

Twitch Human Rights Impact Assessment

April 2023



BSR®

Table of Contents

1. Executive Summary	3
2. Introduction	9
3. Methodology	10
4. BSR Observations for Twitch	15
5. Human Rights Risks	21
6. Human Rights Based Approach to Content Governance for Twitch	40
7. Recommendations	42
Annex 1: Community Guidelines Gap Analysis	57
Annex 2: Human Rights Due Diligence Resources	69

Acknowledgements

BSR, Amazon, and Twitch wish to thank all Twitch employees, 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.

Executive Summary



Purpose

Twitch partnered with BSR and Amazon's central human rights team to undertake a human rights assessment of Twitch. The goal of the assessment is to:

- Identify and prioritize human rights impacts with which Twitch is involved, including both risks and opportunities, and the vulnerable groups impacted.
- Recommend appropriate action for Twitch to address these risks (i.e., avoid, prevent, mitigate, and remedy).
- Describe the roles and responsibilities of other actors in the Twitch value chain, and identify how Twitch could partner with these actors to address these risks.

This assessment focuses on the human rights most relevant to Twitch platform policy, partnerships, and impacts (including safety operations), rather than Twitch's broader operations and supply chain. This focus was agreed with Twitch at the outset given the likely greater salience of these human rights risks and the increasing stakeholder interest in them. Off-service harassment issues were not in-scope for this assessment.

Methodology

BSR's HRIA methodology to identify and prioritize human rights risks is based on the [UN Guiding Principles on Business and Human Rights](#) (UNGPs), which were unanimously endorsed by the UN Human Rights Council in 2011 and provide authoritative guidance on the corporate responsibility to respect human rights. Amazon's [Global Human Rights Principles](#) is based on a commitment to implement the UNGPs.

This assessment was undertaken between December 2021 and April 2022, and subsequently updated in May 2022 in light of the mass shooting in Buffalo, New York.

Consistent with the UNGPs (as well as emerging regulatory requirements in the EU), the prioritization of human rights risks 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.

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

- **Scope**—The number of people affected by the harm, including both users and non-users.
- **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 chance of the human rights risk occurring in the next five years.

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

This assessment makes recommendations for Twitch to address human rights risks using factors contained in the UNGPs:

- **Attribution / Scope of responsibility**—How closely would Twitch be connected to the human rights impact?
 - › **Caused the impact**—Twitch should take the necessary steps to cease or prevent the impact.
 - › **Contributed to the impact**—Twitch 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**—Twitch 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 Twitch have to affect change in the wrongful practices of an entity that “causes” or “contributes to” the harm? How can the company increase leverage, such as by collaborating with other actors?

BSR’s analysis was informed by document review (both publicly available and internal to Twitch), and engagement with rightsholders, civil society organizations, and experts able to provide insights into the human rights risks and opportunities associated with the Twitch platform.

Observations

The following observations arose during this assessment and influenced the BSR analysis of human rights risks and appropriate actions to address them. These observations provide context for the analysis and recommendations that follow:

Current and Future State

- Twitch’s content governance framework for policy creation, implementation, and operations appears thoughtful, robust, and consistent with key elements of human rights guidance—for example, engagement with stakeholders when developing content policy, deploying the concept of severity of harm when implementing the Twitch Community Guidelines, and the ability of users to report content and appeal content decisions.

- The Twitch platform has a lower human rights risk profile than other social media platforms, but this may change over time, especially as Twitch moves beyond the gaming community and becomes popular in more locations. Twitch will need to assess and address the human rights risks associated with expansion and prepare for content-related risks globally—for example, being able to moderate content in even more cultures, contexts, dialects, and languages.
- Twitch will also need to prepare for new human rights risks associated with other changes to the business model and platform, such as the use of recommendation algorithms beyond prominent / trusted creators to surfacing small creators and increased use of targeted advertising.

Live Streaming and Content Policy

- Live streaming presents content policy dilemmas that need to be addressed and that would benefit from further dialogue with stakeholders and experts, such as the privacy rights of those incidentally captured in live streaming and the live streaming of major events (such as protests, conflict, and other gatherings) where violating content may exist in the context of otherwise valuable streaming.
- Live streaming also presents content policy enforcement challenges, such as the limited reliability of tools for automated detection of potentially violating content and recorded content being subsequently uploaded and shared on other platforms, including those not making use of hash sharing resources.
- Community moderators play an essential role in identifying potentially violating content on chat and in live streams, and will benefit from resources, training, and investment that include human rights priorities, such as training on transphobia, gender, and hate speech.

Due Diligence and Access to Remedy

- Twitch can integrate human rights into existing risk assessment processes, such as the safety-by-design review process. Regulatory developments in the EU—such as the General Data Protection Regulation (GDPR), Digital Services Act (DSA), and Artificial Intelligence Act—all require risk assessments of harm to people and society in a manner similar to the human rights due diligence requirements of the UNGPs, and all make direct reference to the EU Charter of Fundamental Rights, which contains a very similar list of rights as the International Bill of Human Rights.
- Twitch can continue to develop a robust stakeholder engagement strategy to engage with affected stakeholders and other experts, such as by building upon the Twitch Safety Advisory Council. Stakeholder engagement underpins ongoing human rights due diligence.
- Child rights require special attention, given the reality (despite Twitch policy) that children below 13 will use Twitch, and children between 13 – 18 will use Twitch without the supervision of a parent or legal guardian. This might include reviewing the accessibility of reporting channels, enhancing age assurance mechanisms, and creating an overall child rights strategy.
- Recent changes to the Twitch appeals process improves access to remedy according to standards set out in the UNGPs; however, further improvements are needed to both the reporting and appeals channels to adhere to the expectations of the UNGPs, such as accessibility based on age, language, and digital literacy, and providing access to non-users without a Twitch account.

System wide Approaches

- Many of Twitch’s human rights risks are beyond the ability of Twitch to address alone, and multi-stakeholder efforts such as the Global Network Initiative (GNI), the Global Internet Forum to Counter Terrorism (GIFCT), the Technology Coalition, and the Digital Trust and Safety Partnership (DTSP) offer important opportunities to collaborate with others.
- Addressing live streaming risks and developing moderation tools and approaches is one area where Twitch can usefully collaborate with others. It is noticeable that multi-stakeholder efforts are at the early stages of addressing human rights risks associated with live streaming, and Twitch can participate in efforts to explore issues, topics, and challenges related to live streaming.
- The increased government interest in technology policy has implications for Twitch’s human rights strategy. It is important that prominent technology policy, laws, and regulations established by governments are designed and implemented in a manner consistent with human rights.

Human Rights Opportunities

- While applying the UNGPs primarily means identifying and addressing adverse human rights impacts, a human rights strategy should also include areas where Twitch can support the enjoyment, realization, and fulfillment of human rights. Promoting access to culture, enhancing freedom of expression, assembly, and association, and facilitating increased political participation are three such opportunities for Twitch.

Human rights risks

BSR identifies the following ten salient human rights issues for Twitch, with more complete descriptions provided in section 5 of this assessment. It is important to note the following:

- The **ten salient human rights issues** listed here are a narrowed list of potential salient risks that **cross a threshold of relevance for Twitch**. In other words, there are other human rights risks (such as the right to marriage or the right to a nationality) that are already excluded from this list of ten salient human rights issues.
- BSR has further **prioritized among these ten salient human rights issues** using the criteria of severity (scope, scale, remediability) and likelihood described in the methodology (above). However, BSR urges caution with this prioritization since (1) it 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.
- Under the UNGPs Twitch 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 Twitch’s responsibility to address all adverse human rights impacts.

Tier One Issues	Tier Two Issues	Tier Three Issues
<ul style="list-style-type: none"> › Bodily Security Rights › Child Rights › Right to Freedom of Opinion and Expression, and Access to Information › Right to Freedom of Thought, Belief, and Religion › Right to Privacy 	<ul style="list-style-type: none"> › Right to Access to Culture and Right to Own Property › Right to Equality and Non-Discrimination › Right to Participate in Government 	<ul style="list-style-type: none"> › Right to Peaceful Assembly and Association › Right to Education

Recommendations

Section 7 of this assessment provides 24 recommendations for how potential impacts may be avoided, prevented, mitigated, or remediated. While all recommendations represent appropriate action for Twitch to take, BSR prioritizes the following 11 recommendations:

- Address the gaps identified in the human rights gap analysis BSR undertook of the Twitch Community Guidelines, specifically (1) definition of harm, (2) clearer prohibition of human trafficking, (3) respect for cultural heritage, (4) educational content, and (5) privacy rights during live streaming. BSR highlights the latter (privacy rights during live streaming) as especially important.
- Establish a Twitch human rights policy and assign a human rights lead.
- Assess and address human rights risks associated with informal location growth, such as how Twitch may be used, user groups and demographics, the political context, existing types of societal discrimination and social tensions that may appear on the platform, and government demands for data and content restrictions.
- Establish a plan for effective content moderation in the likelihood that the Twitch platform becomes more widely used in more cultures and languages, including a consideration of dialect, the distinction between text and audio, and the appropriate use of automated tools to help manage content moderation at scale.
- Centralize and enhance training for content moderators, and consider adding dedicated moderators for large public livestreams. Training should focus on how to apply Twitch policies across geographies and cultures, and include human rights-relevant content, such as transphobia, gender, and different forms of hate speech.
- Review and update reporting channels with vulnerable users, including those under 18, in mind. This should include consulting or engaging with experts to ensure that reporting channels are visible, easily discoverable, recognizable, and accessible to all users.
- Collaborate with civil society actors and industry associates to conduct research and develop rights-based approaches to age assurance—for example, that do not violate a child's right to privacy.

- Integrate human rights into existing assessment processes, such as the Twitch “safety-by-design” process and assessment processes established to comply with the European Union GDPR, DSA, and Artificial Intelligence Act. Key elements include a review against all potentially relevant human rights, assessing impacts on vulnerable groups, and identifying appropriate action.
- Assess and address the human rights risks associated with the expansion of recommendation algorithms beyond prominent / trusted creators to surfacing small creators.
- Assess and address the human rights risks associated with targeted advertising.
- Fund external research into the potential link between gaming, terrorism, and violent extremism, which today is often assumed to exist but isn’t fully understood, and where assertions made are typically anecdotal and qualitative, rather than evidence based and quantitative.

BSR’s other recommendations include investment in government affairs capacity, joining or becoming more active in relevant collaborative initiatives addressing challenges that no single company can address alone (such as the GNI, GIFCT, DTSP, and the Technology Coalition), and various product related initiatives (such as exploring how to mitigate risks associated with mobile streaming, further enhancement to reporting and appeals channels, such as allowing non-users to report potentially violating content, and addressing risks associated with targeted advertising).

Introduction



Twitch partnered with BSR (a global non-profit organization working with companies on just and sustainable business) and Amazon's central human rights team to undertake a human rights assessment of Twitch. The goal of the assessment is to:

- Identify and prioritize human rights impacts with which Twitch is involved¹, including both risks and opportunities, and the vulnerable groups impacted.
- Recommend appropriate action for Twitch to address these risks (i.e., avoid, prevent, mitigate, and remedy).
- Describe the roles and responsibilities of other actors in the Twitch value chain, and identify how Twitch could partner with these actors to address these risks.

This assessment focuses on the human rights most relevant to Twitch platform policy, external partnerships, and impacts (including safety operations), rather than Twitch's broader operations and supply chain. This focus was agreed with Twitch at the outset given the likely greater salience of these human rights risks and the increasing stakeholder interest in them. Off-service harassment was not in-scope in this assessment.

As a result of this project, Twitch should have a deepened understanding of the potential human rights risks with which Twitch is involved, the insights necessary to address them, and the knowledge required to take a human rights-based approach to Twitch products, service offerings, and location growth.

Twitch was chosen within Amazon to test how human rights assessments can assist with the commitment to embedding the [Amazon Global Human Rights Principles](#) across the company, and to address the rapidly evolving human rights risks and opportunities associated with content governance.

¹ UNGPs Principle 11 states that companies should "address adverse human rights impacts with which they are involved."

Methodology



BSR's HRIA methodology is 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. This assessment was undertaken between December 2021 and April 2022, and subsequently updated in May 2022 in light of the mass shooting in Buffalo, New York.

Identifying Human Rights Risks

In line with the UNGPs, companies have a responsibility to respect all internationally recognized human rights. In this assessment, BSR identifies the potential human rights impacts of Twitch using the universe of rights codified in the following international human rights instruments.²

- 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 Labour Organization (ILO) Core Conventions
- The Convention on the Rights of the Child
- ILO Convention 169 on Indigenous Peoples

Furthermore, 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 are in tension with each other (e.g., privacy vs. freedom of expression) and a company needs to make choices when two competing rights cannot both be achieved

² These are the core international human rights instruments and other instruments / conventions that are potentially most relevant for the scope of this assessment

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.

Prioritizing Human Rights Risks

Consistent with the UNGPs (as well as emerging regulatory requirements in the EU), the prioritization of human rights risks 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, 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 risk. Consistent with the UNGPs, severity is not an absolute concept in this context, but is relative to the other human rights impacts with which Twitch is involved:

- **Scope**—The number of people affected by the harm. This is defined as a percentage of the population that could be affected by the harm, which in Twitch’s case includes both the users of the Twitch platform and non-users who may be impacted offline by online content (e.g., victims of incitement to violence).
- **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 chance of the human rights risk occurring in the next five years. Factors involved in an assessment of likelihood include whether the impact has happened in the past or is happening today, whether 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.

Human rights risks 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 are shown in section 5 (human rights risks).

