



# **COMMUNITY-LED HUMAN RIGHTS IMPACT ASSESSMENT ON X (FORMERLY TWITTER, INC.)**

---

---

June 2025

## Table of Contents

1. Abbreviations	2
2. Executive Summary	3
2.1. Introduction	3
2.2. Scope	4
2.3. Methodology	4
2.4. Stakeholder Engagement	4
2.5. Impact Assessment:	4
2.6. Recommendations for X	5
3. Introduction	8
3.1. Introduction	8
3.2. Intent	8
3.3. Incentive	8
4. Disclaimer	9
5. Methodology	9
5.1. Prioritized Human Rights and Causes	10
6. Phase 1: Stakeholder Engagement	10
6.1. Feedback from Rightsholders	10
6.2. Feedback from other relevant stakeholders	13
7. Phase 2: Context Analysis	14
7.1. Company Context	14
7.1.1 Relevant Legal Frameworks and X's Human Rights Responsibilities	14
7.1.2 Company Structures to Implement Human Rights Obligations in China	15
7.1.3 Results of Implementing Human Rights Obligations in Mainland China	16
8. Phase 3: Impact Assessment	17
8.1. Priority 1 Impacts	18
8.2. Priority 2 Impacts	19
9. Phase 4: Recommendations	21
9.1. Improving the Respect and Remedy Mechanisms of Human Rights Within X	21
10. Phase 5: Company Engagement	23
<b>Appendix A – List of Online Sources Consulted</b>	<b>24</b>

## 1. Abbreviations

CLARITI	The Community-Led Assessment of Rights Impacts in the Technology Industry
CSDDD	Corporate Sustainability Due Diligence Directive
EU DSA	Digital Services Act
GDPR	General Data Protection Regulation
HRIA	Human Rights Impact Assessment
HRIC	Human Rights in China
ICCPR	International Covenant on Civil and Political Rights
ICESCR	International Covenant on Economic, Social and Cultural Rights
OHCHR	Office of the United Nations High Commissioner for Human Rights
UDHR	Universal Declaration of Human Rights
UNGP	United Nations Guiding Principles on Business and Human Rights

## 2. Executive Summary

### 2.1. Introduction

Human Rights in China (HRIC) is a nongovernmental organization founded in March 1989 by overseas Chinese students and scientists. HRIC's mission is to support and strengthen domestic civil society actors through the advancement of international human rights and the institutional protection of these rights in the People's Republic of China, including Hong Kong Special Administrative Region (HKSAR) and overseas Chinese in diaspora.

For this project, HRIC worked closely with human rights defenders and dissidents working on China issues who have been and are using X (formerly known as Twitter) as a mode of communication while circumventing the Great Firewall of China. For the purposes of this assessment, this target group of X users will be referred to as "rightsholders." Mediums like X allow rightsholders to access and share information on the Internet which may otherwise be censored on Chinese news sites and social media. It is also a key platform for facilitating free expression and communication among the Chinese-speaking community, including around sensitive topics such as human rights.

In undertaking the assessment, we addressed the following key concerns: 1) X's content moderation policies and their enforcement, which is overly reliant on AI and has resulted in under- and over- moderation of content, in turn leading to arbitrary account suspensions; and 2) X's inconsistent verification system, including the recently revamped Blue Checkmark that allows impersonation and misinformation campaigns, such as coordinated spam, harassment, and bots, to flourish. These issues significantly hinder rightsholders' ability to use X to express their opinions freely, share information on crucial human rights issues happening within the mainland that may otherwise never reach a global audience, and simultaneously access important information that would be censored by the Great Firewall of China. These activities are protected by Article 19 of the Universal Declaration of Human Rights, which declares that "[e]veryone 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."

The importance of X for Chinese human rights defenders cannot be over-estimated. There are no real alternatives to X for these users, since other Chinese messaging platforms and social media apps like WeChat or Weibo are heavily censored, surveilled, or government-linked, and non-Chinese alternatives like Mastodon or Bluesky do not have significant usage or reach. X's arbitrary and non-transparent decisions about content moderation and account suspension create far-reaching consequences for China-based users' rights and safety, especially their right to freedom of expression. This results in defender communities that are already marginalized in China being disproportionately affected.

For this HRIA, HRIC used the CLARITI (Community-Led Assessment of Rights Impacts in the Technology Industry) methodology to conduct an assessment of X (formerly Twitter). The methodology was developed by Ranking Digital Rights in 2023 with support from ARTICLE 19 under the Engaging Tech for Internet Freedom Initiative (ETIF). The assessment is intended to address salient issues experienced by rightsholders, such as arbitrary account suspensions, coordinated spam and attacks, and problems with the Blue Checkmark i.e. lack of transparency of verification, as well as impersonation efforts via Blue Checkmark accounts. The assessment

would be the start of an important process and conversation with X to address and identify key issues, and hopefully enable rightsholders to be able to continue to access X freely and without any encumbrances in the long run, in turn ensuring their right to freedom of expression (including the right of access to information), and right to privacy.

## 2.2. Scope

This HRIA assesses X's content moderation and content visibility. Our target country is mainland China: though X remains banned and does not have an actual office presence in the country, the app is still in use by rightsholders in the mainland. Rightsholders have been using the platform through Virtual Private Networks (VPNs). The timeframe of the assessment is July to December 2024.

## 2.3. Methodology

The CLARITI methodology applies the International Bill of Human Rights as its baseline to define human rights and the UN Guiding Principles on Business and Human Rights (UNGP) as a guideline to assess the content moderation systems and practices of the company. Other legal and non-legal requirements, such as the Global Network Initiative (GNI) Principles, the Santa Clara Principles, and the EU General Data Protection Regulation (EU GDPR), are also applied in the analysis of soft law guidance available regarding how tech companies can respect human rights, as well as the broader legal and regulatory context in which the company operates.

## 2.4. Stakeholder Engagement

In addition to undertaking extensive secondary research (see Appendix A), HRIC directly consulted 20 rightsholders. These interviewees are prolific Chinese human rights defenders and dissidents, based inside and outside of mainland China (in diaspora), including grassroots groups, lawyers, journalists, students, and other civil society actors. We also consulted another stakeholder, an ex-Twitter employee with direct familiarity with X's human rights processes, to gain better insights on the technical, legal, and human rights issues with regards to X's operations and influence in mainland China.

## 2.5. Impact Assessment:

A human rights impact assessment was conducted in line with UNGP Articles 12<sup>1</sup>, 13, 14<sup>2</sup>, 18<sup>3</sup>.

---

<sup>1</sup> UNGP Article 12: "The responsibility of business enterprises to respect human rights refers to internationally recognized human rights – understood, at a minimum, as those expressed in the International Bill of Human Rights and the principles concerning fundamental rights set out in the International Labour Organization's Declaration on Fundamental Principles and Rights at Work."

<sup>2</sup> UNGP Article 14: "The responsibility of business enterprises to respect human rights applies to all enterprises regardless of their size, sector, operational context, ownership and structure. Nevertheless, the scale and complexity of the means through which enterprises meet that responsibility may vary according to these factors and with the severity of the enterprise's adverse human rights impacts."

<sup>3</sup> UNGP Article 18: "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 either through their own activities or as a result of their business relationships. This process should: (a) Draw on internal and/or independent external human rights expertise; (b) Involve meaningful consultation with potentially affected groups and other relevant stakeholders, as appropriate to the size of the business enterprise and the nature and context of the operation."

19<sup>4</sup>, 23<sup>5</sup>, and 24<sup>6</sup> and impact assessment best practices. This assessment highlighted the following impacts and causes which must be addressed by X:

- Article 13(a) of the UNGP states: “The responsibility to respect human rights requires that business enterprises: (a) Avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur.” Through our human rights impact assessment, we found that X’s enforcement of its content moderation policies is overly reliant on AI and has resulted in under- and over- moderation of content, in turn resulting in arbitrary account suspensions, which has a far-reaching impact on the rights of rightsholders, infringing their rights to freedom of expression and access to information. Arbitrary account suspensions, which may be temporary but can last for hours, days, weeks, or even months, mean that rightsholders will lose a valuable tool of communication during that period, while in an environment that already has extensive restrictions on communication.
- In the same vein, X’s inconsistent verification system, including the recently revamped Blue Checkmark, allows impersonation and misinformation campaigns, such as coordinated spam, harassment, and bots, to flourish, which has impacted users’ ability to freely receive and impart information.

