

## **Submission to the Consultation on the Online Harms White Paper**

The Counter Extremism Project (CEP) commends the Department for Digital, Culture, Media and Sport, and the Home Office, for the opportunity to comment on their proposals to make the internet a safer space. This is a great challenge but one that is essential to meet in order to keep people safe, and defend the liberal and democratic principles and values that underpin British parliamentary democracy.

CEP is a not-for-profit, non-partisan, international policy organisation formed to combat the growing threat from extremist ideologies. Led by former world leaders and diplomats, including Senator Joseph I. Lieberman, former US Homeland Security Advisor Frances F. Townsend, and Ambassador Mark D. Wallace. It combats extremism by disrupting extremists' financial, recruitment, and propaganda networks online — including through the development of technological tools — and advocating for improved legislation and policy change. CEP senior advisor, Dr. Hany Farid, developed PhotoDNA technology to combat child sexual exploitation, which is mentioned on page 80 of the Online Harms White Paper.

CEP will address four consultation questions in this submission: question 1, regarding transparency; question 7, regarding private communications online; question 13, regarding UK or EEA representatives; and question 15, regarding technology and innovation. The recommendations made here are based on recommendations made by CEP to the EU and US regulators, and are the result of evidence on online extremism, terrorism, and radicalisation, gathered by CEP.

Question 1: This government has committed to annual transparency reporting. Beyond the measures set out in this White Paper, should the government do more to build a culture of transparency, trust and accountability across industry and, if so, what?

CEP has carried out extensive research into online extremism. This was based on evidence using a unique system developed by CEP and Dr. Hany Farid, called eGlyph – a tool based on ‘robust hashing’ algorithms, and which is capable of detecting known extremist images, videos, and audio files. The evidence will be presented first, followed by recommendations to the regulator. An analysis of the findings leads to recommendations with regard to transparency, trust, and accountability.

Studies by CEP have identified a widespread use of social media platforms by groups or individuals affiliated with extremist organisations. This has played a significant role in the radicalisation of people, some of whom have committed, or attempted to commit, acts of terrorism, or have subsequently joined extremist organisations. The radicalising content produced by extremist groups includes propaganda videos, some showing explicit violence, as well as videos designed to win support for an ideology, and to incite against a common ‘enemy’ (for example the far-right often incites against immigrants, Muslims, and Jews, while Islamist extremists incite against the West or non-Muslims). Speeches, lectures, and sermons by charismatic leaders and recruiters are also widely used, as are videos of demonstrations and marches that show the popularity of the group. There is also a growing use of smaller platforms, closed forums or encrypted messaging platforms for recruitment and discussing terrorist activity in a relatively smaller and low-profile setting.<sup>1</sup> In addition to recruitment, there are online videos and manuals with instructions for carrying out attacks, building bombs, joining extremist groups abroad as fighters, and more operational information.

---

<sup>1</sup> Echison, E. and Knodt, O. Germany’s NetzDG: A Key Test for Combatting Online Hate. *Counter Extremism Project*, November 2018, [https://www.counterextremism.com/sites/default/files/CEP-CEPS\\_Germany%27s%20NetzDG\\_020119.pdf](https://www.counterextremism.com/sites/default/files/CEP-CEPS_Germany%27s%20NetzDG_020119.pdf), P.15; and ‘Extremists exploiting small social media websites, experts warn,’ *BBC News*, 1 September 2018, <https://www.bbc.co.uk/news/uk-wales-45341746>.

