NSPCC

Private messaging and the rollout of end-to-end encryption

The implications for child protection

NSPCC discussion paper

In recent years, we've seen online child abuse increase in its scale and complexity[1], with private messaging a major factor for online abuse.

There is an increasing push to roll out end-to-end encryption on messaging services, crucially including messaging functions that form part of, or are interoperable with, social networks.

End-to-end encryption (E2E) offers many benefits, but poses significant risks to children. Crucially, end-to-end encryption stands to significantly disrupt current mechanisms to detect and disrupt online abuse. At worst, E2E risks online services being able to 'engineer away' the ability of platforms to moderate content and prevent child abuse, substantially weakening the upstream response to online harms.

This discussion paper sets out NSPCC's position on end-to-end encryption. It aims to trigger renewed discussion about how and on what basis decisions about end-to-end encryption should be made, and it argues that any responses must place greater emphasis on children's needs (and the risks they face). This reflects not only their inherent vulnerability, but also that 1 in 3 global internet users are children.

In this discussion paper, we reflect the findings of research commissioned by NSPCC that sets out the contours of the current debate[2]. We also present survey data undertaken by YouGov that underlines strong public support for a more balanced settlement on end-to-end encryption.

NSPCC urges online services to only proceed with end-to-end encryption if and when they can demonstrate that children's safety won't be compromised as a result. Tech firms should be prepared to invest in engineering resource to ensure child abuse can continue to be detected in end-to-end encrypted products.

However, a broader range of responses, including legislative and regulatory action, are ultimately likely to be required.

This discussion paper:

- sets out NSPCC's concerns about the potential impacts that E2E could have on online child protection;
- argues that we should broaden the public policy and civil society discourse on the impacts of E2E, moving away from the often binary positions that typically position user safety and privacy as being in a fixed, unresolvable trade-off;
- explores the potential for a more balanced settlement that protects safety and maximises user privacy, reframing the issue to reflect the safety and privacy of all users, including children.

1 Reported files of abuse content made to the National Center for Missing and Exploited Children (NCMEC) have increased by 15,000 per cent in the past 15 years, and tripled between 2017 and 2019 alone. NCMEC acts as the global clearing house for child abuse reports. The road to safety: equipping industry to end CSAM. Thorn.

2 NSPCC commissioned research undertaken by PA Consulting. NSPCC (2021) End-to-end encryption: understanding the impacts for child safety online. London: NSPCC

Private messaging and child abuse risks

Private messaging provides children with an important space to communicate, socialise and keep in touch with their family and friends. During the COVID-19 pandemic, private messaging has enabled children to stay in touch during lockdowns and periods when in-person schooling has been interrupted.

It can however be readily exploited by abusers because it provides opportunities for abusers to contact children in a seemingly private space; to begin or escalate online grooming; and to produce or share child abuse images, including where children are coerced into sending self-generated photos or videos.

Often, abusers will be able to use other design features such as algorithmic friend requests to identify children prior to sending them private messages, or once a relationship has been established, may seek to escalate the grooming process by moving them to other platforms.

Children and young people have told us:[3]

'I'm in a serious situation that I want to get out of. I've been chatting with this guy who's like twice my age. This all started on Instagram but lately our chats have been on WhatsApp. He seemed really nice to begin with, but then he started making me do these things to 'prove my trust to him', like doing video chats with my chest exposed.' (Girl, aged 15)

'When I was 13, a man in his 30s contacted me on Facebook. I added him because you just used to add anyone on Facebook. He started messaging me and I liked the attention. We'd speak every day, usually late at night for hours at a time. We started using WhatsApp to message on. He started asking for photos and I sent some. Then he asked for some explicit photos and I did that too, and he reciprocated. He told me he'd spoken to other girls online and lied about his age to them, but he didn't lie to me so I felt like I could trust him (Frida, now aged 21)

'I've got a fitness page on Instagram to document my progress but I get a lot of direct messages from weird people. One guy said he pay me a lot of money to do a 'private show' for him. He now messages me almost every day asking for more explicit videos and I'm scared that if I don't do what he says, then he will leak the footage and my life would be ruined' (Boy, aged 17)

Recent data from the Office of National Statistics[4] sets out the broader scale and dimensions of the communication-based risks faced by children and young people. In the twelve months, up until March 2020 (the most recent period for which figures are available), ONS found:

