A1. Call for Evidence questions

Part A:

Questions for industry Services

> *Question 1: Are you providing a UK-established service that is likely to meet the AVMSD definition of a VSP? Please provide details of the service where relevant. The establishment criteria under the AVMSD are set out in annex 5.*

SafeCast's response to Question 1 - No, SafeCast is not a VSP but SafeCast's technology is directly appropriate for VSPs, hence we have included responses where these are relevant.

> *Question 2: Is your service able to identify users based in specific countries and do you provide customised User Interfaces (UI), User Experience (UX) functionality or interaction based on perceived age and location of users? Terms and conditions (Ts&Cs)*

SafeCast's response to Question 2 - SafeCast is able to assist VSPs in complying with all the above mentioned requirements through its metadata labelling technology when embodied in a revision to the international video broadcasting standard TSP 2121.

Question 3: How does your service develop and enforce policies for what is and is not acceptable on your service? (including through Ts&Cs, community standards, and acceptable use policies) In particular, please provide information explaining • what these policies are and whether they cover the categories of harm listed in the AVMSD (protection of minors, incitement to hatred and violence, and content constituting a criminal offence – specifically Child Sexual Exploitation and Abuse, terrorist material, racism and xenophobia); • how your service assesses the risk of harm to its users; • how users of the service are made aware of Ts&Cs and acceptable use policies; and • how you test user awareness and engagement with Ts&Cs.

SafeCast's response to Question 3 - SafeCast can assist VSPs in developing and enforcing policies, without censorship, in respect of what is and is not acceptable on their services through its metadata labelling technology. (See SafeCast's Answer to Question 19)

Question 4: How are your Ts&Cs (or community standards/ acceptable use policies) implemented? In particular, please provide information explaining: • what systems are in place to identify harmful content or content that may breach your standards and whether these operate on a proactive (e.g. active monitoring of content) or reactive (e.g. in response to reports or flags) basis; • the role of human and automated processes and content moderation systems; and • how you assess the effectiveness and impact of these mechanisms/ processes.

SafeCast's response to Question 4 - SafeCast enables VSPs to be transparently proactive in identifying harms; adhering to their Ts&Cs; adhering to their community standards; and maintaining their acceptable use policies. All of these effective measures are produced through the insertion of SafeCast Headcodes in metadata and automatic filtering of inappropriate content without censorship. (See SafeCast's Answer to Question 19 generally and particularly in respect of *click wrap* agreements)

Advertising A1.5

Please note that advertising includes sponsorship, product placement, endorsements and other types of audiovisual commercial communications.

Question 5: Does your service have advertising rules? In particular, please provide information about any advertising rules your platform has, whether they cover the areas in the AVMS Directive, and how these are enforced. See Annex 5 for a copy of the AVMSD provisions.

SafeCast's response to Question 5 - SafeCast is able to assist VSPs in complying with the detailed provisions of Article 28 in respect of national, regional and international advertising rules (including the CAP Code) through its metadata labelling technology together with its proposed amendment to TSP 2121. (See SafeCast's Answer to Question 23)

Question 6: How far is advertising that appears on your service under your direct control, i.e. marketed, sold or arranged by the platform? Please provide details of how advertising is marketed, sold and arranged to illustrate your answer.

SafeCast's response to Question 6 - SafeCast is a metadata labelling technology that can enable infrastructure companies (e.g. Adstream and Peach Media) to implement metadata labelling of advertisements so that their content and the content from their clients (including VSPs) can automatically be filtered to comply with Ofcom and ASA regulations made under the Digital Economy Act 2017 and ASA regulations and regulations made by the ICO in their Age Appropriate Design Code pursuant to Section 123 of the Data Protection Act 2018. As such, SafeCast does not have any direct control as this is done by the advertisers inserting appropriate SafeCast Headcodes in line with the CAP Code and ASA standards. However, SafeCast is concerned that the advertiser-funded ISBA Programmatic Supply Chain Transparency Study[1], which reported in May 2020 following two years of work, found that only a fraction (12%) of 267 million ad impressions, paid for by brands to be served on publishers' websites, could be accounted for or "*matched*" and nearly one in six (15%) of advertising pounds spent on these ads are being lost in the system. This "*unknown delta*" represents money that was spent on programmatic ads but could not be found to have been delivered as an ad that any internet user had actually viewed. SafeCast's view is that all advertising must be trackable and auditable and this can be best achieved through the use of embedded metadata labelling using unique identifiers such as the US Ad-ID[2] commercial service or possibly an equivalent UK or EU service. (Additionally, see SafeCast's Answer to Question 21 and the section headed "**VSPs fail to protect children from commercial exploitation on their platforms"**)

Question 7: What mechanisms do you have in place to establish whether videos uploaded by users contain advertising, and how are these mechanisms designed, enforced, and assessed for effectiveness?

[1] https://www.isba.org.uk/knowledge/digital-media/programmatic-supply-chain-transparency-study

[2] https://www.ad-id.org/

SafeCast's response to Question 7 - SafeCast has a self applied ratings system which, when implemented, labels advertisements as being advertisements thereby enabling automated compliance with Ofcom and ASA regulations in respect of, inter alia, advertising to children. Thus, when implemented, it will enable VSP platforms to comply with the CAP Code. Particulars of its self applied contents rating system are set out in a SafeCast document on the ICO website[3] which exhibits a redacted copy of SafeCast's letter of 25 April 2018 to Facebook[4] and the Entertainment Identifier Registry Association (EIDR) draft Best Practice of 19 March 2018[5]. (See also SafeCast's Answer to Question 23)

Reporting and flagging

Question 8: Does your service have any reporting or flagging mechanisms in place (human or automated)? In particular, please provide information explaining: • what the mechanisms entail and how they are designed; • how users are made aware of reporting and flagging mechanisms; 7 • how you test user awareness and engagement with these mechanisms; • how these mechanisms lead to further action, and what are the set of actions taken based on the reported harm; • how services check that any action taken is proportionate and takes into account Article 10 of the European Convention of Human Rights (freedom of expression); • how users (and content creators) are informed as to whether any action has been taken as a result of material they or others have reported or flagged; • whether there is any mechanism for users (including uploaders) to dispute the outcome of any decision regarding content that has been reported or flagged; and • any relevant statistics in relation to internal or external KPIs or targets for response.

