making communications work
for everyone

Your response

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.

Is this a confidential response? (select as appropriate)

No

I have been heavily involved in online child protection for over 10yrs, working for UK Law Enforcement, Online Safety Education and also within a Social Media platform. I have been at the forefront of reviewing, assessing, investigation all manner of online harms from cyberbullying, violence, sextortion, grooming, self harm, suicide, but my expertise is mainly Child Sexual Exploitation (CSE) and related CSE images (CSAM). I have worked alongside global partners and well as key UK stakeholders and constantly update my LinkedIn profile with current online harm information. My knowledge and expertise would therefore be highly valuable for this response.

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Is this a confidential response? (select as appropriate)

No

Any and all services that have an Internet connectivity would attract minors, from Apps on smartphones/devices, to gaming companies (Xbox/Playstation etc), discussion forums, news related websites and other media sites (TV etc). There is no 1 area that would specifically attract minors, it is purely down to what is used most. It is where ever minors can chat, play games, give responses to topics, share information to name a few, and this can be anyone of the above services.

Question 3: What information do services have about the age of users on different platforms (including children)?

Is this a confidential response? (select as appropriate)

No

Question 3: What information do services have about the age of users on different platforms (including children)?

This will depend on each individual ‘provider’ of the service. Some will ask for date of birth, age, location, phone number, e-mail etc, where as others may just ask for a Username. However, this can then all depend on what information the User chooses to share. For instance, a user might post an image in a school uniform, or at a birthday party – which therefore gives an indication of age (which is some cases contradicts the DoB given!). Some platforms request proof of age, so this information can be confirmed, however this is not the case for the majority. The only information usually held by a platform is a DoB, but this is what the user chooses to input and is not confirmed or verified.

Question 4: How can services ensure that children cannot access a service, or a part of it?

Is this a confidential response? (select as appropriate)

No

On certain platforms, they will ‘hide’ the media with a warning of adult content/disturbing content, however to view the image, all you do is click ‘view’. Some platforms will limit certain content based on the DoB inputted, however many young people are aware of this, so enter false details. The only way currently would be age confirmation to limit what is viewed, or if a platform identified an account for that of a minor, they could add relevant restrictions themselves. However, this would not work as well for gaming platforms (PS/Xbox) for those minors wishing to play more ‘harmful’ games, as usually it’s the parents who have set the account up and are therefore over 18.

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

Is this a confidential response? (select as appropriate)

No

There are currently a multitude of tools and services available to platforms to clarify and confirm age assurance, however its not necessarily the cost which is the factor (although plays a part), it’s the undertaking of such a task. Most platforms are global, so this would have to be introduced on a global scale, with differing laws and legislation, and differing age consents – a highly complicated process. How would this be achieved globally, you would need a confirmed age verification process for each Country! The other point is how could you review current users ages, you would have to introduce the process for every user. There would not only be cost for the technology itself, but the staffing in what ever form could massively increase the financial impact.

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Is this a confidential response? (select as appropriate)

No

You can view harmful content on any platform at any point in time, it can be violence related, self harm, CSE, nudity etc, a lot is shared and posted for various reasons, news, humour, outrage. For most young people who are 'tech savvy', it is very easy to type in phrase or words into a search engine and find the content.

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

Is this a confidential response? (select as appropriate)

No

I have seen and witnessed clear evidence that all manner of harmful content has towards young people. Minors who are highly involved in the sexual element, selling media, selling items etc. I have seen minors glorifying violence (gangs, weapons) and also an increase in fighting with school environments and posting the media, which can also lead to bullying. I have handled cases relating to pro suicide discussions and young people being encouraged to self harm. We may also consider how harmful the risk of certain influencers play towards young people.

Question 8: How do services currently assess the risk of harm to children in the UK from content that is harmful to them?

Is this a confidential response? (select as appropriate)

No

It is for individual services to decide what they consider is harmful and also how they handle the content. Some may report this direct to Law Enforcement (Police), other platforms may just take the content down and not report. As some platforms have differing views on some content, 1 may leave it up and 1 may report, there is no coordinated response for what is harmful content.