A study by CEP looked at 168 individuals who consumed official terrorist propaganda materials online, including on Facebook, YouTube, Twitter, and WhatsApp. Of those profiled, 26 subsequently carried out terror attacks; at least 52 others have attempted to carry out or facilitate attacks; and 57 individuals have attempted to become foreign fighters for an extremist group, with at least 16 of them succeeding.<sup>2</sup>

### ISIS Content on YouTube

A recent study by CEP, which used eGlyph and a web crawler to search for video titles and keywords in videos uploaded to YouTube, found that hundreds of ISIS propaganda videos have been uploaded to the popular video sharing platform between March and June 2018, gathering thousands of views — despite YouTube’s purported content removal efforts.<sup>3</sup> According to the report, which was co-authored by CEP Senior Advisor and UC Berkeley Professor, Dr. Hany Farid, the 1,348 ISIS videos uploaded gathered 163,391 views. This happened even though the majority of videos (76%) remained online for less than two hours before being removed.<sup>4</sup> Despite uploading extremist content, the study also found that 60% of the accounts remained live on YouTube even after the videos have been removed for content violation. Moreover, the study found that 91% of all uploaded videos were uploaded more than once, meaning that YouTube’s hashing systems failed to prevent the re-upload of known terrorist videos.<sup>5</sup>

Based on the findings, CEP recommended that YouTube “should provide clear policy guidelines and take action to consistently and immediately delete accounts that have uploaded ISIS videos. Users should not be allowed to maintain their YouTube account after posting terrorist

---

<sup>2</sup> Extremists and Online Propaganda. *Counter Extremism Project*, April 2018, [https://www.counterextremism.com/sites/default/files/Extremists%20and%20Online%20Propaganda\\_040918.pdf](https://www.counterextremism.com/sites/default/files/Extremists%20and%20Online%20Propaganda_040918.pdf), P.1.

<sup>3</sup> The eGlyph Web Crawler: ISIS Content on YouTube. *Counter Extremism Project*, July 2018, [https://www.counterextremism.com/sites/default/files/eGLYPH\\_web\\_crawler\\_white\\_paper\\_July\\_2018.pdf](https://www.counterextremism.com/sites/default/files/eGLYPH_web_crawler_white_paper_July_2018.pdf).

<sup>4</sup> Ibid., P.1.

<sup>5</sup> Ibid.

content.”<sup>6</sup> The company should also be transparent about how it is implementing hashing technology to find terrorist content, as well as how much data has been removed as a result of a shared database maintained by the Global Internet forum to Counter Terrorism (GIFCT).<sup>7</sup>

### ISIS Support Network on Facebook

A study conducted by CEP, and published in May 2018, found that international ISIS-supporting profiles are commonly found on Facebook, and include activities such as disseminating propaganda, recruitment, and discussing terrorist activity.<sup>8</sup> 1,000 pro-ISIS profiles were found on Facebook in October 2017. By March 2018, 57% of them were still on Facebook, highlighting a failure by Facebook to effectively prevent the spread of ISIS and pro-ISIS content on their platform.<sup>9</sup> During that time, CEP found that official ISIS videos have been viewed thousands of times before being removed. The report also flagged that ISIS supporters used the platform’s livestreaming service, Facebook Live, to host meetings, and linking to banned materials in the livestream’s comments. This practice helped pro-ISIS users avoid Facebook’s automated flagging tools.<sup>10</sup> Additionally, Facebook’s algorithms are designed to connect people who share common interests. Researchers of the CEP study saw that Facebook’s ‘suggested friends’ feature helped introduce ISIS supporters to

---

<sup>6</sup> Ibid., P.8.

<sup>7</sup> Ibid.

<sup>8</sup> Waters, G. and Postings, R. Spiders of the Caliphate: Mapping the Islamic State’s Global Support Network on Facebook. *Counter Extremism Project*, May 2018, <https://www.counterextremism.com/sites/default/files/Spiders%20of%20the%20Caliphate%20%28May%202018%29.pdf>, P.7.

<sup>9</sup> Ibid., P.8.

