UNIVERSITY of TASMANIA

reThink Chatbot Evaluation

FEBRUARY 2024

AUTHORED FOR The Internet Watch Foundation in partnership with the Lucy Faithfull Foundation and Aylo

BY Joel Scanlan, Jeremy Prichard, Lauren C. Hall, Paul Watters, Richard Wortley

IWF Internet Watch Foundation

LUCY FAITHFULL FOUNDATION

.aylo

Safe Online

CHILDHOOD WORLD CHILDHOOD FOUNDATION FOUNDED BY H.M. QUEEN SILVIA OF SWEDEN

Table of Contents

Use the ‹ at the bottom of each page, if you would like to return here.

DECLARATION

This report was funded by Safe Online and World Childhood Foundation. The authors declare that they have no interests, including financial interests, which might influence the approach they have taken to preparing for and writing this document.

UNIVERSITY of TASMANIA

Author biographies

Joel Scanlan

Dr Scanlan is a Senior Lecturer in the College of Business and Economics at the University of Tasmania. Dr Scanlan has a PhD in the twin areas of cybersecurity and machine learning. He specialises in cybersecurity and has been teaching and consulting with the industry for over 20 years. Dr Scanlan is an active researcher on projects relating to child exploitation prevention, maritime cyber security and machine learning. He is currently involved in online harm prevention projects funded by the Australian eSafety Commissioner (chatbot design) and the Australian Institute of Criminology (warning messages).

Jeremy Prichard

Jeremy Prichard is a Professor of Criminal Law at the University of Tasmania, and Adjunct Associate Professor at the University of Queensland. Jeremy's earlier professional roles included appointments at the Australian Institute of Criminology, the Crime and Misconduct Commission, and the Department of Premier and Cabinet in Queensland. His current fields of research include cybercrime, the causes and prevention of crime, and criminal law.

Lauren C. Hall

Ms Hall is a researcher in the College of Business and Economics at the University of Tasmania. Ms Hall has a First Class Honours Degree in Economics, and is a published researcher in factors affecting the health and wellbeing of individuals, such as the health impacts of climate change, gambling mechanisms, and childhood education. With over 12 years experience working in the Tertiary Education sector, Ms Hall has published in world leading outlets Nature Human Behaviour and The Lancet Planetary Health.

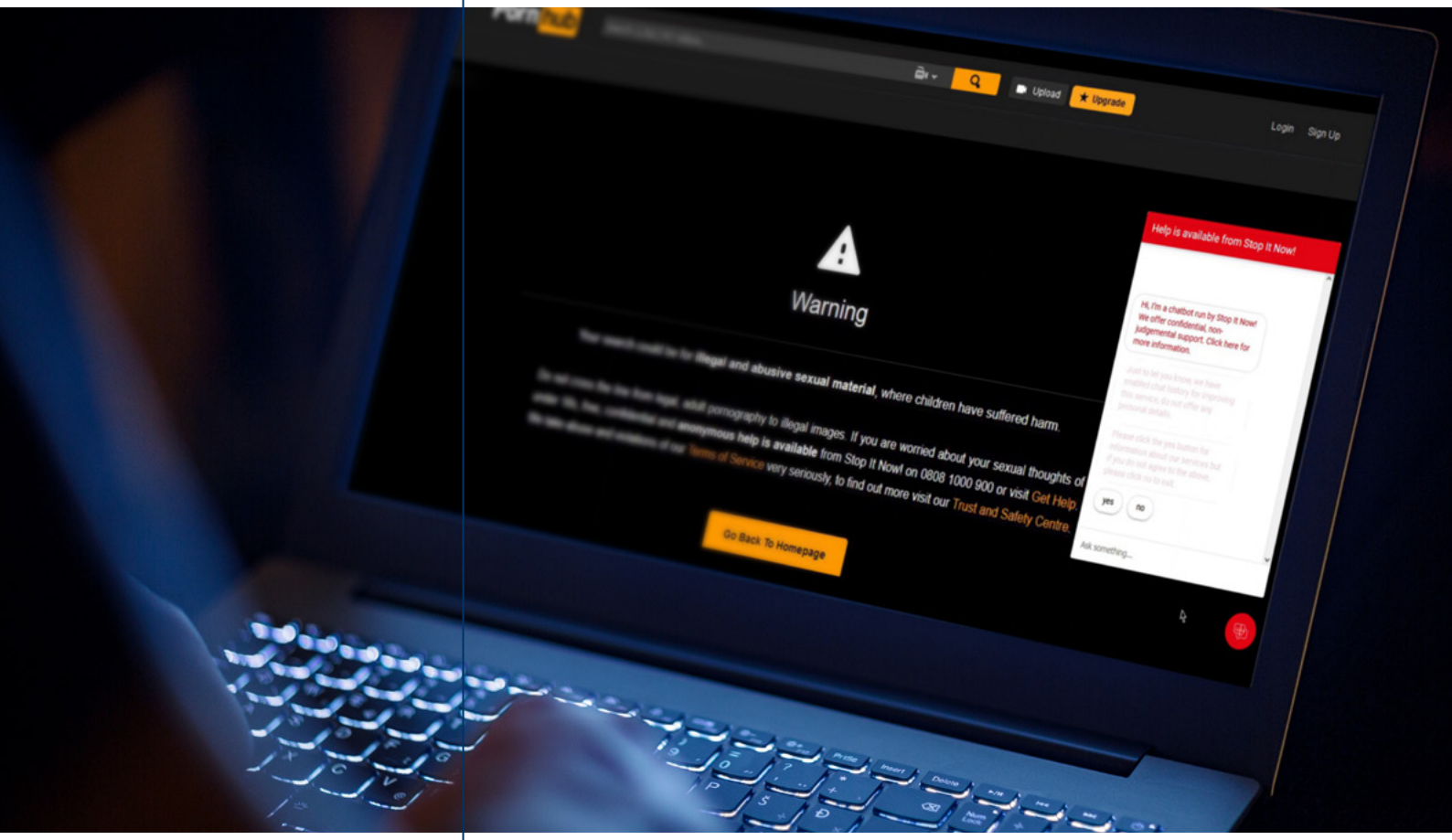
Paul A. Watters

Professor Paul A. Watters is Honorary Professor in Criminology and Security Studies at Macquarie University, Adjunct Professor of Cybersecurity at La Trobe University, and CEO of Cyberstronomy Pty Ltd. Professor Watters is a Fellow of the British Computer Society and a Chartered IT Professional, a Senior Member of the IEEE, and a Member of the Australian Psychological Society. Professor Watters has published more than 200 peer-reviewed research papers in cybersecurity, cybercrime, and cognate fields, which have been cited more than 7,771 times by his peers. He is consistently in the top 10% of all researchers by paper downloads on the Social Sciences Research Network (SSRN).

Richard Wortley

Richard Wortley is currently a Professor of Crime Science at University College London (UCL) and at the University of Waikato. A psychologist by disciplinary background, his research interests centre on the role that immediate environments play in criminal behaviour and the implications this has for situational crime prevention. He has a particular interest in the prevention of child sexual exploitation (CSE), both contact and online offending.

Introduction



Executive summary

This project evaluated the reThink Chatbot, which was deployed on the Pornhub website in the United Kingdom in March 2022. Data was collected until September 2023. The chatbot was designed to direct individuals attempting to search for child sexual abuse material (CSAM) on Pornhub in the UK to support services called Stop It Now, provided by the Lucy Faithfull Foundation (LFF).

A warning message was displayed with the chatbot, which itself had been in operation since March 2021, and also directed individuals to LFF support services.

Data was provided by Aylo (formerly MindGeek), the company which operates Pornhub, covering the search terms entered by sessions during which users were delivered the warning message and chatbot in response to any CSAM-related request; LFF provided user activity data from the Stop It Now website and helpline; and the Internet Watch Foundation (IWF) provided data on chatbot activity and dialogues of users.

Successes

MARCH 2022

99.8%

2.8M

1,656

490

68

AUGUST 2023

During the chatbot data collection period **99.83%-99.87%** of sessions did not contain a search that triggered the warning.

The chatbot was displayed **2.8 million times** between March 2022 and August 2023, resulting in **1,656 requests** for more information and Stop It Now services; **490 click-throughs** to the Stop It Now website; and approximately **68 calls and chats** to the Stop It Now anonymous counselling service.

Prior to the chatbot's launch, the warning message was shown an additional 2,208,864 times, for a total of 4,400,960 warnings across the length of the project.

There is a statistically significant trend showing a decrease in the number of searches for CSAM material on Pornhub in the UK during the length of the intervention.

There are indications that the warning message and chatbot reduced the total number of searches for CSAM material on Pornhub; most sessions which triggered the warning and chatbot once do not appear to have searched for CSAM again and those who see the warning message more than once, tend to undertake non-CSAM searches after receiving the warning.

Sessions during which the first activity on Pornhub was to search for CSAM, continue to use the site, but they did then search for content less than other sessions.

Complexities

The overall number of users requesting information about Stop It Now services from the chatbot decreased over the length of the intervention.

Web traffic data show that the chatbot initially produced a steady number of referrals to the Stop It Now website, however, this rate of referral decreased across the length of the intervention. Data of helpline calls, emails and chats show an increase in usage over the period of the intervention.

An examination of users' typed dialogues with the chatbot showed that a portion of users' experiences with the chatbot were negative.

Limitations

The data analysed in this evaluation comprised primarily of 'sessions' (i.e., a period of time spent on Pornhub by individual users) and fewer were individual users.

The sessions are not distinct, which means that double-counting individuals occurred (although the extent of this cannot be quantified).

A limited number of chatbot conversations were able to be viewed within the evaluation due to the chatbot platform (Google Cloud) having limited export facilities.

Changes made to the warning message text and chatbot in quick succession between May 2023 to August 2023 limited the ability to identify the direct contribution each individual change specifically made to user engagement.

The data examined did not include a period prior to the warning message being present on Pornhub, so there is no direct comparison to the baseline prior to the intervention.

Future directions

The intervention has shown what is possible in this space, and multiple future directions to build on this work exist, including in contexts different from pornography websites.

A possible high impact pathway would be to integrate the LFF video campaigns into the warning message to provide a higher fidelity warning page, which will change over time as different campaigns operate.

The chatbot conversations observed contained some disappointing interactions. An improved chatbot, which is more responsive to user typed inputs, would likely be more successful.

Analysis of video titles could be another avenue to improve existing detection tools, or also be a valuable location for support resources to be advertised.

Future uses of the chatbot should aim to be more integrated with Stop It Now; with direct calling from the chatbot, transferring the chat conversation to the Stop It Now live chat, to enable human-to-human interaction more quickly and facilitate higher retention.

...enable human-to-human interaction more quickly and facilitate higher retention.

1.0 Introduction

Sexual images of children, globally referred to as child sexual abuse material (CSAM), have unfortunately always been easy to find on the internet and may even be accidentally encountered on mainstream websites (Prichard et al., 2013). However, the existence of CSAM on some legal pornography websites (Morgan & Lambie, 2019) has also attracted public consternation. It is seen as a distressing de facto normalisation of CSAM on mainstream parts of the internet which are otherwise law-abiding (see e.g. Dines, 2009: 124; Prichard et al., 2013: 997; Warner, 2010: 395). Deterrence warning messages have been proposed as a viable strategy in responding to this challenge – as detailed in a literature review prepared for this project (Prichard et al., 2022).

