ONLINE SAFETY BILL

IJM Proposed Amendments - May 2022

International Justice Mission fully supports the ambition of the Online Safety Bill to hold regulated services accountable for online sexual exploitation of children occurring on their platforms.

The Bill has great potential to make a global impact and set a world-leading example. The implications of the duties of care introduced by the Bill will be felt around the world in the prevention, disruption and detection of online sexual exploitation of children.

We are encouraged by the prioritisation of tackling the dissemination of child sexual exploitation and abuse (CSEA) content. However, there is room for the Bill to go even further in strengthening child protection online in particular in relation to:

1. The use of online platforms to generate new CSEA content
2. Enforcement to protect children around the world

KEY PROVISIONS

The Bill imposes duties of care on user-to-user services and search services.

CLAUSE 2:

- *"user-to-user service"* means an internet service by means of which content that is generated directly on the service by a user of the service, or uploaded to or shared on the service by a user of the service, may be encountered by another user, or other users, of the service.
- *"search service"* means an internet service that is, or includes, a search engine.

Part 3 of the Bill imposes duties of care on those services.

CLAUSE 8: *A duty to carry out a suitable and sufficient illegal content risk assessment*

This clause also imposes a duty to carry out a risk assessment prior to many significant changes to a service's design or operation.

> *Before making any significant change to any aspect of a service's design or operation, a duty to carry out a further suitable and sufficient illegal content risk assessment relating to the impacts of that proposed change.*

INTERNATIONAL JUSTICE MISSION UK, PO Box 78942, London, SE11 9EB • 0203 405 9080
IJMUK.org • IJMUK:

The illegal content risk assessment must assess the following matters:

A. The the user base;

B. the level of risk of individuals who are users of the service encountering the following by means of the service—
- each kind of priority illegal content (with each kind separately assessed), and
- other illegal content,

taking into account (in particular) algorithms used by the service, and how easily, quickly and widely content may be disseminated by means of the service;

C. the level of risk of harm to individuals presented by illegal content of different kinds;

D. the level of risk of functionalities of the service facilitating the presence or dissemination of illegal content, identifying and assessing those functionalities that present higher levels of risk;

E. the different ways in which the service is used, and the impact of such use on the level of risk of harm that might be suffered by individuals;

F. the nature, and severity, of the harm that might be suffered by individuals from the matters identified in accordance with paragraphs (b) to (e);

G. how the design and operation of the service (including the business model, governance, use of proactive technology, measures to promote users' media literacy and safe use of the service, and other systems and processes) may reduce or increase the risks identified.

IJM UK ASSESSMENT:

This is a positive step forward, particularly in relation to point (g) which effectively amounts to requiring a 'safety by design' approach. Not only does a risk assessment have to be carried out and kept up to date, whenever changes are made, an assessment must be carried out to ascertain whether that will increase the risk of illegal content (i.e. CSEA) and what steps can be taken to mitigate that risk.

RECOMMENDATION: There is scope for it to be strengthened further in relation to newly produced CSEA content and livestreaming.
Amend (d) to read:

the level of risk of functionalities of the service facilitating the production, presence or dissemination of illegal content, identifying and assessing those functionalities that present higher levels of risk;

CLAUSE 8: *A duty to take or use proportionate measures to effectively mitigate and manage the risks of harm to individuals, as identified in the most recent illegal content risk assessment of the service.*

This is a duty to operate a service using proportionate systems and processes designed to:

A. prevent individuals from encountering priority illegal content by means of the service;

B. minimise the length of time for which any priority illegal content is present;

C. where the provider is alerted by a person to the presence of any illegal content, or becomes aware of it in any other way, swiftly take down such content.

