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
SafeCast® is a US and UK patented technology which enables the automatic filtering of inappropriate content away from children and vulnerable people on television and the internet by means of a Self Applied Content Rating system. SafeCast is licensable upon RAND (Reasonable And Non-Discriminatory) terms under a SMPTE compliant standards patent licence. Since 2017, section 104 of the UK Digital Economy Act has required broadcasters and internet service providers to comply with filtering best practices that protect children and vulnerable people from inappropriate content. When uploading a video, the creator inserts a label that indicates how safe the video is for children - whether it contains any pornography, horror, or violence that could harm or could cause distress to children of different ages. Then, when viewing the video, lightweight filters in the device or browser can read the label and filter away content that is inappropriate for a child's age and maturity. This is not censorship; adults are able to view the unfiltered content. The filtering process does not require censorship nor compulsory age verification technology. In the UK for example, the SafeCast Self Applied Content Rating system can be mapped onto the Key Stages of the UK National Curriculum, rather than to the exact age of the child. Any smartphone, tablet, pc, or television can be set to filter in accordance with a child's school age. This makes it ideal for international air travellers where children need to be protected on aircraft travelling across multiple countries and jurisdictions. But its use could be global owing to its respect for "*Digital Sovereignty*". Digital sovereignty is the term generally used for the right of a nation state to control the internet and its use by people within its borders. In this age of globalised trade no state wishes to become a pariah state - constrained by its fellow nation states for failing to protect basic human rights. There is, however, continual debate about the appropriate extent of freedom of speech and human rights in a nation state.

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

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

SafeCast, on its own, does not identify factors which might attract child users. However, SafeCast's *"Self-Applied Content Ratings"* can be included as metadata in ALL video content and can give rise to effective automated filtering of content and advertisements for child and vulnerable person protection purposes. It is thus a technology which can be included to enhance and improve such identification services. Tools can be created around the digital standard previously known as TSP 2121 but now referred to as SMPTE RDD 59-1 IMF Application DPP (ProRes)

SafeCast is a member of OSTIA (the Online Safety Tech Industry Association) and it is happy to licence OSTIA participant companies for the use of its patents in the Safety Tech Challenge Fund Link Sharing project so that it will be possible for the metadata associated with any content or advertisement to be looked at prior to a video being displayed. Any inappropriate content or advertisement can thus be filtered away prior to it being shown to a child or vulnerable person. A filter in a mobile device whose sole purpose is to look at metadata in a video prior to it being shown is a task which involves no material delay and no material processing overhead.

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

Platforms have very detailed information on the age of their users (including children). This information is being misused by Video Sharing Platforms and broadcasters to cause harm to children and their parents. Ofcom needs to know the use of cookies and tracking continues regardless of the legislation.

Currently websites often provide visitors with the opportunity to opt out of data collection. Legal frameworks like Europe's General Data Protection Regulation (GDPR) require websites and associated third parties to get consent before collecting and processing personal data. To help website operators comply with that requirement, vendors like Didomi, Quantcast, OneTrust, and Usercentrics offer what's known as a consent management platform (CMP). These firms provide software that websites use to prompt visitors to accept or reject cookies in order to control how personal information gets handled. They claim their respective CMPs allow companies to comply with privacy laws in the US, EU, UK, Brazil, South Africa, Singapore, and elsewhere.

Recently, computer scientists Zengrui Liu (Texas A&M University), Umar Iqbal (University of Washington), and Nitesh Saxena (Texas A&M University) devised an auditing mechanism to test the effectiveness of CMP-based opt-out controls and found these platforms don't necessarily ensure compliance with GDPR and CCPA requirements. Their paper "Opted Out, Yet Tracked: Are Regulations Enough to Protect Your Privacy?" shows that Opt-out under the law thus is not all that different from "Do Not Track" – a web specification that allowed browser users to declare the desire not to

be tracked, without any consequences for ignoring that preference. Ofcom needs to keep this reality in mind when engaging with VSPs and must be highly sceptical regarding alleged compliance with GDPR requirements.