There is no public information on what steps X is taking to address the impact of these crucial issues. Further, X does not have a dedicated human rights unit to address this impact, with evidence showing that X has deprioritized human rights in its *logos*, *pathos*, and *ethos* overall, such as disbanding its entire Trust and Safety Council in December 2022.

## 2.6. Recommendations for X

We make the following recommendations to X to uphold its responsibility to respect human rights and mitigate the adverse human rights impacts identified above:

- With regards to the Blue Checkmark “for a fee,” X should reconsider only allowing users to receive Blue Checkmark verification through its paid premium subscription. Our research and stakeholders’ experiences have shown that the current system has encouraged impersonation attempts, disinformation campaigns, spear-phishing attacks, and hacking, thus fundamentally preventing the target group from exercising their right to freedom of expression, as well as contributing to information threats and

---

<sup>4</sup> UNGP Article 19: “In order to prevent and mitigate adverse human rights impacts, business enterprises should integrate the findings from their impact assessments across relevant internal functions and processes, and take appropriate action. (a) Effective integration requires that: (i) Responsibility for addressing such impacts is assigned to the appropriate level and function within the business enterprise; (ii) Internal decision-making, budget allocations and oversight processes enable effective responses to such impacts. (b) Appropriate action will vary according to: (i) Whether the business enterprise causes or contributes to an adverse impact, or whether it is involved solely because the impact is directly linked to its operations, products or services by a business relationship; (ii) The extent of its leverage in addressing the adverse impact.”

<sup>5</sup> UNGP Article 23: “In all contexts, business enterprises should: (a) Comply with all applicable laws and respect internationally recognized human rights, wherever they operate; (b) Seek ways to honour the principles of internationally recognized human rights when faced with conflicting requirements; (c) Treat the risk of causing or contributing to gross human rights abuses as a legal compliance issue wherever they operate.”

<sup>6</sup> UNGP Article 24: “Where it is necessary to prioritize actions to address actual and potential adverse human rights impacts, business enterprises should first seek to prevent and mitigate those that are most severe or where delayed response would make them irremediable.”

transnational repression. The current Blue Checkmark verification requirements should be revised to prioritize information accuracy.

- As a stop-gap measure, there should be rapid action focused on due diligence and accurate verification of information, such as the identities of the Blue Checkmark holders, through improving X's current content moderation AI algorithms to precisely address the aforementioned issues that have come about as a result of the Blue Checkmark, as well as an increase in the number of human reviewers which would be able to reinforce these efforts.
  - In the mid and long term, it is recommended that X's Blue Checkmark verification should revert to a system that emphasizes due diligence and accuracy of information with adequate human rights safeguards, to prevent overreach. To do this, X would need to take active steps to ensure and verify that an X account is owned by the person or organization it claims to represent. At the same time, verification requirements should be cognizant of the existing real-name and ID verification regulations under the Cybersecurity Law in China.<sup>7</sup>
- Content moderation should not overly rely on AI models to address both under- and over-moderation, both of which significantly affect rightsholders. An immediate solution would be to utilize more human moderators with specialized training. Such human moderators, not based inside the PRC, should be context-aware, i.e. with a good understanding of issues involved, and linguistically diverse. A long-term solution would be to refine the AI models, which would go towards better trained data sets and models, and enhancing human control over decisions made by AI and ensuring legality, necessity, and proportionality in content moderation decisions.
    - Algorithms should be committed to accuracy of information instead of simply focusing on high engagement with low credibility. Proactive content moderation, by system or human, should only be channelled towards addressing actual issues such as impersonation attempts, disinformation campaigns, spear-phishing attacks, and hacking, and not to the extent that it becomes over-moderation, where relevant and legitimate content gets taken down.
    - Account suspension, especially of human rights defenders, must have a clear basis and should not be done at will without notification. There must be remedies for reinstatement. For instance, this could be done through transparently communicating to users about the content being moderated and providing appeal mechanisms and improving user control mechanisms such as blocking or reporting.
  - Fundamentally, X should be committed to human rights and its indivisibility. As such, X cannot claim to uphold freedom of expression while remaining silent or taking contradictory approaches on other concerns in relation to privacy and access to

---

<sup>7</sup> The Cyberspace Administration of China's draft measures propose a "Network ID" system where Chinese citizens can voluntarily apply for a digital credential linked to their official identification documents to be used for online verification across internet platforms, aiming to enhance real-name registration and data security across the internet. This raises privacy concerns and freedom of expression issues due to the centralized nature of this identity system.

information, amongst others, as these rights are equally important to all users. This requires a holistic change in direction, policies, and systems.

- A human rights unit within X would be able to address some of the above concerns, in particular related to human rights defenders using the platforms to circumvent the Great Firewall of China, on an immediate basis. A dedicated human rights unit would serve as a direct grievance mechanism that would deal specifically with human rights related complaints and issues, whereas the system now does not adopt a human rights-centered approach.

We look forward to collaborating with X in order to enhance its protection, respect, and remedy of adverse human rights impacts on its users in China and the diaspora.

## 3. Introduction

### 3.1. Introduction

Human Rights in China (HRIC) is a nongovernmental organization founded in March 1989 by overseas Chinese students and scientists. HRIC's mission is to support and strengthen domestic civil society actors through the advancement of international human rights and the institutional protection of these rights in the People's Republic of China, including Hong Kong Special Administrative Region (HKSAR) and overseas Chinese in diaspora.

For this project, HRIC worked with Chinese human rights defenders and dissidents who have been and are using X as a mode of communication while circumventing the Great Firewall of China. In addition, mediums like X allow rightsholders to access information on the Internet which may otherwise be censored on Chinese news sites and social media.

### 3.2. Intent

For this assessment, we focused on the following key concerns: 1) X's enforcement of its content moderation policies that are overly reliant on AI and have resulted in under- and over- moderation of content, in turn resulting in arbitrary account suspensions; and 2) X's inconsistent verification system, including the recently revamped Blue Checkmark that allows impersonation and misinformation campaigns, such as coordinated spam, harassment, and bots, to flourish, which has limited rightsholders' free usage of the platform.

This HRIA will address salient issues experienced by rightsholders, such as arbitrary account suspensions, coordinated spam and attacks, and problems with the Blue Checkmark i.e. lack of transparency of verification, as well as impersonation efforts via Blue Checkmark accounts. This would enable rightsholders to be able to continue to access X freely and without any encumbrances, and to ensure their rights to freedom of expression (including the right of access to information), and rights to privacy.

### 3.3. Incentive

- Ensuring rightsholders can exercise their rights to freedom of expression, access to information and privacy.
- Enhanced accountability of the company, especially to be more transparent and responsible towards its users, especially in relation to the over-reliance on AI that has created under- and over- moderation of content.
- Reduced negative rights impact of the company's activities, to address the multitude of issues that have arisen, including the removal of protected speech, while leaving inciting speech that has the potential to silence users unaddressed.
- Improved alignment with international human rights law (as provided for by the UNGP) and ESG legal obligations, some of which are regional but impact companies internationally, such as the Digital Markets Act (DMA) and the Digital Services Act (DSA).

## 4. Disclaimer

This report is prepared by the assessment team on the basis of their professional judgement, based on available primary and secondary information. The assessment team assumes no responsibility or liability for errors or omissions of the information provided by rightsholders or contained in publicly available sources. The assessment team welcomes constructive engagement with X for additional information that may augment and enhance the assessment.

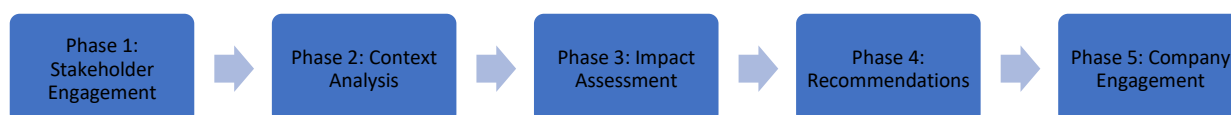
## 5. Methodology

The assessment was conducted using the CLARITI methodology: Community Led Assessment of Rights Impacts in the Technology Industry. The scope was X's content moderation and content visibility. Our target country is China: though X remains banned and does not have an actual presence in the country within mainland China, rightsholders have been using the platform through Virtual Private Networks (VPNs). The timeframe of the assessment was July 2024 through September 2024.