Salience	Levels				
Scope How many people could be affected by the harm? This includes users and other impacted populations.	Smallest Less than 5% of the population is considered impacted	Small 5 – 20% of the population is considered impacted	Medium 21 – 80% of the population is considered impacted	Large 81 – 90% of the population is considered impacted	Largest 91 – 100% of the population is considered impacted
Scale How serious would the impacts be for those affected?	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 those affected to the same or equivalent position before the harm?	Remediable Remedy will return those affected to the same or equivalent position	Likely Remediable Remedy is likely to return those affected to the same or equivalent position before the harm occurred	Possibly Remediable Remedy may help return those affected to the same or equivalent position before the harm occurred	Rarely Remediable Remedies can rarely return those affected to the same or equivalent condition before harm occurred	Not Remediable Remedies will not return those affected to the same or equivalent condition before harm occurred
Likelihood of Occurrence What is the likelihood of the risk occurring considering context (e.g., locations) and business relationships?	Minor Likelihood 0-10% chance of occurrence. Although a risk, it is highly unlikely that impacts on rights-holders may occur.	Some Likelihood 10-40% chance of occurrence. There is some minor risk that the impacts may occur.	Good Likelihood 40-70% chance of occurrence. It's more probable than not that the impacts on rights-holders may occur.	High Likelihood 70-90% chance of occurrence. There is a high likelihood that the impacts on rights-holders may occur.	Certain 90-100% chance of occurrence. Currently occurring or certain to occur.

Determining Appropriate Action

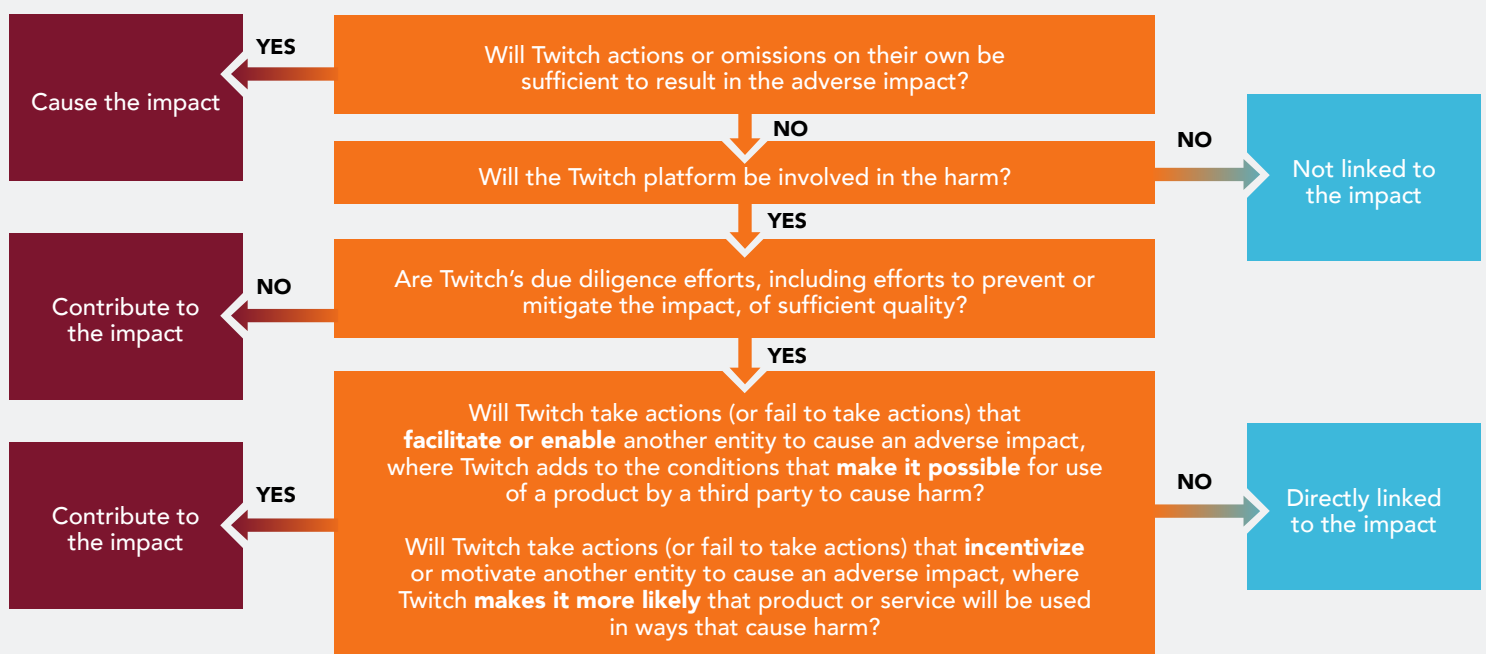
BSR's HRIA methodology considers the appropriate action for a company to address adverse human rights impacts using factors contained in Principle 19 of the UNGPs.:

- **Attribution / Scope of responsibility**—How closely would Twitch be connected to the human rights impact? BSR uses the following definitions and “decision tree”.
 - › **Caused the impact**—Twitch should take the necessary steps to cease or prevent the impact.
 - › **Contributed to the impact**—Twitch 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**—Twitch should determine action based on factors such as the extent of leverage over the entity concerned and the severity of the abuse.

Applying the “cause, contribute, directly linked” framework to social media platforms is challenging due to the complex ways in which platforms interact with, enable, and amplify human behavior.

Given the importance of context, BSR considers this framework to be helpful in setting overall direction in a human rights assessment, rather than providing a definitive “answer” for each impact.³ Further, it is important to note that this decision tree is most useful when assessing individual cases (e.g., identifying whether Twitch caused, contributed to, or was directly linked to specific hate speech content) rather than overall categories (e.g., identifying whether Twitch causes, contributes to, or is directly linked to harassment online more generally), and this makes our analysis necessarily general in nature. However, the factors influencing our analysis for each impact is shown in section 5 (human rights risks).

It should also be noted that methods to apply the “cause, contribute, directly linked” framework to the technology industry are under development.⁴



³ See [Seven Questions to Determine a Company's Connections to Human Rights Abuses](#) for more analysis

⁴ See [UN B-Tech Project](#), especially [Taking Action to Address Human Rights Risks Related to End-Use](#)

- **Leverage**—How much ability would the company have to affect change in the wrongful practices of an entity that “causes” or “contributes to” the harm? How can the company increase leverage, such as by collaborating with other actors?

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 women, children, low income groups, members of the LGBTQI+ community, and marginalized racial, ethnic, and religious groups.

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 informed 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, policy makers, 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.

In this HRIA, BSR engaged with rightsholders, civil society organizations, and experts able to provide insights into the risks and opportunities associated with the Twitch platform. To enable candid dialogue the identities of the individuals and organizations with whom we consulted are being kept confidential.

BSR Observations for Twitch



In this section we provide some observations that arose during this assessment and that influenced the BSR analysis of human rights risks and appropriate actions to address them. These observations provide context for the analysis and recommendations that follow.

Current State

- **Twitch’s content governance framework for policy creation, implementation, and operations appears thoughtful, robust, and consistent with key elements of human rights guidance.** In this assessment BSR reviewed Twitch’s approach to content governance against best practices of a human rights based approach to content governance, and found many key elements already present at Twitch. These provide a very strong foundation upon which to build. For example:
 - › Twitch focuses on vulnerable groups and undertakes stakeholder engagement when developing content policy (e.g., the Policy Authorization Process, or PATH).
 - › Twitch deploys the concept of severity of harm (i.e., the scope and scale of harm) as a key variable when determining suspension length when taking action against users for violating Twitch Community Guidelines.
 - › Key elements of policy implementation and operations (e.g., hiring and training of content moderators; tracking and quality control; use of strikes; user communications when action is taken) all seem designed to address harms in a manner that respects the human rights of users.
- **The Twitch platform has a lower human rights risk profile than other social media platforms today, but this may change over time.** The nature of the Twitch platform as live-streaming focused and community oriented, combined with its current usage patterns, lowers the risk of large-scale societal harm compared to other social media platforms. For example, live streaming combined with the community content moderation model presents a lower risk of harmful content going viral, because the most popular channels are typically more heavily moderated, moderators have a vested interest in the long term success of that channel, content might need to be recorded go viral, and users are highly likely to report any problematic content.⁵ There is also less political or social issues-oriented content than found on other platforms.

⁵ Less risk doesn’t mean no risk—for example, a newer or less popular channel with less community moderation, or a channel created by a bad actor, could go viral.

However, it is important to be mindful of how this may change as Twitch grows beyond the gaming community and gains broader usership. Becoming the most popular live streaming platform around the world may expose Twitch to significantly more large-scale risks, such as those relating to harmful misinformation, online-to-offline violence, election interference, and viral hate speech. More diverse content, larger communities, more recorded content (including content recorded on Twitch and subsequently uploaded elsewhere), and presence in more cultures and languages, combined with the challenges of automated content moderation of live video, will likely increase Twitch's human rights risk profile.

Further, it was noted by many of the stakeholders interviewed by BSR that while human rights risks on the Twitch platform may be of a lower likelihood than other platforms, they may also be of high severity when they occur, such as terrorist recruitment in the context of some types of games (e.g., Viking-themed games) or live streaming of extreme and illegal content, such as Child Sexual Abuse Material (CSAM) or terrorist and violent extremist incidents.

- **There are human rights risks at Twitch that would benefit from stronger management.** This assessment makes recommendations for a more deliberate human rights-based approach to content governance at Twitch, especially related to international growth, the business model shift towards more advertising, and the expansion of recommendation algorithms beyond prominent and trusted creators.

Growth

- **Twitch will need to assess and address the human rights risks associated with location growth, and prepare for content-related risks globally.** It will be important for Twitch to increase content moderator capacity as it scales (both FTE and outsourced) to be able to address content moderation in multiple cultures, contexts, and languages—i.e., to localize policy enforcement effectively, and to develop new policies if needed. Ensuring Twitch has adequate coverage and localization commensurate to usership will be key to avoiding the same mistakes as large platforms have made (such as not deploying sufficient content moderation capacity in high-risk locations as usage grows), and getting ahead of known issues with outsourced moderation, such as ensuring appropriate dialect capabilities and routing of content based on language fluency.
- **Preparation for international growth should look beyond locations that Twitch is “formally” present in (or may choose to invest) by considering unplanned locations where Twitch could become popular.** Twitch should be prepared for the platform to “take off” in different non-US locations than those it is targeting; based on the recent history of social media and user generated content, the absence of evidence of this risk today does not mean it won't arise in the future. Several stakeholders highlighted the importance of “not losing control like Facebook did,” and being alert to the challenges of becoming a more widely used platform in different languages, countries, and cultures. Twitch could be involved with conflict or other offline harm in some of these locations, such as hosting content generated by users that are themselves actors in a conflict (e.g., politicians, influencers, community leaders in Ethiopia, Myanmar, or Ukraine).
- **Twitch will also need to prepare for new human rights risks associated with other changes to the business model and platform.** This might include assessing and addressing the human rights risks associated with (1) the expansion of recommendation algorithms beyond prominent and trusted creators to surfacing smaller creators, (2) the shift toward more targeted advertising and other changes to the business model, (3) the use of Twitch by a wider range of communities, beyond gaming.

Live Streaming and Content Policy

- **Live streaming presents content policy dilemmas.** There are novel content policy challenges raised by live streaming that will benefit from further stakeholder engagement and exploration. For example, the nature of privacy rights for those incidentally captured by live streaming and the definition of private and public spaces (e.g., live streaming a party from a private home) merits further consideration, including human rights principles such as informed consent, especially since options available in other formats (such as face blurring) are unlikely to be available in live streaming. Another dilemma relates to the significance of live streaming during major events—such as protests, conflict, and other gatherings—and the inevitability that violating content will exist in the context of otherwise valuable and important streaming.
- **Live streaming presents content policy enforcement challenges.** One overarching challenge is its immediacy and ephemerality—as opposed to uploaded video, content is “gone” the moment it’s broadcasted, unless it is recorded and uploaded on Twitch or elsewhere. This means that once violating content is viewed, it’s already had some impact. When user reports or automatic detection flag a live stream and it is reviewed, the livestream can be blocked but some level of harm has already occurred. Another enforcement challenge for live streaming is the limited reliability of tools for automated detection, where there is not yet enough advanced machine learning (ML) to effectively detect many types of problematic content in video. Twitch currently only relies on ML tools to flag potential instances of nudity or violence, which work by applying computer vision to screenshots from the live stream; however, there is a high error rate, and all flagged videos must still be human reviewed. This creates particular challenges as Twitch scales and will need to rely more heavily on automated tools for effective moderation.
- **Twitch will need to define how to apply human rights-based principles to the optimization of content policy for belonging and safety, rather than expression.** Over the past decade there has been a significant development in human rights-based approaches for content policy, such as the application of human rights principles (e.g., legality, legitimacy, necessity, and proportionality) for content policy. However, these principles have mainly been applied by social media platforms prioritizing freedom of expression or “voice” as a key value, though the human rights principles appear to be transferable—in other words, are the restrictions on speech necessary, and proportionate for the pursuit of community, belonging, and safety goals.
- **Community moderators will continue to benefit from resources, training, and investment.** The community moderators play an essential role in identifying potentially violating content on chat as well as in live streaming, and the effective achievement of human rights goals will require continued investment in their capacity and capability. Supplementing community moderators during live scale events—i.e., where there is a high volume and / or high visibility of potentially violating content—is one priority.

Due Diligence

- **There are opportunities to integrate human rights into existing risk assessment processes.** For example, the safety-by-design review process (i.e., when product owners and functional experts consider safety implications of design decisions) could be broadened to ensure a deliberate consideration of (1) impact against all potentially relevant human rights, and (2) impact on the most vulnerable users. Data protection impact assessments (DPIAs) required under GDPR requirements present another opportunity as DPIAs require a review against all rights and freedoms contained in the [EU Charter of Fundamental Rights](#) (i.e., not just privacy); the upcoming require-

ments of the Digital Services Act and Artificial Intelligence Act in the EU also present opportunities for integrating human rights into other processes, as both require risk assessments of harm to people and society in a manner very similar to the human rights due diligence requirements of the UNGPs, and both make direct reference to the EU Charter of Fundamental Rights (which contains a very similar list of rights as the International Bill of Human Rights).

