Report on the Second Meeting of High-Level Working Group for Privacy and Safety

Prof Andy Phippen, Bournemouth University
Prof Emma Bond, University of Suffolk

Introduction and Overview

The online harms world is a challenging one, where most parties, we are sure, would agree that it is important that citizens can engage with online platforms and discourses in a manner that mitigates harm and does not expose them to abuse. However, how we achieve this is complex and many stakeholders have conflicting viewpoints. These range on a continuum from the prohibitive ("Harms occur on platforms, therefore platforms need to stop it") to the progressive ("Harms are caused by people, how do we reduce them by being mindful of people's right to participate free from excessive surveillance?"). There are many views between these two positions. However, it is unquestionable that these perspectives all wish to achieve the same goal – that people, particularly young people, can experience the online world while not being harmed.

The 'High-Level Working Group for Privacy & Safety' aims to advocate for a holistic, person-centred approach to online safeguarding that respects people's rights to online participation and to their privacy.

Convened by Prof Andy Phippen and Prof Emma Bond, the Working Group intends to drive discussions where central concepts such as harm, risk, vulnerability, well-being, and the best interest of the child are addressed in a nuanced and contextual manner to move conversations on from the traditional prohibitive narratives that beset the online harms work. In convening this group, Andy Phippen and Emma Bond, who collectively have 40 years' experience working in this area, are hoping to develop a more inclusive and progressive narrative that moves from "someone needs to stop this" to "what can we all do to make online experiences more inclusive while understanding and reducing harm". Current political narratives generally centre around how platforms can reduce or eliminate harms, with little consideration of other stakeholders that might be better placed to mitigate these risks.

The group brings a multi-stakeholder approach, convening experts from regulators, research institutions, private companies, industry associations, non-profit organisations, and academia to better articulate the challenges of tackling online harms in a right based, empowered manner.

As such, the goals of these sessions are:

1. Build a community of stakeholders with a progressive view on tackling online harms

2. Placing a more progressive voice into the public domain with broad stakeholder buy in and a constructive conversation between parties aiming to achieve a common goal mindful of children's rights
3. To develop new approaches that stakeholders might adopt that go beyond technical intervention and prohibitive measures.

Sessions take place under Chatham House rules (although some attendees have consented to being named as attendees). Reporting on each session will be conducted through the publication of a detailed article on the discussions that took place (this being the second report in this series). These documents present the discussion that took place and will result at the end of the first three sessions with a recommendations document that brings together all the discussions that have taken place to articulate what a progressive, holistic, and inclusive approach to tackling online harms looks like. These reports are presented as working documents rather than academic analyses of the events with each output will be made publicly available for free. By placing these reports in the public domain, it is our intention to propose ways we might move conversations on from the current cycle of prohibition and prevention and introduce some new voices into the debates around online harms.

Session 2 – Interests, Rights and Freedoms

This report considers the meeting of the High-Level Working Group for Privacy and Safety, which took place in London on September 14th, 2023. As with the previous meeting it was convened with the organisational support of Meta. Once again, no one was funded to attend, and these are not paid events. Everyone gave their time and expertise for free.

The focus for this second session was to consider the balancing of interest, rights and freedoms. In particular to consider what rights are afforded to people in their engagement with online services, how those rights can be both recognised and respected in scenarios where there is a need to be kept safe, and whether there might be offsets around rights in order to keep the wider population free from harm.

As such the discussions around this balance considers the risks, roles and responsibilities, drawing from ethical, social, legal and political dimensions, as well as the tools that might be used to achieve this balance, exploring issues related to technical, social, psychological and legal dimensions. However, while this introductory context provided the foundations for the discussions among stakeholders, as with the previous session we were also mindful to let conversations take a natural path around these issues, rather than being bound strictly to this approach so that the priorities of stakeholders could be explored without too much restriction. This report is a presentation of the topics discussed and the issues that were raised and, where there were conflicting views, these are explored.

Differing from the previous group meeting, that took place online which was better managed with a single discussion, this session split the meeting into two different discussion groups which would focus on a specific online harms' challenge/use case, before switching

to the other one. With each group facilitated by one of the report authors, the smaller discussions allowed everyone to have a clearer voice to express their views.

The use cases considered in the discussions were *age assurance* and *parental supervision*.

In one of these use cases (age assurance) we were mindful on the level of faith being placed as a "solution" to a lot of online harms challenges – for example in the UK's Online Safety Act 2023[1], it has been proposed that age assurance will be the approach that will prevent young people from accessing adult content and engage with social media platforms while underage. Age assurance, age verification and age estimation appear almost 100 times in the legislation. While it is acknowledged at the time of writing that the regulator (OFCOM) has only just published its update roadmap for the implementation of the Online Safety Act[2] which will not conclude until the end of 2026, we expect to see age assurance to play a major role in platforms demonstrating their duties under the legislation. Therefore, it was important to discuss the views of stakeholders regarding the efficacy of the approach and how it relates to both young people's and the wider population's rights.

