safeguards that are set out in the body of the measures do not fully mitigate the potential impacts.

Privacy and automated processing

In relation to recommended measure 4A, the privacy impact assessment says (para 12.72 of Volume 4):

> "Insofar as services use automated processing in content moderation, we consider that any interference with user's right to privacy under Article 8 ECHR would be slight. Such processing would need to be undertaken in compliance with relevant data protection legislation." (page 33)

A similar point is made in relation to recommendation 4G at paragraph 14.78. In relation to the 4H measure, paragraph 14.201 notes that:

> "any processing of personal data for the purposes of the measure should be limited to the automated analysis of the relevant content to detect whether it includes a URL, and is unlikely to engage users' right to privacy under Article 8 ECHR." (page 127)

For recommendation 4I, paragraph 14.278 says that:

> "insofar as a service processes individuals' personal data for this purpose, any interference with users' right to privacy under Article 8 ECHR would not be significant. Such processing will also need to be

undertaken in compliance with relevant data protection legislation…" (page 142)

From a data protection perspective, we do not agree that the potential privacy impact of automated scanning is slight. Whilst it is true that automation may be a useful privacy safeguard, the moderation of content using automated means will still have data protection implications for service users whose content is being scanned. Automation itself carries risks to the rights and freedoms of individuals, which can be exacerbated when the processing is carried out at scale.

Our guidance on content moderation is clear that content moderation involves personal data processing at all stages of the moderation process, and hence data protection must be considered at all stages (including when automated processing is used, not just when a human looks at content). By way of example, the data protection harms that could flow from automated processing could include the risk of unwarranted surveillance, invisible processing and the loss of control of personal data. The ICO's data protection harms taxonomy sets out more information about data protection harms.

We are pleased to see the references in the privacy impact assessment to the importance of complying with data protection law, but we are unclear how this has been integrated into the measures. We therefore suggest that Ofcom provides a more robust assessment of the relevant data protection considerations. If the intention is to say that compliance with data protection law is a safeguard for privacy and will help to ensure compliance with Article 8 ECHR, this should be made clear and fully explained.

Including privacy safeguards

A more comprehensive privacy impact assessment will ensure that all appropriate privacy safeguards are included as part of the measures. Currently there are no specific privacy safeguards included in the body of the A4 content moderation measures.

The safeguards for measures 4G-I focus on ensuring the security and robustness of the underlying databases and on measures that relate to accuracy. Whilst these are important safeguards, and we are pleased that they have been included, from the perspective of data protection law they are incomplete. We stand ready to engage further with Ofcom to consider what additional data privacy safeguards are required. Relevant areas to consider include transparency; purpose limitation; data minimisation; accuracy; retention of personal data; data protection rights; and rights

related to automated decision making, including under UK GDPR Article 22.

In some respects, measures that Ofcom recommends in other parts of the codes of practice will also function as privacy safeguards. We consider this further below.

Reporting CSEA content to the NCA

Section 66 OSA (which is not yet in force) will require service providers to report detected and unreported UK-linked CSEA material to the NCA.

Ofcom's privacy impact assessment considers the impact where illegal content is incorrectly reported to reporting bodies or other organisations as a result of being detected through its recommended content or search moderation measures. It acknowledges that the measures could result in individuals being incorrectly reported to reporting bodies and states that this would represent a potentially significant intrusion into their privacy (Volume 4 14.80-14.86). We agree that an incorrect report which contains information about an identified or identifiable individual would be a significant intrusion into the individual's privacy.

As a privacy mitigation Ofcom points to the triage processes that reporting bodies will have in place to assess all reports received, ensuring that no action is taken relating to obvious false positives. We agree that it is vital that effective triage systems are in place, but from a data protection perspective, an individual's rights have been significantly impacted as soon as a report is made to the NCA, regardless of any further action taken by that body. We therefore stress that a triage process will not remove the need for services to take all reasonable steps to ensure the accuracy of the personal data that is reported. Some reports will involve the personal data of children. Under data protection law the processing of children's data requires specific protection.

Ofcom refers to the principles and safeguards in the content moderation measures as being safeguards that are designed to help secure that the technology operates accurately in connection with user reports to the NCA. Accuracy is also a relevant consideration in data protection law. The accuracy principle requires that services take all reasonable steps to ensure that the personal data they process is not incorrect or misleading as to any matter of fact. Where content moderation decisions could have significant adverse impacts on individuals, services must be able to demonstrate that they have put sufficient effort into ensuring accuracy.

