Some measures to improve the effectiveness, enforceability and universality of the UK Online Safety Bill

The current bill is ineffective, too costly and biassed in favour of the gambling industry and totalitarian censorship — without properly protecting children and vulnerable adults from online harms

Alistair Kelman[1]

Submitted 13 Sept 2022

Overview

This is a paper to go before the UK regulator Ofcom which in July 2022 published its roadmap to regulation under the UK Online Safety Bill and issued a call for evidence on the matters that will be included in the first phase of its work. It is based upon work which I have undertaken over several years both in practice at the Bar and since as a technologist. I wish to address and overcome the systemic failures in the proposed Bill and put forward the basis of a US-UK consensual approach to child protection.[2] All these proposed measures adhere to Ofcom's transparency requirements and are fair and proportionate responses to the harms and risks involved.
My recommendations will:

- 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

Keywords: metadata labelling, standards, Zachslaw, Online Safety Bill, loot boxes, gambling, pornography, trolling, stalking

---

[1] Alistair Kelman is a retired UK chancery barrister and technologist. He is the CEO and Co-Founder of SafeCast Limited

[2] This approach is also congruent with EU principles and is capable of being adhered to everywhere else in the world including the Middle East, China, Australia, India and Asia

Introduction

The UK Online Safety Bill in its current form has been considerably vitiated by major tech players and their lobbyists to the extent that effectively it can be used as a nontariff barrier by these major players within the supposed protection environment to their long term commercial benefits. Overall it creates a costly set of measures which do not necessarily fulfil the core regulatory requirements laid down in the consultations and reports from the UK Home Office, the DCMS and the ICO. It creates unintended consequences as the Bill will undoubtedly give rise to difficulties in interpretation and proportionate enforcement[3]. Judicial review of regulatory decisions, particularly where there are no precedents, are not only expensive with unpredictable outcomes but also impose huge burdens on any regulator that is given new and expanded powers. Judicial reviews also produce irrecoverable costs to the public purse. Meanwhile children will continue being harmed owing to the lacunae within the legislative processes.

.

A series of relatively small changes to the Online Safety Bill are capable of implementing Baroness Kidron's Children's Code[4]. These changes will enable the Bill to become a timely, effective, standards-based solution to the problems which have been amplified by the pandemic and its aftermath. The solutions presented herein are internationally compliant, are grounded in the Common Law and also in the technical operation of the internet - which has evolved over five decades.[5]

The promotion of standards based solutions for child protection - without censorship and without handing over control to impenetrable Artificial Intelligence (AI) systems that cannot be effectively interrogated and regulated, are constitutionally appropriate for the US Federal system of legislation as well as the UK and Irish legislative systems.[6]

In 2022, the promotion of standards involving the effective use of metadata in internet regulation without having to pass new laws and treaties is a means of regulation that has been adopted by the C2PA[7]. C2PA is a Joint Development Foundation project, formed through an alliance between Adobe, Arm, Intel, Microsoft, Sony and Truepic. C2PA is addressing the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content. On 1 September 2021 the C2PA released its first draft of an open technical metadata labelling standard to provide publishers, creators, and consumers with the ability to trace the origin of different types of media.- thereby providing the key tool for the elimination of fake news and fake advertising without censorship. The C2PA specification aims to repair and restore trust amongst participants on the internet. In this brief paper I aim to build my proposals upon its foundations.

Risky by Design

Baroness Kidron's "*5Rights*" organisation[8] published a series of web pages entitled "*Risky by Design*".[9]

[3] As recently highlighted by Lord Sumption - The hidden harms in the Online Safety Bill | The Spectator

[4] See Age appropriate design: a code of practice for online services

[5] See RFCs - Requests for Comment

[6] They are also appropriate to Australia where the Australian eSafety Commissioner is engaged in detailed regulation of these same issues

[7] See C2PA

[8] See 5Rights Foundation

[9] See Introduction to Risky-by-Design

These pages divide the risks suffered by children online into four broad categories, referred to as the '*4Cs'*

- **Content** - A child or young person is exposed to harmful material (e.g. age-inappropriate content, pornography, extreme and real-life violence, discriminatory or hateful content, disinformation, content that endorses risky or unhealthy behaviours such as anorexia, self-harm, suicide).
- **Contact** - A child or young person participates in activity with a malign actor, often, but not always, an adult (e.g. child sexual exploitation (CSEA), grooming, harassment, stalking, blackmail, unwanted sexual advances, location sharing).
- **Conduct** - A child or young person is involved in an exchange as either a perpetrator, victim, or sometimes both (e.g. bullying, sexting, revenge porn, trolling, threats and intimidation, peer pressure, loss of control of digital legacy/footprint). These exchanges are often, but not always, peer-to-peer.
- **Contract** - (also referred to as commercial risks) A child or young person is exposed to inappropriate commercial contractual relationships or pressures (e.g. compulsive use, gambling, targeted advertising, hidden costs, unfair terms and conditions, loss of control of personal data).