This will require the provider of a service to take or use measures in the following areas, if it is proportionate to do so:

A. regulatory compliance and risk management arrangements,

B. design of functionalities, algorithms and other features,

C. policies on terms of use,

D. policies on user access to the service or to particular content present on the service, including blocking users from accessing the service or particular content,

E. content moderation, including taking down content,

F. functionalities allowing users to control the content they encounter,

G. user support measures, and

H. staff policies and practices.

In determining what is proportionate for the purposes of this section, the following factors, in particular, are relevant:

A. all the findings of the most recent illegal content risk assessment (including as to levels of risk and as to nature, and severity, of potential harm to individuals), and

B. the size and capacity of the provider of a service.

IJM UK ASSESSMENT:

We welcome this duty to ensure that service providers act upon and mitigate the risks identified in the required risk assessment, and to introduce protective systems and processes.

However this duty could be significantly strengthened by not only preventing individuals *encountering* illegal content, but preventing the creation of that content. The Bill has a chance to not simply address what illegal content is seen online, but to address how online platforms are used to perpetrate abuse. This is crucial to addressing the livestreaming of child sexual exploitation.

RECOMMENDATION:

Amend Clause 9 to insert duty to

Prevent the production of illegal content by means of the service illegal content, identifying and assessing

CLAUSE 10: *A duty to conduct a children's risk assessment of a service*

This means an assessment of the following matters, taking into account the risk profile that relates to services of that kind—

A. the user base, including the number of users who are children in different age groups;

B. the level of risk of children who are users of the service encountering the following by means of the service—

- **a.** each kind of primary priority content that is harmful to children (with each kind separately assessed),
- **b.** each kind of priority content that is harmful to children (with each kind separately assessed), and
- **c.** non-designated content that is harmful to children,

giving separate consideration to children in different age groups, and taking into account (in particular) algorithms used by the service and how easily, quickly and widely content may be disseminated by means of the service;

C. the level of risk of harm to children presented by different kinds of content that is harmful to children, giving separate consideration to children in different age groups;

D. the level of risk of harm to children presented by content that is harmful to children which particularly affects individuals with a certain characteristic or members of a certain group;

E. the level of risk of functionalities of the service facilitating the presence or dissemination of content that is harmful to children, identifying and assessing those functionalities that present higher levels of risk, including functionalities—

- **a.** enabling adults to search for other users of the service (including children), and
- **b.** enabling adults to contact other users (including children) by means of the service; the different ways in which the service is used, and the impact of such use on the level of risk of harm that might be suffered by children;

F. the nature, and severity, of the harm that might be suffered by children from the matters identified in accordance with paragraphs (b) to (f), giving separate consideration to children in different age groups;

G. how the design and operation of the service (including the business model, governance, use of proactive technology, measures to promote users' media literacy and safe use of the service, and other systems and processes) may reduce or increase the risks identified.

IJM UK ASSESSMENT:

As above, this provision is positive, however it can be strengthened further to include newly produced CSEA by amending (e) as follows

> ***the level of risk of functionalities of the service facilitating the production of illegal content and presence or dissemination of content that is harmful to children...***

ADDITIONAL AMENDMENT TO CLAUSE 10: Insert specific risk assessment in relation to CSEA content -

a. the level of illegal images blocked at the upload stage and number and rates of livestreams of CSEA in public and private channels terminated; and

b. the number and rates of images and videos detected and removed by different tools, strategies and/or interventions.

The addition of 'public and private livestreamed CSEA content' not only acknowledges first-generation CSEA content, but also recognises that livestreamed CSEA content happens on both public and private channels, which require different methods of detection.

This also details the practical information needed to adequately assess whether action being taken by a regulated service is adequate in countering the production and dissemination of CSEA content, in particular first-generation CSEA content. The separation of rates of livestreams of CSEA in public and private channels terminated is important to note as rates may vary widely dependent on how CSEA content is being generated. By specifying 'tools, strategies and interventions', this addition ensures that the systems in place to detect and report CSEA content are adequate.

CLAUSE 11: Safety Duties protecting children

A duty, in relation to a service, to take or use proportionate measures to effectively:

A. mitigate and manage the risks of harm to children in different age groups, as identified in the most recent children's risk assessment of the service, and

B. mitigate the impact of harm to children in different age groups presented by content that is harmful to children present on the service.