The CLARITI methodology was developed by Ranking Digital Rights in 2023, with support from ARTICLE 19, to foster constructive and collaborative engagement between telecommunications companies and tech platforms with large user bases and the communities these companies serve. The methodology was guided by the impact assessment requirements of the UN Guiding Principles on Business and Human Rights (UNGPs), literature on company-led and community-led human rights due diligence and impact assessment methodologies and extensive stakeholder consultations to help provide companies with a complementary, community-based perspective on their human rights impacts. The methodology is designed to assess the extent to which companies are following the UNGPs while simultaneously embedding the UNGP requirements within the assessment itself.

This methodology uses the International Bill of Human Rights as its baseline to define human rights while adopting business and human rights guidelines, principles, and best practices beyond the UNGPs, such as the GNI Principles, the Santa Clara Principles on Transparency and Accountability in Content Moderation, and the EU General Data Protection Regulation (GDPR). Specifically, the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social, and Cultural Rights (ICESCR) are instructive. It also considers the requirements imposed by the EU Digital Services Act (DSA) and Corporate Sustainability Due Diligence Directive (CSDDD).

The methodology consists of five phases:



During Phase 1, we consulted with potentially affected rightsholders and other relevant stakeholders in line with the requirements of UNGP 18.

During Phase 2, we assessed:

- Industry Context: the human rights record of X's competitors and potential for collaboration
- Company Context: the legal and non-legal human rights obligations X is subject to and the performance of X against these obligations

During Phase 3, we determined the saliency of specific impacts, and X's management of these impacts in line with UNGP 12, 13, 14, 19 and other impact assessment best practices.

During Phase 4, we suggested recommendations to support X in maximizing its positive human rights impacts while mitigating adverse impacts, prioritized in line with UNGP 23 & 24.

During Phase 5, we highlighted efforts we have made to contact X's representatives for potential engagement since we began this assessment.

## 5.1. Prioritized Human Rights and Causes

In line with the CLARITI methodology and the UN Human Rights Council finding that "the same rights that people have offline must also be protected online, in particular freedom of expression," we have prioritized our assessment of the two rights most directly impacted online by X's activities:

- **The right to freedom of expression, including the right of access to information (as protected by Article 19 of the UDHR and the same article in the ICCPR):** Content moderation systems and practices. Over reliance on AI and have resulted in under- and over- moderation of content, in turn resulting in arbitrary account suspensions.
- **The right to privacy (as protected by Article 12 of the UDHR and Article 17 of the ICCPR):** Content & advertising visibility algorithm. X's inconsistent verification system, including the recently revamped Blue Checkmark that allows impersonation and misinformation campaigns, such as coordinated spam, harassment and bots, to flourish, which has limited rightsholders' free usage of the platform and built false trust in the platform's content.

## 6. Phase 1: Stakeholder Engagement

The UNGP 18 requires a human rights impact assessment to be conducted through meaningful consultation with potentially affected groups and other relevant stakeholders. We have accordingly identified and consulted various stakeholders throughout the process, starting with the rightsholders whose concerns and interests we represent.

### 6.1. Feedback from Rightsholders

The methodology emphasized the concerns and perspectives of rightsholders potentially impacted by X's operations. As they are the primary subject of any rights impacts, they are the individuals to be impacted by any mitigation steps X may take, either as part of X's ongoing human rights due diligence or in response to this report. We have paid particular attention to human rights impacts on individuals from groups or populations that may be at heightened risk of vulnerability or marginalization. The following rightsholders or their representatives were consulted:

Total number of rightsholders consulted	20
Total number of rightsholders inside target country:	10

Total number of rightsholders outside target country, in diaspora and/or exile:	10
Total number of marginalized rightsholders:	10 (all of whom are still residing within target country)

The consultation format was largely virtual and included both individual and group interviews.

Breakdown of age range:	18-60
Breakdown of occupation:	Students, lawyers, journalists, and grassroots groups.
Total number of consultation sessions:	2-3 per individual / group

During consultations, rightsholders vividly expressed how X's actions impacted them.

In August 2024, a massive number of prolific Chinese human rights defenders, grassroots groups, and organizations found themselves targets of spear-phishing on X—targeted attacks designed to trick people into handing out information such as passwords. These attacks then allowed hackers to tweet from compromised accounts and access private direct messages. Numerous accounts, mostly within the Chinese dissident community, have been targeted, in one of the largest known attacks amongst the Chinese community on X.

Xiang Li, a well-known US-based human rights activist who was one of the targets during the 709 crackdown in 2015, recounted her experience. It was also thanks to Xiang Li's warning to her followers that more people have been made aware of the spear-phishing attack and did not fall prey to it:

"At around 7 a.m. on August 17, 2024, an X account named Help X @TouchXAnswers that I was not following had sent a private message to my X account @xiangli001, saying that my tweet was under review for possibly copyright violations and needed to be appealed within 24 hours, or the account would be suspended, and a web link was attached. I clicked on the link and filled in some information. Once logged in, it will lead the user to a phishing website that uses a fake X interface to make victims log in. After that, they will steal passwords and then control the accounts. My number was quickly stolen.

**Hello Dear X User,**

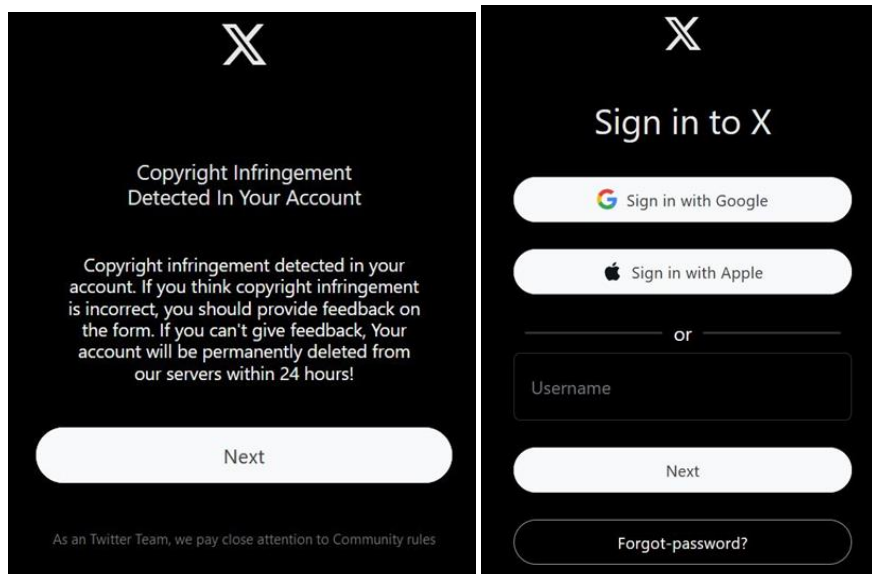
**Your account is under review for possible copyright violations.  
To submit your feedback within the specified timeframe,  
kindly fill out the form provided below.  
If you do not respond within 24 hours, your account will be deactivated.**

**Form: [help.extensions-x-faq.com](https://help.extensions-x-faq.com)**

**Wishing you a wonderful day.  
© X Support Team.**

*Image 1: Example of a direct message received for alleged copyright violations.*

I realised that my account had been stolen when I received an email that said my X account "successfully changed the email address of X account login", and at this point, I was no longer able to change my email address back to my own email address. Although my X account on my iPhone was still logged in, I couldn't modify any information or get my email login information back.



*Images 3 & 4: Once logged in, it will lead the user to a phishing website that uses a fake X interface to make victims log in. After that, they will steal passwords and then control the accounts.*



*Image 5: Example of an independent journalism outlet, Mighty Voice Media, which fell prey to the spear-phishing attack.*

It was only upon contacting X support that I was able to appeal and get my account back. After verifying with some other contacts, including @bingyuicejade, we realized that many other X accounts of famous Chinese human rights defenders had been stolen in the same manner. We then alerted our own followers on our accounts to be aware of the same attacks!”

So-called “shadow banning” of rightsholders on X is also commonplace. Shadow banning refers to the practice when a social media platform significantly limits the visibility of a user’s content covertly without notifying them. Though X denied the existence of this practice, shadow banning appears to be another example of visibility filtering. In November 2024, prolific Chinese user @whyyoutouzhele with 1.8 million followers on X was shadow banned, and it was speculated that it was in relation to the two-year anniversary of the White Paper Movement, in compliance with censorship requests from the Chinese government.

