

Moonshot's Annual Impact Report

2021





Content Warning

This report contains graphic references to violent language and other material that readers may find distressing.

2021 Moonshot's Impact



Moonshot ran 46 projects in over 33 countries in 18 different languages during 2021.¹ These projects addressed six different forms of online harms across over 51 online platforms, social media sites and forums. We developed or improved our in-house proprietary tools to analyse data from 19 online platforms.



PROJECTS



COUNTRIES



LANGUAGES



ONLINE HARMS



ONLINE PLATFORMS

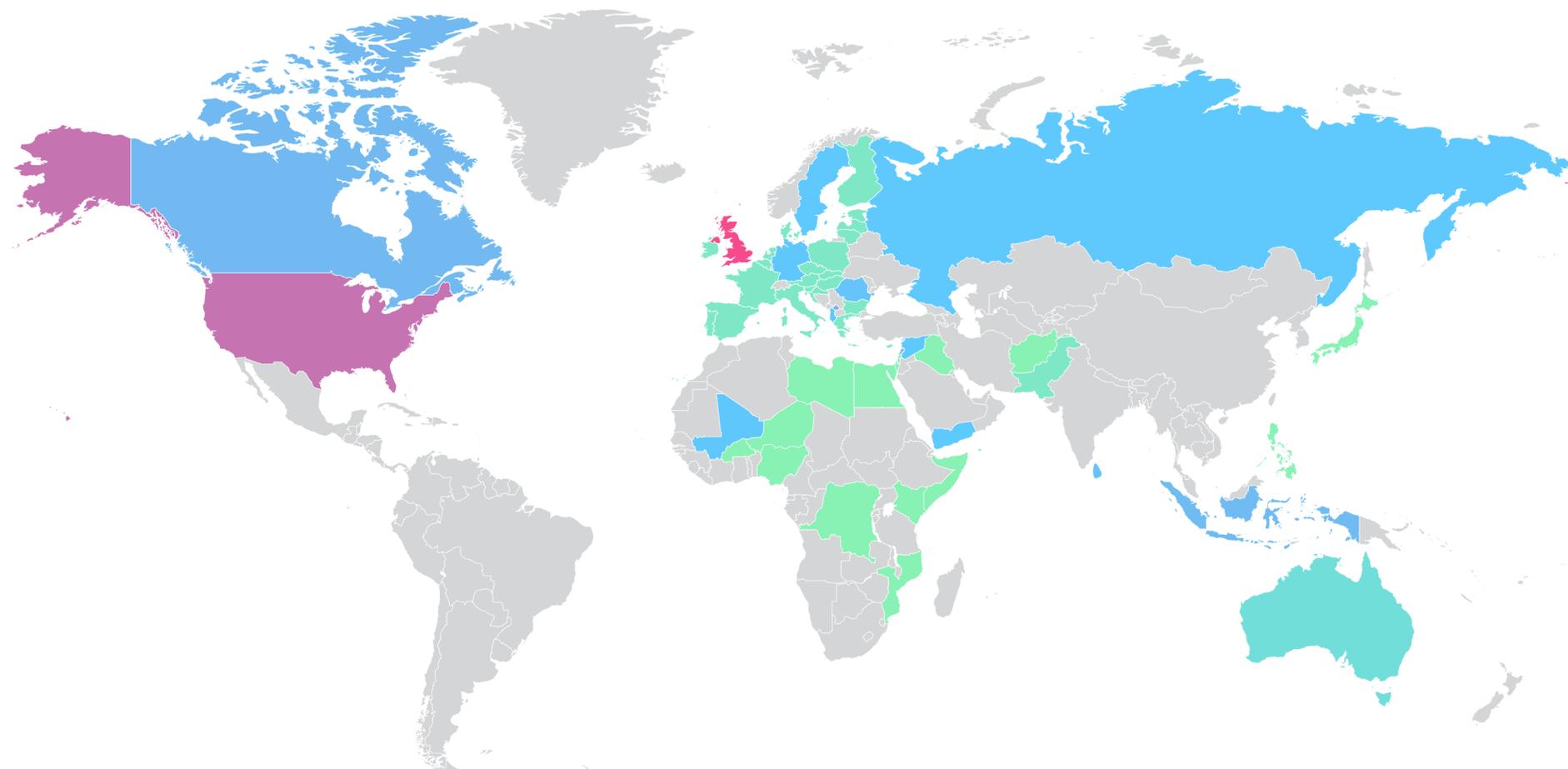
1. This report covers any project that was launched, carried out or completed during the calendar year 2021.