A duty to operate a service using proportionate systems and processes designed to:

A. prevent children of any age from encountering, by means of the service, primary priority content that is harmful to children (for example, by using age verification, or another means of age assurance);

B. protect children in age groups judged to be at risk of harm from other content that is harmful to children (or from a particular kind of such content) from encountering it by means of the service (for example, by using age assurance).

IJM UK ASSESSMENT:

As above, we welcome this duty to ensure that service providers act upon and mitigate the risks identified in the required risk assessment, and to introduce protective systems and processes.

However, there is a sense that the clause in concerned about the harm caused to children by what they will see, rather than the harm caused by using the platform to abuse children, as is the case in the livestreaming of child abuse.

RECOMMENDATION:

Amend Clause 11 to insert a duty to operate a service designed

To prevent the sexual or physical abuse of a child by means of that service

CLAUSE 17: A duty to operate a service using systems and processes that allow users and affected persons to easily report content, including illegal content

IJM UK ASSESSMENT: IJM UK welcomes this provision.

CLAUSE 21 ONWARDS: Introduces duties of care similar to those described above for user-to-user services. The equivalent recommendations apply to these provisions.

CLAUSE 23: Illegal content risk assessment duties for search services

RECOMMENDATION:

Amend Clause 23(5)(b) to read –

the level of risk of functionalities of the service facilitating the production of illegal content, and individuals encountering search content that is illegal content, identifying and assessing those functionalities that present higher levels of risk

CLAUSE 24: Safety duties about illegal content

RECOMMENDATION:

Amend Clause 24(3) to read –

A duty to operate a service using proportionate systems and processes designed to minimise the risk of individuals producing or encountering search content of the following kinds...

CLAUSE 25: Children's risk assessment duties

RECOMMENDATION:

Amend Clause 25(5)(c) to read –

the level of risk of functionalities of the service facilitating the production of illegal content and children encountering search content that is harmful to children, identifying and assessing those functionalities that present higher levels of risk;

CLAUSE 26: Safety duties protecting children

RECOMMENDATION:

Amend Clause 26(3) to include –

A duty to operate a service using proportionate systems and processes designed to

(a) minimise the risk of children of any age suffering harm which amounts to a relevant offence (defined in section 52)...

CLAUSE 37: Requires OFCOM to prepare and issue a code of practice for regulated services describing measures recommended to ensure compliance with duties relating to CSEA content.

CLAUSE 45: A regulated service is to be treated as complying with a relevant duty if the provider takes or uses the measures described in a code of practice which are recommended for the purpose of compliance with the duty in question.

CLAUSE 52: Defines what is meant by 'illegal content' as content that amounts to a relevant offence.

Content consisting of certain words, images, speech or sounds amounts to a relevant offence if:

- **A.** the use of the words, images, speech or sounds amounts to a relevant offence,
- **B.** (in the case of a user-to-user service) the use of the words, images, speech or sounds, when taken together with other regulated user-generated content present on the service, amounts to a relevant offence,
- **C.** the possession, viewing or accessing of the content constitutes a relevant offence, or
- **D.** the publication or dissemination of the content constitutes a relevant offence

"Relevant offence" means—

(b) an offence specified in Schedule 6 (offences related to child sexual exploitation and abuse)

"CSEA content" means content that amounts to an offence specified in Schedule 6.

SECTION 6 makes references to sexual offences in England and Wales, Northern Ireland and Scotland. This includes Sexual Offences Act 2003, section 14: arranging or facilitating commission of a child sex offence. This provision of the 2003 Act is currently being amended by the Police, Crime Sentencing and Courts Bill to: expand the child sex offences included within the offence (i.e. to include rape of a child and others) and to strengthen sentencing (i.e. so that those convicted of section 14 offence will be sentenced as if they had physically perpetrated the abuse of the child).

IJM UK ASSESSMENT:

The alignment of the Online Safety Bill with these provisions of the 2003 Act has the potential to be highly significant in addressing the demand for online child sexual exploitation.