In contrast, parental supervision is not considered to any significant extent in the new legislation[3]. However, given our work around online harms over a long period of time, we have heard much discussion around the role of parental supervision, and parental responsibility, in keeping young people safe online. However, the concept of parental supervision seems to be somewhat intangible outside of some stakeholders stating, "it should be parents that deal with these issues, they are the ones most close to the child" while others will say "this is too complex for parents, providers need to prevent harms so parents do not need to deal with them".

Therefore, these use cases present differing perspectives on rights and responsibilities, while unquestionably both being parts of the different facets of online harms prevention/mitigation.

In each use case, the format of the discussion aimed to firstly unpick what the goal is, and whether this differed depending on the perspective of different stakeholders. And, if this goal is to be realised, whether there is a need to trade-off rights, or whether there is opportunity to innovate in order to balance rights and requirements of the approach. For example, in the case of age assurance, are there rights challenges that emerge from a technical intervention that requires an end user to "prove" their age? And in the case of parental supervision – are there expectations placed upon parents that cannot be met due to their lack of knowledge or engagement with online harms debates. We are reminded of

[1] https://bills.parliament.uk/bills/3137

[2] https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation

[3] We are mindful that the Online Safety Act is one example of legislative implementation tackling online harms, and others exist. However, we use it as a vehicle for discussion here by way of example, rather than definitive approach.

claims by policy makers that it is not the responsibility of parents to prevent online harms, for example[4]:
So I want to reassure every person reading this letter that the onus for keeping young people safe online will sit squarely on the tech companies' shoulders. You or your child will not have to change any settings or apply any filters to shield them from harmful content.

The basic framework for the discussions were:

1. In the scenario, what is the desirable outcome(s) and how does this balance respecting people's rights while keeping them safe?
2. What equities need to be considered around rights, harms, legal duties, and safeguards?
3. What capabilities (technical, social, psychological, legal, etc.) come into play?
4. Are all stakeholders engaged in the ecosystem in this use case?
5. What does applying the best interests of the child look like in practice for this use case?

And finally, an overarching question regarding whether there is a hierarchy of rights (e.g., are some individuals entitled to greater protection at the detriment of others) and do some rights rank higher than others? And if there needs to be a utilitarian approach to tackling online harms (i.e., the greater good for the greatest number), who decides these tradeoffs?

Age Assurance

The discussion groups for age assurance started with a recent quote from the political debates around the online safety bill[5]:
Importantly, user-to-user providers, as well as dedicated adult sites, will now be explicitly required to use highly effective age verification tools to prevent children accessing them. The wording "highly effective" is crucial, because porn is porn wherever it is found, whether on Twitter, which as my right hon. Friend the Member for Chelmsford said is the most likely place for children to find pornography, or on dedicated adult sites. It has the same effect and causes the same harm. It is therefore vital that tech companies will actually have to prevent children from going on their sites, and not just try hard.

This quote was used because it is both recent to the debates around online harms in the UK, and also reflects the somewhat absolutist political view around age assurance when it comes to tackling online harms. Specifically that platforms need to stop children from accessing adult content or going on platforms where age limits should prevent them from doing so.

However, it also raises concerns that there is a view (that it became clear was not shared among our stakeholder groups) that age assurance was a *solution* to this, rather than tool to

[4] https://www.gov.uk/government/publications/michelle-donelan-writes-to-parents-setting-out-how-the-online-safety-bill-will-keep-children-safe/read-the-secretary-of-states-open-letter-to-parents-carers-and-guardians
[5] https://hansard.parliament.uk/Commons/2023-09-12/debates/81853BB7-375E-45C0-8C9D-4169AC36DD12/OnlineSafetyBill

prevent some access and is used to support the input of other stakeholders. And, to return to the quotation above, child and parents should not have to do anything, platforms just need to stop it. Nevertheless, the language does cause us to question whether this is well considered, for example, what does "highly effective" looks like? And are the goals of policy positions like this prohibition, rather than support or risk mitigation? And, as a consequence of this direction, is prohibition possible?

There is clearly a challenge in transforming statements like this into something that can be specified and regulated – we need to be mindful that while the Act has now reached assent there is still a significant job for the regulator to be able to transform what is set out in law into something that can be regulated. And language such as "highly effective" makes it challenging to know what "doing enough" might look like for a regulated provider. Is *highly effective* a system that prevents 100% of access without any false positives or false negatives (i.e., a system that prevents all access to under 18s but enables access to everyone aged 18 and over)? Or should there be some metrics associated with it? And would these metrics include acceptable rates of prevention for those who have a perfectly legal right to access (given that, in this example the politician is talking about pornography, which has no legal restrictions for those past age of majority in the UK).

It was acknowledged by most participants that tools such as age assurance will be effective at preventing accidental discovered by young people. However, there was far less confidence that it would prevent access by a determined teen who would have already considered work arounds to obtain the content they wished to view. While Virtual Private Networks are often proposed as an easy work around to geographically bound age assurance approaches, there was also concern that moving to privacy preserving technology to access adult content, such as Tor browsers, potentially posed the risk that young people could be exposed to illegal content as a result.