This study evaluated the reThink project, which aims to minimise the demand for such material, and to make users aware of the Stop It Now! support service when they attempt to search for CSAM on a legal pornography website, Pornhub in the UK. The reThink project is a collaboration between the company that operates Pornhub, Aylo (formerly MindGeek), and two child protection NGOs: Internet Watch Foundation (IWF) and the Lucy Faithfull Foundation (LFF).

Aylo uses a range of strategies to safeguard the Pornhub platform, including hash scanning against known CSAM, AI to detect potentially under-age material, human moderation of all content before it is published, and returning deterrence messages against a wide range of search terms which could indicate an interest in CSAM or NCII (Non-Consensual Intimate Images). ReThink was launched on 22nd February 2021 as an additional strategy. In short, if Pornhub users enter a search term associated with CSAM, they (1) receive a warning message (from 22 February 2021), and (2) a chatbot (operated by IWF) appears on their screen (from March 2022). Through the information provided in the warning, or by engaging with the chatbot, users are informed about the illegality of CSAM, and they are referred to the LFF's free, anonymous support and advice services, which are provided for people who are concerned about their attraction to CSAM.

The three main research questions for this evaluation are:

1. is reThink reducing CSAM-related searches on Pornhub in the UK?
2. is reThink increasing referrals to anonymous therapeutic services offered by LFF?
3. to what extent is the reThink chatbot contributing to that aim?

To examine these questions the following evaluation framework in Table 1 was proposed, including the range of anonymous data that could be provided by each partner to evaluate the success of the reThink initiative.

Table 1 reThink Evaluation Framework

	STAKEHOLDERS' OBJECTIVES & RESEARCH QUESTIONS	DATA OWNERSHIP	MEASURES
1	Examine whether the reThink chatbot has reduced CSAM searches on Pornhub	Aylo	· Number of sessions including CSAM searches · Number of CSAM searches
2	Determine whether the reThink chatbot is effective at engaging 'at-risk' users who have entered CSAM searches on Pornhub	Aylo	· Number of chatbot views per sessions · Number of chatbot triggers per session · Search types after chatbot trigger
		IWF	· Length of interaction time · Number of LFF URL clicks
3	Assess whether the reThink chatbot increases the number of individuals seeking support from LFF support and assistance	LFF	· Number of individual visits to helpline · Number of individual visits to Get Help info pages

1.1 Intervention overview

An intervention commenced in February of 2021 with a warning message being displayed on Pornhub in the UK whenever a user searched for a term that Aylo had designated as potentially relating to an interest in CSAM. This warning message was displayed for 12 months unchanged.

On the 11th March 2022, the warning message was supplemented with a chatbot (Figure 1) that had been developed by IWF. The chatbot was a simple conversational agent built on the Google Dialogflow ES platform. The chatbot allowed users to click on buttons to select a path forward to information or to manually enter text. The chatbot functionality was designed to minimise the risk of inappropriate responses, and leveraging the buttons provided an efficient way for users to quickly receive information about the support services available. Responses are predefined, and not made using generative AI. Approximately 80% of users preferred to interact with the chatbot via buttons, with the remainder typing in text interacting with the chatbot in a conversational style.

The chatbot conversations were monitored throughout the length of the intervention, which enabled it to be continuously improved to ensure it was as effective as possible at connecting users with support services. This included adding additional responses and modifying how the chatbot operated to ensure appropriate responses were provided to users.

The original warning message operated uninterrupted and unmodified for 14 months, after which it was modified twice to trial different wording and to test the effect of removing the chatbot. The warning message was modified on the 11th May 2023, and then again on the 8th June. The chatbot was disabled from the 6th of July through to the 3rd August. Data collection ended on the 31st August 2023.

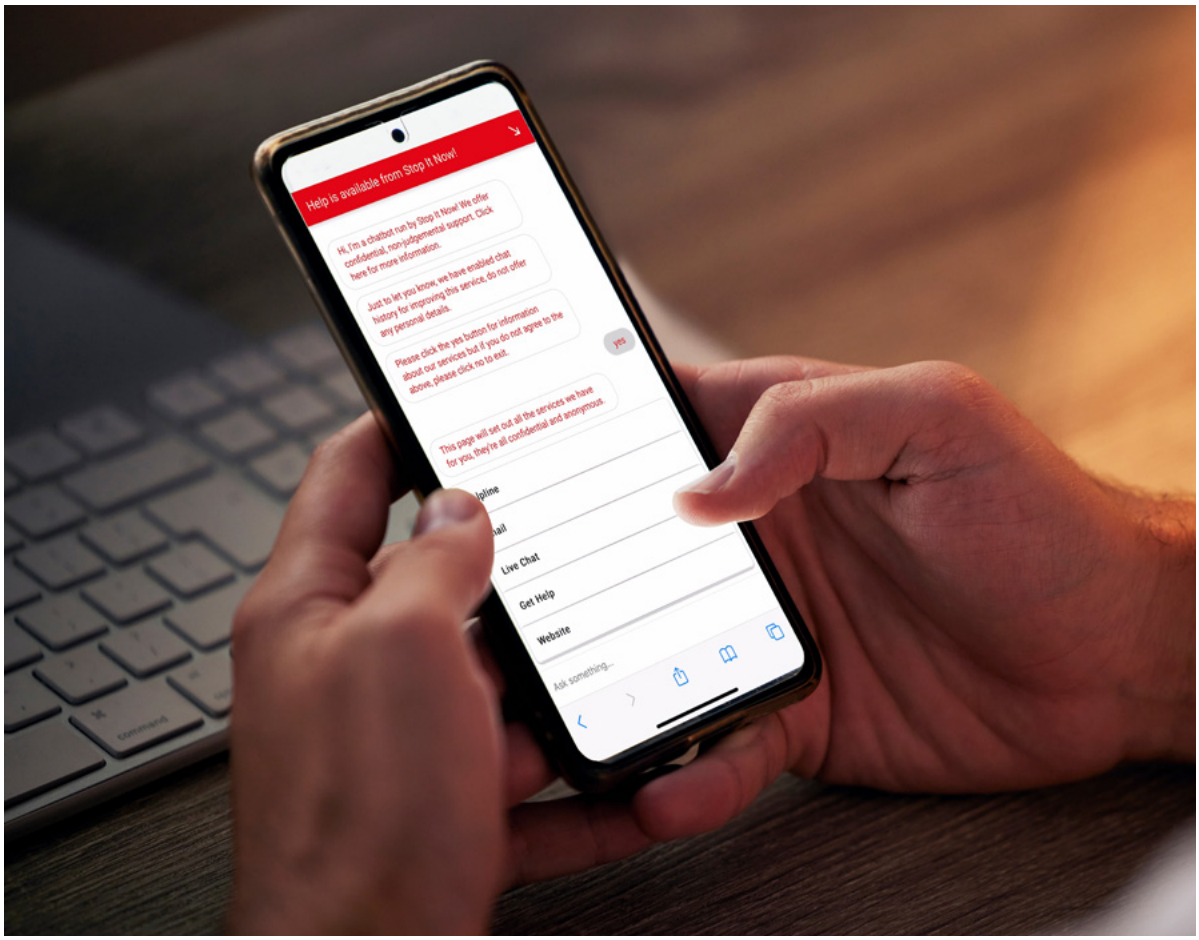


Figure 1
The reThink chatbot was deployed on Pornhub in the UK in March 2022.

1.2 Dataset description

All three partners provided anonymous data for the evaluation. This section will describe the data that were provided in relation to their type and the time periods they cover, in addition to a brief description of the data processing that occurred prior to the analysis.

1.2.1 Aylo

The data provided by Aylo were comprised of two components: search trigger summary data, and an export of user session data. The search trigger summary data included:

- a daily total of the number of searches undertaken on Pornhub in the UK using a potential CSAM-related term that triggered the warning message and chatbot; and
- the daily potential CSAM-related searches as a percentage of all searches on Pornhub in the UK.

This dataset spanned 14th July 2021 through to 14th January 2022, and then from the 8th March 2022 until the 31st August 2023.

The user session data covered all user sessions in which a potential CSAM-related search term triggered the warning message and chatbot. The terms that trigger the warning were selected by Aylo. The session data included the exact text used in each search and the titles of videos viewed within the session history. It is important to note that the session data are based on a tracking cookie stored by the browser on the client side – enabling such a session 'history' in many instances to cover long periods of time, covering weeks or months of access. The session length is not the time spent on the site, but the period between when they first visited the site and the last time they were seen. The number of visits within the session is not recorded in the dataset.

The user session data was exported in the form of a series of tab-separated flat files, where each row covered a single session containing: country, logged-in status, device type, date first seen, and date last seen, followed by a long list of activities including the searches and videos watched, and whether the warning page and chatbot was activated or not (i.e., did the search contain a potential CSAM related term or not). This dataset covered the 1st May 2022 until the 30th April 2023. It totalled 7.72 gigabytes of data containing over 4 billion words of search terms and video title text.

This dataset was processed and transformed into a database of tables to reduce duplication and to enable searches to be undertaken to best understand user behaviour.

...7.72 gigabytes of data containing over 4 billion words of search terms and video title text.

1.2.2 Lucy Faithfull Foundation (LFF)

The data provided by LFF comprised two primary categories relating to Stop It Now services: helpline data and web traffic data.

The helpline data included call volume statistics across a range of categories running from March 2020 through to August 2023. This data included the number of callers, emails, and people using the Stop It Now online chat service. In addition to this, further data were provided about contacts to the helpline that LFF had designated as originating from an interaction with the reThink chatbot.

The web traffic data were divided into three subcategories to enable comparison: visits to Stop It Now Get Help, visits to the Stop It Now UK and Ireland website, and direct referrals from the reThink warning message and chatbot.

1.2.3 Internet Watch Foundation (IWF)

The IWF facilitated access to two Google Cloud web portals (Logs Explorer and Dialogflow ES) that contained data relating to the operation of the chatbot. The goal of this was to enable an examination of the volume and duration of interactions with the chatbot.

Notably, the Logs Explorer (within the Google Cloud platform) only provides access to chat logs from the last 30 days. The evaluation team gained access in early 2023 which unfortunately meant we were unable to export activities for the bulk of the project runtime. Fortunately, LFF was able to share its own export of 132 chatbot conversations (sampled from July, September and December 2022, and March and May 2023). While this isn't a complete export of all chatbot conversations that occurred, it is seen as a meaningful sample of the types of interactions that occurred within the chatbot during the time of the intervention. The difficulty in exporting conversations across long time periods appears to be by design within the Dialogflow ES platform.

The IWF initially shared a file of the number of interactions that were occurring on a monthly basis in 2022 from Dialogflow ES. However, our own inspection of the statistics within Dialogflow ES revealed inconsistencies in this record. As a result, the IWF shared their internal exported values from Dialogflow ES which had been collected on a monthly basis through the project. These values are assumed to be correct for the purposes of this evaluation.

... a meaningful sample of the types of interactions that occurred within the chatbot during the time of the intervention.

2.0 Successes

The high-level summary of key measurable outcomes of the intervention is shown in Figure 2. The warning message and chatbot were displayed approximately 2.8 million times between March 2022 and September 2023.[1] Interactions with the chatbot by users resulted in 1,656 responses asking for more information and Stop It Now services, and there were 490 recorded click-throughs from the chatbot to the Stop It Now website. 68 calls or chats made to the Stop It Now helpline were identified as likely being prompted by the reThink chatbot and/or warning page on Pornhub in the UK. Prior to the chatbot's launch in March 2022, the warning message was displayed an additional 2,208,864 times, totalling 4,400,960 times over the length of the project.