- 17 per cent of children age 10 to 15 (estimated to be 682,000 young people) had spoken to someone online they hadn't met in person before in the previous 12 months. As the ONS correctly sets out, 'children may speak to people online who they have not met in person before to make new friends or join communities, which can have a positive impact on their lives.' However, this also presents significant risks, including children being coerced into sharing sensitive information and self-generated images;
- in 74 per cent of cases where children were messaged first by someone they hadn't met in person, they were initially contacted through private messages;
- 11 per cent of children aged 13 to 15 had received a sexual message in the previous 12 months, including 16 per cent of girls. Among these children, 84 per cent said they had received sexual content through direct messages, images or videos sent to them;
- Children also faced broader contact risks: over one in five children aged 10 to 15 (2 per cent) discussed meeting in person with someone they had first met online. 5 per cent of children went on to do so;
- 1 in 50 children (2 per cent) said they had spoken to someone in the previous 12 months who they thought was that age but later found out were much older. One in 25 children (4 per cent) have experienced this at some point in their lives.

The ONS suggests that, as a result of the probable increase in the time children have spent online in the last year, 'it is likely that the pandemic has had a substantial impact on the degree to which children are involved in these online activities.'

3 Experiences of children and young people shared with the NSPCC's Childline service. All names and potentially identifying details it been changed to protect the identity of the young person. Quotes are created from real Childline service users but are not necessarily direct quotes.

4 Office for National Statistics (2021) children's online behaviour in England and Wales: year ending March 2020. Newport: ONS

The current industry response

It is essential that online services take reasonable measures to identify and prevent both grooming and the production and distribution of child abuse images.

Through better design choices and the proactive use of technology, platforms can disrupt grooming pathways at the earliest possible stage, and prevent the potential for further upstream harm.

At present, most online services take a range of effective but proportionate measures to detect and disrupt child abuse taking place on their sites, including in private messages.

In most unencrypted environments, platforms use photo matching technologies to detect child abuse images, for example using Microsoft's Photo DNA product. These technologies are used to detect and 'hash' child abuse images (giving them a unique digital identifier) so if abusers attempt to upload photos, they can be rapidly identified and removed. Most major platforms have been able to remove an increasing proportion of child abuse images through such automated technologies.

It is important to stress that photo matching technologies only scan an image to determine whether it is a child abuse image, and for no other or more intrusive purposes. In effect, this process is no more intrusive than the use of spam filters.

Platforms are also able to use machine learning tools to identify new child abuse images, including Google's Content Safety API and Thorn's Safer product; and to detect grooming behaviour. These are developing but crucial technologies, with the rapid growth in self-generated content accounting for an increasing proportion of child abuse takedowns.[5]

In many cases platforms will only deploy grooming detection tools in private messaging where there is reasonable suspicion, assessed as a result of other activity taking place on the site.

For example, platforms are able to use metadata to detect suspicious patterns of behaviour that may indicate a grooming risk. If accounts demonstrate unusual patterns of behaviour, such as making a disproportionately high number of friend requests to children and young people, or where friend requests and contacts display clear age and geographical asymmetries, these could constitute reasonable grounds to investigate whether grooming is taking place.

Impact of end-to-end encryption on child abuse detection

In recent years, a growing number of messaging services have introduced end-to-end encryption, for example Apple's iMessage and WhatsApp.[6]

In 2019, Mark Zuckerberg announced that Facebook intended to introduce end-to-end encryption across its entire suite of messaging products. Facebook also intends that its messaging services should become interoperable, so end-to-end encrypted messages could be sent or received across WhatsApp, Facebook Messenger or Instagram's direct messages.

End-to-end encryption poses considerable child abuse risks when integrated into social networks, where abusers may be able to readily exploit other design features on the site to contact children and target them for sexual abuse.

Arguably, the risks of E2E are significantly heightened on social networks because of the central role they play in allowing abusers to identify and contact children, and to initiate well-established grooming pathways, in which children are subsequently abused through messaging or livestreaming products.[7]

Much of the discussion about negative impacts inevitably focuses on the significant challenge that end-to-end encryption presents for law enforcement. E2E is likely to result in considerable negative impacts for policing. For example, E2E is likely to reduce the ability of law enforcement to access evidence, undertake investigations and prosecute offenders that exploit online services to commit abuse.[8]

However, it is arguably in the immediate reduction of a platform's ability to moderate content, and the corresponding reduction in upstream threat capability, that the greatest impact of end-to-end encryption is likely to be felt.

The balance of these risks is often poorly reflected in the broader civil society debate.

