Tech Coalition Human Rights Impact Assessment of the Lantern Program

November 2023

BSR®

Table of Contents

Acknowledgements

BSR and the Tech Coalition wish to thank all Tech Coalition employees, members, rightsholders, and stakeholders who participated in this assessment.

DISCLAIMER:
The conclusions presented in this document represent BSR's best professional judgment, based upon the information available and conditions existing as of the date of the review. In performing its assignment, BSR relies upon publicly available information, information provided by Twitch, and information provided by third parties. Accordingly, the conclusions in this document are valid only to the extent that the information provided or available to BSR was accurate and complete, and the strength and accuracy of the conclusions may be impacted by facts, data, and context to which BSR was not privy. As such, the facts or conclusions referenced in this document should not be considered an audit, certification, or any form of qualification. This document does not constitute and cannot be relied upon as legal advice of any sort and cannot be considered an exhaustive review of legal or regulatory compliance. BSR makes no representations or warranties, express or implied, about the business or its operations. BSR maintains a policy of not acting as a representative of its membership, nor does it endorse specific policies or standards. The views expressed in this document do not reflect those of BSR member companies.

1

Executive Summary

Introduction

The Tech Coalition commissioned BSR to undertake a human rights impact assessment (HRIA) of the Lantern Program, a new signal sharing initiative that will enable technology companies to send and receive online identifiers or indicators ("signals") related to online child sexual exploitation and abuse (OCSEA).

The Lantern Program has been established to address the cross-platform nature of OCSEA, and to protect the rights of children by supporting companies in their efforts to address OCSEA-related harms. However, the use of a cross-platform signal sharing program by companies may have unintended adverse impacts on human rights, such as privacy, nondiscrimination, and freedom of expression. The goal of this HRIA is to:

- Identify and prioritize human rights impacts with which the Tech Coalition is involved through the Lantern Program, including both risks and opportunities, and the vulnerable groups impacted.
- Recommend appropriate action for the Tech Coalition and participants of the program to address these impacts (i.e., avoid, prevent, mitigate, and remedy).

Methodology

BSR undertook this HRIA between April and August 2023 using a methodology based on the UN Guiding Principles on Business and Human Rights (UNGPs), including a consideration of the various human rights principles, standards, and methodologies upon which the UNGPs were built, as well as on the Children's Rights and Business Principles (CRBP).

Consistent with the UNGPs, the prioritization of human rights impacts in this assessment is based on risks to people rather than risks to business.

This HRIA draws upon the human rights concepts of severity (defined as scope, scale, and remediability) and likelihood to inform a prioritization of impacts:

- **Scope**—The number of people affected by the harm.
- **Scale**—The seriousness of the harm for those affected.
- **Remediability**—The extent to which remedy will restore those affected to the same or equivalent position before the harm.
- **Likelihood**—The probability and/or frequency of the adverse human rights impact occurring in the next five years. Factors involved in an assessment of likelihood include whether (or how frequently) the impact has happened in the past or is happening today, whether (or how frequently) similarly situated companies have been involved with a similar impact, and whether the impact has been foreseen during research for the assessment, including during discussions of future trends.

This HRIA makes recommendations for the Tech Coalition to address adverse human rights impacts using factors contained in Principle 19 of the UNGPs:

- **Attribution**—How closely would the Tech Coalition be connected to the human rights impact?
  - ***Caused* the impact**—The Tech Coalition should take the necessary steps to cease or prevent the impact.
  - ***Contributed* to the impact**—The Tech Coalition should take the necessary steps to cease or prevent its contribution and use its leverage to mitigate any remaining impact to the greatest extent possible.
  - ***Directly* linked to the impact through its products, services, or operations arising from its business relationships**—The Tech Coalition should determine action based on factors such as the extent of leverage over the entity concerned and the severity of the abuse.
- **Leverage**—How much ability would the Tech Coalition have to affect change in the wrongful practices of an entity that "causes" or "contributes to" the harm? How much ability does the Tech Coalition have to seek modification of or challenge the wrongful practice? How can the Tech Coalition increase leverage?

Key Observations

BSR's analysis of the human rights impacts and appropriate actions to address them were influenced by the following observations:

- The Lantern Program has the potential to fill a key gap in the industry's approach to child safety and protection. Currently, tech companies primarily address OCSEA risks via individual actions; however, research suggests that OCSEA perpetrators are increasingly operating across multiple platforms. This cross-platform nature of the crime necessitates a collaborative approach. By addressing this critical need, the Lantern Program aims to help companies address some of the most severe adverse human rights impacts associated with their platforms.
- Insights gleaned through the Lantern Program can be valuable for broader efforts to fight OCSEA and protect digital rights. Trends in OCSEA are constantly changing, and stakeholders may find it challenging to stay up to speed on all the ways the crime is shifting and evolving. By enabling the identification and dissemination of OCSEA-related trends and insights, the Lantern Program could be a valuable resource to stakeholders (e.g., civil society, companies, policymakers, academics) working to combat OCSEA globally.

- Signal sharing between companies may exacerbate certain human rights risks that exist within individual company efforts to fight OCSEA, such as those related to freedom of expression (e.g., through over-moderation of content) or privacy (e.g., through overbroad data sharing or monitoring). The cross-platform nature of the Lantern Program may also lead to cumulative impacts, as a result of multiple companies taking action on signals in ways that compound adverse impacts on human rights.
- Human rights risks and/or challenges may arise due to a variety or factors, including:
  - Unsubstantiated signals (i.e., when companies mistakenly identify users or content as potentially harmful and share related signals on the Lantern database);
  - Government requests or involvement with the Lantern Program;
  - The lack of clear or standardized definitions for some types of OCSEA-related behaviors, such as grooming, and the related risk of scope creep;
  - Differences in company policies or enforcement processes for handling edge cases (e.g., legal but harmful content) or gray areas (e.g., self-generated explicit imagery or peer-to-peer offenses);
  - Challenges related to age determination, assurance, and verification;
  - The development of new technologies such as diffusion models and generative AI that enable the creation and distribution of synthetic child sexual abuse material (CSAM).
- Children's interaction with the digital environment and associated impacts on their digital rights are constantly evolving. For example, their experiences with and attitudes toward self-generated explicit imagery are shifting rapidly and there is a lack of in-depth research in the field on how such online behaviors may affect children's development and digital rights. This may create challenges for online platforms to proportionately address the risks and empower children to enjoy their rights in the digital environment.
- Participant engagement with the Lantern Program varies significantly, mainly due to resourcing constraints, such as limited content moderation capacity, as well as philosophical differences between companies' content governance approaches. For example, while most participants conduct manual review and verification of signals shared via the Lantern database, others may automatically action signals without verifying the violating content or conduct on their platform.
- The Lantern Program can serve as a key resource for smaller technology companies with limited capacity to identify and investigate all OCSEA risks or harms occurring on their platform, by facilitating cross-industry knowledge sharing. While signal sharing can help companies prioritize content for review, a company still needs dedicated resources to action the signals and to effectively use the Lantern Program.
- Legal frameworks may complicate the Lantern Program's ethos of voluntary action. Although the Lantern Program is a voluntary program under which participants are not obligated to carry out any moderation activities, legal frameworks may create pressure on participants to take action on all signals received, which could deter participants from fully engaging in the Lantern Program or lead them to automatically action ingested signals.
- Accountability and responsibility related to the Lantern Program is shared between the Tech Coalition and participating companies. Each participating company is responsible for how they share and use signals as part of the program, and is individually accountable for risks associated with their use of the signals. The Tech Coalition, on the other hand, is responsible for putting

in place appropriate measures and mitigations to incentivize and enable participants to use the Lantern Program appropriately, and to identify and act upon instances when they do not.

- Transparency is a critical component of the Tech Coalition's governance approach for the Lantern Program, and it constitutes an important risk mitigation measure in itself. Lack of transparency may limit the effectiveness of the Lantern Program and lead to stakeholder skepticism. On the other hand, transparency may also come with risks—for example, by attracting governments and law enforcement to coerce the Tech Coalition or participating companies to share user information.

Human Rights Impacts

The Human Rights Impacts that the Lantern Program Seeks to Address

This HRIA identifies the human rights risks associated with the Lantern Program and makes recommendations for how those risks should be addressed. However, it is essential to recognize that the Lantern Program addresses existing adverse human rights impacts related to OCSEA.

OCSEA is one of the most severe known adverse human rights impacts associated with technology companies.

In 2022, the National Center for Missing and Exploited Children (NCMEC) received over 32 million reports containing approximately 88 million suspected child sexual exploitation images and videos. There has been an 82% rise in online grooming crimes against children in the last five years, and reports of online sextortion have almost tripled in 2023.

OCSEA often results in serious adverse physical, emotional, and psychological impacts on victims and survivors, with children being a vulnerable group owing to their physical and mental immaturity. These adverse impacts can be lasting and continue long after the incident of abuse has ended, making remediation especially challenging.

The Lantern Program seeks to enhance the capacity of the technology industry to combat OCSEA, and thereby address one of the most egregious known harms that exists in the industry.

The Lantern Program helps companies fulfill their responsibility to address adverse human rights impacts with which they are already associated and will significantly increase their leverage to do so effectively.

BSR identifies the following eight categories of human rights that may be impacted with the use of the Lantern Program. Some of these categories have multiple human rights grouped together.

- **Child Safety and Protection:** The Lantern Program seeks to promote the enjoyment of children's right to protection from sexual abuse and exploitation by establishing effective procedures to identify, prevent, and mitigate threats to children's safety on digital platforms.

- **Civil, Social, Cultural Rights and Freedoms of Children:** While the main purpose of the Lantern Program is to protect children from online sexual exploitation and abuse, there is a risk that the collection, use, and sharing of information in the Lantern Program, as well as actions taken by participants based on signals, may adversely impact the ability of children to enjoy their civil rights and freedoms, such as access to information, privacy, freedom of expression, freedom of thought, or participation in cultural life.
- **Privacy:** Signals shared in the Lantern database may include personally identifiable information about users such as email addresses or account names. The program may be associated with risks to privacy if participants use or share data in ways that result in arbitrary interference with users' privacy, family, home, or correspondence.
- **Freedom of Expression:** Although the Lantern Program has been set up to target content or speech related to child sexual exploitation or abuse, there is a risk that legitimate or non-violative content may be erroneously removed due to overbroad moderation by participants, and/or that users are wrongfully denied access to online platforms where they can exercise their right to free expression and to access information.
- **Equality and Nondiscrimination:** Certain groups or communities may be at greater risk of being wrongly accused of OCSEA and be subject to investigations or punitive actions, due to societal biases (e.g., against LGBTQIA+ people or sex workers), or content moderation tools may not perform as accurately for certain groups (e.g., people of color) and languages. The Lantern Program can proactively address these biases and discriminatory practices by setting guardrails and sharing insights with the broader field.
- **Due Process and Effective Remedy:** When participants take action based on signals shared in the Lantern Program database, this may result in content being removed and user accounts being deleted or flagged for monitoring. In such cases, users may be penalized without notice for conduct or actions taken outside of a platform, and they may not be provided effective or accessible mechanisms for appeals or complaints.
- **Bodily Security:** There is a risk that actions taken on the basis of signals shared in the Lantern Program may wrongfully result in offline harms to users such as arbitrary arrest, detention, or investigation of users, particularly if signals are shared with and misused by governments or law enforcement agencies.
- **Economic, Social, and Cultural Rights:** Digital platforms have become an integral part of social, cultural, and economic life, including the facilitation of work and education and participation in cultural life. The Lantern Program may adversely affect users' ability to enjoy these rights if users are wrongfully denied access to platforms as a result of signals shared in the database.

There are multiple ways in which the Lantern Program may impact the eight categories of human rights listed above. In BSR's analysis, the main pathways that can lead to human rights impacts are:

- **Data Collection, Storage, and Sharing:** The collection, storage, and sharing of data as part of the signal sharing program may be associated with adverse impacts on users' rights, including the right to privacy.
- **Actioning of Signals:** The various actions participants may take on the basis of signals in the Lantern database, such as wrongful account removals, over-moderation of content, or non-compliance with Lantern Program guidelines, may adversely impact users' rights.

- **Government Involvement:** Governments and law enforcement agencies may attempt to gain direct or indirect access to the Lantern database, or influence participants' use of the Lantern Program in ways that adversely impact users' rights.
- **Unintended Consequences on Children:** Although the Lantern Program exists to promote children's safety and protection rights, the use of the database may have unintended consequences on children, adversely impacting their rights.

The Tech Coalition should pay particular attention to individuals from groups or populations that may be at heightened risk of vulnerability or marginalization. These include:

1. **Actual and potential victims of OCSEA:** While all children may be victims of OCSEA, some children are at particular risk, including LGBTQIA+ children, homeless children or children in care, and children with intellectual disabilities.
2. **Users adversely impacted by efforts to fight OCSEA:** While all users of tech platforms are at risk, certain vulnerable populations are at greater risk of being wrongly accused of OCSEA. Also, being wrongfully accused of OCSEA may have more severe impacts on certain vulnerable populations, including LGBTQIA+ people, sex workers, people with intellectual disabilities, people of color, underrepresented linguistic communities, and children.

Recommendations

The Tech Coalition has already made significant efforts to prevent, mitigate, and address the human rights risks associated with the Lantern Program. Section 7 of this assessment provides 19 recommendations across four key areas to supplement and enhance the Tech Coalition's existing efforts, as well as to underline certain areas of focus. The following is a summary of the recommendations:

1. Governance and Participant Engagement

- › Enforce participant commitments by conducting mandatory training and regular check-in meetings with participants.
- › Ensure that participants have a process in place for handling government requests for user data, including applying human rights principles when responding to requests and publishing an annual transparency report.
- › Support smaller participants' ability to comply with the requirements of the Lantern Program by providing additional quality assurance and integration support.
- › Provide guidance and best practice to participants—for example, on how to conduct signal sharing, handle gray areas, and provide educational resources to children.
- › Establish a due diligence process for evaluating new members and ensuring that they have the necessary policies, processes, and culture in place to address the risks associated with signal sharing.
- › Consider participation by "non-tech" industries, such as airlines or financial services companies, that can provide signals related to OCSEA through their own use of technology, with careful consideration of new risks that may arise from such participation.

2. Transparency and Stakeholder Engagement

- Enhance the Lantern Program's approach to transparency to enable stakeholders to understand how the program is being used and whether it is achieving its stated goals, while carefully considering the risks involved with transparency.
- Facilitate an annual dialogue with key stakeholders where participants can share and discuss their use of the Lantern Program, and share insights and lessons learned.
- Establish a strategy for ongoing stakeholder engagement with affected stakeholders and experts, informed by best practices in stakeholder engagement.
- Establish an advisory board with interdisciplinary expertise for the Lantern Program, which can help address important questions related to human rights and child rights.
- Share insights and work with stakeholders to contribute to broader industry, civil society, and policy efforts to fight OCSEA.
- Investigate enabling third-party audits and trusted researcher access to the database as the Lantern Program matures and usage by participants is more established.

3. Policy and Process

- Establish a robust quality assurance (QA) process that enables review of signals for quality and relevance, and feature feedback loops that can alert the Tech Coalition to trends in new signals.
- Monitor the use of the Lantern Program for non-English languages and make improvements as needed.
- Implement additional procedural requirements for "high-risk" signals, such as a manual review by experts, or a checklist that participants need to fill out before uploading signals to the database.
- Take a human rights-based approach to responding to government requests for data, using the Global Network Initiative Principles and Implementation Guidelines as a starting point.
- Conduct ongoing human rights due diligence (HRDD) to identify and assess key developments that may indicate shifts in human rights risks or impacts of the Lantern Program over time.