We are concerned that the safeguards in measure 4G do not differentiate between the level of accuracy that is appropriate for reports to the NCA (which carries a particular risk of serious damage to the rights, freedoms

and interests of a person who is incorrectly reported) and other significant but potentially less harmful actions such as content takedown.

Our reading of measure 4G is that it could allow for the content moderation technology to be configured in such a way that recognises that false positives will be reported to the NCA. Whilst we acknowledge that it may not be possible to completely eliminate false positives being reported, we are concerned that a margin for error could be routinely "factored into" a service's systems and processes as a matter of course. This is unlikely to be compatible with a service taking all reasonable steps to ensure that the personal data it processes is not inaccurate.

We therefore consider that services should be explicitly required to take into account the importance of minimising false positives being reported to the NCA. This should apply both when they configure CSAM hashing technology to strike an appropriate balance between precision and recall and when they decide on what proportion of detected content it is appropriate for human moderators to review. One option would be to add it as a specific factor that services must take into account in paragraphs A4.27 and A4.31 of measure 4G.

Automated content moderation and alignment with UK GDPR Article 22

The automated content moderation measures have the potential to engage UK GDPR Article 22, particularly measures 4G and I.

Article 22 of the UK GDPR places restrictions about when services can carry out solely automated decision-making based on personal information where the decision has legal or similarly significant effects. It provides that services must only take solely automated decisions that have legal or similarly significant effects if they are:

- authorised by domestic law which also lays down suitable measures to safeguard the data subject's rights, freedoms and legitimate interests;
- necessary for a contract; or
- based on a person's explicit consent.

We consider these more fully in our content moderation guidance.

To achieve coherence across the regimes it is important that the recommended code measures are compatible with UK GDPR Article 22 requirements.

The following considerations will be relevant:

- Where a service relies on the Article 22 (2)(b) exception for decisions required or authorised by domestic law, such services should ensure that an individual's rights, freedoms and legitimate interests are safeguarded. In particular, services will need to adhere to s14 of the DPA 2018. This requires services to tell people that they have made the decision as soon as reasonably practicable. It also provides that the data subject may within 1 month of notification request the data controller to reconsider the decision or take a new one that is not solely automated. The OSA and associated codes of practice may also lay down additional measures to safeguard an individual's rights, freedoms and legitimate interests and services should ensure these safeguards are built into their processes.

- A service may also rely on the contract or consent exceptions set out in Article 22 (2)(a) and (c) of the UK GDPR. Where this takes place, Article 22 (3) requires the service to implement suitable safeguards, including at least the right to obtain human intervention on the part of the service, to express their point of view, and to contest the decision.

- The transparency requirements set out in Article 13 (2)(f) and Article 14 (2)(g) of the UK GDPR require services to proactively tell their users where they make solely automated decisions, give them meaningful information about the logic involved in any decisions the system makes and tell them about the significance and envisaged consequences that the decisions may have.

Performance targets (4C)

Measure 4A requires U2U services to have systems or processes designed to swiftly take down illegal content. Measure 4C requires that large or multi-risk services should set and record performance targets for the content moderation function. A4.12 provides that in setting its targets the provider should balance the desirability of taking illegal content down swiftly against the desirability of making accurate moderation decisions.

We suggest that paragraph A4.12 includes a reference to the requirements in data protection law for services to take all reasonable steps to ensure the personal information they use and generate through their content moderation processes is accurate. This is particularly important where CSEA material is detected because of the risk of incorrect reporting to the NCA as outlined above.

These comments also apply to recommendation 4C of the code of practice for search services (Annex 8).

Safeguards for privacy and data protection law

We refer above to the omission of appropriate privacy safeguards from the content moderation and automated content moderation measures. This does not mean that the codes of practice as a whole do not contain privacy safeguards. Some of the these are already provided for by other recommended measures, but they have not been specified as being protections for privacy. We recommend that the final version of the codes (Annexes 7 and 8) collate these measures comprehensively and identify them within the body of the content moderation measures as safeguards for privacy.

Doing this is important for two reasons. Firstly, s 49 of the OSA provides that a service that complies with a recommended measure in a code of practice is to be treated as complying with the privacy duty in sections 22(3) and 33(3) of the OSA where *"the recommended measure incorporates safeguards to protect the privacy of users"*. Setting out the full privacy safeguards will help to confirm that safeguards are incorporated and therefore provide more certainty for organisations.