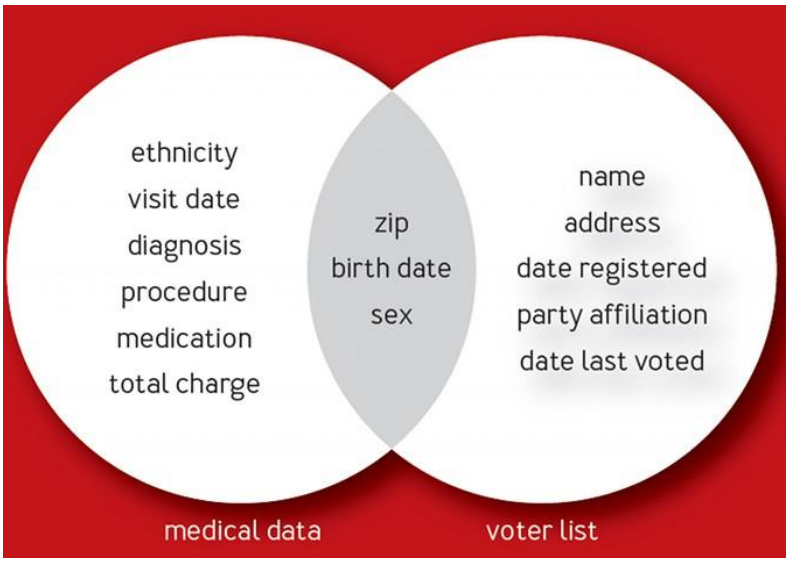
Ofcom is also respectfully directed to consider herein the work of Professor Ross Anderson who in late 2020 published the third edition of "*Security Engineering: A Guide to Building Dependable Distributed Systems*". I would specifically direct Ofcom to Chapter 11 *"Inference Control*" from this edition. I cite below what Professor Anderson says in his introduction to this chapter:

> 11.1 Introduction
>
> Just as Big Tobacco spent decades denying that smoking causes lung cancer, and Big Oil spent decades denying climate change, so also Big Data has spent decades pretending that sensitive personal data can easily be 'anonymised' so it can be used as an industrial raw material without infringing on the privacy rights of the data subjects.
>
> Anonymisation is an aspirational term which means stripping identifying information from data in such a way that useful statistical research can be done without leaking information about identifiable data subjects. Its limitations have been explored in four waves of research, each responding to the technology of the day. The first wave came in the late 1970s and early 1980s in the context of the U.S. census, which contained statistics that were sensitive of themselves but where aggregate totals were required for legitimate reasons such as allocating money to states; and in the context of other structured databases from college marks through staff salaries to bank transactions. Statisticians started to study how information could leak, and to develop measures for inference control.
>
> The second wave came in the 1990s as medical records were computerised. Both health service administrators and medical researchers saw this as a treasure trove, and hoped that removing patients' names and addresses would be enough to make the data non-personal. This turned out to be insufficient because of the richness of the data, which led to tussles in several countries including the USA, the UK, Germany and Iceland. There have since been multiple scandals when inadequately anonymised data were leaked or even sold.
>
> The third wave, in the mid-2000s, came when people realised they could use search engines to identify people in large datasets of consumer preferences such as movie ratings and search engine logs. An advance in theory came in 2006, when Cynthia Dwork and colleagues developed the theory of differential privacy which quantifies the extent to which inferences can be prevented by limiting queries and adding noise, enabling us to add noise where it's needed. This is now being used in the US census, whose experience teaches a lot about its practical limits.
>
> The fourth wave came upon us in the late 2010s with social media, pervasive genomics and large databases of personal location histories collected by phone apps and widely sold to marketers. Ever more companies who sell personal information at scale pretend that it isn't personal because names are somehow tokenised. Ever more press articles show how bogus such claims usually are. For example, in December 2019 the New York Times reported ana-