## 6.2. Feedback from other relevant stakeholders

To gain more insight into X’s processes and systems, we conducted a substantive interview with a former Twitter employee who had direct knowledge of the topics in this report (due to potential professional and legal repercussions, this employee requested to remain anonymous). This employee, Employee A, confirmed the majority of this report’s findings and clarified certain details; for example, confirming that X no longer has a human rights team and that its Trust and Safety team (a separate team) lost the majority of its members following the transition to Elon Musk’s ownership. Previously, the Trust and Safety team had been responsible for tracking and working to counteract threats by state actors, including propaganda and mainland China’s attempts to silence Chinese human rights defenders on the platform, a task which may no longer be covered at all. Employee A emphasized that not only did X’s internal priorities shift substantially away from trust and safety (and human rights), but it also lost a substantial portion of its employees with the highly sophisticated technical and socio-political knowledge necessary to implement key policies in regard to global threats and user safety. These insights provide useful context to the challenges rightsholders have faced over the last two years.

## 7. Phase 2: Context Analysis

### 7.1. Company Context

We first analyze the legal frameworks that may apply to X's content moderation practices, particularly those from some of the most relevant jurisdictions and regions for the purposes of this report, the US and the EU—these are the locations of many of our rightsholders that are located outside of mainland China, and also two of the most significant market regions for X—and unpack to what extent some of these may contain requirements to uphold human rights. Second, we analyze soft law instruments, in particular the UN Guiding Principles on Business and Human Rights as the main framework setting out companies' human rights responsibilities. Third, we assess X's human rights record in mainland China, drawing upon assessments and reports made by journalists and civil society, and X's activities for promoting rights. Due to the limited public availability of some of this information, a best effort assessment has been made, but the assessment team welcomes X's engagement and is ready to update the assessment based on any additional information X may provide.

In November 2022, after Elon Musk purchased Twitter, the entire human rights team was laid off. Another team, known internationally as META, which worked on ethical AI and algorithmic transparency, was also dismissed. Under Musk, X's new content moderation policy, "Freedom of Speech, Not Reach," had begun to limit the visibility of tweets that violated its policies, rather than removing the content from the site as was done previously. Alongside such clear examples of under-moderation, there has also been a surge in over-moderation, many of which are arbitrary account suspensions that are a result of over-intrusive AI.

#### 7.1.1 Relevant Legal Frameworks and X's Human Rights Responsibilities

X, an American company, is not subject to the international laws and regulations that would apply to states. It is instead subject to the domestic laws in the countries in which it operates and provides services. We highlight the United States and the European Union as areas of particular relevance, as these are both large markets for X and the location of many "exiled" Chinese dissidents who were interviewed as rightsholders for this HRIA.

The First Amendment of the U.S. Constitution prevents the United States government from imposing laws on citizens and private actors like X that may constitute an infringement on their freedom of speech. Furthermore, X is protected under Section 230 of the Communications Act of 1934, which allows online platforms to police their sites but does not require them to remove content, and protects platforms from civil liability for third-party content. X has asserted that it has absolute power to ban whomever and whatever it likes, because it is governed only by US law, regardless of where in the world it operates. (Source) US law grants private companies the right to ban speech on their private property. According to X, it can thus ban an entire political party or ideology. It can also ban activist groups or politicians at will. Despite the US government emphasizing how technological progress, specifically artificial intelligence and digital tools, must spur democratic renewal and not undermine it, it is unfortunate that X has adopted the opposite position. The fact that US government regulatory bodies are not taking a stricter regulatory and oversight approach does not help the situation.

However, X is still bound by regional directives or similar legislation. Many Chinese activists and human rights defenders are based in Europe, where the European Union (EU) has taken a proactive approach in

terms of rights protection online. Some have pointed out the so-called “Brussels effect” of regulations such as the EU GDPR, whereby companies adopt the requirements of EU legislation throughout their operations at a global level in order to access the European market and avoid potential conflicts. Yet, it remains to be seen whether this phenomenon will continue with more recent EU regulations, such as the European Commission (EC)’s Corporate Sustainability Due Diligence Directive (CSDDD). The CSDDD, which recently entered into force in July 2024, is an example of such a directive that could have an impact, direct or indirect, on X in relation to China. Even for users of X outside Europe, the effects of the CSDDD extends to beyond European borders due to its supply chain restrictions and requirements. Another relevant piece of European legislation is the EU Digital Services Act (DSA), which is intended to protect the fundamental rights of users online. Currently, the EC is investigating whether X has breached the DSA, particularly whether or not Blue Checkmarks are deceptive, alongside other potential breaches.

X should also be guided by the UN Guiding Principles on Business and Human Rights, which argue that a company has a responsibility to respect human rights, both in terms of foundational principles and operational principles. According to Clause 13, the responsibility to respect human rights requires that business enterprises: (a) Avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur; and (b) Seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts.

Another important source that lays out how online platforms can uphold human rights is the Santa Clara Principles on Transparency and Accountability. The Principles state that all actions that a company takes should be in line with international human rights standards, conducted in a transparent and accountable way, and enforced in a consistent manner. Although X (then Twitter) was one of the twelve companies that endorsed the Santa Clara Principles on Transparency and Accountability in Content Moderation since 2018, after October 2022, the X Transparency Center did not published any further reports on Rules Enforcement or take further steps towards compliance with the Principles; instead, the language on the website changed from “supporting” to being “influenced by” the spirit of the principles. While it is evident that X’s initial declaration to the Principles have not translated into its successful implementation of the Principles in full, civil society organizations like the Electronic Frontier Foundation have called on X’s management to renew the company’s commitment to the Santa Clara Principles. It is clear that X’s actions and operations have not been aligned with the Santa Clara Principles, especially in light of rightsholders’ feedback and user experiences.

### 7.1.2 Company Structures to Implement Human Rights Obligations in China

X’s governance, systems and practices are only partially rights-respecting. While it delivers on its human rights responsibilities in certain aspects (for e.g., it has a comprehensive public data privacy policy), other policies related to public whistleblowing, content moderation, and so on are either partially rights-respecting or do not exist.

The company’s treatment of and commitment to human rights is first illuminated by the fact that post-Musk X does not have a governance structure in relation to its management of human rights (source), in addition to ongoing legal suits and other issues. While X does take certain steps to protect user rights, such as their X Rules (previously the bi-annual transparency reports when it was Twitter), user notice policies, service notices, law enforcement resources, and guidelines on uses of X products, its internal

Trust and Safety Council, a group of human and civil rights advisers formed in 2016, was dismissed in 2022 by Elon Musk.

Notably, under Musk, the platform has fired its entire human rights team and all but two of the ethical AI team, (prompting UN High Commissioner for Human Rights Volker Türk to issue an open letter urging Musk to “ensure human rights are central” to its management), abided by every government request in relation to censorship, sued a group of advertisers for using their rights of freedom of association to not associate with the racist content on the social media site.

For the rightsholders interviewed for this HRIA, X’s content moderation practices are of most significance. With Elon Musk at the helm, content moderation has been done through visibility filtering (i.e., restricting the reach of policy-violating content on the platform through restricting the reach of a post, vis-a-vis AI, but without removing posts or suspending users as frequently as under the human-moderator system). Yet this method has been proven to be poor and ineffective, according to the user experiences of the rightsholders consulted for this HRIA. While stakeholders can report posts and contact support teams to remedy account locking, restrictions, or suspension issues, and can also contact support teams for any other grievance or remedy through email and phone numbers, those who found their accounts suspended did not receive any explanations, and are often not able to reinstate their accounts.

### 7.1.3 Results of Implementing Human Rights Obligations in Mainland China

For X to be considered a rights-respecting or human rights-centered company, there is much more to be done. In November 2022, UN High Commissioner for Human Rights Volker Türk issued an open letter urging Musk to ensure that human rights are central to his management, and noted that: “Twitter [X] needs to understand the harms associated with its platform and take steps to address them. Respect for our shared human rights should set the guardrails for the platform’s use and evolution.”

In July 2024, the European Commission published its preliminary findings that X was in breach of the Digital Services Act (DSA) in areas linked to dark patterns, advertising transparency and data access for researchers. Notably, X was found to have used its “blue checkmark” in a way that does not correspond to industry practices: “Since anyone can subscribe to obtain such a ‘verified’ status, it negatively affects users’ ability to make free and informed decisions about the authenticity of the accounts and the content they interact with. There is evidence of motivated malicious actors abusing the ‘verified account’ to deceive users.” This has become an advantage for those targeting vulnerable groups, including human rights defenders. See Section 6.1, above, where rightsholders recounted how they found themselves targets of spear-phishing on X, targeted attacks designed to trick people into handing out information such as passwords, with the assumption that these hackers had been operating behind Blue Checkmark accounts.