- **Twitch has the opportunity to develop a robust stakeholder engagement strategy and approach.** Engagement with affected stakeholders and other experts underpins a human rights-based approach; avoiding engagement fatigue on the one hand while building strong feedback loops into stakeholder engagement on the other will be an important balance to strike, particularly as Twitch scales across communities and geographies and the number of relevant stakeholders grows. The [Twitch Safety Advisory Council](#) provides a strong foundation to build upon.
- **Twitch should prepare for a future that is different than today.** For example, this might include scenarios where extensive human reviewer involvement of all reported or flagged content is not always possible owing to scale, and where dealing with nuanced problematic content across linguistic and cultural contexts (e.g., hate speech, harassment) needs to happen more rapidly and in more places than at present.
- **Twitch can benefit from experimenting with innovative forms of due diligence.** For example, identifying the challenges associated with managing content-related risks as Twitch grows could include integrating scenarios based and futures methodologies. These methodologies are based on a range of different plausible futures and identify the human rights risks, including first, second, and third order effects, of these different futures. [BSR has found](#) that this can help companies [assess risks more expansively and holistically](#) and avoid blind spots and group think.
- **Child rights risks require special attention.** Despite age limits, the reality is that children below 13 will use Twitch, and children between 13 – 18 will use Twitch without the supervision of a parent or legal guardian, as they also do on other platforms.⁶ Age assurance mechanisms may assist in the identification of children accessing or using Twitch; however, very few mechanisms have proven effective to date, and those that are effective also bring risks to children’s right to privacy. Given the difficulties in preventing children from using the platform, Twitch may have to consider instituting policy and platform changes that will address the harms children face on Twitch, and we make recommendations for this in section 7.

Systemwide Approaches

- **Multi-stakeholder collaboration is essential for addressing the numerous systemic challenges in the technology sector that are beyond the ability of any one company to address alone.** Many of Twitch’s human rights risks can only be addressed by multiple organizations working together and by taking system wide approaches. Multi-stakeholder efforts such as the Global Network Initiative (GNI), the Global Internet Forum to Counter Terrorism (GIFCT), the Technology Coalition, and the Digital Trust and Safety Partnership (DTSP) are key places for companies to increase their leverage to address adverse human rights impacts by collaborating with others, while also providing an effective and efficient channel for stakeholder engagement.

⁶ A [2020 study](#) conducted by Thorn found that, “children are more than twice as likely to use platform blocking and reporting tools than they are to tell parents and other caregivers about what happened.” The study reported that 83 percent of 9- to 17-year-olds who reported having an online sexual interaction reacted with reporting, blocking, or muting the offender, while only 37 percent said they told a parent, trusted adult, or peer.

- **Addressing live streaming risks and developing moderation tools and approaches is one such area where Twitch can usefully collaborate with others.** It is noticeable that the multi-stakeholder efforts listed above are only at the early stages of being able to address human rights risks associated with live streaming, and there are opportunities for Twitch to participate in or help create industry working groups to explore issues, topics, and challenges related to live-streaming. During this engagement BSR heard of strong support for Twitch participation in multi-stakeholder efforts, such as sharing content in hash-sharing databases for other platforms to benefit from (e.g., GIFCT hash sharing database), and exploring how to improve machine learning based approaches to content moderation in live streaming.
- **The increased government interest in technology policy has significant implications for Twitch’s human rights strategy.** It is important that the increasingly prominent technology policies, laws, and regulations established by governments are designed and implemented in a manner consistent with human rights. For this reason it will be important for Twitch to have a larger and more globally dispersed public policy team, especially as Twitch scales. Regional public policy teams are also very helpful for identifying and getting ahead of emerging market-specific risks, and can serve as direct points of contact for local external stakeholders.

Access to Remedy

- **Recent changes to the appeals process at Twitch helps access to remedy according to standards set out in the UNGPs.** For example, streamlining and simplifying the appeals process increases accessibility for stakeholder groups, while the appeals portal provides a more predictable (i.e., clear procedure) and transparent (i.e., keeping parties informed about progress) process for users.
- **Further improvements over time will be needed to both the reporting and appeals channels to maintain adherence to expectations set out in the UNGPs.** For example, reviewing the accessibility of reporting channels for users under the age of 13 (who themselves are breaking the terms of service) and users with minority languages or different levels of digital literacy will be important, and considering how non-users (i.e., those without a Twitch account) are able to report content. Twitch may need to prepare for a future where the volume of appeals presents challenges for addressing all of them in a timely manner, and where prioritization of appeals may be needed.

Human Rights Opportunities

- **Twitch has several important human rights opportunities that it can pursue.** While applying the UNGPs primarily means identifying and addressing human rights risks, a human rights strategy should also include areas where Twitch can support the enjoyment, realization, and fulfillment of human rights. While benefits can’t be used to offset harms (i.e., all risks should be addressed, regardless of benefits), some opportunities include:
 - › **Promoting access to culture:** Article 27 of the UDHR states that “everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.” Twitch supports the realization of this right by helping people interact, stream, and create and participate in their own entertainment, including with communities they may not otherwise access. A human rights-based approach would imply seeking to improve the availability and accessibility of Twitch, and promoting high quality content—in addition to addressing risks to the right to access culture, covered below.

- › **Freedom of expression, assembly, and association:** By hosting and helping to build communities, including across borders, Twitch supports the realization of freedom of expression, especially the right to seek, receive and impart information and ideas through any media and regardless of frontiers. Similarly, Twitch supports the realization of freedom of assembly and association, especially as the diversity of content on Twitch increases.
- › **Political participation:** Twitch has the potential to support the right to participate in government, such as politically-oriented channels or streaming, or the spread of political participation into other content, such as gaming (e.g., AOC / get out the vote). This may be especially important for the voices of people and communities who might traditionally be excluded from political participation.

Human Rights Risks



In this section we assess Twitch’s **potential adverse human rights impacts** (i.e., human rights risks) using the criteria described in the Methodology section (section 3, above) to identify and prioritize **salient human rights issues**.

Our assessment of the likelihood of human rights risks includes a consideration of **actual adverse human rights impacts** (i.e., impacts that have already occurred), though it is beyond the scope of this assessment to provide an authoritative list of impacts occurring in the past—BSR did not conduct an investigation or a detailed review of historical content moderation data or cases, for example. BSR’s focus was primarily forward looking on the potential impacts of the future.

This table summarizes BSR’s analysis, with the “highest” possible outcome provided where there is a range (e.g., “largest” when range is “small to largest”).

Rights Area	Scope	Scale	Remediability	Likelihood	Priority
Opinion Expression, and Access to Information	Largest	Serious	Possibly remediable	High likelihood	Tier One
Bodily Security Rights	Large	Most serious	Possibly remediable	Good likelihood	Tier One
Child Rights	Medium	Most serious	Possibly remediable	Some likelihood	Tier One

Rights Area	Scope	Scale	Remediability	Likelihood	Priority
Freedom of Thought, Belief, and Religion	Medium	Most serious	Possibly remediable	Some likelihood	Tier One
Privacy	Large	Serious	Rarely remediable	Some likelihood	Tier One
Participate in Government	Large	Serious	Possibly remediable	High likelihood	Tier Two
Equality and Non-Discrimination	Medium	Serious	Possibly remediable	Good likelihood	Tier Two
Property and Access to Culture	Small	Serious	Remediable	Certain	Tier Two
Education	Medium	Moderately serious	Likely remediable	Minor likelihood	Tier Three
Peaceful Assembly and Association	Small	Moderately serious	Possibly remediable	Some likelihood	Tier Three

Table: Twitch’s Salient Human Rights Issues

Tier One Issues	Tier Two Issues	Tier Three Issues
<ul style="list-style-type: none">› Bodily Security Rights› Child Rights› Right to Freedom of Opinion and Expression, and Access to Information› Right to Freedom of Thought, Belief, and Religion› Right to Privacy	<ul style="list-style-type: none">› Right to Access to Culture and Right to Own Property› Right to Equality and Non-Discrimination› Right to Participate in Government	<ul style="list-style-type: none">› Right to Peaceful Assembly and Association› Right to Education

There are three important points to note about the prioritization of these human rights risks:

- The **ten salient human rights issues** listed here are a narrowed list of potential salient risks that **cross a threshold of relevance for Twitch**. In other words, there are other human rights risks (such as the right to marriage or the right to a nationality) that are already excluded from this list of ten salient human rights issues.
- BSR has further **prioritized among these ten salient human rights issues** using the criteria of severity (scope, scale, remediability) and likelihood described in the methodology (section 3, above). However, BSR urges caution with this prioritization since (1) it 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.
- Under the UNGPs Twitch 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 Twitch’s responsibility to address all adverse human rights impacts.

The recommendations section of this assessment (section 7, below) provides advice on how human rights risks may be avoided, prevented, mitigated, or remediated. We do not list recommendations alongside each risk as in our experience one recommendation may address multiple risks at the same time.

BSR notes that many of the human rights risks listed below would constitute a violation of Twitch policies by Twitch users.

Bodily Security Rights

Human Rights Risks

- Content shared on Twitch may incite, encourage, exacerbate, facilitate, or coordinate physical or psychological harm and offline harm, or otherwise result in safety and security-related risks to people, including terrorism and violent extremism.
- Terrorist and violent extremist events may be streamed live on Twitch.
- Twitch could be used to recruit people into situations of slavery, such as human trafficking, and facilitate trafficking related activities.
- Twitch may be used to depict or facilitate torture or cruel, inhuman, or degrading treatment.
- Twitch users may be arbitrarily arrested for content they streamed or posted on Twitch.

Relevant Human Rights Instruments

- › **UDHR Article 3 and ICCPR Article 9:** Everyone has the right to life, liberty and security of person.
- › **ICCPR Article 6:** Every human being has the inherent right to life. No one shall be arbitrarily deprived of his life.
- › **UDHR Article 5 and ICCPR Article 7:** No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment.
- › **UDHR Article 9 and ICCPR Article 9:** No one shall be subjected to arbitrary arrest, detention or exile.

Assessment of Severity and Likelihood

- › **Scope—Small to Large:** The number of people adversely impacted will be small compared to the size of Twitch's overall user-base; however, those impacted will likely include non-users who are victims of offline harm.
- › **Scale—Serious to Most Serious:** The impacts may be serious, very serious, or most serious for those affected, including death, direct bodily harm, and / or lasting psychological damage.
- › **Remediability—Possibly Remediable:** Depending on the nature of the impact, the harms are possibly remediable. Some harms (such as death or serious bodily harm) may not be, whereas others may be addressed through rehabilitation. Live streamed or recorded content of terrorist and violent extremist events can be removed, and can also provide evidence for use in criminal trials in pursuit of remedy.
- › **Likelihood—High Likelihood:** These impacts have occurred in the past (e.g., Halle livestream) and so are highly likely to occur in the future. However, investments in content moderation are designed to reduce the likelihood of recurrence.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch's existing efforts to identify and remove content are such that Twitch would likely be considered "directly linked to" rather than "contributing to" the harm. However, Twitch is more likely to be considered "contributing to" the harm if it fails to take appropriate reaction when faced with signals of increased risk in specific geographies or channels.
- › **Leverage:** Twitch's leverage is limited by the live streaming nature of the platform and both the speed and the scale of content. However, leverage exists through Community Guidelines enforcement (e.g., reporting channels, automated tools, community moderators), and can be increased by collaborating with others (e.g., GIFCT Incident Protocol; collaboration with governments in counter terrorism and violent extremism efforts).

Vulnerable Populations

Vulnerable groups are more likely to be the targets of violence, and include:

- Historically marginalized ethnic or racial communities
- Religious and linguistic minorities
- Human rights defenders
- Women
- Members of the LGBTQI+ community

These groups could be affected by content shared on Twitch whether they are platform users or not.

Right to Equality and Nondiscrimination

Human Rights Risks

- Content shared on Twitch may encourage, exacerbate, or facilitate discrimination against people.
- Violating content may go undetected if shared in new locations or less common languages, especially during a sudden conflict or political event.
- Twitch's Community Guidelines could be applied unequally or in a discriminatory manner, for example due to moderator bias, language, culture, resource allocation, or through the use of automated tools.
- Twitch streamers (both prominent and non-prominent) may participate in hate groups offline.
- Disabled users may not be able to secure equal access to the platform.

Relevant Human Rights Instruments

- › **UDHR Article 1:** All human beings are born free and equal in dignity and rights.
- › **UDHR Article 2 and ICCPR Article 2:** Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status.
- › **ICERD Article 5:** Entitlement to civil and political rights free of racial discrimination.
- › **CEDAW Article 2:** Entitlement of women to enjoy freedom from discrimination.
- › **CRPD Article 5:** Entitlement of persons with disabilities to enjoy freedom from discrimination.

Assessment of Severity and Likelihood

- › **Scope—Medium:** The proportion of Twitch’s overall user-base potentially negatively impacted by discrimination is significant. Twitch received 10 million reports in the first half of 2021 alone – while this is not a perfect proxy for incidents of discrimination, it shows that a significant portion of Twitch’s population faces harassment of some kind. Discrimination can also follow users to other platforms or offline, depending on spread of content.
- › **Scale—Serious:** The impacts on those affected could be serious, ranging from lasting psychological harm to offline violence in extreme cases.
- › **Remediability—Possibly remediable:** Twitch has a robust reporting channel, and has shored up capacity for response against on-platform discrimination. However, discrimination taking place against communities with low representation on the platform, or else in geographies with less capacity for moderation, has a lower possibility of remediability (for example, due to a lack of familiarity with the reporting process). Discrimination on or spread via Twitch can also manifest offline and affect participation in society, for instance by affecting access to employment; these cases would also have a lower possibility of remedy. Remediation of psychological harm (e.g., mental health) arising from discrimination and harassment is possible but also challenging.
- › **Likelihood—High Likelihood:** The likelihood of occurrence may be greater in geographies and languages with no or fewer content moderators, and / or no or poorly trained classifiers, and in complex languages (such as Arabic) that have multiple and sometimes mutually incomprehensible dialects. It is also more likely in communities known for discriminatory attitudes, such as gaming.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch has taken several important steps to identify and remove content at scale and with speed during live broadcasts; assuming Twitch continues to take proactive and appropriate action, it would be considered “directly linked” to the harm. However, Twitch could be considered “contributing to” the harm if Twitch fails to take appropriate reaction when becoming aware of specific discrimination risks, or does not effectively resource content moderation in growth locations, geographies, or cultures.
- › **Leverage:** Twitch’s leverage is limited by the live streaming nature of the platform and both the speed and the scale of content. However, leverage exists through Community Guidelines enforcement (e.g., reporting channels, automated tools, community moderators), and can be increased by further educating community moderators on how to prevent harms across cultures and geographies.