4. Technical Measures

- Implement technical barriers to minimize data collection, storage, and sharing, such as preventing bulk download of signals and establishing time limits for personal data to be stored in the database.
- Build a flagging system for signals that are not allowed or are high risk.

2

Introduction

2.1 Background

The Tech Coalition is an alliance of global tech companies working together to combat online child sexual exploitation and abuse (OCSEA), which is one of the most egregious harms that exist in the tech industry. To address the cross-platform nature of OCSEA, the Tech Coalition is launching the Lantern Program, a new signal sharing initiative that will enable technology companies to send and receive online identifiers or indicators ("signals") of suspected or known child sexual abuse materials (CSAM), OCSEA, and/or related activity.

The primary purpose of the Lantern Program is to protect the rights of children by supporting companies in their efforts to address OCSEA-related harms. However, the Tech Coalition recognizes that the use of a cross-platform signal sharing program by companies may have unintended adverse impacts on human rights, such as privacy, nondiscrimination, and freedom of expression. The Tech Coalition aims to embed respect for human rights into the program from the start by undertaking a human rights impact assessment (HRIA) to identify and assess these potential adverse human impacts as part of the development and launch of the Lantern Program.

The increase of OCSEA[1] and associated company and regulatory efforts to address OCSEA have given rise to a public policy debate on how to most effectively pursue online child safety while also protecting human rights.[2] While this debate is sometimes framed as pitting child safety advocates against digital rights advocates in the name of different rights and rightsholders, the reality is much more nuanced. There are a wide range of interconnected human rights impacted by both OCSEA and corresponding CSEA-mitigation efforts, and it is important that the full range of human rights impacts are identified and addressed.

A human rights-based approach considers the full range of human rights issues at stake and allows for an analysis of the nuances and tensions. It gives special consideration to the needs of individuals from groups or populations that may be at heightened risk of becoming vulnerable or marginalized, including children, who are particularly vulnerable to online risks.[3]

[1] Severe Child Sexual Abuse Material Online Has More Than Doubled Since 2020: Internet Watch Foundation Report.

[2] European Commission's Online CSAM Proposal Fails to Find Right Solutions to Tackle Child Sexual Abuse: European Digital Rights network (EDRi).

[3] A 2021 study in the UK showed that vulnerable young people can be up to seven times more likely to experience online harm.

In light of the numerous human rights issues at stake, and in line with international human rights obligations, the Tech Coalition and companies participating in the Lantern Program have a responsibility to take action to assess and address potential harms related to the signal sharing initiative. This HRIA is an essential foundation for the fulfillment of that responsibility and represents a significant milestone in the development of rights-based approaches to the fight against OCSEA.

2.2 About This Project

The Tech Coalition engaged BSR (a global nonprofit organization working with companies on just and sustainable business) to undertake an HRIA of the Lantern Program. The goal of the HRIA is to:

- Identify and prioritize human rights impacts with which the Tech Coalition is involved through the Lantern Program, including both risks and opportunities, and the vulnerable groups impacted.
- Recommend appropriate action for the Tech Coalition and participants of the program to address these impacts (i.e., avoid, prevent, mitigate, and remedy).

As a result of this HRIA, the Tech Coalition should have a deepened understanding of the potential human rights impacts with which they are involved through the Lantern Program, the insights necessary to address them, and the knowledge required to take a human rights-based approach to signal sharing and governance of the program.

3

Methodology

BSR undertook this HRIA between April and August 2023 using a methodology based on the UN Guiding Principles on Business and Human Rights (UNGPs), including consideration of the various human rights principles, standards, and methodologies upon which the UNGPs were built, as well as on the Children's Rights and Business Principles (CRBP).

While the UNGPs and CRBPs were written for use by companies rather than coalitions or multi-stakeholder efforts, their overall spirit and approach can be applied to the Tech Coalition. In addition, the members of Tech Coalition are companies, and so it can be assumed that the UNGPs and CRBPs apply for that reason.

Business, Human Rights, and Child Rights

States have the obligation to respect, protect, and fulfill human rights. However, in 2011, the UN Human Rights Council unanimously endorsed the UN Guiding Principles on Business and Human Rights (UNGPs), which set out the responsibility of companies to respect human rights. The UNGPs apply to all companies, regardless of their size, sector, location, ownership and structure.

The UNGPs set out the expectation that companies should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved. Companies should carry out human rights due diligence to identify, prevent, mitigate, and account for their actual and potential adverse human rights impacts. Human rights due diligence requires assessing actual and potential adverse human rights impacts, integrating and acting on the findings, tracking responses, and communicating how impacts are addressed. Companies are also expected to have in place remediation processes to address adverse human rights impacts which they have "caused" or "contributed" to.

The preamble to the UNGPs clearly states that companies should pay special attention to the rights and needs of, as well as the challenges faced by, individuals from groups or populations that may be at heightened risk of becoming vulnerable or marginalized, which for most companies includes children.

Published in 2012, the Children's Rights and Business Principles (CRBP) build upon the UNGPs by providing an operational framework to guide companies on the full range of actions they can take in the workplace, marketplace, and community to respect and support child rights.

3.1 Identifying Human Rights Impacts

In this assessment, BSR identifies the actual and potential human rights impacts of the Lantern Program using the universe of human rights codified in the following international human rights instruments:[4]

- The Universal Declaration of Human Rights
- The International Covenant on Civil and Political Rights
- The International Covenant on Economic, Social, and Cultural Rights
- The International Convention on the Elimination of All Forms of Racial Discrimination
- The Convention on the Elimination of All Forms of Discrimination Against Women
- Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment
- Convention on the Rights of Persons with Disabilities
- The ILO Declaration on Fundamental Principles and Rights at Work and the eight International Labor Organization (ILO) Core Conventions
- The Convention on the Rights of the Child (CRC)
- ILO Convention 169 on Indigenous Peoples

Child Rights in the International Human Rights Framework

International human rights instruments form the foundation of child rights and protections globally. The underpinning concept of human rights is simple and powerful: people have the right to be treated with dignity. Human rights are inherent to all human beings, regardless of age, nationality, place of residence, sex, national or ethnic origin, color, religion, language, or any other status.

The Universal Declaration of Human Rights (UDHR) enumerates the "fundamental human rights to be universally protected" and establishes a "common standard of achievements for all peoples and all nations." While children have all the rights set out in the UDHR, rights and protections for children were expanded upon in the 1959 Declaration of the Rights of the Child and in subsequent international conventions.

The International Covenant on Civil and Political Rights and the International Covenant on Economic Social and Cultural Rights—adopted in 1966—introduce the rights to education and protection for all children. The International Labor Organization (ILO) set the minimum legal working age at 15 years in 1973.

Recognizing children's specific vulnerabilities and childhood as a unique period in human development, the UN General Assembly adopted the Convention on the Rights of the Child in 1989. The CRC is an international legal framework for the protection and promotion of the human rights and fundamental freedoms of all persons under the age of 18. It incorporates the full range of human rights, including civil, cultural, economic, political, and social rights.

[4] These are the core international human rights instruments and other instruments and/or conventions that are potentially most relevant for the scope of this assessment.

Two optional additional protocols were adopted in 2000 to give specific protections related to the involvement of children in armed conflict and the sale and exploitation of children. A third optional protocol grants the Committee of the Rights of the Child the ability to investigate allegations of child rights violations.

All human rights are indivisible, interdependent, and interrelated. The improvement of one right facilitates advancement of the others; the deprivation of one right adversely affects others. This point becomes especially relevant when rights may be in tension with each other (e.g., child safety and privacy) and a company needs to make choices when two competing rights cannot both be achieved in their entirety. Rather than "offsetting" one right against another, it is important to pursue the fullest possible expression of both rights and identify how potential harms can be addressed.

3.2 Prioritizing Human Rights Risks

Consistent with the UNGPs, the prioritization of human rights impacts in this assessment is based on **risks to people** (i.e., risks to rightsholders) rather than **risks to the business** (i.e., risks to enterprise value creation). This people-oriented approach enables a more meaningful human rights program and a more sophisticated approach to addressing material business risks.

Principle 24 of the UNGPs acknowledges that while companies should address all their adverse human rights impacts,[5] it is not always possible for companies to address them simultaneously, and companies should "first seek to prevent and mitigate those that are most severe or where delayed response would make them irremediable."

This HRIA draws upon the human rights concepts of severity (defined as scope, scale, and remediability) and likelihood to inform a prioritization of impacts. Consistent with the UNGPs, severity is not an absolute concept in this context, but is relative to the other human rights impacts with which the Lantern Program is involved:

- **Scope**—The number of people affected by the harm.
- **Scale**—The seriousness of the harm for those affected.
- **Remediability**—The extent to which remedy will restore those affected to the same or equivalent position before the harm.
- **Likelihood**—The probability and/or frequency of the adverse human rights impact occurring in the next five years. Factors involved in an assessment of likelihood include whether (or how frequently) the impact has happened in the past or is happening today, whether (or how frequently) similarly situated companies have been involved with a similar impact, and whether the impact has been foreseen during research for the assessment, including during discussions of future trends.

[5] While the UNGPs do not explicitly outline what constitutes an "adverse human rights impact" or a "harm," BSR understands this to mean any outcome in which a rightsholder, or group of rightsholders, experiences adverse impacts or is deprived of any of the human rights or fundamental freedoms outlined in international human rights instruments due to the direct or indirect actions or inactions of a state or business. Additionally, this may refer to outcomes in which a rightsholder, or group of rightsholders, is unable to exercise their human rights and fundamental freedoms to the full extent possible.

Human rights impacts prioritized using these concepts are referred to as **salient human rights issues**.

BSR used the following criteria to assess scope, scale, remediability, and likelihood. In all cases these criteria are necessarily directional and reliant on professional judgment, rather than precise calculations, and the factors influencing our analysis is shown in Section 6 (Human Rights Impacts).

Salience	Levels				
Scope How many people could be affected by the harm?	**Smallest** Smallest range of users and/or persons affected	**Small** Limited/smaller range of users and/or persons affected	**Medium** Majority of users and/or persons affected	**Large** Larger majority of users and/or persons affected	**Largest** Significant and/or all users and/or a significant population of qualified range of persons affected
Scale How serious would the impacts be for the impacted stakeholder?	**Least Serious** Unlikely to cause bodily harm / psychological damage / change to standard of living / livelihood	**Moderately Serious** Could result in indirect bodily harm / psychological damage / moderate change to standard of living / livelihood	**Serious** Likely to result in direct bodily harm / lasting psychological damage / major change of standard of living / livelihood	**Very Serious** May result in death or irreversible loss of physical or mental capacities / significant disruption in standard of living / livelihood	**Most Serious** Certain to result in death or irreversible loss of physical or mental capacities / significant disruption in standard of living / livelihood
Remediability / Irreversibility Will a remedy restore the impacted stakeholder to the same or equivalent position before the harm?	**Remediable** Remedy will return the impacted stakeholder to the same or equivalent position	**Likely Remediable** Remedy is likely to return the impacted stakeholder to the same or equivalent position before the harm occurred	**Possibly Remediable** Remedy may help return the impacted stakeholder to the same or equivalent position before the harm occurred	**Rarely Remediable** Remedies can rarely return the impacted stakeholder to the same or equivalent condition before harm occurred	**Not Remediable** Remedies will not return the impacted stakeholder to the same or equivalent condition before harm occurred
Likelihood of Occurrence What is the likelihood of the risk occurring?	**Minor Likelihood** Although a risk, it is highly unlikely that impacts may occur	**Some Likelihood** There is some minor risk that the impacts may occur	**Good Likelihood** It's more probable than not that the impacts may occur	**High Likelihood** There is a high likelihood that the impacts may occur	**Certain** Currently occurring or certain to occur

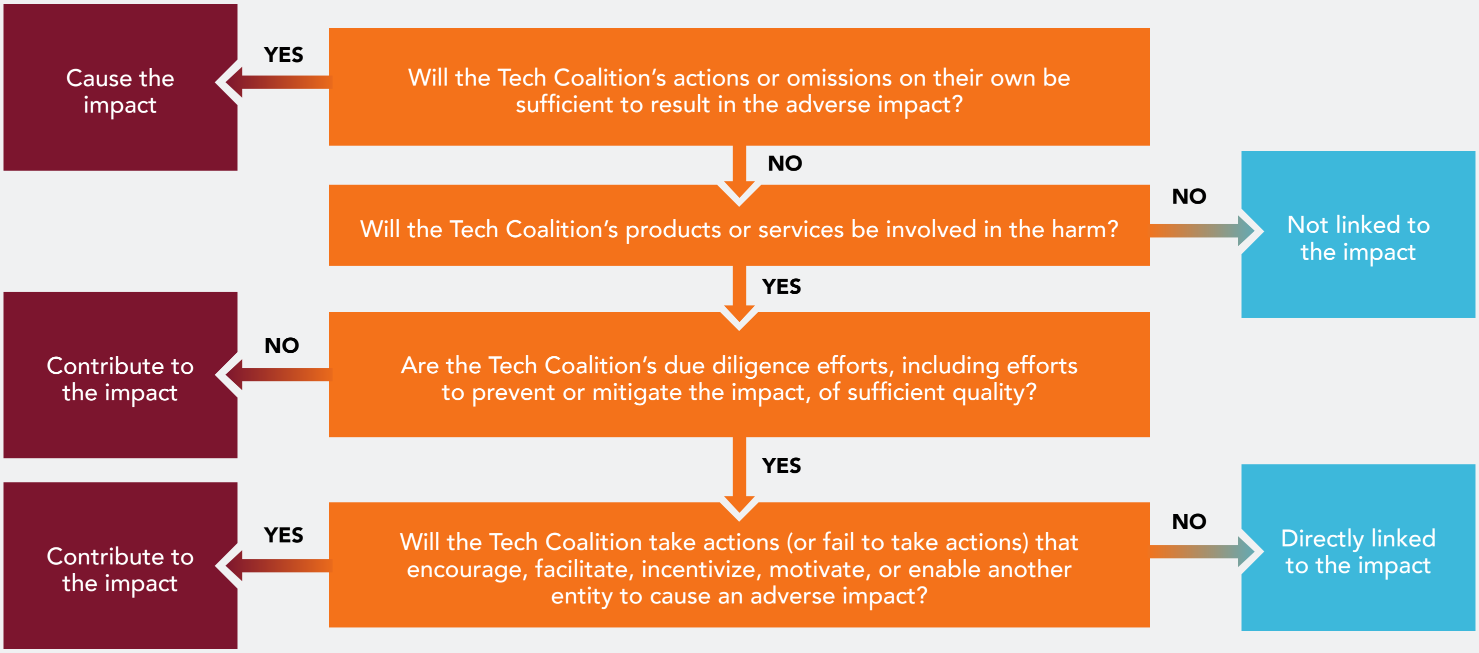
3.3 Determining Appropriate Action

BSR's HRIA methodology considers the appropriate action for a company to address adverse human rights impacts using factors contained in Principle 19 of the UNGPs:

- **Attribution**—How closely would the Tech Coalition be connected to the human rights impact? BSR uses the following definitions and "decision tree":
    - ***Caused* the impact**—The Tech Coalition should take the necessary steps to cease or prevent the impact.
    - ***Contributed* to the impact**—The Tech Coalition should take the necessary steps to cease or prevent its contribution and use its leverage to mitigate any remaining impact to the greatest extent possible.
    - ***Directly linked* to the impact through its products, services, or operations arising from its business relationships**—The Tech Coalition should determine action based on factors such as the extent of leverage over the entity concerned and the severity of the abuse.
- **Leverage**—How much ability would the Tech Coalition have to affect change in the wrongful practices of an entity that "causes" or "contributes to" the harm? How much ability does the Tech Coalition have to seek modification of or challenge the wrongful practice? How can the Tech Coalition increase leverage?