Musk himself is also alleged to have sold user data to a company that facilitates government monitoring, allowing sponsored transphobic content, being anti-diversity, and more. In addition, X has seen a dramatic spike in hateful, violent and inaccurate posts, according to researchers, but there are no known programs or policies on X that educate users on preventing or stopping the spread of harmful online content.

Twitter was once one of the most relevant and useful social media platforms in China for human rights defenders, activists and defenders who were trying to access information outside of the Great Firewall. Because the service has never been legalized in mainland China, it is difficult to estimate the scale of the

user base located there; however, interviews with rightsholders supported HRIC's observations that Twitter, now X, had a thriving, active, and large-scale Chinese-language community that included a large percentage of mainland-based users and political dissidents. However, in recent years, with the disruptions and prohibitions experienced, Chinese users are not able to use Twitter in the same ways as before.

In the latest X Transparency Report, which covers January to June 2024, X noted that its "policies and enforcement principles are grounded in human rights, and [they] have been taking an extensive and holistic approach towards freedom of expression by investing in developing a broader range of remediations, with a particular focus on education, rehabilitation, and deterrence." It is, however, unclear what X is doing for education, rehabilitation, and deterrence.

If rightsholders continue to experience severe disruptions to their work via arbitrary account suspensions it makes it very difficult for them to continue their advocacy and organizing efforts. In addition, with the current state of Blue Checkmarks where users are no longer certain if X accounts with Blue Checkmarks are reliable sources of information, rightsholders will find it harder to obtain reliable, aggregated information from X in the long run. With the lack of any real credible alternatives available, the freedom of expression of rightsholders and the populations they serve will be greatly impacted in the next five years and beyond.

## 8. Phase 3: Impact Assessment

We apply the approach outlined in UNGP 14 and best practice human rights due diligence to determine the saliency of specific impacts and X's management of these impacts. To evaluate the relative priority of the human rights impacts, we used four factors to create a **Salience Score** (based on UNGP 14 and additional sources):

1. **Scope:** The percentage of the population potentially negatively affected by a given impact on human rights
2. **Scale:** The extent to which human rights are impacted
3. **Remediability:** If an individual's rights are impacted, the ease with which these negative impacts can restore the individual to their prior position
4. **Likelihood:** The likelihood of an impact occurring

A high Salience Score is thus assigned to impacts that:

- Affect >50% of the population
- Have a serious impact on physical rights
- Are irremediable
- Are certain to happen or are already happening

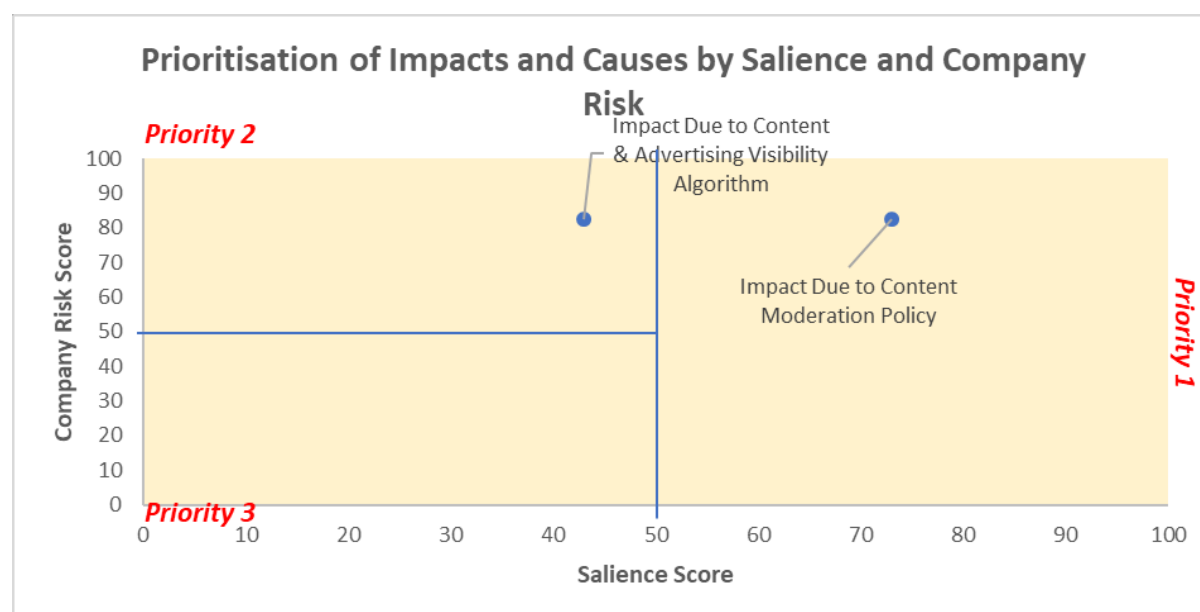
We then evaluate how well X is managing these impacts by combining three factors into a **Company Risk Score**:

1. **Connection:** X's connection with the rights impact (UNGP 13)
2. **Mitigation:** How well X is mitigating the negative rights impact to date (UNGP 19)
3. **Leverage:** The extent to which X can mitigate negative rights impacts further, based on a best-effort consideration of X's internal and external context (UNGP 19)

A high Company Risk Score is thus assigned to impacts and causes that:

- X has caused (as defined by UNGP 13)
- X is not trying to address, or the way X is addressing them is not public
- Exhibit high potential for X to make an improvement

The prioritization chart below plots the Saliency Score against the Company Risk Score for each potential rights impact and cause. The impacts and causes with higher Saliency and Company Risk scores are prioritized over those with lower Saliency and Company Risk scores, in line with UNGP 24.



The table below shows the data used to plot the chart above.

No.	Priority	Impacted Right and Cause	Saliency Score	Company Risk Score
1	1	Freedom of Expression - content moderation policy	73	83
2	2	Freedom of Expression - content & advertising visibility algorithm	43	83

## 8.1. Priority 1 Impacts

Primary Right Impacted	Cause of Impact	Reason for Saliency Score	Reason for Management Score
Freedom of expression	Enforcement of Content Moderation Policies	<ul style="list-style-type: none"> <li>• <b>Scope:</b> Technically &lt;10% of the total Chinese population, since most Chinese citizens are not X users, but affects an</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Connection:</b> Contribute. While X's content moderation policies are in place, they are not necessarily robust, and can incur negative impacts (in the form of an externality) in the</li> </ul>

	overly reliant on AI	<p>overwhelming majority of rightsholders consulted.</p> <ul style="list-style-type: none"> <li>● <b>Scope:</b> Health &amp; safety violation, as the information that rightsholders obtain via Twitter are often crucial to the organizing and advocacy of HRDs and civil society on topics that are time-sensitive and life-threatening. For instance, if under- and over- content moderation led to critical information on COVID-19 during lockdowns in China being inaccessible and led to reduced or halted communication in networks, it could lead to the potential loss of life.</li> <li>● <b>Remediability:</b> Not at all. To elaborate, should rightsholders experience arbitrary detention, enforced disappearances, or even the loss of life in situations such as during the COVID-19 lockdown, as in some real life cases <u>during that period</u>, no remedies would possibly be able to restore the rightsholders to the same or equivalent positions before the harm. This may or may not be a direct result of the lack of information and reduced or halted communication in networks on platforms like X.</li> <li>● <b>Likelihood:</b> Certain. If rightsholders continue to experience severe disruptions to their work via arbitrary account suspensions, it would be unlikely they would be able to continue their advocacy and organizing efforts. In addition, with the current state of Blue Checkmarks where users are no longer certain if X accounts with Blue Checkmarks are <u>reliable sources of information as before</u>, rightsholders will find it harder to obtain reliable, aggregated information from X in the long run. With the lack of any real credible alternatives available, the freedom of expression of rightsholders and the populations they serve would be greatly impacted in the next five years and even beyond.</li> </ul>	<p>case of the rightsholders who are living and working in very challenging country climate and are thus disproportionately affected by over- or under- content moderation.</p> <ul style="list-style-type: none"> <li>● <b>Mitigation:</b> Not trying / unknown. No public evidence of X trying to mitigate the risks of over reliance on AI-based content moderation.</li> <li>● <b>Leverage:</b> High potential. Rightsholders suffer from over content-moderation, which is less of an issue for X compared to under content-moderation - it is probably easier for X to address these issues such as accidental account suspensions if it adheres to its own internal grievance and remedy mechanisms (<a href="#">source</a>).</li> </ul>
--	-------------------------	--	---