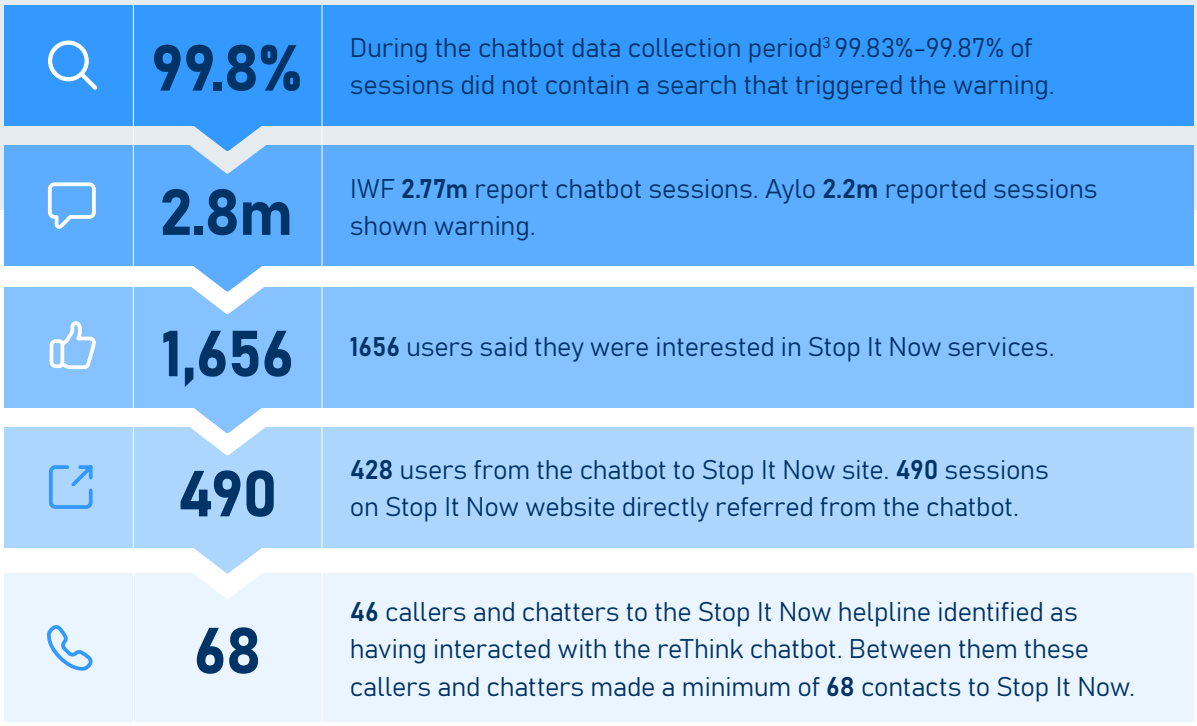


Figure 2
Summary of the conversion rate of chatbot sessions to helpline and web sessions on Stop It Now

This outcome demonstrates that there has been a clear benefit in the reThink project, as individuals have requested the support of the Stop It Now service because of the intervention. The evaluation will also show a clear deterrence effect, with a reduction in CSAM search volume on Pornhub. The remainder of this evaluation will explore the data in depth, and examine what the impact of reThink has been, and the value of the chatbot in line with the research questions. This section will look at several findings that can be considered to be Successes, and we will then examine other findings under the headings of Complexities and Limitations before considering what the Future Directions could be based on what has been learnt in this project.

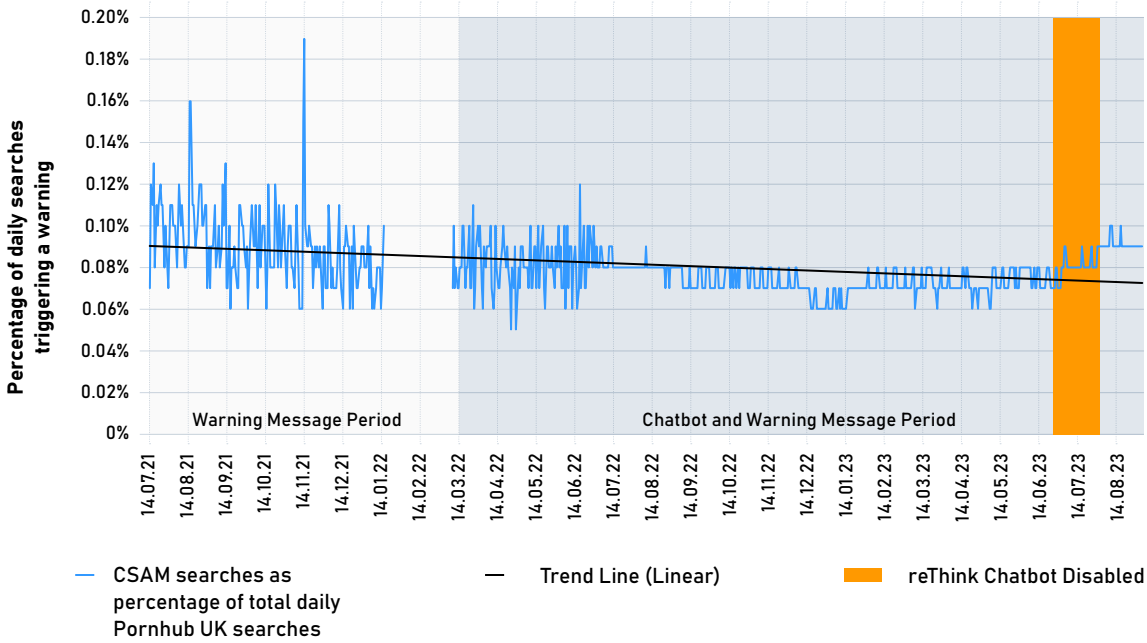
1. The exact number is hard to ascertain, as the data provided by the partners conflicts with one another. IWF reported 2,768,206 chatbot sessions while Aylo's Google Analytics data reported 2,192,096 across the full period. However, this disparity may be explained by Google Analytics reporting the number of users instead of sessions. The data export from Aylo running from May 2022 through April 2023 recorded 2,787,094 times in total that the chatbot was displayed – which is higher than the number IWF reports for a longer time period, further casting doubt on the statistics from Google Cloud – Dialogflow ES.
2. Figure 9 (page 18) shows the number of warning messages delivered before the chatbot commenced operations. However, the graph highlights this data is incomplete, so the total number of warning messages delivered in this period was higher than the 2,208,864 mentioned here.
3. This percentage is for the data collection period between March 2022 and September 2023, as data from the initial months of the warning message period (March to July 2021) or the baseline data prior to March 2021 was not available, which is discussed in the limitations section.

2.1 Targeting offenders or those at risk of offending

The intervention, as seen in Figure 2, resulted in the warning message and chatbot being shown a large number of times. In the broader context of Pornhub usage, site traffic during the intervention ranged between 2,320,000 and 4,514,000 sessions from the UK who searched for content daily. The chatbot was active for between 2,352 to a maximum daily count of 7,044 of these users who searched using potential CSAM related terms – which equates to between 0.05%-0.12% of all searches on Pornhub in the UK (Figure 3). Across the full period from July 2021 to June 2023, there is a downward trend with a standardized estimate = -0.719, p = .001, which does show a statistically significant and meaningful decrease in searches for CSAM over time on Pornhub during the intervention. While there may be several factors that could have contributed to this observed reduction over time, it appears that at least in part, the reThink intervention has contributed to a reduction in CSAM searches on Pornhub in the UK.

The CSAM search data in Figure 3 covers multiple phases, and a statistically significant difference is observed between almost every time period: between stage 1 (warning message only) to stage 2 (chatbot) p = .001; a statistically significant difference between stage 2 (chatbot) to stage 3 (no chatbot) p = .001; a statistically significant difference between stage 3 (no chatbot) to stage 4 (chatbot turned back on) p = .001; although no difference between stage 1 to stage 4.[4] In summary, a reduction in CSAM related searches was observed with the introduction of the chatbot. In addition to the entire project time frame, multiple smaller time windows are examined (Figure 7 and Figure 8), demonstrating a significant difference in the first 30 days of the chatbot's activation compared to the period 12 months prior. It also includes a more in-depth analysis of the period in July 2023 when the chatbot was briefly deactivated and then reactivated.

Figure 3
CSAM content search volume as a percentage of all search volume on Pornhub in the UK during the reThink chatbot intervention.



4. The sample size for stage 3 and stage 4, where the chatbot was turned off for a period and then back on, is a much smaller sample size and so this observed difference should be considered with that in mind.

These numbers occurred with the backdrop of the UK's population declining during the early stages of data collection by approximately 55,000 people (66,971,411 June 2022; The world Bank Group, 2023) before being projected to expand to approximately 67,844,315 by June 2023 (Office for National Statistics, 2022).

While it is encouraging that the proportion of sessions searching for CSAM on Pornhub is ≈0.10% (1 in a thousand), the raw numbers of searches are still quite confronting as can be seen in Table 2. Such numbers far outweigh the number of reports and prosecutions for possession of CSAM in the UK annually. The reThink intervention is, therefore, targeting some internet users who may already be CSAM users. The search terms listed within Table 2 indicate that the intention with these particular searches were unambiguously to find footage of child sexual abuse including terms like "child porn", "kids" and "little girl". This highlights that users being shown the warning message for these search terms and the chatbot are the population in need of the services provided by the Stop It Now program. Other terms that triggered the warning are more ambiguous, and less explicitly CSAM related, highlighting that the users targeted are very much on a spectrum of differing levels of interest in this material.

This likely range of users searching for CSAM aligns with prior work by Seto and Ahmed (2014) and many others (see e.g. Prichard et al., 2022: 9-10) who found that CSAM appears to not be restricted to individuals with paedophilic disorder, and many new CSAM users have had no previous sexual attraction to children.

Table 2
Popular CSAM search terms entered on Pornhub in the UK that resulted in the warning message and chatbot being displayed to the user.

SEARCH TERM	NUMBER OF TIMES SEARCHED	NUMBER OF TIMES SEARCHED COMBINED WITH OTHER TERMS
kids or kid	102,680	268,795
young teen	67,829	177,883
child or children	90,883	196,113
young girl	46,955	281,654
young boy	40,456	185,448
ddlg ('daddy daughter little girl')	40,002	55,365
loli	33,559	113,813
little girl	29,423	87,836
lolita	26,064	43,333
child porn	12,184	14,543

Figure 4, below, shows that, based on the counts of warning messages that have been triggered, three groups of user behaviours appear to be present. These three groups are:

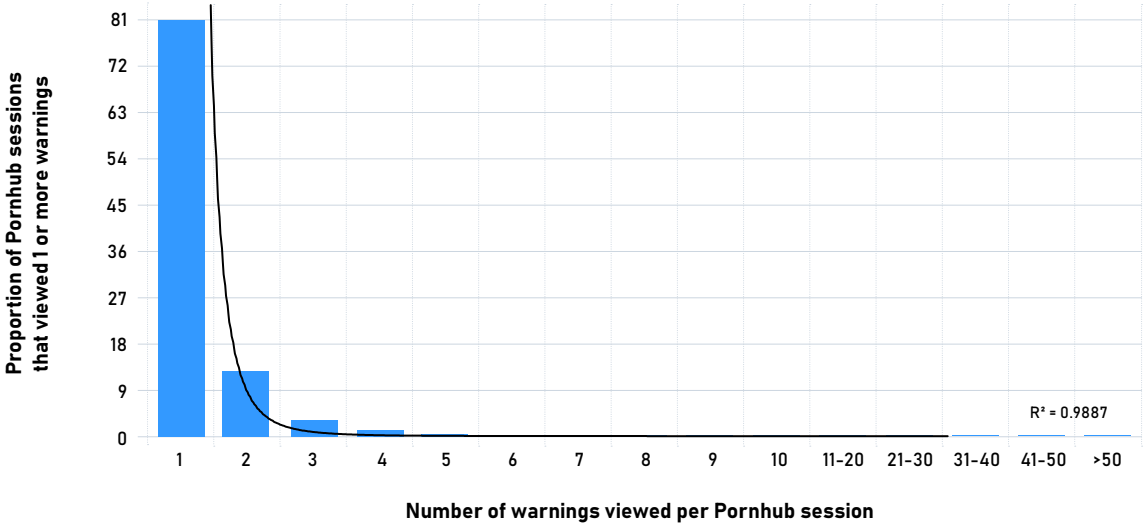
1. Never Group:
These are sessions where search activity never triggers a warning message. Approximately 99.83%-99.87% of sessions on the site each day fall into this category.