However, should the perfect be the enemy of the good? Just because there are work arounds and it is unlikely to be a perfect solution, does that mean that age assurance should be dismissed as an approach? It was generally the view of participants that age assurance certainly had a role to play, as long as the regulation was in place that acknowledged the lack of a perfect solution and that was reflected in the regulation. Providers of age assurance approaches acknowledged that precision for 18 is good, but not perfect. By way of example, it was stated that for a safety critical system, such as a complex Issue like child trafficking, there would not be confidence that age assurance approaches should be considered as a solution to assessing a victim's age.

There was a question raised about whether society "wants" age assurance enough – the view being if there was enough public will, it could be put in place. By way of example, we should consider offline approaches such as the purchasing of knives or alcohol. In these cases, there is little public or stakeholder opposition to an "age gating" approach and the question was raised why it was different for online concerns. The groups raised a number of concerns around using offline examples by way of comparison. Firstly, while there is social acceptance of age assurance approaches for things like the purchasing of knives and alcohol, it was also acknowledged that these were not perfect. Equally, for face to face offline age verification in, for example, self-service checkouts in supermarkets there is a threshold set to

25, not 18 (the member of staff from the store presses a button to approach, for example, the sale of controlled painkillers by stating "this person is clearly over the age of 25.

Furthermore, in online transactions on age restricted items (for example knives) there is also an age verification approach that includes some form of human intervention from the delivery driver (given that age restricted items cannot be delivered to lockers for later collection). Amazon's policy on age restricted items[6] states that if the delivery driver is not convinced the person to whom they are delivering the item is over the age of 25, they have to ask for a document-based form of identification, such as a driving license or passport.

If we instead propose this approach for access to pornography, given the volumes of access to these services (Pornhub states that they have over 100 million visitors per day[7]), we would anticipate that rejecting millions of users who are entitled to access adult content, because they did not look over the age of 25, would be highly problematic and impactful on rights. While there might be a view that this is an acceptable payoff to ensure young people cannot access pornography, there was no agreement that this would be tolerable among the group.

A further observation was made that if platforms can ensure inappropriate adverts are not shared with young people, why cannot they do the same for adult content. Again, this is a simple interpretation of a different use case – young people will not actively seek out ads in the same way they might do with pornography.

We returned to the perennial challenge posed by age assurance that does not use some form of estimation in the UK – there is no universal ID based approach which every citizen holds. Therefore, if document-based approaches are used there will be significant numbers of citizens who cannot verify their age because, for example, they will not hold a passport or driving licence. There is not legal requirement for them to hold these documents and furthermore, there are financial barriers preventing some people from having them. Unless an ID card was introduced in the UK (and there was serious opposition to such proposals given the UK's history around this issue and the potential massive privacy abuses) we must accept that these approaches will not be 100% effective.

There were observations made about the failure of policy makers to learn from history in this regard. The Digital Economy Act 2017 introduced measures for age assurance when accessing adult content that was withdrawn once the technical problems were properly explored and the political view changed to state that the technology was not capable at the time. There were views from some that the technology (and the wider environmental issues such as a lack of uniform ID) was still not sufficiently advanced to achieve the goals of the Online Safety Act. It was also observed that there is a wider history in attempting to use technology to prevent young people from accessing pornography. Various tools (for example filters) have been available for a long time, but have little impact in the home. There was concern that current age assurances will not be any different – while the control has moved

[6] https://www.amazon.co.uk/gp/help/customer/display.html
[7] https://www.pornhub.com/press

from the home to the server, it still cannot account for end user behaviour such as by passing or simply having one user in the home leaving their age gating logged in.

Age estimation, which used machine learning techniques to estimate the age of an end user from either behaviour or physical appearance was also discussed. While clearly there have been technical advances in this area, one of the market leaders, YOTI, still acknowledges that their solutions are not perfect[8] and are also measured in terms of efficacy in identify someone between the age of 13 and 17 being under the age of 25, not 18. While their accuracy is impressive, it could still result in many false positives given the volume of traffic to adult sites which, clearly, would impact upon the rights of someone who is perfectly entitled to access this content.

None of these things would be particularly problematic if age assurance was not viewed as a perfect solution (or at least "highly effective"). At odds with the political rhetoric, within our discussions it was recognised by both providers and regulators that this is not a perfect solution. Which does raise the question on where do policy makers get the impression that technology can be a complete solution on its own? There was a general view that technical voices were lacking from a lot of these policy discussions, with the opinion of some being that the technical voices are ignored because they are not telling the policy makers what they want to hear. Far easier, it was suggested, to say "platforms could do this if they wanted to" rather than "this is a complex area of social policy unless you with to introduce an ID card system with the associated political and financial pressures that brings".

