Evaluation of Safety Tech Challenge Fund Round 2

October 2024

PUBLIC

Department for
Science, Innovation,
& Technology

Contents

Executive Summary

The second round of the Safety Tech Challenge Fund, launched in February 2023, sought to address the issue of the sharing of links to sites that contain child sexual abuse material (CSAM). Issued by the Department for Science, Innovation and Technology (DSIT) and managed by Innovate UK (IUK), the challenge fund gave grants to three companies - CameraForensics, Centre for Factories of the Future (C4FF) and Vistalworks - to develop technology that could detect or disrupt shared links that route offenders to CSAM. Government Communication Headquarters (GCHQ), the Information Commissioner's Office (ICO) and the Home Office (the HMG stakeholders) worked alongside DSIT to support the development of these projects.

In February 2024, PUBLIC was commissioned by DSIT to undertake an independent small-scale evaluation of this round of the Safety Tech Challenge Fund (STCF 2). PUBLIC is a digital transformation partner to the UK Government, specialising in conducting evaluations of major Government technology funds and programmes.

This report contains our findings following a review of available evidence and interviews conducted with DSIT, IUK, the HMG stakeholders and the three funded project teams. Our evaluation comprises a process, impact, value for money and technical evaluation, in line with HMT Magenta Book best practice.

Our evaluation finds that while there are elements of the delivery of the fund that could have been improved, **the research and development conducted by fund recipients has led to novel and potentially impactful technology solutions**. We also found that the fund delivered its key objectives of stimulating research efforts in a sector where there is neither a mature customer demand nor an established market of solutions. The fact that the three funded projects all represented a pivot towards CSAM prevention from previous work also represents a positive benefit to the UK safety tech sector, with the work they have done **reinforcing the UK's position as a global safety tech leader**.

We find that some **inefficiencies in product development** could have been mitigated by a competition phase that focused more on the technical development of projects, rather than the commercialisation of technology. This problem was compounded by the fact that the adoption of standard IUK processes meant that there was not an opportunity for HMG stakeholders to follow-up on applications through further discussion with applicants or an interview. During delivery phases, we found that the principal challenges related to **establishing effective collaboration**, with some misalignment between fund recipients and HMG stakeholders on the propensity for HMG stakeholders to supply data and test the solutions developed. Finally, payment processes were found to have caused challenges for fund recipients, and is a further area where DSIT could consider changes for future funds. Based on these findings, we propose a number of considerations for changes to future funds.

Summary of findings

Process Evaluation

Project phase	Key successes	Key learnings
Market Research	- CSAM link sharing was an appropriate topic for STCF 2 to focus on given the scale of the problem it presents. - The industry roundtable provided a strong basis for understanding the policy landscape of CSAM link sharing.	- Data acquisition could have been anticipated as a challenge for projects. - Technical constraints of developing solutions were not fully explored. - Wider HMG teams and other networks can be used for identification of potential applicants.
Competition	- The IUK process was efficient and familiar to applicants who had experienced it previously. - HMG stakeholders collaborated well to evaluate applications.	- The inability to follow-up after initial applications limited the evaluation and prevented proactive remedying of risks in the fund recipients' plans. - HMG stakeholders and fund recipients felt the application process disproportionately focussed on commercialisation of products as opposed to technical development.[1]
Development	- The support of HMG stakeholders was impactful to project teams and aided development. - Some project teams were able to leverage data and clients they had access to from previous work for testing their solutions.	- There was a misalignment between HMG stakeholders and fund recipients on data access or testing that HMG stakeholders would offer. - Quarterly grant payments in arrears risks precluding smaller entities from participating.
Evaluation	- Those familiar with IUK reporting processes found ongoing reporting straightforward.	- Those unfamiliar with IUK reporting processes found the process was constraining to development.

[1] The commercialisation and technical questions each represented around 20% of the total application.

Impact and Value for Money Evaluation

Intended outcome	Key successes	Key learnings
Innovation and tech development	• The funding of three projects has ensured meaningful R&D in the area of detecting and disrupting link sharing to CSAM that would not otherwise have happened. • All three fund recipients pivoted towards CSAM distribution prevention, supporting the growth of UK safety tech and supported the UK's position as an industry leader.	• Further progress in link detection and disruption in E2EE environments would fill a gap that has not been addressed by fund recipients.
Partnerships and pathways to scale	• New partnerships were established by fund recipients as a consequence of participating in the Safety Tech Challenge Fund. • Projects may not be expected to be in a position to immediately scale, so progress is generally positive.	• There is scepticism about the ability for some fund recipients to immediately scale solutions.
Knowledge sharing and collaboration	• HMG stakeholders have used the learnings from the STCF 2 to inform their wider work. • The collaboration between HMG stakeholders was strong.	• At this time, there is limited perceptible increase in the attention to CSAM link sharing as a result of STCF 2. [2]

[2] It is anticipated that the conclusion of STCF 2 may facilitate further awareness raising of CSAM link sharing and the solutions developed by fund recipients.

Methodology

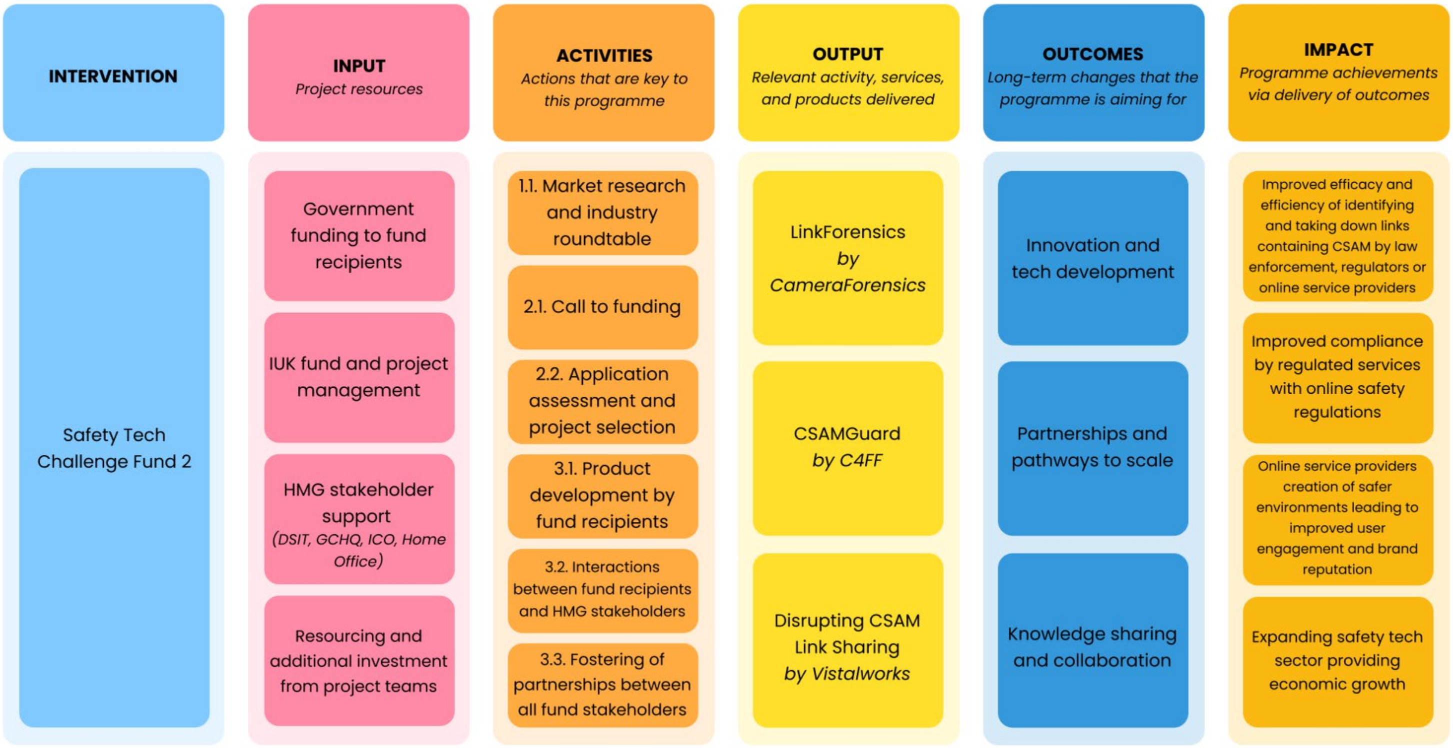
In conducting an evaluation, PUBLIC aimed to assess the delivery of the STCF 2 and identify lessons learned for DSIT and the wider sector.

The evaluation held four key objectives:

1. Assess the impact of project outcomes and learnings, the value for money of each intervention and the effectiveness of STCF 2's processes.
2. Conduct independent technical analysis of the projects, their outputs/tools and their efficacy in tackling CSAM link sharing.
3. Identify key issues, challenges, opportunities and lessons learned from each project.
4. Disseminate learnings with wider sectors through a detailed short evaluation report to spur innovation and adoption in this space.