Applying the "cause, contribute, directly linked" framework to platforms hosting user-generated content (UGC) is challenging due to the complex ways in which platforms interact with, enable, and amplify human behavior.

'Cause-Contribute-Directly Linked' Framework



Will the Tech Coalition's actions or omissions on their own be sufficient to result in the adverse impact?

YES → Cause the impact

NO → Will the Tech Coalition's products or services be involved in the harm?

NO → Not linked to the impact

YES → Are the Tech Coalition's due diligence efforts, including efforts to prevent or mitigate the impact, of sufficient quality?

NO → Contribute to the impact

YES → Will the Tech Coalition take actions (or fail to take actions) that encourage, facilitate, incentivize, motivate, or enable another entity to cause an adverse impact?

YES → Contribute to the impact

NO → Directly linked to the impact

Given the importance of context, BSR considers this framework to be helpful in setting overall direction in an HRIA, rather than providing a definitive "answer" for each impact.[6]

Further, it is important to note that this decision tree is most useful when assessing individual cases (e.g., identifying whether the Tech Coalition caused, contributed to, or was directly linked to a specific adverse privacy impact) rather than overall categories (e.g., identifying whether the Tech Coalition causes, contributes to, or is directly linked to adverse privacy impacts more generally), and this makes our analysis necessarily general in nature. However, the factors influencing our analysis for each impact is shown in Section 6 (Human Rights Impacts).

It should also be noted that methods to apply the "cause, contribute, directly linked" framework to the technology industry are under development. BSR has utilized the most recent resources from UN-led processes exploring this framework in the context of the tech industry.[7]

3.4 Rightsholder and Stakeholder Engagement

Effective human rights due diligence requires meaningful engagement with rightsholders whose human rights may be impacted by the company, as well as external stakeholders such as independent experts, human rights defenders, and others from civil society. Particular attention should be paid to human rights impacts on individuals from groups or populations that may be at heightened risk of vulnerability or marginalization, such as children, women, low-income groups, members of the LGBTQIA+ community, and marginalized racial, ethnic, and religious groups.

While children are always vulnerable, particular attention should be paid to impacts on the rights of children from groups or populations that may be at heightened risk of vulnerability or marginalization.

We define rightsholders and stakeholders as follows:

- **Rightsholders:** Individuals whose rights could be directly impacted by the company. Rightsholders interact with the company and its operations, products, and services, typically as an employee, contractor, customer, user, or member of a particular affected community.
- **Stakeholders:** Organizations knowledgeable about and capable of speaking with informed insight of the needs, interests, and experiences of rightsholders, such as civil society organizations, activist groups, opinion formers, academics, policymakers, or regulators.

Vulnerability depends on context, and someone who may be powerful in one context may be vulnerable in another. Vulnerability can change across geographies, and in relationship to different products and applications of technology.

We identify vulnerable groups based on four dimensions:

- **Formal discrimination**—laws or policies, and/or their application, that favor one group over another
- **Societal discrimination**—cultural or social practices that marginalize some and favor others
- **Practical discrimination**—marginalization due to life circumstances, such as poverty
- **Hidden groups**—people who might need to remain hidden and consequently may not speak up for their rights

[6] See Seven Questions to Determine a Company's Connections to Human Rights Abuses for more analysis.

[7] See UN B-Tech Project, especially Taking Action to Address Human Rights Risks Related to End-Use.

In this HRIA, BSR engaged with rightsholders, civil society organizations, and experts to be able to provide insights into the risks and opportunities associated with the Lantern Program. This included organizations focused on child rights, and those focused on digital rights such as privacy, freedom of expression, and nondiscrimination in the technology industry.

BSR's analysis also benefits from insights shared by companies participating in the Lantern Program, who themselves also consult with stakeholders.

4 Overview of the Lantern Program

The Lantern Program is a framework for sharing information related to harmful threats (known as "signals") for the purpose of combating online child sexual exploitation and abuse ("OCSEA"). The Lantern Program seeks to facilitate cross-industry knowledge sharing, proactive detection of OCSEA, and industry-wide moderation of known or suspected child sexual abuse material ("CSAM"). The goal of signal sharing is to supplement company participants' individual content moderation and trust and safety efforts by providing information that helps companies identify cases of OCSEA or offenders who may be operating across different platforms.

4.1 Signal Sharing

The Lantern Program allows signal sharing among participants using a database built on Meta's ThreatExchange platform. Signals shared on ThreatExchange may include hashes (including of known CSAM), URLs, common keywords or phrases, text fields, search terms, or basic subscriber and user information (such as name, date of birth, gender, or contact information) that may indicate actual or potential cases of OCSEA. Participants are expected to manually review the signals that are shared in the database and establish violation of their own platform policies before taking action based on signals. Signals on ThreatExchange can be shared or accessed via a user interface or an API.

4.2 Participation

Participation in the Lantern Program is currently limited to companies operating in the tech industry such as social media and other user-generated content platforms, messaging apps, and file-sharing services. There are currently 10 companies participating in the Lantern Program; some participants are members of the Tech Coalition, while others are not. To be eligible to join the Lantern Program, prospective participants are required to:

- Identify internal points of contact responsible for various aspects of the company's involvement in the Lantern Program,
- Have publicly accessible platform policies, including privacy policies and user content / conduct policies, as well as public reporting mechanisms for suspected child sexual exploitation or abuse on the platform,

- Have internal processes for alerting users of policy violations and providing appeals processes for affected users,
- Document internal processes for determining what signals will be shared with other participants, and how signals obtained from the Lantern Program will be used,
- Commit to complying with applicable privacy laws and best practices.

Prospective participants may join the Lantern Program by submitting an application to the Tech Coalition, which verifies the company's compliance with the Lantern Program eligibility requirements and shares the application with existing participants for their review, before approving the application. All applicants are subject to a thorough review and must be invited to join the program.

4.3 Governance

The Lantern Program is governed by a multiparty agreement that sets out the framework for signal sharing between participants, and which must be signed by participants before they are able to access the Lantern database. Participants are expected to share signals in accordance with applicable laws such as data protection regulations (e.g., the EU General Data Protection Regulation).

Contributions to the Lantern Program are exclusively conducted by participants. The Tech Coalition serves as the "lead party" in the program with responsibility for overseeing compliance with the terms of the Lantern Program, and reviewing the performance of participants, reviewing the quality and accuracy of signals and collating metrics related to the effectiveness of the Program.

Participants are required to adopt a range of commitments with respect to the Lantern Program; namely that they will:

1. Align with the Lantern Program eligibility requirements listed above,
2. Commit to quality assurance by manually reviewing all signals shared by other participants to establish precision before taking action on them,
3. Refuse contributions to the Lantern Program from external sources such as government agencies and disclose any request or demand for intervention in the Lantern Program received,
4. Support Tech Coalition's transparency efforts by providing metrics and feedback when requested. Participants are also expected to publish their own transparency reports.
5. Comply with applicable data protection and privacy laws.

In the case of a violation of the Lantern Program requirements or participant commitments, the Tech Coalition provides timely notice of the violation to relevant participants and collaborates with participants to correct the violative action. The Tech Coalition is also empowered to remove participants from the Lantern Program for violations of the terms of the multiparty agreement, move the Lantern Program off the ThreatExchange platform to a different host platform, or terminate the multiparty agreement.

4.4 Ecosystem

OCSEA perpetrators are increasingly operating across multiple platforms as part of a tactic to conduct acts without being caught. Given this cross-platform nature of the crime, the Lantern Programs seeks to address the need for industry collaboration to fight OCSEA by allowing tech companies to share signals related to harmful threats. There are existing collaborative efforts that allow companies to share hashes of known CSAM and known keywords to support more effective identification of OCSEA across platforms. These include the National Center for Missing and Exploited Children (NCMEC)'s hash-sharing database, the Internet Watch Foundation, Project Arachnid, and the Thorn / Tech Coalition Keyword Hub.[8]

The Lantern Program has been created to supplement these existing efforts, not to replace them. Participants are encouraged to continue submitting to other industry hash-sharing databases as applicable. Different from existing initiatives, the Lantern Program will allow information sharing related to not only known CSAM, but also to threat vectors that may indicate imminent abuse or exploitation (such as grooming behaviors), thereby allowing participants to proactively prevent the occurrence of OCSEA on their platforms.

The Human Rights Impacts That the Lantern Program Seeks to Address

The Lantern Program seeks to enhance the capacity of the technology industry to combat OCSEA, and thereby address one of the most egregious known harms that exists in the industry.

The Lantern Program helps companies fulfill their responsibility to address adverse human rights impacts with which they are already associated[(1)] and will significantly increase their leverage to do so effectively.[(2)]

This HRIA identifies the human rights risks associated with the Lantern Program and makes recommendations for how those risks should be addressed. However, it is essential to emphasize the severity of the existing adverse human rights impacts that the Lantern Program is addressing related to OCSEA.

Severity is defined by three criteria: *scope*, *scale*, and *remediability*.

- **Scope**: OCSEA impacts a large and expanding population of children globally. In 2022, the National Center for Missing and Exploited Children (NCMEC) received over 32 million reports containing approximately 88 million suspected child sexual exploitation images and videos.[9] There has been an 82% rise in online grooming crimes against children in the last five years,[10] and reports of online sextortion have almost tripled in 2023.[11]
- **Scale**: OCSEA often results in serious adverse physical, emotional, and psychological impacts on victims and survivors. For example, studies show that OCSEA is linked with higher risk of self-harm and suicidal ideation.[12] Such direct physical and psychological harm is likely to be lasting, and may continue even after the incident of abuse has ended. In a survey conducted with survivors of OCSEA, 70% of respondents reported constant worry about being recognized through images of their abuse, and reported feelings of continuing abuse due to the existence of the sexual abuse material online.[13]

[8] Harnessing the Power of Industry Collaboration: Tech Coalition 2022 Annual Report.
[9] 2022 Annual Report: NCMEC.
[10] 82% Rise in Online Grooming Crimes against Children in the Last 5 Years: NSPCC.
[11] Sexual Extortion and Child Abuse Reports Almost Triple: Australian eSafety Commissioner.
[12] Disrupting Harm: End Violence Against Children.
[13] Survivors' Survey Executive Summary: Canadian Centre for Child Protection.

- **Remediability**: While OCSEA content can be removed from platforms, it may still be in circulation online, making remediability very unlikely. The lack of reporting by children[14] makes it more difficult to identify and remediate harms. Furthermore, impacts on children are often difficult to remediate as children are a vulnerable group "by reason of [their] physical and mental immaturity."[15]

Along with severity, the likelihood of impacts is considered to assess saliency.

- **Likelihood**: Reports of online enticement increased by 82% in 2022, with an increase in emerging forms of online victimization such as financial sextortion. Furthermore, in 2020 over 3 million accounts were registered across the 10 most harmful child sexual abuse sites on the dark web.[16]

(1) Principle 11 of the UNGPs states that companies should address adverse human rights impacts with which they are involved.

(2) Principle 19 of the UNGPs states that companies should seek to use and increase their leverage to prevent or mitigate adverse impacts, such as via capacity-building or collaborating with other actors.

[14] Disrupting Harm: End Violence Against Children.

[15] Geneva Declaration on the Rights of the Child.

[16] Global Threat Assessment 2021: WeProtect Global Alliance.

5

Key Observations

This section provides high-level insights and observations that arose during the HRIA and that influenced BSR's analysis of human rights impacts and appropriate actions to address them. These are presented in four categories:

6. The Lantern Program and OCSEA
7. Human Rights Risk Factors
8. Participation and Participant Engagement
9. Governance of the Lantern Program

Many of these observations highlight issues, challenges, and dilemmas that exist for both the Tech Coalition and the technology industry more broadly, and provide context and background for the HRIA.

5.1 The Lantern Program and OCSEA

- **The Lantern Program has the potential to fill a key gap in the approach to child safety and protection.** Currently, tech companies primarily address OCSEA risks via individual actions, including[17]: (1) through their content moderation efforts, including the use of hash-based detection tools and machine learning classifiers to detect CSAM; (2) implementing safety interventions, such as age verification and prevention / deterrence messaging targeting potential victims and those seeking to do harm; and (3) once OCSEA cases are identified, by conducting threat investigations and reporting to relevant authorities as part of their legal obligations in different jurisdictions (e.g., to NCMEC in the US).

  However, research suggests that OCSEA perpetrators are increasingly operating across multiple platforms in an effort to avoid detection. For example, after meeting a child on a social media platform or online forum, a perpetrator may ask the child to move their conversation to a private or encrypted messaging service.[18] This cross-platform nature of the crime necessitates an additional and collaborative approach to the industry's existing content moderation efforts. The Lantern Program aims to address this critical need and strengthen existing efforts to fight OCSEA.

- **The Lantern Program can help companies address some of the most severe adverse human**

[17] Harnessing the Power of Industry Collaboration: Tech Coalition 2022 Annual Report.
[18] Online Grooming: Examining Risky Encounters Amid Everyday Digital Socialization.

rights impacts associated with their platforms. The Program was created in recognition of the cross-platform nature of OCSEA and an acknowledged need for effective cross-platform collaboration to strengthen individual company efforts. The existence of the Lantern Program is an expression of companies seeking to address some of the most severe adverse human rights impacts associated with their platforms, helping to protect the rights of some of their most vulnerable rightsholders. The program has significant potential to enhance company leverage to address risks related to OCSEA, such as by providing more information about the risks they are addressing and enhancing overall industry capability.

- **Insights gleaned through the Lantern Program can be valuable for broader efforts to fight OCSEA and protect digital rights.** By enabling the identification and dissemination of OCSEA-related trends and insights, the Lantern Program could be a valuable resource to stakeholders (e.g., civil society, companies, policymakers, academics) working to combat OCSEA globally. Trends in OCSEA are constantly changing, and stakeholders may find it challenging to stay up to speed on all the ways the crime is shifting and evolving. Aggregated insights gleaned through the Lantern Program can be valuable in informing stakeholder efforts.
- **OCSEA implicates multiple industries, creating opportunities for cross-industry collaboration to address the crime.** OCSEA behaviors such as child sex trafficking and sextortion are often connected to illicit financial flows and dependent on services from multiple industries. This raises concerns that the technology industry may not be able to effectively identify and address these crimes by itself, and cross-sector collaboration may be needed. The financial services and travel / hospitality industries in particular may have signals that can help improve insights and address crime. While the Lantern Program is currently limited to participation by technology companies, it may be valuable to consider the benefits and risks of potential cross-sector collaboration for signal sharing.
- **The Lantern Program is launching at a time when OCSEA-related regulations are on the rise.** The UK's Online Safety Bill, the EU's proposed Child Sexual Abuse (CSA) Regulation, and the proposed Kids Online Safety Act in the US are examples of regulations that require tech companies to proactively detect and report risks related to OCSEA on their platforms. While these regulations are seen as important efforts toward better child protection online, they have been criticized by digital rights groups for their potential to be misused by governments to restrict users' privacy and freedom of expression online,[19] and for potential harmful impacts on vulnerable populations, such as LGBTQIA+ youth.[20] In this context, the Lantern Program may be affected by the politicized and contested nature of such regulatory developments and the debate that surrounds them. For example, the political context may influence participating companies' content moderation practices and lead to biased signals and scope creep in the Lantern database.