I ground my proposals in the requirement to protect children from risk in the '*4Cs*' and a requirement to be able to distinguish between truth and fiction using labels and tools that are to be made available to everyone under the industry led C2PA metadata specification. These labelling solutions could also enable the effective removal of harmful materials from sufferers of epilepsy without censorship if Ofcom and its companion telecommunications and childrens' regulators in other countries, decide to implement a form of *'light touch*' permissive regulation rather than negatively addressing harms after they have happened.

Age Gating versus Age Verification

In 2020, Baroness Kidron's 5Rights organisation made specific recommendations regarding gambling and age verification.[10] These recommendations were adopted and enhanced by the author of this paper[11] in September 2020 in evidence under Ofcom's VSP[12] consultation, The potential use of **Age Gating** in protecting children was raised in the following paragraphs:

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:

- 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.
- 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

[10] Further developed in 2021 by 5Rights in "*Age Assurance and the new regulatory landscape: 5Rights updated report 'But how do they know it is a child?"*

[11] Via SafeCast Limited

[12] Video-sharing platform (VSP)

available[13]. 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.

- 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 UK Government's forthcoming *Document Checking Service*[14], 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

[13] See Security Engineering Chapter 11 Inference Control by Prof Ross Anderson

[14] See Innovative new pilot launched to speed up access to key services - GOV.UK.

be deployed within a revision of Advertising Guidance.

In 2021 I raised these matters with the Broadcasting Authority of Ireland (BAI), the equivalent organisation to Ofcom in the Republic of Ireland. The BAI is overseeing the Irish Online Safety and Media Regulation Bill which is going through the Irish Parliament under their Joint Oireachtas Committee for Tourism, Culture, Arts, Sport and Media. Owing to the historic links between the UK and Ireland (and shared educational enrolment standards) the identical recommendations can be applied in both the Republic of Ireland as well as in the United Kingdom with economies of scale and effectiveness. Recently in private correspondence with the US attorney Damien Riehl,[15] I have formed the view that these same UK and Ireland proposals could also be adopted in the USA.

In the USA the Standards Advancement for the Legal Industry (SALI) enables analytic insights into the underlying law. It does this through identifying and classifying the subject matter of legal cases and statutes so that underlying principles, similarities and differences can be compared and contrasted in a systematic fashion. This approach is similar to the processes in the Chancery Division of the UK High Court under which a latent underlying principle can be extracted from the law which can then be applied to new circumstances without the need for new statutes or legislation; the most famous UK Chancery law analysis of this type is the case called "*High Trees*" where Denning J. (as he then was) held estoppel to be applicable if "*a promise was made which was intended to create legal relations and which, to the knowledge of the person making the promise, was going to be acted on by the person to whom it was made and which was in fact so acted on*." This latent principle - which was subsequently termed the principle of Equitable Estoppel - allowed justice to be done in a very difficult case[16] without overriding historic principles and gave rise to a series of judicial precedents which enabled justice to be done without the need for statutory intervention.

Using this Chancery law approach, Age Gating under standards could perform the same protective functioning as Age Verification for children in the USA without the harms to freedom that arise from a pre-adult identification system.

By the UK Ireland and the USA adopting Age Gating using a generic anonymised token service rather than Age Verification built around proprietary extensions there are a number of key and immediate advantages:

Universal child protection can be brought in quickly across the USA and in the British Commonwealth countries adhering to the Common Law system. These protections would be proportionate measures which would allow new entrants to offer VSP services as do the incumbents, but without there being proprietary *"walled garden*" barriers to the offering of VSP services;

The notorious harm which children are encountering arising from "Contract", as specified by 5Rights "Risky by Design" web pages (namely the harms of gambling, inappropriate advertising, hidden costs, unfair terms and conditions, loss of control of personal data), would be avoided by reason of the fact that under the Common Law system contracts made by children for goods and services that are not necessities are void[17]. By removing the ability of

[15] https://www.linkedin.com/in/damienriehl/

[16] Central London Property Trust Ltd v High Trees House Ltd - Wikipedia

[17] The classic English case in point is *Nash v Inman* in which a tailor sued a minor to whom he had supplied clothes, including 11 fancy waistcoats. It was decided that, as the minor was an undergraduate at Cambridge University at the

VSPs, technology companies and Apps to create commercial contracts with children, the internet would become a safer place for children to play and learn.

Advertising to children to make them aware of goods and services would continue to be possible under the existing Ofcom and ASA regulations without engaging children in contractual relations. As set out in the Bill, advertising to children must be in accordance with the requirements for the conduct of a risk assessment for potential damage to children.