Further challenges were raised around other forms of age assurance. If anything, the challenge around verifying adults is easier than other challenges – there are at least some means for adults to show they are over the age of 18. If we consider the challenge that social media providers should use age assurance to ensure no one under the age of 13 should be on their sites, there are further difficulties, because there is no clear form of ID where someone can prove they are over that age. It was recognised by providers and regulators that for 13 it is even more complex, and the accuracy is not as strong. One potential solution was proposed that given that schools have data on the age of pupils and can "verify" their ages, schools' data might be a means by which the age assurance industry could implement these systems. However, this view was not well supported and concerns around data protection abuses (i.e., collecting data for one purpose then using it for another) were raised.

However, there was a concern voiced that there are some who are too quick to dismiss age assurance due to its inaccuracies and this in some way raised concerns that those who opposed it were happy for young people to access adult content. This was a challenge about the binary nature around a lot of these debates and how the political rhetoric has, along with many other areas of social policy, developed a view that you either support them or you want children to be harmed – parallels with the debates around end-to-end encryption in messaging platforms were drawn. It would seem that critical voices were no invited in these discussions because you have to believe that the legalisation and policy approaches will work. However, among the discussion groups, which comprise stakeholders from

[8] https://www.yoti.com/blog/yoti-age-estimation-white-paper/

different fields, this is not the case, and reflects in some way the goals of this group to bring more critical, progressive views to these debates. The political rhetoric, such as that at the start of this section, highlight a more general prohibitive approach to a lot of the legislation – we need to stop things from happening (for example grooming) and stop young people doing things (for example accessing pornography and social media if they are "underage").

A further problem, which provides a useful segue into the other discussion group, was raised around the role of parental responsibility in the prevention of young people's access to pornography (given this seemed to be the focus of a lot of the political debates around age assurance). Given there is virtually no content in the Online Safety Act related to parental responsibility, there were concerns that poor parental behaviour might further challenge efforts by platforms to ensure their platforms were "highly effective". And if this was being rolled out by the regulator, would the regulator provide any guidance on this, or would be rely on the example set out by the minister's letter that they do not have to do anything? It is an uncomfortable fact in society and in areas of prohibitive social policy that some parents will be accessing pornography through age gates, as they are perfectly entitled to do. If they leave their age assurance logged and therefore accessible by other users in the household, does the provider remain liable to unrelated access? Device sharing by parents was also raised as a potential risk factor if our goal is the prevention of access to pornography. And could a parent who did not conduct due diligence/responsibility be consider liable for abuse?

There was not much discussion around the solution to this, because it would seem that a solution would be hard to come by. However, it was a useful discussion points it highlighting the challenges of age assurance if all the liability it placed with the provider without acknowledging the potentially conflicting behaviours of other stakeholders.

Returning to rights based issues and the challenges of an imperfect age assurance approach, a point raised related to the right to non-discrimination for ALL (as defined in article 2 of the UN Convention on the Rights of the Child): Every child has rights "without discrimination of any kind, irrespective of the child's or his or her parent's or legal guardian's race, colour, sex, language, religion, political or other opinion, national, ethnic or social origin, property, disability, birth or other status"

Given the accuracy of age assurance systems, is the prevention of *some* adults from accessing pornography an acceptable trade-off for ensuring *most* young people will not be able to access it (until they discover work arounds)? Concerns were also raised that if verification systems would not be effective and estimation systems were increasingly seen as the better solution, what are the rights challenges in collecting the sort of training data (i.e., images of young people) that would be used in sufficient numbers. And were we also confident, considering article 2, that such training data would be sufficiently diverse that it would remain accurate for different ethnicities?

In drawing the age assurance discussions to a close, the question was posed that if, as was the clear majority view in the room, if these systems did not provide a solution, what other measures might be effective in supporting young people's online experiences, so they are "safer"? Content moderation was seen as improving in performance but not perfect.

Engagement with the end user around reporting and disclosure would improve things further. It was acknowledged that the additional support for education that came from the requirements on platforms to provide transparency data around how the deal with reports and disclosures would be a useful tool. However, it was also acknowledged that the ambiguity of the expectations on what is harmful, and who decides, meant that detecting ambiguous harms will have major impacts upon training data for such systems and there was still a need to break free of the Minister's expectation that parents and children need to nothing to better consider a broader range of stakeholder responsibilities.

A more contentious question was also raised at this point – why is pornography viewed as the primary use case for age assurance? Is pornography so harmful that the political drive is justified or is this a moral position? Some attendees pointed out that while the literature around young people and pornography raises a number of concerns and it was clearly harmful to some, the literature was also inconclusive on the widespread impacts, and some political claims had little evidence (e.g., young people who look at pornography go on to abuse other children). For example, the second report for the Office of the Childrens Commissioner that explores the impact of pornography[9] showed that in *some* cases young people who have exhibited harmful sexual behaviours have accessed pornography. However, there was no consideration of a causation in opposite direction (i.e., they did not speak to any young people who had accessed pornography who have not engaged in Harmful Sexual Behaviours).

