

for everyone

Call for evidence response form

Please complete this form in full and return to <u>os-cfe@ofcom.org.uk</u>

Title	Call for evidence: First phase of online safety regulation
Full name	Michael Tunks
Contact phone number	07377449342
Representing (delete as appropriate)	Organisation
Organisation name	Internet Watch Foundation (IWF)
Email address	mike@iwf.org.uk

Confidentiality

We ask for your contact details along with your response so that we can engage with you on this consultation. For further information about how Ofcom handles your personal information and your corresponding rights, see <u>Ofcom's General Privacy Statement</u>.

Your details: We will keep your contact number and email address confidential. Is there anything else you want to keep confidential? (delete as appropriate)	Nothing
Your response: Please indicate how much of your response you want to keep confidential. (delete as appropriate)	None
For confidential responses, can Ofcom publish a reference to the contents of your response? (delete as appropriate)	Yes

Your response

Question Your response Question 1: Please The IWF is a UK charity that works in partnership with the internet provide a description industry, law enforcement and government to remove from the introducing your internet (with the co-operation of industry) child sexual abuse organisation, service or images and videos wherever they are hosted in the world and interest in Online Safety. non-photographic images¹ of child sexual abuse hosted in the UK. In addition, the IWF has established reporting portals – places to report online child sexual abuse imagery anonymously and safely - in 49 countries around the world, serving 2.5 billion people. There is a <u>Memorandum of Understanding</u> between the National Police Chiefs' Council (NPCC) and Crown Prosecution Service (CPS) that governs our operations. This ensures immunity from prosecution for our analysts and recognises our role as the "appropriate authority" for the issuing of Takedown Notices in the UK. Operationally, we are independent of UK Government and law enforcement but work closely with both. The IWF also plays a vital role in providing the internet industry with several quality-assured technical services to prevent the spread of known child sexual abuse images and videos online and to stop the upload of images in the first place. These include image hashing utilising Microsoft's PhotoDNA, a URL blocking list of live webpages, keywords list, domain alerts, payment brand alerts, newsgroup alerts and simultaneous alerts (for US companies only). Key to this is our trusted relationship with the internet industry which enables us to act as a broker between them, and government and law enforcement. Our work is funded almost entirely by the internet industry: 90% of our funding comes from our members with the remaining 10% of our funding applying to our role as part of the UK Safer Internet Centre, coming from Nominet, who are a world leading domain name registry based in the UK and responsible for administering the .UK domain.

Please refer to the sub-questions or prompts in the <u>annex</u> to our call for evidence.

¹ Non-Photographic Images include cartoons, drawings, computer generated imagery (CGI) and other nonphotographic depictions of child sexual abuse that are deemed to have breached sections 62-69 of the Coroners and Justice Act (2009).

The IWF has previously received additional Government funding for specific projects and is open to further diversifying its funding mix in the future.

We are a charity registered in England & Wales with an 11-person Board of Trustees of which, eight are independent members and three are industry representatives. The IWF Hotline is <u>audited</u> <u>biennially</u> by an independent team, led by a family court judge, and the report published in full.

Question 2: Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services?

IMPORTANT: Under this question, we are not seeking links to or copies/screenshots of content that is illegal to hold, such as child sexual abuse. Deliberately viewing such images may be a criminal offence and will be reported to the police. The IWF has just marked 25 years as a charity dealing with Child Sexual Abuse Material (CSAM) Online. Over that period, we have assessed **1.8 million** reports and actioned **970,000** reports for removal. Each of these reports can contain from one to thousands of individual images, meaning we have removed millions of child sexual abuse images and videos from the internet in the past 25 years.

In 2021, we investigated more reports of suspected child sexual abuse imagery than the entire first 15 years we were in existence. In the period 1996-2011 we assessed 335,558 reports and in 2021, we investigated **361,000** reports including from members of the public and our analysts own proactive searches. Of these reports, the IWF confirmed **252,000** reports as containing Child Sexual Abuse Material and the number of reports we actioned for removal in 2021, had increased by **64**% on 2020's figures. We have also witnessed over the past decade a worrying increase in the number of Girls appearing in these images. Ten years ago, Girls appeared in 67% of the imagery we removed, in 2021, it was **97%.**