[3] See on the ICO website SafeCast's response to the ICO's Age Appropriate Design Code call for evidence dated 17 September 2018 at SafeCast ICO submission

[4] See redacted letter on the ICO website SafeCast to Facebook -25 April 2018

[5] See redacted document on the ICO website Entertainment Identifier Registry Association (EIDR) draft Best Practice of 19 March 2018

SafeCast's response to Question 8 - SafeCast is able to assist VSPs in complying with reporting and flagging mechanisms, without censorship and in full compliance with Article 10 of the European Convention of Human Rights, through its metadata labelling technology. When deployed, SafeCast expects that uploaders will apply the SafeCast Headcodes themselves or will be assisted by third party businesses to aid uploaders who are unsure of which SafeCast Headcode to assign to their video. Through its deployment, SafeCast will also be capable of producing comparable "big data" that can be used by Artificial Intelligence (AI) which can transparently determine whether content should be reported or flagged. (Additionally, see SafeCast's Answer to Question 19 in respect of "*content restrictions and filtering*" and "*easy to access complaints functions*")

Content rating

Question 9: Does your service allow users to rate different types of content on your platform? Please provide details of any rating system and what happens as a result of viewer ratings.

SafeCast's response to Question 9 - SafeCast's technology has seven classifications to effectively filter the entire spectrum of content as explained in the Demonstrator on the SafeCast website[6]. Using international metadata standards, these seven codes can be used flexibly by different nation states to take account of national mores and digital sovereignty requirements in unique edits of a video (or video advertisement) and in accordance with an ISO standard (ISO 26324).

[6] Labelling Demonstrator – Safecast, protecting children and vulnerable people

Age assurance

Question 10: Does your service use any age assurance or age verification tools or related technologies to verify the age of users? In particular, please provide information explaining: • how your age assurance policies have been developed and what age group(s) they are intended to protect; • how these are implemented and enforced; • how these are assessed for effectiveness or impact; and • if the service is tailored to meet age-appropriate needs (for example, by restricting specific content to specific users), how this works.

SafeCast's response to Question 10 - SafeCast's technology is able to assist VSPs in obviating the need for proprietary age assurance and age verification tools. Its classifications, termed the SafeCast HeadCodes, can be mapped onto the Key Stages of the National Curriculum in the UK and elsewhere (e.g. Republic of Ireland and Australia), thereby facilitating Age Gating through the use of generic school-age tokens on mobile devices. This effective child protection approach to Age Gating could be adopted by other countries via their respective telecommunications regulators without the need for new legislation.

Parental controls

Question 11: Does your service have any parental control mechanisms in place? In particular, please provide information explaining: • how these tools have been developed; • what restrictions they allow; • how widely they are used; and • how users of the service, and parents/ guardians if not users themselves, are made aware of and encouraged to use the parental control mechanisms that are available.

SafeCast's response to Question 11 - SafeCast is able to assist VSPs in obviating the need for proprietary parental control mechanisms. Its classifications, termed the SafeCast HeadCodes, can be mapped onto the Key Stages of the National Curriculum in the UK and elsewhere. In the UK this implementation can support Department of Health and Department of Education control mechanisms on mobile devices used by children as a universal form of parental control that complies with Ofcom's core principles. (See also SafeCast's Answer to Question 19 and in particular the section on *parental controls*)

.

Complaints handling

Question 12: Does your service have a complaints mechanism in place? Please describe this, including how users of your service can access it and what types of complaint they can make. In particular, please provide information explaining: • any time limits for dealing with complaints; • how complainants are informed about the outcomes of complaints; • any appeals processes, how they work, and whether they are independent from the complaints processes; and • the proportion of complaints which get disputed or appealed.

SafeCast's response to Question 12 - In view of the fact SafeCast is not a VSP, this question is not directly relevant. However, SafeCast has made recommendations on this issue in its answer to Question 25.

Media literacy

Question 13: What media literacy tools and measures are available on your service? In particular, please provide any relevant information about: • how you raise awareness of media literacy tools and measures on your service; • how you assess the effectiveness of any media literacy tools and measures provided on your service; and • how media literacy considerations, such as your users' ability to understand and respond to the content available to them feature in your thinking about how you design and deliver your services, for example in the user interfaces, flagging content and use of nudges.

SafeCast's response to Question 13 Since SafeCast is not a VSP this is not directly relevant. SafeCast has made recommendations on this issue in its answer to Question 22.

Reporting and evaluation

Question 14: Do you publish transparency reports with information about user safety metrics? Please provide any specific evidence and examples of reports, information around the categorisation and measurements used for internal and external reporting purposes, and whether you have measures in place to report at country/ regional level and track performance over time.

SafeCast's response to Question 14 - SafeCast does not provide reporting facilities itself since the filtering of content will be transparent within international video standards but has made recommendations on this issue in its answers to Questions 19, 20, 21, 24 and 26. SafeCast also has a '*kite*' mark i.e. the registered trademark '*Safecheck*' which, when used by a VSP under the terms of a Registered User Agreement[7], will inform the public at large that the appropriate SafeCast HeadCodes have been embedded in its video content in accordance with applicable country/regional content level standards laid down by relevant telecommunications authorities and advertising regulators.

[7] Registered User Agreements are filed at the UK IPO under the respective trade mark registration using this form https://www.gov.uk/government/publications/application-to-record-a-licensee

Question 15: What processes and procedures do you have in place to measure the impact and effectiveness of safety tools or protection measures? If not already captured elsewhere in your response, please provide information relevant to all of the measures listed above explaining: • how you test and review user awareness and engagement with each measure (including any analysis or research that you would be willing to share with Ofcom) • how often policies and protection measures are reviewed, and what triggers a review; and • how you test the impact of policies on users and the business more generally, such as how you balance the costs and benefits of new tools.

SafeCast's response to Question 15 SafeCast has made suggestions and recommendations on all these issues in its answers to Questions 19, 20, 21, 24 and 26.

Question 16: How do you assess and mitigate the risk of inadvertent removal of legal or non-harmful content? In particular, please provide any information on: • how freedom of expression is taken into account during this assessment; • how appeals are handled and what proportion are successful; and • audits of automated removal systems and, if you have them, any metrics that relate to their effectiveness.

SafeCast's response to Question 16 - SafeCast has made recommendations on all these issues in its answers to Questions 19, 20 and 25.

Question 17: Have you previously implemented any measures which have fallen short of expectations and what was your response to this? Please provide evidence to support your answer wherever possible.

SafeCast's response to Question 17 - Not Applicable

Question 18: How does your service develop expertise and train staff around different types of harm? (e.g. do you have any partnerships in place?)

