2023 OPEN CALL

REQUEST FOR PROPOSALS

FOR

TECHNOLOGY SOLUTIONS & RESEARCH FOR TACKLING CHALLENGES IN ONLINE CHILD SEXUAL EXPLOITATION AND ABUSE: AGE ASSURANCE & LIVE STREAMING OF ABUSE

Subject: Tech Solutions: Age Assurance & Live streaming of Abuse

Issue date: 15 FEB 2023

Submit Full Proposal: Online Application Form **by 29 MAR 2023 11:59 PM EST**.

I. Background

The End Violence Partnership is a public-private partnership launched by the UN Secretary General in 2016 to accelerate progress towards Sustainable Development Goal 16.2: ending all forms of violence against children by 2030. End Violence comprises 750+ partners, including governments, civil society organisations, UN agencies, the private sector and research institutions, and acts as a global platform for advocacy, evidence-based action, and investments to end all forms of violence against children.

Through its Safe Online investment initiative, End Violence provides funding, policy and advocacy guidance, and coalition-building to significantly advance national, regional and global efforts to prevent and respond to online child sexual exploitation and abuse (CSEA). In 2022, End Violence's Safe Online investment portfolio reached US$71 million in grants to 89 projects achieving tangible results in over 80 countries. Safe Online's grant portfolio as of November 2022 can be found here.

This Open Call follows the 2020 and 2022 Funding Rounds, also focused on Technology Solutions to keep children safe online. Funding is not limited in this Open Call to solely technology solutions, but also includes actionable or exploratory research on social and policy questions for enhancing the capacity of all stakeholders to address challenges around the use of tech solutions for the Open Call's two focus areas – age assurance and live streaming of abuse - in the online CSEA landscape.

II. What We're Looking For

The purpose of this Open Call is to seed and grow innovative solutions and research that leverage existing and new technologies to address challenges around age assurance and live streaming of abuse in the online CSEA landscape. This means we are looking to applicants to identify solutions and research that will be most impactful to tackle these two topics within prevention, detection and response for online CSEA.

End Violence's Safe Online Initiative has selected these topics based on rising prioritisation in global discussions of online CSEA across sectors; upcoming legislation in the EU, UK, US and Australia specifically and emerging regulation/policy focus globally that raises the urgency of proactively engaging with these challenges; as well as following extensive consultation with diverse experts across the online CSEA and related fields.

What we will fund

End Violence invites Technical Proposals for this year's Technology Solutions Open Call that develop solutions that include technological and research approaches in one of two focus areas in preventing and combatting online child sexual exploitation and abuse (online CSEA): **1) addressing age assurance or 2) tackling live streaming of abuse**.

The objective of this Open Call was to encourage Proposals to focus on either age assurance **or** live streaming of abuse, not both. However, we recognise that there is overlap in these topics. If a Proposal does include both topics in some respect, applicants are asked to indicate only one of these topics as the primary focus of the project proposal.

End Violence will award up to USD$200,000 to non-profit or for-profit organisations to work on innovative solutions that **leverage existing and develop new technologies as well as address social and policy questions through research** to address challenges around age assurance or live streaming of abuse in the online CSEA space. The total funding envelope to be awarded through this Open Call is **~US$ 2 million.**[1]

We are interested in two modalities:

1) **Solutions that utilise and implicate technology in innovative, groundbreaking ways** that are scalable and globally applicable for addressing age assurance or live streaming of abuse. Solutions should be applied specifically to exact or related use cases for this Open Call (e.g. if online CSEA data is not available for application, use cases such as adult material could be used instead). Tech solutions we will fund include:
   - **design of a proof of concept** that demonstrates an innovative use of technology to address the specific challenges posed around age assurance or live streaming of abuse, and clear plans for further research and testing;
   - **development of new prototypes or products** – including hardware/software/content components – or of new features within existing products or tools;
   - **adaptation or combination of existing tools** to address current gaps, increase efficacy and/or effectiveness, or apply to new use cases.

2) **Actionable or exploratory research on social and policy questions** to better understand the enabling environment for more effective implementation of solutions, as well as address unintended

[1] The total amount to be awarded will depend on the quality and volume of received applications i.e. End Violence may decide to slightly increase or lower the amount if specific opportunities arise or as a result of the initial assessment of applications.

consequences and structural barriers to impact of tech solutions that address age assurance or live streaming of abuse in the online CSEA landscape, including a focus on any/all of the following:

- key populations (e.g. potential and actual victims, survivors, practitioners, would be and confirmed perpetrators, etc.),
- environment (e.g. tech platforms, geographies),
- and broader ecosystem (e.g. policy, legal, ethical, cultural, economic, etc.).