Another concerning trend seen in our data over the last two years has been the rise in what we call "self-generated" indecent images of children. This is imagery that has been produced by children themselves via webcam and then uploaded, shared, or streamed online. These images are then often harvested from their original upload location in a process known as "capping" and then shared on dedicated child sexual abuse websites, where we then find the content and action them for removal. Compared with our 2019 statistics we have witnessed a **374%** rise in this type of imagery, with the **11-13 age range of young girls** appearing most in this type of imagery. In the first quarter of 2022, we have also seen the children in these images getting younger, with **20,000** incidents identified amongst **7-10 year olds** which is nearly 8,000 incidents more than the same period in 2021. In 2021, **147,900** reports contained self-generated indecent images of children, and this now equates for **59%** of all the reports we actioned. Concerningly, we also saw a three-fold increase amongst **7–10-year-olds** in the last year. In 2020, we saw 8,000 instances of self-generated child sexual abuse material in this age range and in 2021, this had increased to **27,000** a **235%** increase.

The National Crime Agency's recent Strategic Threat Assessment for 2021², estimates that as many as **550,000-850,000** people in the UK pose varying degrees of risk to children either online or offline. Their assessment also states that the Covid-19 pandemic and associated restrictions triggered temporary and permanent changes to the offending landscape.

Despite the concerning figures relating to the scale and nature of the threat, both in the generation of the imagery and the demand for the imagery online, the UK continues to have a robust response to this threat. In 1996, the year the IWF was founded, **18%** of the world's known Child Sexual Abuse Material was hosted in the UK. Today, that is less than **1%** and has been ever since 2003, thanks to the partnership approach that is taken between industry, civil society, Government and Law Enforcement.

25 companies in the UK were responsible for hosting Child Sexual Abuse Material in 2021 and **67 notice and takedown** requests relating to **350 webpages** were sent to these companies in 2021. Only 1 of these 25 companies was an IWF member.

The problem of where these images are hosted is largely an extraterritorial one. In 2021, we found **72%** of the content we actioned for removal was hosted in Europe (including Russia and Turkey) with the Netherlands responsible for **41%** of that content.

This imagery is often harvested from its original upload locations and then discovered by our analysts on image hosting boards and cyberlockers which account for **73%** and **10%** of the CSAM we removed from the internet respectively.

An image hosting board lets users upload images which are then available through a unique URL (webpage). These URLs can be made into online links, or be embedded into other websites, forums, and social networking sites.

² <u>https://nationalcrimeagency.gov.uk/who-we-are/publications/533-national-strategic-assessment-of-serious-and-organised-crime-2021/file</u>

	A cyberlocker is a file hosting service, cloud storage service or online storage provider. They are internet hosting services designed specifically to host users' files.
	We found 2,431 instances on video channels, 1,104 instances on social networks and 948 instances on search services in 2021. None of these types of services were responsible for hosting anymore than 1% of the total amount of CSAM we removed from the internet in 2021.
	Further details on the scale of illegal content identified by the IWF can be found in our <u>annual report</u> published on our website.
Question 3: How do you	Is this response confidential? $-Y/N$ (delete as appropriate)
currently assess the risk of harm to individuals in the UK from illegal	Not applicable to IWF as we are not an online service provider.
content presented by your service?	
Question 4: What are	Is this response confidential? – Y / N (delete as appropriate)
your governance, accountability and decision-making	Not applicable to IWF as we are not an online service provider.
structures for user and platform safety?	
Question 5: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?	Is this response confidential? – Y / N (delete as appropriate)
Question 6: How do your	Is this response confidential? – Y / N (delete as appropriate)
terms of service or public policy statements treat illegal content? How are these terms of service	Not applicable to IWF as we are not an online service provider.
maintained and how much resource is dedicated to this?	
Question 7: What can	Is this response confidential? – Y / N (delete as appropriate)
providers of online services do to enhance the transparency,	The IWF plays an important role in the online safety landscape in the UK, in taking reports from members of the public of
accessibility, ease of use and users' awareness of	suspected child sexual abuse material. We also work with our members to assist them in assessing child sexual abuse content
their reporting and complaints mechanisms?	by taking reports from them which may require clarification and we also support law enforcement in grading through the national
complaints meenamsins:	

Child Abuse Image Database and providing hashes to industry to help them detect illegal child sexual abuse imagery.

Many of the larger companies we work with already have well established systems and process in place to allow users to report child sexual abuse material to them. If the company is an American platform, that company will then be required under US law to report any suspected child sexual abuse material to the National Center for Missing and Exploited Children (NCMEC), based in the US. NCMEC is then responsible for processing the report and filtering actionable intelligence about victims and offenders and disseminating that information to the appropriate law enforcement entity globally to pursue. NCMEC publishes annual figures based on the number of reports it receives from companies.