Theory of Change

To support our evaluation, a Theory of Change (ToC) has been developed to align project activities, outputs and outcomes with measuring impact and benefits:



INTERVENTION

Safety Tech Challenge Fund 2

INPUT
Project resources

Government funding to fund recipients

IUK fund and project management

HMG stakeholder support *(DSIT, GCHQ, ICO, Home Office)*

Resourcing and additional investment from project teams

ACTIVITIES
Actions that are key to this programme

1.1. Market research and industry roundtable

2.1. Call to funding

2.2. Application assessment and project selection

3.1. Product development by fund recipients

3.2. Interactions between fund recipients and HMG stakeholders

3.3. Fostering of partnerships between all fund stakeholders

OUTPUT
Relevant activity, services, and products delivered

LinkForensics *by CameraForensics*

CSAMGuard *by C4FF*

Disrupting CSAM Link Sharing *by Vistalworks*

OUTCOMES
Long-term changes that the programme is aiming for

Innovation and tech development

Partnerships and pathways to scale

Knowledge sharing and collaboration

IMPACT
Programme achievements via delivery of outcomes

Improved efficacy and efficiency of identifying and taking down links containing CSAM by law enforcement, regulators or online service providers

Improved compliance by regulated services with online safety regulations

Online service providers creation of safer environments leading to improved user engagement and brand reputation

Expanding safety tech sector providing economic growth

Methodology

Our evaluation follows Magenta Book best practices, and structures evaluation in three parts:

1. **Process evaluation:** Assessed the approaches taken to deliver the programme, including appraising how funding and selection decisions were made, effectiveness of application approaches, and an assessment of funding by region, technology type and activity. PUBLIC assessed existing accounting data compiled to-date by the STCF 2 delivery partner, and incorporated analysis into the final report.
2. **Impact evaluation**: Measured whether the outcomes outlined in the programme's Theory of Change had been achieved, and assessed the extent to which they can be attributed to the programme. PUBLIC focused on short and intermediate-term success indicators and used primarily qualitative techniques, including exploring counterfactual scenarios with stakeholders.
3. **Value for Money evaluation:** Determined whether the spend on projects to-date is proportionate and represents good value-for-money based on available market data. This evaluation was particularly informed by the assessment of the technical choices made by DSIT, and comparing possible counterfactual costs and outcomes associated with other technology choices.

Additionally, to inform all three phases of the evaluation, we conducted an independent technical assessment of all proof-of-concept products developed as part of the fund. Where data and metrics were available, we have evaluated criteria including performance, robustness, scalability, data privacy and security, and user feedback.

Evidence to inform the evaluation has been taken from three main sources:

i. **Programme documentation**: PUBLIC collected programme and project documentation, as identified as outputs in the ToC for full review. This included DSIT's business case, call of funding, assessment reports, quarterly reports, and project plans.

ii. **Interviews**: PUBLIC conducted three rounds of interviews with the programme stakeholders: UK HMG bodies; fund recipients; and the delivery partner.

Interviews discussed processes, impact and learnings, and counterfactuals. Further information about the interview process can be found in the *Interview process and script* appendix.

iii. **Survey**: PUBLIC distributed a survey to collect evidence for the technical assessment. The survey was issued to fund recipients to collect performance metrics and data on the technical criteria. Further information about the survey can be found in the *Technical survey* appendix.

Limitations

In conducting our evaluation, PUBLIC identified a number of limitations that impacted our delivery approach. The following limitations have been identified and are detailed throughout the report.

1. **Timeline**: As noted in the Magenta Book, ideally evaluation activity begins at the start of the programme to align with all stakeholders on the criteria for evaluation and activities. This evaluation began in March 2024 as the final quarter of the delivery phase was concluding. To mitigate this risk, PUBLIC met with HMG stakeholders and reviewed previously set assessment criteria and programme scoping documentation, to align our evaluation criteria with that used by HMG stakeholders and IUK.
2. **Available data and evidence**: Given the scope of the project and sensitivities in sharing intellectual property or proprietary information, limitations on evidence and data availability was a risk. This would impact the assessment of outcomes and impact achieved. PUBLIC set-up operational mechanisms with project partners to build a secure data sharing pathway to extract impact and learnings from projects. Additionally, throughout the report it is noted when limited evidence has been provided.
3. **Technical testing availability**: To conduct a thorough and robust technical assessment, the proof of concepts would be tested in an independent laboratory environment, otherwise known as a sandbox environment. This evaluation was established as small-scale and as such with limited access to testing data mimicking real life scenarios, further technical assessment may be necessary. PUBLIC conducted a high-level independent technical

assessment with viable data, including metrics on performance of proof of concepts and self reporting surveys.

4. **Technical criteria**: Ideally, the criteria for technical assessment would be defined ahead of the competition phase of the programme and evaluated as part of the application assessment. This would enable consistency in monitoring proof of concept development and for transparency to fund recipients. To overcome this disparity, PUBLIC aligned the criteria with those used in the assessment of applications, those used independently by fund recipients, and the technical principles used in the previous Safety Tech Challenge 1 (STCF 1). The approach was built based on the best practices from the Department for Digital, Culture, Media and Sport's (DCMS) Online Safety Data Initiative.[3]

Summary of solutions

The three projects that were funded by STCF 2 were:

CameraForensics – 'LinkForensics'

CameraForensics

CameraForensics developed a system to automatically analyse the context, routing and destination content of shared links to identify features that are potential indicators of CSAM sharing. Combining new and existing tools, such as media and text classifiers, the system aimed to create an algorithm that could be deployed within a content moderation process to quickly determine the likely intent of a given link so that appropriate action could be taken.

[3] DCMS, Online Safety Data Initiative; The project was a shared outcomes fund project, led by DCMS. DCMS has been superseded by the new department, Department of Digital, Science, Innovation and Technology (DSIT).

Centre for Factories of the Future – 'CSAMGuard'

This project aimed to enhance online safety by addressing the modification and distribution of CSAM links through the utilisation of advanced machine learning-based models. This was to be achieved by developing state-of-the-art systems engineered to detect, disrupt, and report CSAM links to the relevant authorities. The project aimed to encompass the development of two distinct yet interlinked systems, a robust server-based solution and an innovative local router scanning system. These systems were to be strategically designed to collaborate seamlessly, guaranteeing a comprehensive approach to scanning, blocking, and expeditiously reporting of potential CSAM links.

Vistalworks– 'Applying methods & lessons learned from online illicit trade detection to CSA text links'

Vistalworks built on its previous work detecting online illicit trade and focused on tackling the sharing of CSAM links on the open web where links are traded, offenders are sign-posted elsewhere, and potential abuse targets are identified. By applying approaches previously used for combating illicit online trading on public platforms, it aimed to identify CSAM links (including those that have been modified) and generate reports for moderators.

Process Evaluation

For our appraisal of the different phases of the process of STCF 2, we have used the four-stage approach set out in the original DSIT[4] business case for the delivery of the fund. These are:

1. **Market Research.** Appointment of IUK to the challenge fund, and engagement with industry to ascertain appetite and suitability within the sector for this challenge fund.
2. **Competition.** The public launch of the fund with application for the fund from innovators and selection of projects to fund.
3. **Development.** Following the award of the grant, the beneficiaries working to deliver their projects
4. **Evaluation.** Ongoing evaluations of the projects, sharing any appropriate learnings with the sector, and using this to inform potential future challenge funds.

Our interviews and qualitative coding approaches[5] followed this broad structure, which we also use to structure our findings in this report.

Phase 1: Market Research

STCF 2 sought to tackle the issue of offenders' use and sharing of links to illegal CSAM online. Link sharing to CSAM is a significant online safety challenge; the WeProtect Global Alliance, Internet Watch Foundation (IWF) and INHOPE have all highlighted the prevalence of websites that host CSAM and are accessible on the open-web, while the briefing from a joint WeProtect Global Alliance and GCHQ expert roundtable on the link sharing to CSAM stated that there is “little available data on how companies are responding [to link sharing to CSAM], which makes it difficult to assess the

[4] The original business case was submitted by DCMS as it was then. Here, we refer to the department as it is now, DSIT, for clarity and consistency.

[5] Further information about the interview process can be found in the *Interview process and script* appendix

efficacy of responses". Taken together, encouraging innovations that targeted link sharing to CSAM was a worthy and reasonable focus area for STCF 2.

Within the challenge area of link sharing to CSAM, the first phase of the fund focussed on research to assess the viability of the challenge area to attract innovative providers to apply, establish a more specific challenge statement and understand the wider policy and technical landscape around the issue. The engagement was led by DSIT with notable support from GCHQ.

During interviews with stakeholders, it was indicated that an industry roundtable[6] held as part of this phase provided a strong and useful basis for understanding the policy area of link sharing to CSAM. The evidence of the technical challenges to developing solutions was, however, not discussed and understood to the same extent.

Access to data and technical feasibility

During the development phase of some projects, there was a large dependency on access to data for the development and testing of solutions. This dependency and risk to project plans may have been anticipated during the market research stage. Further discussions with key data stakeholders - especially IWF, who maintain an industry-respected list of known links containing CSAM - to provide more readily accessible training and testing data may have ensured that this dependency was more proactively addressed, facilitating project teams to avoid time-consuming data acquisition work. The IWF list does not, self-evidently, contain unknown links that contain CSAM; therefore access to this data would have supported some aspects of product development but would not have ensured that technology was indefectibly robust. It is notable that access to data related to CSAM is a prevailing challenge[7] in the development of safety tech solutions more generally and that the rigorous legal sensitivities around this kind of data are warranted.