2. Desist Group:
These sessions have encountered a single warning during their time on the site based on their search term activity. Of the 2,178,513 sessions during the data period that did receive at least one warning, almost 82% only saw a single warning. A statistically significant downward trend was observed for sessions that only ever received one warning in response to searching for a CSAM term ($p < .001$), in which a greater proportion of users saw a warning in the first quartile of their session than any other time-period (ps < .001), with the proportion of users seeing a warning reducing across quartiles 2, 3, and 4 (ps < .001).

The downward trend across time for this group of users may suggest desistence behaviour occurring in response to observing a single warning. This desist group is the largest proportion of recorded sessions that triggered any warning messages on this site. Later, in the limitations section, it is discussed how we were not able to discount double-counting in the

data. We specifically have session data, which is not tied to an individual. However, the pattern is the predominant pattern in the data and so it is unlikely to be an artifact purely of double-counting of a smaller group of individuals due to its scale.

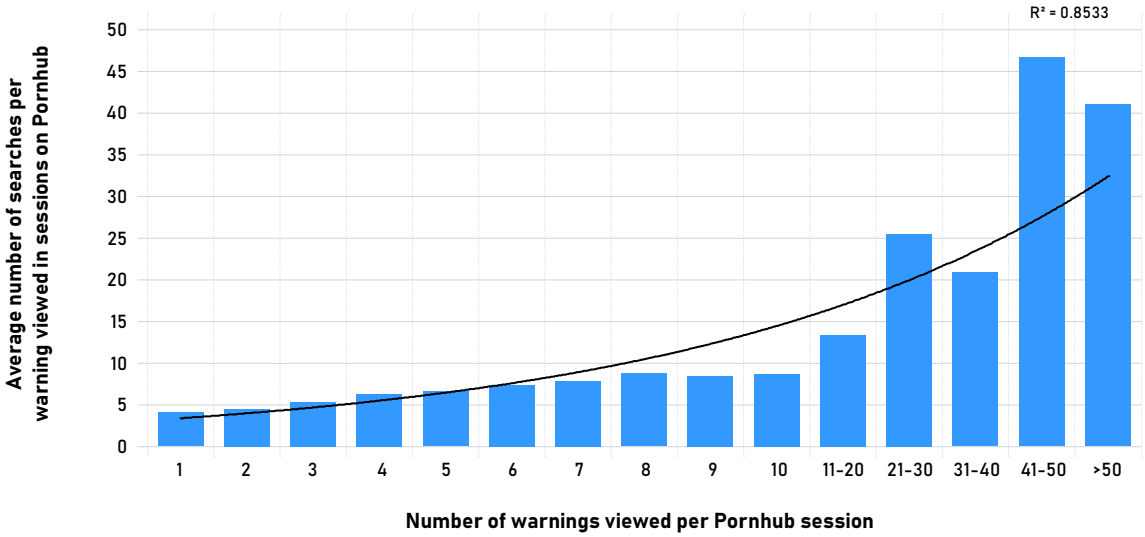
Figure 4
The number of times within a session the reThink chatbot was displayed as a result of a CSAM related search on Pornhub in the UK.



3. Persist Group:

This group of sessions appear to be a very small (n = 5637), but an important group whose search term behaviour does not appear to be affected by triggering and viewing warning messages in response to searching for a CSAM term on this site. This is illustrated in Figure 5 where sessions triggering the highest number of warning messages (e.g., 11 and over) also had a greater proportion of their searches triggering the warning, suggesting more specific interest in CSAM content, and less diversity in search terms, than other users with fewer warnings. In other words, this group tended to search Pornhub with the consistent and primary intention of finding CSAM.

Figure 5
The average number of searches undertaken in a session vs the number of times the reThink chatbot was displayed.



To explore the relationship between the warning messages and search behaviour, their activity was grouped into four time periods ('quartiles'), enabling an order effect element to be considered. If a session undertook 20 searches, then the quartiles would each contain five searches in the order they were undertaken. This then enabled an examination of the number of warning messages displayed following searches, and whether receiving a warning message deterred similar searches in the future.

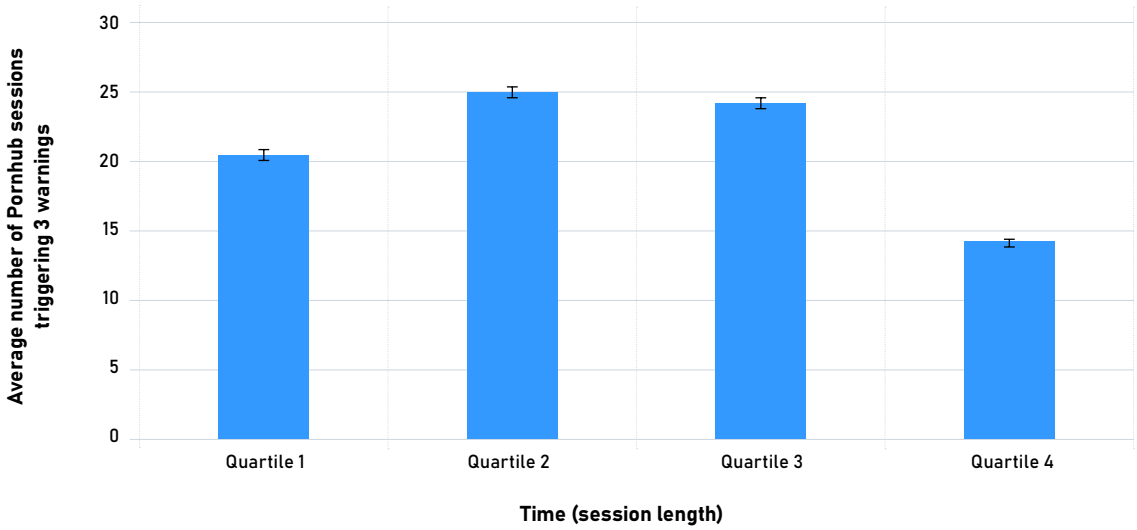
For sessions engaged in searches that triggered only one warning, a statistically significant difference is observed between each time period (time 1, time 2, time 3, time 4, and between time

1-4 p < .001). For sessions with seven or fewer warnings, the ratio of warnings to searches was also significantly lower in quartile 4 than in quartile 1, although the effect was small in magnitude (see Figure 6; the quartile comparison for sessions who triggered three warning messages is included here as an exemplar of the analysis). This suggests that viewing the warning was associated with a slight reduction in engagement with CSAM search terms from quartiles 1, 2, and 3, of time for sessions triggering seven or fewer warnings. For sessions that triggered eight or more warnings, the proportion of searches resulting in a warning in quartile 4 of time was not statistically lower than the proportion of searches resulting in a warning in quartile 1, although both were lower than the 2nd and 3rd quartiles.

This demonstrates that the warning message is having the desired effect of dissuading searches for CSAM material, but that it does not have an effect on a small cohort (0.15% of sessions that received the warning message and chatbot) who keep searching after receiving the warning seven times. This aligns with the desist and persist grouping already described.

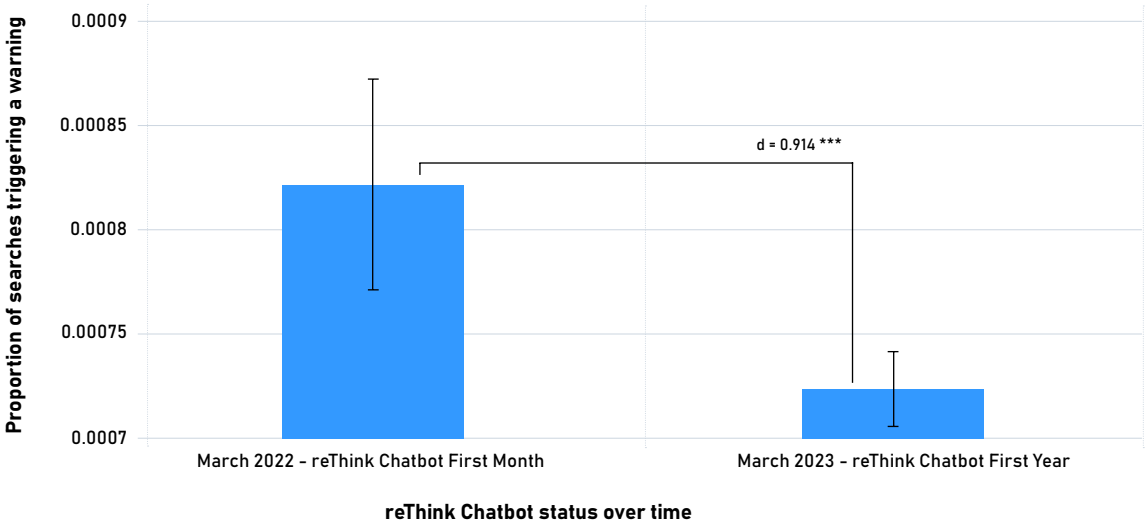
To explore the intentionality of users who visit Pornhub expressly for the purpose of searching for CSAM on Pornhub, an analysis of sessions whose first action was to undertake a search that triggered a warning was undertaken. This group was singled out for examination as the data provided by Aylo did not include timestamps to show when a session started and ended, so a behaviour trajectory cannot be ascertained. However, their very first activity would indicate to some degree the type of content they are interested in, as separate from an escalation after viewing other content. This group represented 36.56% of all sessions that received a warning (796,525 sessions). In comparing this group who did not receive a warning message for their first action, they undertook 80% fewer searches but did still watch a similar number of videos on the site (see Appendix). However, their average session lengths were far shorter, spending 1/5th the amount of time on the site.

Figure 6
Quartile comparison of sessions that triggered three warning messages as a result of their search behaviour.



Examining the effect of the reThink chatbot in the first 12 months of its implementation, a statistically significant effect was observed when comparing the first 30 days of the chatbot's launch (11th March 2022 to 11th April 2022) to the same time period 12 months later (11th March 2023 to 11th April 2023; Figure 7). What was observed between the two time points was the proportion of search terms on Pornhub that triggered a warning was statistically lower in March 2023 after the chatbot's successful operation for a year [t(60) = 3.63, p < .001, d = 0.914, 95% CI = 0.365, 1.46].