<sup>10</sup> Gilberg, D. American ISIS Supporters Are Organizing On Facebook. Here’s How. *Vice News*, 22 May, 2018, [https://news.vice.com/en\\_us/article/59qgaz/american-isis-supporters-are-organizing-on-facebook-heres-how](https://news.vice.com/en_us/article/59qgaz/american-isis-supporters-are-organizing-on-facebook-heres-how).

each other on a routine basis.<sup>11</sup> The location of those ISIS supporters were most commonly from Southeast Asia or Iraq and Syria (17% each), and 6% have been from Europe.<sup>12</sup>

The report concluded that: “Our analysis of online IS communities globally, regionally, and nationally suggests that IS’s online networks, in particular on Facebook, are growing and can be utilized to plan and direct terror attacks as well as mobilize foreign fighters for multiple areas of insurgency. Secondly, IS’s presence on Facebook is pervasive and professionalized, contrary to the tech company’s rhetoric and efforts to convince the public, policymakers, and corporate advertisers from believing otherwise.”<sup>13</sup> It added that, “Facebook remains an important tool for IS supporters and members to spread its propaganda, radicalize others, and recruit new members. This research project revealed Facebook’s ongoing inability to address IS content on its site in a manner that is comprehensive, consistent, and transparent.”<sup>14</sup>

The same study also found that, in a move that can greatly reduce public confidence that the company is trying to keep users safe, profiles were reinstated by Facebook after being suspended,<sup>15</sup> allowing those users to once again disseminate extremist propaganda.

### Far-Right Online Propaganda

CEP has been following far-right, ethno-nationalist, and white-supremacist groups in Europe. It gathered evidence to show that such groups continue to thrive in Europe, and that: “Their propaganda campaigns have allowed them to generate substantial popular support and make gains

---

<sup>11</sup> Evans, M. Facebook accused of introducing extremists to one another through ‘suggested friends’ feature. *The Telegraph*, 5 May, 2018, [https://www.telegraph.co.uk/news/2018/05/05/facebook-accused-introducing-extremists-one-another-suggested/amp/?\\_\\_twitter\\_impression=true](https://www.telegraph.co.uk/news/2018/05/05/facebook-accused-introducing-extremists-one-another-suggested/amp/?__twitter_impression=true).

<sup>12</sup> Waters, G. And Postings, R. Spiders of the Caliphate: Mapping the Islamic State’s Global Support Network on Facebook. *Counter Extremism Project*, May 2018, <https://www.counterextremism.com/sites/default/files/Spiders%20of%20the%20Caliphate%20%28May%202018%29.pdf>, P.10.

<sup>13</sup> Ibid., P.3.

<sup>14</sup> Ibid., P.74.

<sup>15</sup> Ibid., P.76.

in domestic elections. [for example] The AfD came in third in Germany's September 2017 parliamentary elections."<sup>16</sup>

In the UK, CEP has been following the online and offline presence of Combat 18, the UK-founded international neo-Nazi group. Combat 18 has been gathering support by disseminating information using their propaganda magazine, flyers, own websites, and social media platforms such as Facebook and the Russian social network VK – which has been used to promote their ideology and violent activities.<sup>17</sup> CEP has also been following the activities of the English Defence League (EDL), which has a wide social media presence. Despite having some of their accounts suspended on Twitter, EDL was still active on Facebook until after the Christchurch attacks.<sup>18</sup>

Evidence gathered by CEP on online propaganda from the far-right proscribed organisation National Action reveal that despite being a banned organisation, National Action has maintained a robust online presence which is easily accessible.<sup>19</sup> CEP researchers searched websites such as Facebook, Twitter, Vimeo, BitChute, and VK, and found, on 3 October 2018, 44 examples of information that had been uploaded up to a year earlier, was still present online at the time. These included videos taken during marches and protests, blogs, and propaganda videos made by National Action. CEP research also noted that this material was present online even though National Action was recently banned from major platforms.

CEP also found that YouTube, despite having banned neo-Nazi group Atomwaffen Division (AWD) in February 2018, allowed supporters to re-upload known AWD content. In January 2019, CEP researchers found a YouTube channel that uploaded 15 AWD videos with a total of nearly

---

<sup>16</sup> European Ethno-Nationalist and White Supremacy Groups. *Counter Extremism Project*, December 2018, <https://www.counterextremism.com/european-white-supremacy-groups>.

<sup>17</sup> Ibid., P.10.

<sup>18</sup> Ibid., P.16.

<sup>19</sup> National Action Online Propaganda Progress Report. *Counter Extremism Project*, October 2018. Unpublished.