Facebook's proposal to rollout end-to-end encryption is likely to lead to a significant reduction in its ability to detect online child abuse. Last year, the National Center for Missing and Exploited Children (NCMEC)

5 The Internet Watch Foundation reports that in 2020, 44 per cent of actioned content was self-generated. The number of reports tagged as including 'self-generated' child sexual abuse material increased 77% on 2019's total. IWF (2021) Call for experts to help tackle growing threat of 'self generated' online child sexual abuse material

6 Most recently, Google announced its intention to roll out end-to-end encryption in its Android messages app.

7 Europol (2020) Internet organised crime threat assessment. Lyon: Europol

8 Comments made by the FBI and Australian Federal Police in a panel session on end-to-end encryption, as part of the UN Crime Congress in Kyoto, March 2021

reported a record 21.7 million child abuse referrals, of which 20.3 million (93 per cent) came from Facebook platforms.[9]

Facebook deserves credit for the extent to which it has invested in proactive technology that enables it to detect child abuse on this scale. If anything, this data suggests that other platforms are failing to report abuse material on anything comparable to the likely scale of the problem on their sites.

However, if Facebook proceeds with its current E2E plans, much of its ability to detect and disrupt child abuse could be lost:

- Estimates suggest that up to 70 per cent of child abuse reports might no longer be generated, which in the U.K. alone, translates into actionable data that in 2018 led to 2,500 arrests and 3,000 children being safeguarded.[10]
- NSPCC analysis suggests that, in the most recent 12 months for which data is available, Facebook platforms were used in more than half of online child sexual abuse offences (where the platform was recorded.)[11] Facebook owned services were used in 4,903 offences during this period – and it is reasonable to assume many of these offences came to light as a result of actionable data from the company.

When Mark Zuckerberg initially announced Facebook's plans to proceed with end to end encryption, he recognised that the platform's services would be used for 'truly terrible things like child exploitation, terrorism and extortion' and spoke of 'an inherent trade-off because we will never find all of the potential harm we can today when our security systems can see the messages themselves.'[12]

In recent evidence to the parliamentary Home Affairs Select Committee, Facebook acknowledged that the introduction of E2E would lead to a fall in the number of child abuse reports they generate, but insisted they would push ahead anyway. In doing so, they cited an intention to meet an apparent 'industry standard.'[13]

WhatsApp's response to the child abuse threat

As it prepares to rollout end-to-end encryption, Facebook has repeatedly set out its ability to detect and disrupt child abuse on WhatsApp, which is already end-to-end encrypted.

In recent evidence to the Home Affairs Select Committee,[14] WhatsApp set out how it is able to detect child abuse, for example through using photo matching technology and classifiers on unencrypted surfaces, such as group names, photos and descriptions, to scan for abuse.

At present, WhatsApp claims to remove over 300,000 accounts per month for involvement in child sexual abuse and exploitation. However, this does not translate into a comparable number of reports to NCMEC. This is because, as the company's Director of European Public Policy Niamh Sweeney told the Home Affairs Select Committee; 'there would not be evidence against each of those accounts.'

WhatsApp is able to identify suspicious behaviour only where there are signals of illegal behaviour, or where users report it. However, because messages themselves are end-to-end encrypted, the service is significantly less able to generate actionable evidence for law enforcement.

In the UK, the National Crime Agency (NCA) reports that last year it received around 24,000 child abuse referrals from Facebook and Instagram, but only 308 from WhatsApp.[15]

WhatsApp therefore accounts for less than two per cent of Facebook referrals, despite the site being involved in one in ten instances recorded by police where Facebook's sites were used for child sexual abuse.[16]

9 NCMEC (March 2021.) ESP reports 2020. Washington, DC: NCMEC
10 Home Office (2019) Factsheet: encryption. London: Home Office
11 Figures released by NSPCC in March 2021, drawn from a Freedom of Information request to police forces in England and Wales.
12 Zuckerberg, M (2019) A privacy focus vision for social networking. Menlo Park: Facebook
13 Comments given by Monika Bickert, Facebook's VP of Global Policy Management, to the Home Affairs Select Committee's evidence session on online harms, 20th January 2021
14 Comments given by Niamh Sweeney, WhatsApp Europe Director of Public Policy, in front of the Home Affairs Select Committee, 20th January 2021
15 Hamilton, F; Knowles, T (2021) Facebook privacy plans will make it honeypot for child sex offenders'. London: The Times. Published January 26th 2021.
16 Based on data from an NSPCC Freedom of Information request to police forces in England and Wales on child sexual offences where the platform used to commit abuse was recorded. Figures released March 2021.

Reframing the relationship between privacy and safety

Despite the clear risks that end-to-end encryption poses to child abuse, it is important to recognise that for many it is also seen as a means to secure fundamental rights to privacy and freedom of expression.

End-to-end encryption unarguably offers benefits for persecuted groups, human rights advocates and dissidents in authoritarian regimes, and helps to maintain free expression through the protection of journalistic sources. E2E has therefore been particularly embraced by privacy and data rights activists.