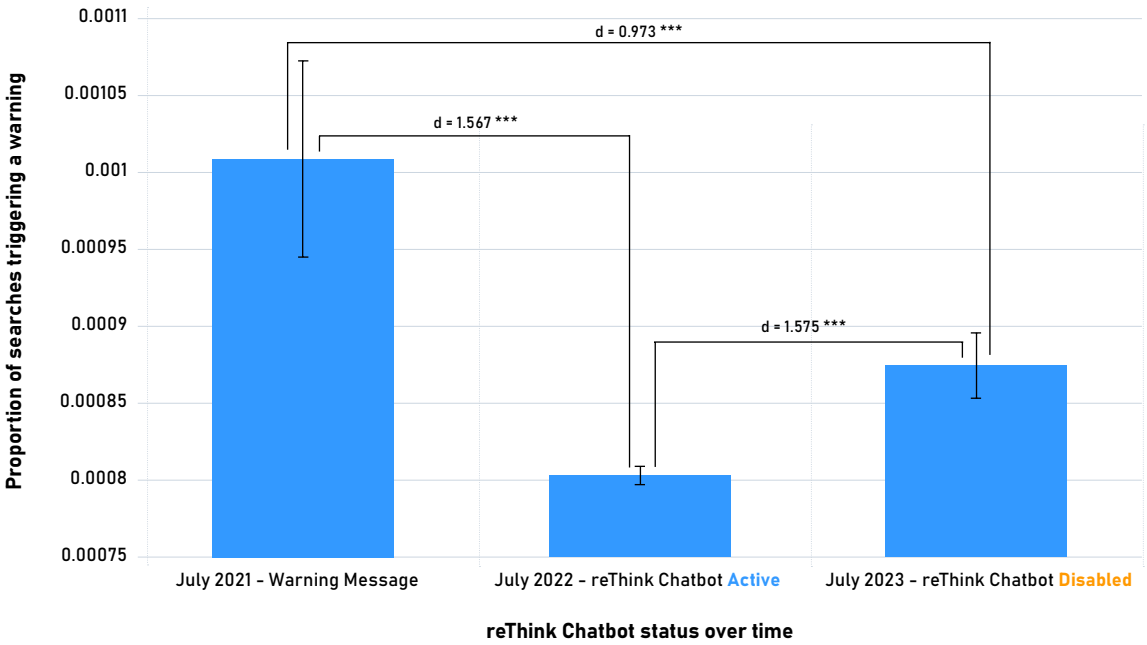
Figure 7
Proportion of CSAM searches on Pornhub in the UK triggering a warning during first year of the reThink chatbot intervention.



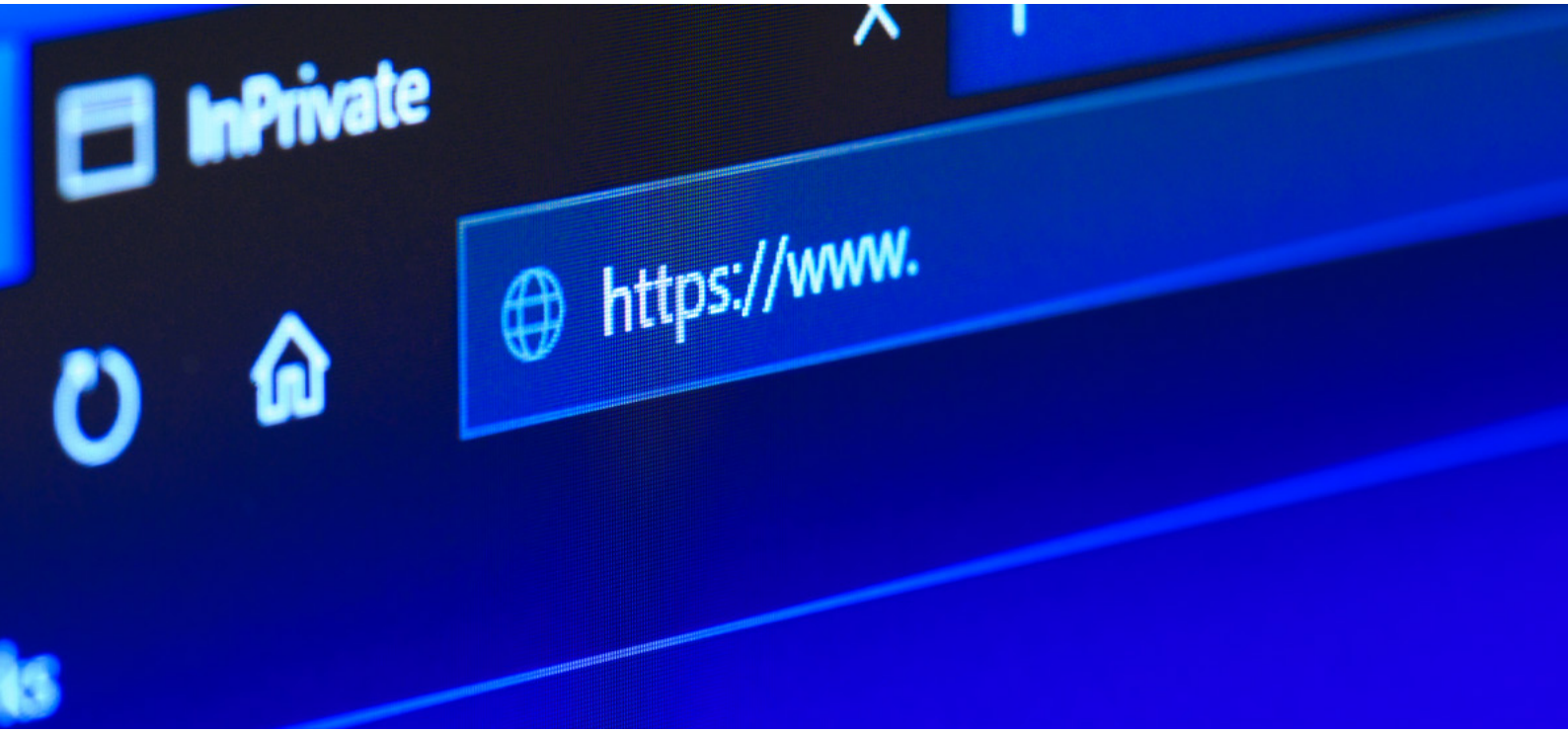
In July 2023, when the chatbot was temporarily deactivated for nearly one month, it created an opportunity to compare the effectiveness of the rethink chatbot on search term behaviour relative to the solely providing warning messages. This A-B-A design meant it was possible to compare the period in July 2022 when the chatbot was operating in conjunction with warning messages relative to the same duration of time in July 2021 and July 2023 when the chatbot was not operating (Figure 8). In short, the presence of the chatbot operating did have a statistically significant effect in reducing the proportion of searches that triggered a warning for CSAM [$F(2, 44.7) = 37.8$, $p < .001$].

The proportion of searches that triggered a warning was significantly higher before the chatbot was active in July 2021 ($M = 0.00101$, $SD = 0.000186$) than in July 2022 ($M = 0.000803$, $SD = 0.0000177$) after the chatbot was activated, $p < .001$, $d = 1.567$. In contrast, the proportion of searches that triggered a warning message was significantly lower in July 2022 than in July 2023 ($M = 0.000875$, $SD = 0.0000622$) when the chatbot was temporarily deactivated for approximately 1 month, $p = .034$, $d = 1.575$. The proportion of searches that triggered a warning message was also significantly lower in July 2023 than in July 2021, $p < .001$, $d = 0.973$.

Figure 8
Proportion of CSAM searches on Pornhub in the UK triggering a warning across three separate settings of the reThink Chatbot status.



These findings indicate that users who came to Pornhub looking for CSAM did not typically end their session upon receiving the warning, as one may expect, but instead proceeded to watch videos on the site prior to their session ending. It may demonstrate that this cohort doesn't perceive the search functionality to be useful, and instead browses based on existing categories or general content surfacing filters. Their sessions were shorter, but this could be indicative of higher usage of privacy modes within browsers by this cohort to hide activity, resulting in session tracking being less effective across longer time periods. Further investigation is needed to ascertain what behaviour they undertake after seeing a warning message, and whether they go to other websites or the dark web as a result.



2.2 Successes summary

In summary, the warning message and chatbot are being displayed to a cohort that is demonstrating an unambiguous interest in CSAM. Over the length of the intervention, there was a reduction in searches for CSAM on Pornhub in the UK. This deterrence effect is a key aim of the project.

In most cases, after seeing the warning message and chatbot, there were fewer subsequent searches for CSAM, which has been labelled here as desistance behaviour. However, there were some sessions where individuals persisted in searching for CSAM content despite repeatedly receiving the warning message and chatbot. Sessions that contained 7 or fewer warning messages tended towards a pattern of viewing fewer warning messages as their session progressed.

Finally, a sub-cohort of cases who could be seen to be motivated to search for CSAM were shown to not leave the website after receiving the warning message and chatbot, but instead to continue using the website watching legal alternative sexual content.

3.0 Complexities

While there are demonstratable positives out of the intervention, as the previous section has shown, there do remain some outcomes that are less clear, or raise questions to be answered about the overall efficacy of the chatbot, which could be resolved through further research.

The project contained datasets of quite different scales, formats and definitions of a user or session. Establishing the impact of millions of warning messages being displayed and the subsequent helpline contacts and web page visits in the 100s or 1000s does present an analysis challenge.

Figure 9 displays an overview of the project, with the multiple periods visible and their different scales visible across the two y-axes. This section will aim to describe what can be demonstrated amidst this complexity.