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

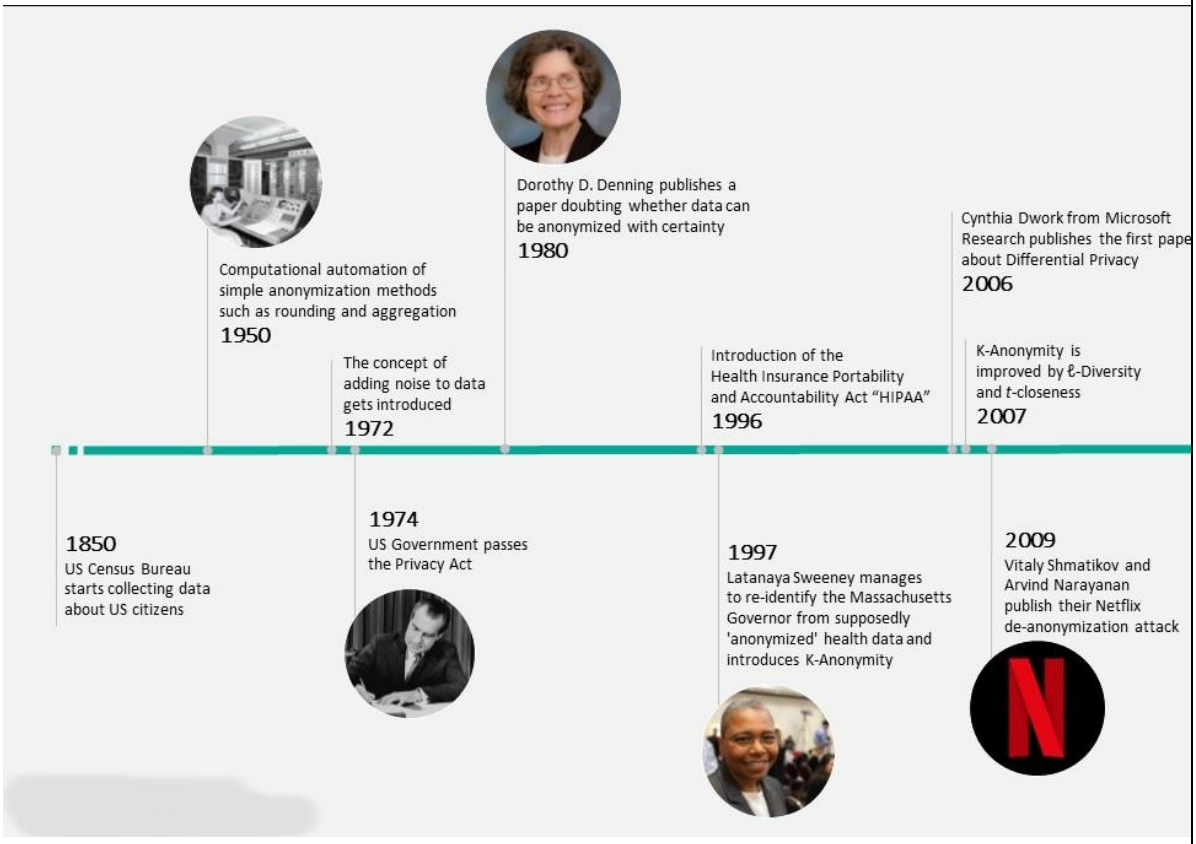
> lysing the mobile-phone location history of 12 million Americans over a few months, locating celebrities, rioters, police, Secret Service officers and even sex-industry customers without difficulty.
>
> We face a yawning gap between what can be done using anonymisation and related privacy technologies, and what stakeholders from medical researchers through marketers to politicians would like to believe is possible. This gap has been the subject of much discussion and, as with tobacco and carbon emissions, political argument. As our knowledge of the re-identification risks becomes ever more detailed and certain, so the hopes of both governments and industry become ever more unrealistic. Governments repeatedly call for proposals, and data users call for contractors, to create services that cannot be created; all too often, contracts for privacy services are won by the more ignorant or unscrupulous operators.
>
> It must be said that not all governments have simply been ignorant. Both the UK and Ireland, for example, annoyed other EU member states for years by allowing firms to pretend that data were anonymous when they clearly weren't, and this was one of the factors that led the EU to pass the General Data Protection Regulation (GDPR), as I will discuss later in section 26.6.1. Since it came into force, the wriggle room for wishful thinking has become less – though even the European institutions have sometimes had a rosy view of what can be achieved by de-identification.