7,000 views.<sup>20</sup> Again in April, CEP researchers reported additional AWD propaganda videos for violating YouTube's hateful and abusive content policies. However, the videos were still available 48 hours after CEP reported them to YouTube.<sup>21</sup> Even though the company had banned this group, YouTube is failing to prevent re-uploads or act with urgency to remove the content. YouTube owes lawmakers and the public detailed explanations about their content moderation process.

Large social media companies have consistently failed to remove extremist content as well as users who advocated for violent extremism. Dr. Farid stated that they “have time and again failed to recognize misuses of their platform by online extremists and have failed to respond quickly or install safeguards to prevent said misuse from happening again.”<sup>22</sup>

### **Recommendations for increasing transparency, trust and accountability:**

1. Tech firms must be transparent about their efforts to tackle content that glorifies and supports extremist views and violence. They should be clear about their use of hashing technology, and whether they are deploying at the point of upload. This is crucial because a video that remains online for even two hours, can gather hundreds of views, and spread onto smaller, encrypted platforms where it will be nearly impossible to find and remove. Companies should also provide a detailed explanation of how each contributes to, and participates in, the so-called “hashing coalition” announced in December 2016. Each company should state how much content they have contributed to this shared database, and whether there is an agreement that all content in the database be removed across industry platforms and websites, that are members of the hashing coalition and the Global Internet Forum to Counter Terrorism (GIFCT). Companies

---

<sup>20</sup> Press Release: Extremist Content Online: Atomwaffen Division Reuploads Itself Onto YouTube. *Counter Extremism Project*, 23 January 2019, <https://www.counterextremism.com/press/extremist-content-online-atomwaffen-division-reuploads-itself-youtube>.

<sup>21</sup> Press Release: Extremist Content Online: Atomwaffen Division Content Reuploaded to YouTube Despite Ban. *Counter Extremism Project*, 9 April 2019, <https://www.counterextremism.com/press/extremist-content-online-atomwaffen-division-content-reuploaded-youtube-despite-ban>.

<sup>22</sup> Farid H. Press Release: Misuse. *Counter Extremism Project*, 16 May 2019, <https://www.counterextremism.com/press/cep's-dr-hany-farid-tech-companies-have-failed-to-recognize-abuses-their-platform>

should state how much content has been removed from their platform as a result of the database, and explain how the database is updated.

2. CEP research found that tech companies' current content moderating systems are inadequate. The regulator should issue minimum requirements for tech companies' training of content moderators, who are instrumental in monitoring and removing online extremism. Such minimum requirements would help make sure that these people receive thorough training, and acquire specialised knowledge, for example knowing how to recognise alt-right symbols and codes, especially as they shift and evolve. Content reviewers should also have a good and clear understating of what material violates the law, violates the company's terms of service, and constitutes extremist or terrorist content. Further, content reviewers must be proactive in their monitoring, rather than wait for content to be reported to them. The regulator must also mandate that tech companies provide content reviewers with adequate pay and emotional support, because constant exposure to harmful content can have an adverse emotion effect or even radicalise moderators.
3. Companies have to be transparent about their content removal policies, including, but not limited to, information about how many and which offenders have had their content and/or accounts removed. Additionally, companies must report on which profiles, which have suspended or banned due to dissemination of terrorists and violent extremist content, have been allowed to post again, and explain why they were allowed to do so. Such a reporting requirement can encourage companies to act more proactively to permanently ban extremists, and restore their users' trust.
4. A study by CEP of the German Network Enforcement Act, or NetzDG, found that a lack of uniform reporting system made it difficult for users to flag NetzDG violations.<sup>23</sup> The regulator should therefore establish guidelines and quality standards for reporting and make sure that all

---

<sup>23</sup> Echison, E. and Knodt, O. Germany's NetzDG: A Key Test for Combatting Online Hate. *Counter Extremism Project*, November 2018, [https://www.counterextremism.com/sites/default/files/CEP-CEPS\\_Germany%27s%20NetzDG\\_020119.pdf](https://www.counterextremism.com/sites/default/files/CEP-CEPS_Germany%27s%20NetzDG_020119.pdf)

companies have a clear, simple, and standardised reporting system, and that reports are treated seriously. This will also help increase the public's trust.