We should stress, once again, that the view of participants was not that it was perfectly fine for young people to access pornography and we should just let them get on with it. However, there was a view by some that this is more of a moral position that is being argued as a harm based one (which could also be levelled at other areas of social policy such as drugs harms). Can parliament define moral stances as the representatives of society and how does that tackle different moral positions? And perhaps equally importantly given the political views, how can code implement a moral position?

Which furthermore raises concerns about whether Parliament and regulator will decide what is harmful, whether there is any route to challenge, and whether there is confidence this comes from an evidence-based perspective that draws from a broad range of sources, not just those that simply align with the policy direction.

We finished by looking at education, which is the primary call from young people themselves and there was a lot who asked why there so much focus on this particular aspect of technical intervention when it is at odds with what young people call for. In the case of online harms was it really the case that adults felt they knew better and therefore young people should just do as they are told and not have a voice to inform policy? (do adults "know better"?). Why was there less focus on better care around the child (i.e., informed stakeholders, trauma informed approaches, focus on resolution) rather than prohibition? Again, there was a view that perhaps proposing technical solutions and claims that no one except the providers need to do anything if far easier social policy to communicate even though it would achieve little.

[9] https://www.childrenscommissioner.gov.uk/resource/pornography-and-harmful-sexual-behaviour/

There was clear concern that is education virtually absent from online harms legislation and there was a clear view that online harms prevention or risk mitigation needs to have its heart in good quality social and relationships education. As one participant articulated, LGBT young people will go to pornography to learn about sex because there is literally no education about it in formal or informal educational settings. Education remains the most important aspect for young people, alongside educational and home environments with supportive adults, yet there is nothing in current policy directions to address this.

In closing, the single stakeholder model around age assurance was viewed as problematic for several reasons. But a fundamental truth also remains - even if successful, this is content that is legal for adults (regardless of the moral perspectives of parliamentarians). They can access as much as they like as soon as they turn 18. If this content is as harmful as claimed, surely this is a major social concern? Why would it be harmful for a 17 year old but fine for an 18 year old? There is nothing in current policy directions that attempt to tackle this social issue, there is no evidence to suggest the prohibition to 18, then free reign to consume as much content as one likes is a successful policy direction, and we can draw parallels from alcohol policy in this regard.

Parental Supervision

The focus on the parental supervision group also considered prohibition against more progressive viewpoints. The group explored *parental supervision* and highlighted that there is a distinct difference between parental supervision and *parental control* that is often overlooked. Often adults fear what young people may be doing online and therefore restrict their online activity. Restriction is often couched as protection but does little to protect the child. The more restrictive parenting more likely child is to hide activity and less likely to seek from parents is risk arises. It was also observed that while the parent might be viewed as a key stakeholder in keeping children safe, and therefore engagement was important, we also know from wider safeguarding that the person most likely to harm a child is someone known to them, including parent. Therefore, as with most of our discussions around online harms, this is not simply a case of looking at the responsibilities of a single stakeholder and assuming if they are well equipped to tackle online harms, all will be well. A child who is in a controlling or abuse household will need routes for disclosure and support outside of the home.

For example, attendees raised cases of children contacting ChildLine, the child safeguarding helpline, after experiencing online risk rather than telling parents as they were frightened of consequences of disclosing to a parent – for example having their device taken away. Removing a device is a restriction, not supervision.

It was raised that the focus of a lot of parental knowledge comes from what comes out of the media, and often conflicts with young people's needs. For example, the unhelpful concept of screen time used by policy makers and schools and parent as attempt to control but in reality, screentime as a stand along measure has little impact on wellbeing and certainly and does not ameliorate risk. However, it is perhaps easier to focus on something with a simple metric (i.e., "you've been on that too long") rather than the more complex "is

there anything I can help you with to better understand the risks around online harms and how to mitigate them".

What keeps children safe is not a parent there all the time or digital spying using all manner of safety tech platforms with dubious claims for efficacy and rights abuses. There is a far greater need for tools for children to keep safe, and to understand how to disclose if they are concerned, and teaching them the skills to know how to use them and navigate online spaces safely. Adult control does not keep children safe, the best thing for a child to know is when to seek help and support. The challenge for any piece of prohibitive legislation is that they are not tackling a fixed harm – there is shifting landscapes of risk and shifting landscapes of digital risk. For example, LGBTQI+ young people being outed to parents or wider community by tech might then have to face the consequences of parental reactions. By way of further example, parental reaction to a child seeing intimate images of peers under the age of 18 could potentially escalate rather than ameliorate risk, as a parental overreaction might result in bring in the police and a young person being charged, rather than protected. Parental supervision is not control and surveillance, it requires informed parents to create an environment of support and risk mitigation. However, there is little current social policy that looks to address this.

Resilience is important and more likely to keep child safe rather than restriction and this is developed through education and understanding. The group also discussed the relationship between the role of parental supervision and educating young people about social media and virtual environments and that if education for young people could be improved was parental supervision so important? There was general agreement that self-governance of a child was more effective than parental control and this was a central aspect of resilience. However, this education might have a parental dimension. Teaching children resilience is what is important but there was conflict between the politics of restriction (which could be considered easier in social policy terms) versus rights and resilience as life skills (which took longer to achieve and did not tie in with political terms of office and "quick wins").