In the UK, the IWF welcomes the introduction of the mandatory reporting body. This will ensure that companies that are within scope of the Online Safety Bill, but not currently reporting through other international structures, such as NCMEC, will be able to report incidents of child sexual abuse material on their platforms and services.

At present, it is difficult for anyone to provide an estimate of the scale of reports the designated body hosted by the National Crime Agency will receive and the Online Safety Bill's impact assessment provides rough estimates from two of the IWF's members who are based in the UK, BT and Jagex, which would suggest that these reports will be relatively low in number and cost. The IWF would suggest, however, that Ofcom, NCA and IWF carry out further research into these figures with the companies, given that there are an estimated **24,000** companies within the scope of the legislation and **17,000** of those are estimated to be small and medium sized companies, there is much that remains unknown about the scale of reporting that we could see from these companies.

Ofcom will need to carefully consider what steps many of these smaller and micro companies will be required to take and how they may be assisted with the reporting process, given that many of these companies will not have the same level of resources available to them as larger companies within the scope of the regulation.

In terms of transparency reports, many of the larger platforms in scope of the legislation are already producing transparency reports. With the new reporting obligations for CSE/A in the UK, to complement structures already in place in the US and anticipated in the EU, it is hoped that we will gain a more informed picture of companies' response to CSE/A with regular annual reporting from designated bodies being able to be compared with company transparency reports.

	In terms of accessibility of reporting and user awareness, one of the biggest challenges we see from our work as being one third of the UK Safer Internet Centre, is that children don't want to report to platforms because they feel it will not help. According to Childnet's project DeShame project, and referenced in a <u>recent</u> <u>report</u> of the APPG for Social Media 43 % of children wouldn't report to a company for this reason.
	There is also a wider issue with children and young people seeking help, with many feeling embarrassed (52%), worried about what might happen next (42%), worried about being targeted by those involved (42%), worried that they are to blame (39%) or would rather sort it out themselves (39%). Project DeShame also stated that children and young people were more likely to block (82%) than report to a platform.
	We believe that further research needs to be carried out into the reasons and motivations behind why children don't report and are currently planning to investigate this with the All-Party Parliamentary Group for Social Media as part of its next inquiry.
	The IWF would also like to see companies doing more to promote and support independent reporting mechanisms to help children and young people. The IWF in partnership with the NSPCC has established <u>Report/Remove</u> a project which, for the first time, enables children and young people to report to us an image that they may have generated themselves and shared with others, that they are now worried may be shared online. This project ensures through the work of the NSPCC and Childline that safeguarding advice and support can be offered to the child, whilst the IWF ensures that the child's imagery, if deemed illegal can be hashed and added to the list that we distribute to companies to prevent its wider distribution online.
Question 8: If your service has <i>reporting or</i> <i>flagging</i> mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?	Is this response confidential? – Y / N (delete as appropriate) Not applicable to IWF as we are not an online service provider.
Question 9: If your service has a <i>complaints</i> mechanism in place, how are these processes designed and maintained?	Is this response confidential? $-Y/N$ (delete as appropriate) Not applicable to IWF as we are not an online service provider.

Question 10: What action does your service take in response to reports or complaints?

Question 11: Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what? *Is this response confidential? – Y / N (delete as appropriate)*

Not applicable to IWF as we are not an online service provider.

Is this response confidential? -N (delete as appropriate)

The IWF provides a range of <u>technical services</u> including a webpage blocking list (<u>URL list</u>), an <u>image hash list</u>, <u>keywords list</u> and a range of other services that support the wider internet industry in preventing the circulation of known child sexual abuse material and the ability to deindex in search services known keywords that return child sexual abuse material.

For these services to be effective, they require hosts, user-to-user services, search services and other companies beyond the scope of the regulation to deploy these services as broadly and widely as possible online. We would like to see as many companies as possible taking steps to prevent the distribution of this imagery online, by becoming a member of the IWF and taking all the technical services that they can to deliver better protections for their users. Currently the IWF has a membership of 180 members, and we are keen to further grow and expand our reach and expertise to assist Ofcom and the companies within scope of the Online Safety Bill to protect their users.

We have seen just how effective these services can be, when in April 2020, just three of the IWF's members serving the UK market blocked **8.8 million attempts** to access known child sexual abuse websites that were on our webpage blocking list.