Alternatively, the market research phase may have interrogated the feasibility and impact of using synthetic data for the training and testing of solutions. In future Safety Tech Challenge Funds, a more proactive assessment of potential data needs

[6] Activity 1.1. in the Theory of Change

[7] Perspective Economics, The UK Safety Tech Sector: 2023 Analysis, commissioned by DSIT

of projects would be welcome, with the value of this being discussed further in the Value for Money Evaluation.

Engagement with market and government stakeholders

As well as expanding the knowledge base around the challenge area, the market research phase was also intended to conduct early market engagement activities with potential fund applicants. This may include identifying potential applicants or ecosystem convenors who are relevant to the challenge and preparing the ground for the competition phase by engaging these organisations. While DSIT, the HMG stakeholders and IUK were rightly mindful to not disclose information about the fund to potential fund applicants ahead of the competition opening, engagement with firms new to the safety tech space but relevant to the innovation challenge may have attracted a wider set of applicants.

During interviews with stakeholders, wider engagement with a broader set of government departments, trusted networks and providers known to government colleagues was emphasised as activities which may have benefited the market research phase. This is likely to bolster market engagement in future funds.

Phase 2: Competition

The competition phase included the official launch of the challenge fund, the call for applications, the submission of applications and the evaluation of applications[8].

Use of Innovate UK competition process

This phase was managed by IUK and leant, largely, on the standard IUK processes for launching innovation competitions.

Using established IUK processes had key advantages, namely:

- The competition brief was compiled and launched in a professional and accessible manner;
- IUK's networks - through the Knowledge Transfer Network (KTN) and other channels - are extensive, meaning that the competition was made clear to a

[8] Activities 2.1. and 2.2. in the Theory of Change

large and diverse pool of innovators beyond the typical span of online safety tech providers that may have otherwise applied;

- The process was familiar to companies experienced in applying for or participating in IUK-ran innovation programmes previously.

However, during interviews, stakeholders highlighted elements of using the standard IUK process that may highlight adaptations to the standard IUK process for future rounds of the challenge fund. These were:

- The original, agreed scope for STCF 2 project did not include the ability to hold interviews or follow-up conversations after the submission of applications with applicants. The inability for interviews or asking of follow-up or clarifying questions to project teams limited assessor's ability to evaluate the proposals;
- Disallowing of follow-up conversations after the submission of applications also meant that concerns of the assessors to projects that were awarded funding could not be addressed earlier;
- The eligibility bar[9] for consideration of funding may have contributed to fewer projects being funded than may otherwise have been possible, although - as discussed in our *Value for Money Evaluation* - may also have ensured funding was not awarded to low-quality projects;
- Innovators new to the IUK process may have been disadvantaged compared to those who had participated previously[10].

HMG stakeholders reported that the application, which was IUK's standard and unaltered application template, did not ask enough technical questions with the technical questions asked not specific to the context of the fund. HMG stakeholders and fund recipients suggested too much focus was placed on questions about the commercialisation of solutions, with fund recipients also suggesting that this focus was not in line with their motivations for applying to the STCF 2. The scores for

[9] The eligibility bar was set at a fixed score, with an applications overall score being derived from the assessment of ten criteria which included topics such as market awareness, understanding of risks and project management approach.

[10] Applicants unfamiliar to an application process may have a natural disadvantage, irrespective of the fund management body and is therefore not unique to IUK.

questions evaluating commercialisation plans reflect this, and scored consistently lower than other questions for selected projects.

Additionally, the areas regarding the commercialisation of solutions were referred to in the Competition and Development phase as "exploitation". Given the topic area of the fund - child sexual exploitation and abuse - project recipients reported that this was an inappropriate use of language. Future challenge funds may be mindful of the choice of language used during operational elements of the fund.

As highlighted in the *Key Findings and Lessons Learned* section, the IUK standard process can be adapted following consultation with departments setting up challenge funds. The points raised here may form the basis of considering the scope of fund management services in future similar challenge funds. Similarly these points are likely to be inapplicable to other funds where the IUK standard process has been employed and do not serve as an evaluation of this process itself.

HMG stakeholders' role in application assessment

The assessment of applications received mixed reviews from stakeholders. As this challenge fund was on a relatively specific topic, stakeholders interviewed suggested that the briefing of initial assessors could have been more detailed. That being said, collaboration within organisations and between organisations to fairly and robustly evaluate applications was highlighted as a success and is something that should be replicated in future funds. From a policy perspective, the blend of HMG stakeholders involved resulted in a strong, holistic evaluation of application that included review of data protection approaches as well as addressing the online safety challenge.

The technical evaluation of projects fell largely to GCHQ. The strength and quality of GCHQ's technical review meant that outcomes are likely to have been unaffected on STCF 2 by having a single organisation responsible for the technical assessment. However, the fact that technical approaches were so significant to this challenge fund and are likely to be central to future challenge funds means that diversifying the sources of technical assessment would be welcome.

As aforementioned, IUK's processes are subject to modification on a case-by-case basis following discussions on scoping with the customer, in this case DSIT. In future Safety Tech Challenge Funds, the lessons learned from STCF 2's competition phase should be incorporated into the scoping of requirements for future fund managers. In line with feedback from HMG stakeholders, the ability to hold follow-on conversations or conduct interviews with applicants is the most impactful change required.

Phase 3: Development

The development phase incorporates a variety of phases. We have evaluated each in turn.

Onboarding

Once successful applicants were notified that they would be participating in STCF 2, they were onboarded onto the programme.

Due diligence and supplier assessment

The rigour of financial and operational checks that IUK conducted when project teams when first onboarding to the fund was raised as disproportionately intrusive by some fund recipients. However, this was not a view shared by all fund recipients and is likely caused by familiarity with the IUK process.

Project planning and scoping

Fund recipients universally welcomed the input from HMG stakeholders, especially GCHQ and ICO during the initial stages of the project. More specific and tailored support regarding developing a process for dealing with potentially illegal content was raised as one area where further guidance could be offered during onboarding. Provision of this support is likely to ensure that time is used developing technology rather than establishing legal security. Furthermore, advertising that this support will be provided may provide extra assurances to those new to safety tech that their legal risk will be mitigated and increase the number of applicants to the fund by lowering the cost of participating.

One of the selected projects required a change to their project plan as a result of feedback from assessors. The identification of the challenge and a course correction to update the plan was conducted efficiently and the operational project change request was reported as being straightforward. Recognition of the need to update development plans when issues are identified is welcome and the ease of this should be embraced in future Safety Tech Challenge Funds.

Product development

During the main product development phase[11] of STCF 2, the fund recipients presented and met with the HMG stakeholders to get feedback and guidance on their progress[12]. When the purpose of meetings to discuss progress was clear and had a specific desired outcome, they were successful for both fund recipients and HMG stakeholders. The ICO reported use of a questionnaire at the outset of the project to ascertain the data protection knowledge and understanding of project teams, which enabled clear and impactful support. Fund recipients also noted the value of ICO support, given that getting guidance on data protection compliance is either difficult or expensive to get from a trusted source. The technical support of GCHQ was also welcomed by fund recipients, but support sessions were more efficient when an agenda was established, something that should be encouraged in future funds.

Access to data and product development

The access and use of data was a key feature of the product development of some of the projects. While some fund recipients were able to use data acquired through their existing solutions, others had to acquire data for training and development of AI models used. Multiple HMG stakeholders reported that more clearly outlining, from the Competition phase, the data needs and acquisition plans would have enabled more efficient product development. The application form for STCF 2 only made passing reference to the impact of data acquisition, asking applicants under the risks section to highlight "any project inputs that are critical to completion, such as resources, expertise, data sets." This was unlikely to be sufficient for capturing full data needs and could have been anticipated as a risk.

This issue was compounded by a misalignment between HMG stakeholders and fund recipients on the extent to which the data held by the HMG stakeholders could be used or whether they would be able to test solutions. Clarifying whether or not HMG

[11] Activity 3.1. in the Theory of Change

[12] Activity 3.2. in the Theory of Change

stakeholders were able to provide any data for training or testing would be of benefit to future challenge funds. It is expected that legal sensitivities around sharing data related to CSAM would preclude HMG stakeholders sharing any data, but this should be made more explicitly clear if it is the case.

IWF holds a lot of information that can be used for training and testing of solutions to CSAM. In future funds, where the data they hold can be anticipated to be useful to applying projects - as it was for STCF 2 - their active participation in the fund may make product development more efficient.

Coordination with other projects

Some fund recipients also highlighted the lack of coordination between the other projects funded by STCF 2. As the projects' solutions targeted different elements of the CSAM link sharing value chain, there may have been some value in further coordination between them. Quarterly workshops where project teams discuss their projects and explore opportunities for integrations or more general support may lead to greater advances and the development of more holistic solutions.

Testing and identifying a route to market

The fund recipients had mixed results with regards to testing their solutions in the real world. While some projects were able to leverage existing contacts and secure live trials[13], other projects which were more research focussed did not secure external partners for testing. There appears to have been some misalignment between fund recipients and HMG stakeholders on the role that HMG stakeholders would play in supporting acquisition of early adopters of developed solutions. There was hope from some fund recipients that the HMG stakeholders would be early adopters of the solutions or that they would more proactively connect project teams with potential leads. Here again, clarifying whether this level of support would be beneficial for project teams, either at the competition or onboarding phases would help fund recipients plan accordingly. This is not to say that HMG stakeholders may not test or promote the use of developed tools following the conclusion of the fund and the development of products.

