Letting the Sun Shine In

Transparency and Accountability in the Digital Age

In the series:

World Trends in Freedom of Expression and Media Development

Key trends:

- There is growing global multistakeholder support for enhanced transparency as a means of increasing the accountability of internet companies for their operations.

- Greater transparency would strengthen issues central to UNESCO's work, from freedom of expression and privacy, through to education, the sciences and culture.

- Existing corporate transparency reports provide a lot of data, but still have significant gaps and cover different issues and in different ways, making comparisons difficult and good practice hard to identify.

- Existing initiatives to promote greater internet transparency have so far remained largely aspirational, lacked substantive impact and occurred in relative isolation of each other.

- Vast differences in types, sizes, business models and engineering of internet platform companies show a need for high-level principles, focused on achieving outcomes.

- The illustrative high-level principles in this brief can form a basis for future discussions towards a framework for transparency to guide companies, policy makers and regulators.

Summary

This brief sets out the rationale for exploring a framework for greater transparency of internet companies as part of UNESCO's ongoing multi-stakeholder work to promote and protect human rights in the digital ecosystem.[1] Focused especially on companies that provide services that manage content and which impact on expression in various ways, the brief outlines the potential impact of these companies on concerns such as freedom of expression and privacy, through to the impact on democracy more generally, as well as issues of the free flow of information and knowledge, media and information literacy, scientific research and cultural identity.

The paper reviews existing transparency mechanisms of a selection of companies as well as some of the initiatives that seek to promote greater transparency. Looking at the advantages and limitations of current approaches, it discusses the value of a unifying initiative to elaborate the meaning of transparency for key digital actors.

The brief sets out a preliminary selection of illustrative principles that could enhance the transparency of internet platform companies. These proposed, high-level, principles can form the basis of future discussions with companies and other stakeholders. Such a dialogue could explore how the principles can be further developed, and then transformed into a more detailed operational framework. The resulting output could serve as a global point of reference for companies, policy makers and regulators.

The relevance of this initiative to UNESCO is related to how the Organization's work increasingly intersects with a range of digital issues and their many opportunities and challenges. Through its international role in advancing education, science, culture, freedom of expression and access to information, UNESCO has a direct interest in greater transparency about the workings of the digital ecosystem. The discussion that follows elaborates on this interest, what improved transparency can entail, and what benefits can result. It identifies a common approach to the topic based on agreed UNESCO positions and with a view to informing the Organization's 2022-2029 medium-term strategy and implementation of a proposed UNESCO standard-setting instrument on the ethics of artificial intelligence.

Introduction

There is growing global discussion that supports enhanced transparency as an optimum means of making internet companies more accountable for their operations which impact increasing areas of social life.

Improved transparency by the companies would provide more information to users as well as help to give an evidence-base to the wider public debate about the impact of the companies on democracy, free expression and privacy. Current discussions about transparency embrace everything from regulatory approaches (there are legislative initiatives in at least 40 countries which seek to regulate online content), through to self-regulation and multi-stakeholder approaches, social responsibility initiatives, funding incentives and legal cases.

For their part, internet companies, reporting to their shareholders and balancing this with user interests and public opinion, are negotiating a cacophony of ideas. Some of these ideas reflect limited knowledge of the operational architecture as deployed, and the character of artificial intelligence in regard to corporate operations. Civil

---

1 The brief has benefited enormously from a series of bilateral informal consultations with several internet companies, regulators and experts from different countries both in the Global North and in the Global South. The views expressed herein are not necessarily the views of those consulted. This publication provides a foundation for further processes of consultation by UNESCO.

society groups, meanwhile, are increasingly active in questioning the credibility of the companies as good corporate citizens. UNESCO therefore envisages a consensus-building approach that adds value to the global debate, underlining inter alia why companies themselves should embrace a comprehensive approach to transparency that goes beyond existent commitments in this area.

When considering the transparency of internet companies, there are different approaches possible. One is related to the structure of the business itself – the terms of service, data policies, advertisements, etc., as well as associated processes of implementation, monitoring, appeal and review. An alternative approach is to set out issues in relation to the potential impact of internet companies—for example by their impact upon democracy, freedom of expression or privacy. This brief seeks to bridge these, and to initiate a discussion that can be further developed through a more elaborated set of principles and a detailed framework for operationalization.

It is noted that there is a vast array of internet companies, across all of whom transparency in general can aid accountability and international consensus about the application of universal standards for human rights and freedoms. These digital actors range from those providing the basic telecommunications infrastructure, through internet exchange points, those dealing with connectivity, internet security providers, e-commerce operators, gaming, search, browser software providers, digital advertising brokers, and app-stores, through to those which provide a platform for sharing content, and a growing number of specific applications companies related to content distribution.