The IWF's image hash list now contains over **1.5 million unique hashes** of child sexual abuse imagery and is one of the largest data sets in the world. The other unique aspect of our hash list is that it conforms to UK law standards with every image on the list meeting the UK Sentencing Council Guidelines 2014 sentencing criteria of Category A, B or C. Recently, we have also been working hard to improve our classification system to align with other international assessment criteria and have developed Intelligrade, which allows us to add additional contextual metadata to a hash, which enables companies to deploy these hashes globally to other international criteria.

In time, we would like to get to a stage where we can further improve content moderation for both our analysts and for those working in companies, where we can move towards an element of automation on 100% hash matches, where we have seen and

	graded a specific image before, and issue a notice straightaway without an analyst having to view the image again, in order to improve the welfare of our staff and also free them up to view imagery that hadn't previously been detected. However, at present, it remains IWF policy that every image must be reviewed by an analyst and seconded as well as being subject to a quality assurance process to ensure that our technical services meet the highest possible standard.
Question 12: What automated moderation systems do you have in place around illegal content?	Is this response confidential? – N (delete as appropriate) Not applicable to IWF as we are not an online service provider.
Question 13: How do you use human moderators to identify and assess illegal content?	<i>Is this response confidential? –N (delete as appropriate)</i> Not applicable to IWF as we are not an online service provider.
Question 14: How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?	Is this response confidential? -N (delete as appropriate) The IWF has systems and processes in place which enables members of the public, hosts, and owners of content we are requesting to be removed from the internet or contained within one of our technical services to be challenged. The <u>complaints section</u> of our website provides details on the process. In summary, when a complaint is received it will be sent to our Hotline Director in the first instance and the content will be reassessed. We then reply to the person who has made the appeal and inform them of the decision we have reached. If the person appealing isn't happy with the response they have received, the IWF will then refer that decision to the relevant policing agency. The relevant agency will then decide and inform the IWF of the decision that has been made and IWF will act accordingly. Ultimately, if the person appealing remains dissatisfied, they are then able to appeal for Judicial Review. The IWF also conducts regular reviews of the policies, processes, and procedures that we have in place and are regularly audited by a retired High Court Judge on a bi-annual basis. These reports are published on <u>our website</u> and recommendations are adopted in full. The IWF also in 2014 carried out a <u>Human Rights Audit</u> led by Lord Ken Macdonald, the former Director of Public Prosecutions to ensure that our systems and process were fully compliant with human rights

	In terms of how these processes apply to online service providers, we believe it is important that companies have in place clear appeals mechanisms for any restrictions that they may be putting in place. We also believe that companies should be carefully striking a balance between the protection of fundamental rights and removing, blocking, or preventing the spread of illegal content on their platforms.
	We believe that this will be an important part of the risk assessment process that both Ofcom and companies are required to carry out. This careful balance will also play out elsewhere in the implementation of the legislation, particularly around the use of technology notices and the accreditation criteria used and in skilled person reports, ensuring that a balanced and credible report can be relied upon by Ofcom.
	It is important to remember that whilst privacy and unwarranted sanctions are important to safeguard against, children too also have a right to privacy and to ensure that images of their sexual abuse are not freely available online.
Question 15: In what instances is illegal content removed from your service?	<i>Is this response confidential? – N (delete as appropriate)</i> Not applicable to IWF as we are not an online service provider.
Question 16: Do you use	Is this response confidential? – N (delete as appropriate)
other tools to reduce the visibility and impact of illegal content?	Not applicable to IWF as we are not an online service provider.
Question 17: What other	Is this response confidential? – N (delete as appropriate)
sanctions or disincentives do you employ against users who post illegal content?	Not applicable to IWF as we are not an online service provider.
Question 18: Are there	Is this response confidential? – N (delete as appropriate)
any functionalities or design features which evidence suggests can	Please see our response to question 11.
effectively prevent harm, and could or should be deployed more widely by industry?	On design features specifically, we would like to see companies moving towards a safety by design approach, ensuring that the products and services they provide have safety built in during their design and we are supportive of concepts such as regulatory sandboxes where companies can work with regulators and others with expertise to trial their products and services in a secure environment before bringing them to market.
	We have seen how companies have been taking steps to comply with the age appropriate design code for example, by disabling <u>Autoplay</u> on YouTube Kids, <u>Instagram</u> introducing new safety setting making accounts for under 16s private by default, blocking