As Ofcom data shows, privacy has become an increasing concern over recent years,[17] with public concern over how personal data is used for commercial purposes. Concern is likely to remain high, and particularly after the Cambridge Analytica allegations, tends to be focussed on how personal data is used for profiling purposes. There have also been a number of high profile hacking and security breaches affecting social networks and messaging services, largely involving commercial and profiled data.[18]

At the same time, companies such as Google have announced a significant refocusing of their advertising business to adopt more privacy preserving approaches.[19]

Facebook has stressed it does not scan or use the contents of private messaging for any commercial purposes. It has arguably sought to emphasise its commitment to privacy through the rollout of E2E, while at the same time actively resisting other privacy preserving approaches which are more intrinsic to its business model.[20]

With debates over the privacy, security and safety implications of end-to-end encryption well established, many positions have become entrenched. Discussions about end-to-end encryption often start with reference to an apparent trade-off between privacy and user safety. The argument suggests that it is a zero-sum game – that dialling up user safety somehow means dialling down or eliminating user privacy altogether.

This tendency towards absolutist positions has arguably been reinforced by some tech companies, who have sought to frame E2E as a simplistic trade-off[21]; suggest that any discussion of E2E safeguards is an existential threat to online privacy; or consistently seek to focus the debate on the implications for law enforcement, rather than the impact on their own ability to detect and disrupt harmful content.[22]

Such framing arguably acts to undermine attempts at interventions to secure the safety of users, and legitimate debates about what form this should take.[23]

Public views on the rollout of E2E and private messaging

Polling conducted for NSPCC, by YouGov, which surveyed 2,125 adults in December 2020 and January 2021,[24] shows there is a strong public interest in reaching a more appropriate balance between child safety and privacy.

Our findings suggest that public opinion is altogether more balanced than the approaches currently being taken by some tech firms, and that is sometimes projected in the dynamics of public policy discussions on E2E.

17 Ofcom (2020) Online Nation: 2020 Report. London: Ofcom.

18 For example, as Ofcom's Online Nation report sets out (ibid), during 2019 there were three major security incidents involving WhatsApp, including one incident where it was claimed a security flaw would allow 'hackers to intercept media files being sent between users and potentially alter them'. Most recently, personal data of 533 million Facebook users has been compromised, which Facebook claims was scraped from the site but this is contested by cybersecurity experts. Manancourt, F (2021) 'Misleading' Facebook data claims questioned. Brussels: Politico

19 In early 2021, Google published its Privacy Sandbox proposals, including the phasing out of third party tracking cookies and the introduction of more privacy preserving methods, including the use of Federated Learning of Cohorts (FLoC). For a summary of Google's position and criticisms from privacy advocates, see Oremus, W (2021) Pattern Matching: a newsletter by One Zero. San Francisco: Medium.

20 For example, Facebook has actively pushed back on constraints to its profiling of users through cookies on third party sites, for the purposes of targeted advertising (which is a core part of Facebook's business model.)

21 Earlier this month Facebook's VP of User Integrity Guy Rosen wrote that in respect of end to end encryption the 'stakes are not just a matter of personal, financial or reputational risk for the few.' and that 'if nothing online is private, and every conversation today is online, then no conversation is private. Either we communicate face-to-face, or we surrender any expectation that we're alone.' Rosen, G (2021) *Encryption has never been more essential – or threatened*. Op-ed published in Wired

22 Inevitably focussing the discussion on so-called 'backdoors', rather than the proposals for lawful access proposed by the UK National Cybersecurity Centre

23

24 Polling undertaken by YouGov for NSPCC. Total sample size was 2125 adults, fieldwork was undertaken between 31 December 2020 – 4 January 2021. The survey was carried out online. Figures have been weighted and are representative of all UK adults (aged 18+)

Our findings suggest:

- substantive public concern about the impacts of E2E on child abuse detection, but also strong support for a settlement that balances safety and privacy issues;
- a real incentive for platforms to introduce end-to-end encryption with child protection safeguards in place; and,
- a number of areas where governments and civil society groups could usefully inform public understanding about the nature of current responses to the child abuse threat, in order to inform discussions about the balance between proportionality and user privacy.

There is strong concern about online child abuse, which translates into support for legislative and technical measures: in the survey, we found broad public concern about the risks of online abuse against children, with 87 per cent of UK adults concerned about online grooming, and 86 per cent concerned about inappropriate contact between children and adults.

This translates into very high levels of support for legal requirements on social networks and messaging to detect and disrupt abuse on their services:

- 90 per cent support social networks and messaging sites having a legal requirement to detect child abuse on their services;
- 92 per cent support social networks and messaging services having a technical ability to detect child abuse images; and,
- 91 per cent support a technical ability to detect adults sending sexual images to children.