Countries covered by Moonshot's work in 2021



Moonshot worked in over 33 countries in 2021. Moonshot's objective is to build longer-term, deeper partnerships and expertise in countries where we work. Our geographic maturity score is a way of measuring our progress in this objective. It is calculated as a weighted index of number of projects, number of indicators, number of partners and number of subject matter experts (SMEs) with whom we have worked.



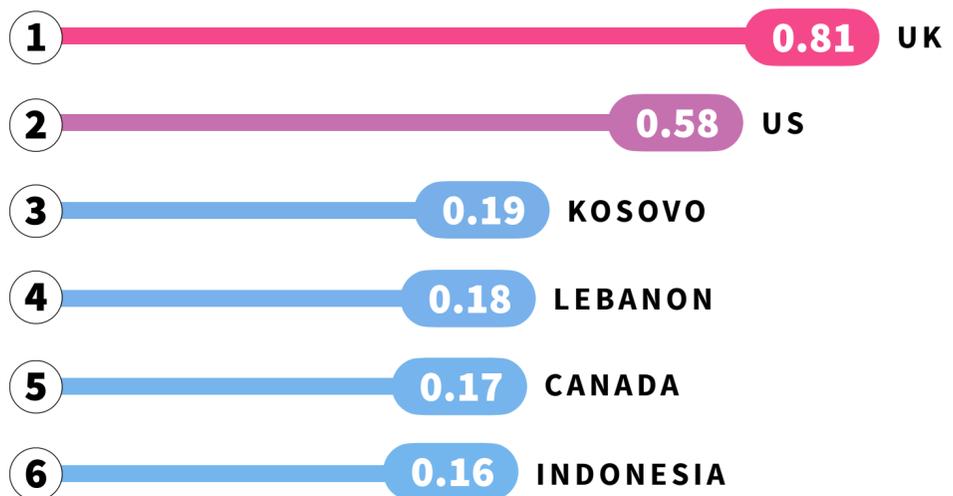
KEY: WEIGHTED MATURITY SCORE



STANDARD

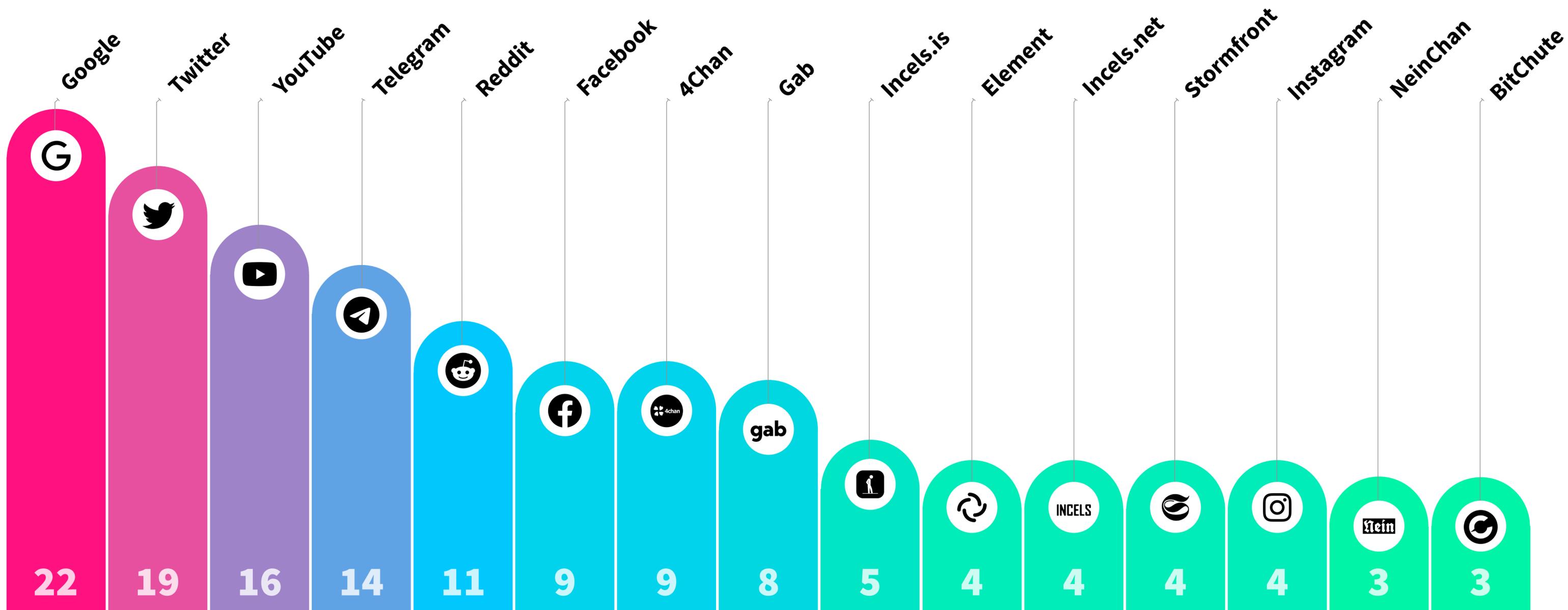
ADVANCED

TOP 6 COUNTRIES





Top 15 platforms covered by Moonshot's work in 2021



NUMBER OF PROJECTS



Types of online harm covered by Moonshot's work in 2021

Moonshot defines online harm as content or activity that endangers individuals and/or public safety. While Moonshot typically focuses on online harms which can cause physical harm, either to individuals or society, the online harms in scope contribute to a wide array of negative impacts on society, including the degradation of democracy and human rights.

Moonshot worked across six online harm types in 2021. Moonshot's objective is to build deeper knowledge of these online harm types.



ONLINE HARM TYPE

OF PROJECTS

Violent Extremism

29

Disinformation and Misinformation

13

Targeted Violence and Hate Speech

6

Gender-Based Violence (GBV)

7

Human Trafficking and Modern Slavery

2

Child Sexual Exploitation and Abuse (CSEA)

3

Our impact in 2021

How do we measure our impact?



Moonshot works to end online harms applying evidence, ethics and human rights. We have broken down our mission into six tangible objectives to measure the impact of our programming across the organisation. These objectives are as follows:

 **Building the evidence base:** Moonshot gathers evidence through its programming to provide a more informed picture of the nature of online harms. This allows us and our clients to address the threat posed by these harms in a more data-driven way.

 **Safeguarding at-risk users:** Moonshot connects people searching online for harmful content with constructive and safer alternative messages in order to reduce their risk of causing harm to themselves and others.

 **Protecting targeted communities:** Moonshot offers tailored support to vulnerable audiences targeted by online harms, such as sharing social services resources to victims of domestic violence. This contributes to protecting targeted communities and matching them with offline support.

 **Provision of services:** Moonshot connects users vulnerable to carrying out online harms with support from specialist service providers to reduce their risk of perpetrating harm.

 **Disrupting online harms:** Moonshot disrupts the activities of extremists online by reporting harmful content and partnering with organisations to break the link between harmful online behaviour and offline violence.

 **Increasing capacity:** Moonshot builds the capacity of CSOs and local activists to respond to online harms, so as to increase the possibility for sustainable, locally-driven, long-term impact.

OBJECTIVE 01 BUILDING THE EVIDENCE BASE



Moonshot gathers evidence through its programming to provide a more informed picture of the nature of online harms. This allows us and our clients to address the threat posed by these harms in a data-driven way.

Number of projects

36

Number of countries covered

32

Number of harms covered

6

Number of reports delivered

212

Number of incidents covered in real time

14

Moonshot conducted 36 projects in 2021, writing 212 reports, to build the evidence base on the nature, prevalence and threat posed by online harms. We have included some example insights below by online harm.



Violent extremism

We conducted 23 projects building the evidence for seven different types of violent extremism in 2021. These include far-right extremism, Salafi-Jihadist extremism, incel-motivated extremism, Buddhist ultranationalism, Sikh violent extremism and violent far-left extremism.

Far-right extremism

We conducted research into right-wing violent extremism in the US to proactively monitor and respond to threats of violence and violence-inciting disinformation during the US Presidential Election. Moonshot found that searches relating to conspiracy theories were particularly popular around the election; with more than 35,000 searches identified across the US for harmful conspiracy theories between mid-October 2020 and mid-December 2020. Moonshot also identified more than 11,000 searches across the US indicating a desire to join armed groups around the November 2020 election. The Three Percenters was the most searched-for group nationwide, with nearly 6,500 searches identified.

Salafi-Jihadist extremism

Analysis of the online presence of international Salafi-Jihadist groups found that Telegram was the most popular platform among supporters. Content hosted on the platform was diverse in nature, with official and unofficial AQ, Daesh and HTS media outlets, online influencers and supporter spaces all present on the platform. While on the whole this form of extremism is more highly moderated by platforms, non-English language content tends to remain online for longer than English language content. This was found by analysis into Albanian-language violent extremist (VE) content.