We recognise that there can be overlap in these modalities. If a Proposal does include some research aspects for technology solution development or is focused on research using or with applications in technology tool development, applicants are asked to indicate only one modality and ensure the scope of the outputs, activities and budget are clearly aligned.

Principles

We will fund tech solutions and research that meets the following standards:

Child Rights-Centred

Tech solutions and research must recognise and uphold the spectrum of all human rights - civil, political, economic, social and cultural rights of every child, regardless of their race, religion or abilities - in their approaches to achieve child safety. This includes respecting user privacy – including children and young people – and ensuring data protection laws/guidance are incorporated into development of solutions and research.

Inclusive

Tech solutions and research outputs consider all users in their design, including the most underrepresented or marginalised groups. Clear consideration of downstream implications of solutions and research should be reflected, with nuanced understanding of contextual factors and differential user-experiences.

Collaborative

Tech solutions and research demonstrate multi-stakeholder collaboration within and across sectors - including digital technology industry, academia, CSOs, companies or organisations in parallel fields (e.g. cybersecurity, gender-based violence), national partners, etc. - as well as holistic approaches including interlinkages between online violence and in-person abuse against children.

Outcomes-Driven

Tech solutions have demonstrable feasibility and efficacy in various environments and use-cases. Research includes outcomes that promote understanding of structural barriers and unintended consequences in order to more feasibly or effectively operationalise tech solutions.

Transparent

Accountability mechanisms are in place for ensuring governing principles are adhered to and communicated clearly. Projects create open-source outputs and engage in open working techniques to the extent possible.

Who is eligible to apply

We are actively seeking submissions of proposals from **non-profit organisations**, such as research institutes and academic institutions, civil society organisations (CSOs), non-governmental organisations (NGOs), and international organisations as well as **private sector companies**.

Consortia are also highly encouraged to apply; however, the organisation submitting the application will be considered the main grantee, bearing all the contractual responsibilities *vis-à-vis* End Violence. Organisations are asked to list partners and advisors. Collaboration with other stakeholders such as law enforcement, government, etc. is encouraged, however the primary applicant must be a non-profit organisation. We encourage maximising synergies across jurisdictions/ sectors/ communities, as well as awareness of and exchange with existing tech solutions and research.

We strongly encourage applications from non-ODA countries, as well as applications listing partners in these countries. Demonstration of matching funding, and a strong evaluation and impact assessment component are encouraged as well.

Funded projects will be connected to similar projects from the Safe Online global community, which should enable projects to develop faster and better. More information about the Safe Online portfolio of grantees is available on the End Violence Partnership website.

End Violence's Safe Online initiative encourages and will give preference to projects that are open-source. It also aspires to have any research or technical outputs that it invests in made available to the widest range of actors possible. Proposals should consider how to implement "open working techniques"[2], which could include all or some of the following:

- use of open standards, common components or patterns;
- ways of working in the open, for example sharing learnings with stakeholder groups or publishing elements of assets produced;
- use of open and reusable source code.

For purposes of clarification, there is no requirement for release of open-source products. Safe Online will not own any part of the IP / solution[3].

What we will not fund

End Violence does not provide core organisational funding for non-profit organisations (including such as research institutes and academic institutions, civil society organisations (CSOs), non-governmental organisations (NGOs), and international organisations) nor contribute to new business lines for large private sector companies.

Multiple applications from the same institution are acceptable; however, consideration will be given to ensure diversity of grant recipients and therefore applications from the same institution will be carefully evaluated with this in mind.

---

[2] Sourced from UK Safety Tech Challenge Fund

[3] All intellectual property and other proprietary rights including, but not limited to, patents, copyrights, and trademarks, with regard to products, processes, inventions, ideas, know-how, or documents and other materials which the Grantee develops using the Grant will be managed in a way that maximises public accessibility and allows the broadest possible use.

I. General Conditions

Eligibility criteria

End Violence's Safe Online Initiative is looking to provide project grants to nonprofits and investment-style, equity-free seed funding to private companies. Only entities that fulfill these mandatory requirements will be considered eligible:

1. **legally registered** entity, either as nonprofit or private company;
2. the tech solution or research addresses the aims of the 2023 Open Call;
3. the organisation has a safeguarding policy or similar in place or is willing to develop a policy[4];
4. solutions and research must be compliant with the relevant/applicable legislative, regulatory and enforcement frameworks, including data protection policy in alignment with relevant laws/guidance;
5. the proposed solution or research responds to a **clear need/gap**, does not duplicate existing tools or research, and builds upon and/or interacts with existing solutions and research.