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

Is this a confidential response? (select as appropriate)

No

It is understanding what is a ‘risk’ or ‘harmful’. What maybe viewed as harmful to 1 person, may not be seen the same by another, hence this can cause confusion. We also must consider context with regards to risks and harm, this is vitally important, but overlooked. If the content is on a global platform, how can we confirm the risk is solely for UK minors, as it maybe legal in other Countries.

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

Is this a confidential response? (select as appropriate)

No

This is varied across all services and there is not a specific criteria that fits all. Some platforms have specific child safety teams, others do not. Some have specialist experts, other are more generalists. You can be a manager with management experience, but not child safety experience, which therefore can impact on the relevant and correct decisions being made or implemented. Current accountability is currently being argued under the UK Online Safety Bill and the US Sec 230 debate. At present, there is no singular accountability within platforms. Governance can depend on Country regulations, internal policies and reporting regulations (CSE only)

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Is this a confidential response? (select as appropriate)

No

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Most ToS are long winded and to be honest very few users either read or take notice of! The majority of users know what is illegal and what isn't, but there is confusion around legal yet harmful. Services should add 'reminders' when Users log in, and/or ask them to review updated ToS when changed. There should also be clear messages about what is acceptable on 'front pages'. There are 2 major platforms who are also involved in giving direct safety advice (in various ways) to young users and talking with them face to face. This action should be applauded and encouraged further.

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[1]

Is this a confidential response? (select as appropriate)

No

The main focus here is what is or would be classed as a priority, and in simple terms it would fall under either 'immediate risk of harm' or 'potential risk of harm'. Therefore "Primary Priority" could be CSE related content, suicide, self harm – where there is an immediate risk;
"Priority Harmful Content" – would be classed as 'potential', where there is clear abuse, bullying, certain violence related content etc..

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Is this a confidential response? (select as appropriate)

No

Current procedures for reporting and/or complaining is confusing and complicated, and in some cases very little to zero response received back. User often can feel neglected or not listened too when raising any concerns. These procedures must be simplified and rigid processes in place, however, there must also be an understanding that platforms can receive hundred/thousands of reports a day which requires more staff to handle these matters.

[1] See A1.2 to A1.3 of the call for evidence for more information on the indicative list of harms to children.

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Is this a confidential response? (select as appropriate)

No

The current best practice is to report direct to the particular platform and it is solely down to the current policies in place within that platform.

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

Is this a confidential response? (select as appropriate)

No

Actions are currently taken based current policies, however in some circumstance, if further ‘pressure’ is applied then further action can be taken.
As per above, what is harmful to 1 person, may not be harmful to another, therefore context is important when reviewing complaints. Such reports may also have a knock on effect as reports can be received by minors who are under the relevant age for being on the platform anyway, so action can be taken against them too. Although very hard to determine, there should be a ‘common sense’ approach to certain online harms and decisions should be based on this approach as well.

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Is this a confidential response? (select as appropriate)

No

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.
Age restricted content is currently available, if the user has added a DoB indicating under 18 (may or may not be correct). In some cases, a warning of harmful content may be displayed or content hidden behind a blurred image until clicked.

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?
Is this a confidential response? (select as appropriate) No
Age verification is a possible solution, but as mentioned, difficult to implement. Better communication between Agencies and Platforms identifying and handled recognised harmful content. Increased training for moderators to identify and flag concerning content.

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?
Is this a confidential response? (select as appropriate) No
This issue is not the sole responsibility of services/platforms to try to manage. They have a huge part to play and there is much more that can be done, however, there must be better online safety education for young people and adults, and more conjoined reporting service for UK users. There must be more closer working and interaction between UK Agencies and platforms. Only working together will help improve this issue and improve safety for UK Child users.

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

Is this a confidential response? (select as appropriate)

No

The basic premise of any platform is based on ‘interactions’, this maybe family, friends, people you may know etc – however the same algorithm will identify what you have looked at, what you have liked and then suggest similar interests. For example, if you have looked at self harm, it will then suggest further users or groups also involved or interested in that area. This will happen across any and all genres. It is possible, algorithms could be changed to identify certain content and not to recommend it if looked at, however, this would have to be constantly updated as topics and identifiers change repaidly.