Incel extremism

Our research analysed global incel communities posting in several languages. Our network analysis into these communities found that they are truly global in nature but also have local hubs. An analysis of engagers with incel YouTube content found the most popular geographies where there is consumption of incel content were the US, the UK and Canada.

Moonshot conducted 36 projects in 2021, writing 212 reports, to build the evidence base on the nature, prevalence and threat posed by online harms. We have included some example insights below by online harm.



Targeted violence and hate speech

In 2021, we ran six projects looking into the nature of targeted violence and hate speech globally as well as specifically in the UK, US, Sri Lanka and Lebanon. Of particular note, we analysed the online presence of those interested in mass shootings.

Moonshot analysed how people explore mass shootings online in order to identify ways to reduce the risk of those vulnerable to committing mass shootings. We analysed anonymous search data in California related to mass shootings and identified some concerning behaviour, including watching footage of shootings, idolising shooters, and expressing intent to harm students. Links to downloadable versions of games created by perpetrators or emulating school shootings were also found. The exhibition of these behaviours online presents an opportunity to offer support to those who may be at risk.



Disinformation and misinformation

We conducted eight projects on the threat posed by online disinformation and misinformation in over 12 different geographies, publishing 37 reports. We were particularly interested in the impact of changing COVID-19 restrictions on conspiracy narratives globally. Analysis of search trends found that the introduction of localised restrictions coincided with upticks in conspiracy-related searches.



Gender-based violence

We conducted five projects addressing gender-based violence online. In a landmark research study, Moonshot explored gendered abuse and disinformation across six social media platforms over two months in late 2020. Our analysis found that gendered abuse affected 12 out of

Moonshot conducted 36 projects in 2021, writing 212 reports, to build the evidence base on the nature, prevalence and threat posed by online harms. We have included some example insights below by online harm.



13 influential female politicians analysed, while nine of these politicians were impacted by gendered disinformation. Online gendered abuse and disinformation was often intersectional in nature, with abusers engaging with both sex- and race-based narratives, compounding the threat for women of colour.

Our work in 2021 also furthered our understanding of how perpetrators and victims of domestic violence use the internet. In London, Moonshot conducted research on the prevalence of GBV and domestic violence searches online and uncovered that a majority of searches were actively looking for help or shelter. Such findings demonstrate the importance of advertising support services and helplines to ensure vulnerable individuals have access to the support they need.

Notably, searches related to domestic abuse on the day of the Euro 2020 final were almost 300% higher than the rest of 2021. Searches for 'domestic violence helpline' were 891% higher than the average search volumes for that term for the rest of the year.



Human trafficking and modern slavery

We conducted two projects developing our understanding of human trafficking, modern slavery and forced labour, globally. Analysis into human trafficking found that online platforms are frequently used by perpetrators to advertise their services, often opening and managing several accounts across platforms to maximise their online footprint. Perpetrators utilise specific slang in their advertisements to avoid account identification and suspension.

OBJECTIVE 02

SAFEGUARDING AT-RISK USERS



Moonshot connects people searching online for harmful content with constructive and safer alternative messages in order to reduce their risk of causing harm to themselves and others.

Number of projects

7

Number of countries covered

5

Number of harms covered

5

Number of reports delivered

25

Total offers of safer alternatives²

122,959,772

Total number of searches for dangerous content successfully disrupted³

1,069,631

Hours of counter-content watched

> 454

Highest CTR achieved

8%

MOONSHOT AIMS TO AMPLIFY EXISTING CONTENT CREATED BY CREDIBLE MESSENGERS ACROSS THE GLOBE. MOONSHOT ONLY CREATES CONTENT WHERE THERE ARE GAPS TO FILL OR EXISTING CONTENT DOES NOT MEET A SPECIFIC NEED.

Moonshot shared over **243 pieces** of curated counter-content

Moonshot created or co-created: **13 videos, 2 websites, 2 quizzes, 1 song** and **1 video game** to address online harm

2. We calculated the total offers of safer alternatives from the total number of impressions from our campaigns.

3. We calculated the number of searches for dangerous content successfully disrupted from the number of clicks on our campaign content.

Moonshot conducted seven projects in 2021 to connect people searching online for harmful content with constructive and safer alternative messages. We have included some of our most interesting findings from these projects below.



