

ANNEX D: GAPS BETWEEN OFCOM'S ANALYSIS OF CAUSES AND IMPACTS OF ONLINE HARM PROPOSED CODE OF PRACTICE MITIGATIONS: ILLEGAL HARMS AND PROTECTION OF CHILDREN COMBINED

In our response to Ofcom's illegal harms consultation, [we provided a table](#) analysing how far harm arising from the functionalities that it identified in its risk register (volume 2) were mitigated by specific measures in the codes (annex 7). The approach Ofcom took in its protection of children's consultation was broadly similar to that proposed in the illegal harms consultation - though caveated by many references throughout the documents that the responses to the latter had not yet been taken into account and further updates will follow. We therefore carried out the same analysis on the children's consultation and [updated our table](#) to combine the results from both for ease of reference. We resubmitted this to Ofcom as part of their consultation process.

This version is updated to take account of changes to the two final codes following the consultations, which are minimal. Where changes have been made, these are highlighted in the far right-hand column of the table. In summary, they amount to:

- A caveat added to both the illegal harms and children's codes in response to business feedback that content takedown is only required "if technically feasible to do so". Or, as Ofcom say in their [Summary of our decisions](#) document: "We have made changes to the measures for providers of services that cannot take action on content identified as harmful." We wrote about this change - and the anger felt amongst civil society groups, particularly those working to prevent CSAM - in [our illegal harms statement](#) in January.
- Further consultation launched on extending the ability to mute or block accounts to smaller services in the illegal harms codes, to bring them into line with scope of measure in children's codes.

DN: Ofcom have made some additional changes and added new measures to other non-risk-related code sections (eg complaints and reporting, age assurance) but these are not our focus here. Where there have been changes (namely in the first few rows of the table) we have updated Ofcom's reference numbers for both the measures in the consultation versions and in the final versions. *HEALTH WARNING: this table has been updated at haste in the short period of time since the codes were published on 24 April so there may be omissions and errors.*

Commentary

As we set out in the earlier versions of this document, we would expect that Ofcom's decisions on which measures to include in their codes of practice would reflect the level of risk threat that the functionalities identified in the risk register pose. We would also reiterate here our acknowledgement that the work that has gone into the risk registers themselves - [volume 3](#) in the children's consultation, [volume 2](#) in the illegal harms - is thorough and analytical. But in neither of the children's codes of practice do these risks flow through to the mitigation measures for user-to-user services ([code as laid in Parliament](#)) and search ([code as laid in Parliament](#)), which focus primarily on content takedown or, in the children's code, measures to deal, ex-post, with primary priority content (PPC), priority content (PC) or non-designated content (NDC). The exception to this is the measures relating to recommender systems, which are welcome and go some way to addressing the scale and impact of harm caused by the recommendation and promotion of PPC, PC or NDC content to children.

Despite the representations of ourselves and others during both consultations, the rules-based nature of the Codes (which is NOT required by the definition of “measures” in the Act¹) - specifying narrow recommended measures rather than describing desired outcomes - and the fact that the Codes are designed as a “safe harbour” (eg if companies follow the measures they will be judged to have complied with their duties under the Act²), means that there is no incentive for companies to implement mitigating measures to protect children beyond those described in the codes, even if their risk assessment has flagged that their service poses particular risks from other ex ante functionalities (such as design choices). This is particularly notable in relation to the omission of any measures relating to livestreaming³ - which is mentioned in relation to seven out of the nine types of content in the children’s risk profiles; and in relation to two new functions that are covered in the children’s consultation: stranger pairing and ephemeral messaging. Furthermore, smaller companies are in many instances exempt from implementing particular mitigating measures due to Ofcom’s proportionality analysis; following both consultations, and further requirements were removed from smaller services as a result of industry feedback. (See for example [Volume 4](#), page 55 of the children’s codes documentation: “We have concluded that the safety benefits for some of the reporting and complaints measures would be small, if any, when applied to small, low-risk services. On the other hand, users - including children - would lose out if these services withdrew from the UK because of the regulatory burden.”)