[19] European Commission's Online CSAM Proposal Fails to Find Right Solutions to Tackle Child Sexual Abuse: European Digital Rights network (EDRi).

[20] Congress Is Pushing an Online Safety Bill Supported by Anti-LGBTQIA+ Group: Vice Digital.

5.2 Human Rights Risk Factors

- **Signal sharing between companies may exacerbate certain human rights risks.** When individual companies' fight OCSEA alone they are faced with risks to other human rights, such as those related to freedom of expression (e.g., through over-moderation of content) or privacy (e.g., through overbroad data sharing or monitoring). The severity and likelihood of these risks may increase with the cross-platform nature of the Lantern Program. For example, signal sharing may increase the likelihood of erroneous content removal or blocking of user accounts because signals are further decontextualized when shared across platforms (e.g., a company flags all text exchanges with a certain keyword and uploads them to the database; these signals are then interpreted differently by the other participants as they are not able to determine the context of communications that took place on another platform). This risk of exacerbating human rights impacts is more likely for content and behaviors that have less clear definitional boundaries, such as grooming and sextortion, rather than CSAM, which has clearer definitions.

- T**here may be cumulative impacts associated with cross-platform signal sharing.** The human rights impacts of company action to fight OCSEA can be assessed at two different levels: (1) individual company-level impacts, and (2) cumulative impacts arising from, or exacerbated by, the actions taken by more than one company. Due to its cross-platform nature, the Lantern Program may lead to multiple companies taking action on signals in a way that results in compounded adverse impacts on human rights. For example, if multiple companies take action on unsubstantiated signals, users may be wrongly denied access to online services across multiple platforms.

  The Tech Coalition will seek to mitigate this risk by requiring companies to make their own independent decisions in accordance with their own policies and understanding of their legal obligations. However, there remains a risk that participants may ignore this requirement and action signals based on other companies' decisions without sufficient due diligence on their own part.

- **Human rights risks may arise from unsubstantiated signals.** Companies may mistakenly identify users or content as potentially harmful and share related "unsubstantiated signals" on the Lantern Program database. Further, companies may take subsequent action to erroneously remove content or shut down a user's account based on unsubstantiated signals. While the risk of taking action on unsubstantiated signals already exists within individual company practices, its likelihood may increase with cross-platform signal sharing because signals are further decontextualized when shared across platforms. Similarly, the resulting human rights impacts may be more severe when multiple companies take action—for example, if users are wrongly denied access to online services across multiple platforms.

- **Government requests or involvement with the Lantern Program may be associated with human rights risk.** Although the Tech Coalition is not planning to directly engage with governments, governments or law enforcement agencies may gain access to signals in the database by engaging with individual companies. Participating companies often have direct interaction with law enforcement agencies and may receive legal requests, or may otherwise be encouraged or coerced into sharing intelligence or signals, depending on the level of control governments exercise over technology companies in different countries. This may impact users' right to privacy and lead to surveillance and threats to bodily security or due process.

  Governments or law enforcement agencies may also make requests to the Tech Coalition or participants of the Lantern Program to share user data. Where such requests are overly broad or not legitimate (e.g., requests for data with respect to threat vectors that do not rise to the level of criminal conduct under applicable laws), they may impact users' right to privacy. The likelihood of this risk may increase with the Lantern Program, as a centralized and collaborative database with user information may draw attention from law enforcement agencies around the world.

- **OCSEA-related behaviors with less clear definitions create challenges for signal sharing.** While CSAM is relatively easy to define (and there are industry efforts to harmonize the definition and classification of different types of CSAM, such as INHOPE's Universal Classification Schema), other behaviors, such as grooming, do not have standardized definitions. Grooming is a poorly defined "gray area" in most legal frameworks,[21] and attempts to define grooming have proven challenging due to cultural differences and the increasingly politicized nature of the term. For example, members of the LGBTQIA+ community who discuss issues related to sexual orientation or gender identity have been accused of grooming children online in some jurisdictions.[22]

  The age of consent (or "the minimum age at which a person is considered legally competent to consent to sexual acts") also varies by jurisdiction. Jurisdictions have different cut-off points for what constitutes a "minor" who can provide legal consent, ranging from 11 to 21.[23] As a result, determining boundaries or guardrails for signal sharing based on age can be challenging.

- **Children's interaction with the digital environment and associated impacts on their digital rights are constantly evolving.** For example, children's experiences with and attitudes toward self-generated explicit imagery are shifting rapidly[24] and there is a lack of in-depth research in the field on how such online behaviors may affect children's development and digital rights. As a result, the impacts of efforts to address this behavior may be difficult to ascertain, and it may be challenging for tech platforms to establish content policies that proportionately address all risks and empower children to enjoy the full spectrum of their rights in ways that are appropriate for their development.

- **The lack of clear definitions increases the risk of scope creep.** The Tech Coalition draws a clear line that signals shared in the Lantern Program must be for the purpose of combating OCSEA. However, given that some OCSEA behaviors do not have universally accepted and clear definitions (e.g., grooming), signals outside of the intended scope may also be shared in the database. This may happen as a result of government pressure related to child safety and protection more broadly, general political environment and regulations, or biases that exist within companies' content governance structures. For example, companies in the US may institute policies banning content related to transgender health in the name of child safety, as a result of the changing political environment and regulatory proposals such as the Kids Online Safety Act. Subsequently, signals related to transgender health content may be shared in the Lantern database, leading to more platforms taking action on such content.

- **Companies have different policies and enforcement processes against OCSEA, complicating the effectiveness of signal sharing.** While the industry has largely standardized approaches to handling certain OCSEA cases, such as CSAM, there are "edge cases" that companies handle differently and "gray areas" that the industry has not yet decided how to handle.

  Edge cases include legal but harmful content, such as grooming "manuals" (i.e., materials that are not explicitly CSAM, but that may be related to the production of CSAM). While some companies have policies that cover this type of content, most companies only have illegal content policies and/or CSAM policies, neither of which would cover this edge case.

[21] Online Grooming of Children for Sexual Purposes: Model Legislation & Global Review.

[22] Accusations of Grooming Are the Latest Political Attack — with Homophobic Origins: NPR.

[23] Age of Consent by Country 2023: World Population Review.

[24] New Thorn Research Monitors Evolution of Youth Attitudes and Experiences with SG-CSAM.

Gray areas include increasingly prevalent OCSEA-related behaviors such as self-generated explicit imagery and peer-to-peer offenses. These cases don't have agreed-upon principles or guidelines, and the field has not yet decided how to handle them. For example, some companies choose to deprioritize self-generated explicit imagery when they enforce their content policies, while others treat it the same as third-party CSAM.

The lack of standardized policies and enforcement approaches may lead to inconsistencies in the sharing and actioning of signals, limiting the effectiveness of the program. For example, signals related to edge cases and gray areas may not be uploaded to the Lantern Program database or they may go unaddressed even if they are uploaded by another company.

Further, since the field has not yet established the best way to address gray areas, the use of the Lantern Program may increase the likelihood of human rights risk. For example, if companies share signals related to self-generated explicit imagery, there may be an increase in action taken against minors across platforms for distributing self-generated images, which may have adverse impacts on them, especially if they are criminalized.

- **Age determination, assurance, and verification are key challenges that may lead to increased risk for children.** Age determination continues to be one of the biggest challenges in the OCSEA field. Determining the age of an individual in a sexual, explicit, or abusive content is challenging because adolescent children can be difficult to distinguish from adults. This increases the likelihood that the industry underestimates the volume of child sexual abuse material because some images of children are classified as adults. Furthermore, it should be noted that some models currently used by tech companies have lower accuracy rates on children of color compared to white children, and children of color are more likely to be flagged as older than they actually are, which may mean the image is not flagged as CSAM.[25]

  Conversely, attempts to verify the age of users, particularly children, also come with risks, including potential adverse privacy impacts and inaccuracies based on race, gender, ethnicity, or culture. Currently, companies conduct age assurance by using biometric models or asking users to upload identification documents—both of which may lead to risks on privacy and data protection. Along with age determination, age assurance and verification are also major technical challenges for the field.

  Although the Lantern Program discourages participants from sharing personal information about children (e.g., in cases where a child is the offender), challenges around age assurance and verification may limit the ability of companies to effectively adhere to this rule.

- **Computer-generated CSAM brings new challenges to moderating OCSEA.** Developments in diffusion models and generative AI technologies have enabled increased creation and distribution of synthetic CSAM, and it is proving increasingly challenging for the field to effectively distinguish between genuine or photorealistic computer-generated materials. Although computer-generated CSAM currently constitutes a small portion of online CSAM, its prevalence may increase rapidly.

  At the time of writing, tech companies do not have a clearly defined approach to moderating computer-generated CSAM. One important factor when prioritizing action is determining whether a real-world victim is depicted in the materials or not. Companies may take different actions when moderating computer-generated CSAM if it does not involve real victims because their priority is to address real-world harm, though this is increasingly difficult to distinguish. The proliferation of computer-generated CSAM may not only increase the scale and speed of OCSEA, but it may also lead to inconsistencies and challenges in content moderation and signal sharing.

- **Signal sharing may complicate the user appeals process and raise barriers to access remedy.**

[25] Stakeholder interview.

The Tech Coalition requires each participating company to have a user appeals process in place to allow users to object when they believe their content was erroneously removed or their account was unfairly shut down. The signal sharing program may lead to complications that impact the effectiveness of user appeals processes when multiple companies act upon a signal. For example, if multiple companies independently decide to deactivate the account of an individual because of an unsubstantiated signal shared in the database, that individual will need to appeal to each company separately, placing the onus on the user and creating a higher barrier to access remedy.

5.3 Participation and Participant Engagement

- **Participant engagement with the Lantern Program varies significantly.** Some participants use or plan to use the Lantern database more actively than others, and some participants are or plan to be more diligent than others in uploading and actioning signals. These differences are often based on resourcing constraints, such as limited content moderation capacity, or philosophical differences between companies' content governance approaches.

  The Tech Coalition expects all participants to conduct human review and verification of signals shared via the Lantern database and to identify violations of terms and conditions or content policies on their own platforms before carrying out actions against identified users or accounts. The Tech Coalition has taken measures to ensure that participants comply with this requirement, but differences in participants' content moderation capacity and approach may lead some participants to implement this requirement less diligently than others.

- **The Lantern Program can serve as a key resource for companies with capacity constraints.** Smaller technology companies often don't have the technical expertise or capacity to identify and investigate all OCSEA risks or harms occurring on their platform, whereas large tech companies are more likely to have well resourced, dedicated teams and sophisticated approaches. This difference can result in bad actors gravitating toward smaller company platforms, including those outside the US and EU, increasing the likelihood of harm because these companies may not be able to effectively moderate their platforms. The Lantern Program can provide an avenue of collaboration and support to smaller companies by helping them prioritize content and accounts that are more likely to require immediate action, and by facilitating cross-industry knowledge sharing.

- **Companies need dedicated resources to effectively utilize the Lantern Program.** Signals shared on the Lantern Program database should only be actioned by companies after sufficient due diligence is conducted to ensure that a signal is associated with a violation of the platform's own terms of service. While signals can help companies prioritize content for review, a company still needs to have their own content moderation capacity to be able to action the signals.

  Similarly, before uploading a signal to the Lantern Program database, companies need to ensure that signals are accurate, which also requires resource capacity. Without sufficient resourcing, participating companies may share and/or action signals without review, which would increase the likelihood of unsubstantiated signals and adverse human rights impacts.

- **Increasing the number of participants will increase the positive human rights impact of the Lantern Program, but it may also increase the likelihood of human rights risks.** The Tech Coalition has established a high bar for entry into the Lantern Program; requirements for participating in the program are more stringent than those that qualify companies for Tech

Coalition membership (e.g., participants are required to have content policies and user appeals processes). Companies that are able to satisfy these requirements are naturally larger companies mostly based in the US or Europe.

Growing the participant base and including smaller companies from outside the US and EU will increase the positive impact of the Lantern Program to fight OCSEA. However, including these companies may also increase human rights risk. The Tech Coalition will need to strike the right balance in expanding participation to increase the impact of the Lantern Program while addressing the risks that may be associated with more participants.

- **Over-dependence on the Lantern Program may weaken individual company efforts to fight OCSEA.** There is a risk that less well-resourced or smaller organizations participating in the Lantern Program over-rely on the program as their sole child sexual abuse or exploitation moderation tool and neglect individual efforts or investments in platform moderation. This could ultimately lead to failure by companies to identify and address OCSEA instances unique to their own platforms, outside of the Lantern Program.
- **Quality concerns and lack of context about signals may decrease participant engagement.** The Tech Coalition has set clear parameters regarding the types of signals that can be uploaded to the database. However, because the Lantern Program has not fully come into effect, participants have alluded to the fact that they are not yet confident about the quality of the signals. Similarly, some participants noted that signal metadata shared by other participants often lacks the appropriate context needed to understand, verify, or trust the signal and conduct investigations based on it, which limits the actionability of a signal. For instance, companies may not be comfortable acting upon an upload of a sexualized image of an older minor without accompanying evidence of the age of the victim.

  Concerns about the quality of signals, or a lack of labeling and context, may lead to a decline in participant engagement. For example, if participants are concerned that there might be unsubstantiated signals, this may disincentivize them from engaging with or utilizing the database. It will be important for the Tech Coalition to establish robust mechanisms for quality control.
- **Legal frameworks may complicate the Lantern Program's ethos of voluntary action.** Although the Lantern Program is a voluntary program under which participants are not obligated to carry out any moderation activities, the overarching legal frameworks in the various jurisdictions where company participants operate may create pressure on participants to take action on all signals received. In particular, legal or policy frameworks that require the proactive detection and removal of child sexual abuse or exploitation materials by online platforms may either deter participants from fully engaging with the Lantern Program or lead them to automatically action signals ingested, with both scenarios arising out of fear for the legal consequences of not acting upon signals (i.e., that the company has violated its duty of care, or similar).

  Similarly, concerns related to privacy laws may also deter participants from engaging with the program. Though the Tech Coalition requires participants to have an appropriate privacy policy and comply with applicable privacy laws, some participants noted that sharing personal data about users may conflict with their company's existing privacy policies. As a result, some legal teams advise against engaging with the Lantern Program.

5.4 Governance of the Lantern Program

- **Accountability and responsibility related to the Lantern Program is shared between the Tech Coalition and participating companies.** The Tech Coalition clearly states that each participating company is responsible for how they share and use signals as part of the program, and each participant is individually accountable for risks associated with their use of the signals. The Tech Coalition, on the other hand, is responsible for putting in place appropriate measures and mitigations to incentivize and enable participants to use the Lantern Program appropriately, and to identify and act upon instances when they do not. This structure of shared responsibility has implications for the Tech Coalition's attribution to human rights impacts and its leverage to address those impacts, which is discussed in the Human Rights Impacts section (Section 6, below) in more detail.

- **The use of Meta's ThreatExchange platform may have implications for the governance and autonomy of the Lantern Program.** Meta serves as an infrastructure provider to the Tech Coalition and makes adjustments to the signal sharing database and its technical features because the Lantern Program is built on Meta's ThreatExchange platform. This may take away some independence and control from the Tech Coalition, and may lead to challenges if the Tech Coalition needs more resources than Meta can provide.