Vulnerable Populations

Vulnerable groups are at a higher risk of being subject to discrimination on Twitch's platform, as well as the discriminatory impacts of improperly resourced content moderation. These groups include:

- Historically marginalized ethnic or racial communities
- Religious and linguistic minorities
- Women
- Members of the LGBTQI community
- Twitch users or streamers with cultural / linguistic backgrounds poorly represented among moderators

Child Rights

Human Rights Risks

- › Twitch may not be able to verify the age of users, limiting the ability to identify if / when individuals under the age of 13 access or use the platform.
- › Attempts to deploy age assurance mechanisms may violate children's right to privacy.
- › Children may be exposed to hateful, harmful, age-inappropriate, or illegal content, or mis-disinformation. Children may participate or engage in hateful, harmful, illegal, or otherwise problematic user-generated behavior.
- › Children may be the recipients or targets of harmful actions, including online harassment, cyber-bullying, grooming, child sex trafficking, and sextortion.
- › Children may be adversely impacted through their exposure to marketing and advertising on Twitch platforms.
- › Access to and use of Twitch may adversely impact children's mental and physical well-being and / or hinder social skills and development.
- › Children between 13 – 18 may use Twitch without the supervision of a parent or legal guardian.
- › Children below the age of 18 may not understand how or when to use reporting channels, and therefore may not be able to use Twitch's safety features as intended.

Relevant Human Rights Instruments

- › **CRC Article 6:** Children have the inherent right to life.
- › **CRC Article 12:** Children have the right to express their views freely [...] the views of the child being given due weight in accordance with the age and maturity of the child.
- › **CRC Article 13:** Children have the right to freedom of expression; this right shall include freedom to seek, receive, and impart information and ideas of all kinds.
- › **CRC Article 16:** Children have the right to privacy.
- › **CRC Article 17:** Children have the right to access to information and material from a diversity of national and international sources.
- › **CRC Article 19:** Appropriate measures must be taken to protect children from all forms of physical or mental violence, injury or abuse [...], maltreatment or exploitation, including sexual abuse.
- › **CRC Article 31:** Children have a right to rest and leisure, to engage in play and recreational activities appropriate to the age of the child and to participate freely in cultural life and the arts.
- › **CRC Article 32:** Children have the right to be protected from economic exploitation.

Assessment of Severity and Likelihood

- › **Scope—Small to Medium:** Twitch users under the age of 18 represent a relatively small proportion of Twitch's overall user base. This could increase if Twitch / livestreaming "takes off" among young people.
- › **Scale—Serious to Most Serious:** Impacts can range from psychological harm and infringement on childhood development to physical exploitation (e.g., risk to bodily security). International human rights guidance is to always consider risks to children as severe.
- › **Remediability—Possibly Remediable:** Depending on the harm, Twitch can possibly provide remedy through existing reporting channels, especially if they are designed with vulnerable users in mind (see recommendation). However, Twitch does not currently differentiate or prioritize handling reports from between those 13-17 years old and those 18 and older. No recourse is available for children under the age of 13.
- › **Likelihood—Likely:** These impacts have occurred in the past, and thus are likely to occur in the future.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch has taken several important steps to make its platform safe for children; assuming Twitch continues to take proactive and appropriate action on the risks listed above (e.g., implementing the recommendations in this assessment) then it would be considered “directly linked” to the harm. If Twitch does not take proactive action, including action to identify or remove underage accounts, and / or does not design reporting channels for use by users under 18 then Twitch may be considered “contributing to” the harm.
- › **Leverage:** Twitch’s leverage is limited by the live streaming nature of the platform and both the speed and the scale of content. However, leverage exists through Community Guidelines enforcement (e.g., reporting channels, automated tools, community moderators), and can be increased by creating separate channels for reports from underaged users. Leverage can also be increased through participation in collaborative efforts, such as those to address CSAM and sexually exploitative material.

Vulnerable Populations

Children are considered a vulnerable group. Within this category, children may experience other dimensions of vulnerability that increase their risk of harm.

- Children aged 13 – 18
- Children under 13
- Children from Historically marginalized ethnic or racial communities
- Children from religious and linguistic minorities
- Members of the LGBTQI community
- Children from cultural / linguistic backgrounds with limited representation on Twitch or community moderators

Right to Privacy

Human Rights Risks

- The privacy rights of non-users may be violated during live streaming. Twitch could be used in a way which violates the individual right to privacy of individuals who do not give consent to be included in live streams or whose private information is shared without consent.
- Users and non-users may be victims of doxxing and / or the sharing of personal information without consent.
- Governments may make overbroad demands for user data.
- Twitch may collect, utilize, and share user data in ways that are inconsistent with privacy rights.
- Twitch employees may abuse access to user data to violate the privacy rights of individuals (i.e., the insider threat).
- Twitch could suffer from a data breach that results in the exposure of personal data about users.

Relevant Human Rights Instruments

- › **UDHR Article 12, ICCPR Article 17, and CRC Article 16:** Nobody shall be subjected to arbitrary interference of their privacy.

Assessment of Severity and Likelihood

- › **Scope — Small to Very large:** The vast majority of streams on Twitch include individuals consenting to being streamed, which means in general, the privacy of streamers and Twitch users depicted in live streams is well protected. Improper data collection and use practices by Twitch could impact all users. Similarly, a data breach could result in the leakage of data about all users, leading to a very large scope.
- › **Scale — Least serious to Serious:** When the privacy of streamers or other users is intentionally violated, such as by doxxing, harm can escalate to offline stalking, harassment, or violence. Abuse of user data access by Twitch employees can similarly lead to serious harm. Compliance with over-broad government requests can lead to suppression of expression, jail time, and a host of other knock-on effects. Data breaches and improper user data collection and use practices may not have serious consequences, but can result in more significant harm in certain cases.
- › **Remediability — Rarely Remediable:** Depending on the impact, violations of privacy can result in serious, non-remediable harm; for instance, if a streamer has to move or flee their home after being identified on the platform. Similarly, leaked user data once out in the world can rarely be taken back.
- › **Likelihood — Likely:** The broadcasting of individuals without their consent (e.g., from inside private space and property) will increase in likelihood as the Twitch platform becomes used for a wider range of purposes and in a wider range of contexts; the “gray zone” nature of this scenarios likely increases its likelihood too. Twitch’s privacy practices make it unlikely that there will be any significant or widespread improper collection and use of user data; however, this risk grows over time as Twitch expands its advertising and algorithmic curation practices. Twitch has already had severe data breaches, and in the current cyber security context additional breaches are likely to occur.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch’s attribution to privacy-related issues depends on the nature of the harm. For non-consensual streaming, which is prohibited by Twitch Community Guidelines, Twitch may be considered as “directly linked” to the harm; for compliance with over-broad government information requests, Twitch may be considered “contributing to” the harm if it does not have in place effective procedures to identify or challenge overbroad requests. For cases of improper data collection and use or employee abuse of data access, Twitch would “cause” the harm. For data breaches, Twitch would either “cause or contribute” to harm.
- › **Leverage:** Twitch’s Community Guidelines clearly state that violating the privacy of users or non-consenting individuals can result in removal from the platform. Twitch has some leverage to push back on overbroad government requests by demanding proper legal channels and processes be used. Twitch has total leverage over its own privacy and security policies and practices.

Vulnerable Populations

Certain vulnerable groups faced heightened risk of consequences as a result of privacy violations. They include:

- Human rights defenders
- Members of the LGBTQI+ community
- Religious and cultural minorities
- Marginalized racial and ethnic groups
- Women

Freedom of Thought, Belief, and Religion

Human Rights Risks

- Content shared on Twitch may encourage, exacerbate, or facilitate discrimination against people based on their beliefs and religion.
- Content shared on Twitch may incite, encourage, exacerbate, facilitate or coordinate physical or psychological harm and offline harm against people based on their beliefs and religion.
- Attacks on religious groups could inhibit Twitch users from utilizing Twitch to practice their religion and participate in religious community gathering.

Relevant Human Rights Instruments

- › **UDHR Article 18 and ICCPR Article 18:** Everybody has the right to freedom of thought, conscience, and religion.

Assessment of Severity and Likelihood

- › **Scope—Medium:** As a platform with global reach, Twitch hosts users from all cultural and religious backgrounds. Religious or cultural discrimination hosted on Twitch platforms, or else perpetrated off platform by Twitch streamers, has the potential to harm millions of individuals on and off the platform.
- › **Scale—Moderate to Most Serious:** Harms could range from psychological damage to offline harms in extreme cases.
- › **Remediability—Possibly Remediable:** Correct Community Guidelines enforcement constitutes one form of remedy for content-related harms. Discrimination taking place against ethnic communities or members of religion with low representation on the platform, or else in geographies with less capacity for moderation, has a lower possibility of remediability due to unfamiliarity with reporting channels or cultural and language barriers.
- › **Likelihood—Likely:** The likelihood of this risk will likely increase as Twitch becomes popular in new locations and geographies, especially those with existing religious tensions, or where religious groups suffer from discrimination.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch has taken several important steps to identify and remove content at scale and with speed during live broadcasts, and effective implementation of these would result in Twitch being considered “directly linked” to the harm. However, Twitch’s reliance on community moderation to fill in gaps in employed moderation are such that Twitch may be considered “contributing to” the harm, especially if moderation decisions lead to a failure to take appropriate action when dealing with discrimination in under-represented geographies or cultures.
- › **Leverage:** Twitch’s leverage is limited by the live streaming nature of the platform and both the speed and the scale of content. However, leverage exists through Community Guidelines enforcement (e.g., reporting channels, automated tools, community moderators), and can be increased by further educating community moderators on how to prevent harms across cultures and geographies.

Vulnerable Populations

- Religious minorities or marginalized religious groups
- Twitch users or streamers with cultural / linguistic / religious backgrounds poorly represented among moderators

Freedom of Opinion and Expression, and Access to Information

Human Rights Risks

- Twitch may over-enforce its Community Guidelines.
- Twitch may under-enforce its Community Guidelines in ways that hinder the creation of safe spaces for expression
- Misinformation and disinformation may spread on Twitch.
- Governments may make overbroad demands for content removal.

Relevant Human Rights Instruments

- › **UDHR Article 19 and ICCPR Article 19:** Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.
- › **CERD Article 4:** Outlaws the dissemination of ideas based on racial superiority or hatred.

Assessment of Severity and Likelihood

- › **Scope—Very Small to Largest:** Twitch’s moderation decisions affect every single user on their platform. Decisions on which content to regulate or leave up can also impact non-users and extend offline.
- › **Scale—Least Serious to Serious:** Harmful content on Twitch can lead to impacts ranging from serious psychological harm to offline violence and death.
- › **Remediability—Possibly Remediable:** In most cases, infringement on freedom of expression can be remedied by Twitch by restoring uploaded content, live streams, or accounts. However, under especially repressive regimes, or else in conflict contexts, where access to information is heavily limited and can mean life or death, over- or under-moderation on Twitch can lead to non-remediable harms. Similarly, in situations where information sharing is timely, such as during an election or conflict, remediation may come too late.
- › **Likelihood—Very Likely:** The likelihood of this risk will likely increase as Twitch becomes popular in new locations and geographies, especially those with weak rule of law, or else in areas likely to descend into conflict.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch’s existing efforts to identify and remove harmful content are generally appropriate, such that Twitch would likely be considered “directly linked to” rather than “contributing to” the harm. Efforts will need to become more robust as Twitch scales in order for Twitch to avoid “contributing to” harm associated with both over and under enforcement and the spread of mis / disinformation.
- › **Leverage:** Twitch allows communities to set their own moderation guidelines above their baseline, allowing users to find their own balance of expression vs. safety. However, in conflict affected areas or in geographies with repressive regimes, Twitch could also educate users on possible safety measures—for instance, directing users to reputable news sites, etc.

Vulnerable Populations

- Twitch users or streamers with cultural / linguistic backgrounds poorly represented among moderators
- Human rights defenders and other users living in conflict-affected areas, or else in geographies with weak rule of law
- Any member of a group at risk of having policies applied in an inconsistent manner due to moderator bias or unfamiliarity

Right to Own Property and Access to Culture

Human Rights Risks

- Content that violates intellectual property rights may be shared on Twitch.
- Theft of property could be facilitated on Twitch.
- Twitch may inhibit Indigenous groups’ access to culture by not addressing the reproduction / use of their cultural heritage on the platform.

Relevant Human Rights Instruments

- › **UDHR Article 17:** Everybody has the right to own property. Nobody shall be arbitrarily deprived of their property.
- › **UDHR Article 27, ICESCR Article 15 and ICCPR Article 27:** Everybody has the right to practice their culture among their own communities.
- › **UDHR Article 27:** Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary, or artistic production of which they are the author.
- › **UNDRIP Articles 11 and 13:** Indigenous peoples have the right to practice and protect their cultural traditions.
- › **UNDRIP Article 31:** Indigenous peoples have the right to develop and protect intellectual property over their cultural heritage.

Assessment of Severity and Likelihood

- › **Scope—Smallest to Small:** Copyrights for music and television, the most common violations, are generally concentrated among very few owners. Use of indigenous art and culture will affect a slightly larger amount of people, including non users.
- › **Scale—Least Serious to Serious:** Harms from intellectual property are likely not to cause psychological or physical harm in most cases. In the case that an independent or smaller rights holder has their copyrights infringed, harm is still relatively mild. Harms resulting from property theft can be more serious, and include bodily harm.
- › **Remediability—Likely Remediable to Remediable:** Twitch’s takedown policy for DMCA violation is robust. Property theft facilitated on Twitch would be difficult to remediate.
- › **Likelihood—Certain:** DMCA takedown requests are very common on Twitch, and will certainly continue to be.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch’s management of DMCA-related issues is well-handled. Twitch policies are clear, and in case of infringement, Twitch would likely only be “directly linked” to harm. Without an adequate policy to govern use of indigenous art and cultural heritage, Twitch may be considered “contributing to” related harms. Twitch would likely be considered “directly linked” to any harm connected to property theft facilitated on the platform.
- › **Leverage:** Twitch clearly communicates their policies on DMCA violations. Twitch has licensed certain “soundtracks” for streamers to use, in order to provide another avenue for content creators. Twitch has slightly less leverage to address use of indigenous art and culture because it lacks a clear legal IP framework.