5. There should be a clear definition of what constitutes 'terrorist content.' Based on this, a database can be created where this content could be 'hashed' and shared with service providers so as to ensure it is permanently removed. CEP acknowledges the difficulty in achieving consensus on the definition of 'terrorist content' and supports the definitions provided in the EU proposal for regulation on 'preventing the dissemination of terrorist content online' (Art 2.5). The definitions: **(a)** inciting or advocating, including by glorifying, the commission of terrorist offences, thereby causing a danger that such acts be committed; **(b)** encouraging the contribution to terrorist offences; **(c)** promoting the activities of a terrorist group, in particular by encouraging the participation in or support to a terrorist group within the meaning of Article 2(3) of Directive (EU) 2017/541; and **(d)** instructing on methods or techniques for the purpose of committing terrorist offences. These are sufficient in their scope and CEP believes they could contribute significantly to the removal of terror content online.<sup>24</sup> Additionally, CEP recommends that governments ensure that 'terrorist content' includes content promoting, or produced by groups on UK, US, EU, and UN sanctions list—which helps promote cohesion across the tech industry and jurisdictions—as well as individuals and specific pieces of content within proven links to violence. CEP has highlighted several notorious propagandists and pieces of content that the tech industry should be flagging for permanent removal due to their links to violent extremist individuals and attacks, specifically: Muslim Brotherhood leader Yusuf al-Qaradawi, neo-Nazi book *Siege*, violent extremist ideologue Abdullah al-Faisal, influential Salafist preacher Ahmad Musa Jibril, and a white supremacist 'bible' *The Turner Diaries*.<sup>25</sup>

---

<sup>24</sup> Preventing the Dissemination of Terrorist Content Online – CEP Position. *Counter Extremism Project*, <https://www.counterextremism.com/sites/default/files/Preventing%20The%20Dissemination%20Of%20Terrorist%20Content%20Online%20-%20CEP%20Position.pdf>, P.2.

<sup>25</sup> Tech and Terrorism: Tech's Self-Regulation Fails to Set Standards for Removing Extremist Content. *Counter Extremism Project*, 27 March 2019, <https://www.counterextremism.com/press/tech-terrorism-tech%E2%80%99s-self-regulation-fails-set-standards-removing-extremist-content>.

Question 7: Which channels or forums that can be considered private should be in scope of the regulatory framework?

Question 7a: What specific requirements might be appropriate to apply to private channels and forums in order to tackle online harms?

The use of technological tools to find child sexual abuse material as well as terrorist and extremist content could be useful in platforms that host private channels and forums, and platforms that have end-to-end encryption.

Last year, the US-based National Center for Missing and Exploited Children (NCMEC) received more than 18 million reports of apparent child sexual abuse images, online enticement, child sex trafficking, and child sexual molestation.<sup>26</sup> Most major technology companies have deployed technology that has proven effective at disrupting the global distribution of known child sexual abuse material. This technology, PhotoDNA, developed by CEP Senior Advisor Dr. Hany Farid and Microsoft in 2009, works by extracting a distinct digital signature from known harmful content and comparing these signatures at the point of upload. Flagged content can then be instantaneously removed and reported. This type of robust hashing technology is similar to that used to detect other harmful digital content like viruses and malware. Since its development and its eventual worldwide deployment, PhotoDNA remains one of the most effective strategies for combatting child sexual abuse online. The efficacy of this technology, however, is under threat.

In March 2019, Facebook CEO Mark Zuckerberg announced that he would move Instagram, WhatsApp, and Facebook Messenger to use end-to-end encryption, preventing anyone—including Facebook—from seeing the contents of any communications. Mr. Zuckerberg conceded that this move comes at a cost, stating: “At the same time, there are real safety concerns to address before we can implement end-to-end encryption across all of our messaging services. Encryption is a powerful tool for privacy, but that includes the privacy of people doing bad things. When billions of

---

<sup>26</sup> NCMEC Data. *National Center for Missing and Exploited Children*, <http://www.missingkids.com/ourwork/ncmecdata#bythenumbers>.

people use a service to connect, some of them are going to misuse it for truly terrible things like child exploitation, terrorism, and extortion.”<sup>27</sup>

Despite Facebook’s insistence, even end-to-end encryption does not provide users with as much privacy as users may be led to believe. Even without the ability to read the content of users’ private messages, Facebook will still know with whom a user is communicating, when a user is communicating, from where a user is communicating, as well as a user’s online activity.