Secondly, the existence of effective privacy safeguards is relevant to whether services can rely on the exception in UK GDPR Article 22 (2)(b). This permits solely automated decision making with a legal or similarly significant effect where the processing "*is authorised by a law which lays down suitable measures to safeguard the data subject's rights, freedoms and legitimate interests"*. Setting out the privacy safeguards will enable services to feel more confident about making the assessment of whether the exception is available to them.

Privacy safeguards that are provided for in the codes of practice but are not currently referenced as safeguards for the content moderation measures include the following (this list is not intended to be exhaustive):

- **Measure 4B - Internal content policies**: The privacy impact assessment notes that "*preparing a policy would tend to improve internal scrutiny, and improve the consistency and predictability of decisions, in a way which we think would also tend to protect users' privacy and personal information rights" (paragraph 12.95 page 37)*. We agree.

- **Measure 4C - Performance targets**: The privacy impact assessment notes that "*we consider that the setting and monitoring of accuracy targets as a part of this option, also acts as a safeguard for users' rights to freedom of expression" (paragraph 12.113 page 40).* We agree and also think that accuracy targets could safeguard privacy if they make systems more accurate and hence fairer.

- **Measure 4F - Provision of training and materials to moderators**: The privacy impact assessment notes "*We consider that the training of moderators would be a further safeguard for users' privacy, against the possibility that services may incorrectly report detected illegal content to reporting authorities" (paragraph 12.207 page 58).* We agree. This could also act as a wider privacy safeguard (not just in relation to minimising incorrect reports). For example, training and materials for moderators would support moderators in making fairer and more accurate moderation decisions.

- **Measure 6A - Terms of Service - Substance of the terms**: Services are required to tell people how they are protected from illegal content and provide information about any proactive technology used (including the kind of technology, when it is used, and how it works). They are also required to say how their complaints will be handled. From a data protection perspective this can help services to comply with the transparency principle. It will also help people to understand how their personal data is used and provide a route for challenging moderation decisions about their content.

- **Measure 5A - Enabling complaints**: Services are required to have complaints processes which enable UK users and affected persons to make each type of relevant complaint in a way which will secure that the provider will take appropriate action in relation to them. This may help service users to contest a content moderation decision which may support the exercise of data subject rights under data protection law.

Search moderation (Volume 4 sections 13 and 15, Annex 8 section A4)

Measure 4A in the search moderation code requires search services to have a moderation function designed to deindex or downrank illegal search content. It is our understanding that search services implementing this measure would not need to process additional personal data to do so. The exception would be where a service processes user personal data connected to a user complaint about illegal search content.

The privacy assessment of the recommendation (13.58-63) refers to the potential privacy impact of service users being reported to reporting bodies. It is not clear to us why this is a possible outcome of this recommendation. We would expect that determining whether search content is illegal would primarily involve moderation of web page content. Where personal data processing occurs as part of reporting illegal search

content, we would expect that it would be limited to third party personal data contained on web pages, rather than that of search service users. We would welcome further clarification about the circumstances in which this measure may require search services to process the personal data of users, particularly in relation to the reporting of users to reporting bodies. If service users may be reported to the NCA the concerns that we express above in relation to measure 4G in Annex 7 will also be relevant to this measure.

Reporting and complaints (Volume 4 section 16, Annexes 7 & 8 section A5)

Data retention

Paragraphs 16.26-27 of the consultation document state that Ofcom decided not to include a recommendation for services to keep complaints data to facilitate appeals as part of this measure. However, other consultation measures require or recommend the further use of complaints data, for example the risk assessment guidance, illegal content judgements guidance, and the recommendation that services track signals of new and increasing illegal harms (recommended measure 3E). We think that it is important that the overall package of measures make clear what information Ofcom considers necessary for services to retain and use to comply with online safety obligations. This will help services to feel confident about complying with their data protection obligations.

Accessibility of complaints systems (recommended measure 5B)

Ofcom recommend that all providers have easy to find, easy to access and easy to use complaints systems. This complements ICO guidance on transparency under data protection law, including the Transparency standard of the Children's Code, which states that privacy information must be concise, prominent and in clear language suited to the age of the child.

Timelines - sending indicative timelines (5C) and appropriate action for relevant complaints which are appeals (5E(i) and 5E(ii)).