SafeCast's response to Question 18 - Through the use and deployment of the '*Safecheck*' trademark as a '*kite*' mark, VSPs (under the terms of the Registered User Agreement[8]) could be required to establish partnerships involving the development of expertise and the training of staff to ensure that the appropriate SafeCast HeadCodes have been embedded in a VSP's video content in accordance with applicable country/regional content level standards laid down by relevant telecommunications authorities and advertising regulators.

[8] See also SafeCast's response to Question 14

Part B.

Questions for all stakeholders A1.6 The following questions are intended for industry, academics, experts, and consumer groups, as well as any other interested stakeholders.

> ***Question 19:*** *What examples are there of effective use and implementation of any of the measures listed in article 28(b)(3) the AVMSD 2018? The measures are*
>
> - *terms and conditions,*
> - *flagging and reporting mechanisms,*
> - *age verification systems,*
> - *rating systems,*
> - *parental control systems,*
> - *easy-to-access complaints functions, and*
> - *the provision of media literacy measures and tools.*
>
> *Please provide evidence and specific examples to support your answer.*

SafeCast's Answer to Question 19

Since click-wrap (or their predecessor "*shink-wrap*") agreements came to be used in the licensing of dematerialised goods and services around forty years ago, all the reported case law involving "*click-wrap*" terms and conditions has been between business users and companies invoking or relying upon the terms and conditions. Consequently, a general understanding has emerged that nobody, other than academics and lawyers, actually reads the terms and conditions in a click-wrap agreement. As a matter of legal practice, every click-wrap agreement must be construed upon its reasonableness in the circumstances of each case. This requirement undermines the utility of click-wrap agreements because they create a lack of certainty in the law to the detriment of all parties involved. Nobody can be sure how a click-wrap agreement will be construed by the courts until it is litigated and then the decision will only apply to that exact set of terms and circumstances.

SafeCast proposes that, in accordance with Ofcom's approach to VSP regulation based around the core principles: *protection and assurance; freedom of expression; adaptability; transparency; enforcement; independence; and proportionality,* all click-wrap terms and

conditions used by VSPs should contain a clause which expressly makes Ofcom a party to the agreement under the *Contracts (Rights of Third Parties) Act 1999*. Under this clause Ofcom would be granted the right to be notified of all cases when click-wrap terms and conditions are invoked and the right to intervene should Ofcom wish so to do.

The right to be notified and the opportunity to intervene, if Ofcom so wished, would be a *proportionate* check upon misuse of contractual terms and conditions by VSPs against consumers, particularly children and parents of young children. As a means of promoting best practice; inhibiting abusive behaviour and establishing fair dealing on VSP platforms, Ofcom's third party contractual rights would be a form of light touch regulation with minimal cost burdens on all parties. Ofcom would not be required to intervene in any dispute under the terms and conditions but, at its sole discretion, could choose to do so.

Such a contractual term, making Ofcom a party to the contract, would also address the question of the suitability and appropriateness of a VSP's *flagging and reporting mechanisms* because all such mechanisms and their operations would be referred to in the terms and conditions of service and would fall within the scope of review and interpretation of the agreement. Ofcom would, at its sole discretion, be able to review and comment upon any VSPs *flagging and reporting mechanisms* in furtherance of Ofcom's core principles of protection & assurance and freedom of expression. This again would be a *proportionate* means of regulating VSPs.

Age Verification Systems

The COVID-19 lockdown has highlighted the urgent need for some form of Age Gating on systems which are used by children. However, SafeCast's view is that age verification systems based upon individual attributes of a user are not an appropriate and proportionate response to the harms they seek to eliminate. This is for the following reasons:

1. Age verification systems are not a structural part of the internet. Thus restricting access to specific age groups is not the *default* in its current implementation.
2. Age verification enrolment systems, which are based upon the exact age of an internet user, automatically give rise to privacy risks which can lead to stalking, grooming and bullying. Safe use of these systems requires additional controls and measures which may not always be available. Thus the trade-off between the design of the internet being open to all militates against the use of exact age systems in its current implementation. For example, if Facebook or YouTube were to establish an

age verification enrolment system based upon the exact age of a Facebook or YouTube user, this could be used for the commercial benefit of Facebook or YouTube respectively and their "*walled gardens*" of commercial services.

3. Unlike some EU countries, the UK does not have a centralised digitally accessible register of births and deaths, In consequence, any age verification enrolment system for UK children based upon their exact age will be a proprietary age verification system. Proprietary systems can become non-tariff barriers to new competitors wishing to enter the market.

Rather than requiring age verification systems based upon the actual age of a child to be used to support Age Gating of content on VSP systems, Ofcom's long experience in maintaining the Television Watershed restrictions on regulated television services based on children's age range and time that programmes are shown, suggests a better way of addressing the need for Age Gating of content without giving rise to new privacy risks. Following a revision of the Ofcom Broadcasting Code to bring it into line with modern practices, Age Gating could be implemented on mobile devices and tablets using the school-age of a child.

Such a measure could be deployed by teachers and parents enrolling a child through the use of an anonymised token embedded in the phone or mobile device. A school age token could be generated and loaded as middleware on the child's mobile device following the completion of a secure webform by the child's parent or guardian or teacher. The school age token would be cryptographically signed with the date and time of its installation on the child's mobile device and this information would be logged. Primary schools, nurseries and public libraries would be able to enrol children of identified parents as well as parents and guardians directly from their homes through use of the Government's forthcoming Document Checking Service, which is to give people easier and safer access to digital services that require identity checks.

This change should be reflected in a revision to the Advertising Guidance given by CAP in relation to non-broadcast marketing communications. This would allow school age metrics and restrictions to be substituted for the highly subjective "*120 Index"* system, which is not suitable for use in multichannel non-linear broadcasting environments. Marketeers bear principal responsibility for the marketing communications they produce and must be able to prove the truth of their claims to the ASA; they have a duty to make their claims fair and honest and to avoid causing serious or widespread offence. Agencies have an obligation to create marketing communications that are accurate, ethical and neither mislead nor cause

serious or widespread offence. Publishers and media owners recognise that they should disseminate only those marketing communications that comply with the Code. That responsibility extends to any other agent involved in producing, placing or publishing marketing communications. They accept the rulings of the ASA Council as binding. School age metrics and restrictions should be deployed within a revision of Advertising Guidance.