Limited awareness of the current child abuse response: Although there is strong support for platforms adopting technical measures to detect and disrupt online abuse, there is limited public awareness of how industry currently tackles abuse.

Just under half (49 per cent) of UK adults think that social networks and messaging services detect child abuse images on their sites. However, almost a quarter (24 per cent) think that social networks and messaging services never detect abuse. A slightly larger proportion (27 per cent) are unsure.

This suggests there is an important exercise needed to raise awareness of the current nature of the child abuse threat response, including the nature and extent of the often sophisticated approaches undertaken by larger firms.

Governments and civil society groups should do more to set out what tech firms currently do to protect child users – and by implication, what could be lost if platforms were to roll out end-to-end encryption before demonstrating they had effective child safety safeguards in place.

Strong public support for a balanced E2E settlement: Public support for end-to-end encryption grows considerably if platforms are able to demonstrate they have effective child safety measures in place.

At present, there is narrow support for Facebook to use end-to-end encryption on its services. 40 per cent support the proposals, and 35 per cent oppose them. A further quarter of respondents (25 per cent) said they were unsure. However, support for Facebook's plans grows markedly if the company proceeds with end-to-end encryption, once there are appropriate child safety safeguards built in.

If Facebook could prove children were at no greater risk than they are currently, almost three fifths of respondents (57 per cent) back its proposed rollout of end-to-end encryption. Less than one in five (19 per cent) would oppose it.

Our survey found a broadly similar pattern for other companies that might be looking to rollout end-to-end encryption, including Google.[25]

If platforms are able to demonstrate that children's safety is protected, support for the rollout of end-to-end encryption almost doubles, from 33 per cent to 62 per cent of respondents. This underlines a clear incentive for tech firms to proceed with end-to-end encryption only once there are appropriate child protection arrangements in place.

As a number of tech firms continue to develop their rollout of E2E, it suggests that platforms should be investing considerable engineering resource to develop new child safety solutions, and ensure existing threat responses can be adapted to work in end-to-end encrypted environments.

25 Google has announced plans to introduce end-to-end encryption on its RCS Android Messages app

Support for a balanced approach to safety and privacy: there is considerable public support for a balanced approach to scanning for child abuse threats on social networks and messaging services, with a majority of respondents recognising the importance of detecting online abuse.

When respondents were asked to decide whether the right to privacy or the ability to identify child abuse images were more important:

- More than half (55 per cent) said the ability to detect child abuse images was more important than the right to privacy;
- One third (32 per cent) said the right to privacy and the ability to detect child abuse images were equally important;
- 4 per cent said that the right to privacy was more important.

This suggests there is considerable public consensus for both safety and privacy outcomes to be considered in the development of end-to-end encrypted environments, and in any legislative or regulatory proposals.

Support for a proportionate basis for scanning: Our findings suggest there is significant public support for a proportionate approach to scanning of private messages, but that government and civil society need to do more to inform the debate about what constitutes proportionate versus more intrusive forms of activity.

When respondents were asked their views on what types of content should be scanned on social networks and messaging services to detect child abuse:

- 40 per cent said both text and images in direct messages should be scanned;
- 16 per cent said that images should be scanned, but text should remain private;
- 13 per cent said that it should never be permitted for either text or images in private messages to be scanned;
- three in ten (29 per cent) didn't know.

Among respondents who support the practice of scanning text and images, 70 per cent support this where there is reasonable suspicion of criminal activity, and 26 per cent support it in all instances.

This suggests that governments, civil society groups and tech firms need to do more to set out the current arrangements for detecting child abuse images, where all images are scanned using photo matching technology for the sole purpose of assessing whether it contains known child abuse imagery.

Tech firms and child safety advocates should stress that the current use of PhotoDNA and other photo matching tools is proportionate, and no more invasive than the use of tools such as spam filters. Similarly, child safety advocates should actively support more invasive approaches only where there are appropriate safeguards in place.

In reaching a balanced settlement on the proactive detection of child abuse images, debates over proportionality need to reflect the parameters in which scanning takes place; set out the purposes of proactive scanning (including whether data could be used for other purposes); and reach consensus that scanning should always be no more invasive than a proportionate threat response demands.

Broadening the terms of the end-to-end encryption debate

NSPCC wants to see a broader public policy and civil society discourse on end-to-end encryption, moving away from binary and absolutist positions and towards a more balanced settlement that protects user safety and maximises privacy.