Professor Anderson then goes deeper into the analysis of the issue of 'inference control' and cites a variety of sources. The science shows that anonymised data can be recovered by combining two or more databases and the legislative protection of sensitive personal data is worthless because it can be recovered through these combinations. Below is a graphic which illustrates the overlap between medical data and voter lists. It is now too late for society to recover from this risk.



In his inference control chapter Professor Anderson outlines the history of inference control. I set out below a version of Professor Anderson's timeline which I have adapted from a graphic produced by Nicolas Sartor at Aircloak.

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



In 2010 Professor Paul Ohm (now Professor of Law and Associate Dean, Georgetown Law School) published a highly influential paper, "Broken promises of privacy', I enclose a link to a copy of this very important paper. Professor Ohm noted that '*scientists have demonstrated they can often 'reidentify' or 'deanonymize' individuals hidden in anonymized data with astonishing ease*' and confessed *'we have made a mistake, labored beneath a fundamental misunderstanding, which has assured us much less privacy than we have assumed. This mistake pervades nearly every information privacy law, regulation, and debate, yet regulators and legal scholars have paid it scant attention*.'

As Professor Anderson expressed when discussing Professor Ohm's paper:

> For the previous thirty years, computer scientists had known that anonymisation doesn't really work, but law and policy people had stopped their ears. Here at last was an eminent lawyer spelling out the facts, telling the story of AOL and Netflix, in a law journal and using lawyer-accessible language. Among other things he ridiculed Google's claim that IP addresses were not personal information so that its search logs should fall outside the scope of data protection, denounced the binary mindset of data as either personal or not, and called for a more realistic debate on privacy and data protection.
>
> In 2012, a report from the Royal Society called for scientists to publish their data openly where possible but acknowledged the reality of re-identification risks: '*However, a substantial body of work in computer science has now demonstrated that the security of personal records in databases cannot be guaranteed through anonymisation procedures where identities are actively sought*'.

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

> In that year, the UK Information Commissioner also developed a code of practice on anonymisation; as the ICO is the privacy regulator, such a code can shield firms from liability, and it was the target of vigorous lobbying. The eventual code required data users to only describe their mechanisms in general terms, and shifted the burden of proof on to anyone who objected. This was a less stringent burden than the ICO applies in freedom-of-information cases, where a request for public data can be refused on the presumption that the data subjects' *'friends, former colleagues, or acquaintance*s may know relevant context. …

In the light of this evidence Ofcom is asked to follow the reasonable requests of parliamentarians to protect children by mandating that Age Gating based upon school age rather than actual age is required on all Video Sharing Platforms. (The case for Age Gating as a safer alternative to Age Verification was set out in my response to the First Ofcom consultation on the Online Safety Bill which is cited in my reply to Question 4 in this second consultation.)

Ofcom needs to *"follow the science"* to adequately protect children. (Further information on how this could be done in the UK is set out in my answer to Question 5).

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

In the First Ofcom consultation on the Online Safety Bill I set out the case for Age Gating and opened with the statement "*The current bill is ineffective, too costly and biased in favour of the gambling industry and totalitarian censorship — without properly protecting children and vulnerable adults from online harms"*

Our proposals would:

- Enable the key requirements of this Bill to come into force and protect children **within this Parliament** rather than many years from now since the necessary enabling changes can all be made in software/middleware in existing phones, tablets, set top boxes and PCs;
- Assist the broadcasting and internet industries in the **elimination of fake news and fake advertising to children and vulnerable people**.
- Assist the content and advertising industries in **developing high quality child protection services and content around the world**;
- **Reduce the enforcement cost of the Online Safety Bill** — and make it more **effective**.
- **Protect epilepsy sufferers from being harmed by trolling** — see #ZachsLaw — **without the need for new legislation in each country of the world.**

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

Ofcom are respectfully requested to publish my evidence to the First Ofcom consultation on the Online Safety Bill so that others could consider and comment upon my proposals. In the interim a copy of my September 2022 evidence is being made available at this weblink.

Additionally, Ofcom is respectfully requested to engage with GCHQ where in November 2022 I lodged a confidential response to *"Thoughts on Child Safety on Commodity Platforms"* by Dr Ian Levy and Crispin Robinson. Dr Ian Levy was the Technical Director of the National Cyber Security Centre, and he wrote the paper with Crispin Robinson, Technical Director of Cryptanalysis at GCHQ.