	some adults from interacting with children and restricting how advertisers can target teenagers and <u>Tik Tok</u> preventing children from receiving notifications past their bedtime are all welcome steps to improving the use of social media and experiences of children and young people online. One challenge that remains is how to age verify a child online. Whilst these steps are welcome, the difficulty that remains for these companies is how do they know that a child has not lied
	about their age to gain access to an adult account that doesn't have these design features as default?
	We are also particularly concerned by the impact pornography has on the development of children and young people. We have been supportive of the attempts to introduce age verification for people wanting to access pornography online. We welcome that Ofcom's roadmap includes this as one of its first four priorities when it seeks to implement the Online Safety regime, but we are concerned that on its present timeline, we may not see this important measure being brought forward before 2024 at the earliest. This is disappointing, when the Government had previously legislated on this issue in 2017 through the Digital Economy Act, Part 3, which wasn't implemented.
Question 19: To what	Is this response confidential? – N (delete as appropriate)
extent does your service encompass functionalities or features designed to mitigate the risk or impact of harm from illegal content?	Not applicable to IWF as we are not an online service provider.
Question 20: How do you	Is this response confidential? – N (delete as appropriate)
support the safety and wellbeing of your users as regards illegal content?	Not applicable to IWF as we are not an online service provider.
Question 21: How do you	Is this response confidential? – N (delete as appropriate)
mitigate any risks posed by the design of algorithms that support the function of your service (e.g. search engines, or social and content recommender	Not applicable to IWF as we are not an online service provider.
systems), with reference to illegal content specifically?	
specifically:	

Question 22: What age	Is this response confidential? – Y / N (delete as appropriate)
assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?	The IWF is not best place to comment on Age Verification technologies as this is not our core business. However, we are aware that the Digital Policy Alliance and British Standards Institute (BSI) has developed and published a standard- PAS 1296:2018 entitled: Online Age Checking- Provision and use of online age checking services- Code of Practice, which has been widely adopted by many of the providers working within the sector. The Age Verification Providers Association (AVPA) are also working closely with the Institute of Electrical and Electronics Engineers (IEEE) on further standards to be published this year. The European Union are also considering further research and steps on age verification with their <u>EU Consent project</u> .
Question 23: Can you identify factors which might indicate that a service is likely to attract child users?	Is this response confidential? – N (delete as appropriate)
Question 24: Does your	Is this response confidential? – N (delete as appropriate)
service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?	Not applicable to IWF as we are not an online service provider.
Question 25: If it is not	Is this response confidential? – N (delete as appropriate)
possible for children to access your service, or a part of it, how do you ensure this?	Not applicable to IWF as we are not an online service provider.
Question 26: What	Is this response confidential? – N (delete as appropriate)
information do you have about the age of your users?	Not applicable to IWF as we are not an online service provider.
Question 27: For purposes of	Is this response confidential? – N (delete as appropriate)
transparency, what type of information is useful/not useful? Why?	The IWF was part of the UK Government's Ministerial working group on Transparency which produced a <u>report</u> in 2020 which made 38 recommendations for shaping the transparency regime that Ofcom and Government will be seeking to implement. The Technology Coalition in the United States has also recently published <u>its own</u> <u>voluntary framework</u> for transparency at the WePROTECT Global Alliance summit in June 2022.
	We believe that whilst providing quantitative data is important it is also important for the regulator and companies to provide a narrative

	behind the data. For example, a question we often grapple with, is a company reporting it has removed vast amounts of CSAM a bad player in comparison with a company that hasn't reported removing any? The company removing a vast amount of CSAM, might be detecting it at the point of upload before it is publicly viewable on the platform and then reporting it in line with its lawful obligation. Whilst the company not reporting any CSAM could have lots present on its services, but currently cannot detect it (for example it is in an End-to-End Encrypted environment). This context will be crucial in trying to distinguish for the public between which services are doing all that they can to detect CSAM vs the ones where clear improvements need to be made.
	We believe it will be important for Ofcom as part of its risk assessment process to understand how platforms work and operate and how they are enforcing their Terms and Conditions. If a platform states it doesn't allow something on its services, particularly illegal content, then what systems and process does it have in place for preventing it? How effective are these services and what more could they possibly be doing? In the case of CSE/A given the nature of the content, not all of this information will be able to be made public, as we do not want to direct members of the public inadvertently towards CSAM, but it is important that a positive dialogue and iterative approach to the regulation of these platforms is established early in the process.
	This is also an area that the IWF has offered to provide assistance to Ofcom with through our regulatory model and we would welcome the opportunity of providing further information.
Question 28: Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?	Is this response confidential? – N (delete as appropriate) No Further comments to add.

Please complete this form in full and return to OS-CFE@ofcom.org.uk