Any responses to the challenges of end-to-end encryption, whether coming from tech firms, governments or regulators, need to be driven by highly informed, nuanced public policy discussions.

But end-to-end encryption is also a societal issue, not just a technology one. As a result, the debate must involve a discussion of protecting children's best interests and rights. To do this, industry and policy responses must be informed by the voices of users – including those of marginalised and vulnerable groups that often don't get a seat at the policy making table or in the C-suites of Silicon Valley.

Civil society groups and governments should be prepared to lead such dialogue – and tech companies should join it in good faith – to facilitate a nuanced and thoughtful discussion which sets out how best to proceed with any E2E rollout, and what safeguards are legitimate to put in place.

The intention should be to flesh out a way ahead which informs the responses of the tech firms themselves and, in turn, the legal and regulatory routes that are necessary to embed user safety and privacy.

A broader debate on end-to-end encryption that allows us to reach a balanced settlement on both safety and privacy must:

1. Consider the needs of all users, including children

It's vital that discussions on end to end encryption consider the needs of all users, including children. All too often children have been left out of, or underserved by, discussions on internet governance.

However, it's essential that children's needs are seen as a valid and central part of this debate – not only because of the vulnerabilities they face online, but because they are a significant constituency of internet users in their own right.[26]

Too often, end-to-end encryption is seen purely in terms of adults, and so the discussion tends to focus on issues of personal privacy above all else. Where children do feature in the debate, children's views are often typically seen as being diametrically opposed to those of adults; or worse, it is sometimes implied that child abuse concerns are introduced primarily as a subtext to deliver broader interventions.[27]

We will only be able to deliver a balanced settlement, and deliver outcomes that protect the interests of all internet users, if we are able to move beyond such positions.

2. Understand the interplays between children's safety and privacy

Arguments that characterise end to end encryption as a trade-off between adult privacy and child safety are often unhelpfully simplistic. In reality, the rollout of end-to-end encryption presents a complex set of interplays between privacy and safety that should be carefully balanced to ensure corporate and government actions are taken in the best interests of the child.

For example, although end-to-end encryption may offer children immediate privacy benefits, there is also a risk that it may significantly weaken privacy among the most vulnerable of children, including those children that have been sexually abused.

If end-to-end encryption were to significantly frustrate current efforts to detect and remove child abuse material, it is likely that abuse images and videos would be able to circulate much more freely on private messaging services. In turn, the children abused in these photos and videos are likely to face significant and continuing re-victimisation.

The victims seen in this material continue to suffer, with some reported to spend huge amounts of time searching for and reporting their own pictures. In the Canadian Centre for Child Protection's Survivor Survey, 20 per cent of respondents reported having been identified by someone who had seen their abuse online.[28]

If platforms were to introduce end-to-end encryption with appropriate safeguards in place, it could be possible to both protect the safety of children (including child abuse survivors), and maximise the privacy benefits to other vulnerable groups, for example LGBTQ+ young people.

3. Reflect children's digital rights under international law

It is essential that decisions on the future rollout of end to end encryption fully take account of the specific protections available to children and young people under international law. General Comment 25 sets out that children's rights, as set out in the UN Convention on the Rights of the Child, expressly apply in the digital world.

As 5Rights note, 'there has been considerable tension between those protecting adult privacy particularly as it regards their privacy from the state, and the abuse of that privacy by those who spread and or consume child sexual abuse material.'[29]

However, the privacy offered to users by private companies must not protect or enable those who commit child abuse or consume child abuse content; nor interfere with mechanisms to detect and disrupt child abuse taking place on online services.

In addition, the European Convention on Human Rights (ECHR) implies positive obligations on the state to take proactive measures to protect children from abuse, which are grounded in the prohibition

26 Data from the Information Commissioner's Office suggests that 1 in 5 internet users are children. Globally, this rises to 1 in 3. ICO (2020) Age appropriate design: a code of practice for online services.

27 See for example, Pfefferkorn, R (2019) Banning strong encryption does not mean you catching criminals. It only makes you less safe from them. Blog posted on the Center for Internet and Society blog, Stanford University. Hruska, J (2020) Congress floats spectre of child exploitation to kill legal encryption. Article posted on extreme tech.com

28 Canadian Centre for Child Protection (2017) Survivors Survey. Winnipeg: CCCP.

29 5Rights Foundation (2021) Explanatory notes on General Comment 25. London: 5Rights Foundation.

of inhuman or degrading treatments and the right to privacy.

