Government funds new tech in the fight against online child abuse

- Five winning projects in £555k competition exploring new ways to stop the spread of child abuse material in encrypted online communications

- Safety Tech Challenge Fund will boost innovations in AI and other tech that can scan, detect and flag illegal child abuse imagery without breaking end-to-end encryption

- Comes as the UK hosts G7 Summit calling for global collaboration on tech that makes the internet safer

The five projects – including tech companies in Edinburgh, Poole, St Albans and London – are the winners of the Safety Tech Challenge Fund, which aims to encourage the tech industry to find practical solutions to combat child sexual exploitation and abuse online, without impacting people's rights to privacy and data protection in their communications.

The winners will each receive an initial £85,000 from the Fund, which is administered by the Department for Digital, Culture, Media and Sport (DCMS) and the Home Office, to help them bring their technical proposals for new digital tools and applications to combat online child abuse to the market.

Projects include new AI plug-ins that can be run in the background of existing encrypted messaging services to identify images of child abuse and flag them to moderators. Others will use age estimation and facial recognition technology to scan for and detect child abuse images before they are uploaded. Another project will look specifically at how to prevent livestreaming of violence and child pornography.

The winners, who met with the Digital Minister Chris Philp yesterday to showcase their projects, will spend the next five months developing and evaluating their solutions. Additional funding of £130,000 will be made available to the strongest projects, bringing the total funding to £555,000.

The announcement comes as G7 and invited guest countries met virtually for the G7 Safety Tech Summit yesterday to discuss ways in which global partners can collaborate to promote continued innovation in safety tech and help to deliver safer online environments for people all over the world.

Digital Minister Chris Philp said:

> It's entirely possible for social media platforms to use end-to-end encryption without hampering efforts to stamp out child abuse. But they've failed to take action to address this problem so we are

> stepping in to help develop the solutions needed. It is not acceptable to deploy E2EE without ensuring that enforcement and child protection measures are still in place.
>
> We’re pro-tech and pro-privacy but we won’t compromise on children’s safety. Through our pioneering Online Safety Bill, and in partnership with cutting-edge safety tech firms, we will make the online world a safer place for children.

End-to-end encryption (E2EE) is a technology which encrypts communication data – including messages, images and recordings – between sender and recipient to prevent third parties accessing them.

It is widely used across a range of services including banking but is particularly prevalent on messaging services such as Whatsapp and iMessage. It significantly impacts the ability of tech companies to protect children from being groomed for sexual abuse and help law enforcement track down and arrest criminals who share child sexual abuse material.

Based across the UK and Europe, and in partnership with leading UK universities, the winners of the Safety Tech Challenge Fund are:

- Edinburgh-based Cyan Forensics and Crisp Thinking, in partnership with the University of Edinburgh and Internet Watch Foundation, will develop a plug-in to be integrated within encrypted social platforms. It will detect child sexual abuse material (CSAM) – by matching content against known illegal material.
- SafeToNet and Anglia Ruskin University will develop a suite of live video-moderation AI technologies that can run on any smart device to prevent the filming of nudity, violence, pornography and CSAM in real-time, as it is being produced.
- GalaxKey, based in St Albans, will work with Poole-based Image Analyser and Yoti, an age-assurance company, to develop software focusing on user privacy, detection and prevention of CSAM and predatory behavior, and age verification to detect child sexual abuse before it reaches an E2EE environment, preventing it from being uploaded and shared.
- DragonflAI, based in Edinburgh, will also work with Yoti to combine their on-device nudity AI detection technology with age assurance technologies to spot new indecent images within E2EE environments.
- T3K-Forensics are based in Austria and will work to implement their AI-based child sexual abuse detection technology on smartphones to detect newly created material, providing a toolkit that social platforms can integrate with their E2EE services.

The government’s forthcoming Online Safety Bill will transform how illegal and harmful online content is dealt with. It will place a new duty of care on social media and other online companies towards their UK users.

This will mean there will be less illegal content such as child sexual abuse and exploitation online and when it does appear it will be removed quicker. The duty of care will still apply to companies that choose to use end-to-end encryption.

ENDS

Further project details

Cyan Forensics

Cyan’s technology matches content in messaging against a database of known CSAM using a new technique that splits the matching process between the user device and a cloud service. The communications between device and cloud use an innovative Privacy Assured Matching protocol that ensures the content of messages remain private and secure in normal use (with no “back door” access), while allowing appropriate action to be triggered when known Child Sexual Abuse Material is being exchanged.

Ian Stevenson, Cyan Forensics CEO said:

> We’re delighted to be invited to take part in this challenge along with our partners at Crisp, the Internet Watch Foundation and the University of Edinburgh. Cyan is already delivering for law enforcement and, building on Contraband Filter technology, we will work together to demonstrate how it can be applied to deliver a new approach to detecting Child Sexual Abuse Material while protecting user privacy – extending the range of solutions available with new and much-needed capabilities.