- **Transparency is a critical component of the Tech Coalition's governance approach for the Lantern Program, and it constitutes an important risk mitigation measure in itself.** If the Tech Coalition and participants of the Lantern Program fail to provide adequate information about the use of the database, the lack of transparency may lower the effectiveness of the program. For example, companies may upload or action signals based on biased assumptions and this may go unchecked. This would not only limit the impact of signal sharing, but may also lead to stakeholder skepticism about the program.

  Transparency may also come with risks. For example, transparency about the Lantern Program may expose participating companies to legal scrutiny and disincentivize them from engaging with the database (see above). Similarly, it may attract governments and law enforcement to coerce the Tech Coalition or participating companies to share user information. The Tech Coalition will need to continuously consider these risks and strike a balance between openness and potential risks.

- **External communications about the Lantern Program will be critical to the success of the program.** The ongoing public policy debate about the relationship between child safety, privacy, and other rights necessitates that the Tech Coalition and companies participating in the Lantern Program carefully consider the external communications around the program. The program may be perceived to skew on either side of this debate if messaging and communications are not undertaken in a balanced and nuanced manner. In turn, perception issues may disincentivize companies from engaging with the Lantern Program, create distrust among certain stakeholder groups, and ultimately limit its impact.

6

Human Rights Impacts of the Lantern Program

In this section we assess the potential human rights impacts associated with the Lantern Program and use the criteria described in the Methodology section to prioritize salient human rights issues.

6.1 Impacted Human Rights

Below we list eight categories of human rights that may be impacted. Some of these categories have multiple human rights grouped together. Child Safety and Protection (Articles 19, 34, 36 Convention on the Rights of the Child): The Lantern Program seeks to promote the enjoyment of children's right to protection from sexual abuse and exploitation by establishing effective procedures to identify, prevent, and mitigate threats to children's safety on digital platforms.

- **Civil, Social, Cultural Rights and Freedoms of Children (Articles 12, 13, 14, 15, 16, 17, 24, 27, 28, 31, Convention on the Rights of the Child)[26]:** While the main purpose of the Lantern Program is to protect children against online sexual exploitation and abuse, there is a risk that the collection, use, and sharing of information in the Lantern Program, as well as actions taken by participants based on signals, may adversely impact the ability of children to enjoy their civil rights and freedoms, such as access to information, privacy, freedom of expression, freedom of thought, or participation in cultural life.
- **Privacy (Article 12, Universal Declaration of Human Rights; Article 17, International Covenant on Civil and Political Rights):** Signals shared in the Lantern database may include personally identifiable information about users such as email addresses or account names. The program may be associated with risks to privacy if participants use or share data in ways that result in arbitrary interference with users' privacy, family, home, or correspondence.
- **Freedom of Expression (Article 19, Universal Declaration of Human Rights; Article 19, International Covenant on Civil and Political Rights):** Although the Lantern Program has been set up to target content or speech related to child sexual exploitation or abuse, there is a risk that legitimate or non-violative content may be erroneously removed due to overbroad moderation by participants, and/or that users are wrongfully denied access to online platforms where they can exercise their right to free expression and to access information.

[26] Note that these are the rights of children specifically listed in the Convention of the Rights of the Child. However, children are also considered rightsholders under the International Bill of Human Rights, so all of the other rights listed in this section would apply to children as well as adults.

- **Equality and Nondiscrimination (Articles 1, 2, 7, Universal Declaration of Human Rights; Articles 3, 26, International Covenant on Civil and Political Rights; Articles 2, 3, International Covenant on Economic, Social, and Cultural Rights):** Certain groups or communities may be at greater risk of being wrongly accused of OCSEA and be subject to investigations or punitive actions, due to societal biases (e.g., against LGBTQIA+ people and sex workers), or content moderation tools may not perform as accurately for certain groups (e.g., people of color) and languages. The Lantern Program can proactively address these biases and discriminatory practices by setting guardrails and sharing insights with the broader field.
- **Due Process and Effective Remedy (Article 8, 9, 10, Universal Declaration on Human Rights; Articles 2, 14 International Covenant on Civil and Political Rights):** When participants take action based on signals shared in the Lantern Program database, this may result in content being removed and user accounts being deleted or flagged for monitoring. In such cases, users may be penalized without notice for conduct or actions taken outside of a platform, and they may not be provided effective or accessible avenues for appeals or complaints mechanisms.
- **Bodily Security (Articles 3, 5, 9, Universal Declaration of Human Rights; Articles 6, 7, 9, International Covenant on Civil and Political Rights):** There is a risk that actions taken on the basis of signals shared in the Lantern Program may wrongfully result in offline harms to users such as arbitrary arrest, detention, or investigation of users, particularly if signals are shared with and misused by governments or law enforcement agencies.
- **Economic, Social, and Cultural Rights (Articles 23, 25, 26, 27, Universal Declaration of Human Rights; Articles 7, 11, 13, 15, International Covenant on Economic, Social and Cultural Rights):** Digital platforms have become an integral part of social, cultural, and economic life including the facilitation of work and education, and participation in cultural life. The Lantern Program may adversely affect users' ability to enjoy these rights if users are wrongfully denied access to platforms as a result of signals shared in the database.

There are three important points to note about the prioritization of these potential adverse human rights impacts:

- The eight salient human rights categories listed here are a narrowed list of potential salient risks that cross a threshold of relevance for the Tech Coalition. In other words, there are other potential adverse human rights impacts that are already excluded from this list of salient human rights issues.
- BSR has reviewed these eight salient human rights issues using the criteria of severity (scope, scale, remediability) and likelihood described in the Methodology section. However, BSR notes that (1) prioritization is more directional than precise and (2) all human rights are indivisible, interdependent, and interrelated, and the deprivation of one right adversely affects others. In other words, the connections between these rights can be as important as their relative salience. For example, all the rights listed are child rights (even though child rights are also listed as a salient issue), and violation of freedom of expression might impact the right to education.
- According to the UNGPs, the Tech Coalition has a responsibility to address all adverse human rights impacts. Prioritization on the basis of severity is encouraged when it is not possible to address all impacts simultaneously, but it does not remove the Tech Coalition's responsibility to address all adverse human rights impacts.

6.2 Description of Impacts

There are multiple ways in which the Lantern Program may impact the eight human rights listed above. In BSR's analysis, the main pathways that can lead to human rights impacts are:

- **Data Collection, Storage, and Sharing:** The collection, storage, and sharing of data as part of the signal sharing program may be associated with adverse impacts on users' rights, including the right to privacy.
- **Actioning of Signals:** The various actions participants may take on the basis of signals in the Lantern database, such as wrongful account removals, over-moderation of content, or noncompliance with Lantern Program guidelines may adversely impact users' rights.
- **Government Involvement:** Governments and law enforcement agencies may attempt to gain direct or indirect access to the Lantern database, or influence participants' use of the Lantern Program in ways that adversely impact users' rights.
- **Unintended Consequences on Children:** Although the Lantern Program exists to promote children's safety and protection rights, the use of the database may have unintended consequences on children, adversely impacting their rights.

In the tables below, we describe the specific human rights impacts (both risks and opportunities) related to each of these pathways.

The Recommendations section of this assessment provides advice on how potential impacts may be avoided, prevented, mitigated, or remediated. We do not list recommendations alongside each risk because, in our experience, one recommendation may address multiple risks at the same time.

Impacts Related to Data Collection, Storage, and Sharing

The collection, storage, and sharing of data as part of the signal sharing program may be associated with adverse impacts on users' rights, including the right to privacy.

Potential Adverse Impacts / Risks

- The Lantern Program may lead to increased data collection and storage by participants, including increased amounts of contextual data to support the signals in the database, which may be associated with adverse impacts on the privacy and data protection rights of users. For instance, larger amounts of personal communications may be collected and shared to provide context to make signals more useful or actionable by other participants.
- Participants in the Lantern Program may use personal data contained in signals in ways that adversely affect the privacy and data protection rights of users.
- Participants in the Lantern Program may collect, store, and share data that is not necessary or proportionate for the detection of online child sexual exploitation and abuse. This may include signals outside the defined scope of OCSEA (i.e., scope creep).
- Unauthorized persons may obtain access to personal data contained in the Lantern Program database as a result of a breach, hack, or unauthorized access.

Potential Opportunities

- The Lantern Program may provide insights into the presence and evolution of risks to children on online platforms, including new threat vectors, emerging types of OCSEA, etc., that may enable participants to develop early intervention processes that minimize the amount of data they collect or store for harm-detection purposes; or reduce their reliance on content moderation to identify potential threats (i.e., as a result of the availability of metadata in the Lantern Program database).

Impact Factors

- **Severity**: The scope of users that could be affected by potential adverse impacts arising from data collection, storage, and sharing associated with the Lantern Program is large, with the scale of impact ranging from least to most serious. Unauthorized disclosure of personal data contained in the Lantern database could, for instance, be associated with threats to the life, liberty, or security of affected users in some jurisdictions if individuals take "vigilante" action against users, or law enforcement agencies unduly detain affected users. Unauthorized access to the Lantern database may be difficult to reverse if private information has already been disclosed, so impacts may not be remediable.
- **Likelihood**: Adverse impacts on users' right to privacy may be more likely with the Lantern Program compared to individual company processes due to increased data sharing. Increases in the amount of data collected and shared by participants may provide greater opportunities for unauthorized access to data or lead to disproportionate and/or inappropriate storage of personal data. The adverse impacts arising from the data practices of participants may be effectively mitigated by adequate data protection and privacy safeguards.
- **Attribution**: Some privacy risks may arise as a result of the data practices of individual participants, while others may be associated with the cross-platform nature of the Lantern Program. For example, data security risks may arise as a result of the transfer or sharing of data that occurs as part of the Lantern Program. Other risks such as unauthorized access to the program as a result of hacks or breaches may be directly connected to the data security infrastructure of the Lantern Program. Data related risks may therefore be affected by, or connected to, the data protection and privacy safeguards implemented by both the Tech Coalition and participants.
- **Leverage**: The Tech Coalition's position as lead party in the Lantern Program enables it to provide guidance to participants on signal sharing and data practices associated with the program and review compliance with applicable privacy laws and recommended practices. However, the Tech Coalition's leverage may be limited by the fact that privacy or data protection risks may arise on participants' individual platforms or on ThreatExchange.

Impacts Related to Actioning of Signals

The various actions participants may take on the basis of signals in the Lantern database, such as wrongful account removals, over-moderation of content, or noncompliance with Lantern Program guidelines, may adversely impact users' rights.

Potential Adverse Impacts / Risks

- Actions taken against content based on signals shared in the Lantern database may adversely impact users' freedom of expression. This may happen in a number of ways:
  - › Participants may erroneously remove content based on unsubstantiated signals, or
  - › Participants may over-moderate and remove content that does not violate their content policies based on signals that they have not independently verified.
- Actions taken against users based on signals shared in the Lantern database may adversely impact users' freedom of expression, access to information, and bodily security, as well as their economic, cultural, and social rights. This may happen in a number of ways:
  - › Participants may apply penalties or restrictions on users, such as warning strikes or removal of accounts, based on unsubstantiated signals, or
  - › Participants may take action against users for actions committed on other platforms based on signals that they have not independently verified.
- Actions taken by participants to identify and prevent OCSEA may be associated with bias or discrimination against certain populations due to content moderation tools underperforming for non-Western linguistic communities or underrepresented populations, or due to societal biases (e.g., against LGBTQIA+ people or sex workers). Such biases may be present in signals shared in the Lantern database, which may lead to a higher likelihood of unsubstantiated signals associated with these groups, and result in a disproportionate amount of wrongful actions taken against them.
- Signals outside the defined scope of OCSEA may be shared in the Lantern database (i.e., scope creep). Actioning of such signals by participants may lead to wrongful actions or actions that are not necessary and proportionate for addressing OCSEA, which may be associated with adverse impacts on users' right to privacy, freedom of expression, and other rights.
- Following erroneous actions, participants may fail to provide effective remedy to users. For example:
  - › Participants may fail to adequately provide explanations to users for actions taken on the basis of signals shared in the Lantern database that enable users to exercise rights of appeal, or
  - › Users may not be provided appropriate mechanisms to appeal actions taken on the basis of signals shared in the Lantern database, or
  - › Participants may fail to provide appropriate remedy for harm suffered as a result of erroneous actions, or institute remedial steps (e.g., account reinstitution or deletion of an erroneous signal from the Lantern database) in a timely manner to effectively remedy the harm.

Potential Opportunities

- Signals shared in the Lantern database may enable participants to more accurately identify exploitation or abuse of children on their platforms, leading to fewer erroneous actions.
- Information deduced from Lantern Program signals may improve participants' OCSEA moderation efforts. Insights about emerging trends, patterns of conduct, key markers that reliably indicate potential or actual child sexual abuse or exploitation and gray areas such as self-generated explicit imagery or minor-to-minor interactions may enable participants to improve their manual and automated moderation processes and further advance best practices in the field.

Impact Factors

- **Severity**: The scope of users that could be affected by potential adverse impacts arising from the actioning of signals is large, with the scale of impact ranging from serious to potentially life or liberty threatening. For example, users may be unable to access digital platforms necessary for their livelihood, or may be arrested, interrogated, or detained by law enforcement agencies as a result of erroneous actions. Adverse impacts arising from such actions may range from possibly remediable (e.g., via restoration of a user account that was inaccurately removed) to not remediable (e.g., where an arrest or loss of employment has occurred).
- **Likelihood**: The likelihood of adverse impacts related to the actioning of signals may be greater with the Lantern Program compared to individual company practices, and will vary depending on the participant taking action, the type of platform they have, its size, and the company's content moderation resources. Large platforms focused on user-generated content and interactions may have higher volumes of content to moderate and therefore a higher likelihood of unsubstantiated signals and erroneous action. Better resourced companies may be more capable of conducting reviews of the signals shared in the Lantern database, verifying their accuracy before taking action, and mitigating risks in a timely manner. Additionally, the likelihood of unsubstantiated signals and erroneous participant action may be higher for underrepresented languages, dialects, or markets where participants have fewer content moderation resources.
- **Attribution**: Although risks related to erroneous actions already exist within individual companies' OCSEA moderation efforts, these risks may be amplified or exacerbated by the cross-platform nature of the Lantern Program. For instance, impacts may be compounded if users lose access to multiple platforms that serve social or economic purposes as a result of an unsubstantiated signal by one participant. In such cases where the impact is compounded due to signal sharing among platforms, the Tech Coalition may be more closely associated with the harm.
- **Leverage**: The Tech Coalition has developed guidelines and conditions for participation in the Lantern Program that require participants to conduct independent verification of signals and identify violations on their own platform prior to taking any action. However, participants' approach to content moderation and capacity for verification vary, and some participants may take enforcement action against users for violations that occur outside of their platform that may further amplify the potential adverse impacts associated with the actioning of signals.

Impacts Related to Government Involvement

Governments and law enforcement agencies may attempt to gain direct or indirect access to the Lantern database, or influence participants' use of the Lantern Program in ways that adversely impact users' rights.

Potential Adverse Impacts / Risks

- The Tech Coalition may receive pressure from governments and law enforcement agencies to provide access to the Lantern database in ways that contradict its guidelines and adversely affect users' rights to privacy and bodily security.
- Participants of the Lantern Program may, as a result of participation, receive an increased amount of requests from governments or law enforcement agencies to store or share user data in ways that are not necessary nor proportionate for the purposes of addressing OCSEA and that adversely impact users' right to privacy, freedom of expression, and bodily security.
- Participants may receive and comply with overbroad government demands for access to signals in the Lantern database, thereby adversely impacting the rights of users to freedom of expression and bodily security.
- If they get access to data or signals shared in the Lantern database, governments or law enforcement agencies may subsequently use this information to identify, track, or monitor individuals in a manner that adversely impacts their rights to privacy or bodily security, or take punitive steps such as deprivation of liberty against users for conduct that is not legally prohibited.
- Governments or law enforcement agencies may influence participants in ways that lead to scope creep or biases in the identification or definition of signals indicating OCSEA, which may result in unnecessary or biased signals being shared in the Lantern database.