Collectively these measures concern timelines for deciding complaints and appeals. We note that online safety complaints may, in some instances, also constitute complaints or requests under data protection legislation. For example, a complaint could include a request for a service to erase personal data it holds about an individual. Services will need to ensure that they are able to identify where an individual is also exercising their data protection rights, and that they comply with the timeframes set out by data protection law where this is the case. Data protection law sets time limits for responding to a request to exercise data protection rights.

The need to comply with data protection timeframes will only apply to the parts of the complaint that fall within data protection law.

We therefore recommend that the measures remind services of the need to comply with relevant response time limits that are laid down by other areas of law.

Terms of service and publicly available statements (Annex 7 and 8 section A6)

Ofcom's terms of service recommendations require services to provide information specifying how individuals are to be protected from illegal content, about the use of proactive technology for compliance with the illegal content safety duties, and about complaints processes.

These recommendations support and complement the transparency provisions of data protection law in informing individuals about how their personal information may be used by services fulfilling their online safety obligations.

Services will also need to comply with the transparency requirements of data protection law, including the right to be informed and, where applicable, standard 4 of the Children's code.

Default settings and child user support (Volume 4 section 18, Annex 7 section A7)

This measure sets out recommendations around how U2U services with a high risk of grooming, and large U2U services with a medium risk of grooming, should implement default settings and user support measures for child users. The ICO's Children's code takes a similar approach to safeguard the data protection rights of children online. The Children's code requires online services to:

- implement high privacy settings for children by default unless services can demonstrate a compelling reason for a different default setting, taking account of the best interests of the child;
- turn geolocation off by default unless services can demonstrate a compelling reason for geolocation to be switched on by default, taking account of the best interests of the child; and
- not to use nudge techniques to encourage children to provide unnecessary personal data or turn off privacy protections.

The code's transparency standard also requires services to provide child users with privacy information that is concise, prominent and in clear language suited to the age of the child, which aligns with the recommendation about child user support information at A7.11. The transparency standard also includes a requirement for services to provide 'bite-sized' explanations about personal data use at the point that use is

activated, a similar approach to the child user support recommendations at A7.6 to A7.9.

Paragraph 18.12 of Volume 4 sets out Ofcom's view that for services in scope of the measure, the recommendations apply to all child users under the age of 18. The Children's code also applies to services likely to be accessed by children under the age of 18. Ofcom's recommendations are therefore consistent with the code in ensuring that all under 18s should benefit from these protections.

We support Ofcom's approach to this measure, which complements the Children's code and forms part of a consistent and coherent regulatory framework covering the protection of children online.

We note that the Volume 4 assessment for these measures acknowledges that services that rely on self-declaration of age should continue to use this to indicate where a user is a child for the time being (for example paragraph 18.79 of Volume 4). We note that this may change when Ofcom sets out proposals for the use of age assurance technology on U2U services as part of its consultation on protecting children. We share the reservations that self-declaration would not provide an adequate level of certainty given the severity of the illegal harms that the measure is intended to mitigate. Our updated Commissioner's Opinion on age assurance provides more information about the ICO's data protection expectations for age assurance for the purposes of the Children's code.

Recommender system testing (Volume 4 section 19, Annex 7 section A8)

Ofcom's privacy assessment for this recommendation (19.44 b) refers to the need for services to obtain consent for the processing of personal data for on-platform testing. Consent is only one of the six lawful bases for processing personal data available under Article 6 of the UK GDPR, and it is unlikely to be an appropriate lawful basis for recommender system testing, primarily because consent can be refused, or withdrawn at any time, which may prevent effective testing. Furthermore, the privacy assessment suggests that services could obtain consent for processing personal data as part of consent to overall terms of service. Consent for data processing must be specific and freely given, which means that it cannot be bundled along with consent to terms of service. This is covered in more detail in ICO guidance on consent. We recommend that this paragraph is removed, and services are directed to ICO guidance for information about what they need to consider under data protection law.

This measure may also require services to process personal data as part of the safety metrics specified. Services adopting this measure should ensure that they comply with the purpose limitation and data

minimisation principles, and update their privacy information, as necessary.

Search functionalities (Volume 4 section 21, Annex 8 section A7)

Annex 8 measures 7B and 7C require all large general search services to provide crisis prevention information in response to search requests regarding suicide, and to provide warnings in response to search requests which clearly suggest the user is seeking to encounter CSAM. For both of these measures, the privacy assessment states that it does not consider there to be any impact on the right to privacy as there is no requirement that services retain information about searches that trigger these warnings (paragraphs 22.66 and 22.92).