Violent extremism

Sri-Lanka

Moonshot ran an 18-month programme in Sri Lanka from September 2019 - July 2021, to counter Buddhist ultranationalism and build digital resilience to anti-minority hate speech. As part of our digital campaigns to counter hate speech, we ran a successful behaviour change experiment on Twitter that measurably reduced volumes of ultranationalist harmful speech posted by a sample of Twitter users in Sri Lanka. Moonshot delivered targeted positive counter messaging to a treatment group of individuals who were otherwise posting hate speech on Twitter.

Moonshot analysed the harmful speech posted by these individuals both before and after the treatment, in comparison to a control group not provided the treatment. Moonshot found that tweets from the treatment group were half as likely than tweets in the control group to contain harmful speech after receiving our treatment. Results were statistically significant at the 95% confidence level ($p < .05$).

United States

Moonshot launched a new model for election-violence prevention in the U.S. Over four months, we sought to counter and reduce the threat of violence in the aftermath of the 2020 US presidential election. We did this by safeguarding internet users seeking or consuming content that could incite violence, and offering alternative pathways that supported shared values, inclusive citizenship, and mental health. In total, we delivered more than 105 million violence prevention messages across platforms, with more than 1.7 million user engagements.

Our “Calm” themed messaging, focused on general mental health and taking a “breather”, were the most successful across all target audiences. Audiences looking for content by groups, particularly those searching for Proud Boys and Oath Keepers content, also engaged at high rates with our Empathy-Anger content. This offered an empathetic response to feelings of anger, and highlighted the potential impact of acknowledging and defusing feelings of anger among at-risk audiences.

Moonshot conducted seven projects in 2021 to connect people searching online for harmful content with constructive and safer alternative messages. We have included some of our most interesting findings from these projects below.