[13] Activity 3.3. in the Theory of Change

Consistent with the Competition phase, the commercialisation plans of project teams have been of lower priority compared to product development. As will be discussed in the Impact Evaluation, it is not possible to be confident in the scalability of solutions at this stage and whether a clear route to market has been established.

Fund management

During the product development phase, IUK managed the payment of funds to project teams. Funds were paid quarterly and in arrears for work completed. While this is a reasonable process that is manageable for larger companies with strong cash flow, it proved to be a challenge for smaller fund recipients. One project team reported that staff had to defer salaries at one point as a result of this funding approach.

IUK are able to adjust their funding approach if a need for this is identified during the scoping phase. To encourage new entrants to the safety tech space, where cash flow may be a challenge, having a portion of funds paid in advance of work would be welcome.

Phase 4: Evaluation

While this work forms the key element of the evaluation phase, IUK also ran their own monitoring of projects during project delivery. IUK reported that all reports were gathered from project teams on time and to expectation, but there were some mixed responses from fund recipients on this process.

Those that were familiar with IUK reporting found the process straightforward and familiar, while those with less experience found it disproportionately resource-intensive to complete. The focus of reporting on commercialisation was also something that was highlighted as a challenge for fund recipients, as this was not the focus of projects.

Impact Evaluation

The impact of the STCF 2 has been assessed against the core objectives and desired outcomes outlined in DSIT's original business case for funding. We have interpreted these desired outcomes to be:

1. **Innovation and tech development.** Supporting growth and innovation of the UK safety tech sector by developing more effective solutions for the detection of links to child sexual abuse material by platforms.
2. **Partnerships and pathways to scale.** Helping companies to establish partnerships, identify potential adopters of their technology, and develop pathways to scale (e.g., go to market strategy).
3. **Knowledge sharing and collaboration.** Generating and disseminating insights on the challenge of link sharing to CSAM and possible solutions that can be leveraged by the online safety ecosystem. In doing so, supporting the development of collaboration between domestic and international stakeholders.

These outcomes duly form the basis of the relevant aspect of the Theory of Change for STCF 2.

Innovation and tech development

On a fundamental basis, STCF 2 has been successful in stimulating innovation in the safety tech industry as the products developed are in direct response to the challenge and to the funding. In particular, the fund was successful in providing funding and support for a technology area that lacks other forms of funding.

All of the project teams reported that they would not have conducted the work they did either to the extent it has been done over the last nine months or at all had it not been for STCF 2 funding. In an area where the technological approaches to preventing link sharing to CSAM are not well known or tested, the fact that three projects have assessed the area and tried to develop solutions is notable and underpins this outcome being delivered. The technical progress of the fund recipients is discussed further in the Technical Evaluation section of this report, but all stakeholders interviewed have reflected that at least some progress has been made.

While not a stated as a specific objective of the fund, exploring solutions that could be used in end-to-end encrypted (E2EE) environments was included as a theme that STCF 2 could explore. None of the fund recipients explicitly explored the impact of E2EE on their solutions. Indeed two of the funded projects rely on the link being supplied to their product as opposed to being automatically retrieved from public or private online spaces. Future interventions, through challenge funds or otherwise, may wish to explore the impact of E2EE on detecting and disrupting link sharing to CSAM, especially as these environments may be likely to be used by offenders. It is important to note that due consideration to maintaining user's privacy must be maintained in any such project, if the commercial imperative for solutions to be developed by providers does not emerge.

The STCF 2 has also been successful in attracting new providers into the safety tech ecosystem. While Vistalworks and CameraForensics' previous work has been adjacent to the safety tech space, the STCF 2 marks a pivot for all three providers into a core area of safety tech development to tackle CSAM. IUK's delivery and use of their networks to promote STCF 2 has contributed to attracting new entrants to the safety tech space and future challenge funds should seek to identify and access networks where new entrants may be attracted.

The UK's position as a leader in safety tech has also been reaffirmed by STCF 2. Innovators have held discussions or piloted their solutions with organisations in Europe and North America.

Partnerships and pathways to scale

As discussed in the Testing and identifying a route to market element of the process evaluation, the ability of project teams to scale their solutions is not assured. However, HMG stakeholders noted they think this is to be expected at the end of a nine month innovation period, especially as this development was new ground for the fund recipients. The balance of tech development from STCF 2 funded projects, compared to commercialisation strategy is largely appropriate on the basis that further product development and establishing firmer pathways to scale happen in the future, which can be expected.

The development that has been possible has been facilitated in many cases by establishing new partnerships with key stakeholders, domestically and internationally. Fund recipients reported that they felt they were able to develop these partnerships as a result of participating in STCF 2 and due to the strength of the 'Safety Tech Challenge Fund' brand. This provides clear evidence of the fund delivering on its objective to help establish new partnerships and, importantly, this can support providers in future developments beyond the STCF 2.

Fund recipients also did not report any direct link between participation on STCF 2 and new interest from investors. However, whilst this would have been a positive outcome from STCF2, it was not an explicit objective, nor was it something that participating businesses sought, so it is not reasonable to consider this an unfulfilled outcome.

Knowledge sharing and collaboration

In general, the collaboration of STCF 2 stakeholders - particularly HMG stakeholders and fund recipients - was reported to be strong. As discussed in the Process Evaluation, this knowledge sharing was supportive of product development for fund recipients, but there is also evidence that this process was impactful for HMG stakeholders also. For example, the ICO reported that participation helped inform their approach to content moderation and data protection, while other stakeholders reflected that the learnings from the STCF 2 will be used to shape future similar work.

Knowledge sharing between HMG stakeholders was also something that was highlighted as successful. While there were some initial inefficiencies with the governance and process for the engagement between HMG stakeholders, these were quickly resolved and ensured that productive discussions could take place. Some fund recipients and HMG stakeholders indicated that Ofcom could be better integrated into the delivery of the fund. However, given their role as the independent online safety regulator and what this means regarding their future ability to enforce the proactive use of specific technologies in certain circumstances, there are reasonable grounds for them to not be a formal delivery partner on an innovation competition like this. In the wrap-up of STCF 2 and future funds, developing ways - in concert with Ofcom - that learnings can be shared between Ofcom, other HMG stakeholders and fund recipients would be of value.

The *Knowledge sharing and collaboration* impact also included the outward raising of the fund and the issue of the link sharing to CSAM. There is limited evidence that there has been significant knowledge sharing to the wider trust and safety ecosystem. Google Trends does not indicate significant traffic to the search of "safety tech challenge fund", while engagement on platforms such as X (formerly Twitter) and LinkedIn with STCF 2 is low. In general, it is not possible to attribute any increase in the prominence of link sharing to CSAM as an issue to STCF 2, although more rigorous methods to detect issue awareness may be required to validate this. Fund recipients suggested that some trade-offs from their communication and socialisation plans had to be made to allocate sufficient resources to technical development. There is also no evidence of a formal external communications plan by IUK or other HMG stakeholders. The conclusion of STCF 2 and this evaluation report, however, may form a good basis for further external communications. In future challenge funds, allocating a central resource in either DSIT or the fund management function for a single communications plan on behalf of the whole fund may provide an efficient way to share knowledge to the wider ecosystem.

Value for Money Evaluation

Given the focus of the programme was on developing early-stage technology projects, it is not possible to conduct a typical VfM evaluation in line with Magenta or Green Book recommended practices.

Instead, our Value for Money evaluation has focused on answering the following research questions (RQ):

- **Research Question 1:** Has STCF 2 followed good practices in maximising value for money?
- **Research Question 2:** Has STCF 2 funded technology projects in a proportionate way?
- **Research Question 3:** Has funding these projects represented a good use of HMG funding?

Following qualitative and quantitative data analysis, the findings have been captured below for each research question.

Assessment of STCF 2’s Value for Money

RQ1: Has STCF 2 followed good practices in maximising value for money?

This question aims to assess the extent to which funding and resources dedicated to the programme have been used effectively. Overall the evaluation found that the good practices were conducted in developing and delivering the challenge fund to develop new novel proof of concepts.

Critically, the practices that supported effective value for money were:

1. **Market research:** Market research defined specific challenge areas relevant to the sector and identified opportunity areas for startup growth addressing a specific threat.
2. **HMG stakeholder expertise:** HMG stakeholder support throughout project delivery was invaluable to ensure robust technical approaches and user privacy was at the forefront of proof of concept development.

During Phase 1 of the fund, the market scoping laid the groundwork for the challenge fund to be specific to the growing threat area of CSAM link sharing. Specifically, the roundtable with experts and market actors was key to ensure the challenge fund was both topical and addressing priority areas for the market, and engaged a wider network of actors to gain support and share learnings. This informed definition of the challenge area and understanding market maturity, set up the fund to appropriately select startups and address a growing threat where current solutions are nascent.

Throughout all interviews with stakeholders, from fund recipients to HMG stakeholders, the value of support and advice of HMG stakeholders was emphasised. Stakeholders shared how this directly informed development to better protect user and data privacy, as well as overcome unexpected technical challenges. This represents especially good value for money as the only cost of the support was the relatively limited time dedicated to it by HMG stakeholders, who themselves derived some value from the interactions in line with their job roles.