Awards

Modality	Total funds available	Who can apply	Max duration	Allocation amount
Technology solution: Design of a proof of concept, development of new prototypes or products, or adaptation / combination of existing technology tools to address age assurance or live streaming of abuse	Up to USD $2 million	Legally registered nonprofits: NGOs, CSOs, academic institutions, research institutes and international organisations (to receive a project grant) Or Legally registered private companies (to receive equity-free seed funding)	12 months	Up to $200,000
Research: Exploratory or actionable social / policy research to facilitate or strengthen the use, adaptation or scale-up of tech solutions in addressing age assurance or live streaming of abuse			18 months	

Please note that if you are successful in your application and therefore selected for a Grant award, your organisation will be asked to submit two years of the latest financial audit reports completed by an independent auditor and written/translated in English. If your organisation does not have this readily

[4] End Violence is committed to supporting organisations to improve their safeguarding capacity and practice. As part of this, we will ask all grantees to comply with all the applicable End Violence safeguarding requirements. Collaborators receiving less than 50% of the funds will answer questions around safeguarding. Collaborators receiving more than 50% of the funds, will be required to undergo a microassesment and due diligence similar to the main applicant. The primary applicant is responsible for all issues that may arise with partners or contractors in regards to safeguarding. End Violence's Safeguarding Framework is available for all applicants to consult here on our website.

available, a description of why audits are not available and further financial documentation will be requested for the required due diligence by End Violence.

As End Violence is hosted administratively by UNICEF, organisations without a risk rating within UNICEF's financial management system may be required to undergo a financial micro-assessment during the grant period. If the funding amount for the project is less than USD $100,000, there might be an exemption from a financial micro-assessment.

II. Submission of Proposals

All submissions must be made in English. Interested entities that meet the eligibility criteria are required to complete and submit the full Application Form and provide the information and supporting documents indicated in the Form. Completed Applications Forms must be submitted through this online form.

Proposals will be reviewed on an ongoing basis. However, the last day for submissions of Proposals is **29 March 2023 11:59 PM EST**.

A response to this Request for Proposals does not automatically guarantee that submitting entities will be selected for a grant award.

End Violence's Safe Online Initiative reserves the right to change or cancel the requirement at any time during the Request for Proposals and/or solicitation process. End Violence's Safe Online Initiative also reserves the right to require compliance with additional conditions as and when issuing the final contracting/request for proposals document.

If you have any questions about the Request for Proposals, please submit to the FAQ online form. Answers to FAQ submissions can be found on the FAQ document which will be updated weekly. Please review in detail the Technical Guidelines at the end of this document for more details on the focus and scope of this call around challenges in the online CSEA ecosystem in addressing age assurance or live streaming of abuse. You can find further information on online CSEA and End Violence's response on our website.

III. Timeline & Dates

Date	Milestone
15 February 2023	Open Call Launch
29 March 2023	Proposal Submission Deadline
June 2023	Awards Selection
September 2023	Awards Announcement

IV. Terms & Conditions

1. By submitting this Proposal, you are authorising End Violence and external experts to evaluate the Proposal for potential award, and you agree to the terms herein.
2. You agree and acknowledge that personal data submitted as part of the Proposal, including name, mailing address, phone number, and email address of you and other named team members in the Proposal may be collected, processed, stored and otherwise used by End Violence for the purposes of administering the website, reporting to donors and evaluating the contents of the Proposal.

3. You acknowledge that neither party is obligated to enter into any official agreement as a result of the Proposal submission, End Violence is under no obligation to review or consider the Proposal, and neither party acquires any intellectual property rights as a result of submitting the proposal. End Violence reserves the right to withdraw at any time and the applicant agrees to not take any action to bring End Violence into disrepute.
4. Applicants represent and warrant that they have authority to submit a Proposal in connection with this Open Call and grant the rights set forth herein on behalf of their organisation. Any problems that arise related to IP or data privacy are solely the responsibility of the applicant.
5. A sample grant confirmation letter with its legal stipulations and conditions is available here for interested applicants.

V. Review & Award Process

End Violence awards grants through an open, fair and competitive process. **All Proposals will be assessed on their overall quality with attention paid where applicants have clearly explained the contextual challenges, the specific and measurable results that they expect to deliver, the strategies to achieve them with a focus on tailored approaches and interventions.** In addition, applications are expected to acknowledge any risks to delivery and demonstrate plans to mitigate as such.

Under this Open Call, eligible proposals may result in signing of a Grant Confirmation Letter for up to US$ 200,000 and a period of up to 12 or 18 months depending on the project focus – technology solutions or research respectively. Considerations of Proposals that require more funding than the indicated amount will be considered by End Violence at its sole discretion and only if this is in the best interests of achieving the goals of the Open Call. In addition to the relevant costs for the implementation of their project, applicants are strongly encouraged to make provisions for evaluation of their projects (10-15% of the total direct costs) and contingencies (i.e. fluctuations of exchange rates and unforeseeable circumstances, up to 5% of the total direct costs).