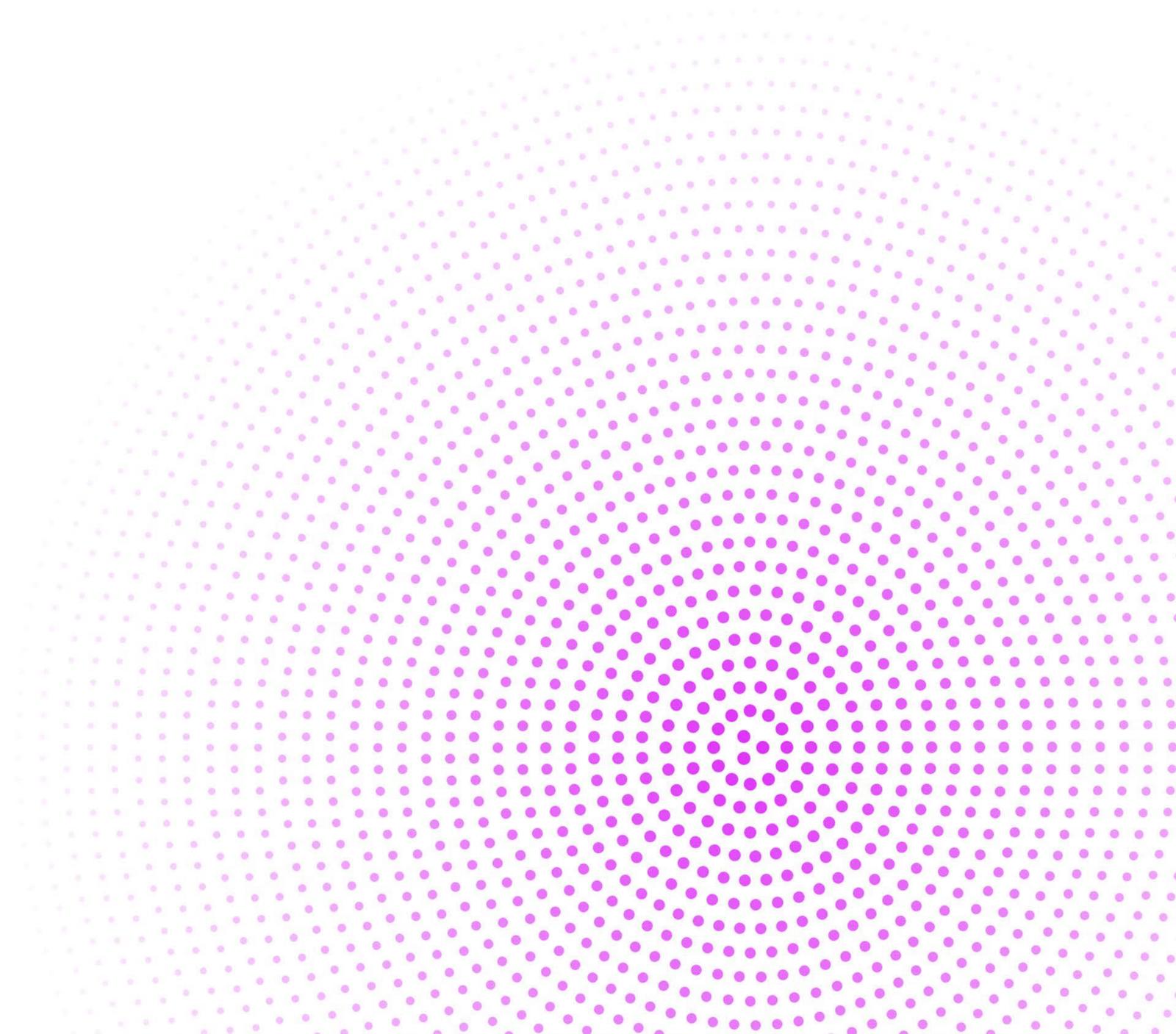
Disinformation and misinformation

Indonesia

In Indonesia, Moonshot has worked to advance media literacy amongst new digital arrivals. Initially, Moonshot analysed the disinformation environment in Indonesia. Moonshot then designed and deployed digital campaigns to reach these audiences with a media literacy website and a disinformation inoculation game.⁵ Both the website and game proved effective at taking user attention away from disinformation content towards positive media literacy content.

The media literacy game, designed for an Indonesian audience, was particularly popular. The goal of the game is to inoculate people who are otherwise vulnerable to disinformation and equip them with the media literacy skills necessary to identify it in the real world, both online and off. Users who played the game spent around 12 times as long engaging with the content than those who only visited the project's website.

5. Moonshot tested the Inoculation theory to combat mis/disinformation. The Inoculation theory explains how individuals can build resilience to harmful ideas or beliefs in a similar way to how bodies build resilience to diseases through pre-exposure to weakened versions of the threat.



OBJECTIVE 03

PROTECTING TARGETED COMMUNITIES



Moonshot offers tailored support to vulnerable audiences targeted by online harms, such as sharing social services resources to victims of domestic violence. This contributes to protecting targeted communities and matching them with offline support.

Number of projects

4

Number of countries covered

7

Number of harms covered

4

Number of reports delivered

7

Total of crisis campaigns run

2

Total number of times help was signposted⁶

12,121

Highest CTR achieved

11%

6. We calculated the total number of times help was signposted from the number of clicks on our campaign content.

Moonshot conducted four projects across eight countries in 2021 to safeguard vulnerable audiences targeted by online harms and match them with support services. We have included some key highlights from an example project below.



Gender-based violence

Our analysis into how perpetrators and victims of domestic violence use the internet in Miami and London found that victims of domestic violence were often unable to find local services. Moonshot launched a crisis campaign to advertise local and tailored GBV services to those searching for them, specifically services offering in-person interactions. Through our crisis campaigns in Miami, Moonshot found that Spanish terms, despite their lower proportion in our databases, had a higher CTR (11%) than for their English counterparts, suggesting an appetite for greater support in Spanish language.

OBJECTIVE 04

PROVISION OF SERVICES



Moonshot connects users vulnerable to carrying out online harms with support from specialist service providers to reduce their risk of perpetrating harm.

Number of projects

6

Number of countries covered

8

Number of harms covered

3

Number of reports delivered

19

Number of interventions campaigns run

12

Total offers of safer alternatives⁷

3,849,317

Total number of searches for dangerous content successfully disrupted⁸

40,669

Number of users reached

25,251

Number of 1:1 messages with at-risk users

>892

Number of individuals receiving support via our messaging

59

7. We calculated the total offers of safer alternatives from the total number of impressions from our campaigns.

8. We calculated the number of searches for dangerous content successfully disrupted from the number of clicks on our campaign content.



Moonshot conducted six projects across eight countries in 2021 connecting users at-risk of carrying out online harms with support services, often working with partners with specialist or local expertise. We have included some key highlights from these projects below.



Violent extremism

2021 was a year when we made significant advancements in the approach of offering behavioural health services to counter extremism. We built from the learning of pilots delivered in previous years in countries as diverse as Bangladesh, Australia, and the United States. This year we tested the offering of text-message based services for the first time - something which showed promise and we intend to build on.