Ensuring accountability for sex offenders in the UK who livestream child sexual abuse around the world is extremely challenging. This is in part due to the lack of detection, particularly of livestreamed and newly produced CSEA content. There has also been a trend of lenient sentencing for those convicted of directing and livestreaming this abuse. However, the duties imposed by the Online Safety Bill ought to improve prevention, detection and reporting, and the amendments to the 2003 Act have the potential to improve accountability.

CLAUSE 83: OFCOM's register of risks and risk profiles

OFCOM must carry out risk assessments to identify and assess the following risks of harm presented by Part 3 services of different kinds:

- **A.** the risk of harm to individuals in the United Kingdom presented by illegal content;
- **B.** the risk of harm to children in the United Kingdom, in different age groups, presented by content that is harmful to children;
- **C.** the risk of harm to adults in the United Kingdom presented by content that is harmful to adults present on regulated user-to-user services.

IJM UK ASSESSMENT:

It will be a significant advantage for OFCOM to oversee the risk of harm presented by the regulated services. However, harm should not be limited to those in the UK. Online harms are global in nature, and IJM has seen the abuse perpetrated by British nationals upon children in the Philippines.

RECOMMENDATION:

Amend Clause 83 to include –

The risk of harm posed by individuals in the United Kingdom in relation to adults and children in the UK or elsewhere through the production, publication and dissemination of illegal content

CLAUSE 103: Notices to deal with terrorism content of CSEA content (or both)

If OFCOM considers it necessary and proportionate to do so, they may give notice to a regulated service provider to:

use accredited technology to identify CSEA content, whether communicated publicly or privately by means of the service, and to swiftly take down that content

use accredited technology to identify search content of the service that is CSEA content and to swiftly take measures designed to secure, so far as possible, that search content of the service no longer includes CSEA content identified by the technology

OFCOM may give such a notice after giving a warning notice to the provider that they intend to give such a notice relating to that service or that part of it.

CLAUSE 104: Matters relevant to a decision to give a notice under section 103(1):

- **A.** the kind of service it is;
- **B.** the functionalities of the service;
- **C.** the user base of the service;
- **D.** in the case of a notice relating to a user-to-user service (or to the userto-user part of a combined service), the prevalence of relevant content on the service, and the extent of its dissemination by means of the service;
- **E.** in the case of a notice relating to a search service (or to the search engine of a combined service), the prevalence of search content of the service that is relevant content;
- **F.** the level of risk of harm to individuals in the United Kingdom presented by relevant content, and the severity of that harm;
- **G.** the systems and processes used by the service which are designed to identify and remove relevant content;
- **H.** the extent to which the use of the specified technology would or might result in interference with users' right to freedom of expression within the law;
- **I.** the level of risk of the use of the specified technology resulting in a breach of any statutory provision or rule of law concerning privacy that is relevant to the use or operation of the service (including, but not limited to, any such provision or rule concerning the processing of personal data);
- **J.** whether the use of any less intrusive measures than the specified technology would be likely to achieve a significant reduction in the amount of relevant content.

IJM UK ASSESSMENT:

The move away from requiring CSEA content to be prevalent and persistent before enforcement action can be taken is a positive. It is good that OFCOM will have the opportunity to consider a range of factors.

IJM is still concerned at the inclusion of 'prevalence' as a factor owing to the difficulty in detecting newly-produced CSEA content, especially livestreamed abuse.

RECOMMENDATION: Amend subsections (d) and (e) to replace *'prevalence'* with *'presence'*.

As stated above, IJM is concerned about limiting the focus to the risk of harm facing individuals in the UK. Rather the Bill should recognise the harm which UK nationals cause to people around the world, including children in the Philippines.

RECOMMENDATION:

Amend Clause 104 to include –

> ***The risk of harm posed by individuals in the United Kingdom in relation to adults and children in the UK or elsewhere through the production, publication and dissemination of illegal content***

CLAUSE 112: Confirmation decisions

If OFCOM are satisfied that there has been a failure to comply with a requirement having been notified, OFCOM may issue a "confirmation decision".

Under **CLAUSE 113** the confirmation decision can require that certain steps be taken to remedy a failure.