UNESCO's concerns are especially relevant to those companies that act as platforms for users to share or access content. Not least of these are those whose operations also include information produced through journalism—and at the same time, provide oxygen for disinformation and hate speech. Most of these entities make money through advertising based upon the harvesting of user data. Consequently, it is this category of platform companies that are the focus of this issue brief. But even within platform companies there is wide diversity. Some specialize in video, others in photos, audio chat, text, email, and memes; many operate in hybrid formats. Some serve youth content markets. Some cover many topics; others are thematically focused. Some offer relatively closed groups and messaging services, others function more as public fora. Many of these actors provide search functions in greater or lesser degrees, and some are involved in e-commerce too. Another set are entities that combine user-generated content with centrally provided content, and those which have different engagements of users in moderating the prioritisation of content. However, all curate the content experience of users, and most also intervene further in addition to curation through moderating content (often via what the industry calls 'trust and safety' policies and staffing).

In addition to insufficiently capturing the nuances among these varied platform companies, initiating comprehensive and detailed transparency requirements in the absence of wider high-level principles could favour larger companies. The result could be a perverse outcome of restricting competition by placing too high a burden on small or new entrants, working against the importance of pluralism as a factor for accountability. High-level principles for transparency should therefore be workable for companies of all sizes.

This brief therefore identifies a selection of such high-level transparency principles that could be relevant generally to all platform companies. It recognizes, nevertheless, that many proposals have been made for elaborating a tiered set of operational requirements according to the size of companies, such as their numbers of users or strategic market power.

Why transparency?

Transparency has become a buzzword in the field of digital technology, fuelled by the increasing dominance of internet companies in public life and, alongside their benefits, also their potential for causing harm to human rights. Transparency, drawing from the use of the term in relation to the universal right to freedom of information, has a dual dimension. It covers arrangements for access to information and data, as well as those for proactive disclosure.

Such arrangements make it possible for external stakeholders to gain insight into the impact of companies upon people's ability to express themselves, protect their privacy, access journalism, recognise and counter hate speech and disinformation, and share and exchange knowledge. Too often the workings of the companies (and their monitoring of impact) are opaque and therefore difficult to assess, making it difficult for outsiders to respond to problems and opportunities from the point of view of human rights and sustainable development concerns like the right to health, the right to equality (including gender equality) and the right to environmental protection.

Given that several of these internet corporations have the status of mega-institutions, the externalities of their business models and associated internal operations, have large significance for public life. A strong case can therefore be made from the point of view of external stakeholders to want greater transparency if there is to be societal trust that commercial actors are not putting profit above respect for human rights and progress in sustainable development. Many of the companies themselves already have degrees of transparency reporting, and some have committed to increased openness. The main debate is therefore not whether there should be transparency, but rather issues like transparency over what, how much, to whom, and for what reasons and expected outcomes.

Transparency is closely linked to companies' legal and ethical obligations to respect human rights, for example as set out in the **UN Guiding Principles on Business and Human Rights** ('Ruggie Principles'), and the **UN Global Compact**. Companies in other industries, such as pharmaceutical and automotive, face extensive regulatory requirements in the public interest, including regulation for transparency. Many commercial entities also practice voluntary regimes of openness consistent with international norms. In principle, internet companies as private entities have at least the same social obligation as other sectors which impact upon the public. For platform companies, there are particular balances to be struck in fulfilling human rights obligations. An example is that between enabling openness and transparency, and protecting personal privacy. Another case is balancing how the right to property and the protection of commercial secrets can nevertheless allow for the public interest in open access. Striking these balances merits continuous exploration and recalibration.

Transparency is highly relevant to UNESCO. As part of the United Nations, UNESCO operates within the frameworks of universal human rights, the 2030 Agenda for Sustainable Development, and the strategic objectives decided by its 193 Member States. Broadly, transparency is recognized as necessary for building peaceful, just, and inclusive societies. Sustainable Development Goal 16.10 calls for public access to information and fundamental freedoms. The right to free expression includes the right to seek and receive information, and can apply not only to the accountability of states, but also that of private actors performing activities contracted for, or impacting upon, public interest purposes.

The idea of transparency is also integral to UNESCO's Member States' agreement on the concept of Internet Universality and the framework of Rights, Openness, Accessibility and Multistakeholder governance (ROAM). Without transparency of how human Rights are being balanced online, abuses and violations can take place without knowledge or redress. A digital ecology that is not Open to all entrants can lead to hidden dominance and unforeseen data flows and use. In order to ensure Accessibility, transparency is needed as a factor for pricing policies, and for the public to be empowered with knowledge of the digital environment, such as the provision 'free services' which disguises how the user's own information is being monetised, as well as obscuring disinformation actors and counter-measures, etc. Multi-stakeholder Governance issues, ranging from decisions on values through to regulations, depend on the degree to which there is transparency of process and participants (including corporate lobbyists).