Parental control systems

Parental control systems which are deployed on systems used by young children provide the young child with reassuring care and concern from their parents. However, with older children, parental controls can become stifling constraints on children's innate curiosity and development. There are two main types of parental controls and most systems provide both of them together:

- Controls which limit the usage time and which can invoke a shut-off time for the child's device
- Controls which restriction content viewable by a child

Usage time controls

Usage time controls are used by parents to ensure that their children do not spend too much time on the internet to the detriment of normal social activities or having sufficient sleep. Usage time controls can be presented to parents as a way in which they can prevent their children from developing an obsessive-compulsive disorder or an impulse control disorder termed *"internet addiction"*. The World Health Organisation (WHO) has suggested that *"internet addiction*" could be a form of addictive behaviour that might be harmful. Medical evidence from mental health practitioners in support of usage time controls is very limited - children tend to grow out of any obsessive compulsive behaviours that usage time controls are meant to address. In extreme cases behavioural problems appear to be resolvable by NHS supported psychological interventions and by cognitive behavioural therapy (CBT).[9]

One growing area of usage time control is computer use management which is focused on empowering parents to balance the computing environment for children by regulating gaming. The main idea of these applications is to allow parents and teachers to introduce a learning component into the computing time of children, who must earn gaming time while working through educational contents. This time management measure could be a fruitful

[9] See for example Pages 6-7 of Insight Issue 5 2018 - the journal of the Royal College of Psychiatrists

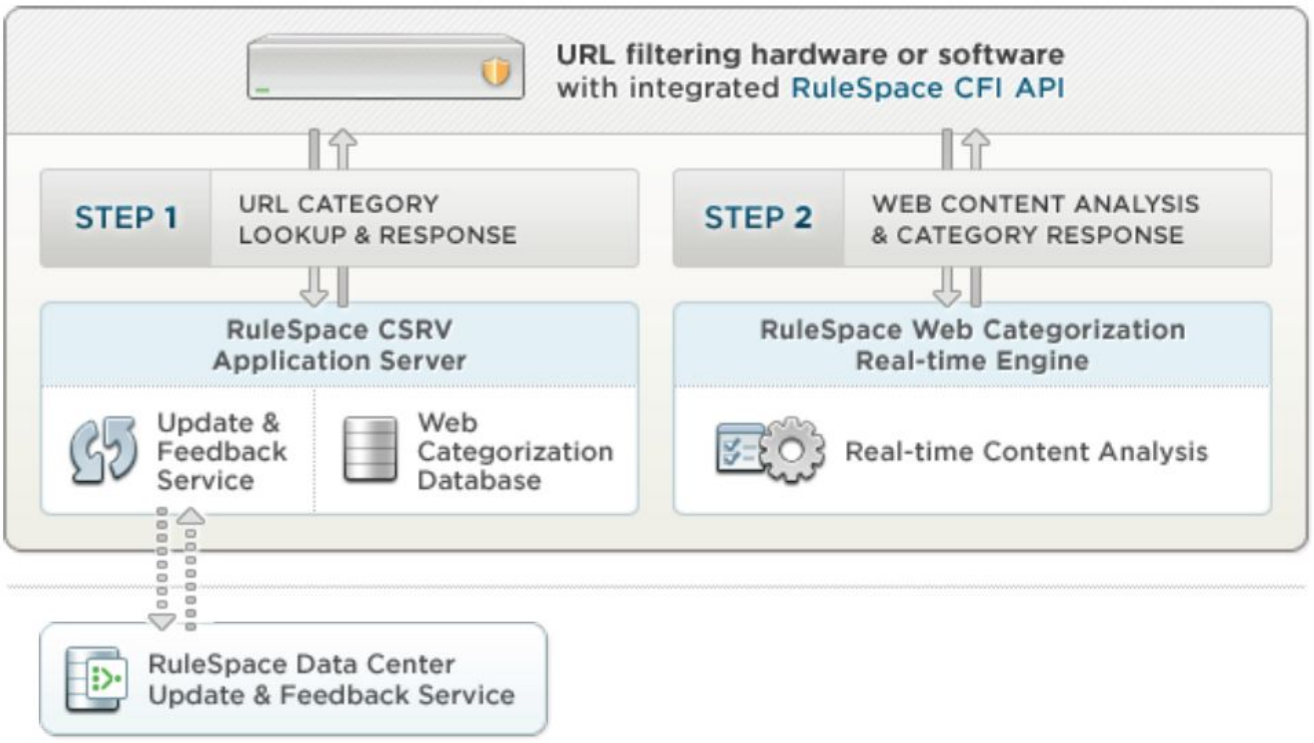
area of development in the years ahead. However, it would be a mistake to assume that educational learning and computer games are always conflicting pastimes. There is emerging support amongst mental health practitioners that the playing of specific computer games can be beneficial to mental health and development of children.[10]

Content restrictions and filtering

In the UK, content filtering for the protection of children on the internet is considered by parents, politicians, teachers and academics to be essential. Today it is deployed through privately administered ISP-level content filtering systems generically referred to as "*Cleanfeed*" and sometimes under various proprietary trade names (e.g. *BT WiFi Protect, Sky Broadband Shield* etc). Over 90% of households in the UK receive their domestic broadband service via one of the "*Big Four*" ISPs (BT, Sky, Virgin, TalkTalk). Each supplies customers with a set of family filters to restrict access to content considered unsuitable for children. These filters are widely used - the industry claims that 80% of users with "*Cleanfeed*" available use it.[11]