United States

One project aimed to reduce the threat of election-related violence in the US. Over four months, Moonshot offered pathways to partner websites that provided mental health support, offered guides on having difficult conversations about politics, and promoted active listening skills - notably Crisis Text Line (CTL). Our objective was to raise awareness of the service and offer support to individuals who may be experiencing psychosocial stresses, such as election-related anxiety, loneliness, anger, despair, or emotional abuse, and thereby reduce the risk of violence in the

post-election period. In total, we advertised Crisis Text Line to at-risk users more than 4 million times, and redirected more than 22,000 users to websites featuring mental health resources and the CTL helpline. At least 33 individuals texted CTL using via our referral pathway, resulting in a total of 39 conversations during the project. Our partnership with CTL continues, and we have facilitated 150 conversations between at-risk individuals and the service to date.

Albania

Moonshot built a multi-platform solution to provide online support to users at-risk of Salafi-Jihadist radicalisation in Albania. Moonshot partnered with a local counselling service to engage with at-risk individuals on a 1-1 basis to reduce their vulnerability to and increase their resilience against violent extremism. The pilot interventions model was live for ten months. In total, 19 at-risk individuals proactively contacted the services and 1,463 people saw self-help content from our pages. We found that emotion-based messaging can be effective in reaching the at-risk audience and that those at risk of extremism can be provided online counselling.

OBJECTIVE 05

DISRUPTING ONLINE HARMS



Moonshot disrupts the activities of extremists online by reporting harmful content and partnering with organisations to break the link between harmful online behaviour and offline violence.

Number of projects

7

Number of countries covered

8

Number of harms covered

4

Number of reports delivered

72

Number of incitements to violence and hate speech reported to authorities or social media platforms

2,997

Moonshot conducted seven projects to disrupt the activities of extremists online through reporting harmful content and partnering with organisations to break the link between harmful online behaviour and offline violence. This makes it harder for those perpetrating online harms to operate on the internet and attempts to reduce the threat posed. We have included some key highlights from these projects below.



Violent extremism

Moonshot conducted a platform audit with a large technology company to reduce the hateful content on their platform. As part of this content, we blacklisted over 1,177 pieces of violent extremist content and therefore reduced the amount of hate on this platform. We are currently engaging with the platform to launch a safety-by-design initiative to offer positive alternatives to users seeking hateful content.



Child sexual exploitation and abuse (CSEA)

Moonshot conducted projects to analyse child sexual exploitation and abuse on online platforms. Our work uncovered multiple specific and direct threats towards children. Moonshot reported these threats to the authorities, contributing to the protection of children in danger. Moonshot will continue to report direct threats towards children in any country we deliver this work, as we grow this portfolio.

OBJECTIVE 06 INCREASING CAPACITY



Moonshot builds the capacity of CSOs and local activists to respond to online harms, so as to increase the possibility for sustainable, locally-driven, long-term impact.

Number of projects

11

Number of countries covered

8

Number of harms covered

4

Number of reports delivered

36

Number of training sessions run

86

Number of people trained

335

Number of partner organisations trained

31

As part of 11 projects, Moonshot built the capacity of CSOs and local activists to respond to online harms. We trained participants in over nine different languages and across four types of online harm: violent extremism, disinformation and misinformation, targeted violence and hate speech, and gender-based violence. In total, we ran 86 sessions. We have included some key highlights from these projects below.



Disinformation and misinformation

In Mali, Moonshot worked with several partners as part of a project to improve prospects for long-term peace, security, and reconciliation in Mali by strengthening the positive resilience of communities working in partnership with their elected officials. As part of this work, we analysed Malian COVID-19 misinformation narratives online, provided recommendations for resilience building campaigns and ran training sessions on strategic communication skills for youth. As part of these training sessions, Moonshot delivered a four-day capacity building programme, in French, to 28 community activists on techniques to identify and counter disinformation. Post training data showed a 52% increase in overall skills and confidence in tackling disinformation and creating digital campaigns.



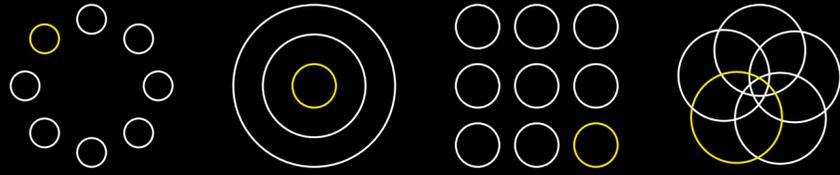
Violent extremism and gender-based violence

Moonshot delivered capacity building, Redirect campaigns and online intervention pilots to counter violent extremism and gender-based violence in Kosovo. In addition to our digital campaigns, we ran 23 tailored workshops to upskill 12 local partners and 67 participants on running digital campaigns.