Proposals will be closely evaluated for alignment of the scope and activities outlined with the proposed budget. Payment will be made to the applicant's institution, and in the case of a consortium, to the main grantee organisation. Grantees' instalments are determined based on their proposed budgets, with 1-2 instalments depending on project duration and budget. Indirect costs are limited to 7% for grants.

End Violence's Safe Online initiative will actively monitor the progress of all supported projects during the period of the grant, and periodic evaluation of progress. Specifically, all grantees will be required to:

- Report on project progress during annual reporting periods using the Safe Online's reporting templates, which will be provided to grantees;
- Establish and report on key milestones according to qualitative and quantitative indicators selected by the grantee based on their project proposal in the submitted Results Framework Monitoring and Project Implementation Plan;
- Report on key potential barriers or obstacles included in the Proposal in the related question on the Application Form. Identify challenges encountered and steps taken to address them throughout the project; and,
- Attend ad hoc webinars, bilateral (online) meetings or other discussions relevant to the project, including field visits by Safe Online team members, as applicable.

VI. Scoring Criteria

Criteria	Specific Criteria	Possible Score & Key Data Points
MANDATORY CRITERIA		
1. Legal registration	Registered as a legal entity (nonprofit organisation or a private company)	Yes/No
2. Alignment with focus areas	The proposed technology solution or research addresses one of the two focus areas of the End Violence's Safe Online Initiative 2023 Open Call	Yes/No
3. Safeguarding	The entity has a Safeguarding Policy or similar and procedures in place or is willing to develop a policy	Yes/No
4. Compliance	The proposed project is compliant with the relevant/applicable legislative, regulatory and enforcement frameworks, including data protection policy in alignment with relevant laws/guidance	Yes/No
4. Builds on existing work	The proposed solution responds to a clear need/gap, does not duplicate existing tools, and builds upon and/or interacts with existing solutions	Yes/No
SCORING CRITERIA		
1. Problem-solution fit & principles: Relevance of tech solution or research for tackling age assurance or live streaming of abuse and accessibility	i. Clearly defined technological solution or research questions to address a specific challenge/gap related to age assurance or live streaming of abuse in the online CSEA landscape	(20 points) Is the proposed tech solution or research specific and understandable for different audiences? Is there a clear challenge / gap identified and described to reflect a high technical understanding of the need?

	ii. Relevance of the solution for tackling age assurance or live streaming of abuse iii. Alignment between need/gap identified and tech solution or research proposed	Is there close alignment of the proposed solution or research with the problem?
2. Innovation: Novelty and impact of tech solution or research	i. Innovative design, methodology or application of technology ii. Generating open-source[5] outputs iv. Clarifying and ensuring actionability of the tech solutions or research for key stakeholders	(20 points) Is the project working on the development of a new technology, a new application / use case of existing technology, a new combination of technologies, or new research questions or approaches? Is the project developing / expanding existing open-source technology or existing open access research? Will / can the research be widely shared in the field? If outputs cannot be widely shared due to sensitivity of research, is the impact clear for specific stakeholders? Is the research project expected to result in actionable insights and applications for the technology industry and related practitioners?

[5] End Violence will consider occasional exceptions from the open-source rule as justified by the nature and/or sensitivity of the proposed solution.

3. Robustness & feasibility: Performance of the tech solution	i. Robust design, methodology and applications described ii. Well defined testing environments or research questions, data, and other resources required by solution iii. Overall probability of successful delivery of the tech solution or research products and likelihood that the predicted impact and results will be realised ii. Organisation's relevant experience and proof of capacity to implement the project successfully, including solid enumeration of risks and assumptions	(20 points) Is there a solid methodology for the implementation of the project and clearly defined outputs? How thorough is the definition of the methodology including existing technologies, data sources and other resources required for the tech solution or research? Does the project implementation seem feasible? Does it seem like outputs are viable for applications to the problems of age assurance or live streaming of abuse? Does the team have the relevant prior technical or research experience to successfully execute the project to completion? Have they carefully considered all risks and clearly defined key assumptions?
4. Fairness & Inclusion: Suitability of the team to implement the project	i. Alignment of team members' proficiency and experience with skills and time commitment needed to implement project ii. Team is diverse, including across gender iii. Team is primarily composed of individuals with direct local knowledge and connections to the country where the solution is being built and piloted	(20 points) - Does the team have the right skills and experience to implement the project? (inc. technical relevant for the tech product or research, business strategy, UX, software / hardware, programmatic or multidisciplinary expertise) - Is there gender diversity in leadership team and in project team?