Under article 8, states have a positive obligation to secure the physical and psychological integrity of an individual from other persons.[30] This applies particularly to the well-being of vulnerable groups, and in order to protect their rights to a private life, includes the protection of a child from physical and mental harm.[31] In a scenario where images of a young person's abuse are allowed to remain in circulation on platforms, and this causes acute psychological distress or harm to the child, questions would arise about compatibility with Article 3 (which includes the right not to be subjected to inhuman or degrading treatment.)

This has significant implications for, and arguably recasts, the end to end encryption debate. Whereas Government intervention has typically been described as 'encryption coming under attack,[32] states are under a duty to take action to address known risks and ensure adequate legal structures and sanctions are in place to protect children from sexual abuse and broader harms.[33]

If states need to take steps to fulfil their positive obligations to protect children from abuse that occurs online, it is essential we have a measured debate about what are considered proportionate and commensurate interventions, particularly in the context of end-to-end encryption.

4. Move towards a more inclusive adoption of human rights principles

In recent years, there has been a move towards the adoption of a human-rights based framework to underpin the content moderation strategies of major platforms.[34]

This is hugely welcome, although there remains a tendency for some companies to position their approaches in an interpretation of human rights frameworks in ways that are advantageous or reinforce their existing approaches to certain issues, including end-to-end encryption.[35]

The NSPCC supports the adoption of human rights frameworks, but it is vital that the position taken in response to end to end encryption fully reflects the range of rights at stake.

Any assessment of commercial decisions against a human rights framework must appropriately balance the range of fundamental rights at stake; including freedom of expression, privacy and safety. It must also carefully assess the impacts of policy decisions on particular groups of users, including children.

In March 2021, Facebook published its Corporate Human Rights Policy,[36] drawing heavily on the UN's Guiding Principles on Business and Human Rights. As part of that policy, Facebook commits to 'depending on the circumstances [...] utilise other widely accepted human rights instruments, including the UN Convention on the Rights of the Child.' . It is unclear how Facebook intends to utilise the UNCRC in its decision-making on the further rollout of end to end encryption.

However, what is clear is the policy seems to privilege certain rights over others in its description of how E2E is already used in relation to WhatsApp. End-to-end encryption is actively used as an exemplar of how human rights are promoted by the company:

> *'by upholding the privacy and security of people's messages via end to end encryption – so that only the people who are communicating with each other can read or listen to what is sent, we help protect the most vulnerable groups from surveillance and abuse.'*[37]

The policy concludes that end to end encryption enables Facebook to deliver 'privacy as an enabling right, which underpins freedom of expression, freedom of association, and the safeguarding of human dignity.' However, no reference is given to the safety implications of end-to-end encryption, including the adverse implications for child abuse.

30 European Court of Human Rights (2020) Guide to article 8: right to respect for private and family life, home and correspondence. Strasbourg: ECHR.
31 KU vs Finland. European Court of Human Rights (2015) Internet case law and the ECHR. Strasbourg: ECHR.
32 Burns, H (2020) Online Harms: encryption under attack. London: Open Rights Group
33 O'Keefe vs Ireland. European Court of Human Rights, Grand Chamber. Application no: 35810/09, 28th January 2014
34 A human-rights based approach was strongly advocated by David Kaye during his tenure as UN Special Rapporteur on freedom of expression. See for example Sander, B (2020) Freedom of expression in the age of online platforms; the promise and pitfalls of a human-rights based approach to content moderation. Fordham International Law Journal 55. Fordham, NY: Fordham University
35 Evelyn Douek has raised concerns that the adoption by tech firms of international human rights law as a framework for their content moderation strategies could represent 'bluewashing' – with 'companies prepared to wrap themselves in the language of human rights, co-opting its legitimacy at little cost. She suggests that as non-binding norms, 'there is no mechanism to force a company into compliance.' Douek, E (2020) The limits of international law in content moderation. UCI Journal of International, Transactional and Comparative Law, forthcoming.
36 Facebook (2021) Corporate human rights policy. Menlo Park, CA: Facebook
37 ibid

5. Emphasise the impact of end-to-end encryption on specific services

In order to secure a more balanced debate, it may increasingly be helpful to shift away from an overarching focus on the impacts of end-to end-encryption, towards a focus on the particular risk profile associated with specific services.

The NSPCC is particularly concerned about the rollout of E2E on messaging functions that form part of, or become interoperable with, social networks.