Impact Factors

- **Severity**: The scope of users that could be affected by potential adverse impacts arising from government involvement in the Lantern Program is large, with the scale of impact ranging from serious to potentially life or liberty threatening—for example, when government involvement leads to surveillance and adversely impacts users' right to privacy. Adverse impacts arising from government involvement may range from remediable (e.g., where law enforcement access to the Lantern Program is revoked) to not remediable (e.g., where an individual has been wrongfully arrested or detained).
- **Likelihood**: The likelihood of adverse impacts related to government involvement may be higher in the Lantern Program compared to individual platforms, and will vary depending on the participant engaging with the government, the type of platform they have, the markets they operate in, and the platforms' existing relationships with governments or law enforcement agencies. Platforms with close existing relationships with government or law enforcement agencies may be more likely to receive requests to share data. Companies operating in countries with authoritarian governments or where there are fewer human rights protections may also face a higher likelihood of government involvement associated with their participation in the Lantern Program.
- **Attribution**: Individual participants of the Lantern Program may have existing interactions with government or law enforcement agencies with respect to their content moderation efforts, and they may already receive requests to share user data. However, risks to users may be compounded by the cross-platform nature of the Lantern Program because the program may draw interest from government and law enforcement agencies and lead to demands for the signals and data shared in the Lantern database, creating new risks to users.
- **Leverage**: The Tech Coalition's leverage may be limited by the fact that government involvement in the Lantern Program may occur via individual participants (e.g., law enforcement requests to receive data from or contribute signals to the Lantern database may be received by participants rather than the Tech Coalition). However, leverage exists through reinforcing Lantern Program guidelines and principles, providing clear communication on database access restrictions, and requiring increased transparency from participants with respect to their response to government and law enforcement requests.

Impacts Related to Unintended Consequences on Children

Although the Lantern Program exists to promote children's safety and protection rights, the use of the database may have unintended consequences for children, adversely impacting their rights.

Potential Adverse Impacts / Risks

- Participants may collect, store, and share personal information of children in the Lantern database, adversely impacting children's right to privacy. Technical difficulties with age assurance or verification may result in information about children and victims being included in the database.
- Participants may over-moderate content created or shared by children (i.e., remove content beyond what is prohibited in their terms of service) based on signals shared in the Lantern database in a manner that is not proportionate or necessary for addressing OCSEA, thereby limiting children's rights to freedom of expression.[27]
- Participants may take punitive actions against child users, such as account deletion or restrictions, based on signals shared in the Lantern database, related to OCSEA behaviors in which the child user took part (e.g., self-generated explicit imagery, minor-on-minor interactions). In some cases, child users may be reported and criminalized for such behaviors. Such punitive actions may adversely impact children's rights to education, access to information, and freedom of expression and association, and ultimately may negatively impact their social, cognitive, and emotional development.
- Signals shared in the Lantern database related to self-generated explicit imagery or peer-to-peer interactions may be disproportionately with respect to certain groups of children (e.g., related to gender, racial identity, or sexual orientation), leading to an increased likelihood of punitive action taken against these children, and resulting in the exclusion of these children from online platforms, exacerbating the digital divide.
- Participants may have fewer content moderation resources or underperforming content moderation tools for underrepresented languages, dialects, markets, or populations. As such, participants may fail to identify and share signals of OCSEA for certain populations of children, adversely impacting their protection rights and their right to nondiscrimination.
- In cases where action is taken against them or the content they create or share, children may be unable to enjoy the full range of their civil, social, economic, cultural, or political rights. For example:
  - › Children's ability to participate in online platforms may be hindered (e.g., by account restrictions that remove opportunities to exchange messages or interact with peers, thereby limiting their freedom of expression and the ability to seek, receive, or impart information and ideas).
  - › Lack of access to digital platforms may impede children's social, cognitive, or behavioral development.

[27] The saliency of this risk may change as children's interaction with online platforms, and the field's interpretation of children's freedom of expression online, evolve over time.

- › Children's right to education may be restricted if they are not able to use online platforms for educational purposes.
- › Children's right to health may be restricted if, for example, nude photos shared for medical purposes are erroneously removed.
- › Children may be unduly subject to criminal investigation, arrest, or detention, which may have adverse impacts on their development and mental health.

- Subsequent to erroneous actions, participants may fail to provide effective remedy to children. For example:
  - › Participants may fail to adequately provide child-friendly explanations for enforcement actions taken on the basis of signals shared on the Lantern Program that enable users to exercise rights of appeal.
  - › Children may not be provided easily accessible or appropriate mechanisms to appeal enforcement actions taken on the basis of signals shared on the Lantern Program.
  - › Participants may fail to provide appropriate remedy for harm suffered as a result of erroneous actions, or institute remedial steps (such as account reinstitution or deletion of an erroneous signal from the Lantern Program) in a timely manner to effectively remedy harm to children.

Potential Opportunities

- The Lantern Program may provide insights into challenges associated with balancing the protection rights of children against the enjoyment of their civil rights and freedom. For instance, the Lantern Program may enable participants to develop innovative approaches to address issues of self-generated explicit imagery, minor-to-minor interactions on digital platforms, and age assurance and verification, and may enable participants to further engage with and empower children on digital platforms.

Impact Factors

- **Severity**: The scope of impacts related to unintended consequences on children is small, as children are a smaller population than adults. However, the severity of potential adverse impacts on children are more serious because they are a vulnerable group. The Geneva Declaration on the Rights of the Child notes that as a result of their physical and mental immaturity, children need special safeguards and protection. Children's vulnerability may be heightened as a result of factors such as age (with younger children likely to be more vulnerable than older teenagers), gender or sexual identity, level of education, cognitive abilities, or socioeconomic class. As such, children may be more vulnerable to the adverse impacts identified or may suffer more severe harms from such impacts. For example, the impact of exclusion from digital platforms may be more severe for children than adults. Impacts on children are also less likely to be remediable once they occur.
- **Likelihood**: The adverse impacts outlined above may be likely to occur as a result of challenges associated with age assurance and verification on online platforms, and challenges related to the lack of universal approaches in handling gray areas, such as self-generated explicit imagery. The likelihood of harm may vary across different platforms. For example, some platforms are able to dedicate more resources and technical tools to conduct age verification of users, or some platforms have different content policies based on the nature of the platform (e.g., gaming platform vs. email service) and different enforcement mechanisms when it comes to child users. Furthermore, platforms with a higher number of child users (e.g., social media platforms) or those with large amounts of user-generated content may be more likely to share children's personal data as signals in the Lantern database.
- **Attribution**: Risks associated with the erroneous transfer of children's personal data in the Lantern Program arise directly from participation in the Program. Other risks to children associated with the loss of access to digital platforms may be exacerbated by the cross-platform nature of the Lantern Program, where children are excluded from multiple platforms as a result of signals erroneously shared in the Lantern database.
- **Leverage**: Ensuring the full enjoyment of all of the rights of children on digital platforms remains a challenge for all industry participants due to the need to balance the digital protection and safety of children against their self-actualization and development. Furthermore, the process of addressing gray area issues such as self-generated explicit imagery via processes such as age verification may itself exacerbate existing or create new risks (e.g., data protection and privacy risks). The Tech Coalition may have limited ability to prevent these risks from occurring; but it can provide guidance and insights to participants to help prevent such unintended consequences on children.

6.3 Counterbalancing Human Rights in Tension

Human rights can come into tension with one another for legitimate reasons, and it is important to deploy rights-based methods when two competing rights cannot both be achieved in their entirety. Rather than "offsetting" one right against another, it is important to pursue the fullest expression of both and identify how potential harms can be addressed. Counterbalancing is a methodology that the Tech Coalition can use to navigate human rights trade-offs when making decisions related to the Lantern Program, such as whether to allow a new signal type or implement a new policy.

Counterbalancing can be done using the following international human rights principles:

- **Reverting to principle**—Can the core principle of the restricted right still be upheld in different ways?
- **Legitimacy**—Is there a legitimate aim in pursuing the restriction of this right?
- **Necessity and proportionality**—Is the restriction of the right necessary or can the legitimate goal be achieved through other means? If it is necessary, is it the least intrusive way to restrict this right?
- **Nondiscrimination**—Can the restriction of the right be done in a nondiscriminatory manner?

Because the Lantern Program is a signal sharing program with the goal of combating OCSEA, the rights most likely to be in tension are (1) the right of children to be protected from sexual exploitation and abuse, and (2) the rights to privacy and freedom of expression / self-actualization.

These rights may be in tension when making decisions about the scope of information that can be included in the signal sharing database. For example, the decision to allow or not allow information about minors in the database involves a tension between the right of children to be protected from sexual exploitation and abuse, and the privacy and freedom of expression / self-actualization rights of the minors whose data is involved. UNICEF—the leading authority on child rights from an international human rights perspective—has stated that a child's right to privacy and protecting children from abuse and exploitation must be equally upheld, and that privacy cannot be viewed as secondary. Because neither right can necessarily be privileged over the other, counterbalancing can be used to work through the trade-offs when making this decision.

The Tech Coalition has decided that, at least for the time being, no information about children or teen users may be included in the signal sharing database, with the exception of hashes corresponding to CSAM. However, with the growth in self-generated explicit imagery and minor-on-minor offenses, sharing data about minors is an option that may be discussed in the future. Below we utilize the counterbalancing principles to examine two choices—allowing data about minors vs. not allowing data about minors.

Counterbalancing Exercise: Allowing / Not Allowing Data About Minors in the Lantern Program

- **Reverting to principle**: The core principle behind the Lantern Program's role in protecting children from sexual abuse and exploitation is to better equip online platforms to prevent, detect, and respond to OCSEA, which the program will likely achieve without the inclusion of data about minors. In other words, by not allowing data about minors, the Tech Coalition can likely uphold the principle of the right of children to be protected from sexual abuse and exploitation, even though the right may be somewhat limited in practice. On the other hand, if the Tech Coalition decides to allow data about minors to be shared, this would be a limitation of the right to privacy, and it would be difficult for Lantern to fully uphold the principle of privacy in this case. However, it could still uphold the principles behind free expression / self-actualization of minors through stringent requirements or guidance around data about minors that reduce the likelihood of unsubstantiated signals or disproportionate enforcement of legitimate signals by companies.

- **Legitimacy**: Limiting the privacy of children in order to protect them from sexual abuse and exploitation is a legitimate aim, particularly because most OCSEA practices are nearly universally recognized as crimes. Conversely, protecting the privacy and free expression of minors involved in OCSEA (whether as victims or perpetrators) by not allowing the sharing of data about minors is also a legitimate aim. For example, if unsubstantiated signals about minors are shared and actioned by participants, it could result in the loss of access to online accounts, which can have a significant adverse impact on minor users' right to free expression, access to information, and right to education, and inhibit their ability to participate in daily life.
- **Necessity and proportionality**: It is not clear whether data about minor victims shared across platforms is necessary for enabling platforms to better prevent, detect, and respond to OCSEA, or whether a focus on data about offenders is sufficient. Evidence would be required to make this determination. However, if the Lantern Program were to also target minor-on-minor OCSEA, then it could be considered necessary. If it is not necessary, then data about minors should not be included in the database. If it is necessary, then the Tech Coalition should take steps to protect the privacy and free expression / self-actualization rights of the minors in question as much as possible. For example, the Tech Coalition could implement privacy preserving approaches to data sharing and require rigorous review for signals that involve data about minors to prevent actioning of unsubstantiated signals or disproportionate enforcement responses.
- **Nondiscrimination**: The inclusion of data about minors would in theory apply to all minors equally. However, in practice there may be minors from certain groups that are overrepresented as both victims and perpetrators of OCSEA for a variety of reasons, which could result in unintentionally discriminatory outcomes.

The counterbalancing exercise indicates that Lantern should not allow the sharing of data about minors unless evidence indicates it is necessary to enable companies to prevent, detect, and respond to OCSEA on their platforms. If evidence indicates that it is necessary, then the Tech Coalition should take a variety of steps to limit the potential risks associated with sharing and acting upon data about minors.

6.4 Impacted Rightsholders and Vulnerable Populations

A human rights-based approach requires a clear understanding of which rightsholders are impacted by the Lantern Program, with a particular attention to individuals from groups or populations that may be at heightened risk of vulnerability or marginalization. A clear understanding of which rightsholders and vulnerable groups are impacted by the Lantern Program will inform the Tech Coalition's stakeholder engagement strategy.

BSR identifies the following categories of impacted rightsholders and vulnerable populations that may be impacted through the Lantern Program:

1. **Actual and potential victims of OCSEA:** The goal of the Lantern Program is to help children who are or may be victims of online child sexual abuse and exploitation—making them the primary rightsholder group impacted through the program. Children at risk of harm across multiple platforms are likely to directly benefit from signal sharing and the cross-platform identification and moderation of OCSEA.

Although the Lantern Program helps promote the safety of OCSEA victims, risks associated with signal sharing may have adverse impacts on them and exacerbate some of the harms they already experience. While all children may be victims of OCSEA, some children are at particular risk, including:

- **LGBTQIA+ children** are three times as likely to experience risky interactions online as they seek out and rely on online communities for exploration and perceived safety. They are also more likely to not report unsafe interactions out of fear of being cut off from online interactions completely.[28]
- **Homeless children or children in care** are more vulnerable to attempts by offenders trying to bond with them. According to research, children who are in foster care or involved with the child welfare system make up the majority of child sex trafficking victims.[29]
- **Children with intellectual disabilities** may not be aware of what is appropriate or inappropriate sexual behavior by an adult, making them more vulnerable to sexual abuse and exploitation.[30]

2. **Users adversely impacted by efforts to fight OCSEA:** Users of tech platforms may be adversely impacted by efforts to address OCSEA, such as overbroad or wrongful content removal and other actions (e.g., account suspensions). While all users of tech platforms are at risk, certain vulnerable populations are at greater risk of being wrongly accused of OCSEA, and/or being wrongfully accused of OCSEA may have more severe impacts on certain vulnerable populations, including:

   - **LGBTQIA+ people** have historically faced unfounded assertions that they are child molesters or pedophiles.[31] In the current political landscape, certain OCSEA behaviors, such as grooming, are weaponized against LGBTQIA+ people.[32] Similarly, medical procedures associated with LGBTQIA+ people, such as gender-affirming care, are characterized as child abuse.[33]

     Such biases about LGBTQIA+ people may lead to a higher likelihood of LGBTQIA+ people being wrongly accused of OCSEA on tech platforms, and a higher likelihood of unsubstantiated signals associated with them being shared on the Lantern Program. Similarly, there is a higher likelihood of LGBTQIA+ content (e.g., content related to transgender medical care) being signaled and removed.

     Special attention should be paid to **LGBTQIA+ youth**, who may be disproportionately signaled as a result of self-generated explicit imagery and minor-on-minor interactions because they are particularly dependent on online platforms to explore their sexual identity[34] and particularly vulnerable to harm.

   - **Sex workers** utilize online platforms as an important avenue for advertising their services.[35] There is a higher likelihood of sex workers being wrongfully accused of OCSEA because they share more content that is sexual in nature, and due to political and social biases against sex work.