Our view is that this assessment does not fully capture the relevant data protection considerations. As stated in paragraph 22.54 of Volume 4, services will need to detect the nature of search terms entered by a user. Depending on how services implement these warnings, this could result in services processing personal data to deliver warnings to individual identifiable users, and as a result processing of user personal data could occur when search terms are analysed. This could be the case for both the delivery of warnings and crisis prevention information to users. Analysing searches to provide crisis prevention information may also require services to process special category data relating to the health of users.

Paragraph 22.66 (CSAM content warnings) recognises that services may choose to retain this information, advising services that do so that they will need to comply with applicable privacy and data protection laws.

We recommend that Ofcom should review its privacy assessment of these measures to take the impact on data protection rights into account. We recommend that the measures refer to the need for services to identify if they are processing personal data and if so to familiarise themselves with the requirements of data protection law.

Record keeping and review guidance (Annex 6)

Where service providers take alternative measures to those set out in Ofcom's codes of practice to comply with their online safety duties, A6.33 of this guidance clarifies that they must keep a written record of how they have had regard to protecting the privacy of users. The ICO supports this recommendation, which will help to ensure that services taking alternative measures comply with the privacy duties within the OSA.

Services will also need to be able to demonstrate their compliance with the UK GDPR under the accountability principle. Measures that services may need to take to meet this requirement include maintaining documents of their processing activities, carrying out data protection

impact assessments, and putting written contracts in place with third parties processing personal data on their behalf.[3] Whilst taking these measures is part of the obligations that services have under data protection law, services may also find that these measures are useful in meeting Ofcom's record-keeping recommendations and the privacy duty within the OSA.

Guidance on content communicated "publicly" and "privately" (Annex 9)

The OSA Schedule 4(13) constraint on Ofcom's powers to recommend use of proactive technology where content is communicated privately is an important safeguard for privacy.

The Annex 9 guidance requires services to make their own assessment about whether content is communicated publicly or privately by means of the service. This is not a requirement that the OSA places on services. The consultation documents do not explain why this is Ofcom's preferred approach. We consider that this is a significant omission, and we encourage Ofcom to provide an explanation of its reasons.

The guidance suggests that organisations should apply the statutory factors in s232 OSA to their service to determine whether content is being communicated publicly or privately. This is a complex exercise. It is therefore important that the guidance provides sufficient direction and certainty to empower services to make the assessment with confidence. If this is lacking, there is a risk that some services will default to assessing content as being communicated publicly. This would undermine the effectiveness of the privacy safeguard in practice.

We have been in discussion with Ofcom about including hypothetical examples which would help services to make the assessment. There may be particular benefit in providing examples which are obviously at either end of the public/private spectrum. This will enable services to recognise clear-cut situations in the context of their own processes. We recognise the constructive approach that Ofcom has taken to considering this.

Specific areas of the guidance that we have identified as requiring clarification are:

- The definition of "a substantial section of the public" in paragraph A9.23 . The guidance provides that "Where [content] is accessible to a substantial section of the public, it should be considered as communicated publicly" and footnote 7 says "This is the case irrespective of the second and third statutory factors" (page 6). The guidance therefore seems to suggest that statutory factor 1 alone

[3] Accountability and governance | ICO

can be determinative where this threshold is met and clarity about this threshold is therefore key.

- Clarifying paragraph A9.23 which, on our reading, suggests that if there is no access restriction in place on a service, the content should be considered accessible to all UK internet users. We do not believe that this is intended to include services that are configured to have low maximum capacity thresholds (but have no formal access restrictions). If so, we recommend that this is made clear.

In addition to these specific comments the guidance would benefit from providing more instruction about how services should go about making a holistic assessment taking all three statutory factors into account, particularly where the factors do not all point to the same conclusion. Some assessments will be "clear-cut". But others will not be, and it is important that services know what approach to take in such circumstances.

Because of the foundational importance of the public/private distinction for safeguarding privacy, we advocate that where a service can demonstrate that they have engaged fully with making the assessment but are unable to come to a firm conclusion, the presumption should be that the service considers content to be communicated privately. This would be in line with the spirit of the OSA privacy duty in s22 which requires services to have particular regard to the importance of protecting users from a breach of privacy law when deciding on, and implementing, safety measures and policies.

Record keeping and review and Annex 9

It is important that services document and keep records of how they have conducted the public/private assessment. To the extent that this is not already provided for, we suggest that this should be a specific record keeping requirement which should be included in the guidance on record keeping and review (Annex 6).