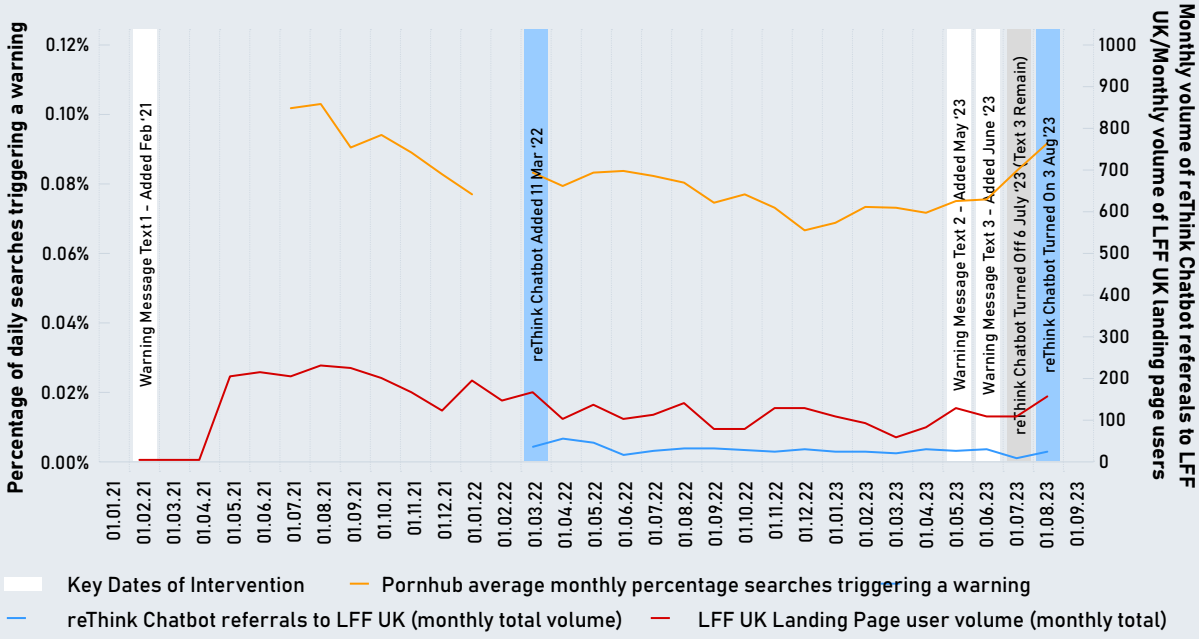
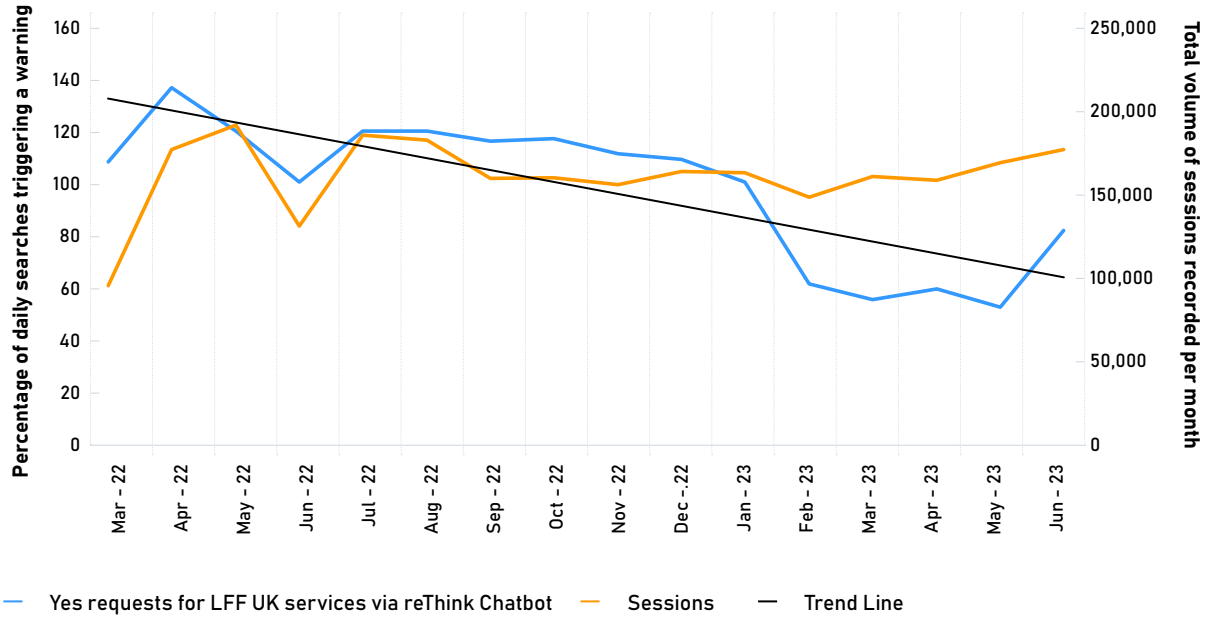


Figure 9
Comparison of the web traffic access for the Stop It Now website in the UK vs referral data from the reThink chatbot.

3.1 Evaluation of web traffic data

Over the length of the intervention the search volume of CSAM related terms decreased, however, the rate of requests for information about the services Lucy Faithfull Foundation provides also decreased (Figure 10, over page), and at a faster rate than the decrease in overall sessions of the chatbot active on Pornhub. This infers that the chatbot was less effective over time at garnering the interest of users and engaging with them to request information about LFF services. While there were 1,656 requests, it is worthwhile noting that some degree of 'saturation' possibly occurred in time. For example, over the time period studied, to the extent repeated sessions with CSAM searches may have originated from the same user, such users may have been less and less likely to engage with the warning message or chatbot each time they were displayed.

Figure 10
Overall trend in requests via the chatbot for information about the services offered by the Lucy Faithfull Foundation.



Further to the decline in requests for information from the chatbot in **Figure 10, Figure 11** shows the number of web referrals to Stop It Now website from the chatbot. After an initial surge, it settles into a fairly stable pattern of referrals. A drop-off occurred while the chatbot was disabled (which is to be expected); however, referrals return to the previous rate once it is turned back on. In total, there were 490 direct referrals to the Stop It Now website from the reThink chatbot. Also shown clearly in **Figure 11** are the periods where the chatbot was briefly changed and deactivated for a brief period in 2023. While an interesting experiment, the data collected in the window is too low for rigorous analysis of its impact, and it does appear to remain at an equivalent level.

Figure 11
Referral data from the reThink chatbot to Stop It Now website.

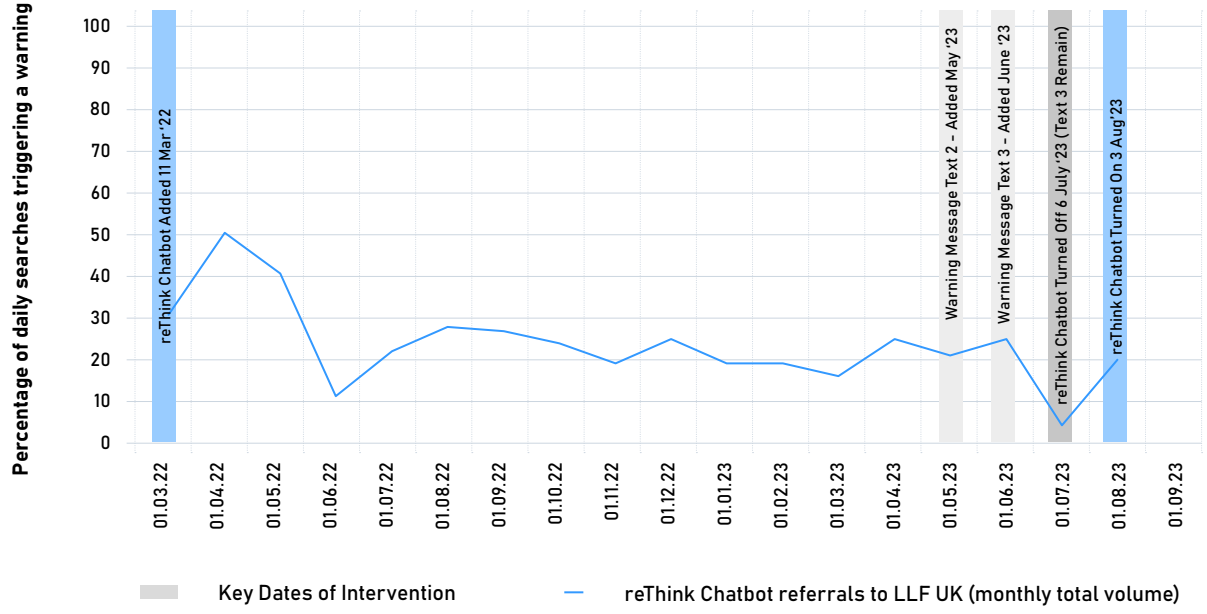


Figure 12
Monthly user referrals from the reThink chatbot to LFF website as a percentage of LFF Landing Page new users.

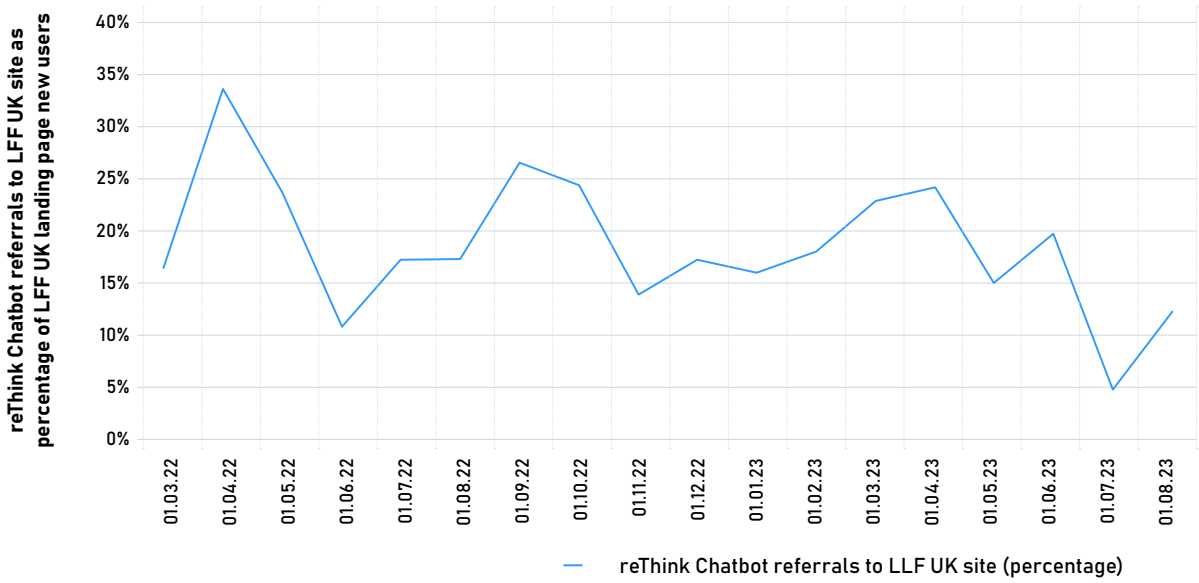


Figure 12 shows the number of new referrals from the chatbot to the LFF site as a percentage of LFF site's new users. The number of referrals occurring during the intervention from the reThink chatbot hovered around 20% of new users to the LFF site, which is a notable increase in the number of individuals seeking information from the LFF website. In Figure 9 (on page 18) a slight decrease in overall UK LFF traffic over the period of the study can be seen, but the role in countering this decrease with new referrals from reThink is challenging to firmly prove.

In summary, the web traffic analysis is positive when considering the number of referrals to the Stop It Now website. However, this is in the context that requests for information on LFF services within the chatbot decreased across the intervention. As there is still a substantial volume of CSAM searches on the site, future interventions must investigate how to increase the deterrence effect and raise engagement with preventative support materials to decrease the CSAM search behaviour further.

3.2 Evaluation of helpline data

One of the core aims of the interventions is to increase referrals to the anonymous therapeutic services LFF offers (the other being deterrence, to reduce CSAM search volume). The previous section showed that there was some decrease in visits to the Stop It Now website, there was a substantial percentage of referrals from the reThink intervention. This section will explore how these translated into direct engagement with LFF services such as the helpline.

The volume of helpline services during the intervention is shown to be increasing in Figure 13. Total helpline services include callers, emailers and chatters. Across the 32 months of helpline data available, the average monthly service volume was 919.84 (SD = 114.23). The lowest number of monthly helpline services occurred in December 2021, prior to the reThink chatbot being deployed. In comparison, the highest volume of helpline services engaged was observed in March 2023, following a year of the chatbot operating. In fact, in the final 12 months of helpline data, 10 of those months experienced above-average engagement with helpline services. However, reports produced by LFF have broadly demonstrated year-on-year increases in helpline service volume since its introduction, so establishing the direct impact of the reThink project is challenging.

Figure 13
Total Monthly volume of LFF helpline services.