In addition, transparency has emerged in UNESCO's work to elaborate a global standard-setting instrument on the ethics of AI. Proposed as a core principle in the **first draft of the Recommendation on the Ethics of Artificial Intelligence**, transparency is noted as "often a crucial precondition to ensure that fundamental human rights and ethical principles are respected, protected and promoted" and "necessary for relevant national and international liability legislation to work effectively".

Increased transparency of internet companies would be valuable across UNESCO's programmes to advance education, the sciences, culture, and communication and information. For example, greater insight into how companies identify and deal with online hate speech could strengthen UNESCO's work to promote global citizenship education, media and information literacy, inclusion and non-discrimination. Transparency related to reports to companies made by media workers who have experienced online abuse, and actions taken in response to these reports, would reinforce efforts to enhance the safety of journalists. Through greater transparency into the origins, extent, formats, themes and flow of disinformation, UNESCO (and other UN actors) could design evidence-based strategies to counter such content and anticipate emerging challenges well in advance.

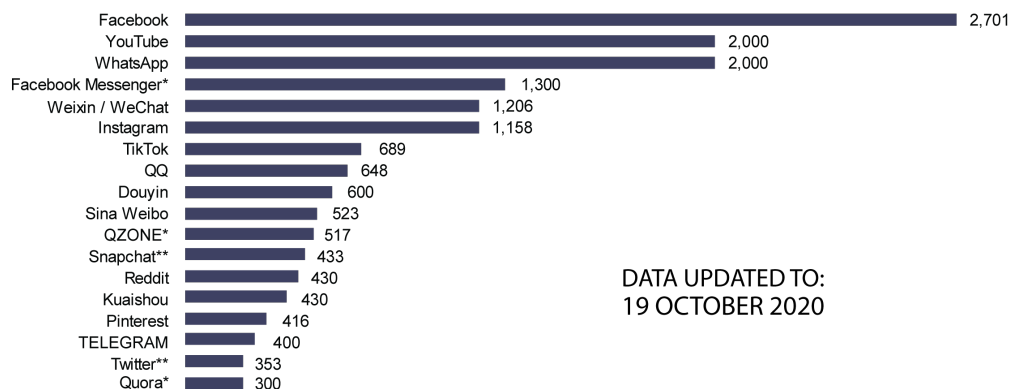# Analysing the internet companies and their impact

The COVID-19 pandemic has driven home just how powerful the internet economy has become. Social, political and economic life has significantly moved online, and many aspects of societies are now dependent upon internet companies. As the public and societal importance of internet companies becomes more apparent, questions about how an ecosystem, upon which we all depend, is made democratically accountable have become more urgent.

When this paper refers to 'platform companies', it means those with the facility to bring together two or more groups, for example, users and users, users and advertisers, consumers and producers, producers and suppliers, voters and candidates, as well as those who have questions and those whose content is prioritised as relevant answers. Such platforms in turn produce network effects, in that the more users there are on the platform, the more others are drawn into using that platform and usually find their accumulated data locked into 'walled gardens' even if they decide to leave. Platforms then gather data from the users, analyse the data and produce new services or products that reinforce use as well as draw in more users, feeding the network effect. The mounting data is monetised through sales, or through selling targeted marketing opportunities to advertisers. These network effects facilitate the emergence and continued dominance of a small number of mega-platforms that have achieved a monopoly market position. In addition, a number of platform companies model their operations on reinforcing content and communications based on users' past patterns, particular profiles and allocation to specific cohorts, which can impact on what people see as important, normal and legitimate.

The data on the dominance of a small number of platforms is striking. Four of the top six social media platforms (Facebook, WhatsApp, Instagram and Messenger) are owned by the same company which also entails enormous power concentrated in one individual.

THE WORLD'S MOST-USED SOCIAL PLATFORMS
BASED ON MONTHLY ACTIVE USERS, ACTIVE USER ACCOUNTS, OR ADDRESSABLE ADVERTISING AUDIENCES (IN MILLIONS)

Platform	Users (millions)
Facebook	2,701
YouTube	2,000
WhatsApp	2,000
Facebook Messenger*	1,300
Weixin / WeChat	1,206
Instagram	1,158
TikTok	689
QQ	648
Douyin	600
Sina Weibo	523
QZONE*	517
Snapchat**	433
Reddit	430
Kuaishou	430
Pinterest	416
TELEGRAM	400
Twitter**	353
Quora*	300

DATA UPDATED TO:
19 OCTOBER 2020

Source: https://blog.hootsuite.com/social-media-demographics/

The dominant internet platform companies have the advantage that they can scale very quickly compared to, for example, traditional retail or manufacturing operations. As contemporary cases highlight, growth is often achieved at the expense of competitors, through contested practices ranging from buyouts, to copying of innovative features, and in some cases, even national legal measures which ban foreign competitors outright. At the same time, it is the lack of regulatory constraints which has provided the frictionless potential enabling the astonishing growth of these internet giants. This lack, however, has also often resulted in scant attention being paid, and little budget share allocated, towards monitoring challenges, creating guardrails like independent oversight, or commissioning human rights impact assessments. Scale then becomes cited as an inherent obstacle in addressing challenges, even though it has been mastered in a number of areas such as copyright control, spam and child abuse imagery—where indeed there has been regulation. On the other hand, statutory restrictions of other areas of content brings with it several risks from the point of view of freedom of expression.