The replacement of the UK CAP "*120 Index*" system[18] by filtering content against a school age token produced by a generic *"token based"* enrolment system in primary schools, would remove a serious administrative cost burden imposed upon legitimate advertisers and broadcasters. It would allow new entrants to engage in advertising on VSPs and regulated broadcasting services without the need to comply with an outdated regulatory requirement;

It is important for legislatures in all countries to be fully aware of the huge cost of measures, such as the UK Online Safety Bill, which, in their current form, do not address the underlying problems. In the UK the original "*Full Economic Assessment*" to the Online Safety Bill said that *"A duty of care for user generated content and activity addressing illegal harms and safeguarding children from both illegal and harmful content activity*" will cost £1.7 Billion ($2 Billion) with a new Ofcom branch in Manchester costing at least £46 Million per year ($54 Million) and employing 150 staff. This bureaucratic regulatory regime is going to take many years to implement - with British children remaining unprotected while it is rolled out. The same level of costs and delays will be necessary in the USA and elsewhere - unless these alternative technical solutions are considered and deployed.

#ZachsLaw - the Epilepsy Society

In June 2022 the UK Government performed a U-turn and accepted local MP Paul Maynard's[19] amendment to the Online Safety Bill making the deliberate sending of flashing images a criminal offence, as this can cause seizures for those living with epilepsy.

I believe that this UK Parliamentary amendment is well intentioned but misguided - the better solution and one which is capable of universal global deployment is through the use of metadata labelling standards and lightweight filtering so long as Ofcom is prepared to change its regulatory policy so as to enable technology companies to do the right thing in the face of scientific evidence. Such a change in UK regulatory policy would be acceptable to the BAI and to the FCC in the USA, all of whom have similar regulatory responsibilities and neither of which wish to be content censors,

As a special case, in the UK there has been a long standing practice whereby television advertisements are tested to see if they contain flashing images which could trigger an epileptic seizure prior to being approved for broadcast on the UK television

---

time, the clothes were suitable according to the minor's station in life. Unfortunately for the tailor, however, it was further decided that they were not necessary, as he already had sufficient clothing. **Similar US case law states** that "A contract of a minor is not void ab initio, but merely voidable at the election of the minor upon his attaining majority," when "a person may either disaffirm or ratify a contract that he entered into while he was still a minor." *Fletcher v. Marshall*, 632 N.E.2d 1105, 1107 (1994) (cited by *Nitka v. ERJ Dining IV, LLC* (N.D. Ill. 2018).

[18] Explained within the CAP document Identifying TV programmes likely to appeal to children

[19] https://www.paulmaynard.co.uk/

networks. The original company which developed the flashing images identification service for use in commercial advertising only test the flashing images with big television screens that are viewed across a room. They do not test their technology with tiny mobile devices which a child might be viewing a few inches from their faces. Put simply, the original company (and all other potential providers of similar services) are afraid of Ofcom doing what it normally has done in the past, which is prosecuting a company who Ofcom considers to be in breach of regulations rather than leading the safety industry towards better practices.

This heavy handed regulatory approach has led to companies which are capable of testing flashing images, taking the commercial decision not to test in different environments or investigating whether their technology can be more widely used than in the approval of advertisements shown on large television screens in the home. By not certifying that their technology can prevent harm they avoid the risk of being sued by Ofcom. But children with epilepsy continue being harmed.

As an alternative remedy to prosecuting for harms it would be better for Ofcom to regulate the use of new technologies to protect epileptics from flashing images by means of a "*comfort letter*" procedure whereby Ofcom would absolve developers and licensees from prosecution if they could certify that they were using the best and latest technologies to protect epileptics from flashing images.

Such an approach would involve there being a research programme to investigate how flashing images on a small mobile device could harm epileptics. This research programme could be international - and would be implemented through the use of video digital standards.

Comfort letters are frequently used in competition law situations where parties can make a voluntary submission to a regulator to check the legality of vertical and other arrangements. Here the competition regulator considers the facts and issues a letter saying that it does not consider that the proposed course of action would be harmful to competition. The issuance of a "*comfort letter*" is binding upon the competition authorities - who are then unable to resile from their comfort letter so long as there has been full disclosure by the party seeking approvals.

A comfort letter in respect of labelling and filtering of flashing images would only be issued after consideration of independent and high quality research that comes to firm conclusions.

Such an approach could also lead to protection of epileptics who are currently unable to safely make use of Virtual Reality headsets - by establishing a safe route to market these technologies.

I am very grateful to Damien Riehl for reading this paper in advance of its circulation and for making a number of highly constructive suggestions - nonetheless all mistakes and errors in this paper are my own.