However, in addition to the overall volume data of LFF services, data was recorded about contacts to the helpline. Within the notes taken of calls and chats during the study period, LFF was able to identify 68 contacts that appear to have been motivated to some degree by the reThink intervention, by the user mentioning interacting with a chatbot or receiving messaging of some kind on Pornhub. This list of contacts is considered a sample and not the actual number of contacts, as within this data, users state they interacted with chatbots or warning pages on a range of platforms, including 4Chan and Facebook. In some instances, the user could be withholding the actual location they saw the chatbot; in other warning message contexts, they could be from other sites and not part of the reThink project.

In a brief analysis of the sample, a few themes seemed to be present in terms of the types of calls/encounters with Stop It Now helpline. These were:

a) **Panic and distress:** these users seemed to have encountered a single warning and seemed worried, panicked, stressed and/or greatly distressed by encountering the warning. Several of these users seemed to call LFF worried they were going to be on a watch list, reported to the police or go to jail. Several users in this group also said "it's never happened before" and that they will never interact with porn again.

b) **Tech support:** these users seemed to be contacting LFF more for tech support to get rid of the warning rather than engage with the process. The general interactions described for these types of users appeared to be more about trying to remove the warning to enable them to go about the users' viewing intentions than anything about behaviour change. Anger and frustration seemed to be emotions expressed toward some LFF staff as a result of LFF staff not being able to remove the warning for the user.

c) **Help seekers:** these users seemed to be reaching out for help. The range of issues they were seeking help for varied. However, there appeared to be a portion of the users that were reaching out, sometimes on multiple occasions, for help, guidance, counselling, or other support to change their behaviour.

d) **Information seekers:** these appeared to be users who encountered a direction to contact LFF and wanted more information. Some of these seemed to be first-time encounters with a warning; many expressed disgust and/or disdain towards having searched for CSAM, coupled with stating they never had or would engage with CSAM, and wanted or took help for how to stay safe online. Some overlap appeared to occur between this group and the panic group.

e) **Time wasters:** some users appeared to contact LFF just to waste time. Some even appeared to seek self-gratification by calling the LFF line, and then proceeding to continue to use other materials while on the phone for personal satisfaction, likely at the discomfort of the LFF staff member. These users did not seem interested in the support, help, and behavioural change services available and offered to them.

These are quite interesting, although not broadly within the scope of work of this evaluation. However, these themes could be of value to LFF, or other organisations, to consider in the operation of helpline services, and further developments undertaken to improve the service. These themes could be investigated and be directly addressed with their own research questions.

3.3 Evaluation of recorded chatbot conversations

While the chatbot was able to engage with some users, resulting in 685 web referrals, and a number of contacts to the helpline (discussed in 3.2), there were a number of unsuccessful interactions with the chatbot worth examining. The successful interactions were typically by users who interacted only with the buttons within the chatbot (approximately 80% of users), whereas the examples of unsuccessful interactions were with users who corresponded with the chatbot via typing. This examination, however, was limited due to the Google chatbot platform having limited capacity to export chat dialogues en masse, resulting in 10% of all interactions being viewed within this evaluation. While this limitation is not ideal, this sample has revealed useful information about chatbots shortcomings. Of the 20% of free text conversations that were examined, half of these were handled appropriately, with the other half having a shortcoming in some form.

Many typed chatbot interactions were clearly motivated by frustration with not getting the results that they searched for, or that they were being informed that what was searched for was harmful (demonstrating a lack of understanding that the requested content was illegal). Here are some examples excerpts of such conversations, each numbered speech bubble is a conversation with commas separating what was entered by the user.

1. *'Can I please watch what ever I want on PornHub?'*
2. *'this filtering is shit i didnt searh anything illegal', 'what did i do stupid chatbot', 'i did nothing wrong', 'I HAVE STOPPED DUMBASS I DD NOTHING WRONG'*
3. *'why i like young girls?'*
4. *'Piss off', 'What the hell are ya', 'What the hell are you on about I don't need help ok', 'Fuck off'*
5. *'Hi', 'Do you love god?', 'sex?'*
6. *'I accidently searched child how do I get off this'.*

The chatbot is purposefully limited in how it responds, and it was ill-suited to deal with some of the responses entered into the chatbot, whether they were genuine or fake, with users trying to see how the chatbot responded. Either way, its inability to respond appropriately either did not provide assistance when needed or was taken to be a joke by users, neither of which are positive outcomes. This demonstrates there is room for improvement here in future implementations. Further research and technical development may be required to adequately detect and respond to the range of questions being asked by users.

4.0 Limitations

In any evaluation, it is vital to be aware of the limitations of the data that you have access to and the impact of the assumptions being made by those undertaking the evaluation. Within the reThink project there are several key factors relating to the data which impact how it has been interpreted and how the value of the intervention is considered.

A key element relating to how the data are perceived and compared between the different datasets provided by the partners, is the nature of what has been recorded. Multiple datasets, for example the Aylo data export, contained session data not user data. This data is from the same user, across a non-fixed time period, but only from a single device. The most common session length within this dataset is less than 24 hours (comprising 69.4% of all Pornhub sessions that triggered the chatbot). It is tempting to treat these like individual users; however, it is a safe assumption to believe many of these are repeat users, or the same users accessing the site across multiple devices.

Other data provided by Aylo and LFF make use of the Google Analytics platform – which includes a mixture of session and individual user data. But even this individual user data would not account for a user using multiple devices, or a browser privacy window. There is no 100% accurate way to observe traffic to a website on a true user-by-user basis, there will always be a degree of inaccuracy and double counting.

So, with consideration of this, when the effect of showing a warning message or chatbot to the population is calculated, it is not showing it to a completely naïve population, who have not seen the message before. Therefore, the diminishing impact over time is likely a result of the same users viewing the warning message even though they are a "new session" within the data.

The analysis describes a "desist" behaviour, which is encouraging to see that users were searching for CSAM terms less after seeing the warning message. However, this assumes that they didn't view the warning message, acknowledge that it meant that the content did not exist on Pornhub, and proceed to another website or service online to search for content there. As such, in some cases, possibly not desisting, but moving themselves to another outlet to continue their behaviour.

The evaluation examined the usage of a chatbot to provide support and referral to the Stop It Now helpline. However, due to issues with the chatbot platform, there was limited ability to export chatbot conversations for analysis. Only 132 of the 1650 conversations of note were able to be analysed. This did reveal that the chatbot did respond poorly in many typed interactions, so it is likely that more value would not have been extracted from a fuller sample. However, it is a clear limitation of the evaluation to only review a sample of 10%; although, it is worth noting that 80% of users opted to interact via buttons and not text entry.

Towards the end of the intervention, a series of changes (shown in Figure 11, on page 19) were made to the warning message text and chatbot in quick succession between May 2023 to August 2023. While the trialling alternatives were well intended, the small amount of data collected in the time windows had limited ability to identify the direct contribution each individual change specifically made to user engagement.

Finally, when the evaluation was first proposed, three time periods of data analysis were suggested. They were:

1. Before the introduction of the static Stop It Now warning (pre-February 2021);
2. During the static Stop It Now warning (February 2021-March 2022); and
3. During the reThink chatbot trial (March 2022-February 2023).

However, the data provided for analysis only covers the period July 2021 onwards – that is, periods 2 (partial) and 3. A comparison of the warning message to a true baseline, to separate the effect of the warning message from the chatbot, is not possible.

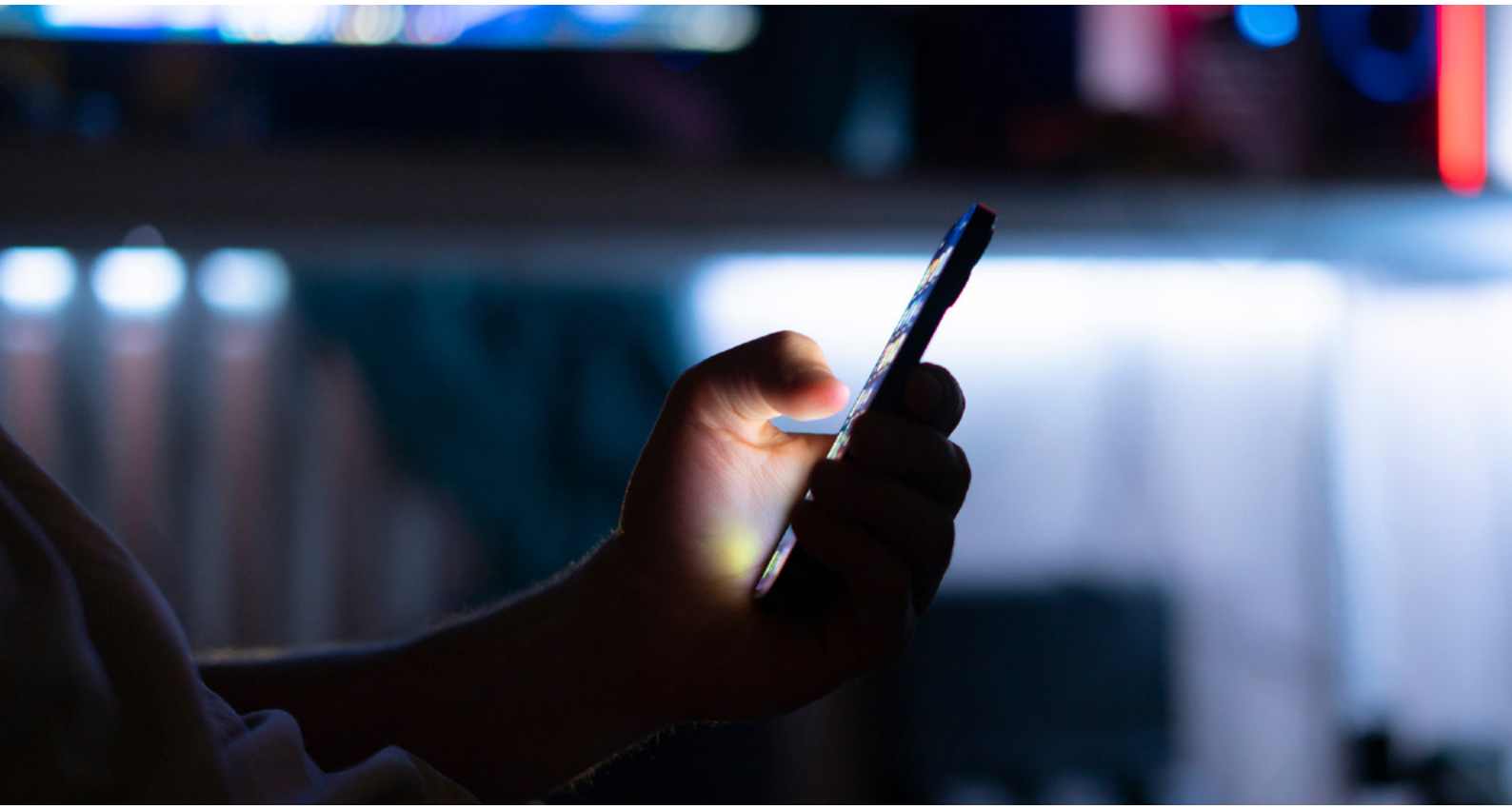
5.0 Future directions

The evaluation has shown that the warning message and chatbot on Pornhub did reduce CSAM searches during the period of the intervention. This point alone suggests that the intervention could be usefully replicated on other pornography sites, including those owned by Aylo. Similar interventions could also be valuable in non-pornography contexts, when tailored to the context.