CLAUSE 114 enables OFCOM to require action to be taken in relation to risk assessments.

CLAUSE 116: Proactive technology

A proactive technology requirement may only be imposed in a confirmation decision only for the purpose of complying with, or remedying the failure to comply with, any of the duties relating to (amongst others) illegal content or children's online safety.

Before imposing a proactive technology requirement in relation to a service OFCOM must consider: the kind of service it is;

- **A.** the kind of service it is;
- **B.** the functionalities of the service;
- **C.** the user base of the service;
- **D.** the prevalence of relevant content on the service and the extent of its dissemination by means of the service, or (as the case may be) the prevalence of search content of the service that is relevant content;
- **E.** the level of risk of harm to individuals in the United Kingdom presented by relevant content present on the service, or (as the case may be) search content of the service that is relevant content, and the severity of that harm;
- **F.** the degree of accuracy, effectiveness and lack of bias achieved by the kind of technology specified in the decision;
- **G.** the extent to which the use of the kind of proactive technology specified in the decision would or might result in interference with users' right to freedom of expression within the law;
- **H.** the level of risk of the use of the kind of proactive technology specified in the decision resulting in a breach of any statutory provision or rule of law concerning privacy that is relevant to the use or operation of the service (including, but not limited to, any such provision or rule concerning the processing of personal data);
- **I.** whether the use of any less intrusive measures than the proactive technology specified in the decision would be likely to result in compliance with, or would be likely to effectively remedy the failure to comply with, the duty in question.

IJM UK ASSESSMENT:

Providing OFCOM with the authority to require service providers to adopt proactive technology has significant potential to mitigate harm to children. Ground-breaking technology is being developed which will help to detect and prevent the creation of CSEA. It is hoped that service providers will adopt these technologies in the implementation of their safety duties, but the power to require service providers to do so will hopefully help raise standards across the board.

Once again, IJM would like to see the matters to be considered expanded so as not to be limited to harm caused to people in the UK, but recognising the harm which individuals in the UK can cause to others around the world.

RECOMMENDATION: Amend Clause 116(6) to include -

The risk of harm posed by individuals in the United Kingdom in relation to adults and children in the UK or elsewhere through the production, publication and dissemination of illegal content

IJM would also like to avoid reliance on 'prevalence' amongst the factors to be considered. Not only is there a challenge in ascertaining prevalence of illegal content, especially CSEA content, 'prevalence' indicates that there is some threshold (to be determined) beyond which the material will not be tolerated. This is inappropriate in relation CSEA content, where any presence of CSEA should be cause for alarm.

RECOMMENDATION:

Amend Clause 116(6)(d) replacing 'prevalence' with 'presence'.

CLAUSES 117 TO 122: Relate to the penalties which can be imposed.

IJM UK ASSESSMENT: IJM supports these provisions.

CLAUSE 123: Service Restriction Orders

A service restriction order can be imposed where there has been a failure to comply with a requirement, the failure is continuing and other factors are present (e.g. failure to comply with confirmation decision or failure to pay penalty).

Service restriction orders are orders that require providers of "ancillary services" (persons providing, for example, payment or advertising services) to take steps aimed at disrupting the business or revenue of a non-compliant provider's operations in the United Kingdom. For example, an order could require an advertising service to cease the provision of its service to a noncompliant provider's service.

CLAUSE 125: Access Restriction Orders

An access restriction order can be applied when it is deemed that a service restriction order is insufficient to prevent harm to individuals in the United Kingdom.

An access restriction order can require third parties who provide an "access facility" to take steps to impede access to a noncompliant regulated service, by preventing, restricting or deterring individuals in the United Kingdom from accessing that service. Examples of access facilities are internet service providers and app stores which may be required to restrict access to a service provider's website or app via their service.

IJM UK ASSESSMENT:

IJM welcomes the stronger enforcement provisions of the Bill, most notably the power to restrict services from operating due to a failure to address CSEA content. The potential risk of not being able to operate fully in the UK can be a powerful motivator to ensure that action is taken by service providers to address illegal content.