The following tables provide detailed analysis on the individual functionalities, the number of offences (for the illegal harms codes) or types of content (for the children’s codes) where Ofcom identifies that particular functionality is a contributory factor, and the appearance (or not) of mitigating measures relating to this functionality in the codes of practice for user to user and search services for both duties. A summary “at a glance” table is provided for U2U (pages 3-9) and search (p9-10). We have divided the measures in both sets of codes into “ex ante” and “ex post”, the latter largely applying to measures relating to content moderation and takedown when either illegal content or PPC, PC or NDC has been identified on a service. While we have used the term “ex ante” in relation (generally speaking) to the non-takedown measures, the measures identified are focused on the presence of specific content (either illegal or designated) on the service (or the search functionality enabling users to find it) so are not what we would term “safety by design” measures. These we would classify as biting at a systemic level separate to the nature of the particular types of content (e.g. business model, default settings or measures that are not directed to a particular type of content for eg rebalancing weighting in recommender tools).

¹Section 236(1) Online Safety Act

² “Services that choose to implement the measures we recommend in Ofcom’s Children’s Safety Codes will be treated as complying with the relevant children’s safety as well as their reporting and complaints duties. This means that Ofcom will not take enforcement action against them for breach of that duty if those measures have been implemented. This is sometimes described as a “safe harbour.” However, the Act does not require that service providers adopt the measures set out in the Children’s Safety Codes, and service providers may choose to comply with their duties in an alternative way that is proportionate to their circumstances. (Vol 5, para 13.4)

³ Livestreaming will now be included in the subsequent Ofcom consultation on additional measures, [due in June](#).

COMPARISON OF RISK REGISTER FUNCTIONALITIES WITH USER-CODE OF PRACTICE MITIGATIONS: SUMMARY TABLE

Functionality	Illegal harms offences	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations		Changes after consultation
	15 in total	9 in total	Illegal harms	Children	Illegal harms	Children	Both codes
Content: posting, commenting, hyperlinks, including images and video	15	9	Limited to user controls measures (eg muting, blocking): 9A, 9B Final codes: ICU J1 & J2	Limited to user controls measures (eg muting, blocking, disabling comments): US2, US3 Final codes: PCU J1 & J2	Content moderation & takedown: 4A-F (consultation version) Final codes: ICU C2.1-2.5	Content moderation & takedown: CM1-CM7 Final codes: PCU C1-C8 Limited: Signposting children to support when they a) report content (all services); b) post or repost content (large, risky services); US3, US4	Yes: on content moderation - in response to business feedback - to weaken the content takedown measures in both codes with caveat " <i>unless it is currently not technically feasible for them to achieve this outcome</i> " (ICU C2.2 & PCU C2) In the children's codes, the first measure has been separated out into two: the requirement to have a system to review and assess suspect harmful content; and a moderation

Functionality	Illegal harms offences	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations		Changes after consultation
	15 in total	9 in total	Illegal harms	Children	Illegal harms	Children	Both codes
							<p>function that allows for "swift action". (PCU C1 and C2)</p> <p>On blocking and muting measures, Ofcom is now consulting further on expanding these controls to smaller services under the illegal harms duties to bring them into line with the children's duties.</p>
Reposting or forwarding content	5	4	None	None	Limited: reference to "limiting time"	None	No
Livestream & live audio	9	7	None	None	None	None	Livestream measures are to be included in Ofcom's next consultation, due in June.
Use of hashtags	5	8	None	None	None	None	No