[28] New Research from Thorn: LGBTQIA+ Minors are 3X More Likely to Experience Unwanted and Risky Online Interactions.

[29] Human Trafficking and Child Welfare: A Guide for Child Welfare Agencies: US Department of Health & Human Services.

[30] Children with Disabilities and Sexual Abuse: Risk Factors and Best Practice: American Bar Association.

[31] The Long, Sordid History of the Gay Conspiracy Theory: New York magazine.

[32] What Is "Grooming?" The Truth Behind the Dangerous, Bigoted Lie Targeting the LGBTQIA+ Community: ADL.

[33] How Medical Care for Transgender Youth Became 'Child Abuse' in Texas: New York Times.

[34] Online Communities and LGBTQIA+ Youth: Human Rights Campaign.

[35] Posting into the Void: Studying the Impact of Shadowbanning on Sex Workers and Activists: hacking // hustling.

- **People with intellectual disabilities** have less knowledge about sexual health and consent, and are more vulnerable to victimization,[36] making it potentially more likely that they are wrongfully accused of OCSEA.
- **People of color** have a higher chance of being wrongfully accused of OCSEA because image classifiers do not perform as well on people of color, which means that age assurance and verification mechanisms don't work as accurately (e.g., underperforming classifiers may indicate individuals as underage when they are not), which may lead to an increased number of unsubstantiated signals.
- **Underrepresented linguistic communities** also have a higher chance of being wrongfully accused of OCSEA because text classifiers do not perform as well on non-Western languages, which may lead to an increased number of unsubstantiated signals. Additionally, tech companies typically have fewer resources for content moderation in non-Western languages, increasing the likelihood of errors in content moderation.
- **Children** who perpetrate OCSEA offenses themselves, or those who create and distribute self-generated explicit imagery, may be adversely impacted through efforts to address OCSEA. For example, overly restrictive or punitive actions may be taken against them. Since children are a vulnerable group, being wrongfully accused of OCSEA may have more severe impacts on them.

[36] Personal and Sexual Boundaries: the Experiences of People with Intellectual Disabilities: BMC Public Health.

7 Recommendations

The Tech Coalition has already made significant efforts to prevent, mitigate, and address the human rights risks associated with the Lantern Program. In addition to the general governance mechanisms and participant commitments set out in Section 4 above, the Tech Coalition is also taking the following actions, among others:

- **Quality assurance plan to ensure that signals shared are applicable and relevant to OCSEA violations and abuse.** The plan includes a commitment from participants to manually review all signals that do not originate from their platforms and establish precision before taking action on them. Participants are encouraged to provide descriptions, confidence metrics, and review status when uploading signals, as well as "reactions" to uploaded signals indicating the extent to which signals mirror the description provided.
- **Parameters around what data can be included in the Lantern database to ensure only information necessary for the detection and prevention of OCSEA is shared by participants.** Information about children or teenage platform users are precluded from inclusion in the Lantern database. Furthermore, to ensure respect for users' rights of privacy, personal data issued by public authorities (such as passport or social security numbers), information related to employment or education, or sensitive information about personal characteristics are not permitted in the Lantern database.
- **Measures to prevent external involvement in the database**. Participants commit to not accept direct contributions to the Lantern database from government actors or representatives, and to challenge efforts by government actors to engage with or access the Lantern database. Participants are required to disclose any government requests, influence, or attempts to access the Lantern Program within 14 days of its occurrence.
- **Publication of a transparency report to provide information on the purpose and policies of the Lantern Program, demonstrate accountability for the use of signals by participants, and provide insights applicable to the child safety and protection field for the prevention of OCSEA.** The Tech Coalition will collect metrics from participants annually on the types of signals used, and the number of actions taken on the basis of signals in the Lantern database. The Tech Coalition also recommends that participants publish individual transparency reports.

The recommendations that follow are intended to supplement and enhance the Tech Coalition's existing efforts, as well as underline certain areas of focus. The recommendations are divided into four categories:

1. Governance and Participant Engagement
2. Transparency and Stakeholder Engagement
3. Policy and Process
4. Technical Measures

For each recommendation, BSR has written (1) explanations of how they can be implemented, (2) why they are important, and (3) a rationale based on the UNGPs and CRBPs.

These recommendations include actions for the Tech Coalition, Lantern Program participants, and the broader field:

- 🔵 **Blue dots** indicate recommended actions that the Tech Coalition can take alone (or in some cases, in collaboration with Meta).
- 🟡 **Yellow dots** represent recommended actions the Tech Coalition can take in collaboration with participating companies.
- 🟢 **Green dots** represent recommended actions the Tech Coalition can take in collaboration with the broader technology industry and other stakeholders.

The suggested time frame and priority level are indicated next to each recommendation:

- **Higher Priority:** These are fundamental steps that should be prioritized for implementation because they form the basis of a strong program and address the most severe impacts.
- **Medium Priority:** These recommendations build on higher priority recommendations and/or can be implemented in the medium- to long-term.

Governance and Participant Engagement

RECOMMENDATION 1 **PRIORITY: HIGHER**

Enforce participant commitments.

The Tech Coalition has developed a set of commitments for the participants of the Lantern Program (see Section 4.3), and it is important that these commitments are enforced. While commitments #1 (alignment with eligibility requirements), #3 (challenging external involvement), #4 (transparency), and #5 (compliance with privacy laws) are easier to enforce, some aspects of commitment #2 (quality assurance) may be difficult to enforce (e.g., participants may not be manually reviewing signals before actioning them, contradicting the guidance provided by the Tech Coalition).

In order to enforce compliance with commitments, the Tech Coalition should ensure that all participant commitments and guidelines are included in the Lantern Program agreement. In addition to the agreement, the Tech Coalition should also put in place mechanisms to help increase the likelihood that participants comply with the commitments of the program. This could include:

- **Mandatory training:** As part of the onboarding process, the Tech Coalition should conduct training for (1) the primary point of contact at each participating company and (2) the content moderators who will engage with the Lantern database. The training would teach Lantern database users how they should handle signals and describe the potential risks of not handling them according to program guidelines. Access to the Lantern database should only be allowed for individuals who have completed this training.
- **Regular check-ins:** The Tech Coalition should organize regular (e.g., quarterly) check-in meetings with each participant to understand how they are using the database and handling the signals. These meetings would also help surface learnings and best practices, which can then be shared with the broader group of participants.

Some of the other recommendations listed below (e.g., recommendation #14 on quality assurance, #15 on high signal thresholds) can also help enforce participant commitments.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with participating companies.

RATIONALE

- Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways to increase it—for example, by offering capacity-building or other incentives to the related entity, or by collaborating with other actors.
- Principle 16 of the UNGPs states that human rights policy should be "supported by any necessary training for personnel in relevant business functions."

RECOMMENDATION 2 PRIORITY: MEDIUM

Ensure that participants have a process in place for handling government requests.

One of the commitments that companies make to participate in the Lantern Program is to challenge external involvement by governments—specifically, "never to accept direct contributions to the program by anyone acting on behalf of a government and to challenge efforts by government officials to engage with or access the database."

For participants to be able to fulfill this commitment, the Tech Coalition should require as part of the eligibility requirements that companies have a robust process in place for handling government requests. This includes:

- Applying human rights principles (such as the Global Network Initiative Principles and Implementation Guidelines) when responding to government requests for user data or content removals.
- Publishing an annual transparency report to inform stakeholders about the company's approach to handling government requests.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with participating companies.

RATIONALE

- Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways to increase it—for example, by offering capacity-building or other incentives to the related entity, or by collaborating with other actors.
- Principle 17 of the UNGPs states that "in order to identify, prevent, mitigate, and account for how they address their adverse human rights impacts, business enterprises should carry out human rights due diligence. The process should include assessing actual and potential human rights impacts, integrating and acting upon the findings, tracking responses, and communicating how impacts are addressed."

RECOMMENDATION 3 PRIORITY: HIGHER

Support smaller participants to be able to comply with requirements.

Participants of the Lantern Program have different levels of resourcing and capacity for content moderation and trust and safety efforts, and it may be difficult for some of the smaller companies to fulfill the requirements for participation, leading to a higher likelihood of human rights risks. For example, some companies may not be able to manually review all signals before actioning them, or they may not be able to track the signal sharing metrics that the Tech Coalition requires them to report on a regular basis.

For these smaller companies with resource constraints, the Tech Coalition can consider offering additional support, such as designating a Tech Coalition staff member who would provide quality assurance and integration support to smaller companies in their first year of participation in the Lantern Program.

- **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

- Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways to increase it—for example, by offering capacity-building or other incentives to the related entity, or by collaborating with other actors.
- Principle 16 of the UNGPs states that human rights policy should be "supported by any necessary training for personnel in relevant business functions."

RECOMMENDATION 4 **PRIORITY: HIGHER**

Establish a human rights policy and assign a human rights lead.

Signal sharing is a new practice for most tech companies, and due to the sensitive nature of OCSEA, there are many questions that participants will need to deal with. The Tech Coalition can provide useful guidance on how to conduct signal sharing and how to handle gray areas, such as self-generated explicit imagery, and how to best provide educational resources to child users.

There are a few different ways in which the Tech Coalition can provide guidance, including:

- Create a user guide for signal sharing, including examples.
- Hold office hours or optional one-on-one meetings with participants to answer questions, understand how they are using the program, identify best practices, etc.
- Organize best practice sharing sessions for participants.

- **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

- Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways to increase it—for example, by offering capacity-building or other incentives to the related entity, or by collaborating with other actors.

RECOMMENDATION 5

PRIORITY: HIGHER

Establish a due diligence process for evaluating new members.

When evaluating potential new participants for the Lantern Program, the Tech Coalition should undertake rapid due diligence of these companies to ensure they have the necessary policies, processes, and culture in place to address the risks associated with signal sharing. In addition to securing assurance of compliance with the eligibility requirements, due diligence would help:

- Identify any potential barriers to implementation and make recommendations to ensure that participants can leverage the Lantern Program effectively and in a sustainable way.
- Identify contextual risks. For example, is the company particularly active in a jurisdiction without rule of law? Might there be risk of scope creep, based on the regulatory and political environment in the company's operating context?

Due diligence can be undertaken by the Tech Coalition, or in collaboration with an independent partner.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

- Principle 17 of the UNGPs states that due diligence "should be initiated as early as possible in the development of a new activity or relationship."
- Principle 18 of the UNGPs states that human rights assessment should be undertaken "prior to a new activity or relationship; prior to major decisions or changes in the operation (e.g., market entry, product launch, policy change, or wider changes to the business); in response to or anticipation of changes in the operating environment (e.g., rising social tensions)."

RECOMMENDATION 6

PRIORITY: MEDIUM

Consider participation by "non-tech" industries.

The Lantern Program is currently limited to participants from the tech industry. In the future, the Tech Coalition should consider expanding to other industries that can provide signals related to OCSEA through their own use of technology, such as airlines (e.g., for signals related to child sex trafficking) or financial services companies (e.g., for signals related to the financial sextortion of children).

To explore whether participation by non-tech industries makes sense for the Lantern Program, the Tech Coalition can:

- Conduct interviews with companies from different industries to gauge interest and viability.
- Conduct a pilot with interested companies that is limited to a specific OCSEA harm type (e.g., financial sextortion).

While an expansion into non-tech industries may help increase the effectiveness and impact of the Lantern Program through system-wide approaches, it would require careful consideration of new risks that may arise. In BSR's experience, "non-tech" industries have a significantly lower awareness of the human rights risks associated with data use and sharing than tech companies (although some, like the financial services industry, are more advanced than other non-tech industries); it should not be assumed that these industries have the same cautious instincts that existing members of the Tech Coalition have.

Therefore, in case of expansion into non-tech industries, a key role for the Tech Coalition would be capacity-building to ensure the responsible integration of these other industries into the child protection ecosystem. Specifically, the Tech Coalition can consider putting in place a special onboarding and mentoring system for "non-tech" participants.

- **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

- Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways to increase it—for example, by offering capacity-building or other incentives to the related entity, or by collaborating with other actors.
- Principle 17 of the UNGPs states that due diligence "should be initiated as early as possible in the development of a new activity or relationship," while Principle 18 states that human rights assessment should be undertaken "prior to a new activity or relationship; prior to major decisions or changes in the operation (e.g., market entry, product launch, policy change, or wider changes to the business); in response to or anticipation of changes in the operating environment (e.g., rising social tensions)."

Transparency and Stakeholder Engagement

RECOMMENDATION 7 PRIORITY: HIGHER

Enhance the Lantern Program's approach to transparency.

The Tech Coalition's transparency approach for the Lantern Program should be geared toward enabling stakeholders to understand how the program is being used and whether it is achieving its stated goals. Transparency reporting should provide information that is sufficient to evaluate the adequacy of the Tech Coalition's approach while not posing risks to stakeholders or to the success of the Lantern Program.

The Tech Coalition is planning regular transparency reporting around the Lantern Program. In addition to the metrics and narratives that are in the existing Transparency Template, BSR recommends including information about the following aspects of the program:

- **Standards for the inclusion of signals:** A description of thresholds and criteria that participants use when evaluating which signals to upload to the database.

- **Information related to quality assurance:** For example, how many signals in the database were identified as "unsubstantiated signals"? What is the count of erroneous account removals?
- **Information about government requests:** A summary of requests that the Tech Coalition and participants have received from governments or law enforcement agencies, including the number of requests, what the request was for (e.g., user data, content removal, etc.), and how the Tech Coalition or participants have responded to the request.
- **Descriptions of lessons learned:** A discussion of challenges faced as part of the signal sharing process and how these have been or are planning to be addressed.
- **Insights for the field:** Information about OCSEA trends, and what works or what doesn't work in OCSEA content moderation (e.g., whether certain keywords are helpful in identifying OCSEA practices, or details about the cross-platform dynamics of live streaming).

While increased transparency is generally positive, the Tech Coalition should also carefully consider the risks involved with being transparent. For example:

- Transparency about the Lantern Program and its use may provide insights for malicious actors to manipulate the system.
- It may expose participating companies to legal scrutiny (and therefore disincentivize them from engaging).
- It may attract governments and law enforcement to coerce the Tech Coalition or participating companies to share user information.

In order to prevent such risks, the Tech Coalition can consider a layered approach to transparency. For example, some insights would only be shared with a select group of stakeholders, while others would be shared publicly.

The case of the Global Internet Forum to Counter Terrorism's (GIFCT) Hash-Sharing Database for terrorist and violent extremist content may provide helpful insights about effective transparency. Following critique from civil society stakeholders, the GIFCT conducted extensive stakeholder engagement about how the database worked, created an informational video, and released detailed transparency reports. These efforts have ultimately helped gain the trust of stakeholders about the integrity of the GIFCT's Hash-Sharing Database.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

- Principle 21 of the UNGPs states that companies should be prepared to communicate how they address their human rights impacts externally, providing a measure of transparency and accountability to impacted individuals or groups and other relevant stakeholders. Communication can take various forms, such as in-person meetings, online dialogues, consultation with affected stakeholders, and formal public reports. Formal reporting is expected where there are risks of severe human rights impacts due to the nature of business operations or operating contexts.
- Principle 1 of the CRBPs further states that companies should "communicate externally on their efforts to address the business impact on children's rights in a form and with the frequency that

reflects such an impact and that is accessible to its intended audiences. The business should provide sufficient information to evaluate the adequacy of its responses. Such communication should not pose risks to affected stakeholders, personnel, or to legitimate requirements of commercial confidentiality."