Illegal content judgement guidance (Annex 10)

Making illegal content judgements (ICJs) is an area of potential tension between online safety and data protection law. Section 192(2) OSA states that such judgements are "to be made on the basis of all relevant information that is reasonably available to the provider". The data protection principle of data minimisation requires organisations to limit personal data processing to what is adequate, relevant and necessary to achieve their purpose. Ofcom acknowledge these tensions in the section on reasonably available information (A1.64-67) and clarify that services

should only process as much personal data as is necessary to make ICJs. The ICO agrees with Ofcom's approach.

Making ICJs will also require services to use the personal data they hold for the purpose of deciding whether there are reasonable grounds to infer that content is illegal. This could have a significant privacy impact. We welcome that Ofcom has recognised this and has highlighted the need to comply with data protection law. ICO guidance will support data protection compliance by services making ICJs. In particular, our guidance products on content moderation, the data protection principles and lawful basis are likely to be relevant.

We also support the pragmatic approach that Ofcom has taken to setting out how services can make ICJs. Each chapter considers the offences in order of the likely ease of making reasonable inferences of whether content amounts to an offence, and the guidance states that once content has been identified as illegal content under one offence, there is no need to consider other offences (A1.75). Section A3 (Threats, abuse and harassment offences) states that services should look first at the offences of this type with the simplest criteria for illegality. Even where content may be a very serious offence, it may be possible to apply a less serious but simpler offence (A3.3-A3.4). This approach should support data minimisation as services are likely to use less personal data when making ICJs about "simpler" offences.

There are however some areas of the guidance where we have found it to be less clear about the approach that services should take to balancing the need to make ICJs with the need to comply with data minimisation. We discuss these further below.

Relationship between consultation document and guidance

In paragraphs 26.151 and 26.166 of Volume 5, Ofcom states that using information relating to account activity to infer the age of the subject of a CSAM image or of a victim of grooming is likely to constitute a "very significant interference with all users' right to privacy", and that this information should not be considered to be reasonably available. However, this restriction is not replicated in the draft guidance (Annex 10), which simply says that services should have regard to data protection law when using such information. Given that services will refer primarily to the guidance, Ofcom should ensure that the messaging is consistent and compatible with the data minimisation principle.

Use of information beyond that covered in guidance

Paragraph 26.27 in Volume 5 discusses the use of information beyond the five types of reasonably available information that is specified in 26.26. It

states that where services have access to this information, they should have:

> "reasonable regard to any other relevant information to which they have access, above and beyond what is set out in the Content Judgements Guidance but only so long as this information is processed lawfully, including in particular in line with data protection laws." (paragraph 26.27 page 9)

However, the Annex 10 guidance is phrased more equivocally, stating at A1.67 that:

> "Where such information is relevant to content judgements as set out in this guidance, services should consider this information as appropriate. Services will need to ensure that they comply with data protection law when processing this information." (page 17)

The data minimisation principle requires that personal data being processed be relevant, adequate, and limited to what is necessary. Where an ICJ can be made accurately without the need to process the additional personal data held by a service it would not be necessary for a service to process this information under data protection law. Our preference would be for the text in Volume 5 to appear in the guidance itself (Annex 10).

The text could also clarify that services may not always need to consult all available information in every instance, if it is possible to make an accurate judgement using less information.

Criminal offence data

A1.68-A1.70 of Annex 10 concerns data provided to services by law enforcement. A1.70 states that services will need to ensure that they comply with data protection law.

We welcome that reference. Specifically, services will need to comply with Article 10 of the UK GDPR when processing personal data relating to criminal offences and convictions. Where law enforcement provides a service with information relevant to an ICJ, it is likely that any personal data included will be criminal offence data. Because of this, we suggest that A1.70 includes a more specific reference to the need to comply with Article 10 UK GDPR.

Age assurance data

Paragraphs A4.22 and A5.18 of Annex 10 state that services should have regard to the privacy implications of reviewing a user's account activity and information to determine their age (in relation to CSAM and grooming offences respectively), and that services should have regard to their data protection obligations when doing so. The reference to account

information may be intended to include the use of data derived from age assurance technologies but this is not clear to us. We recommend that if the intention is for such information to be referred to when making an ICJ that this is made clear and that services are directed to the ICO's Commissioner's Opinion on age assurance for more information about the data protection requirements.