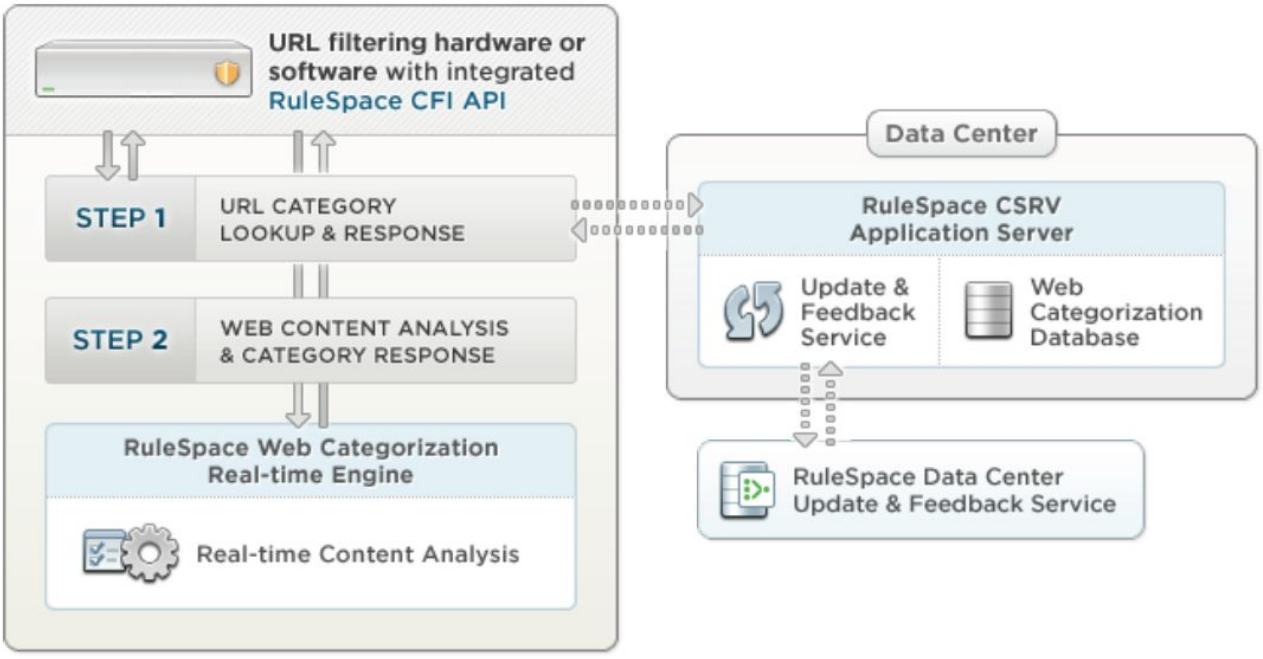
All the "*Cleanfeed*" systems in the UK would appear to be implementations of a technology called RuleSpace from Symantec which today is part of Broadcom. The OEM documentation of Symantec RuleSpace contains the following two diagrams and statements.[12]

[10] Ibid Insight Issue 5 2018 - the journal of the Royal College of Psychiatrists
[11] Note - the forthcoming rollout of encrypted DNS services threatens to undermine Cleanfeed in all its forms - the best solution would appear to be to pre-label and post-filter content for child protection purposes outside of the embedded encryption processes.
[12] See https://docs.broadcom.com/doc/ga-rulespace-en



STEP 1
URL category lookup from the RuleSpace categorization database(s)

STEP 2
For "unknown" or "dynamic content" URLs, perform real-time content analysis



STEP 1: URL category lookup from the RuleSpace categorization database(s)
STEP 2: For "unknown" or "dynamic content" URLs, perform real-time content analysis

> *"Symantec RuleSpace incorporates proven and globally respected URL Web categorization databases with real-time analysis from one of the pioneers in the industry with 18+ years of expertise delivering highly effective categorization solutions to multiple markets. RuleSpace offers a unique combination of database lookups and real-time analysis for accurate categorization of unknown or dynamic web sites, especially useful in dealing with sites containing user generated content like Facebook, Twitter and YouTube. Our business model is designed around providing superior service quality with a dedicated engineering and technical support team. Symantec's RuleSpace is powering dozens of 3rd party web filtering solutions serving over 350 million end-users worldwide, from subscribers at ISPs/mobile operators, to businesses and individual consumers."*

How RuleSpace works today is a proprietary secret. There is some documentation which is five years old that indicates, in broad terms, how RuleSpace then worked.[13] This contains a Question and Answer section that includes the following:

[13] https://docs.broadcom.com/doc/rulespace-en

How are URLs assigned to categories?

Both human analysis and machine algorithms are used for assigning classifications at a domain, host, path or page level as appropriate to the content.

- Specialists (graders) categorize URLs received through category dispute requests, unknown URLs showing high frequency numbers in our customers' traffic and during manual category maintenance.

- Automated systems use machine learning technology which assigns one or more categories to incoming unknown URLs

The last statement confirms that five years ago RuleSpace was using Artificial Intelligence to filter content. It can reasonably be assumed that it is doing the same today. Additionally, within the same Question and Answer section, is the following:

Do you support Whitelist / Blacklist?

Sites are categorized based on published categories and associated definitions. Our partners decide whether to allow/deny categories. Partners can also create their own pre-post lists which can be used for their own blacklist/whitelist purposes.

The above confirms that five years ago '*partners*' (as Broadcom calls its corporate clients) were able to decide which websites could get through the RuleSpace filtering. Without evidence to the contrary, it can reasonably be assumed that this remains RuleSpace's position today.

Another content filtering solution which is relied upon by millions of families all over the world is Google "*SafeSearch*". It is a feature, in Google Search and Google Images, that acts as an automated filter of pornography and potentially offensive and inappropriate content. How Google "*SafeSearch*" works is also a proprietary secret.

Both Google and Broadcom have chosen to use the law of trade secrets and confidence to protect their proprietary means of protecting children and vulnerable people. However their solutions appear to breach two of Ofcom's core principles: *freedom of expression and transparency*. Proprietary systems such as Google SafeSearch and Symantec RuleSpace do not explain how their AI-based decisions are made, what those decisions are based on, and why they were taken. They are not **transparent** in making any underlying values explicit

and they do not make their "*partners*" take responsibility for AI-based decisions made on their behalf. Neither Google SafeSearch nor Symantec RuleSpace produce reports for inclusion by their partners in their annual reports on how their services are functioning, what they are filtering away and what they are allowing through; thus they are restricting "**freedom of expression**"[14]. In short, neither Google SafeSearch nor Symantec RuleSpace are services which are auditable and reviewable by any government regulator concerned about censorship or abuse of market conditions or creation of non-tariff barriers to competition in the marketplace.

In this area there are also third party parental control services which allow parents to put further controls on top of content which may have passed through Google SafeSearch and Symantec RuleSpace filtering. These third party proprietary services (e.g. NetNanny[15]) use Artificial Intelligence to remove content which might not have been picked up by Google SafeSearch and Symantec RuleSpace filtering. They can also be used to totally control all functions on a child's mobile device including monitoring all communications via the device and location tracking (including geo-location gating). Whilst third party parental control services may be appropriate for the protection of young children, their use with older children can give rise to trust issues between parents and children that can lead to dysfunctional relationships and harms. UNICEF has stated in setting out its principles of the rights of a child that "*Children's privacy and freedom of expression should be protected and respected in accordance with their evolving capacities*". Third party parental control services can be deployed in ways which do not allow children to develop freedom of expression and might in due course breach the ICO Age Appropriate Design Code under Section 123 of the Data Protection Act 2018 when these come into force in September 2021.