Our analysis has identified three main ways that delivery processes could have been improved to ensure best value of money:

1. **Lack of technical assessment during selection:** The limited focus on technical aspects of project plans during the selection process undermined the technical development of products and led to some inefficient use of time that could have been redeployed on higher-value tasks, if the selection process had uncovered weaknesses earlier.
2. **Low safety tech network engagement:** There was a lack of community and partnership building due to the absence of a dedicated communications and socialisation workstream.
3. **Delivery risks due to payment schedules:** Quarterly payments made in arrears risked limiting innovation and resourcing capabilities for projects. This includes the risk that smaller recipients may have faced real cash flow challenges, putting DSIT’s grant investment at risk.

During the selection process, IUK deployed a standard application and selection process which helped streamline the process. If the lack of specific questions relevant to the technical aspects of projects and absence of interviews as part of the competition phase had been mitigated, the initial product development phase of some projects may have also been more efficient. However, the use of off-the-shelf processes ensured an overall efficient and smooth execution.

The programme did not have a designated marketing or socialisation workstream or resource, leading to little engagement during the development of solutions. Given that one of the target outcomes of the Theory of Change was sharing learnings across the sector, the programme may have gained additional value for money from developing a low-cost central resource for sharing learnings iteratively with the wider sector on the behalf of all fund recipients. In previous challenge funds and innovation programmes, communications, marketing, and campaigns efforts proved to be beneficial in establishing a positive brand reputation. For example, BridgeAI, the UK's innovation fund to foster startup's development of AI solutions in target sectors, included an event launch, and BSI communities website for interested ecosystem actors, driving positive press and a flagship branding for AI innovation in the UK. The positive perception of the Safety Tech Challenge Fund as a programme has been cited by STCF 2 stakeholders as assisting them to secure new partnerships, demonstrating the long-term value for money central communications activity can offer.

Investing in a central communications function of this sort would also offer additional value for money by supporting the efforts to fulfil additional outcomes. This may include attracting safety tech investors and providing a basis to promote the whole safety tech sector.

IUK used its standard quarterly in arrears payment schedule for STCF 2. A different payment structure may have been more cost-efficient as some startups had little runway and, in some cases, even reported delaying salary payments due to the delayed nature of payments in arrears. It was also noted by a fund recipient that the reporting structure as part of the grant monitoring process was extensive and on a quarterly basis. While we recognise monitoring processes are critical to ensure proper use of funding, there could be opportunities to explore more efficient payment processes.

Payment structures where portions of payments are made up front to kick off R&D workstreams would be more welcoming for early stage startups looking to engage in the fund. It is suggested to de-risk misuse of funding that the portion of funding provided up front is minimal. A cost-benefit analysis would have to be conducted to determine the deal amount that is not high risk, but enables team to overcome the kick off threshold of cost. A risk assessment should be assessed for providing a portion of the funding up front, and it shall be split across While there is risk to provide funding up front, the challenge fund should assess the level of maturity of solutions in the market to assess what funding structure would incentivise participation and spur development, not hinder it.

RQ2: Has STCF 2 funded technology projects in a proportionate way?

This question aims to assess whether the amount of funding for the programme was appropriate and proportionate, in order to execute the programme's target outcomes. To ascertain whether grant funding was proportionate and appropriately used, we asked all stakeholders (including fund recipients and HMG delivery partners) to consider two counterfactual scenarios, in which:

1. the fund did not exist; and
2. there was a 50% reduction in the total grant.

It is apparent that without STCF 2, or a significant reduction in the funding for each recipient, the outcomes delivered through the programme would not have happened, or would have been reduced. In particular, it was noted by stakeholders that there would not have been a significant commercial imperative for similar technology to be produced and all fund recipients indicated that any reduction in funding would not have made participation viable. As such, due to STCF 2 being successful in leading to the development of technology against a high-priority online safety challenge, it can be broadly said to have delivered value for money. Overall, the proportionate funding achieved in developing solutions that were tested in laboratory or relevant environments to achieve TRL of 3 or higher.

As mentioned, a significant aspect for some fund recipients' product development was access to data. From a VfM perspective, having a central sandbox, trusted research environment, secure access to synthetic datasets or other mechanisms for permitting tools to access CSAM-related data for training and testing may offer good value for money. However, centralising data access in this way is a feature for a number of DSIT and other HMG stakeholder projects. As such, it may be the case that access to safe and secure data for a future safety tech challenge fund may be included as part of a larger business case for a project to develop one.

STCF 1 funded more projects (five) at a lower budget each (£85,000), while STCF 2 funded fewer projects (three) at a higher budget each (up to £120,000). Projects for STCF 1 received a per project budget of £85,000 with stretch funding for add-on workstreams of £129,500 total across two projects. This led to a total budget spent on STCF 1 was £555,000, while it was £700,000 for STCF 2. While the overall delivery outputs across the two programmes (STCF 1 and 2) are broadly similar, differences in budget can mostly be explained by the fact that STCF 2 had an additional focus on commercialisation, compared with STCF 1. A number of project teams were required to redeploy budgets originally allocated for communications or commercialisation towards technical development. While these changes were approved and made good value for money, additional support that is appropriately timed towards the end of any future challenge fund may ensure that budget reserved for these activities is optimised.

RQ3: Has funding these projects represented a good use of HMG funding?

This question aims to assess whether the funding and resources attributed to the programme is a good use of public funding. This research question is more difficult to answer, given the phase of the projects in question. It is not possible, at this stage, to assess the return on investment that may be generated by the fund, as fund recipients' products have not yet been made commercially available and have only been trialled.

STCF 2 made funding available for up to five companies, but only three were funded. Stakeholders reported that this was due to only three projects meeting the assessment criteria to receive funding and therefore funded fewer projects that met the criteria, over funding weaker proposals at risk of low-impact projects. Given the assessment committee deemed only three were high-impact to receive funding, this was an appropriate approach and likely ensured positive value for money. In the absence, however, of definitive evidence of an economic return on the spend, it is not possible to ascertain whether a marginal difference would have materialised - positive or negative - by funding two additional projects.

Instead, we can answer the question by assessing the original rationale for the programme, and its funding. The challenge fund fills the need for funding to develop child safety focused solutions where there is a lack in the commercial market. While investor activity has increased slowly over the years, there is still very little engagement, relative to other markets, in safety tech investment generally. More specifically, even more limited funding and development resourcing for solutions delivering link analysis and detection of CSEA harms has been made available. This is mainly due to the specificity of technical development needed, and commercial R&D budgets addressing other priority areas such as generative AI threats. The challenge fund goes beyond legislative goals promoting innovation and competition in new markets. The challenge fund also delivers technological development fit for a wide array of actors in safety tech as it engages all actor types from commercial partners, law enforcement, and third sector reporting bodies.

In conclusion, while a full Value for Money evaluation with a robust quantitative analysis could not be conducted, there is sufficient evidence to suggest STCF 2 represented good value for money, across all three research questions. Although there may have been some additional value for money possible through improved delivery processes, or targeted additional investment, no significant financial inefficiencies have been identified.

Recommendations for future Value for Money Evaluations

As discussed, a full Value for Money evaluation could not be conducted due to the developed products being pre-commercialisation. However, here we recommend a framework that would facilitate a full economic value for money evaluation at an appropriate stage, pulling on the HMT Magenta Book guidance. The Magenta Book outlines two key methods for conducting full value for money evaluations. These are:

- A **social cost-effectiveness analysis**, which “compares the costs of alternative ways of producing the same or similar outputs” without attributing a monetary value to the outputs as doing so is not viable.[14]
- A **social cost-benefit analysis**, which “goes further to assess the impact of different interventions on social welfare with all relevant costs and benefits valued in monetary terms (where proportionate and possible)”.[15]

[14] HMT, Magenta Book

[15] ibid

A future Value for Money evaluation would likely combine these two approaches, monetising costs and benefits for process improvements directly, but also using social cost-effectiveness approaches for more complex outcomes relating to the safeguarding of children, where benefits are less directly monetisable.

Drawing on the work of Ipsos and Perspective Economics' Trust and Safety and the Digital Economy, we recommend the use of some of the following benefit types to conduct a future Value for Money evaluation.[16] These metrics are in line with the *Impact* outline in STCF 2's Theory of Change:

Benefit	Description and applicability to STCF 2
Increased efficiency of content moderation processes	The products developed are likely to reduce the time it may take to conduct an analysis of links to see if they contain CSAM. An increase in the efficiency due to using tools may be captured by the actor using tools or used to review a greater quantity of links. Here, users of tools may include law enforcement agencies or regulators monitoring illegal or non-compliant activity or by online service providers on their own services.
Increased efficacy of content moderation processes	The products developed are likely to facilitate an increase in the accuracy and efficacy of content moderation and the detection of links to CSAM content. An increase in accuracy, especially in combination with an increase in efficiency, may lead to more site take down and a reduction in online CSAM.
Reduced non-compliance with online safety regulation	Products developed may help regulated services to comply with online safety regulation (such as the Online Safety Act and Digital Services Act) by proactively preventing CSAM, a priority illegal harm. In addition to the benefit of improving the efficiency and efficacy of moderation processes, this may also lead to a reduction in fines paid by firms.
Improved user engagement	The Trust and Safety and the Digital Economy cites the ability for improved online safety to deliver improved user engagement, both online and offline. Deploying tools may therefore lead to monetisable benefits from creating safer online environments.

[16] Ipsos and Perspective Economics, Trust and Safety and the Digital Economy

Improved brand reputation	An online service provider’s ability to create safe online environments may also ensure that their brand is strengthened. This can lead to the attraction and retention of advertisers.