## 8.2. Priority 2 Impacts

Primary Right Impacted	Cause of Impact	Reason for Salience Score	Reason for Management Score
Freedom of expression	Paid Blue Checkmark verification system after November 2022	<ul style="list-style-type: none"> <li>● <b>Scope:</b> Less than 10% of the country's population, but within the rightsholders we are consulting with (i.e. human rights defenders and civil society groups), an overwhelming majority. Especially with the new Blue Checkmark, these changes have eased impersonation, and allow accounts spreading misleading information to feign credibility. These have affected</li> </ul>	<ul style="list-style-type: none"> <li>● <b>Connection:</b> Contribute. X's low-credibility algorithms are a hotbed for China state-sponsored trolls and bots and misinformation campaigns behind the guise of Blue Checkmarks "for a fee", even if the majority of the Chinese population is not on X. These campaigns can reach the masses outside of China as part of their influence campaigns.</li> <li>● <b>Mitigation:</b> Not trying / unknown. With the introduction of the Blue Checkmark ownership</li> </ul>

		<p>our—rightsholders to a large extent, especially with CCP-backed accounts and misinformation campaigns masquerading behind the false legitimacy of the Blue Checkmarks.</p> <ul style="list-style-type: none"> <li>● <b>Scope:</b> Severe Mental Violation. Impersonation and misinformation campaigns such as coordinated spam, harassment, and bots, could lead to threats, intimidation, harassment or stalking, humiliation, cyberbullying, and defamation, and so on. For instance, in August 2024, a massive number of prolific Chinese human rights defenders, grassroots groups, and organizations have found themselves targets of spear-phishing on X, targeted attacks designed to trick people into handing out information such as passwords, with the assumption that these hackers had been operating behind Blue Checkmark accounts. These attacks then allowed hackers to tweet from compromised accounts and access private direct messages. Hundreds, if not thousands of accounts, mostly within the Chinese dissident community, have been targeted, in one of the largest known attacks amongst the Chinese community on X.</li> <li>● <b>Remediability:</b> To some extent. The problems could be addressed if steps are taken within a reasonable time. In the recent case of rightsholders whose accounts were compromised or hacked into due to the hackers that may have been operating behind Blue Checkmark accounts, they were able to reverse the situation quite quickly after proceeding with the standard grievance mechanisms, namely, that they still had access to their accounts on other devices (such as on a second mobile phone), which meant that time is of the essence in the remediability of this issue at present. However, it does not mean that the remedy could restore the rightsholder to the same or equivalent position to a large extent, if more extensive damage had been done in the short period of time the accounts were compromised, which was not unlikely.</li> <li>● <b>Likelihood:</b> Certain. If rightsholders continue to experience severe disruptions to their work via coordinated spam and harassment by CCP-backed hackers, such as the massive spear-phishing attacks in August 2024 on the assumption that they were via Blue Checkmark accounts, it would be unlikely they would be able to continue their advocacy and organizing efforts or to obtain information from X in the long run. With the lack of any real credible alternatives available, the freedom of expression of</li> </ul>	<p>via a premium membership, X reported selling 44 million Blue Checkmarks in a single day, earning up to \$660 million (<a href="#">source</a>). However, the increase in sponsored verification was actually found to heighten the spread of false information on the platform, with countless impersonation accounts (<a href="#">source</a>), an increase in disinformation (<a href="#">source</a>), phishing attacks, like those experienced by our rights holders, and many other examples of abuse and misuse.</p> <ul style="list-style-type: none"> <li>● <b>Leverage:</b> High potential. X would easily be able to alter and improve the Blue Checkmark “for a fee” verification system and revert to a more authentic verification system like in the past, if it wanted to.</li> </ul>
--	--	--	--

		rightsholders and the populations they serve would be greatly impacted in the next five years and even beyond.	
--	--	--	--

## 9. Phase 4: Recommendations

Based on the context analysis and impact assessment above, and in line with the fundamental UNGP principles of *protect* (UNGP 18, 23), *respect* (UNGP 11) and *remedy* (UNGP 22, 29, 31), we make the following recommendations to support X in maximizing its positive human rights impacts and mitigating adverse impacts. Notably, UNGP 24 requires businesses to prioritize the most severe mitigations or where delayed response would make the rights impact irremediable, while UNGP 23 indicates that businesses should comply with all applicable laws and respect internationally recognised human rights, to honour internationally recognized human rights principles when faced with conflicting requirements. Therefore, recommendations are made upon two dimensions:

1. **Timeline:** A best-effort consideration of the change possible within the short (within 1 year), medium (within 3 years) and long term (within 5 years). The timeline reflects what should be possible within the country's context to maximize respect for international human rights laws and principles, in line with UNGP 23; and
2. **Positive Impact of Recommendation:** Impact of recommendation on mitigating or remedying any negative rights impacts. Our prioritization considers UNGP 24, which requires prioritizing mitigations based on their severity.

### 9.1. Improving the Respect and Remedy Mechanisms of Human Rights Within X

According to UNGP 11, companies should respect human rights by not infringing on the human rights of rightsholders and mitigating adverse human rights impacts. Furthermore, UNGP 29 states that companies should establish or participate in grievance mechanisms, while UNGP 31 establishes eight criteria for effective grievance mechanisms.

Our specific recommendations to meet these UNGP requirement(s) are as follows:

Recommendation	Timeline	Impact
With regards to the Blue Checkmark "for a fee," X should reconsider only allowing the Blue Checkmark verification through its paid premium subscription. Our research and stakeholders' experiences have shown that the current system has encouraged impersonation attempts, disinformation campaigns, spear-phishing attacks, and hacking, thus fundamentally preventing the target group from exercising their right to freedom of expression, as well as contributing to information threats and transnational repression. The current Blue Checkmark verification requirements should be revised to prioritize information accuracy.		
<ul style="list-style-type: none"> <li>As a stop-gap measure, there should be rapid action focused on due diligence and accurate verification of information, such as the identities of the Blue Checkmark holders, through improving X's current content moderation AI algorithms to precisely address the aforementioned issues that have come about as a result of the Blue Checkmark, as well as an increase in the number of human reviewers which would be able to reinforce these efforts.</li> </ul>	Within 1 year	High
<ul style="list-style-type: none"> <li>In the mid and long term, it is recommended that X's Blue Checkmark verification should revert to a system that emphasizes due diligence and accuracy of information with adequate human rights safeguards, to prevent overreach. To do this, X would need to take active steps to ensure and verify that an X account is</li> </ul>	Within 5 years	High

actually owned by the person or organization it claims to represent. At the same time, verification requirements should be cognizant of the existing real-name and ID verification regulations under the Cybersecurity Law in China. <sup>8</sup>		
Content moderation should not overly rely on AI models so as to address both under- and over-moderation, both of which significantly affect rightsholders. An immediate solution would be to utilize more human moderators with specialized training. Such human moderators, not based inside the PRC, should be context-aware, i.e. with a good understanding of issues involved, and linguistically diverse. A long-term solution would be to refine the AI models, which would go towards better trained data sets and models, and enhancing human control over decisions made by AI and ensuring legality, necessity, and proportionality in content moderation decisions.		
<ul style="list-style-type: none"> <li>Algorithms should be committed to accuracy of information instead of simply focusing on high engagement with low credibility. Proactive content moderation, by system or human, should only be channelled towards addressing actual issues such as impersonation attempts, disinformation campaigns, spear-phishing attacks, and hacking, and not to the extent that it becomes over-moderation, where relevant and legitimate content gets taken down.</li> </ul>	Within 1 year	High
<ul style="list-style-type: none"> <li>Account suspension, especially of human rights defenders, must have a clear basis and should not be done at will without notification. There must be remedies for reinstatement. For instance, this could be done through transparently communicating to users about the content being moderated and providing appeal mechanisms and improving user control mechanisms such as blocking or reporting.</li> </ul>	Within 5 years	High
Fundamentally, X should be committed to human rights and its indivisibility. As such, X cannot claim to uphold freedom of expression while remaining silent or taking contradictory approaches on other concerns in relation to privacy and access to information, amongst others, as these rights are equally important to all users. This requires a holistic change in direction, policies, and systems.		
<ul style="list-style-type: none"> <li>A human rights unit within X would be able to address some of the above concerns, in particular related to human rights defenders using the platforms to circumvent the Great Firewall of China, on an immediate basis. A dedicated human rights unit would serve as a direct grievance mechanism that would deal specifically with human rights related complaints and issues, whereas the system now does not adopt a human rights-centered approach.</li> </ul>	Within 5 years	High

---

<sup>8</sup> The Cyberspace Administration of China's draft measures propose a "Network ID" system where Chinese citizens can voluntarily apply for a digital credential linked to their official identification documents to be used for online verification across internet platforms, aiming to enhance real-name registration and data security across the internet. This raises privacy concerns and freedom of expression issues due to the centralized nature of this identity system.