The prohibitive approach can also be seen in school settings. Supervision and monitoring in schools was typical (and, indeed, a statutory duty for schools in the UK) and the use of filtering and monitoring was not generally viewed in a positive light because it reinforced the view that things could be stopped and therefore risk did not need to be understood. There were views that rather than over reliance on these tools it would be better to educate young people. For example, while they were not able to use dating sites until they were 18, there would be no harm in getting them to understand how such sites might be used safely and are an ideal model for understanding risk and harm and how to develop resilience.

More broadly there was a view that education, wherever participants lived, was generally poor and not well considered by policy makers. There was a view that if there was a more comprehensively, *developmentally appropriate,* digital skills framework that was also transparent to parents so they could also follow and understand what is age appropriate in relation to young people's social media use (and therefore reduce the risk of overreaction in the event of a disclosure) would be more effective than prohibitive legislation. Although as the group pointed out the approach of assuming stages in child development is of itself problematic. What is age/stage appropriate and who makes those decision? Co-creation of

approaches with clear youth voice was considered a gold standard in approach, but there was little evidence that this was something that regularly emerged from the policy world.

There was much discussion around rights and how it related to parental supervision and a view that many of the debates that are occurring neglect children's rights, whereas they should be at the foundation of any policy approach. Notion of 'best interests' of child are not always easy to apply but we need to improve understanding of rights and best interests of children. Youth participation and role in design is essential in achieving this. Co-design and co-production are essential to ameliorate the risk of wrong assumptions in tech design and delivery, and reflect the anticipated "adults know best" narrative that many young people find problematic. It was raised that Meta have a best interests of child framework – but equally it was acknowledged that this needed to be mindful of cultural relativism by design and again highlights the problematic nature of nation states trying to regulate global platforms. That is not to say that governments do not have a role to say but, to take the UK government's claim to "make the UK the safest place to go online in the world", there is a strong ethnocentric perspective on this. Any best interests framework needs to be constantly evolving – the best interests of *a* child may not be best interests of *all* children. In recent times there have been many examples of this, for example LGBTQI+ issues and the example of US pregnancy information and abortion law. There is still a view by many that parents' views trumps children's views, and that hardly reflects a best interests or youth voice based approach.

It was agreed that children's rights need to be promoted more in educational messages and there was frustration that the almost universally agreed (by nations) rights framework, the UN Convention on the Rights of the Child was not more embedded in education approaches and that parents were not more aware of it. Indeed, in some cases there have been statements by politicians that are at odds with the UN CRC[10] – for example that children need to accept an erosion in privacy in order to ensure that are safe.

There are parallels that also highlight conflicts between the rights of the child and the views of the parent and the role of parental supervision. For example, parental discourse around sexuality - thinking that they need to protect children from information on sexual health or sexuality is not acceptable. We should, instead, consider the role of the parent in helping the child exercise their rights (which requires a great understanding of children's rights by parents) and protection from harm – balancing freedom of expression with protection from sexual abuse or trafficking. Which, in reality, relates equally strongly to online harms – the parent needs to provide a supportive environment why a young person can explore the online world and be mindful of the risks, while still having reliance on the parent to mitigate those risks. Furthermore, the concept of agency is crucial in children's rights. Children should be able to exercise agency especially aged 13-18, with the support of the parent, rather than being in conflict.

An example of this that arose, which is often applied in child protection, is the Gillick 1983 competence case – which defined the seminal case law around young people and

[10] https://www.dailymail.co.uk/news/article-2265583/Snoop-childs-texts-Its-bizarre-parents-treat-youngsters-internet-mobile-exchanges-private-says-PMs-childhood-guru.html?ito=feeds-newsxml

contraception. This is an important case to bear in mind in any discussion of parental supervision as is child consent over age of 13. Some parents have an issue with that, but parental views are not always in the best interests of the child, as the Gillick case law highlighted.

There were also areas of tension between parental behaviour and children's rights being highlights in our discussions, many of which relate to privacy. For example, there can be tension between parent and child around device access if the parent is paying the phone bill or owns the contract. There is potentially a concern around the right to privacy should the parent, through ownership of the device/contract, be able to see who the child has been contacting. Similarly, there was general agreement that parents "stop checking" devices and demand access achieve little in terms of safety and have the potential to drive the child to hide content and communication and they will be less likely to disclose concerns if they feel the outcome of the disclosure would be further surveillance. Another familiar parental practice of posting information and pictures of children on their own social media platforms without the consent of the child should also be viewed as an invasion of their privacy.

As a result of concerns around parental behaviour and role modelling, there was also a more progressive discussion around what good parenting would look like in relation to supporting young people.