Targeted violence and hate speech

Moonshot conducted a series of projects to promote religious freedom in Lebanon, aimed at countering religious-based hate speech in the online space. As part of these projects, Moonshot delivered a series of virtual workshops in Arabic to 17 local CSOs to train them in conducting effective digital campaigns. In addition to group sessions, Moonshot offered tailored one-on-one mentoring sessions to meet the needs of each participating organisation. Following the workshop completion, 85% of survey respondents reported finding the sessions useful or very useful.



MOONSHOT

CASE STUDIES

Moonshot responded to over 14 different incidents, mostly related to violent extremism, over 2021. Our response to these incidents showcases the breadth of our impact from informing policy makers on how to respond to the threat to launching digital crisis campaigns.



Washington D.C. January 6th insurrection

On January 6th, a violent insurrection took place in Washington, D.C. aiming to overturn the results of the 2020 presidential election.



Moonshot response

- ▶ **Safeguarding at-risk users:** Moonshot was already delivering crisis monitoring and crisis campaigns across the United States to reach audiences engaging VFR content. On January 5th, our monitoring team picked up on a surge in VFR activity in the Washington Metropolitan Area. 24 hours before the insurrection, we surged our crisis campaigns in the region to ensure we could be there to offer safer alternatives to those in need. This work continued during the entire insurrection and through the period that followed.
- ▶ **Building the evidence base:** Moonshot monitored US and UK extremist audiences' reactions to the storming of the Capitol, producing several live reports highlighting key developments and other longer-term analyses examining the impact of the incident on at-risk audiences.

Data insight

- ▶ Globally, extreme right-wing audiences discussed the attack in high volumes, including several discussions relating to copy-cat attacks. In the UK, several posts were identified encouraging users to “*storm Westminster*”.
- ▶ Analysis of global audiences engaging with conspiracy theories found that references to conspiracy theories on Twitter dropped significantly after February 2021. This was likely due to a change in Twitter’s moderation policies in response to the insurrection in January 2021. Our analysis identified that Tweets referencing conspiratorial sentiment decreased by over 75% compared with the previous month.
- ▶ On January 5th, 24 hours before the insurrection, we surged our crisis campaigns across the US in response to an escalation of violence online. Within 48 hours, our advertisements were shown **over 4.9 million times, including to over 160,000 users in the DC Metro area**. These resulted in 46,000 engagements and 38,000 clicks through to crisis support.



Afghanistan

Taliban takeover

On 14 April 2021, US President Joe Biden announced that he would fully withdraw US troops from Afghanistan by 11 September 2021. Taking advantage of decreased security in the wake of the withdrawal, the Taliban began advancing across Afghanistan before finally entering Kabul on 15 August 2021 - thereby taking control of all key cities.



Moonshot response

- ▶ **Building the evidence base:** Moonshot conducted social media analysis of narratives from Salafi-Jihadist groups, extreme right-wing audiences and incel audiences between August and September 2021.

Data insight

- ▶ **Global violent Salafi-Jihadism:** Responses to events in Afghanistan differed by militant group:
 - **Daesh:** Daesh supporters sought to undermine the Taliban, and criticise its “Islamic legitimacy” for collaborating with the US. Daesh celebrated its own attacks in the region and threatened more attacks to bolster its recruitment efforts.
 - **Al Qaeda (AQ):** AQ supporters launched a coordinated and extensive media campaign following the US announcement of its withdrawal. AQ’s media efforts were intended to galvanise supporters, recruit new supporters, and project its “staying power”.
 - **Haya Tahrir al-Sham (HTS):** HTS and its supporters championed the Taliban’s takeover of Afghanistan. The group also regularly attempted to liken itself to the Taliban as a legitimate Islamic government.



- ▶ **Extreme right-wing groups in the US and UK:** Among US racially/ethnically motivated violent extremist militia violent extremists and UK-based extreme right-wing audiences, Moonshot's analyses consistently identified anti-government, anti-migrant and anti-refugee narratives from these audiences.

Moonshot's analysis further identified attempts by some groups to use Afghanistan-related narratives to stir recruitment efforts - particularly those aimed at veterans, active military personnel and current/former members of law enforcement.

- ▶ **Incel extremism:** Moonshot's analyses also identified a number of incel narratives relating to the Taliban's return to power in Afghanistan. These posts generally celebrated the Taliban's advance into Kabul and takeover of Afghanistan, in light of a prospective 'decline in gender equality'. Users consistently utilised misogynistic language and expressed support for former Taliban policies towards Afghan women, with some users encouraging more restrictive measures on women globally.