	iv. Existence of key advisers filling team's expertise gaps v. Existence of relevant partners	- Are local people leading the project? Are they from the country where solution or research is being developed? If not, does the team have strong local networks showing in-depth understanding of the context and required for implementation of project? - If skills are missing in the core team, do advisors help fill the gaps? Is this enough? - Does the entity have partnerships in place that they need for tool or research development?
5. Capacity: Alignment between budget ask and project goals	i. Matching of overall budget ask for investment with main objectives of the project ii. Balance of funding sources: entity's own capital contribution to the project (human, capital, assets) and other investments	(20 points) Is the budget ask consistent with the cost of employing/developing this technology or research? If the ask is not enough, does the team have enough resources or partners to cover gaps?
TOTAL SCORE		100 Points

VII. Technical Guidelines

Currently, the landscape of technology approaches and evidence-based understanding of how to tackle age assurance and live streaming of abuse in the online CSEA field is fragmented and often reactive to current trends in online harms. The purpose of this Open Call is to push for a more cohesive, proactive approach across the ecosystem to current and upcoming legislation/regulation, emerging technology and constantly evolving threats.

CHALLENGE

Age assurance techniques and live streaming of child abuse are at the centre of many recent online child safety discussions, specifically in legislation and engagement with tech industry. These topics have surfaced

as key actionable areas of focus at the intersection of child rights and industry sectors. Technology solutions that reimagine the current privacy/safety dichotomy are needed. Addressing these issues requires a holistic look at the ecosystem as well as wider research and policy considerations.

Children are targets of harm and victims of abuse online at alarming rates, and evidence and trends increasingly reflect children and young people exposed to violent material and exhibiting risky and potentially harmful sexual behaviours online. Data from the multi-country Disrupting Harm project says that up to 20% of 12–17-year-olds across 13 countries were subjected to online sexual exploitation and abuse in the past year alone.[6] Data from a recent global survey by Finnish organisation Protect Children on the dark web shows 70% of the nearly 20,000 respondents said they were minors the first time they encountered CSAM.[7]

Age Assurance

Legislation is coming into effect globally to **require companies to age assure on their platforms.** Balance is needed between child and data protection as well as proportionality of risk and child rights in digital environments. We have heard challenges around training data, cost of tools, age differentiation especially for age difficult stages, bias and implications of robustness of tools for inclusion. We have also heard the need to better classify risks and understand what good practice looks like beyond initial age assurance – what are the implications for children's access and experience online as well as responses from industry.

We aim to address some of these points in the design of the Call, such as exploration in this Open Call of facilitation of partnerships with organisations in the online CSEA community for access to data for training and validation purposes based on selection criteria for awarded grantees.

Live streaming of Abuse

We want to ensure that we include a scope that addresses CSEA in live streamed environments **beyond only the moment of live streaming including wider context considerations such as victimisation and offending pathways**. We recognise that live streaming intersects various offending behaviours and profiles: demand side (buyers), supply side (sellers), hands-on abuse, and coerced self-generated sexually explicit instances including through extortion, grooming and production of sexually explicit content for financial gain by children. A shift needs to be made in current approaches to prevention, detection and reporting of live streaming to focus more on making digital environments less hospitable to bad actors. There are several technical challenges to live detection and video content moderation. We have heard technical approaches focusing on chat or audio rather than video detection using existing NLP and other tools.

There are specific considerations for multistakeholder collaboration and how to enable more cross-platform and cross-sectoral work given the nature of this crime.

DEFINITIONS

Age assurance is necessary and legally required under certain service provision and data protection laws. Governments are beginning to require online platforms and services to offer age-appropriate content and experiences. However, significant opportunities remain for developing accountability and transparency

[6] End Violence, Disrupting Harm https://www.end-violence.org/disrupting-harm#findings

[7] Protect Children, https://www.suojellaanlapsia.fi/en/post/protect-children-s-research-in-the-dark-web-is-revealing-unprecedented-data-on-csam-users

standards and mechanisms for industry platforms and services around age assurance and age appropriateness online.

Age assurance definition: Umbrella term for service-level means of checking the age of users with various degrees of certainty.[8] The word 'assurance' refers to the varying levels of certainty that different solutions offer in establishing an age or age range. Under age assurance falls a range of methods from age verification to age estimation.

Age verification: A system that relies on hard (physical) identifiers and/or verified sources of identification, which provide a high degree of certainty in determining the age of a user. It can establish the identity of a user but can also be used to establish age only.[9] Verification often is based on official or government issued identification documents and has considerations of inclusion for groups without access to this type of documentation.