Easy-to-access complaints functions

The best examples of complaints functions are those which are clear and concise; are accessible (which means that they explain how to complain in plain language with functional visuals); detail how the complaints will be handled and what timescales will be used as well as the complainants right to appeal if they do not agree with the result. They should also

[14] See also discussions on Google cloud services for use in "*sovereignty sensitive markets*" such as China and the EU reported at https://www.scmp.com/news/world/united-states-canada/article/3092416/google-scraps-cloud-initiative-china-and-other

[15] https://www.netnanny.com/

contain tracking information - a ticket based system as used in software support can often be a form of best practice.

Problems can arise when there are failings in any of these categories, particularly when complaints are not tracked. The tracking of complaints needs to be audited and reported upon. Good tracking of complaints improves users willingness to engage with the company and can improve the brand experience in commercial situations.

Difficulties can arise when the user is making a complaint against an automated decision. One frequent issue on YouTube arises because the service's Content ID system will automatically try to figure out if there is any copyrighted music or video in it. Artists and labels can choose to then either mute that audio; block the video from being seen; or monetise the video by running ads against it. Aggressive copyright claims from record companies and movie producers can cause user-created content to be automatically taken down when in fact the use by the uploader should be lawful, by reason of the fact that it is fair use, parody or criticism - all of which are valid defences to infringement of copyright. YouTube addresses this issue reasonably well with a policy of "*copyright strikes*" (three copyright strikes and the user is barred). However problems can arise in appealing a decision using their '*counter notification*' process where the user is able to raise fair use and other defences. YouTube states in its advice to users who wish to file a counter notification:

> If your counter notification meets all legal requirements and has a clear explanation of how you have the rights to the removed content, we'll forward it to the claimant. The claimant will have 10 business days to provide evidence that they've initiated a court action to keep the content down. This time period is a requirement of copyright law, so please be patient.

The above-mentioned YouTube procedure appears to be biased in favour of claimant copyright holders against uploaders. Furthermore it can be argued that the '*counter notification*' process might in due course breach the ICO Age Appropriate Design Code under Section 123 of the Data Protection Act 2018 when these come into force in September 2021, since it is not written in terms which would be understood by a user who is an adolescent.

Provision of media literacy measures and tools

The ability to access, analyse, evaluate, create, and act using all forms of communication is a key development skill and VSPs have been particularly successful in developing tools to

allow children and young adults to make their own video and audio content without the need for professional (or teacher) help and assistance. Indeed it can be said that the competition within the VSP marketplace has led to the development of tools that are extremely easy to use and are free. A highly competitive market has led to short videos made by children reaching millions of viewers. However, there are concerns that there can be exploitation of young creators who use these tools to make content and then do not benefit from the commercial exploitation of their creative works. (Further discussion on this point is set out in SafeCast's Answer to Question 21 under "*VSPs fail to protect children from commercial exploitation on their platforms*".

Question 20: What examples are there of measures which have fallen short of expectations regarding users' protection and why? Please provide evidence to support your answer wherever possible.

SafeCast's Answer to Question 20

- **Click-wrap agreements** which create a lack of certainty in the law to the detriment of all parties involved. Nobody can be sure how a click-wrap agreement will be construed by the courts until it is litigated and then the decision will only apply to that exact set of terms and circumstances. There might be a case for model forms of "*click-wrap*" agreements which are approved by Ofcom for use by VSPs. These could require different versions for children and adults.

- **Age verification systems** are not a structural part of the internet. Thus restricting access to specific age groups is not the default in its current implementation. This is not something that can be easily solved because, unlike many countries, the UK does not have a centralised digitally accessible register of births and deaths. Therefore age verification enrolment systems based upon the exact age of an internet user will be proprietary age verification systems, which could become non-tariff barriers to new competitors wishing to enter the market.

- **Parental control systems** can become stifling constraints on children's innate curiosity and development. Proprietary systems such as Google SafeSearch and

Symantec RuleSpace do not explain how their AI-based decisions are made, what those decisions are based on, and how they function. They are not **transparent** in making any underlying values explicit and they do not make their "*partners*" (as Broadcom and Google calls their corporate clients - see our answer to Question 19) take responsibility for AI-based decisions made on their behalf. Neither Google SafeSearch nor Symantec RuleSpace produce reports for inclusion by their partners in their annual reports on how their services are functioning, what they are filtering away and what they are allowing through; thus they are restricting "**freedom of expression**". In effect, neither Google SafeSearch nor Symantec RuleSpace are services which are auditable and reviewable by any government regulator concerned about censorship or abuse of market conditions or creation of non-tariff barriers to competition in the marketplace.

- Difficulties can arise when the user is making a **complaint against an automated decision**. YouTube's "*counter notification*" procedure (as set out in SafeCast's answer to Question 19) appears to be biased in favour of claimant copyright holders against uploaders. Furthermore, it can be argued that the '*counter notification*' process might in due course breach the ICO Age Appropriate Design Code under Section 123 of the Data Protection Act 2018 when these come into force in September 2021, in that it is not written in terms which would be understood by a user who is an adolescent.

Question 21: What indicators of potential harm should Ofcom be aware of as part of its ongoing monitoring and compliance activities on VSP services? Please provide evidence to support your answer wherever possible.

SafeCast's Answer to Question 21

SafeCast considers that Ofcom should be aware of the following indicators of potential harm and, as part of its ongoing monitoring and compliance activities, could reasonably require VSPs to address them:

- **VSPs fail to adequately police sexual abuse imagery** on their platforms and fail to cooperate sufficiently with the authorities when they find it. On 29 September 2019[16] The New York Times published an in-depth analysis which found an insatiable criminal underworld that had exploited the flawed and insufficient efforts to contain it. SafeCast is of the opinion that the best lightweight regulatory measure to encourage VSPs to devote serious resources to addressing and eliminating these harms is for Ofcom to be mandated as a third party to all click-wrap agreements made between VSPs and their users. SafeCast also considers that Ofcom could require VSPs to produce quarterly audit reports to Ofcom on the the incidents of sexual abuse imagery on their platforms and the measures they have taken to eliminate and remove the same.