In the non-confidential parts of my response to GCHQ, I said:

Metadata labelling if done in accordance with a global standard can enable the quick and effective removal of potentially harmful content without censorship through the use of lightweight filters. This would greatly reduce the areas which the security services and the NCA need to actively police and review content. It is also the only way in which there could be an effective UK "CyberTipline" service which adhered to Ofcom transparency and openness requirements given the numbers involved . Furthermore, the need for "Outcome21" peer to peer protections, so that children are not criminalised for just being curious and social amongst their peers, can only be implemented in accordance with global standards. Client side protections require economies of scale which can only be deployed in accordance with a standard that does not create commercial barriers to new entrants or protected silos for the incumbents. Failure to implement these measures could also result in some long tail risks as youthful behaviour resurfaces from web archives in a child's adult life - this has already been identified as a long-term security risk by unfriendly foreign state actors building dossiers for blackmail at a future time.

More recently Professor Ross Anderson published a non-confidential response to this same paper entitled "Chat Control or Child Protection". In his response Professor Anderson said:

In short, the data do not support claims of large-scale growing harm that is initiated online and that is preventable by image scanning. Yet real harm has been done by false positives. The first wave of prosecutions for illegal abuse images, Operation Ore, swept up many innocent men who were simply victims of credit card fraud, their cards having been used to pay for illegal material. A number were wrongly convicted, and at least one innocent man killed himself. The organisation responsible, CEOP, became part of SOCA and then of the NCA. It still uses as a metric the number of children 'safeguarded'. This term is elastic; it can mean that a child has been taken into care (whether rightly or wrongly); it can mean that a parent or carer has been arrested, or accepted a caution, or signed the Sex Offenders' Register, or (in the context of 'Prevent', which we discuss later) that society has been safeguarded from a child considered to be dangerous. The police have since acknowledged that too much effort has been put into indecent images and not enough into preventing actual abuse of minors.

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

Our context should therefore be crimes of sexual violence against minors. The European Parliament decided in 2017 not to use the American term "child pornography" but "child sexual abuse material" (CSAM) instead. Rather than getting into a terminological dispute, we will use the term "CSAM" in what follows, but we prefer to read it as "Crimes of Sexual violence Against Minors". Germany considers children to be those under 14, while Denmark considers sex with minors under 15 to be a serious offence; yet the CSA Regulation will apply to images not just of children but of young people under 18. The term 'minor' is thus more accurate. Similarly, when we come to discuss terrorism, we will prefer the more general and less emotional terms "violent political extremism" for the kinetic variety and "violent online political extremism" for its online aspects.

Professor Anderson then rightly highlights the dangers of "*false positives*" in NLP and establishes beyond a shadow of doubt that this cannot work if Ofcom's fundamental requirements in respect of transparency and openness genuinely to be part of a regulatory system. Professor Anderson says:

"It is hard to see how anyone could trust an NLP text scanning tool that was trained on data to which the public and even public-interest technologists have no access. There are too many ways in which machine-learning pipelines can be subverted: manipulating the inference engine, or the training data, or its labelling, or even its batching.

Independently, in the USA, within the C2PA which is attempting to develop an open technical standard to provide publishers, creators, and consumers with the ability to trace the origin of different types of media, there is now a move to clarify and contextualize two new C2PA inference engine flags:

- **"Do not train"** (i.e. Do not use this document or object to generate subsequent works through the use of artificial intelligence.)
- **"I am an inference result"** (i.e. The use of this document or object to generate subsequent works is redundant and may worsen systematic bias.) It is apparent that poisoned AI/ML generative applications are able to create thousands of new images that already are biased. Consequently, society needs a way to automatically keep malicious images out of services such as ImageNet and other popular training grounds that are producing well-intended machine learning neural networks.

However, the scientists and engineers who are working together in the C2PA are universally adamant that the implicit objective remains unreachable because the technology of explainable AI is still in its infancy. Notwithstanding the various participants impressive corporate engineering and scientific resources, (C2PA members include Adobe, ARM, BBC, CBC, INTEL, Microsoft, New York Times, Sony, Trupic and Twitter) the C2PA board cannot compile a compelling positioning statement that says that they consider that explainable AI is ever going to exist.