Age estimation (AE) A process that establishes a user is likely to be of a certain age, fall within an age range, or is over or under a certain age. Age estimation methods include automated analysis of behavioural and environmental data; comparing the way a user interacts with a device or with other users of the same age; metrics derived from motion analysis; or testing the user's capacity or knowledge.[10] Estimation has a lower level of certainty and some of the emerging common methods include biometric – such as facial, voice, dexterity, based methods – or contextual information – such as search history. Under proposed legislation would potentially apply to platforms deemed lower risk under standard risk assessments.

Legislation

Governments are making progress in improving regulation of digital spaces - e.g. the European Union (EU) released a new legislative proposal that has the potential to make an impact beyond the EU and help advance the global fight against online CSEA. The Digital Services Act (EU)[11] for example, mandates protection of minors and explicitly mentions the need for civil society input to elaborating sufficient measures, the DSA does not mandate age verification explicitly but suggests as a measure and opens a nuanced conversation about due diligence for risk mitigation for online child safety. euCONSENT[12] is aiming to create a safer digital world for children focusing on age assurance and parental consent. The Online Safety Bill (UK)[13] is to be put forward next year and discussions are currently taking place about the inclusion of an age assurance code. California introduced a new Age-Appropriate Design Code[14] and Australia recently started issuing legal orders to digital service providers under the AU Online Safety Act 2021[15] to ensure digital services' transparency and accountability.

---

[8] WeProtect Global Alliance & Yoti, *The role of age verification technology in tackling child sexual exploitation and abuse online,* https://www.weprotect.org/library/the-role-of-age-verification-technology-in-tackling-child-sexual- exploitation-and-abuse-online/

[9] 5RIghts, How Do They Know It is a Child, https://5rightsfoundation.com/uploads/But_How_Do_They_Know_It_is_a_Child.pdf

[10] ibid

[11] Digital Services Act, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[12] euCONSENT, https://euconsent.eu/download/understanding-of-user-needs-and-problems-a-rapid-evidence-review- of-age-assurance-and-parental-controls/

[13] Online Safety Bill, UK Parliament, https://bills.parliament.uk/bills/3137

[14] California Age-Appropriate Design Code, 5Rights, https://californiaaadc.com/

[15] eSafety Comissioner, https://www.esafety.gov.au/newsroom/media-releases/new-online-safety-laws-come-force

Standards are being developed in the space to support accountability and consistency in reporting such as the IEEE Standard for an Age-Appropriate Digital Services Framework[16]. However, challenges still exist in resolving concerns from privacy advocates in taking some measures for online child safety forward as well as complexity to this emerging legislative landscape that must be considered holistically and with respect to all human rights concerns.

The standard across most tech industry services currently is an age gate – that typically asks only for the date of birth or age to be entered by the user. Conversations for how to move forward range from age verification as a more technically robust mechanism to exploring age estimation using AI with greater margins of error. Technical tools exist and are currently used for microtargeting[17] and are working well for this purpose on certain platforms. There is an intersection for these mechanisms as well with age-appropriateness and verifiable parental consent. There is a balance needed between child and data protection as well as proportionality of risk and child rights in digital environments.

A consideration raised by industry stakeholders is that solely spotlighting age assurance could divert from innovative, holistic approaches. There is emphasis on the balance of risk and safety with privacy and participation. Proportionality along with data minimisation and purpose limitation are highlighted in industry discussions.[18] Other stakeholders such as CSOs and child safety advocates have offered to expand upon these core principles to ensure approaches are child-friendly, and to focus on inclusivity and data sharing. Opportunities exist in mapping standards and accountability mechanisms such as child rights impact assessment – similar to data protection impact assessments in industry – to drive decision making for features and behavior modification and support regulatory oversight.

Age verification tools exist or are being developed, focusing on decentralisation for privacy-preservation, but it is not yet clear if child rights standards are being taken into account upfront in these methods.[19] Discussions around age or developmental appropriateness, broader implications of digital identity for children and young people, and data / technology ethics such as algorithmic accountability are crucial to include in these dialogues going forward as well.

Live streaming is raising significant technical challenges along with conceptual ones as live streaming falls at the intersection of child sexual abuse material (CSAM) and online CSEA and so solutions must bridge existing silos in the ecosystem.

Live streaming definition: There is no universally agreed definition for the offence of live streaming child sexual exploitation and abuse. However, examples include live streaming abuse of a child occurring offline or children coerced into performing sexually explicit acts in front of a webcam.[20] Within the scope of this

[16] IEEE Standard Based on the 5Rights Principles for Children, https://standards.ieee.org/ieee/2089/7633/

[17] *Microtargeting (also called micro-targeting or micro-niche targeting) is a marketing strategy that uses consumer data and demographics to identify the interests of specific individuals or very small groups of like-minded individuals and influence their thoughts or actions.*

[18] Family Online Safety Institute, FOSI 2022 Annual Conference, Panel: *Making Age Assurance a Reality*, 23 June 2022, https://cdt.org/event/fosi-2022-panel-making-age-assurance-a-reality/

[19] Simon van der Hof, London School of Economics, *Age assurance and age appropriate design: what is required?*, 17 Nov 2021, https://blogs.lse.ac.uk/parenting4digitalfuture/2021/11/17/age-assurance/

[20] WeProtect Global Alliance, WPGA Global Threat Assessment 2021, https://www.weprotect.org/global-threat- assessment-21/#report

call, any video content shared live – whether broadcast or in direct communication channels – is within the definition of this form of abuse.