## 10. Phase 5: Company Engagement

The core purpose of the CLARITI methodology and this report is to enhance engagement with X to enable rightsholders in China to be able to continue to access X freely and without any encumbrances, and to ensure their rights to freedom of expression (including the right of access to information), and rights to privacy.

After this assessment, we hope to engage with X to learn more about its commitment to human rights and how X is managing human rights issues. Since our context analysis, impact assessment, and recommendations are limited to information available in the public domain, we welcome the opportunity for a more in-depth discussion with X to understand the opportunities and challenges X faces, and to discuss potential avenues for constructive collaboration.

This report is intended to facilitate open dialogue around our constructive criticisms and suggestions in order to help X consider the real-world rights impact of its policies, and hopefully take steps to adjust said policies as needed. X's status as a vital platform for the free flow of information makes it a uniquely valuable ally in the fight for human rights and freedom of expression. Therefore, we seek to work collaboratively with X to maximize its positive human rights impacts and mitigate any adverse impacts.

## Appendix A – List of Online Sources Consulted

### Online Sources Consulted for Country Context Analysis

No.	Organization	Page Title	URL
1	Amnesty International	Countries	<a href="https://www.amnesty.org/en/countries/">https://www.amnesty.org/en/countries/</a>
2	Baker McKenzie	Global Data Privacy & Security Handbook	<a href="https://resourcehub.bakermckenzie.com/en/resources/data-privacy-security">https://resourcehub.bakermckenzie.com/en/resources/data-privacy-security</a>
3	Constitute	Constitutions	<a href="https://constituteproject.org/constitutions?lang=en">https://constituteproject.org/constitutions?lang=en</a> .
4	Council of Europe	Complete list of the Council of Europe's treaties	<a href="https://www.coe.int/en/web/convention/full-list">https://www.coe.int/en/web/convention/full-list</a>
5	DataReportal	Complete Report Library	<a href="https://datareportal.com/library">https://datareportal.com/library</a>
6	Freedom House	Countries and Territories	<a href="https://freedomhouse.org/countries/freedom-world/scores">https://freedomhouse.org/countries/freedom-world/scores</a>
7	Global Data Privacy & Security Handbook		<a href="https://resourcehub.bakermckenzie.com/en/resources/data-privacy-security">https://resourcehub.bakermckenzie.com/en/resources/data-privacy-security</a> .
8	Global Network Initiative (GNI)	Country Legal Frameworks Resource (CLFR)	<a href="https://clfr.globalnetworkinitiative.org/">https://clfr.globalnetworkinitiative.org/</a>
9	Office of the United Nations High Commissioner for Human Rights (OHCHR)		<a href="https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?Lang=en">https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?Lang=en</a>
		Assessing the Effectiveness of National Human Rights Institution	<a href="https://www.ohchr.org/sites/default/files/Documents/Publications/NHRIen.pdf">https://www.ohchr.org/sites/default/files/Documents/Publications/NHRIen.pdf</a>
10	Organisation of American States	Basic Documents in the Inter-American System	<a href="https://www.oas.org/en/IACHR/jsForm/?File=/en/iachr/mandate/basic_document_s.asp">https://www.oas.org/en/IACHR/jsForm/?File=/en/iachr/mandate/basic_document_s.asp</a>
11	South Asian Association for Regional Cooperation (SAARC)	Agreements & Conventions	<a href="https://www.saarc-sec.org/index.php/resources/agreements-conventions?limit=20&amp;limitstart=20">https://www.saarc-sec.org/index.php/resources/agreements-conventions?limit=20&amp;limitstart=20</a>
12	The African Union	OAU/AU Treaties, Conventions, Protocols & Charters	<a href="https://au.int/en/treaties">https://au.int/en/treaties</a>
13	The Association of Southeast Asian Nations (ASEAN)	Human Rights	<a href="https://asean.org/our-communities/asean-political-security-community/rules-based-people-oriented-people-centred/human-rights/">https://asean.org/our-communities/asean-political-security-community/rules-based-people-oriented-people-centred/human-rights/</a>
14	United Nations Conference on Trade and Development	Data Protection and Privacy Legislation Worldwide	<a href="https://unctad.org/page/data-protection-and-privacy-legislation-worldwide">https://unctad.org/page/data-protection-and-privacy-legislation-worldwide</a>
15	US Department of State	Country Reports on Human Rights Practices	<a href="https://www.state.gov/reports-bureau-of-democracy-human-rights-and-labor/country-reports-on-human-rights-practices/">https://www.state.gov/reports-bureau-of-democracy-human-rights-and-labor/country-reports-on-human-rights-practices/</a>
16	World Bank	Classification of Fragile and Conflict-Affected Situations	<a href="https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/harmonized-list-of-fragile-situations">https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/harmonized-list-of-fragile-situations</a>
17	World Justice Project	WJP Rule of Law Index	<a href="https://worldjusticeproject.org/rule-of-law-index/global/2022">https://worldjusticeproject.org/rule-of-law-index/global/2022</a>

## Online Sources Consulted for Company Context Analysis

No.	Organization	Page Title	URL
1	AALEP - Association of Accredited Public Policy Advocates to the European Union	Biggest tech associations representing digital companies	<a href="https://www.aalep.eu/biggest-tech-associations-representing-digital-companies">https://www.aalep.eu/biggest-tech-associations-representing-digital-companies</a>
2	Amnesty International	What the EU's Digital Services Act means for human rights and harmful Big Tech business models	<a href="https://www.amnesty.org/en/wp-content/uploads/2022/07/POL3058302022ENGLISH.pdf">https://www.amnesty.org/en/wp-content/uploads/2022/07/POL3058302022ENGLISH.pdf</a>
3	ARTICLE 19	ARTICLE 19 recommendations for the Digital Services Act Trilogue	<a href="https://www.article19.org/wp-content/uploads/2022/02/A19-recommendations-for-the-DSA-Trilogue.pdf">https://www.article19.org/wp-content/uploads/2022/02/A19-recommendations-for-the-DSA-Trilogue.pdf</a>
4	ARTICLE 19	EU: Digital Services Act does not provide a green light for platform blocking	<a href="https://www.article19.org/resources/eu-dsa-does-not-provide-a-green-light-for-platform-blocking/">https://www.article19.org/resources/eu-dsa-does-not-provide-a-green-light-for-platform-blocking/</a>
5	Business & Human Rights	Mandatory Due Diligence	<a href="https://www.business-humanrights.org/en/big-issues/mandatory-due-diligence/">https://www.business-humanrights.org/en/big-issues/mandatory-due-diligence/</a>
6	CIRIGHTS	CIRIGHTS	<a href="https://cirights.com/">https://cirights.com/</a>
7	Digital Trust & Safety Partnership	Best Practices Framework	<a href="https://dtspartnership.org/best-practices/">https://dtspartnership.org/best-practices/</a>
8	Electronic Frontier Foundation	Who Has Your Back? Censorship Edition 2019	Santa Clara: <a href="https://www.eff.org/wp/who-has-your-back-2019">https://www.eff.org/wp/who-has-your-back-2019</a>
9	Fierce Telecom	Who are the wireless and wired telecom trade associations?	<a href="https://www.fiercetelecom.com/telecom/who-are-wireless-and-wired-telecom-trade-associations">https://www.fiercetelecom.com/telecom/who-are-wireless-and-wired-telecom-trade-associations</a>
10	Ranking Digital Rights	The 2022 Telco Giants Score Card	<a href="https://rankingdigitalrights.org/tgs22/">https://rankingdigitalrights.org/tgs22/</a>
		The 2022 Big Tech Score Card	<a href="https://rankingdigitalrights.org/bts22/">https://rankingdigitalrights.org/bts22/</a>
		2020 Ranking Digital Rights Corporate Accountability Index	<a href="https://rankingdigitalrights.org/index2020/">https://rankingdigitalrights.org/index2020/</a>
		Governance and Management Oversight	Tech: <a href="https://rankingdigitalrights.org/bts22/indicators/G2">https://rankingdigitalrights.org/bts22/indicators/G2</a>  Telco: <a href="https://rankingdigitalrights.org/tgs22/indicators/G2">https://rankingdigitalrights.org/tgs22/indicators/G2</a>
		Governments and regulations	<a href="https://rankingdigitalrights.org/index2020/indicators/G4a">https://rankingdigitalrights.org/index2020/indicators/G4a</a>
		Processes for policy enforcement	<a href="https://rankingdigitalrights.org/index2020/indicators/G4b">https://rankingdigitalrights.org/index2020/indicators/G4b</a>
		Targeted advertising	<a href="https://rankingdigitalrights.org/index2020/indicators/G4c">https://rankingdigitalrights.org/index2020/indicators/G4c</a>
		Algorithmic System	<a href="https://rankingdigitalrights.org/index2020/indicators/G4d">https://rankingdigitalrights.org/index2020/indicators/G4d</a>