Vulnerable Populations

- Groups or individuals lacking the resources required to police copyright violations of their own intellectual property
- Indigenous groups

Right to Peaceful Assembly and Association

Human Rights Risks

- Twitch may over-enforce its Community Guidelines (e.g., suspend streamers and users).
- Twitch may under-enforce its Community Guidelines in ways that hinder peaceful assembly and association.

Relevant Human Rights Instruments

- › **UDHR Article 20 and ICCPR Article 21:** Everybody has the right to freedom of peaceful assembly and association.

Assessment of Severity and Likelihood

- › **Scope—Small:** As a community-oriented platform, Twitch focuses on creating safe spaces for association. The proportion of users who would be impacted on the platform is relatively small.
- › **Scale—Least Serious to Moderate:** Those unable to find a safe space on the platform, especially due to over or under moderation by Twitch, could face psychological harm. Additionally, as livestreaming grows in popularity, those not wanting to be captured on video while protesting or otherwise assembling may be disincentivized from joining movements publicly.
- › **Remediability—Possibly Remediable:** In most cases, Twitch’s robust reporting channels afford users with the ability to appeal moderation decisions. However, many of those who would be discouraged from public association due to Twitch’s gaining popularity would likely be non-users—and Twitch is less likely to be able to provide sufficient remedy for those individuals. Additionally, in time sensitive contexts, remedy may come too late.
- › **Likelihood—Likely:** The likelihood of this risk will increase as Twitch gains popularity and grows into new locations, especially those with existing tensions regarding the right to freely associate with others.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Due to Twitch’s robust appeals process, they would likely be “directly linked” to harms related to policy enforcement dampening freedom of association for groups on the platform.
- › **Leverage:** Twitch’s Community Guidelines prevent streamers from violating the privacy of others. However, clearer communication of Twitch’s policy to those off platform who may be worried about their own privacy may help mitigate any chilling effects on association related to Twitch’s platform.

Vulnerable Populations

- Twitch users or streamers with cultural / linguistic backgrounds poorly represented among moderators
- Human rights defenders and other users living in conflict-affected areas, or else in geographies with weak rule of law

Right to Participate in Government

Human Rights Risks

- Misinformation and disinformation may adversely impact participation in government and elections.
- Governments may retaliate against users and creators for content streamed or posted on Twitch.

Relevant Human Rights Instruments

- › **UDHR Article 21, CERD Article 5, ICCPR Article 25, and CEDAW Articles 7, 8:** Everyone has the right to take part in the government of their country.

Assessment of Severity and Likelihood

- › **Scope—Smallest to Large:** The vast majority of Twitch users live in democracies. As streams relating to political affairs grow in number around the world, more and more individuals will be affected—both on and off the platform. Similarly, as Twitch grows to become a more general purpose live streaming platform, it is likely to host a large amount of civic and election related content.
- › **Scale—Serious:** As Twitch streams include more political content and thus attract a more political audience, disinformation and misinformation will become a bigger concern – especially in the context of ongoing elections. Harms can potentially include targeted disinformation related to voter registration procedures or misinformation related to candidate positions. These impacts can fundamentally alter the result of an election, and can thus result in a slew of knock on effects after the election.
- › **Remediability—Possibly Remediable:** Depending on the impact and the extent to which Twitch alters an election, Twitch may be able to provide remedy for those who experience human rights abuses stemming from election results.
- › **Likelihood—Very Likely:** As seen in the 2020 elections, and across other social media platforms, misinformation and disinformation campaigns are growing in sophistication and breadth. The likelihood that this phenomenon will continue on Twitch is very high.

Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Depending on the extent of the impact Twitch specifically has on an election or political discourse, Twitch would likely be “directly linked” or “contribute to” harm.
- › **Leverage:** Moderation for disinformation and misinformation on live broadcasts at scale is a very difficult problem. While Twitch’s new misinformation policy outlaws the spread of harmful misinformation on the platform, it does not define what constitutes harmful, which means Twitch lacks a clear framework for drawing lines as new forms of misinformation emerge. One definition of harmful could be “misinformation with a risk of resulting in significant harm to human rights.”

Vulnerable Populations

Certain vulnerable groups are more likely to be disproportionately by election-related information:

- People living in contexts without a robust free media
- People lacking digital literacy

Right to Education

Human Rights Risks

- Content may be streamed or hosted on Twitch with claims that it originates from legitimate educational bodies, but it isn't.

Relevant Human Rights Instruments

- › **UDHR Article 26, CEDAW Article 10, and ICESCR Articles 13 and 14:** Everyone has the right to an education.
- › **CRC Article 28 and 29:** Every child has the right to quality education, directed towards the development of their fullest potential.

Assessment of Severity and Likelihood

- › **Scope—Small to Medium:** The student-aged population on Twitch is a relatively significant proportion of Twitch’s user base.
- › **Scale—Least Serious to Moderate:** The impacts of disseminating false education content depends on the nature of the content shared, but in general, harms would most likely be limited to possible psychological harm.
- › **Remediability—Likely Remediable:** Twitch’s reporting channels serve as an effective method for removing harmful content and providing remedy to the rightsholder. However, Twitch’s policies do not address this risk.
- › **Likelihood—Unlikely:** Educational content makes up a relatively small proportion of Twitch’s content base. False educational content is even more rare – more importantly, education is not a huge driver for Twitch views at this time. This likelihood may increase, however, as Twitch evolves and continues to expand across geographies.

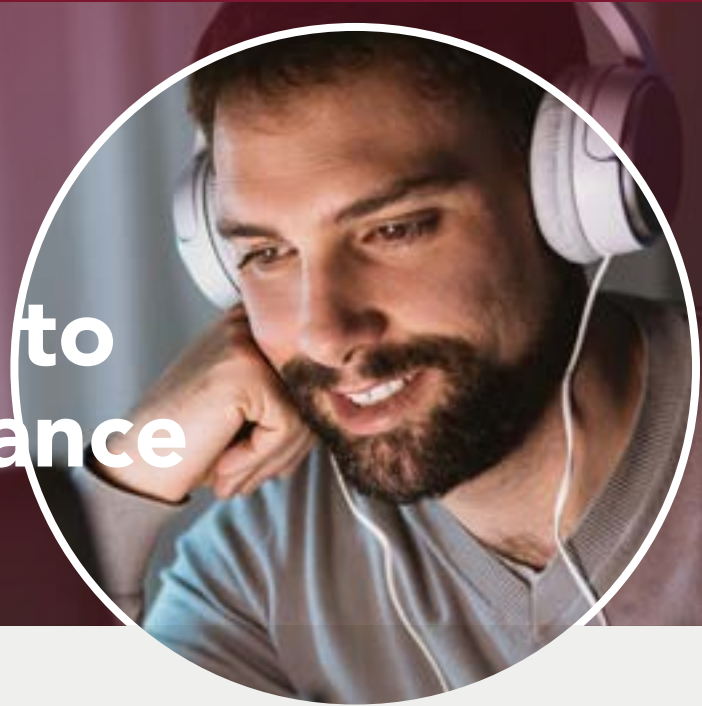
Assessment of Appropriate Action

- › **Attribution (Cause, Contribute, or Directly Linked):** Twitch’s existing efforts to identify and remove content are such that Twitch would likely be considered “directly linked to” education-related harms.
- › **Leverage:** Twitch’s harmful misinformation policy and impersonation policy likely would include enforcement on falsified educational materials. While the definition of harmful is currently undefined in the policy, it seems clear enough to be applied successfully in the educational context.

Vulnerable Populations

- Children

Human Rights Based Approach to Content Governance for Twitch



We have used BSR's paper [A Human Rights Based Approach to Content Governance](#) to shape our recommendations for Twitch (section 7, below). This paper is segmented into four parts:

- **Content policy**—statements about what content is and is not allowed on a social media platform, as well as about the visibility of content. Content policy should:
 - › **Encompass all human rights.** BSR has undertaken a gap analysis for Twitch on this; the results are contained in the annex, and covered by recommendations, where BSR proposes potential additions to the Twitch Community Guidelines (recommendations 1 – 3).
 - › **Be founded upon human rights standards and instruments.** BSR has made a recommendation that Twitch adopt a human rights policy (recommendation 4).
 - › **Be informed by stakeholder engagement.** Twitch's PATH process already encompasses this, and BSR has made recommendations for continuous improvement in stakeholder engagement methodology (recommendation 7).
- **Content policy implementation**—how content decisions are executed in practice. Content policy implementation should:
 - › **Be informed by engagement with affected stakeholders and experts that understand the relevant context.** Twitch's Safety Advisory Council already encompasses this, and BSR has made recommendations for continuous improvement (recommendation 7).
 - › **Seek ways to honor the principles of internationally recognized human rights when faced with conflicting requirements.** BSR has recommended that Twitch join the Global Network Initiative and adopt their principles (recommendation 20) and invest in government affairs and public policy capacity (recommendation 19).
 - › **Prioritize based on severity to affected stakeholders and pay heightened attention to conflict-affected areas.** BSR has made recommendations relating to Twitch's global content moderation efforts, especially in higher risk locations (recommendations 5 – 6) and children (recommendations 10, 13, and 23).
 - › **Use appeals mechanisms that meet minimum effectiveness criteria established by the UNGPs, and provide effective remedy when mistakes are made.** BSR has made several recommendations to further enhance Twitch's appeals channels (recommendations 11 – 13).

- **Product development**—how new features, services, and functionalities are introduced and evolve. Product development should:
 - › **Include human rights assessments of new or evolving features, services, and functionalities.** BSR has made recommendations for the integration of human rights into Twitch’s existing assessment processes (recommendation 15, and annex) and made specific product-related recommendations (recommendations 15 – 17).
 - › **Assume that new or evolving features, services, and functionalities may have different, unintended, or more severe consequences in higher risk locations.** BSR has made recommendations for increasing preparedness for scenarios where Twitch use grows in new locations (recommendations 5 – 6).

- **Tracking and transparency**—how the outcomes and effectiveness of a human rights-based approach is measured and communicated. Tracking and transparency should:
 - › **Publish annual public reports that provide sufficient quantitative and qualitative information to evaluate the adequacy of their approach.** BSR has made recommendations for the further improvement of Twitch’s transparency reporting, such as adding appeals data. (recommendation 25)
 - › **Be transparent about the rationale for important content decisions.** Twitch already provides proactive communication around content policy decisions, but BSR has made recommendations for the further improvement to both transparency reporting and communications with users (recommendations 11 and 25).



Recommendations

BSR makes the following recommendations for how Twitch should avoid, prevent, mitigate, and remediate the human rights risks associated with its platform. The BSR recommendations are divided into the following five segments: (1) content policy; (2) implementation of content policy; (3) product development; (4) system wide approaches; (5) tracking and transparency.

Current Policy

RECOMMENDATION 1

Address the gaps identified in the human rights gap analysis BSR undertook of the Twitch Community Guidelines.

Twitch can explore whether additional content policies are needed in the following areas:

- Definition of harm (e.g., a taxonomy of harm).
- Prohibition of human trafficking (e.g., sex trafficking).
- Respect for cultural heritage (e.g., content containing indigenous people's art and culture).
- Educational content (e.g., misrepresentation of formal accreditation or licensing).

BSR notes that human trafficking is already prohibited (e.g., "Breaking the Law"; "Spam, Scams, and Other Malicious Conduct"; "Nudity, Pornography, and Other Sexual Content"), but we believe it would benefit from clearer expression and specificity.

An additional important gap relating to privacy is covered in the next recommendation.

EXPLANATION

As described in BSR's [Human Rights-Based Approach to Content Governance](#), the highly diverse mix of content posted on social media platforms means that any human rights contained in international human rights instruments can be impacted by user-generated content, and for this reason content policy should encompass all human rights.

Principle 16 of the UNGPs states that, as the basis for embedding their responsibility to respect human rights, companies should express their commitment to meet this responsibility through a statement of policy, and work towards policy coherence in their wider activities.

Principle 18 of the UNGPs states that companies should include all internationally recognized human rights as a reference point, since they may potentially impact any of them.

RECOMMENDATION 2

Explore whether new content policies are needed relating to privacy rights during live streaming.

There are novel content policy challenges raised by live streaming that will benefit from further stakeholder engagement and exploration.

For example, the nature of privacy rights for those incidentally captured by live streaming and the definition of private and public spaces (e.g., live streaming a party from a private home) merits further consideration, including human rights principles such as informed consent.

The risk of privacy violations during live streaming will likely grow as new devices and more mobile devices are used for streaming (e.g., smart phones, glasses, XR), and as the content of live streaming diversifies over time.

BSR reviewed content policies at other companies (e.g., YouTube, Facebook) and did not find examples referencing the privacy of non-streamers (or “inadvertent participants”) in a livestream—hence the recommendation to explore further.

EXPLANATION

Article 12 of the UDHR and Article 17 of the ICCPR state that no one shall be subjected to “arbitrary interference with his privacy, family, home or correspondence.”

Stakeholders and experts interviewed for this assessment highlighted this risk as emerging, lacking in consensus, and requiring further exploration.

As described in BSR’s [Human Rights-Based Approach to Content Governance](#), the highly diverse mix of content posted on social media platforms means that any human rights contained in international human rights instruments can be impacted by user-generated content, and for this reason content policy should encompass all human rights.

Principle 16 of the UNGPs states that, as the basis for embedding their responsibility to respect human rights, companies should express their commitment to meet this responsibility through a statement of policy, and work towards policy coherence in their wider activities.

RECOMMENDATION 3

Explore whether exceptions to content policy should exist during certain live events.

A dilemma exists relating to the significance of live streaming during major events—such as protests, conflict, and other gatherings—and the inevitability that violating content will exist in the context of otherwise valuable and important streaming. Twitch may need additional content policy to address these scenarios.

EXPLANATION

According to UN General Comment 34, restrictions to expression should be necessary and proportionate—i.e., only restricted when the same goal cannot be achieved by other means, and using restrictions that are the least intrusive to achieve the legitimate purpose.