However, reThink's impact diminished over the length of the intervention, and most markedly after the first three months. As a result, future efforts should focus on trying to recapture and lengthen the period of higher response rate in relation to the warning message and chatbot. One possibility could be to change the content of the warning message or chatbot initial message to users who have previously seen the warning.

A possible future direction is to leverage existing prevention efforts, to integrate the most recent Stop It Now advertising campaign into the warning message shown on Pornhub. The videos produced for these campaigns are of high quality and are targeting the population in question. A key takeaway from this intervention is that it is very much targeting a population searching for this material, whether intentional or not, and in sizeable numbers. Engaging them directly with existing campaigns to leverage the accessibility the platform provides to the target population could be a strong opportunity. This could add to the novelty of the warning page, as it changes with different campaigns over time, which could increase engagement with the warning. This would also add a situational crime prevention element to how these campaign videos are viewed, likely also enhancing their effectiveness.

Such an initiative would also serve to build in more changes to the warning message to ensure that it has a different visual appearance periodically. Habituation to seeing the same warning message could be a factor in the decreased engagement overtime. Making periodic changes to the warning message and introducing novelty may result in more sustained engagement with the messaging of the warning page.



The chatbot had a positive impact, gaining referrals to the Stop It Now website and calls to the helpline. However, the evaluation demonstrated that some users who interacted with the chatbot with typed messages had a negative experience or did not gain the support needed. Future versions of the chatbot should aim to be more complex, enabling them to be empathetic without being therapeutic in unexpected situations. It needs to be less one-size-fits-all and more adaptive to situations that arise. Hopefully the data that has been extracted from this intervention can be used to train, or at least inform the design, of an improved future version.

Several of the interactions seemed to relate to individuals who would be better served by a different service to Stop It Now – and perhaps they would be more receptive to assistance for depression or suicidal tendencies if a service like Samaritans was also included on the warning page. The visuals of including their logos, which are more commonly seen than that of LFF or IWF, may also have a positive effect in nudging users towards support services.

The chatbot could be improved by enabling it to connect directly to the Stop It Now live chat service, enabling a user to switch from the chatbot to a real person, enabling a smoother transition, and greater retention. Similarly, the chatbot could enable a call directly to the helpline through the computer/device the user is on. However, it is noted that the peak usage time for accessing Pornhub is 11pm-Midnight (2022) and the peak usage time of the reThink chatbot at 10 pm is when the Stop It Now helpline is not in operation[5]. Perhaps this is an opportunity for longer hours to leverage the recently launched helpline service in Australia to provide extended hours into the night.

Another possibility for future work could consider the themes of videos that are present on pornography sites and not only be limited to searches. A recent study examined the 'tags' (basically a keyword designated by the video uploader) of videos on a large mainstream pornography website (not Pornhub and not operated by Aylo) and had a panel of international experts undertake a thematic categorisation activity with 150 terms used (Gane et al.,2024). Further analysis of video titles could yield additional terms or phrases that could then be used to improve existing detection tools, or also be a valuable location for support resources to be advertised, even earlier on the trajectory, for individuals displaying an interest in CSAM content.

Finally, by design, the intervention had a separate data stream from the three partners. However, a more fully integrated intervention could track a user from end to end. Such a proposal would also include far greater privacy concerns and some technical hurdles in integrating systems; however, these are not insurmountable. Our own honeypot project (Scanlan et al., 2022) juggled such issues while appearing to the user as three separate platforms.

Hopefully the data that has been extracted from this intervention can be used to train, or at least inform the design, of an improved future version.

5. Helpline: Mondays – Thursday: 9am – 9pm, Fridays: 9am – 5pm, closed on weekends. Live chat: Mondays – Tuesday: 9am – 12pm, Wednesday: 1pm – 4pm and 6pm – 9pm, Thursday: 9am – 12pm, Fridays: 9am – 12pm, closed on weekends.

References

Dines, G. (2009). Childified women: How the mainstream porn industry sells child pornography to men. In S. Olfman (Ed.), The sexualization of childhood, 121–142. Praeger.

Gane, G., Watters, P., Wortley, R., Prichard, J. Blurring the lines: The vague boundary between mainstream and deviant Internet pornography tags for at-risk viewers. (Currently under review – expected in 2024).

Morichetta, A., Trevisan, M., Vassio, L., & Krickl, J. (2021). Understanding web pornography usage from traffic analysis. Computer Networks, 189, 107909.

Morgan, S., & Lambie, I. (2019). Understanding men who access sexualised images of children: exploratory interviews with offenders. Journal of Sexual Aggression, 25(1), 60-73. https://doi.org/10.1080/13552600.2018.1551502

Office for National Statistics (2022). Overview of the UK population: 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/2020

Pornhub (2022). 'The 2022 Year in Review'. https://www.pornhub.com/insights/2022-year-in-review

Prichard J, Scanlan J, Watters P, Wortley R, Hunn C, Garrett E,. (2022) 'Online messages to reduce users' engagement with child sexual abuse material: a review of relevant literature for the rethink Chatbot', University of Tasmania, Hobart. ISBN 978-1-922708-21-2. https://figshare.utas.edu.au/articles/report/Online_messages_to_reduce_users_engagement_with_child_sexual_abuse_material_a_review_of_relevant_literature_for_the_reThink_chatbot/23171162

Prichard, J., Spiranovic, C., Watters, P., & Lueg, C. (2013). Young people, child pornography, and subcultural norms on the Internet. Journal of the American Society for Information Science and Technology, 64(5), 992-1000.

Seto, M. C., & Ahmed, A. G. (2014). Treatment and management of child pornography use. Psychiatric Clinics of North America, 37(2), 207-214. https://doi.org/10.1016/j.psc.2014.03.004

Scanlan J, Watters PA, Prichard J, Hunn C, Spiranovic C, Wortley R. Creating Honeypots to Prevent Online Child Exploitation. Future Internet. 2022; 14(4):121. https://doi.org/10.3390/fi14040121

Warner, K. (2010). Sentencing for child pornography. Australian Law Journal, 84(6), 384–395.

The World Bank Group (2023). Population, total – United Kingdom. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=GB

Appendix

This appendix displays the details for the difference discussed in the body text in the behaviour between all sessions who received a warning, the subgroup who received a warning for their first activity undertaken on Pornhub, and the counter subgroup who received a warning after undertaking other searches first. The first group being the superset containing the combined data of the second two.

The dataset of session activity shown below is substantial but also skewed by high-volume users or possible bot activity (Pornhub does put considerable effort into minimising the activity of bots), as can be seen by the sizable standard deviation values for each column. Apparent skewed data of behaviour exist across the three subgroups shown. For example, the volume of videos that are listed as being viewed by users within a session seems to be quite high, especially in the context of Pornhub's own reporting in documents such as their 'Year in Review' (2022). Their reporting of an average duration of 10 minutes and 5 seconds in the UK in 2022 does not seem to align with the 100s of videos being watched. The primary cause of this is the data examined in this evaluation is video pages that have been loaded to represent a 'watch history', but it does not record how much they watched any of the videos. The 'Year in Review' is based on the actual time that videos played. However, in a scenario where a user 'clicked' on 10 videos, opening them all into new tabs but only watched one, it would still be recorded as 10 within the evaluation dataset. Further, the mode for watch counts is much lower at 3, with a median in the tables below ranging from 50-70. This demonstrates that it is a highly skewed dataset, with a small number of users viewing far more content than the rest of the population. The skewed nature of the viewing behaviour across the population aligns with Morichetta et al. (2021), who found that a small subgroup accessed web pornography far more frequently than the rest of the population.

Table 3
All sessions who received the warning page with the chatbot, grouped by the number of warnings received.

NUMBER OF WARNINGS	NUMBER OF SESSIONS	AVERAGE NUMBER OF SEARCHES	STANDARD DEVIATION	AVERAGE SESSION LENGTH IN DAYS	STANDARD DEVIATION	AVERAGE NUMBER VIDEOS WATCHED	STANDARD DEVIATION
1	1784093	24.79	57.47	14.94	34.16	151.28	192.24
2	279406	44.70	76.37	25.81	41.51	160.24	196.27
3	68635	56.30	87.4	30.60	43.56	170.26	199.8
4	26247	64.39	92.45	32.86	44.07	174.11	198.49
5	9725	76.64	103.05	36.97	45.17	187.57	205.28
6	4781	81.59	106.31	37.79	45.25	185.14	204.57
7	2195	89.94	108.44	40.46	45.65	198.27	209.97
8	1347	91.52	106.41	38.71	44.72	191.61	202.61
9	637	106.85	118.7	44.30	45.98	199.58	199.88
10	399	115.18	121.13	41.61	44.63	231.55	248.31

Table 4
The subgroup of sessions who received the warning page with the chatbot on their first activity upon visiting Pornhub in the UK, grouped by the number of warnings received.

NUMBER OF WARNINGS	NUMBER OF SESSIONS	AVERAGE NUMBER OF SEARCHES	STANDARD DEVIATION	AVERAGE SESSION LENGTH IN DAYS	STANDARD DEVIATION	AVERAGE NUMBER VIDEOS WATCHED	STANDARD DEVIATION
1	729712	2.81	7.87	0.68	6.82	154.57	192.74
2	49540	8.40	17.77	3.17	14.81	161.31	195.84
3	11037	11.57	24.36	4.65	17.74	161.95	194.14
4	3727	16.10	31.00	6.11	20.24	164.40	195.68
5	1245	19.90	41.51	7.65	23.20	166.60	198.72
6	608	21.93	38.86	7.72	22.47	156.81	183.01
7	250	26.16	51.55	11.97	28.68	167.02	194.77
8	144	30.89	51.75	12.67	27.87	174.01	192.41
9	66	25.86	35.03	9.27	22.26	152.58	182.70
10	50	29.34	29.91	12.42	29.86	212.20	261.98

Table 5
The subgroup who received the warning page with the chatbot after first undertaking other searches that did not result in a warning message, group by the number of warning messages received.

NUMBER OF WARNINGS	NUMBER OF SESSIONS	AVERAGE NUMBER OF SEARCHES	STANDARD DEVIATION	AVERAGE SESSION LENGTH IN DAYS	STANDARD DEVIATION	AVERAGE NUMBER VIDEOS WATCHED	STANDARD DEVIATION
1	1054381	40.01	70.56	24.80	41.29	149.00	191.86
2	229866	52.52	81.70	30.69	43.74	160.01	196.36
3	57598	64.88	92.37	35.57	45.24	171.86	200.82
4	22520	72.38	96.71	37.29	45.36	175.72	198.91
5	8480	84.96	106.69	41.27	46.00	190.64	206.05
6	4173	90.28	110.16	42.17	46.06	189.27	207.20
7	1945	98.14	111.08	44.12	46.13	202.29	211.51
8	1203	98.78	108.93	41.83	45.33	193.72	203.69
9	571	116.21	121.37	48.35	46.30	205.02	201.07
10	349	127.47	124.26	45.80	44.83	234.32	246.17

UNIVERSITY of TASMANIA

Published by the University of Tasmania
ISBN: 978-1-922708-67-0

Suggested citation:
Scanlan J, Prichard J, Hall LC, Watters P, Wortley R, 'reThink Chatbot Evaluation', University of Tasmania, Hobart. ISBN 978-1-922708-67-0 (2024)

Author contact details:
Joel Scanlan – joel.scanlan@utas.edu.au
Jeremy Prichard – jeremy.prichard@utas.edu.au

General enquiries:
media@iwf.org.uk