		Access to terms of service	<a href="https://rankingdigitalrights.org/bts22/indicators/F1a">https://rankingdigitalrights.org/bts22/indicators/F1a</a>
		Process for terms of service enforcement	<a href="https://rankingdigitalrights.org/bts22/indicators/F3a">https://rankingdigitalrights.org/bts22/indicators/F3a</a>
		Scorecards and Rankings	<a href="https://rankingdigitalrights.org/rankings-report-cards/">https://rankingdigitalrights.org/rankings-report-cards/</a>
11	The Santa Clara Principles		<a href="https://santaclaraprinciples.org">https://santaclaraprinciples.org</a>
12	United Nations Global Compact	Our Participants	<a href="https://unglobalcompact.org/what-is-gc/participants">https://unglobalcompact.org/what-is-gc/participants</a>
13	World Benchmarking Alliance	Digital Inclusion Benchmark	<a href="https://www.worldbenchmarkingalliance.org/publication/digital-inclusion/companies/">https://www.worldbenchmarkingalliance.org/publication/digital-inclusion/companies/</a>
14	World Benchmarking Alliance	Digital Inclusion Benchmark	<a href="https://www.worldbenchmarkingalliance.org/publication/digital-inclusion/companies/">https://www.worldbenchmarkingalliance.org/publication/digital-inclusion/companies/</a>
15	World Favor	The ultimate guide to Human Rights Due Diligence laws – who’s affected and how to comply	<a href="https://blog.worldfavor.com/the-complete-list-of-national-human-rights-due-diligence-laws-whos-affected-and-how-to-comply">https://blog.worldfavor.com/the-complete-list-of-national-human-rights-due-diligence-laws-whos-affected-and-how-to-comply</a>

## Online Sources Consulted for Impact Assessment

No.	Organization	Page Title	URL
1	Access Now	How’s your country on Net Neutrality?	<a href="https://www.accessnow.org/how-your-country-on-net-neutrality/">https://www.accessnow.org/how-your-country-on-net-neutrality/</a>
		Five excuses governments (ab)use to justify internet shutdowns	<a href="https://www.accessnow.org/five-excuses-governments-abuse-justify-internet-shutdowns/">https://www.accessnow.org/five-excuses-governments-abuse-justify-internet-shutdowns/</a>
		The impact of forced data localisation on fundamental rights	<a href="https://www.accessnow.org/the-impact-of-forced-data-localisation-on-fundamental-rights/">https://www.accessnow.org/the-impact-of-forced-data-localisation-on-fundamental-rights/</a>
2	ARTICLE 19	Watching the Watchmen: Content moderation, governance, and freedom of expression	<a href="https://www.article19.org/wp-content/uploads/2023/02/Watching-the-watchmen-UPDATE-Jan2023-P04.pdf">https://www.article19.org/wp-content/uploads/2023/02/Watching-the-watchmen-UPDATE-Jan2023-P04.pdf</a>
		Taming Big Tech: A pro-competitive solution to protect free expression	<a href="https://www.article19.org/wp-content/uploads/2023/02/Taming-big-tech-UPDATE-Jan2023-P05.pdf">https://www.article19.org/wp-content/uploads/2023/02/Taming-big-tech-UPDATE-Jan2023-P05.pdf</a>
2	CSO	The 15 biggest data breaches of the 21st century	<a href="https://www.csoonline.com/article/534628/the-biggest-data-breaches-of-the-21st-century.html">https://www.csoonline.com/article/534628/the-biggest-data-breaches-of-the-21st-century.html</a>
3	Daily Dot	What is zero-rating?	<a href="https://www.dailydot.com/debug/zero-rating/">https://www.dailydot.com/debug/zero-rating/</a>
4	European Commission	What personal data is considered sensitive?	<a href="https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-">https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-</a>

			<a href="#">data/what-personal-data-considered-sensitive_en</a>
5	Facebook	Our Approach to Facebook Feed Ranking	<a href="https://transparency.fb.com/features/ranking-and-content/">https://transparency.fb.com/features/ranking-and-content/</a>
6	Freedom House	Countries	<a href="https://freedomhouse.org/countries/freedom-net/scores">https://freedomhouse.org/countries/freedom-net/scores</a>
7	GDPR Text	Article 29 Working Party Guidelines on transparency under Regulation 2016/679	<a href="https://gdpr-text.com/guidelines/transparency/">https://gdpr-text.com/guidelines/transparency/</a>
8	General Data Protection Regulation (GDPR)	GDPR Overview	<a href="https://gdpr.eu">https://gdpr.eu</a>
9	Global Data Privacy & Security Handbook		<a href="https://resourcehub.bakermckenzie.com/en/resources/data-privacy-security">https://resourcehub.bakermckenzie.com/en/resources/data-privacy-security</a>
10	Global Network Initiative (GNI)	Country Legal Frameworks Resource (CLFR)	<a href="https://clfr.globalnetworkinitiative.org/">https://clfr.globalnetworkinitiative.org/</a>
		Defining Direct Access: GNI calls for greater transparency and dialogue around mandatory, unmediated government access to data	<a href="https://globalnetworkinitiative.org/defining-direct-access-2/">https://globalnetworkinitiative.org/defining-direct-access-2/</a>
		Implementation Guidelines For The Principles of Freedom Of Expression And Privacy	<a href="https://globalnetworkinitiative.org/wp-content/uploads/2018/08/Implementation-Guidelines-for-the-GNI-Principles.pdf">https://globalnetworkinitiative.org/wp-content/uploads/2018/08/Implementation-Guidelines-for-the-GNI-Principles.pdf</a>
11	Internet Society	Global Internet Shutdowns	<a href="https://pulse.internetsociety.org/shutdowns">https://pulse.internetsociety.org/shutdowns</a>
12	NetBlocks		<a href="https://netblocks.org/reports">https://netblocks.org/reports</a>
13	PayPal	List of Third Parties (other than PayPal Customers) with Whom Personal Information May be Shared	<a href="https://www.paypal.com/ie/legalhub/third-parties-list">https://www.paypal.com/ie/legalhub/third-parties-list</a>
14	Ranking Digital Rights	The 2022 Telco Giants ScoreCard	<a href="https://rankingdigitalrights.org/tgs22/">https://rankingdigitalrights.org/tgs22/</a>
		The 2022 Big Tech Score Card	<a href="https://rankingdigitalrights.org/bts22/">https://rankingdigitalrights.org/bts22/</a>
15	TechTarget	Top 10 customer data privacy best practices	<a href="https://www.techtarget.com/searchcustomerexperience/tip/Top-customer-data-privacy-best-practices">https://www.techtarget.com/searchcustomerexperience/tip/Top-customer-data-privacy-best-practices</a>
16	The Santa Clara Principles		<a href="https://santaclaraprinciples.org/">https://santaclaraprinciples.org/</a>
17	Twitter	A new era of transparency for Twitter	<a href="https://blog.twitter.com/en_us/topics/company/2023/a-new-era-of-transparency-for-twitter">https://blog.twitter.com/en_us/topics/company/2023/a-new-era-of-transparency-for-twitter</a>
18	United Nations Conference on Trade and Development	Data Protection and Privacy Legislation Worldwide	<a href="https://unctad.org/page/data-protection-and-privacy-legislation-worldwide">https://unctad.org/page/data-protection-and-privacy-legislation-worldwide</a>