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Is this a confidential response? (select as appropriate)

No

There has to be closer working between experts in the field, agencies, regulators and platforms. At present it is very convoluted and a serious lack of joined up working. We need to be aware of what is happening on platforms, what is trending, what new harms have appeared and how to deal with them. This doesn’t happen at present!

There are technical capabilities available that can and could assist with improving content moderation, but it’s down to choice, cost and system integration whether it is utilised or not. It would also require constant updating to be fully effective. Increased human moderation is also highly advantageous, however this is another financial impact on any company.

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Is this a confidential response? (select as appropriate)

No

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

There are numerous systems either in existence or in development that can identify and action harmful content. The most utilised and well known in the field is CSE media detection capabilities, widely used across Law Enforcement and Platforms – however, there are still issues, which sharing such tools, which could assist quicker identification and results.

There are tools that can identify certain keywords (harmful), tools that can identify objects (guns/knives etc) and also voice detection tooling. The UK is very much leading the way with a lot of these techniques, but its still very under utilised or shared to be able to have the maximum effect

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Is this a confidential response? (select as appropriate)

No

There are various agencies and teams that could be classed as 'human moderators' who view and assess content deemed harmful. This ranges from UK charity help centres, Law Enforcement Teams, Corporate Moderation Centres (CMC) who work on behalf of platforms and platform specialists themselves. CMC moderators usually are viewing mass information with a short period of time to make a review and decide on relevant action, so certain factors can be over looked. Content may then get flagged to a specialist to review depending on internal policies. UK LE may also receive direct reporting from various sources and assess for appropriate action.

Question 23: What training and support is or should be provided to moderators?

Is this a confidential response? (select as appropriate)

No

Most training for moderators is done 'in house', usually by another employee who either has previous skills and experience, who has been there the longest! There isn't any external Company that I'm aware of who does or currently can provide this training. Although there is certain legal controls when it comes to training around CSE/CSAM content.

There should also be suitable and appropriate welfare and wellbeing support and training, which should start before any moderator has started the job, and this should be an ongoing review (at least 6months). You never know when you will see that 1 post, media or content that may affect your mental wellbeing as everyone is different.

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

Is this a confidential response? (select as appropriate)

No

Depending how systems have been set up, there are various ways moderators and automation can work together. Some automation will automatically action certain content, some times it will flag it for human review and unless the automation is fully up to date, then content will be missed and it will be for a human to identify and action. Automation will never take into account 'context', so its possible false positives maybe identified. It is also worth noting, especially re CSE, user are aware of certain automation and will try to use detection avoidance techniques, which means only human moderation is possible,

Question 25: In what instances is content that is harmful to children, that is in contravention of terms and conditions, removed from a service or the part of a service that children can access?

Is this a confidential response? (select as appropriate)

No

Content will be removed if it is perceived that it would or could cause risk of harm to a minor user. Such as pro self harming, certain imagery, certain sexualised content. It may, depending in the Country and platform include certain glorification of violence content.

Question 26: What other mitigations do services currently have to protect children from harmful content?

Is this a confidential response? (select as appropriate)

No

Question 26: What other mitigations do services currently have to protect children from harmful content?

None apart from those mentioned above; warning label, blurring media, age restriction if undertaken

Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?

Is this a confidential response? (select as appropriate)

No

There will always be an attempt by children to circumvent these mitigations, depending usually on the age of the user. Teenagers would be more likely than those younger. It is possible to identify a User who has been previously banned or actioned by certain information that can be identified from the account information. Each platform is different, but the general idea is the same. Using these identifiers would assist platforms in stopping young users re-accessing harmful content.

Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?

Is this a confidential response? (select as appropriate)

No

This evidence is based purely on what is known predominantly in the UK, there maybe systems and other services being utilised in other Countries that may assist in reducing and identifying harmful content.
There is a huge opportunity the UK could lead the way in protecting children on the internet, however, all suggestions must be listened too and those people who have actually done the work, seen the harm and would be deemed experts must be involved