Users will benefit from only a marginal increase in privacy, but end-to-end encryption would significantly hamper the efficacy of much needed technologies like PhotoDNA. Without PhotoDNA operating on the private messaging platforms of Instagram, WhatsApp, and Facebook Messenger, it would be impossible to detect harmful content shared through those platforms.

Recent advances in encryption and robust hashing technology, however, mean that technologies like PhotoDNA can be adapted to operate within an end-to-end encryption system. Specifically, when using certain types of encryption algorithms (known as partially- or fully-homomorphic encryption), it is possible to perform the same type of robust image hashing on encrypted data. This means that encrypted images can be analysed to determine if they are known harmful material without the need, or even ability, to decrypt the content. For all other images, this analysis provides no information about its contents—preserving content privacy. Alternatively, technologies like PhotoDNA can be implemented at the point of transmitting a message as opposed to the current approach where it is implemented upon receipt. In this client-side implementation, the hash is extracted prior to encryption and transmitted alongside the encrypted message. Because no identifying information can be extracted from this signature, it does not reveal any details about the encrypted content while allowing for the monitoring of known harmful material.<sup>28</sup>

---

<sup>27</sup> Zuckerberg, M. A Privacy-Focused Vision for Social Networking. *Facebook*, 6 March 2019, <https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/>.

<sup>28</sup> Farid, H. Facebook’s plan for end-to-end encryption sacrifices a lot of security for just a little bit of privacy. *Fox News*, 16 June 2019, <https://www.foxnews.com/opinion/hany-farid-facebook-end-to-end-encryption-security-privacy>.

### **Recommendations for flagging illegal or harmful content on encrypted platforms:**

1. CEP encourages the regulator to mandate that end-to-end encryption platforms use specific encryption protocols (i.e., partially- or fully-homomorphic encryption) that allow for the use of hashing technology or deploy hashing technology at the point of transmitting a message. These two scenarios allow companies to detect known harmful material, such as terrorist and extremist content as well as child abuse and exploitation content, while maintaining the privacy of communications.

Question 13: Should the regulator have the power to require a company based outside the UK and EEA to appoint a nominated representative in the UK or EEA in certain circumstances?

It is important to require companies based outside the UK and EEA to appoint a nominated representative in the UK or EEA. Following the launch of CEP's study on the NetzDG, CEP hosted a panel discussion in Berlin in November 2018, which included participation from Daniel Holznagel, legal officer at the German Federal Ministry of Justice and Consumer Protection. Mr. Holznagel contributed to the drafting the NetzDG, and his comments about requiring foreign businesses to appoint an in-country representative helped shed light on how such a requirement would help empower Internet users and victims of harmful content on social media platforms.

Users have become victims of the spread of harmful content on the Internet. In the case of extremist and terrorist material, such content has resulted in the radicalisation of individuals as well as lives lost. Mr. Holznagel suggests that, through regulation, it is possible to empower victims to use legal solutions by making it easier to sue large companies. He states, "If you're here in Germany and you want to sue a company in California, [it] can be very complicated. But it becomes easier once you have persons authorised to receive service here in Germany—contact points. So

that’s what we did with the NetzDG. We made it obligatory for big social networks to have a contact point here in Germany.”<sup>29</sup>

### **Recommendations for a representative in the UK or EEA:**

1. The regulator should require a company based outside the UK and EEA to appoint a nominated representative in the UK or EEA. This stipulation will help empower Internet users as well as victims of the loss, injury, or damage that stems from harmful content online.