DragonflAI

The solution combines AI nudity blocking and age assurance technologies to keep children safe in E2EE social networks. An age estimation model (using facial recognition) will allow nudity and age to be detected together within E2EE. The result is a solution that can detect unseen and new indecent images with a high degree of accuracy, and thus offenders, without this content leaving users’ devices for detection. Upon detection of potentially indecent content, the user can be flagged.

Hanah Mercer CEO and Founder, DragonflAI said:

> The sharing of CSAM and indecent images of children is abhorrent, and the partnership between Yoti and DragonflAI will allow us to combine our technologies to detect unseen, illegal content within end-to-end encrypted (E2EE) messages. We have proven the technology is viable, and given the solution will work entirely on-device, the content can be detected without images leaving the user‘s device. We hope that this can play a part in the future of social media moderation.

Galaxkey

This solution proposes a system that provides instant E2EE messaging and explicit content detection and will allow users to safely register with the platform using age estimation by Yoti. When the user sends content, it will be scanned prior to encryption. If the content is detected as explicit or not allowed, it will be reported securely to the moderators to take action. This will prevent the livestreaming of and sharing of CSAM images and videos in E2EE environments. When the content is detected, the moderators will receive the information securely and it will be audited and tracked to prevent any misuse. Currently there are no products in the market that provide this kind of model of pre-content filtering with end-to-end encryption.

Sir George Zambellas, Chairman Galaxkey said:

> On behalf of the whole Galaxkey team and our technology partners Yoti and Image Analyser, I’m delighted we have won this prestigious win. It reflects our commitment to online child safety, by combining high quality encryption and specialist content detection, in support of the UK’s Online Safety Bill.

SafeToNet

SafeToNet’s project will develop a suite of live video-moderation safeguarding technologies that can run on any smart device such as a mobile phone, tablet, gaming platform and wearable. This technology can prevent the filming of nudity, violence, pornography and CSAM in real-time, as it is being produced. Using efficient algorithms that utilise the processing power of the device, the technology runs locally on the device and without the need for cloud interaction. This helps maximise the privacy rights of the user whilst automatically keeping them safe.

SafeToNet founder and Group Chairman Richard Pursey said:

> SafeToNet’s entire mission is to safeguard children online – however, we cannot do it on our own. Tackling online harms requires a collaborative approach with expert minds from the safety tech industry, academia, Governments and more coming together to share their learnings and skills to improve online safety. Therefore, we are delighted to have been awarded a DCMS safety tech challenge grant. It means we can work closely with our friends in the Policing Institute for the Eastern Region of Anglia Ruskin University to develop technology that minimises the negative impacts of end-to-end encryption and positively improves the safety of children online both in the UK and across the world. SafeToWatch can have a significant impact in preventing the creation, distribution, and consumption of Child Sexual Abuse Material.

T3K Forensics

This project will use AI Classifier technology to detect CSAM. This means that instead of the more commonly used method of comparing a picture’s unique hash or PhotoDNA value (the “fingerprint” of the image) with a database of previously known data, it will look at the actual content of a picture or video. It will use biological features to find visible children and place emphasis on the situation those children are in, and in this way be able to detect newly created material that is not yet in a database and may otherwise be missed. It will access media files on devices to spot content in E2EE messaging apps that is suggestive of abuse or the livestreaming of illegal content.

Because the CSAM Classifier will be integrated within the communication apps, it will not depend on sending data to remote locations for checking, or intercept and lever out end-to-end encrypted communication. Instead, it will look at the data that a user has on their own device – thereby preserving their privacy.

Martina Tschapka, Director Operations & Content Manager Online Child Safety at T3K-Forensics said:

> Smart and easy ways to share harmful data at any time require equally smart and easy ways to find and stop this data spread. Let’s not forget that numerous stories of suffering and vulnerability lie behind this material. It is T3K’s mission to play our part in stopping the vicious circle of suffering and revictimisation, and we are thrilled that our solution was picked to be a part of this crucial project.

Notes to Editors

- The Safety Tech Challenge Fund, supported by DCMS, HO and GCHQ, was launched in September.
- For more details on the winning projects, see below or visit www.safetytechnetwork.org.uk
- Companies participating in the Challenge Fund will receive tailored advice on privacy protection from the Information Commissioner’s Office (ICO) to ensure privacy is respected and built in from an early stage. All solutions developed as part of the fund will be assessed by a team of independent academic experts working for the National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online (REPHRAIN).
- At the end of the five month Delivery Phase, the projects will be evaluated on success criteria by the external evaluator. Project solutions will also be assessed on this commercial viability to determine deployability into the market, and long term impact.
- The Safety Tech Challenge Fund is one example of how the UK is helping to deliver on the G7 Internet Safety Principles under its commitment to share information, research and best practice in the development and adoption of safety technology.