There was broad agreement that what is more important is good communication and trust, along with respect. There was also agreement that parental supervision related to online harms should not be a 'one off' but ongoing negotiations and interactions considering the measures that could be put in place to support the young person, and some of these might be technical tools, such as blocking, accessing contacts, managing access to certain platforms and apps and similar. However, such technical interventions must be both age appropriate and also with the consent of the child, once they understand why measures are being put in place. "Because I say so" was not considered to be good parenting when tackling online harms and supporting young people's online lives.

Unsurprisingly there was consensus that having open conversations around online harms and risk that take place regularly and are not confrontational lay a foundation for young people to be confident that should they encounter a harm, or even simply something that concerns them, they know they can disclose to the parent without risk of punishment or judgement.

However, the language of online harms and the potential misunderstanding by parents might result in more prohibitive approaches taking place. For example, if a professional (for example a teacher or a social worker) talks to parents about "monitoring and supervision" it is important to understand this does not necessarily mean "you need to install tools to do this", particularly given that there are plenty of safety tech projects that use similar language.

The notion of a 'good parent' was raised and stressed this is a socially and culturally constructed concept – what is understood to be good parenting, for example, restricting access or screen time in one context may be viewed as controlling and bad parenting in

another or by another. Digital environments are very much a part of children growing up – so should be a key role of everyday parenting.

The challenge for service providers in supporting parents was also a key topic of conversation. One of the issues that platforms, who have to provide consistent services, is the diversity of parent styles and beliefs. While some parents might wish to see more detail about the behaviour of their children on the platform, others will not. Platforms have a balancing act in providing tools while being mindful of parental need and children's rights.

There was a point raised that often that even parents of an individual child may have very different approaches to parental supervision which may cause conflict especially in the event of separation. Examples were given of court orders in the UK mandating access to mobile phone for contact with absent parent but no regard to access of content and contact with other parties. Again, this can present challenges to the provider if they are mindful of orders that might not be in the best interests of the child. There was a view by some participants that the family courts can potentially negatively impact on child/parent relationships if, for example there was access to the absent parent granted via mobile device, but no consideration given to other forms of communication, for example social media. This was viewed as another knowledge gap by a stakeholder group that could negatively impact on young people's experiences.

The concept of parental knowledge, and its development, was a prevalent feature of the discussions. It was noted that there firstly a great deal of variability in knowledge among parents. Equally, there was a great deal of diversity in parental supervision and the level of control in the home. This was not something where you either had engaged or disengaged parents, this was a complex and dynamic phenomenon varying according to socioeconomic status, the age of the child, knowledge and understanding of parent, family situation, trust and parenting styles. Furthermore, there was an observation that for some parents there was a threat dualism – some parents believed their child was perfect and would never be an abuser, but they might become a victim of a less "perfect" child.

When considering the knowledge of parents, the question was raised regarding who is responsible for growing parental knowledge and how can we be confident that the information they are getting is of good quality? There is little in emerging legislation that identifies a role for any stakeholder in public education, whereas there was a view that governments are best positioned to mandate formal and informal education approaches. Platforms that operate on a global scale are certainly able to provide materials and tools that allow parents to develop knowledge around their children's use of their platforms and broader online harms issues, but they could not mandate education. The dearth of legislation or regulation that aimed to mandate educational approaches, there was a view that a mix of civil society, NGOs and the private sector are providing resources instead, but there is no quality check on any of these approaches and it needs to be acknowledged that some organisations (for example ones who sell services or technology to ensure young people are "safe") might have a vested interest.

Returning to the issue of safety services and technologies, which we might broadly refer to as tools to support parental supervision, there was a general view that the focus on technical

solutions (which is arguably driven by policy direction) was problematic. While tools could be useful (for example a platform providing the means to complain about a post or individual, or tools the allow an end user to better manage the content they will see), there was agreement that parental controls in general give a false sense of security and the infamous "Ranum's law"[11] was referred to once again – you do not solve social problems with software.