There are several elements[21] that can distinguish trends or dynamics of this phenomenon such as:

- the form of in-person abuse including,
    - coerced self-produced abuse – cases when the child or teenager has been groomed or is being coerced into creating the live stream. In such cases, only the victim may be visible, but there might be indicators that they were interacting with or being instructed by someone remotely such as the victim pausing to read or respond to typed chat.
    - coerced peer-to-peer abuse – cases when one child or teenager has been groomed or is being coerced into abusing another child or teenager during a live stream. In such cases, there might again be indicators that one subject is interacting with or being instructed by someone remotely.
    - distant live streaming - cases where another adult is present, orchestrating or involved hands-on in the abuse live – with sexual and/or financial motivations. Adult viewers might just watch the abuse, or could interact with the child being abused, orchestrating and directing the abuse, or could have ordered the abuse remotely.
    - perceived first person - cases where a child or teenager creates the live stream themselves, and does so voluntarily i.e. without coercion from another party and then this kind of live streamed material might be intended for a girlfriend/boyfriend, or be without sexual intent, but might be accessed by those viewing it for sexual purposes.
- if the live streaming takes place on private or public platforms/channels;
- if there is a 2-way communication aspect or if the abuse is broadcast;
- the profiles of offenders involved – e.g. financially or sexually motivated, rape by proxy, and/or CSAM viewing.

Material of all kinds of live streaming of abuse (such as recordings or screenshots post-event which are shared online) is treated as CSAM.

Live streaming intersects various offending behaviours and profiles: demand side (buyers), supply side (sellers), hands-on abuse, and coerced self-generated sexually explicit instances including through extortion, grooming and production of sexually explicit content for financial gain by children. A key aspect of this form of child sexual exploitation and abuse online is it thrives on economic inequality.[22] There is often an element of transaction or payment that accompanies this form of online CSEA.

Instances of live streaming entail activities that often cross several platforms such as financial services for payment and various online platforms for grooming, procurement and viewing of abuse. Coordination across industry platforms and services as well as across sectors – such as between tech industry and law enforcement are critical and complex in addressing this issue. Limited tools and capacity currently exist with significant legal barriers and technical challenges for effective cross-platform, cross-sectoral, and cross-jurisdictional collaboration.

[21] INHOPE, What is Livestreamed abuse, https://inhope.org/EN/articles/what-is-live-streamed-abuse
[22] WeProtect Global Alliance, https://www.weprotect.org/issue/live streaming/

Multistakeholder Collaboration

Few efforts currently exist specifically targeting live streamed child sexual abuse. Some examples of tools in this space are emerging from international law enforcement and non-profit organisations to address the time it takes to identify and capture necessary information for effective victim identification and prosecution of such instances. For example, Thorn14 has helped to reduce investigation times for law enforcement significantly through their investigative tools and work with CSAM classifiers. Child Rescue Coalition15 has created tools to help investigators process data from multiple sources and in multiple formats at quicker speeds. INHOPE16 has begun creation of a common language/ontology for the categorisation of child sexual abuse material schemas to, in turn, facilitate automated translation of these schemas with the purpose of more effective processing of CSAM reports by hotlines, law enforcement and industry.

Opportunity exists in building national capacities and international collaboration which is currently under- resourced and limited by policy and technical challenges. DevOps17, for example, is the only multi- stakeholder platform – bringing together law enforcement, academia, and IT - dedicated to designing technical solutions for and providing access for law enforcement in Interpol member states to resources to prevent, detect, and investigate online CSEA crimes. This initiative has designed a proof of concept for monitoring live streaming platforms as well as work on AI classifiers for age detection, self-generation media detection, and broad geolocation detection.

More and more online content is video-based, which presents unique challenges to detection and prosecution of online CSEA. Existing technology to detect, moderate or prevent live streamed child sexual abuse material faces significant technical challenges given the speed and real-time nature of live streaming of abuse, and this problem is exacerbated by end-to-end encryption. In addressing live streaming, solutions around more dynamic emerging technologies that are host to constantly evolving and often blurry edge cases when it comes to risks are challenging. Content moderation faces substantial challenges in live streamed environments. Video content includes several layers of data and technical challenges arise in training AI to determine meaning.