Principle 23 of the UNGPs deals with issues of context, and states that “although particular country and local contexts may affect the human rights risks of an enterprise’s activities and business relationships, all business enterprises have the same responsibility to respect human rights wherever they operate. Where the domestic context renders it impossible to meet this responsibility fully, business enterprises are expected to respect the principles of internationally recognized human rights to the greatest extent possible in the circumstances, and to be able to demonstrate their efforts in this regard.”

RECOMMENDATION 4

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

While a part of Amazon (and therefore covered by [Amazon’s Global Human Rights Principles](#)), Twitch’s human rights impacts are distinct enough to merit its own human rights policy—for example, Amazon’s human rights principles are geared towards Amazon’s role as an employer, whereas Twitch’s policy would be geared more towards content governance issues. This policy could reference the Twitch Community Guidelines, and explain how they form part of an overall human rights-based approach.

Further, Twitch should consider assigning a human rights lead to drive the implementation of the human rights policy and oversee the ongoing identification and addressing of human rights risks at Twitch.

EXPLANATION

Principle 16 of the UNGPs states that “as the basis for embedding their responsibility to respect human rights, business enterprises should express their commitment to meet this responsibility through a statement of policy.”

Principle 19 of the UNGPs states that “the horizontal integration across the business enterprise of specific findings from assessing human rights impacts can only be effective if its human rights policy commitment has been embedded into all relevant business functions. This is required to ensure that the assessment findings are properly understood, given due weight, and acted upon.”

Content Policy Implementation

RECOMMENDATION 5

Assess and address human rights risks associated with formal location growth prior to moving forward.

BSR notes that Twitch does not currently have a formal international growth plan specifying locations, but this recommendation has been retained in case one is adopted at a future date.

Twitch should identify and prepare for safety-related issues prior to formally expanding to new locations, rather than reactively responding to problems.

This should include an assessment of how Twitch is likely to be used in a given location, user groups / demographics, existing types of societal discrimination and social tensions that may appear on the platform, the political context, and the governments history related to requests for data and content takedowns.

EXPLANATION

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 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).”

RECOMMENDATION 6

Establish a plan for effective content moderation in the likelihood that the Twitch platform becomes more widely used in more cultures and languages.

Twitch already has content moderation capability for around 20 languages, and the community-based approach to content moderation allows for improved understanding of culture, nuance, and language for any particular channel. However, Twitch would benefit from a plan that incorporates the following:

- Languages where improved content moderation capacity is likely to be needed over time, including a consideration of dialect (e.g., the multiple dialects of Arabic) and the distinction between text and audio (e.g., moderators may be able to comprehend text but not audio, or vice versa).
- Ability to anticipate upcoming risks (e.g., elections, major events) and scale rapidly if required (e.g., conflict).
- Estimates of scale (e.g. FTE) and reach (e.g., cultures, countries) needed.
- Consideration of staff location, and balancing (1) the benefits of content moderation staff “close” to the users with (2) the increased risks to freedom of expression and privacy that can arise from in-country presence (e.g., “hostage taking risk” for FTEs).
- Appropriate use of automated tools to help manage moderation at scale.

Twitch can also use futures methodology to help prepare for and stress test its content moderation policies, process, and structure against plausible future scenarios.

EXPLANATION

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 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.”

RECOMMENDATION 7

Establish a strategy for ongoing stakeholder engagement.

This strategy should be informed by best practices in stakeholder engagement, and can build upon the foundation made by the Safety Advisory Council to include a wider range of stakeholders.

Engagement can inform:

- Development of content policy, such as building upon the existing Policy Authorization Process (PATH).
- Implementation of content policy, such as engaging with affected stakeholders and experts in higher risk locations, groups, or communities.
- Location risks.
- Vulnerable groups risks.
- Effectiveness of reporting and appeals systems (i.e., operational grievance mechanisms).

Best practices in a social media context include 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 multi-stakeholder forums where possible).

EXPLANATION

Principle 18 of the UNGPs states that human rights assessment should “involve meaningful consultation with potentially affected groups and other relevant stakeholders.”

Principle 20 of the UNGPs states that when tracking the effectiveness of their human rights approach, companies should “draw on feedback from both internal and external sources, including affected stakeholders.”

Principle 31 of the UNGPs states that operational grievance mechanisms should be “based on engagement and dialogue: consulting the stakeholder groups for whose use they are intended on their design and performance, and focusing on dialogue as the means to address and resolve grievances.”

RECOMMENDATION 8

Centralize and enhance training for community moderators, and consider adding dedicated moderators for large public livestreams.

BSR recommends that Twitch holds dedicated community moderation training sessions, focusing on sensitively applying their channel's policies across geographies and cultures. For example, this might include training on transphobia, gender, or different forms of hate speech.

Training could be targeted at community moderators based on the frequency, severity, and type of content violations found in channels.

Additionally, BSR recommends that Twitch staff experienced moderators for public facing-streams (e.g. politicians or celebrities unfamiliar with the platform).

EXPLANATION

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 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.”

RECOMMENDATION 9

Create training modules on relevant content-related topics.

Twitch has an opportunity to provide training resources to creators, streamers, and content moderators on topics where violating or “borderline” content is prevalent in gaming communities (and beyond), such as transphobia, gender, and specific forms of hate speech.

EXPLANATION

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

Develop a comprehensive child rights strategy for Twitch.

Twitch currently has policies and enforcement practices in place to address actual and potential adverse impacts to children; however, these measures focus primarily on identifying and addressing risks to child safety.

Twitch should assess and address the full spectrum of child rights (including children's right to freedom of expression and access to information, right to rest and leisure and participation in arts and culture) that it may be involved with.

This includes taking further action to address risks associated with users under the age of 13. Despite age limits, the reality is the children below 13 will use Twitch, and children between 13 – 18 will use Twitch without the supervision of a parent or legal guardian.

BSR recommends that Twitch consult with or engage experts and children to conduct a child rights impact assessment of the platform and use the findings and recommendations from the report to develop a comprehensive strategy on child rights.

EXPLANATION

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.”

In accordance with the UNGPs, Twitch should ensure that it has appropriately assessed the risks to these users and takes action to avoid, prevent, and mitigate harm; this requires action beyond stating that users under the age of 13 are not allowed on the platform.

RECOMMENDATION 11

Increase the number of scenarios where users submitting reports and appeals receive responses.

At present, users reporting potentially violating content do not hear back if (1) action was taken for a different reason than the one reported, (2) action had already been taken for the same content reported, (3) the report relates to alleged criminal behavior, or (4) no action was taken.

BSR recommends that Twitch review the consistency of this approach against expectations set out in Principle 31 of the UNGPs, which suggest more communication with rightsholders.

EXPLANATION

Principle 31 of the UNGPs states that operational grievance mechanisms should be “predictable” (providing a clear and known procedure with an indicative time frame for each stage, and clarity on the types of process and outcome available and means of monitoring implementation) and “transparent” (keeping parties to a grievance informed about its progress, and providing sufficient information about the mechanism’s performance to build confidence in its effectiveness and meet any public interest at stake).

Principle 31 also states that “communicating regularly with parties about the progress of individual grievances can be essential to retaining confidence in the process.”

RECOMMENDATION 12

Investigate feasibility of options for non-user reporting.

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.

BSR recommends that Twitch review the consistency of this approach against expectations set out in Principle 31 of the UNGPs, which suggest availability for all rightsholders.

EXPLANATION

Principle 31 of the UNGPs states that “a grievance mechanism can only serve its purpose if the people it is intended to serve know about it, trust it and are able to use it.”

RECOMMENDATION 13

Review and update reporting channels with vulnerable users, including those under 18, in mind.

Existing Twitch reporting channels may be difficult for certain vulnerable users to access or use, either because they are unsure of how to file a report or because the language and design of the reporting channel is not friendly for all users. BSR recommends that Twitch consult or engage with experts to review and update reporting channels to ensure that they are visible, easily discoverable, recognizable, and accessible to all users—including Twitch’s most vulnerable users such as those under the age of 18, with a clear infrastructure and established process to ensure speedy review and appropriate action.

BSR interprets “under 18” to include (1) users between the ages of 13 and the age of legal majority in the jurisdiction of residence using Twitch services under the supervision of a parent or legal guardian, and (2) users under the age of 13 that are using Twitch services in violation of Twitch Terms of Service.

EXPLANATION

Principle 29 of the UNGPs states that “To make it possible for grievances to be addressed early and remediated directly, business enterprises should establish or participate in effective operational-level grievance mechanisms for individuals and communities who may be adversely impacted.”

Principle 31 of the UNGPs specifies that State-based and non-State-based operational grievance mechanisms should be: legitimate; accessible; predictable; equitable; transparent; rights-compatible; and a source of continuous learning.

RECOMMENDATION 14

Continue seeking to detect discernable patterns of harmful conduct and content.

Detecting discernable patterns of harmful conduct and content (e.g., indicators of grooming, violent extremism, or spam) can assist with the identification of users, accounts, and channels more likely to violate Twitch Community Guidelines, and therefore the prioritization of resources.

However, Twitch should also be attentive to the privacy and non-discrimination risks associated with this approach (e.g., identification of false positives, collection and analysis of new types of user data), and exercise caution on actions taken based on these patterns, especially in the absence of violating content.

EXPLANATION

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.”

Product Development

RECOMMENDATION 15

Integrate human rights into existing risks assessment processes.

Twitch can enhance its human rights due diligence by integrating a human rights-based approach (e.g., consideration of all internationally recognized human rights as a reference point; identification of vulnerable groups; identification of appropriate action to address human rights risk) into existing risk assessment processes.

BSR highlights in particular Twitch’s “safety-by-design” approach (which could become a “human rights-by-design” approach) and Data Protection Impact Assessments (required under the GDPR), as well as assessments that may be required in the EU with the Digital Services Act and AI Act.

As examples, this human rights due diligence approach may assess the human rights risks associated with growing in new locations, introducing new product features (e.g., encrypted private chat), or introducing new products (e.g., paid contracts with creators, rather than ad-based income). It is important that these assessments include impacts on both users and non-users.

BSR provides sample human rights assessment resources in the annex.

EXPLANATION

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 17 of the UNGPs states that 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 18 of the UNGPs states that “in order to gauge human rights risks, business enterprises should identify and assess any actual or potential adverse human rights impacts with which they may be involved.”

Principle 18 of the UNGPs states that “while processes for assessing human rights impacts can be incorporated within other processes such as risk assessments or environmental and social impact assessments, they should include all internationally recognized human rights as a reference point, since enterprises may potentially impact virtually any of these rights.”

RECOMMENDATION 16

Assess and address the human rights risks associated with the expansion of recommendation algorithms beyond prominent, trusted creators to surfacing small creators.

Currently, algorithmic curation poses a low risk of surfacing harmful content because it is only used for trusted streamers, who are highly unlikely to violate Twitch's Community Guidelines.

However, as Twitch plans to utilize algorithmic curation to surface smaller curators, this increases the risk that Twitch will recommend problematic / harmful content to users. Twitch should assess this risk and identify ways to mitigate it.

EXPLANATION

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 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).

RECOMMENDATION 17

Assess risks unique to mobile streaming.

Twitch does not currently place any limits (e.g. subscriber-count, prior watch-hour requirements) on the ability to stream from a mobile device.

There are nuanced and complex pros and cons to adding friction to mobile streaming—for example, friction may deter streaming of potentially violating content, but may also discourage "positive" content more than it deters content posted by bad actors determined to evade controls. It is for this reason that BSR recommends a review of potential minimum requirements before mobile devices (or similar) can be used for streaming, rather than advocating one way or another.

This review should pay particular attention to the safety of children, especially those under the age of 13, who may use mobile devices to evade parental supervision.

EXPLANATION

Article 17 of the Child Rights Convention emphasizes "the development of appropriate guidelines for the protection of the child from information and material injurious to his or her well-being."

Principle 5 of the Child Rights and Business Principles states that companies should "ensure that products and services are safe, and seek to support children's rights through them."

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.”

RECOMMENDATION 18

Assess and address the human rights risks associated with targeted advertising.

Targeted advertising comes with a wide range of human rights risks, including privacy, discrimination, freedom of thought, and the amplification of harmful mis / disinformation.

Twitch should identify and address the risks that are likely to arise from its shift toward more advertising.

EXPLANATION

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 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).”

System Wide Approaches

RECOMMENDATION 19

Increase the general capacity and human rights capability of government affairs and public policy teams.

Governments are increasingly establishing policies, laws, and regulations for online platforms that have significant impacts on human rights, both positive and negative. It is important that these policies, laws, and regulations are consistent with international human rights law and address the human rights issues associated with live streaming platforms, not just other types of social media platform.

BSR recommends that Twitch increase the capacity and capability of government affairs and public policy teams to address this challenge, including advocacy for rights-respecting approaches to government policy, laws, and regulations. These staff should be appropriately trained in human rights.

Europe and the United Kingdom are obvious locations to start with given how much these jurisdictions are shaping the future of technology policy, but a presence in Asia (e.g., Singapore, Tokyo) would help address developments in different countries across the region, including Hong Kong and India.

EXPLANATION

Principle 16 of the UNGPs states that human rights policy should be “supported by any necessary training for personnel in relevant business functions.”

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 20

Join the Global Network Initiative (GNI).

BSR recommends that Amazon and / or Twitch join the GNI—perhaps initially as an observer in order to assess suitability and relevance.

Twitch’s participation in the GNI would involve collaborating with other companies, civil society organizations, investors, and academics to help address situations where government action places the freedom of expression and privacy rights of Twitch users at risk.

EXPLANATION

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.

Many of the human rights risks highlighted in this assessment cannot be addressed by Twitch alone, but instead require system wide and multi-stakeholder approaches.

RECOMMENDATION 21

Become a more active participant in the Global Internet Forum to Counter Terrorism.

Twitch’s participation in GIFCT should be designed with three outcomes in mind:

- **Technology:** Assisting with the development of machine learning, computer vision, and other technology-based approaches to counter terrorist and violent extremist activity in live streaming, as well as contributing relevant hashes of live content to the GIFCT hash sharing database.
- **Incident response:** Collaborating with other technology companies to enable more nimble, effective, and agile responses to incidents as they occur, including both the Hash-sharing Database and Content Incident Protocol.
- **Learning:** Contributing to learning about trends in terrorism and violent extremism in the context of gaming so that these threats can be more effectively addressed.