Beyond these core benefit areas, a future Value for Money assessment of the programme may investigate ancillary benefit types such as:

- Increased retention of staff and users within target organisation
- Improved insights about users, leading to better services
- Better alignment of organisational values with user expectations

The benefits detailed in this section would likely be realised across the private sector (especially online service providers), public sector (especially CSAM law enforcement agencies), and the third-sector (especially reporting bodies, CSAM advocacy and research groups).

Technical Evaluation

Approach

Based on best practices from DCMS' Online Safety Data Initiative, technical evaluation criteria from STCF 1 and PUBLIC's past experience, we have used five technical criteria (see table below) for our analysis.[17]

Criteria	High-Level Definition	Source(s) of Analysis
Performance	The extent to which the product performs accurately and effectively.	• Supplier technical survey • Supplier one-to-one interviews
Robustness	The extent to which the product performs equally effectively when faced with perturbations and variations in content.	• Supplier technical survey
Scalability	The extent to which a product can maintain stable, effective performance, including during or after a steep increase in workload.	• Supplier technical survey
Data Privacy / Security	The ability of the product to maintain data privacy in design, development and deployment. This includes compliance with data protection legislation and cyber security standards and best practices.	• Supplier technical survey • Supplier one-to-one interviews • Quarterly reports
User Feedback	The extent to which the product was designed, built and tested in a user-centric way.	• Supplier technical survey • Quarterly reports

We also considered technical limitations, either identified proactively by the supplier or through our own independent assessment.

[17] DCMS, Online Safety Data Initiative.

PUBLIC then used industry-standard **Technical Readiness Levels (TRL)** to assess the maturity level reached by products at the end of STCF 2.[18]

Using these criteria, we assessed the technical proof of concepts against three research questions:

- **Research Question 1:** Were appropriate methods used to develop proof of concepts?
- **Research Question 2:** Did the projects achieve what was set out to be built?
- **Research Question 3:** Are these proof of concepts new and novel across the safety tech sector?

The findings for each research question are summarised below. Overviews of the projects can be seen in the *Summary of solutions* section and the acknowledged limitations of the technical evaluation can be found in the *Limitations* section for reference.

RQ1: Were appropriate methods used to develop proof of concepts?

Specific methodologies were set out by each project plan which emphasised development of operational proof of concept that could be scalable for real world implementation. Each project plan emphasised the need to put data security and privacy at the forefront of solution exploration.

All project plans included a thorough research phase to inform development specific to real-world environments and shape datasets to build and test the proof of concepts. Following acquiring testing data, solutions tested their data in laboratory or real world environments with partners. Commercialisation or scalability was explored but was not a focus of the programme given the short development timeline of nine months.

Findings across the relevant criteria of Scalability, Data Privacy/Security, and User Feedback are below.

[18] UK Research and Innovation, Eligibility of technology readiness levels (TRL)

Scalability

Scalability approaches were largely similar across suppliers, building on cloud architectures for rapid scalability of compute and deployment. All three suppliers measured throughput as part of scalability testing, with Vistalworks applying the broadest range of metrics (Uptime, Throughput, Latency), leveraging Microsoft Azure cloud infrastructure.

While Vistalworks claimed their solution can scale easily from just 1,000 rows of test IWF data to 100,000s of rows, further work is needed to assess this claim against a large, representative test dataset, ideally in a live production environment. Despite that, they take a mature approach to model scalability and deployability, building on their commercial experience, monitoring uptime, downtime, speed, false positives and match rates.

Similarly, C4FF's solution currently offers good throughput speeds of 5 seconds per request and average 7 seconds for bulk URLs processing in its current environment. 5 seconds per request and average 7 seconds for bulk URLs processing.

Based on information available, we agree with their claim that "these results suggest that the system can efficiently handle moderate workloads within acceptable response times."

The team has also considered future steps to improve the scalability when faced with a steep increase in workload:

1. Load Balancing
2. Optimisation
3. Monitoring and Alerting

Data Privacy/ Security

As part of the application process, all three suppliers were reviewed on their proposal's ability to demonstrate transparency, data protection and protect user privacy. As such, we would expect the organisational and project approach to data privacy and security to be robust from the outset. This is largely supported by our assessment.

CameraForensics appears to have taken proactive steps to embed data protection recommendations into system design, as well as apply best practice use of encryption in transit as well as at rest. One key area of improvement is in establishing access control and identity management.

C4FF demonstrated the most complete approach to data privacy and cybersecurity, implementing robust information and security controls to ensure the protection of sensitive data within our solution. This includes:

1. **Authentication and authorisation** using role-based access control (RBAC) to ensure that users only have access to the data and functionalities relevant to their roles.
2. **All data transmission and storage are encrypted using industry-standard encryption protocols** such as HTTPS for secure communication and robust encryption algorithms for protecting sensitive data at rest.
3. **Robust data backup and disaster recovery mechanisms** to ensure data availability and resilience in the event of system failures or data breaches. Regular backups are performed, and recovery procedures are in place to minimise downtime and data loss.

They also took a proactive approach to compliance with GDPR and other data protection legislation, implementing the GDPR principles. Based on our findings, C4FF was the only supplier to integrate privacy-enhancing technologies into the design and architecture of our solution, including the encryption of data transmission by default to protect the confidentiality of user data and prevent unauthorised access or disclosure.

Finally, Vistalworks' project was de-risked by not processing personal data, alongside their robust organisational cyber posture from working with MoD and NCSC.

The ICO has also indicated that they are aware of good data protection approaches used as part of the development of their technology and the plans for scaling. While these comments are a positive indicator of robust data protection protocols used, they do not provide final assurances as such assurances are beyond the scope of the ICO's role in STCF 2.

User Feedback

Suppliers considered user experience and gathered user feedback as part of the product development. Notably, CameraForensics secured a partnership to test a prototype in a law enforcement stakeholder's operational environment following engagement with relevant bodies and agencies in the UK, Europe and elsewhere. They are continuing to work with a number of these stakeholders to explore the further testing and implementation of their solution.

Unlike CameraForensics, Vistalworks did not get access to government users despite efforts, which appears to be due to a combination of user access, availability, customer readiness and product development dependencies. Instead, they conducted robust user testing with the private sector, following commercial product management best practices.

Finally, although we have limited details on the user testing approach, we understand that C4FF took a systematic, user-centric approach to the design and development process, focusing in particular on enhancing the user experience. They have also identified opportunities to drive further product improvements through further user feedback. For example, they are currently in discussion with IWF to validate the system with live URLs.

RQ2: Did the projects achieve what was set out to be built?

All solutions achieved their development plans and built a proof of concept that was operational and tested.

Findings across the relevant criteria of **Performance, Robustness,** and **Tech Readiness Level** are below.

Performance

Given their AI-driven approach, C4FF and Vistalworks took similar approaches to technical performance evaluation. On industry-standard performance metrics, C4FF scored slightly higher.

Given performance metrics were not standardised across projects, companies took varied approaches to measuring and tracking performance. Metrics reported here were **self-reported** by companies through the technical assessment survey. Given the information is self-reported, there are limitations in the standardisation of measuring and reporting out performance. As a result, comparison of the reported measurements is not appropriate without further context; the results here serve to share the metrics used to track performance as well as general progress. Technical performance metrics were not applicable to CameraForensics' project as it is not developing an AI algorithm, as self-reported, and therefore not included in the table.

Self-Reported Performance Metrics:

Company	Precision	Accuracy	Other Scores
C4FF	Unknown	97%	● **Kappa:** 94% ● **RMSE:** 19%
Vistalworks	~100% (on real, semi-real and plausible synthetic data)	93%	● **False negatives:** <2% ● **False positives:** 5% ● **Speed:** Processing test datasets reduced from 15 seconds to 0.2 seconds from start to end of two week demoing process (data volumes unknown)

Based on the accuracy scores of over 90%, both solutions are capable of early piloting as part of further model refinement. However, there is a need for further human review of results, rather than a fully automated process, with further scope for optimisation to reach a target benchmark of >99% accuracy. We also recognise that precision and accuracy scores reported here require further independent verification and additional information on factors including the size of the sample dataset tests were performed on.

Robustness

While all three suppliers considered robustness of their solutions as part of design and build, robustness was the area of greatest differentiation in quality and maturity of approach between the suppliers. CameraForensics and Vistalworks focused on handling various types of content and links effectively, while C4FF incorporates robustness measures like subpopulations and adversarial testing to ensure effectiveness against evasive tactics.

CameraForensics implemented a range of enrichments based on initial assumptions and was the only supplier we are aware of who employed user feedback to identify enrichments. Their emphasis on robustness focused on the variety of link journeys and enrichments.

By contrast, Vistalworks built on their proven robustness approach from experience in illicit trade detection. They ensured robustness in handling various text types and characters, targeted open web text and links and focusing on standard formats.

The most mature approach to robustness testing was demonstrated by C4FF. They implemented a range of robustness steps as part of data pre-processing and testing such as data augmentation, cross-validation, adversarial testing, sensitivity analysis, and real-world testing. Their range of best-practice AI/ML robustness metrics, including subpopulations, transformations, distributional shift, and uncertainty reflects best practices identified by the DCMS Online Safety Data Initiative. They also considered various circumvention tactics used by adversaries.