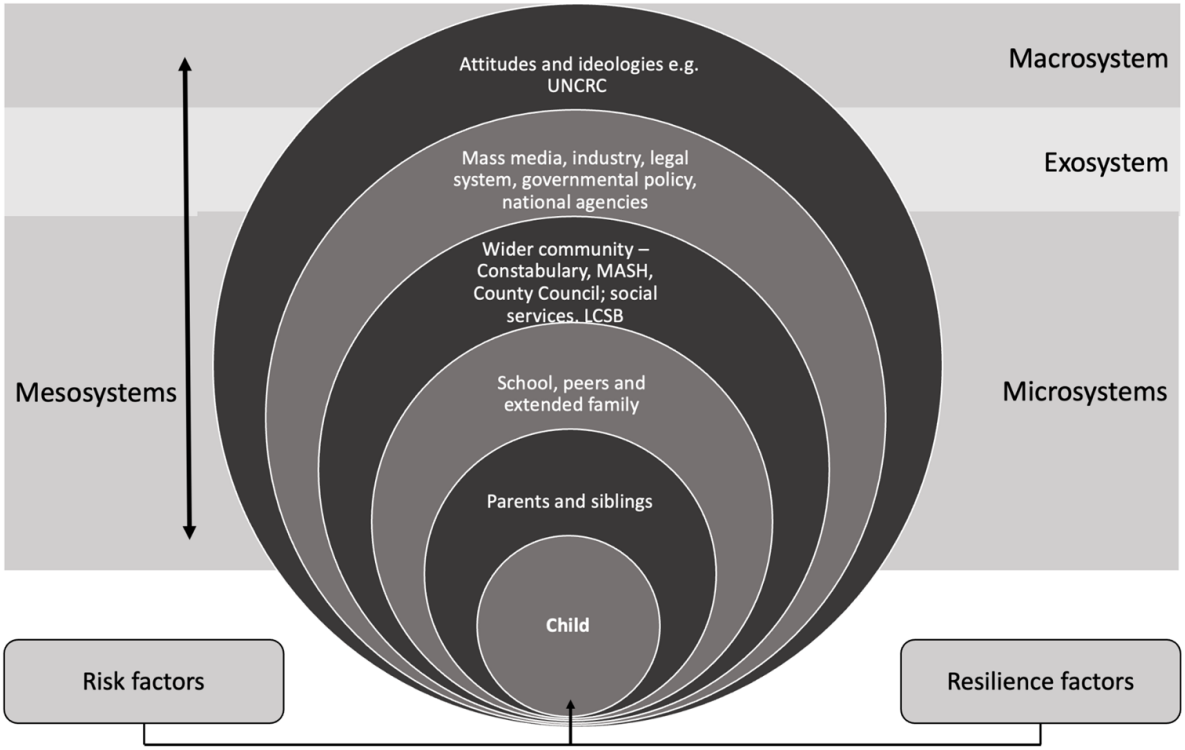
There was broad agreement that these tools that could be installed in the home (and on children's devices) were at best a blunt way to manage parent/child matters to mitigate online harms, and open communication was seen as a far more powerful approach. Furthermore, some tools have significant impacts on a young person's rights (such as privacy and access to information). Monitoring and tracking tools in particular were viewed as problematic.

Perhaps to most extreme example of this is the normalisation of tracking. It was observed that while a parent might view tracking as a means of ensuring their child is "safe" in reality all they know is where the child's device is. And a child who knows they are being tracked and wishes to do something they do not want their parents to know about will leave the device elsewhere.

The difference between consensual and non-consensual tracking was discussed and again highlighted the need for critical thinking when tackling online harms and what the goals of the technical intervention are. A view the tracking means a child safe is naïve, but a consensual discussion around why the family might like to see where each other are might be more progressive. Young people commonly track each other (for example via SnapMaps) and can see the positive value in tracking, but will also acknowledge that this too can be problematic (for example seeing all of your friends are at a party you were not invited to). Which, again, highlights the need to development knowledge and education around these issues, particularly related to rights and privacy.

[11] Cheswick, W. R., Bellovin, S. M., & Rubin, A. D. (2001). Firewalls and Internet security: repelling the wily hacker. Addison-Wesley Professional. pp. 202–. ISBN 978-0-201-63466-2.

Bringing Things Together



In the first session of these workshops, we presented the stakeholder model above. Adapted from Bronfenbrenner's Ecology of Childhood[12], this *Online Harms Ecosystem* attempts to show, through a clear model, that tackling online harms is not something that can be achieved by a single stakeholder. All stakeholders around the child have a role to play. What is clear from these discussions is that tackling online harms, from whatever stakeholder perspective, is not easy or straightforward. When this is considering the role of a particular technical intervention (age assurance), which has the potential to be an effective tool to support some goals of tackling online harms, it still requires its application in legislation and regulation to be conducted with the knowledge of its capabilities and to be mindful of the rights of citizens using online services.

If we are considering parental supervisions and the challenges therein, again, we can see from these discussions that, in supporting parents to make better choices around helping their children in tackling online harms, there is no simple approach.

In both use cases the importance of knowledge around rights is essential in both realising the potential of an approach and doing so in a manner that achieves positive outcomes for all users of online services. There should be no trade offs between an erosion of rights and intervention to ensure someone is "safe", as this is contradictory to the goals of rights-based approaches which have, almost universally, been agreed by nation states. While utilitarian approaches are generally viewed as the most realisable in the tech policy world, it was clear from our discussions that removing a right to privacy in order to reassure yourself that your

[12] Bronfenbrenner, U. (2000). Ecological systems theory. Oxford University Press.

child is safe, or preventing large numbers of end users from accessing services they are perfectly entitled to make use of, is no acceptable.

We remain convinced that a knowledgeable ecosystem is the starting point for progressive policy around online harms, and that is the responsibility of all stakeholders, from parents to policy makers.

In the next session we will, again through use cases, be developing the findings of this workshop (e.g., reliance on a single stakeholder cannot achieve policy goals) to see how we might get a more effective balance among stakeholders, whether current policy approaches established good foundations for this, and if not, what we can do to change.