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

At a practical engineering level, when looking at AI systems we need to distinguish between what is sometimes called the "Dog v Cat" problem when compared to the "Dog v Muffin" problem. Today AI systems in daily use are very good at distinguishing photographs of the faces of Dogs from the faces of Cats. This has led to their successful adoption in medicine in reading MRI scans, cancer screening etc as means of diagnosis since the training data can be very similar and very good diagnosis results have been generated. But today's AI systems find it a lot harder to distinguish between a photograph of the face of a dog and the photograph of a blueberry muffin. The American attorney, Damien Riehl, has recently written about this in https://www.linkedin.com/pulse/centaurs-machines-humans-efficacy-cost-matrix-damien-riehl/

Damien Riehl's work suggests to me that Ofcom should be highly conservative in allowing the adoption of AI based analysis where the logical proof of how the inference is generated remains unknown and unknowable. Open research programmes should be mandated with extensive testing and independent peer reviewed analysis before Ofcom approves any use of AI which is not explainable.

However, it may be necessary for there to be an **Explainable AI Commissioner** who certifies, after enquiry on behalf of Ofcom, the deployment of an AI service which is not explainable if the perceived utility of the AI outweighs the fact that nobody understands how the AI comes to its conclusions. While this might lead to some *"black swan"* events it is likely that the benefit will outweigh the risks to society at large.

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

In respect of *"impact and cost"* of age assurance and age verification I would like to highlight a concern in respect of *"regulatory capture"* originally identified as an issue in *"Thoughts on Child Safety on Commodity Platforms"* by Dr Ian Levy and Crispin Robinson of the NCA and GCHQ.

Finally, we consider regulatory capture. This is the phenomenon where a regulator acts in the interests of a small number of those it seeks to regulate at the expense of a much larger population. Many of the child protection charities that manage these databases are at least partly funded by the big tech companies whose services are the subject of this discussion and so it is reasonable to ask how we can be certain that the service owners are not manipulating the curation of the database.

SafeCast considers that this is something which Ofcom should be very aware of. The COVID lockdown has undermined transparency in lobbying. Civil servants and business people have been having to work from home with no administrative or peer supervision. The opportunity for secret, biased lobbying and secret special pleading has never been greater with Parliament and parliamentarians unable to control its misuse.

There will naturally be a bias in Government in favour of *"free"* technology where the costs of delivery are hidden through sale (or misuse) of personal information. Conversely there is a good case to be made for Government to provide a universal interoperable platform for age verification based upon its essential national records which are **independent** of any third-party sponsor (be it Facebook, Yoti, Google, et al) or any captured *"charity"*.

I understand that in Local Authorities in England and Wales since 2008 for births and deaths and since 2011 for marriages, the country (England and Wales) has run a dual system. There is a paper copy of everything which, in law, is the formal register for which the County Council is responsible. There is also a centralised online system which is held at the General Registrar Office for England and Wales. This central system is called RON (Registration on Line). The input into RON and the data security issues around the paper copies lies exclusively with local authorities. There is significant variation between local authorities in how that is managed. For instance, I understand that Staffordshire County Council is unique in having all their paper documents uploaded to RON going as far back as 1837. Other councils are not quite as diligent or up to date.

Currently verification of a birth is through an NHS system which the midwives input on the birth of a child. The Council Registrars can verify that the individual making the application

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?

for registration is *"bone fide"* by mapping what they are being told against the NHS database. The details on a birth certificate plus the NHS number go onto RON. A birth certificate is then issued to the applicant.

I am of the view that the County Council system which is centralised and online and safe and well managed forms the basis of the best possible age assurance and age verification system which could be deployed in the United Kingdom.

Ofcom should require that this existing and working system forms the basis of a universal age gating, age assurance and age verification system in the UK – in preference to any "*special pleading*" from technology companies and the like. The existing County Council system can easily be enhanced so that an applicant could phone in for age gating, age assurance and age verification rather than appear in person. Online applications are a clear and low-cost enhancement to this service which needs to be delivered by accountable Local Government rather than being hived off to some unaccountable and secret VSP service provider - no matter how well connected that VSP is to the UK charity sector and with friends in Westminster and Whitehall.