This reflects that the risk profile will be determined by a range of factors, including:

- The interplay between end-to-end encrypted services and other design features. For example, grooming risks may be significantly increased if abusers are able to take advantage of algorithmic friend suggestions to contact large numbers of children at scale, and in turn, to expedite grooming pathways;
- End-to-end encryption being bundled with other high risk design choices, for example WhatsApp's proposals to auto-delete all messages by default. Europol has cited this design feature as being particularly problematic for child abuse detection,[38] and;
- Multiple design features being placed under a single end-to-end encrypted cloak. Under such circumstances, groomers could potentially message a child and then coerce them into producing self-generated material on video chats, without a platform being able to identify or disrupt this abuse at any stage of the grooming process.[39]

A greater focus on risk assessment should enable emphasis to be placed on the actual potential for harm to be caused. This could be undertaken by platforms themselves, but arguably should form an essential requirement of any legislative and regulatory response.

Next steps and conclusions

To achieve a broader discourse on end-to-end encryption and arrive at a balanced set of public policy and technological responses, governments and child safety advocates should:

- Actively resist the overly simplistic framing of end-to-end encryption as a fixed trade-off between safety and privacy;
- Do more to reframe the terms of the debate, arguing unapologetically that decisions on end-to-end encryption must take into account the best interests of children, not only because of their inherent vulnerability, but because they are a significant constituency of internet users in their own right;
- Push for decisions on end-to-end encryption to have due regard to the range of fundamental rights at stake (rather than, as seems to happen now, privileging certain rights over others). In the very welcome move towards the technology industry's adoption of a human rights decision-making framework, children often seem to be left behind;
- Lead a broad dialogue with tech firms to establish how best to proceed with end-to-end encryption in a way that reflects the needs of all users, including a discussion of the technical, legal and regulatory safeguards that are needed to protect them.

Our polling data demonstrates there is strong public support for a balanced settlement that reflects the full complexity of the issues, and that doesn't reduce the contours of decision-making to an unhelpful zero-sum game.

The public want tech firms to introduce end-to-end encryption in a way that maximises user privacy and the safety of vulnerable users. Indeed, if platforms can demonstrate that children's safety will be protected, there is significant support for end-to-end encryption to go ahead – a clear incentive for tech firms to invest the necessary engineering resource to ensure child abuse threat responses can continue to work in end-to-end encrypted products.

38 Europol (2020) Internet organised crime assessment. The Hague: Europol.

39 In this respect, Facebook's proposal to end-to-end encrypt both its private messages and Messenger Rooms product represents a hugely problematic product offer. Facebook Rooms allows up to 50 participants to join a call, who do not need Facebook accounts to join

But it is precisely because end-to-end encryption involves big societal issues that it is ultimately appropriate for governments to set the guardrails, to ensure that decisions made by tech companies protect the needs of vulnerable users.

Legislative and regulatory requirements should be nuanced and proportionate – but just as it is important that there are requirements on companies to detect child abuse through the most minimally invasive of means, it is similarly reasonable to anticipate that companies should not be able to 'engineer away' their abilities to protect child users altogether.

End-to-end encryption is ultimately a design choice, much like any other. Across the world, there is growing understanding that it is reasonable to expect companies to design their products with children's safety in mind.

In the UK, the NSPCC have been advocates of a Duty of Care – an overarching requirement to require online services to identify reasonably foreseeable risks to children, and to take reasonable and proportionate steps to mitigate them. End-to-end encryption is a test of whether tech companies are prepared to move towards a more responsive and systemic approach, that hardwires children's safety into their products and wider corporate decision making.

Ultimately, end-to-end encryption is a child protection issue, and it deserves to be considered as such.

The NSPCC urges tech companies to refocus their approach to end-to-end encryption – recognising that the privacy and safety of all users should be maximised, including children and young people who are so frequently poorly served by decisions taken about the products they use.

But if tech companies will not protect the needs of their child users, and work towards a balanced settlement, it is entirely appropriate that legislative and regulatory remedies should be used to secure it – with companies required to ensure their upstream capability to detect and disrupt child abuse is not lost.

There is a particular opportunity for the UK to drive international consensus on the issue, including during its ongoing Presidency of the G7. Through the introduction of proportionate but child-centred regulation, that reflects the contours of the re-framed debate as set out in this discussion paper, the UK can create a global model for how we create legal and regulatory safeguards to achieve the optimal balance between safety and privacy.

NSPCC

Everyone who comes into contact with children and young people has a responsibility to keep them safe. At the NSPCC, we help individuals and organisations to do this.

We provide a range of online and face-to-face training courses. We keep you up-to-date with the latest child protection policy, practice and research and help you to understand and respond to your safeguarding challenges. And we share our knowledge of what works to help you deliver services for children and families.

It means together we can help children who've been abused to rebuild their lives. Together we can protect children at risk. And, together, we can find the best ways of preventing child abuse from ever happening.

But it's only with your support, working together, that we can be here to make children safer right across the UK.

nspcc.org.uk

EVERY CHILDHOOD IS WORTH FIGHTING FOR