Understanding and solutions that address the wider context of live streaming of abuse as well as victimisation and offending pathways are critical. A shift needs to be made in current approaches to prevention, detection and reporting of live streaming to focus more on making digital environments less hospitable to bad actors. Considerations in addressing live streaming include not only content risks and moderation challenges but also behavioural ones, such as patterns of offenders - including rises in supply-side offending - and the complexity of grooming interactions online, as well as intersections with age verification to assess the problem. This requires multidisciplinary and cross-sectoral collaboration.

SCOPE

Examples – *these are not comprehensive in any way, they are simply a sample of more specific areas of exploration and interest* – of the types of questions that could be answered under each of these topics and around which we would welcome proposals include the following:

	Technology Solutions	Policy / Social Research
Age assurance	- Where do gaps exist in the technology landscape for addressing age assurance? - How can age differentiation at a more granular level be built into tools? - With better access to training data, what technical solutions could be further explored by other stakeholders? - How can accuracy and effectiveness of tools or intersections of multiple tools be developed and assessed? - How can tools better address circumvention? - How can design of tools ensure inclusion and/or improve existing bias? - How will age assurance and digital identity look in the decentralised web?	- What does good practice look like for age-appropriate action / response after age assurance? - How can solutions better differentiate across geographies and populations? - What are proactive preventative or educative measures that can sit alongside age assurance for increased efficacy? - How can better accountability measures, risk assessments and comparability standards of tools be developed? - How can the necessary level of age assurance be assessed for different cases, such as online pornography versus wagering, and a unified rather than issue-by-issue approach be taken?
Live streaming	- What are tools that exist for addressing live streaming of abuse and how effective are they? - Where do gaps exist in the technology landscape for addressing live streaming? - What tools should be prioritised to build to address live streaming? - What are feasible technical solutions to cross-platform sharing? - How can tools used for related challenges, such as grooming detection, be applied to live streaming? - What are feasible solutions in end-to-end encrypted environments?	- What happens after detection of live streaming of abuse? - At what points are interventions most effective? What kind of interventions? - How can multidisciplinary collaboration be effectively facilitated? - What are the legal and policy challenges, such as jurisdictional barriers, to deploying tools in direct communication environments? - What are specific legal challenges around client side detection of content? - How does radicalised violence against certain groups (misogyny, homophobia, racism) intersect with live streamed abuse trends and how can proactive educative approaches complement technical solutions effectively?

Across any of these four intersections above there are significant cross-cutting considerations for enhancing impact of solutions. Whether Proposals focus on age assurance or live streaming of abuse, we encourage technology or research outputs to incorporate / address questions such as:

- How can the development of open-source tools be promoted?
- What are tradeoffs in approaches and tools that operate at different levels of the technology stack?
- What are existing methods that have proven useful in specific use cases for online CSEA or other fields that can be applied to further online CSEA use cases?
- How is the heterogeneity of digital technology companies taken into account across solutions?
- How can downstream impacts of different interventions be assessed - e.g. implications for harm reduction, behaviour change, or advancing investigations?
- What are the implications of new and evolving technology and trends in tech industry, such as end-to-end encryption, distributed web, deep fakes and synthetic imagery, etc.

COLLABORATION

We would like to consider more effective ways to engage and facilitate collaboration between child safety and privacy communities around these issues as well as other major fields or agendas that have significant overlap in thinking about and developing solutions to age assurance and live streaming challenges – including topics such as gender-based violence or radicalisation.

We would like to ensure that we are using this Call as an opportunity to involve key communities who might not be actively engaged in these topics currently or who are not the usual suspects in online CSEA efforts. The topics covered by this Open Call are complex and nuanced and will require new and interesting constellations of partnerships to address effectively.

The strongest bids are therefore likely to come from organisations who:

1. are able to demonstrate skills and expertise across a number of disciplines - for example, social science, data science, knowledge of online harms, experience with child participation, privacy and security issues;
2. can demonstrate the effectiveness of solutions by testing on real use cases – in the case that online CSEA data is not available for research or testing tools, testing on other datasets to show operational feasibility and plans for adaptation to online CSEA use cases is acceptable (this is not an essential requirement);
3. propose solutions that focus on detection rather than prevention of activities around Child Sexual Abuse Material.

Safe Online is able to facilitate partnerships between organisations that are willing to apply to the Open Call, but currently do not have a consortium in place. Safe Online will identify these types of opportunities during the Proposal evaluation phase (not during the application stage), and this opportunity would be facilitated only for organisations shortlisted for award. Please indicate in your application in the related Partners question if this would be of interest in the case of shortlisting. You can consult Safe Online portfolio here.