Areas of Similarity	Key Robustness Measures
● **User Feedback Incorporation:** CameraForensics and Vistalworks incorporated user feedback into their development process, indicating a user-centric approach to solution design. ● **Testing and Validation:** All three solutions recognise the importance of rigorous testing and validation before deployment. Vistalworks explicitly mentions planning thorough security testing	● **Approach to Robustness:** While all three solutions prioritise robustness, they differ in their specific approaches. CameraForensics emphasises a diverse range of enrichments, Vistalworks focuses on handling various text types and characters, and C4FF employs a combination of robustness measures including data augmentation and sensitivity analysis. ● **Focus and Target Domain:** Each solution has a different focus and target domain. CameraForensics seems to focus on link navigation and enrichments, Vistalworks targets illicit trade detection in open web text and links, and C4FF is concerned with

before the commercial production of their software, while C4FF outlines specific testing measures like cross-validation and real-world testing.	detecting illegal content like CSAM and considers circumvention tactics employed by adversaries. ● **Testing Stage:** The readiness for commercial production varies among the solutions. While Vistalworks mentions that their software is not yet in a commercial production environment and plans thorough testing, CameraForensics and C4FF do not provide explicit information regarding their production readiness or testing plans.

Tech Readiness Level

According to UK Research and Innovation (UKRI), the TRL levels and definitions are as follows:[19]

- **TRL 1:** Basic principles observed and reported
- **TRL 2:** Technology concept or application formulated
- **TRL 3:** Analytical and experimental critical function or characteristic proof-of-concept
- **TRL 4:** Technology basic validation in a laboratory environment
- **TRL 5:** Technology basic validation in a relevant environment
- **TRL 6:** Technology model or prototype demonstration in a relevant environment
- **TRL 7:** Technology prototype demonstration in an operational environment
- **TRL 8:** Actual technology completed and qualified through test and demonstration
- **TRL 9:** Actual technology qualified through successful mission operations

Our ability to assess each project's TRL is limited by factors set out in the Limitations section. As such, while we have cited a rationale for our TRL assessment of each

[19] UK Research and Innovation, Eligibility of technology readiness levels (TRL)

project, our assessment is based only on evidence available and is not exhaustive. This evidence includes feedback provided by project teams on the TRL they believe their product has reached. In some cases there is a discrepancy between the TRL self-reported and the TRL we have been able to qualify in our assessment. Further independent testing of solutions is required to confirm the TRL of each technology.

Company	TRL	Rationale
C4FF	**TRL 4** Technology basic validation in a laboratory environment	● Proof of concept validated in a controlled laboratory environment with engagement from an external partner. ● Adapted solution based on performance monitoring. ● Metrics established and tested.
Camera Forensics	**TRL 5** Technology basic validation in a relevant environment	● Technology tested with a partner in a relevant environment. ● Used simulated conditions for a real world or operational environment. ● Identified technical and data privacy issues. Co-developed to overcome challenges to improve proof of concept. ● Tested integration, adapting model to partner needs.
Vistalworks	**TRL 5** Technology basic validation in a relevant environment	● Technology deployed is consistent with Vistalwork's illicit trade detection technology and has therefore been validated in a *relevant* environment. ● Specific CSAM detection proof of concept tested in-house in a laboratory environment. ● Initial experimentation demonstrated feasibility and basic validation of effectiveness. ● Focus of work was to conduct thorough research to inform CSAM detection tooling and build datasets that use illicit trade detection capabilities as a basis.

RQ3: Are these proof of concepts new and novel across the safety tech sector?

In conducting market analysis, these solutions appear to be novel solutions in a nascent application area. Based on PUBLIC's deep knowledge of the safety tech sector, including our annual International State of Safety Tech report, we have identified only up to six safety tech solutions on the market that claim to detect harmful links, however these are not specific to CSAM material.[20] These solutions include detecting fraudulent phishing links and links used for mis/disinformation and those developed by large technology firms like Google and Microsoft. Based on our quick-turn analysis, we believe that the solutions developed as part of STCF 2 are novel and address an emerging gap in the market.

Given the sensitivity and complexity of CSAM material and perpetrator behaviour, novel solutions built to address these are required to be developed to be effective to this threat. The challenge fund proves that the model to define a CSAM-specific challenge area, and provide funding with technical development guidance is successful in developing novel solutions.

Technical Challenges and Limitations

While in general the suppliers delivered against the original technical scope of the proposals, we identified issues faced during development and current limitations to the solution.

Company	Technical Challenges	Technical Limitations
C4FF	● **Data access:** Timely access to real data for training and testing was a technical, legal and process challenge for C4FF within the project timescale. While C4FF did manage to get IWF data through IWF membership, this long-standing issue could have been better anticipated ● **Technical feasibility:** C4FF faced technical and feasibility challenges in their local CSAM implementation. This was flagged during evaluation and, based on advice from stakeholders, led to	● **Accuracy:** Beyond the scope of the project, the accuracy could potentially be enhanced by incorporating additional features, such as page metadata. C4FF plans to implement in future iterations of the model.

[20] PUBLIC, International State of Safety Tech 2023

	C4FF descoping the local CSAM workstream and pivoting effort solely towards development of a web-based CSAM detection application for URLs.	
Vistalworks	● **Data access:** Access to law enforcement data for cross-verification ● **Access to representative government users:** Access to government users to provide feedback, due in part to lack of readiness by law enforcement/HMG to enforce on text link sharing yet	● **Data representativeness:** All training data comes from the open web and private sector, mainly search engines, social media and ecommerce). ● **Human review overheads:** All automated results required manual verification by Vistalworks researchers and, to a lesser extent, the IWF.
Camera Forensics	● **Legal Constraints:** Testing and deployment while complying with privacy and legal constraints. Issue of avoiding potentially accessing illegal content, while tackling the issue effectively. ● **Challenging Media Types:** Potential technical challenges in tackling video and URL-embedded text.	● **Deployment on Video:** We understand further system enrichments are planned to tackle video processing ● **Generalisability:** CameraForensics are working to ensure the generality of the system for deployment to different operational environments.

Through technical assessment of the three projects, it can be concluded that a challenge fund model can successfully drive development in a nascent sector for developing proof of concepts when they have the ability to test in relevant or laboratory environments. This is particularly important to commercialisation where there is testing in partnership with end users to foster co-development. Access to data for the safety tech sector continues to be a critical and clear development barrier across product use cases that also needs consideration in the context of the STCF programme. Initial research phases are key to build robust evidence that shapes technical development to align with the real-time threat landscape. Performance of models was not consistently measured, yet across the metrics used, early testing of models proved to be promising. Future challenge funds would benefit in aligning on technical evaluation criteria and metrics, or principles ahead of development.

Conclusion

Key Findings and Lessons Learned

Process Evaluation

The process followed by STCF 2 was in line with the standard process of the UK's national innovation agency, Innovate UK. This evaluation has rightly not assessed the IUK standard process itself, but the applicability of the standard IUK process to STCF 2. In line with that context, this evaluation has found:

- The use of the IUK standard process provided reliability and consistency that enabled the fund to be executed on schedule and with good outcomes - as highlighted by the Impact Evaluation.
- Data access for projects could have been more proactively considered during the market research phase, although this is a consistently difficult issue to contend with in the area of combatting CSAM.
- A more comprehensive understanding of the technical approaches of projects, as opposed to commercialisation plans, during the competition phase may provide a better basis for assessment and identify issues that can be addressed more efficiently.
- Quarterly and in arrears payment of grants risks precluding smaller companies from participating in funds.
- The contributions of all HMG stakeholders was highly impactful and utilised by fund recipients, especially when more structured and targeted towards specific needs that projects had.

Impact Evaluation

Through the impact evaluation of STCF 2, we found:

- Funding R&D in the area of detecting and disrupting link sharing to CSAM has ensured that meaningful lessons have been learned that can form the basis of further technological development.
- The Safety Tech Challenge Fund brand is strong and an important facilitator in the formation of partnerships and connections between safety tech providers

and other stakeholders in the ecosystem. This helps to reinforce the growth of UK safety tech and its world-leading position.

- While further knowledge sharing and dissemination would be welcome, the interactions between all stakeholders was positive and meaningful. This evaluation report may form the basis for that knowledge sharing to the wider trust and safety ecosystem, while engagement between HMG stakeholders and fund recipients with Ofcom - as the regulator of online safety - would be encouraged.
- The ability for solutions to be scaled up, commercialised and become sustainable is yet to be seen and there is not sufficient evidence to suggest whether this is possible yet.

Technical Evaluation

The technical assessment has also raised a number of broader takeaways and lessons learned for technical development:

- Access to real data, regardless of source (ie. law enforcement, platform or hotlines) for model training and testing is a long lead time item with high risk of failure within development timelines. As such, platforms should be encouraged to start this engagement process earlier (e.g., C4FF) or consider a multi-pronged mitigation strategy, including synthetic data. DSIT and HMG stakeholders establishing datasets that can be used by fund recipients may also help centrally mitigate this risk.
- Based on the Vistalworks experience, we recognise the importance of social and behavioural research to characterise and map circumvention techniques and threat actor behaviours, before creating the products to recognise risk signals and detect them.
- DSIT has an opportunity to use the Safety Tech Challenge Fund programme to gain a better understanding of state-of-the-art AI testing and evaluation. For example, C4FF's testing approach offers learning for practical cross-validation, bootstrap sampling and hyper-parameter tuning.