Functionality	Illegal harms offences	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations		Changes after consultation
	15 in total	9 in total	Illegal harms	Children	Illegal harms	Children	Both codes
Editing visual content	9	4	None	None	None	None	No
Screen capturing or recording	1	2	None	None	None	None	No
User tagging	5	3	None	None	None	None	No
User profiles	10	4	Limited to user controls: 9A, 9B	Limited to user controls: US2, US3	None	None	No
User connections	8	8	Limited to default settings, user controls: 9A, 9B	Limited to default settings, user controls: US2, US3	None	None	No
Stranger pairing	N/A	1	N/A	N/A	None	None	No
User search	2	1	None	None	None	None	No
User groups	9	4	None	None	None		No
User base profile	3	7	None	Significant measures via age assurance (AA1-6) though no differentiation for age ranges within this	Limited: references in 4E, 5B	None	No

Functionality	Illegal harms offences	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations		Changes after consultation
	15 in total	9 in total	Illegal harms	Children	Illegal harms	Children	Both codes
Recommender systems	11	8	None	Significant new measure (RS1-3) covering PPC and PC, and feedback	Limited: A6 ("limited time"), A9 safety metrics	Not applicable: ex-ante design choice	No
Group messaging	6	6	None	US1: option to accept or decline an invite to a group chat	None	None	No
Encrypted messaging	10	3	None	None	None		No
Direct messaging	15	6	Limited to user controls: 9A, 9B Plus 7A: Default settings for child users where services are high risk for CSAM	Limited to user controls: US2, US3	None		No
Ephemeral messaging	N/A	2	N/A	None	N/A	None	No
Anonymous user profiles	15	5	9C has recommendation re user labelling schemes, but this is only	None	None	None	No

Functionality	Illegal harms offences	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations		Changes after consultation
	15 in total	9 in total	Illegal harms	Children	Illegal harms	Children	Both codes
			limited to services at risk of fraud or the foreign interference offence				
Fake user profiles	13	4	As above 9C	None	None	None	No
Business model - inc small, fast-growing services; ad revenue	5	3	None	None	None	None	No
Payment facility	2	0	None		None		No
User location	4	1	Included in A7 default settings measures, but only limited to services at high risk of grooming		None		No
UGC search facility	3	3	None		None	Limited: Signpost children to support services when they search for harmful content (high or medium risk): US5	No
Posting goods or services for sale	7	0	None		None		No

Functionality	Illegal harms offences	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations		Changes after consultation
	15 in total	9 in total	Illegal harms	Children	Illegal harms	Children	Both codes
Building lists or directories	2	0	None		None		No

COMPARISON OF FUNCTIONALITIES WITH SEARCH CODE OF PRACTICE MITIGATIONS: SUMMARY TABLE

NB the analysis of the search functionalities that cause harm is less detailed and presented in a different way to the evidence in the user-to-user sections of both consultations.

Functionality	Illegal harms	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations	
			Illegal harms	Children	Illegal harms	Children
Typing in searches for illegal / specified content	8	Not defined	Limited: provision of warnings for CSAM searches; and provision of suicide prevention information in relation to suicide/self-harm searches	None	Search moderation & takedown: 4A-F - these measures largely replicate the user-to-user content moderation measures but with 4A applying to deindexing or deranking illegal content. An additional deindexing measure applies to CSAM URLs (4G)	Equivalent as for illegal harms: Measures SM1-7
Ranking	-	N/A	None	None	As above	As above.
Reverse image search	1	Not defined	None	N/A	None	N/A
Search prediction or personalisation	3	Not defined	None	N/A	Limited: requires action when there is a user report that predictive	Limited: offer users means to easily report predictive search

Functionality	Illegal harms	Children's PPC, PC or NDC	Code of practice: ex ante mitigations		Code of practice: ex post mitigations	
					search suggestions are directing users to priority illegal content	suggestions relating to PPC and PC (SD1); provide crisis information in response to searches relating to suicide, self-harm and eating disorders (SD2)
Revenue models	2	Not defined	None	None	None	None
Commercial profile/size	-	Not defined	None	None	None	None
Gen AI/chat bots	-	Not defined	None	None	None	None