- **VSPs fail to protect children from commercial exploitation on their platforms.** The Wall Street Journal on 11th April 2019[17] published an article about the 'lack of accountability' of children's content on YouTube *"The most popular YouTube channel for children features a video of animated toddlers in a bubble bath singing along to the "Baby Shark" melody. The video, on the CoCoMelon channel, has been viewed 1.4 billion times and has generated millions of dollars in advertising revenue*." A further example of this, which was cited in respect of Instagram and YouTube's disruption of child labour laws, was published in the Media Guardian on 24th April 2019[18]. As an express condition of licensing by Ofcom, VSPs could be required to produce annual audit reports to Ofcom on the accounting of advertising earnings generated by children on their platforms. Furthermore, to assist in this process and to ensure that the video content and earnings from advertisements are auditable, all video content on VSPs should contain information on ownership, provenance and product placement/advertising embedded as metadata within the video. This information can easily be included in the video content when it is uploaded to a VSPs platform. Such a measure would enable the equitable remedy of tracing to be used to track and account for earnings by children on VSPs. Such a measure should not require new legislation because there is already the power under English law for a court to order the setting aside of a percentage of a child performer's earnings in a blocked trust where it would remain untouched until the child reaches adulthood. Additionally, Section 104 of the Digital Economy Act 2017 allows ISS providers (and

[16] New York Times - 29 September 2019, The Internet Is Overrun With Images of Child Sexual Abuse. What Went Wrong?

[17] *"Kids love these YouTube channels. Who Created Them is a Mystery …"*

[18] **"***'Kidfluencers' are earning millions on social media, but who owns that money?***"**

hence VSPs) to filter content for child protection purposes. The inclusion of provenance, ownership information and product placement information within standard metadata labelling in a video would facilitate the removal of a video through the use of lightweight filters without censorship and should, therefore, be considered to be in compliance with the provisions of the ICO Age Appropriate Design Code which will come into effect in September 2021. Furthermore SafeCast also brings to Ofcom's attention the ISBA Programmatic Supply Chain Transparency Study[19], which reported in May 2020 following two years of work. This found that only a fraction (12%) of 267 million ad impressions, paid for by brands to be served on publishers' websites, could be accounted for or "*matched*" and nearly one in six (15%) of advertising pounds spent on these ads are being lost in the system. This "*unknown delta*" represents money that was spent on programmatic ads but could not be found to have been delivered as an ad that an internet user had actually viewed. SafeCast is therefore recommending that all advertising must be trackable and auditable and this can be best achieved through the use of embedded metadata labelling using unique identifiers such as the US Ad-ID[20] commercial service or possibly an equivalent UK or EU service.

Question 22: The AVMSD 2018 requires VSPs to take appropriate measures to protect minors from content which 'may impair their physical, mental or moral development'. Which types of content do you consider relevant under this? Which measures do you consider most appropriate to protect minors? Please provide evidence to support your answer wherever possible, including any age related considerations.

SafeCast's Answer to Question 22

Evidence and research that exposure to large amounts of graphic content and pornography amongst teenagers is increasing violence against women by normalising what would otherwise be aberrant behaviour is well documented in extensive literature. There is a

[19] https://www.isba.org.uk/knowledge/digital-media/programmatic-supply-chain-transparency-study

[20] https://www.ad-id.org/

useful introduction to the topic of normalisation of aberrant behaviour which is set out in a BBC Future website.[21]

SafeCast considers that the most effective and unobtrusive measure to address this issue is through self-applied content labelling on uploading and lightweight filtering so that content can be filtered away from children whose age and maturity make it unsuitable for them to view it.

The Department of Education publishes comprehensive statutory government guidance **Keeping Children Safe in Education** (often abbreviated to KCSiE). The latest version of the KCSiE document came into effect in September 2020 and reflects safeguarding changes and advancements. Part 1 plus the school's behaviour policy must be read by all staff. This includes teachers, headteachers, and everyone else working in an education environment (including admin and support staff) as well as governing bodies, proprietors, and management committees. All schools and colleges have a legal duty to ensure all staff have read the new guidance before the start of the academic year.

According to Cybersafe Scotland, internet predators have targeted a child in '*every class in every school in Scotland.*'[22] The National Curriculum for Computing includes elements of e-safety and school age children should receive information on the importance of keeping personal information (including images) safe; how to use technology responsibly and respectfully; how to recognise unacceptable behaviour and know how to get help in case of a problem arising. Many schools choose to give more information.

SafeCast considers that VSPs should take a proactive approach to self-applied content labelling on uploading and should work with Ofcom and the Department of Education in ensuring that everyone who uploads content is able to classify it appropriately and has access to services which enable them to do so.

[21] https://www.bbc.com/future/article/20170314-how-do-we-determine-when-a-behaviour-is-normal

[22] https://www.dailyrecord.co.uk/news/scottish-news/online-predators-target-scots-kids-15002592

Question 23: What challenges might VSP providers face in the practical and proportionate adoption of measures that Ofcom should be aware of? We would be particularly interested in your reasoning of the factors relevant to the assessment of practicality and proportionality.

SafeCast's Answer to Question 23

SafeCast considers that the best practical and proportionate measure to enable VSPs to comply with Ofcom's core principles would be for VSPs to support the amendment of the TSP 2121[23] technical specification so that self-applied content labelling on uploading and lightweight filtering could be deployed on the internet, without censorship. Such a change would be straightforward to implement and is already in working draft format. The Entertainment Identifier Registry Association (EIDR) is a not-for-profit industry association that was founded by MovieLabs, CableLabs, Comcast and TiVo to meet a crucial need across the entertainment supply chain for universal identifiers for a broad array of audio visual objects. Metadata labelling for child protection can be included within it without any difficulties. On 19 March 2018 a revision of the EIDR best practice for registering EIDR IDs was produced by the Technical Director of EIDR. This version[24] contains the following draft provision for the inclusion of Self-applied content ratings:

> 2.3.3 Add Self-Applied Ratings Data to the EIDR Edit Record
>
> Self-applied ratings systems generally lack a central registry of all rated works, so the necessary ratings information is recorded in the EIDR Edit record and a flag is added referring back to the applicable ratings authority.
>
> - Add an Edit Details field that contains the self-applied rating information.
>   - Set the value according to the instructions of the ratings authority.
>   - Set domain for the ratings authority (e.g., "safecast.co.uk").

[23] Full title SMPTE TSP 2121-1:2018 IMF Application DPP (ProRes) a standard which is jointly managed by the DPP in the UK and SMPTE in the USA

[24] On the ICO website at https://ico.org.uk/media/about-the-ico/consultation-responses/2018/age-appropriate-design-code-responses/2260237/safecast-additional-attachment-2.pdf

- Add "CA:[ratings authority domain];" (e.g., "CA:safecast.co.uk;") to Registrant Extra to indicate that the Edit Details contains a Content Advisory.