EXPLANATION

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.

Many of the human rights risks highlighted in this assessment cannot be addressed by Twitch alone, but instead require system wide and multi-stakeholder approaches—for example, sharing hashes so that violating content originating on Twitch can be identified on other platforms, and vice versa.

There are no industry best practices for detecting terrorist and violent extremist content in live streaming, and participation in GIFCT would be in service of this goal (e.g., technology tools to help humans take action against violating content).

RECOMMENDATION 22

Fund or participate in external research into the potential link between gaming, terrorism, and violent extremism.

The link between gaming, terrorism, and violent extremism is often assumed to exist but isn't fully understood, and assertions made about this link are typically anecdotal and qualitative, rather than evidence based and quantitative. The actual measurement of scale, impact, or prevalence is very underdeveloped, and this gap can result in ill-informed policy making and narrative on matters relevant to Twitch and Twitch users.

Funding or participating in research on this topic would assist the achievement of Twitch's content policy goals, while making a contribution to the overall field.

EXPLANATION

This recommendation is derived from the stakeholder and expert engagement undertaken to inform this assessment.

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 23

Collaborate with civil society actors and industry associations (such as the Technology Coalition) to conduct research and develop rights-based approaches to age assurance.

Age verification and age assurance mechanisms have been challenging across the industry. Existing age assurance / age verification mechanisms are often ineffective or risk violating children's right to privacy. More research and collaboration is needed on how to take a rights-based approach to age verification.

BSR recommends working with civil society experts and existing industry coalitions to explore new and innovative solutions to age assurance and actively contributing to the field's understanding of risk to users under the age of 18 and ability to address these risks with a human rights lens.

EXPLANATION

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. Leverage may be increased by, for example, offering capacity-building or other incentives to the related entity, or collaborating with other actors.”

RECOMMENDATION 24

Consider joining the Digital Trust and Safety Partnership (DTSP).

The DTSP exists to develop industry best practices on trust and safety, verified through internal and independent third-party assessments, to ensure consumer trust and safety when using digital services.

Twitch's participation should be geared towards informing industry-wide best practices that incorporate the unique challenges faced in a live streaming context.

EXPLANATION

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.

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.”

Many of the human rights risks highlighted in this assessment cannot be addressed by Twitch alone, but instead require system wide and multi-stakeholder approaches.

Tracking and Transparency

RECOMMENDATION 25

Undertake continuous improvement on Twitch's transparency reporting, including through the use of international reporting standards.

Twitch's transparency report is already highly developed. Improvements over time could include:

- Data for appeals.
- More insights into the governance of content policy creation (e.g., PATH process) and enforcement.

- More information about the nature of government requests, particularly when they are overbroad.
- Reviewing and potentially using proposed standards for transparency reporting (e.g., SASB, OECD).
- Quantitative and qualitative review of specific cases and incidents to share insights gained and lessons learned.

EXPLANATION

BSR received positive feedback on Twitch existing transparency during the assessment; these recommendations are made in the spirit of continuous improvement.

Principle 21 of the UNGPs states that “In order to account for how they address their human rights impacts, business enterprises should be prepared to communicate this externally, particularly when concerns are raised by or on behalf of affected stakeholders.”

Annex 1: Community Guidelines Gap Analysis



This gap analysis assesses the extent to which Twitch Community Guidelines address all human rights in the Universal Declaration of Human Rights (UDHR) and other international human rights instruments, to the extent they are relevant for Twitch content policy. These instruments include:

- International Covenant on Civil and Political Rights (ICCPR)
- International Covenant on Economic, Social, and Cultural Rights (ICESCR)
- International Convention on the Elimination of Racial Discrimination (ICERD)
- Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW)
- Convention against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment (CAT)
- Convention on the Rights of Persons with Disabilities (CRPD)
- Convention on the Rights of the Child (CRC)
- ILO Core Conventions
- UNESCO Convention Against Discrimination in Education
- UN Declaration on the Rights of Indigenous Peoples (UNDRIP)

Several rights that BSR assessed were not directly relevant for the Twitch Community Guidelines, and therefore do not need to be addressed. However, these rights may still be impacted by Twitch in other ways, such as through Community Guidelines enforcement. These rights include:

- Right to recognition as a person before the law
- Right to equality before the law
- Freedom from arbitrary arrest, detention, or exile
- Right to a fair trial
- Right to be innocent until proven guilty
- Freedom of movement

- Right to asylum
- Right to nationality
- Right to marriage and family
- Right to social security
- Labor rights
- Right to rest and leisure
- Right to an adequate standard of living

This gap assessment is intended to proactively identify places where Twitch’s human rights risks could be addressed through additional content in the Twitch Community Guidelines.

This gap assessment suggests that Twitch has good coverage overall, but there are a few areas that merit further consideration—notably, content that (a) is intended to be educational, (b) facilitates trafficking, (c) may violate the privacy rights of bystanders in a livestream, and / or (d) contains indigenous people’s art and culture. Additionally, Twitch could provide further clarification to address ambiguities in several policies, such as the violence and threats policy and the harmful misinformation policy.

Right to Equality and Nondiscrimination

Related Human Rights

UDHR Art. 1, 2 • ICCPR Art. 3, 26 • ICESCR Art. 3 • UDHR Art. 13 • CEDAW Art. 3, 15, 16 • CRPD Art. 5

Relevance to Twitch

- Twitch could be used to facilitate or incite discrimination against people.
- Twitch’s Community Guidelines could be applied unequally or in a discriminatory manner.

Related Twitch Community Guidelines Content

Existing Policy:

- The “Breaking the Law” policy requires all users to “respect all applicable local, national, and international laws while using our services.”
- The “Hateful Conduct and Harassment” policy prohibits hateful content, which is defined as “any content or activity that promotes or encourages discrimination, denigration, harassment, or violence based on the following protected characteristics: race, ethnicity, color, caste, national origin, immigration status, religion, sex, gender, gender identity, sexual orientation, disability, serious medical condition, and veteran status. We also provide certain protections for age.” It also states that this policy is applied equitably: “We afford every user equal protections under this policy, regardless of their particular characteristics.”

- The “Harmful Misinformation Actors” policy states that Twitch will remove users whose presence is dedicated to persistently sharing misinformation, including “misinformation that targets protected groups.”
- The “Account Usernames and Display Names” policy prohibits account names that violate Community Guidelines, including usernames and display names that include hateful conduct, harassment, and sexual harassment.
- Twitch’s Off-service Conduct policy enforces against users found to be engaging in severe offenses outside of Twitch, hate group membership.

Gaps: No major gaps identified

Bodily Security Rights

Related Human Rights

UDHR Art 3, 4, 5 • ICCPR, Art. 6, 7, 8, 9, 20 • CRPD Art. 10, 14, 15, 16 • CRC Art. 6, 11, 19, 21, 34, 37 • UNDRIP Art. 7 • CAT Art. 13,14 • ILO C29, C105, C138, C182 • Article 51 of the UN Charter

Relevance to Twitch

- Twitch could be used in ways that encourage, exacerbate, facilitate, or incite physical or psychological harm, or otherwise result in safety and security-related risks to people.
- Twitch could be used to recruit people into situations of slavery, such as human trafficking, and facilitate trafficking related activities.
- Twitch may be used to depict or facilitate torture or cruel, inhuman, or degrading treatment.

Related Twitch Community Guidelines Content

Existing Policy:

- The “Breaking the Law” policy requires all users to “respect all applicable local, national, and international laws while using our services, Any content or activity featuring, encouraging, offering, or soliciting illegal activity is prohibited. This includes committing or aiding in the malicious destruction, defacement, or theft of public or another person’s private property without permission on stream.”
- The “Self Destructive Behavior” policy states, “Activity that may endanger your life, lead to your physical harm, or encourage others to engage in physically harmful behavior is prohibited. This includes, but is not limited to: suicide threats, glorification or encouragement of self-harm, intentional physical trauma, illegal use of drugs, illegal or dangerous consumption of alcohol, and dangerous or distracted driving.”
- The “Violence and Threats” policy prohibits acts and threats of violence as zero-tolerance violations, meaning accounts associated with violence and threats will be indefinitely suspended. In the policy, violence and threats includes but is not limited to: attempts or threats to physically harm or kill others, attempts or threats to hack, DDOS, or SWAT others, and the use of weapons to physically threaten, intimidate, harm, or kill others.

- The “Violence and Threats” policy prohibits content that “depicts, glorifies, encourages, or supports terrorism, or violent extremist actors or acts. This includes threatening to or encouraging others to commit acts that would result in serious physical harm to groups of people or significant property destruction. You may not display or link terrorist or extremist propaganda, including graphic pictures or footage of terrorist or extremist violence, even for the purposes of denouncing such content.”
- The “Violence and Threats” policy states, “In exceptional circumstances, we may preemptively suspend accounts when we believe an individual’s use of Twitch poses a high likelihood of inciting violence. In weighing the risk of harm, we consider an individual’s influence, the level of recklessness in their past behaviors (regardless of whether any past behavior occurred on Twitch), whether or not there continues to be a risk of harm, and the scale of ongoing threats.”
- The “Hateful Conduct and Harassment” policy prohibits hateful conduct, which can lead to physical violence against targeted groups and individuals.
- The “Hateful Conduct and Harassment” policy prohibits harassment, which can lead to physical safety and psychological well being risks. Harassment is defined as “stalking, personal attacks, promotion of physical harm, hostile raids, and malicious false report brigading.” Sexual harassment is included under this policy, and is defined as, “unwelcome sexual advances and solicitations, sexual objectification, or degrading attacks relating to a person’s perceived sexual practices.”
- Sharing of personal information can lead to physical safety risks for the affected parties, as well as the risk of harassment. The “Unauthorized Sharing of Personal Information” policy prohibits the sharing of “content that may reveal private personal information about individuals, or their private property, without permission.” The policy defines personal information as including but not limited to, PII, restricted or protected social profiles, or content that violates a person’s reasonable expectation of privacy, such as streaming from a private space without permission.
- Impersonation can lead to physical safety risks and psychological harm for the impacted persons. Twitch’s “Impersonation” policy prohibits “content or activity meant to impersonate an individual or organization” and considered impersonation to be a zero-tolerance violation that will result in indefinite suspension.
- The “Account Usernames and Display Names” policy prohibits account names that violate Community Guidelines, including usernames and display names that include terrorism, child exploitation, violence and threats, harassment and sexual harassment, unauthorized sharing of private information, impersonation, glorification of natural or violent tragedies, self-destructive behavior, and references to recreational drugs.
- The “Harmful Misinformation Actors” policy prohibits “harmful misinformation superspreaders who persistently share misinformation on or off of Twitch. We seek to remove users whose online presence is dedicated to (1) persistently sharing (2) widely disproven and broadly shared (3) harmful misinformation topics. The policy includes both an assessment on on-platform and off-platform behavior.
- The “Nudity, Pornography, and Other Sexual Content” policy prohibits “Content or activities that threaten or promote sexual violence or exploitation,” which may involve torture or cruel, inhumane or degrading treatment.
- The “Extreme Violence, Gore, and Other Obscene Conduct” policy prohibits “content that exclusively focuses on extreme or gratuitous gore and violence is prohibited.”
- Twitch’s Off-service Conduct policy enforces against users found to be engaging in severe offenses outside of Twitch, which include physical harm to others, such as hate group membership, terrorist recruitment, sexual assault, and child grooming.

- The “Breaking the Law” policy requires all users to “respect all applicable local, national, and international laws while using our services, Any content or activity featuring, encouraging, offering, or soliciting illegal activity is prohibited. This would in theory include laws prohibiting slavery and human trafficking.
- The “Spam, Scams, and Other Malicious Conduct” policy prohibits “Any content or activity that disrupts, interrupts, harms, or otherwise violates the integrity of Twitch services or another user’s experience or devices is prohibited. This includes some activities that could be associated with slavery and human trafficking, such as defrauding others and misinformation.
- The “Nudity, Pornography, and Other Sexual Content” policy prohibits “Content or activities that threaten or promote sexual violence or exploitation,” which would include sex trafficking related activities.

Gaps:

- The “Violence and Threats” policy does not mention any distinctions based on context, or situations in which depictions of and calls to violence may be legitimate from a human rights perspective. For example, would users in a conflict setting be prohibited from utilizing Twitch to document war crimes? Or would users in an invaded territory be allowed to call for people to take up arms against an invading force, as enabled by Article 51 of the UN Charter? Would users be allowed to praise a country’s military actions?
- Many Twitch policies indirectly address child exploitation risks, but not explicitly. Being unambiguous about child exploitation content, such as grooming and solicitation, as well as sexualization of children, is a gap.
- The harmful misinformation policy could more clearly define what constitutes harmful to ensure Twitch has a clear framework for drawing lines as new forms of misinformation emerge. One definition of harmful could be “misinformation with a risk of resulting in significant harm to human rights.”
- There is no specific prohibition on utilizing Twitch to recruit potential victims and facilitate human trafficking, including sex trafficking, sales of children / illegal adoption, orphanage trafficking, forced marriages, labor exploitation, domestic servitude, organ trafficking, forced criminal activity, and recruitment of child soldiers.
- Although the Community Guidelines do not address torture or cruel, inhumane, or degrading treatment directly, it is likely sufficiently covered by existing policies that address risks of physical harm to people.

Right to Remedy

Related Human Rights

UDHR Art 8 • ICCPR Art. 2 • CRC Art. 37

Relevance to Twitch

- When Twitch’s platforms are used to harm others, Twitch may not be able to offer an effective remedy to the rightsholder depending on the harm in question.

- Twitch may not retain removed content that has evidentiary value in a remedy process, such as documentation of war crimes or other human rights abuses.

Related Twitch Community Guidelines Content

Existing Policy:

- Twitch’s Community Guidelines states that Twitch reserves the right to suspend or take action at any time against accounts for any “inappropriate or harmful” conduct. Such actions may include “removal of content, a strike on the account, and / or suspension of accounts(s).”

Gaps:

- In the case of extreme harm, such as violence or death stemming directly from use of Twitch’s platform, it is unclear whether suspension of the account in question will suffice as an effective remedy under international human rights law and the UN Guiding Principles.
- The Community Guidelines do not say anything about retention of blocked live streams or removed uploaded content with evidentiary value in remedy processes.