RECOMMENDATION 8 **PRIORITY: MEDIUM**

Facilitate an annual dialogue with key stakeholders where participants can share and discuss their use of the Lantern Program.

In addition to the transparency reporting that the Tech Coalition is planning for the Lantern Program, it is important for the participants to also be transparent about their participation in the Lantern Program and their use of the database. Requiring transparency by participants is (1) a natural extension of emphasizing that each company makes its own decisions on the basis of information received, and (2) an additional mechanism to assure compliance with the multiparty agreement.

BSR recommends that participants of the Lantern Program take part in an annual dialogue with select and informed stakeholders to discuss their use of the signal sharing program, as well as share any insights and lessons learned. We believe this would facilitate a candid and open dialogue with informed stakeholders and serve the twin goals of facilitating accountability and strengthening the effectiveness of the Lantern Program.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with participating companies.

RATIONALE

- Principle 20 of the UNGPs states that "to verify whether adverse human rights impacts are being addressed, business enterprises should track the effectiveness of their response. Tracking should: (a) Be based on appropriate qualitative and quantitative indicators; (b) Draw on feedback from both internal and external sources, including affected stakeholders."
- Principle 18 of the UNGPs states that human rights assessment should "involve meaningful consultation with potentially affected groups and other relevant stakeholders."
- Principle 1 of the CRBPs states that companies should "monitor and track the effectiveness of the business's responses in order to verify whether an adverse impact on children's rights is being addressed, using appropriate qualitative and quantitative indicators and drawing on feedback from internal and external sources, including affected children, families, and other stakeholders."

RECOMMENDATION 9 **PRIORITY: HIGHER**

Establish a strategy for ongoing stakeholder engagement.

Engagement with affected stakeholders and other experts underpins a human rights-based approach; building strong feedback loops through stakeholder engagement while avoiding engagement fatigue will be an important balance to strike, particularly as the Tech Coalition launches the Lantern Program and scales usage to a wider group of participants.

Stakeholder engagement can inform:

- Standards for the inclusion of signals
- Location / geographic risks
- Risks to particular vulnerable groups
- Effectiveness of reporting and appeals systems
- Improvements to transparency approach

Stakeholders should include a range of potentially affected rightsholder groups (see Section 6.4 above) and experts. The strategy should be informed by best practices in stakeholder engagement, including strong contact management (e.g., avoiding ad hoc and duplicate requests to stakeholders), clear feedback loops (e.g., reporting back what happened), and being strategic in engagement to avoid "engagement fatigue" (e.g., predictable cycle, or using multistakeholder forums where possible).

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways to increase it—for example, offering capacity-building or other incentives to the related entity, or collaborating with other actors.

RECOMMENDATION 10

PRIORITY: MEDIUM

Establish an advisory board for the Lantern Program.

There are important questions around how the human rights and child rights implications of the Lantern Program are understood, anticipated, and addressed by the Tech Coalition and Lantern program participants. An advisory board with interdisciplinary expertise can help the Tech Coalition understand and address these risks.

The advisory board for the Lantern Program can be an extension or sub-committee of the existing Tech Coalition Board. It would include individuals from both the child safety and digital rights fields.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with the broader technology industry and other stakeholders.

RATIONALE

- Principle 18 of the UNGPs states that human rights assessment should "involve meaningful consultation with potentially affected groups and other relevant stakeholders."
- Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways

to increase it—for example, by offering capacity-building or other incentives to the related entity, or by collaborating with other actors.

- Many of the potential adverse impacts highlighted in this assessment cannot be addressed by the Tech Coalition alone, but instead require system-wide and multistakeholder approaches.

RECOMMENDATION 11 PRIORITY: MEDIUM

Share insights and work with stakeholders to contribute to the broader field.

Insights gleaned through the Lantern Program can be useful for broader industry, civil society, and policy efforts to fight OCSEA. In addition to the information shared in annual transparency reports, the Tech Coalition should explore other avenues to share insights and collaborate with stakeholders.

There are specific areas where Lantern Program insights may be particularly helpful:

- Gray areas that the field has not yet agreed how to handle, such as self-generated explicit imagery and minor-on-minor offenses.
- Industry-wide challenges, such as age assurance and verification, or the identification of intent in sexual content.
- New or evolving risk areas, such as synthetic or AI-generated CSAM.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with the broader technology industry and other stakeholders.

RATIONALE

- Principle 19 of the UNGPs states that where a company has leverage to prevent or mitigate adverse impacts, it should exercise it; and where a company lacks leverage, it should seek ways to increase it—for example, by offering capacity-building or other incentives to the related entity, or by collaborating with other actors.

RECOMMENDATION 12 PRIORITY: MEDIUM

Investigate enabling third-party audits and trusted researcher access to the database.

At present, only registered Twitch users can report potentially violating content, rather than casual viewers. In practice, many casual viewers (or users not logged in to their account) may see and want to report potentially violating content, and Twitch's ability to review and address potentially harmful content would benefit from receiving these reports.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with the broader technology industry and other stakeholders.

RATIONALE

- Principle 20 of the UNGPs states that companies should "verify whether adverse human rights impacts are being addressed … by tracking the effectiveness of their response." This should "draw upon feedback from both internal and external sources, including affected stakeholders."

- Principle 1 of the CRBPs states that companies should "monitor and track the effectiveness of the business's responses in order to verify whether an adverse impact on children's rights is being addressed, using appropriate qualitative and quantitative indicators and drawing on feedback from internal and external sources, including affected children, families, and other stakeholders."

Policy and Process

RECOMMENDATION 13 **PRIORITY: HIGHER**

Establish a robust quality assurance (QA) process.

A robust QA process is vital to addressing the most salient human rights risks associated with the Lantern Program. The QA process should enable regular, timely, and prioritized review of signals for quality and relevance, and feature feedback loops that can alert the Tech Coalition to trends in new signals. The QA process can also help ensure that the Lantern Program is useful and effective at achieving its goal—for example, by identifying what signals are most useful over time for what types of companies.

As part of the QA process, BSR recommends conducting regular check-in meetings with each participating company to gather qualitative insights about how they are using the database, technical / operational issues they have encountered, and feedback on the quality of the signals shared (e.g., signals shared by a particular participant being more or less useful than others).

As Lantern matures and the Tech Coalition gains more clarity about the utility of different signals, it will be important to develop metrics to monitor quality and use over time.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with participating companies.

RATIONALE

- Principle 19 of the UNGPs states that companies should "integrate the findings from their impact assessments across relevant internal functions and processes, and take appropriate action."
- Principle 20 of the UNGPs states that in order to verify whether adverse human rights impacts are being addressed, companies should track the effectiveness of their responses based on appropriate qualitative and quantitative indicators, as well as feedback from both internal and external sources.
- Principle 1 of the CRBPs states that companies should "integrate the findings from their impact assessments across relevant internal functions and processes and take appropriate action."

RECOMMENDATION 14 PRIORITY: HIGHER

Monitor the use of the Lantern Program for non-English languages and make improvements as needed.

The Tech Coalition should monitor the extent to which non-English signals are being shared, stored, and used in the Lantern database for two reasons: (1) a large portion of CSAM and OCSEA behaviors such as sextortion or livestreaming are increasingly originating from non-English regions, and (2) tech companies typically have fewer content moderation resources for underrepresented languages and regions, which may increase the likelihood of risks associated with unsubstantiated signals and erroneous actioning of signals.

The Tech Coalition should ensure that it has insights into the quality of signals in all languages and that its quality assurance process applies to non-English languages. The Tech Coalition can also consider developing guidelines for sharing signals in different languages, to support companies that have fewer resources for underrepresented languages.

- **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with the broader technology industry and other stakeholders.

RATIONALE

› The UNGPs state that companies should pay "particular attention to the rights and needs of, as well as the challenges faced by, individuals from groups or populations that may be at heightened risk of becoming vulnerable or marginalized."

› The CRBPs highlight that one of the core principles enshrined in the Convention of the Rights of the Child is nondiscrimination, which "provides for equal treatment of an individual irrespective of race, color, sex, language, disability, religion, political or other opinions; national, social, or indigenous origin; and property, birth, or other status."

RECOMMENDATION 15 PRIORITY: HIGHER

Implement additional procedural requirements for "high-risk" signals.

To minimize the risks associated with signal sharing (e.g., unsubstantiated signals, scope creep), the Tech Coalition should establish high thresholds for signals that are added to the database. Particularly for high-risk signals—e.g., signals that include personally identifiable information (PII) or signals that might include information about a child—the Tech Coalition should consider putting in place additional procedural requirements. For example:

› Requiring manual review by an expert before uploading high-risk signals into the database. The Lantern interface can include a specific tag for signals that have gone through this additional manual review.

› Creating a checklist that participants need to use before uploading high-risk signals. This checklist can prompt participants to consider whether the inclusion of the signal in question is necessary and proportionate to the intended purpose. A copy of the completed checklist would be uploaded to the Lantern database along with the signal.

As the Lantern Program scales and the volume of signals shared in the database increases, the Tech Coalition should regularly review the criteria for high-risk signals to ensure that new signal types are included and that the manual review requirement can be effectively implemented by participants.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition in collaboration with participating companies.

RATIONALE

- Principle 17 of the UNGPs states that due diligence "should be initiated as early as possible in the development of a new activity or relationship."
- Principle 18 of the UNGPs states that human rights assessment should be undertaken "prior to a new activity or relationship; prior to major decisions or changes in the operation (e.g., market entry, product launch, policy change, or wider changes to the business); in response to or anticipation of changes in the operating environment (e.g., rising social tensions)."

RECOMMENDATION 16 **PRIORITY: HIGHER**

Take a human rights-based approach to responding to government requests for data.

As the Lantern Program launches and becomes more widely known, the Tech Coalition may start receiving requests for user data from governments and law enforcement agencies. To avoid adverse human rights impacts that may be associated with overbroad government requests, the Tech Coalition should establish a human rights-respecting process for responding to such requests and inform stakeholders about its approach through the Lantern Program transparency report.

The process should use the Global Network Initiative Principles and Implementation Guidelines as a starting point, and may include investigating the nature of the request, challenging jurisdictional claims (where relevant), and not acting on law enforcement requests for user data until the nature of the request is clear.

When building this process, it will be important for the Tech Coalition to consider its interaction with different legal regimes—especially since many OCSEA threats are more prevalent in non-Western countries. Signals shared as part of the Lantern Program may be seen as a valuable source of information that can help governments with fewer resources better identify and address OCSEA crimes; on the other hand, there is a risk that authoritarian governments or those in countries with weak rule of law may use this information in ways that violate individuals' rights.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

- Principle 17 of the UNGPs states that "in order to identify, prevent, mitigate, and account for how they address their adverse human rights impacts, business enterprises should carry out human rights due diligence. The process should include assessing actual and potential human rights impacts, integrating and acting upon the findings, tracking responses, and communicating how impacts are addressed."

- Principle 19 of the UNGPs states that companies should "integrate the findings from their impact assessments across relevant internal functions and processes, and take appropriate action."
- The GNI Principles state that "If national laws, regulations and policies do not conform to international standards, ICT companies should avoid, minimize, or otherwise address the adverse impact of government demands, laws, or regulations, and seek ways to honor the principles of internationally recognized human rights to the greatest extent possible."

RECOMMENDATION 17 **PRIORITY: MEDIUM**

Conduct ongoing human rights due diligence (HRDD).

New human rights risks may emerge over time as participation in the Lantern Program increases, use of the Lantern database expands, and global regulatory frameworks for the prevention and detection of OCSEA evolve. The Tech Coalition should conduct ongoing human rights due diligence to identify and assess key developments that may indicate shifts in human rights risks or impacts of the Lantern Program over time.

Due diligence should be conducted before key milestones and decision points, including the following:

- Proliferation of new technologies that may impact user behavior on online platforms (e.g., generative AI technologies)
- Proliferation of new OCSEA threats
- Enactment of new OCSEA regulations
- Expansion of the Lantern Program to new participants or industries
- Expansion of the signal types included in the Lantern database

Ongoing HRDD could include methods such as adversarial testing (i.e., exercises where the product or program is stress tested to discover the ways in which it might be misused, abused, or associated with harmful outcomes).

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition.

RATIONALE

- Principle 17 of the UNGPs states that "in order to identify, prevent, mitigate, and account for how they address their adverse human rights impacts, business enterprises should carry out human rights due diligence." Further, Principle 17 of the UNGPs also states that human rights due diligence, "should be ongoing, recognizing that the human rights risks may change over time as the business enterprise's operations and operating context evolve."
- Principle 1 of the CRBPs states that companies should conduct ongoing human rights due diligence "for assessing its actual and potential human rights impact, including on children's

rights" and "integrate the findings from their impact assessments across relevant internal functions and processes and take appropriate action."

- Principle 5 of the CRBPs states that companies should "seek opportunities to support children's rights through products and services, as well as their distribution."

Technical Measures

RECOMMENDATION 18 **PRIORITY: MEDIUM**

Implement technical barriers to minimize data collection, storage, and sharing.

To minimize risk associated with data collection, storage, and sharing, the Tech Coalition should uphold the following privacy principles:

- Data Minimization: The processing of personal data should be adequate, relevant, and limited to necessity of the purpose for which it is being processed.
- Storage Limitations: Personal data should only be retained for the period of time that is necessary for the purposes for which it was processed.

Technical measures that can be implemented to uphold these principles include the following:

- Configure the database such that users cannot bulk download all signals. This can help minimize the risk that the signals are shared with external parties (e.g., government officials) or are shared with too many people.
- Configure the database such that, by default, certain signals are shared only with participants for whom those signals may be relevant.
- Establish time limits on how long sensitive or personal data is present in the database.
- Establish time limits for false or unactioned signals so that these are automatically removed after a certain period of time.

● **Implementing this recommendation:** This recommendation would be implemented by the Tech Coalition and Meta.

RATIONALE

- Principle 19 of the UNGPs states that "If the business enterprise has leverage to prevent or mitigate the adverse impact, it should exercise it. And if it lacks leverage there may be ways for the enterprise to increase it."
- Principle 5 of the CRBPs states that companies should "seek opportunities to support children's rights through products and services, as well as their distribution."
- Privacy and data protection principles based on the human rights framework include data minimization and storage limitation, among others.

RECOMMENDATION 19

PRIORITY: MEDIUM

Build an automatic flagging system for signals that are not allowed or are high risk.

The Tech Coalition does not allow certain signals to be shared in the Lantern database, such as information about children and information about sensitive characteristics. Together with the Meta ThreatExchange team, the Tech Coalition can build a classifier system that flags when participants try to upload such signals to the database.

The same system can also flag high-risk signals such as those that involve text from private communications, and prompt participants to complete and upload the checklist described in Recommendation #14.

- **Implementing this recommendation:** This recommendation would implemented by the Tech Coalition and Meta.

RATIONALE

- Principle 17 of the UNGPs states that due diligence "should be initiated as early as possible in the development of a new activity or relationship."
- Principle 18 of the UNGPs states that human rights assessment should be undertaken "prior to a new activity or relationship; prior to major decisions or changes in the operation (e.g., market entry, product launch, policy change, or wider changes to the business); in response to or anticipation of changes in the operating environment (e.g., rising social tensions)."

BSR®

BSR™ is an organization of sustainable business experts that works with its global network of the world's leading companies to build a just and sustainable world. With offices in Asia, Europe, and North America, BSR™ provides insight, advice, and collaborative initiatives to help you see a changing world more clearly, create long-term business value, and scale impact.

www.bsr.org