- The party responsible for the self-applied rating is strongly encouraged to add their internal/house ID for this version of the work as an Alternate Identifier to the EIDR Edit record

The 19 March 2019 draft has not yet been implemented by EIDR - the current draft on the EIDR website is dated 24th March 2014[25] and makes no reference to self-applied content ratings.[26] The EIDR ID is an implementation of an ISO standard (ISO 26324) and hence there would be no technical problems associated with this change being required by Ofcom for child protection purposes so that all video content from whatever source contained appropriate metadata labelling to enable lightweight content filtering. Furthermore there is no technical limit on the number of videos with an EIDR ID because EIDR can identify 1,208,925,819,614,630,000,000,000 (i.e. 1.20892581961463 x 10^{24}) unique audiovisual assets.

[25] http://eidr.org/documents/EIDR_BP_Ratings.pdf

[26] For full details concerning this proposal see SafeCast's response to the ICO's Age Appropriate Design Code call for evidence dated 17 September 2018 at https://ico.org.uk/media/about-the-ico/consultation-responses/2018/age-appropriate-design-code-responses/2260209/safecast.pdf

Question 24: How should VSPs balance their users' rights to freedom of expression, and what metrics should they use to monitor this? What role do you see for a regulator?

SafeCast's Answer to Question 24

SafeCast considers that, in accordance with Ofcom's core principle of "transparency", VSPs should allow their users to classify their video uploads through the use of a self-applied content rating system. This would maintain the users' freedom of expression while at the same time limiting the harm arising from children seeing inappropriate video content. Users who are unsure as to how they should classify their video content could be assisted by Artificial Intelligence (AI) reviewing their video and by third party professional video content experts who could review an uploaded video on behalf of the uploader.

As part of its lightweight regulatory processes, SafeCast suggests that every regulated VSP should supply Ofcom with an annual report on the operation of its self-applied content rating system together with metrics on its use of AI and third party professional video content experts. Ofcom would be able to review this annual report, compare it to reports from other VSPs and make suitable recommendations.

Question 25: How should VSPs provide for an out of court redress mechanism for the impartial settlement of disputes between users and VSP providers? (see paragraph 2.32 and article 28(b)(7) in annex 5). Please provide evidence or analysis to support your answer wherever possible, including consideration on how this requirement could be met in an effective and proportionate way.

SafeCast's Answer to Question 25

SafeCast considers the policy on YouTube of "*copyright strikes*" (three copyright strikes and the user is barred) could be an effective and proportionate model redress mechanism for the

impartial settlement of disputes between users and VSP providers so long as the bias in favour of one party is removed from the current system. An out of court redress mechanism could make use of Artificial Intelligence (AI) to balance competing rights provided that Ofcom's core principle of **transparency** is maintained, so that there is full information on how AI-based decisions are made; what those decisions are based on; and why they were taken the way in which they are taken. Ofcom could invite VSPs to develop and put forward appropriate solutions for its approval as being effective and proportionate ways of providing users with redress mechanisms.

Question 26: How might Ofcom best support VSPs to continue to innovate to keep users safe?

SafeCast's Answer to Question 26

By making use of the *Contracts (Rights of Third Parties) Act 1999* to give Ofcom the right to be notified and the opportunity to intervene, if it so wished, in disputes between VSPs and their users. This would be a *proportionate* light touch regulatory measure which would not inhibit innovation by VSPs. Reporting on metrics by means of quarterly reports for incidents of sexual abuse imagery and annual reports for everything else would not be an unacceptable burden on consumer facing businesses such as VSPs.

Question 27: How can Ofcom best support businesses to comply with the new requirements?

SafeCast's Answer to Question 27

The production of an Ofcom VSP Code of Practice which could be derived from the Ofcom Broadcasting Code but which includes all the light touch measures and reporting

requirements for regulating VSPs, would provide a substantive measure to assist businesses in complying with its new requirements. Such a Code of Practice would be consistent with Ofcom's core principles of *protection and assurance; freedom of expression; adaptability; transparency; enforcement; independence; and proportionality,* would serve as a roadmap for VSPs and for new entrants who wish to innovate and become VSPs.

Question 28: Do you have any views on the set of principles set out in paragraph 2.49 (protection and assurance, freedom of expression, adaptability over time, transparency, robust enforcement, independence and proportionality), and balancing the tensions that may sometimes occur between them?

SafeCast's Answer to Question 28

SafeCast considers that the core principles set out in paragraph 2.49 can be best addressed through the application of digital sovereignty to the regulation of VSPs. This concept, which is sometimes referred to as "*tech sovereignty*" (as in the speech by Ursula von der Leyen, President of the European Commission, on 19th February 2020)[27], allows nation states to harmonise technical issues in a consensual manner through the use of interoperable standards as a proxy for legislation with local variations. During the second decade of this century the emergence of standards within the broadcasting community has greatly simplified the global exchange of video content across all platforms no matter where they are located and in whatsoever jurisdiction. On both sides of the Atlantic work is going on in establishing universal standards to reduce friction, improve interoperability and increase trade. Universal interoperable standards can automate processes which would otherwise cause failure or error, such as the automatic selection of the correct version of a video for display on any device. Technical standards, such as embedding specific metadata, could enable effective child protection without censorship on the internet and in full compliance with national and international norms.[28]

[27] https://ec.europa.eu/commission/presscorner/detail/en/AC_20_260

[28] For further information on the concept of digital sovereignty see the Plum Consulting report Digital Sovereignty: the overlap and conflict between states, enterprise and citizens

The heart of the problems with video content from VSPs is that, unlike commercial movies, VSP content does not have to go through a pre-classification process before being made available to the public at large. Over the past three years the volume of video coming from a plethora of devices (smartphones, tablets, etc) has overwhelmed society's capability of protecting children from unlabelled content. Every minute of every day four hundred hours of new video content is uploaded to YouTube alone. Sharon White, the then chief executive of Ofcom, said that "*The sheer volume of text, audio and video that is generated or shared online far outstrips the output of traditional media. That means, for example, that it could be impractical to review platforms' decisions about content case-by-case*." .

During 2018/19 the UK Government's response to this situation was to require "*parity of protection*" for children, whether they be on or off line, whilst answering the challenges of censorship and capacity. Sharon White refers to this concept as "*the standards lottery*" which illustrates some of the difficulties in transplanting traditional broadcast regulation, unamended, into the online corpus.

Ofcom, with its increased powers as the designated Online Harms Regulator, is now in a position to establish improved child protection on the internet via an amendment to the international standards processes and its forthcoming light touch regulation of VSPs.

Alistair Kelman **Diana Kelman**

Directors and Co-Founders of SafeCast and SafeCast Global
(Telephone: -[redacted])