Right to Privacy

Related Human Rights

UDHR Art. 12 • ICCPR Art. 17 • CRC Art. 16 • CRPD Art. 22

Relevance to Twitch

- Twitch could be used in a way which violates the individual right to privacy of individuals who do not give consent to be included in live streams or whose private information is shared without consent.

Related Twitch Community Guidelines Content

Existing Policy:

- The “Breaking the Law” policy requires all users to “respect all applicable local, national, and international laws while using our services, Any content or activity featuring, encouraging, offering, or soliciting illegal activity is prohibited. This would in theory include laws against violations of reasonable privacy.
- The “Hateful Conduct and Harassment” policy prohibits harassment, which includes stalking, unwanted advances, and other violations of personal privacy.
- The “Unauthorized Sharing of Personal Information” policy prohibits the sharing of “content that may reveal private personal information about individuals, or their private property, without permission.” The policy defines personal information as including but not limited to, PII, restricted or protected social profiles, or content that violates a person’s reasonable expectation of privacy, such as streaming from a private space without permission.
- The “Account Usernames and Display Names” policy prohibits account names that violate the Community Guidelines, including usernames and display names that include unauthorized sharing of private information.

Gaps:

- The Community Guidelines do not address including people in live streams who may not be aware without their consent. There is a general lack of clarity on where lines are drawn on privacy in live streams—for example, of a publicly accessible space vs. a party held in a private home.

Right to Own Property

Related Human Rights

UDHR Art. 17 • ICESCR, Art. 15 • CRPD Art.12 • UNDRIP Art 12, 28, 31, 40

Relevance to Twitch

- Content violating intellectual property rights could be shared on Twitch.
- Theft of property could be facilitated on Twitch.

Related Twitch Community Guidelines Content

Existing Policy:

- Unauthorized sharing of personal information can lead to property theft. The “Unauthorized Sharing of Personal Information” policy prohibits the sharing of “content that may reveal private personal information about individuals, or their private property, without permission.”
- Impersonation can be used to carry out property theft. The “Impersonation” policy prohibits “content or activity meant to impersonate an individual or organization.” And treats misrepresentation as a member of Twitch representatives is a zero-tolerance violation resulting in indefinite suspension.
- The “Account Usernames and Display Names” policy prohibits account names that violate the Community Guidelines, including usernames and display names that include unauthorized sharing of private information or impersonation.
- The “Spam, Scams, and Other Malicious Conduct” policy prohibits any content or activity that disrupts, interrupts, harms, or otherwise violates the integrity of Twitch services or another user’s experience or devices, and includes a number of activities that may be associated with property theft, such as phishing, defrauding others or spreading viruses or malware.
- The “Intellectual Property Rights” policy states that users can only share content on their Twitch channels that they own or are authorized to share. Unauthorized shared content can include “any third party content included in your content, derivative creations, or performances of others’ copyrighted content.” The policy also encourages users to “assess your content for adherence to applicable intellectual property laws and the proper application of principles such as fair use, and to secure all appropriate rights needed, before sharing your content on Twitch.” The policy states that “Any unauthorized content you share on Twitch violates our Terms of Service and is subject to removal. Multiple violations of our policies may lead to a permanent suspension of your account.”

Gap:

- A policy governing use / depiction of cultural heritage and intangible property, particularly that which belongs to indigenous peoples.

Freedom of Thought, Conscience, and Religion

Related Human Rights

UDHR Art.18 • ICCPR Art. 18 • UNDRIP Art.12 • CRC Article 14

Relevance to Twitch

- Attacks on religious groups could inhibit Twitch users from utilizing Twitch to practice their religion and participate in religious community gathering.

Related Twitch Community Guidelines Content**Existing Policy:**

- The “Hateful Conduct and Harassment” policy prohibits hateful content, which is defined as “any content or activity that promotes or encourages discrimination, denigration, harassment, or violence based on the following protected characteristics: race, ethnicity, color, caste, national origin, immigration status, religion, sex, gender, gender identity, sexual orientation, disability, serious medical condition, and veteran status. We also provide certain protections for age.” It also states that this policy is applied equitably: “We afford every user equal protections under this policy, regardless of their particular characteristics.”

Gap:

- No gap identified. Risks to this right are adequately covered by the hateful conduct and harassment policy.

Freedom of Expression and Opinion, Access to Information

Related Human Rights

UDHR Art. 19 • ICCPR Art. 19 • CRPD Art. 7, 21 • CRC Art. 12, 13 • UNDRIP Art. 13, 16

Relevance to Twitch

- Twitch policies may unduly limit or fail to expect freedom of expression, especially if enforcement is not applied consistently across issues or groups.
- At the same time, by failing to take action against harmful content and behavior, Twitch could fail to create a safe space for free expression.
- Access to information may be put at risk by misinformation and misrepresentation.

Related Twitch Community Guidelines Content

Existing Policy:

- The “Breaking the Law” policy requires all users to “respect all applicable local, national, and international laws while using our services. Any content or activity featuring, encouraging, offering, or soliciting illegal activity is prohibited, including threats of violence. This is a justifiable restriction on free expression according to international human rights law.
- The “Violence and Threats” policy prohibits acts and threats of violence as zero-tolerance violations, meaning accounts associated with violence and threats will be indefinitely suspended. In the policy, violence and threats includes but is not limited to: threats to physically harm or kill others, hack, DDOS, or SWAT others, and the use of weapons to physically threaten or intimidate others. This is a justifiable restriction on free expression according to international human rights law.
- The “Hateful Conduct and Harassment” policy prohibits hateful content, which is defined as “any content or activity that promotes or encourages discrimination, denigration, harassment, or violence based on the following protected characteristics: race, ethnicity, color, caste, national origin, immigration status, religion, sex, gender, gender identity, sexual orientation, disability, serious medical condition, and veteran status. We also provide certain protections for age.” This is a justifiable restriction on free expression according to international human rights law. It also states that this policy is applied equitably: “We afford every user equal protections under this policy, regardless of their particular characteristics.”
- The “Impersonation” policy prohibits “content or activity meant to impersonate an individual or organization.”
- The “Harmful Misinformation Actors” policy prohibits “harmful misinformation superspreaders who persistently share misinformation on or off of Twitch. We seek to remove users whose online presence is dedicated to (1) persistently sharing (2) widely disproven and broadly shared (3) harmful misinformation topics. The policy includes both an assessment on on-platform and off-platform behavior. This is a justifiable restriction on free expression according to international human rights law.
- Twitch’s Off-service Conduct policy enforces against users found to be engaging in severe offenses outside of Twitch, including hate group membership.

Gaps:

- The harmful misinformation policy does not clearly define what constitutes harmful, which means Twitch lacks a clear framework for drawing lines as new forms of misinformation emerge. Without such a framework, Twitch risks moderating misinformation in a way which dampens the freedom of expression of users.
- Twitch’s Off-service harm policy fails to provide a standardized example of what hate group membership looks like. As hate speech evolves online, consensus on which groups espouse hateful ideology will change as well, which creates risk of under or over-enforcement of content policy.

Freedom of Assembly and Association

Related Human Rights

UDHR Art. 20 • ICESCR, Art. 8 • ICCPR, Art. 21 • CRC Article 15

Relevance to Twitch

- Freedom of assembly and association can be adversely impacted by harmful content or behavior on Twitch that leads to users not feeling safe to gather.

Related Twitch Community Guidelines Content

Existing Policy:

- Being targeted by threats of violence can make people feel unsafe to gather on Twitch. The “Violence and Threats” policy prohibits acts and threats of violence as zero-tolerance violations, meaning accounts associated with violence and threats will be indefinitely suspended. In the policy, violence and threats includes but is not limited to: attempts or threats to physically harm or kill others, attempts or threats to hack, DDOS, or SWAT others, and the use of weapons to physically threaten, intimidate, harm, or kill others.
- The “Violence and Threats” policy prohibits content that “depicts, glorifies, encourages, or supports terrorism, or violent extremist actors or acts. This includes threatening to or encouraging others to commit acts that would result in serious physical harm to groups of people or significant property destruction. You may not display or link terrorist or extremist propaganda, including graphic pictures or footage of terrorist or extremist violence, even for the purposes of denouncing such content.”
- The “Violence and Threats” policy states, “In exceptional circumstances, we may preemptively suspend accounts when we believe an individual’s use of Twitch poses a high likelihood of inciting violence. In weighing the risk of harm, we consider an individual’s influence, the level of recklessness in their past behaviors (regardless of whether any past behavior occurred on Twitch), whether or not there continues to be a risk of harm, and the scale of ongoing threats.”
- Being targeted by hate speech or harassment can make users feel unsafe to gather on Twitch. The “Hateful Conduct and Harassment” policy prohibits hateful conduct, which can lead to physical violence against targeted groups and individuals.
- The “Hateful Conduct and Harassment” policy prohibits harassment, which can lead to physical safety and psychological well being risks. Harassment is defined as “stalking, personal attacks, promotion of physical harm, hostile raids, and malicious false report brigading.” Sexual harassment is included under this policy, and is defined as, “unwelcome sexual advances and solicitations, sexual objectification, or degrading attacks relating to a person’s perceived sexual practices.”
- Doxxing activities can make users feel unsafe to gather on Twitch. The “Unauthorized Sharing of Personal Information” policy prohibits the sharing of “content that may reveal private personal information about individuals, or their private property, without permission.” The policy defines personal information as including but not limited to, PII, restricted or protected social profiles, or content that violates a person’s reasonable expectation of privacy, such as streaming from a private space without permission.

Gaps: No gaps identified

Right to Political Participation

Related Human Rights

UDHR Article 21 • ICCPR, Art. 25 • CEDAW Art. 7 • CRPD Art. 29 • UNDRIP Art.18

Relevance to Twitch

- Civic and election related misinformation on Twitch could adversely impact people's right to political participation.

Related Twitch Community Guidelines Content

Existing Policy:

- The "Harmful Misinformation Actors" policy prohibits "harmful misinformation superspreaders who persistently share misinformation on or off of Twitch. We seek to remove users whose online presence is dedicated to (1) persistently sharing (2) widely disproven and broadly shared (3) harmful misinformation topics. The policy includes both an assessment on on-platform and off-platform behavior. The policy includes "civic misinformation that undermines the integrity of a civic or political process," such as " the promotion of verifiably false claims related to the outcome of a fully vetted political process, including election rigging, ballot tampering, vote tallying, or election fraud."

Gap:

- During elections and other periods of political crisis, individual users who may not be considered harmful misinformation actors may share civic-related information in ways that aid harmful misinformation actors and have an adverse impact on political processes. Because the harmful misinformation policy is focused on actors who are devoted to this kind of misinformation, it does not address these compound risks.

Right to Education

Related Human Rights

UDHR Art. 26 • ICESCR, Art. 13, 14 • CRC Art. 17, 28, 29 • CRPD Art. 24 • UNESCO Convention against Discrimination in Education Article 5 • UNDRIP Art 13, 14, 16

Relevance to Twitch

- The right to education could be adversely impacted if there is educational content on Twitch framed as being from legitimate educational bodies when it isn't.

Related Twitch Community Guidelines Content

Existing Policy:

- The "Impersonation" policy prohibits "content or activity meant to impersonate an individual or organization."

Gaps:

- The Community Guidelines do not address more general risks of misrepresentation of being formally accredited or licensed in certain fields, including education. They also do not address the quality of ostensibly “official” educational material.

Right to Science and Culture

Related Human Rights

UDHR Art. 27 • ICCPR, Art. 27 • CRPD Art. 30 • ICESCR Art. 15 • UNDRIP Art. 11, 13, 31

Relevance to Twitch

- Twitch may inhibit Indigenous groups’ access to culture by not addressing the reproduction / use of their cultural heritage on the platform.

Related Twitch Community Guidelines Content**Existing Policy:**

- The “Intellectual Property Rights” policy states that users can only share content on their Twitch channels that they own or are authorized to share. Unauthorized shared content can include “any third party content included in your content, derivative creations, or performances of others’ copyrighted content.” The policy also encourages users to “assess your content for adherence to applicable intellectual property laws and the proper application of principles such as fair use, and to secure all appropriate rights needed, before sharing your content on Twitch.” The policy states that “Any unauthorized content you share on Twitch violates our Terms of Service and is subject to removal. Multiple violations of our policies may lead to a permanent suspension of your account.”

Gap:

- A policy governing use / depiction of cultural heritage and intangible property, particularly that which belongs to certain indigenous peoples.

Annex 2: Human Rights Due Diligence Resources



BSR recommends that Twitch integrate human rights into existing risk assessment processes, such as safety-by-design, data protection impact assessments, and other forms of risk assessment. As examples, this human rights due diligence approach may assess the human rights risks associated with growing in new locations, introducing new product features (e.g., encrypted private chat), or introducing new products (e.g., paid contracts with creators, rather than ad-based income).

This annex is intended to provide some resources to support the implementation of this recommendation.

Key Questions

What human rights risks may arise from the use of the new technology / product feature / location, for both users and non-users? [Refer to list of human rights, below]

Which vulnerable groups and marginalized populations (both users and non-users) might be adversely impacted by the new technology / approach? [Refer to non-exhaustive list of vulnerable groups, below]. Consider:

- › Formal discrimination—laws or policies 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, such as undocumented migrants.

Of the impacts identified, which are most salient, taking into account the following criteria?

- › The scope of the negative impact: How widespread would the impacts be on the population impacted?
- › The scale of the impact: How grave is the negative impact on those affected?
- › The remediability of the impact: Is it possible to counteract or make good of the resulting harm?
- › The likelihood of the impact: What is the chance of the impact happening?

What actions can be taken to avoid, prevent, or mitigate adverse impacts? Examples include:

- › Technical limitations (e.g., limited functionality)
- › Policy limitations (e.g., service terms, AUPs)
- › Limitations on who can use the new technology / approach (e.g., gating process to allow / block users)
- › Best practice guidance (e.g., user manual, training)
- › Transparency (e.g., public reporting; communications with users)

What further research is needed?

Who else should be consulted (e.g., affected stakeholders, experts) in determining Twitch's approach?



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

Copyright © 2023 by Business for Social Responsibility (BSR)

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.