Question 15: What are the greatest opportunities and barriers for (i) innovation and (ii) adoption of safety technologies by UK organisations, and what role should government play in addressing these?

The use of automated tools is essential for preventing harmful terrorist and extremist content from appearing online. Technology that enables locating online terrorist and extremist content has been available for several years and is constantly improving. However, as demonstrated by CEP’s findings, tech companies have been reluctant to deploy effective technological tools to combat the issue. They have also failed in preventing the re-upload of images and videos that have been previously removed because their content included violent extremist ideologies.

CEP believes that hashing technology is a highly effective method for finding and removing re-uploaded extremist content. To help tackle the proliferation of extremist content online, CEP and Dr. Hany Farid developed eGLYPH — a technological solution that can help reduce the ability of extremists and terrorists to spread their content. The technology is capable of detecting known extremist images, video, and audio files through “robust hashing” technology, which was originally deployed to identify and flag images of child pornography online (PhotoDNA), by extracting a distinct digital signature from an image, and comparing it against all other images encountered

---

<sup>29</sup> Panel discussion: The Regulation of Tech Companies – A Key Test for Combatting Hate Speech. *Counter Extremism Project*, 29 November 2018, <https://www.youtube.com/watch?v=cRgyODQp1VY&t=1s>, minute 51:42.

online. eGLYPH expands on this existing technology, and is able to analyse video and audio content quickly and accurately, making it particularly impactful in combatting the proliferation of extremist propaganda. CEP makes this technology available, free of charge, to anyone who wants to use it to fight extremism.

### **Recommendations for technological innovation:**

1. The use of technological tools to detect harmful content at the point on upload should be mandatory. The regulator would need to make sure that companies are using the most up-to-date, effective tools available, and encourage larger companies to make their technologies and knowledge available to smaller platforms with less financial resources, to help them combat extremist content. If necessary, smaller companies should receive financial aide to combat harmful content, or offered technological tools, such as the eGlyph, free of charge.
2. Tech companies should be transparent about the technology that they are using, and about how successful this technology has been. This can encourage companies to invest more in innovative tools.
3. A mandatory database for sharing ‘hashes’ should be established and made accessible to hosting service providers. A voluntary approach, such as through the Global Internet Forum to Counter Terrorism, has proven to be insufficient and resulted in re-uploads of known terrorist and extremist content previously identified to have violated companies’ terms of service.<sup>30</sup>
4. Human oversight and verification is important. However, the regulator should keep in mind that using this method alone without the help of automation would be less effective and could hinder innovation. It is also important to consider that exposing human moderators to large quantities of harmful content over a long period of time can have an adverse effect on their mental and emotional wellbeing. Human moderators should therefore work in collaboration with

---

<sup>30</sup> Extremist Content Online: Videos from New Zealand Shooting Remain Online. *Counter Extremism Project*, 1 April 2019, <https://www.counterextremism.com/press/extremist-content-online-videos-new-zealand-shooting-remain-online>.

technological tools to limit exposure to violent content, make the process more efficient, and encourage innovation.

5. CEP researchers evaluating Germany's NetzDG Law have made the following recommendation based on their findings, which can encourage learning and innovation:

'Understanding the complexities of social media and how algorithms function is critical for lawmakers, law enforcement authorities, and the cyber-competence of society. All tech companies should continue to allow research on their APIs so that progress (e.g. on the quick removal of ISIS propaganda) and trends (e.g. the spread of false information during elections) can be investigated. Facebook has disabled this option for independent researchers. Since aggregated numbers are difficult to verify, governments should require Facebook, Twitter, and Google to allow access to 'raw' aggregated data for the purpose of analysis.'<sup>31</sup>

---

<sup>31</sup> Echison, E. and Knodt, O. Germany's NetzDG: A Key Test for Combatting Online Hate. *Counter Extremism Project*, November 2018, [https://www.counterextremism.com/sites/default/files/CEP-CEPS\\_Germany%27s%20NetzDG\\_020119.pdf](https://www.counterextremism.com/sites/default/files/CEP-CEPS_Germany%27s%20NetzDG_020119.pdf), P.16.