Recommendations

Process

To ensure the process of future Safety Tech Challenge Funds is efficient as possible and contributing to desired impacts three key recommendations could be considered:

- It is clear that the **standard IUK process could be adapted to better meet the needs of a future Safety Tech Challenge Fund**, following consultation with the relevant fund owner during the initial scoping of the fund management. These **findings and lessons learned should therefore be considered at this early scoping phase** in any future challenge funds, especially if managed by IUK.
- Adopting a **more agile fund management approach** that can identify and leverage **opportunities for knowledge sharing and awareness raising** or achieve any **economies of scale regarding commercial or partnership opportunities** through the course of the fund may ensure more efficient achievement of desired outcomes.
- In conducting a **technical evaluation**, the **assessment approach** and **testing criteria** should be established ahead of the competition phase. Throughout the development and delivery of M & E activities, experts across **safety tech product development,** including **AI solution developments and auditing** should be engaged to leverage their expertise. This could provide opportunities for deeper collaboration with the UK's AI Safety Institute and the Responsible Tech Adoption Unit.

Impact

The impact evaluation demonstrated that many of the desired outcomes were achieved by STCF 2. To ensure that outcomes can be similarly achieved in future funds and that opportunities to go further can be realised, it is recommended that:

- **Contributions of HMG stakeholders to fund recipients is structured** to ensure it is **relevant to the projects' status and challenges** and extracting the insight and expertise of HMG stakeholders as effectively as possible.
- Future challenge funds **develop a plan** to make sure that either projects **integrate the impact of E2EE** into their development or the influence of E2EE is

duly considered through other mechanisms. It is **vital that maintaining user's privacy is central** to this recommendation.

Technical

Based on survey findings, interviews and technical analysis, a number of recommendations have emerged:

- **Establish formal, comparable technical performance metrics/KPIs** for suppliers as part of the fund delivery, especially for non-AI solutions. CameraForensics did not share formal performance metrics with us, which makes assessment and comparison challenging. To ensure a diversity of technical approaches to a challenge, DSIT could provide a range of options or require the supplier to define 1+ technical performance metric.
- When working on an illegal harm (ie., CSAM), allow **sufficient time for social and behavioural discovery** to ensure the product is designed in a way that maximises efficacy, while **limiting the need to access raw data**. This also has secondary benefits where DSIT can require the supplier to present back findings for policy learnings.
- Review **long lead time technical tasks and data access dependencies** upfront during supplier evaluation. While C4FF did finally receive real data for development and testing, it required a lengthy application for IWF membership and nearly became a high-impact issue.
- Consider the realism of the **deployment context and technical dependencies**. In the case of C4FF's local CSAM implementation concept, the technical restrictions of encrypted web traffic and implementation challenges of deploying a model on routers were highlighted by GCHQ in their initial supplier evaluation. In a future fund, DSIT and its delivery partner should consider how to ensure technical feedback from GCHQ feeds into evaluation and/or is factored into timely project scope revisions.

Appendices

Glossary

Acronym	Meaning
CSAM	Child Sexual Abuse Material
C4FF	Centre for Factories of the Future
DCMS	Department of Digital, Culture, Media and Sport *[former department]*
DSIT	Department for Science Innovation and Technology
E2EE	End-to-end encrypted / End-to-end encryption
GCHQ	Government Communications Headquarters
HMG	His Majesty's Government
ICO	Information Commissioner's Office
IUK	Innovate UK
IWF	Internet Watch Foundation
KTN	Knowledge Transfer Network
RBAC	Role-based access control
ROI	Return on investment
RQ	Research question(s)
STCF 1	Safety Tech Challenge Fund Round 1
STCF 2	Safety Tech Challenge Fund Round 2
TRL	Technology Readiness Level
ToC	Theory of Change
UKRI	UK Research and Innovation

Acknowledgements

This evaluation report has been made possible by the generous contributions of stakeholders who gave their time and considered thought to aid us in collecting the necessary evidence. We also recognise that combatting CSAM is a difficult area to work in and has the potential to be highly traumatic for individuals dedicating their efforts to it. The ambition, integrity and wholeheartedness with which all stakeholders have worked on this project should not be underestimated.

Interview process and script

PUBLIC conducted semi-structured interviews with representatives of the following stakeholders:

- DSIT [as fund owner and HMG stakeholder]
- Home Office [as an HMG stakeholder]
- Information Commissioner's Office [as an HMG stakeholder]
- GCHQ [as an HMG stakeholder]
- Innovate UK [as manager of the fund]
- CameraForensics [as a fund recipient]
- Centre for Factories of the Future [as a fund recipient]
- Vistalworks [as a fund recipient]

We developed a core interview script that could be used to gather insights consistent to each stakeholder group and made bespoke changes to the scripts for questions relevant to a single user group. Additional questions were posed as follow-ups from insights shared, in keeping with a semi-structured approach.

Below is the outline of themes for our core interview script.

Section 1: Introduction and Roles

Questions focused on understanding the stakeholders role, responsibilities, and level of engagement throughout the fund.

Section 2: Understanding Processes

Questions gathered evidence on the different types of activities, outputs, outcomes, and challenges throughout the phases of the challenge fund and the stakeholders specific engagement.

Section 3: Understanding Outcomes and Impact

Questions gathered evidence on viewpoints in how the challenge fund achieved the programme objectives and impacted stakeholder groups.

Section 4: Assessing the Counterfactual

Hypothetical scenarios in delivering a challenge programme was discussed to ruminate how results would have varied under different circumstances.

Interviews were recorded and transcribed. Relevant findings were grouped according to topic - in line with the structure of the interview - and by the user group of the interview participant. These insights formed the basis for the analysis included within this report and are cited throughout.

Technical survey

The technical survey, sent to project teams, asked the following questions:

1. What TRL is your solution?[21]
2. What metrics and criteria did you use to measure technical performance? Are you able to share any performance data over time?
3. What are the current limitations of your proof of concept (efficacy, technical, administrative, other)?
4. Please can you share current performance scores against these metrics (if available)
5. How did you collect and incorporate user feedback?
6. What steps did you take to test whether the solution performs equally effectively when faced with perturbations and variations in content?
7. What approach(es) did you take to test robustness?
8. What circumvention tactics did you take into consideration (if any) and how did you test for this?
9. What data did you use for solution development and testing?
10. Please describe how you got access to this data.
11. What testing methods and technical environments did you use (if any)?
12. What technical scalability testing have you performed (if any)?
13. Please can you share current scalability testing scores against these metrics

[21] The final TRL assessment of projects also considered other factors besides what was self-reported.

14. How do you expect the solution to perform when faced with a steep increase in workload in a production environment?
15. What monitoring processes and practices do you have in place or plan to implement to track performance of the model over time?
16. What information and security controls do you have in place for the proof of concept?
17. What steps did you take to comply with GDPR and data protection legislation?
18. Please can you share any performance scores of your solution against the metrics and criteria mentioned in Question 2 (if available)?

Data Collection and Data Privacy Notice

PUBLIC recorded interviews and took written notes to collect the required evidence for the evaluation. The notes and recordings were used purely by the evaluation team and were used only to inform the final evaluation report. Ahead of recording we received consent from interviewed stakeholders.

Relevant minutes, notes, and recording(s) were available to stakeholders upon request.

The documentation of the interviews, evidence collected, and any other personal information was collected in full compliance with the UK's General Data Protection Regulation.

Among other things, this required PUBLIC to:

1. Process personal data in good faith,
2. Be transparent to stakeholders regarding how we process personal data,
3. Process personal data only for the specific purposes communicated,
4. Minimise the personal data we collect and store,
5. Treat personal data confidentially.

At any time, stakeholders have the right to:

1. Request information about how PUBLIC processes personal data and why it has been do so,

2. Receive information within one month's time about what personal data PUBLIC holds about stakeholders,
3. Inform PUBLIC about mistakes in the personal data they hold about stakeholders and to see these mistakes corrected,
4. Request deletion of personal data and the termination of PUBLIC's processing of personal data.

Bibliography

- DCMS, Online Safety Data Initiative
- His Majesty's Treasury, 2020, Magenta Book
- ICO, 2024, Content moderation and data protection
- INHOPE, 2022, Annual Report 2022
- Internet Watch Foundation, 2022, The Annual Report 2022
- Ipsos Mori and Perspective Economics, 2022, Trust and Safety and the Digital Economy
- Perspective Economics, commissioned by DSIT, 2023 The UK Safety Tech Sector: 2023 Analysis
- PUBLIC, 2022, The International State of Safety Tech: 2022
- PUBLIC, 2023, International State of Safety Tech 2023
- UK Research and Innovation, 2022, Eligibility of technology readiness levels (TRL)
- WeProtect Global Alliance, 2023, Global Threat Assessment 2023
- WeProtect Global Alliance, 2023, Briefing from a joint WeProtect Global Alliance and GCHQ expert roundtable

About PUBLIC

PUBLIC conducts independent evaluations of major public sector digital, technology and innovation programmes, to help authorities to measure the impact of their digital services. Combining expertise in technology, data science, statistics and economics, we have partnered with teams like the Evaluation Task Force, Government Digital Service (GDS), Innovate UK, and Department for Levelling Up, Housing and Communities (DLUHC) to evaluate some of the UK public sector's most important digital projects and programmes.