An early pilot local authority programme to examine and test the roll out Age Gating is therefore recommended – a programme which could then be scaled-up to a national solution.

Additionally, such a system could also become an exemplar for the Republic of Ireland whose equivalent regulatory authority, the soon to be established *Coimisiún na Meán* (being the body responsible for overseeing the regulation of broadcasting and video-on-demand services and introducing the new regulatory framework for online safety, implementing the revised Audiovisual Media Services Directive into Irish law and supporting the development of the wider media sector in Ireland under their Online Safety and Media Regulation (OSMR) Act 2022) will share the identical concerns as those of Ofcom.

Ofcom and *Coimisiún na Meán* working together could give the British Isles significant benefits arising from economies of scale in sharing UK and Irish child protection services.

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

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?

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise. SafeCast would nevertheless recommend that Ofcom considers Nicholas Kristof's feature in the 2020 New York Times article – The Children of Pornhub when considering whether presence of content is harmful to children on general search services.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise. SafeCast has suggested areas for further research and recommends that Ofcom further engages with various OSTIA members who are likely to have confidential access to this evidence.

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

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise. However Ofcom is undoubtedly aware that currently the US Supreme Court is considering the "Gonzalez v Google LLC" case. Nohemi Gonzalez was one of the 129 people killed by ISIS in coordinated attacks in Paris in 2015. The family wants Section 230 of the 1996 Communications Act to be revoked. This section is described as giving "sweeping immunity" to internet providers from liability for harmful content posted by their users. The family believe that the internet provider is, by default, a "recruiting sergeant" for terrorists. However, section 230 is also credited with allowing the internet to flourish without the fear of litigation for users posts.

Ofcom will also be aware that British and American companies insure their directors against litigation both in defending the case and any damages which may ensue. It would likely therefore be the case that as robust sanctions against directors of tech companies have won the day in Westminster and, most importantly, if the US Supreme Court find in favour of the Gonzalez family and revoke section 230, then that will be an issue for the insurance companies who may choose to "rescind" (void, essentially exclude) policies for internet providers who engage in their current behaviour. They will likely insist that certain standards be maintained in order for their directors to be assured of cover.

Consequently, in colloquial terms, Ofcom regulatory function should be to *"follow the money"* via insurance coverage in persuading directors of British and American companies to effectively assess the risks of harming children from content on their systems.

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

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?

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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)?

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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 SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

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

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise. However SafeCast would recommend that Ofcom engages with ODI Fellow Georgia Meyer who is a recently appointed Research Fellow at the Open Data Institute (ODI) and as an MPhil/PhD student (Information Systems) at the London School of Economics and Political Science (LSE), supervised by Dr Edgar Whitley.

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 SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise. However SafeCast would recommend that Ofcom engages with ODI Fellow Georgia Meyer who is a recently appointed Research Fellow at the Open Data Institute (ODI) and as an MPhil/PhD student (Information Systems) at the London School of Economics and Political Science (LSE), supervised by Dr Edgar Whitley.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise. However SafeCast would recommend that Ofcom engages with ODI Fellow Georgia Meyer who is a recently appointed Research Fellow at the Open Data Institute (ODI) and as an MPhil/PhD student (Information Systems) at the London School of Economics and Political Science (LSE), supervised by Dr Edgar Whitley.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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 SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

Is this a confidential response? (select as appropriate)

No

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

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?

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

Is this a confidential response? (select as appropriate)

No

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

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?

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.

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

The SafeCast response to the **Second Consultation on the Online Safety Bill** is principally about the merits of a global standards based system led by the UK to protect children and the vulnerable whilst preserving free speech and adopting an age gating system of age authentication to protect the identity of children and the abuse of their rights.. The global standards based system is grounded in the regulatory and legal concept of *"digital sovereignty"* which SafeCast has outlined to Ofcom and the then DCMS in earlier consultations. Our evidence in this Second "call for evidence" is given under Questions 2 to 5 inclusive, in their associated weblinks and in SafeCast response to Ofcom's First Consultation on the Online Safety Bill

This particular question is not within SafeCast's area of expertise.