Challenges arise from the logics of the companies' architectures, and the role of users, although this does not mean there is a symmetry of power or obligation on the two sides. Many platforms enable users to upload and share content, within the frame of terms of service, which may not always be effectively or universally applied, and they index and prioritise content in particular ways related to their business models. In some cases, this extends to external use of content such as limits on scraping, or even blocks on researcher-user relationships.

Platforms operate with less direct editorial power than news media, where content is mediated by editors or publishers. Online, there are more opportunities for users to create and edit content, but this is by no means without boundaries. Users may think they are talking to each other, but they are also talking to the company hosting them. Those who think they are simply searching data are also revealing much of themselves. In short, no platform is a content-neutral stage, independent of company values, motives and algorithms related to its raison d'être and financing model.

Understanding this means recognising curation models, which affect both content prioritisation and user activity (such as whether anonymity is allowed, age limits are set and enforced, bots are permitted, automated activity is allowed, or if there are caps on the extent of sharing content).

The automated aspect of curation works at a systemic level through, amongst other things, the order of content presented to each user, the signalling of trending topics, auto-completion of search queries, and recommendations to join or follow groups and individuals. There are further controls over content through active moderation efforts, although some moderation is outsourced (such as to fact-checkers) or is influenced by parties recognised as 'trusted flaggers' or user-enlistment to identify and act upon apparent problems.

Within all these parameters, the platform services have allowed many new and previously excluded voices to be heard. They have also facilitated the

sharing of knowledge through search, as well as through content platforms like Wikipedia. The opportunity has facilitated citizen mobilisation, as we have seen in many protests that have taken place worldwide since the Arab Spring. The #metoo and #blacklivesmatter movements have used content platforms to connect people with each other, share experiences and organise for their rights.

On the other hand, these same technological affordances have also enabled and encouraged terrorist and racist groups to recruit and organise, as well as for a host of groups to intimidate and abuse online those whom they see as their enemies. Anonymity, though a benefit in some circumstances, often appears to allow people to express themselves in anti-social ways they would not dare if their identity was known. Closed groups may be used to advance or oppose human rights. The environment is also vulnerable to manipulation through the creation of false identities and the dissemination of disinformation, which is often funded and well-organised, and which is increasingly used to affect democratic elections globally as well as to attack critical journalists and civil society actors.

Yet, instead of being agnostic about such content—which itself would be a source of concern—many platforms, driven by their business model to accumulate users and their data, function to amplify the problem. The 'engagement-enragement' relationship linked to these models has given prominence to such content and to groups trafficking in racism, hate and conspiracies. The content expressed in these cases is often also wielded against people who benefit from the enhanced freedom of expression opportunities offered by the expansion of the internet. For many users, social media platforms have become extremely unpleasant places, particularly for women, migrants and minorities amongst others.

Underlying widespread abusive content is a mix of curation through defined algorithms and architectures, combined with insufficient active moderation. Historically, investors in technology and leaders of many companies have been more concerned with how quickly a platform acquires users, than considering unintended consequences, such as the spread of violent extremist views and fuelling content that violate people's rights or misleads them on key issues such as health.

The speed of digital communication has allowed disinformation and misinformation to spread rapidly and virally, exploiting trusted networks of family and friends whom users increasingly rely upon. Social networks of the people we like and trust, perhaps too easily, can become enclosed 'bubbles', which exclude alternative points of view or information, limit democratic debate and exchange, exacerbate divisions and increase polarisation within society.

At the same time, many platform users are not even aware that they are being categorised within particular cohorts (also known as 'Lookalike Audiences'), whose evolving demographics and psychographics are sold to enable the highest bidder to micro-target individuals with particular messages, embedding world views that are antipathetic to human rights and sustainable development. Nor are many users aware of the dark patterns that nudge particular behaviours

within a platform interface. Particular concerns have been raised in relation to children's vulnerabilities in terms of data harvesting and targeted advertising.

These phenomena can impact adversely on a range of rights. The right to express oneself becomes subject to attack; the right to privacy becomes vulnerable to 'outing' and 'doxing'; the right to safety and security can be jeopardised by hate speech. Dignity and reputation can be casualties, and the right to political participation and the right to public health can be damaged. Autonomous development of the human personality becomes subject to subtle manipulation.

In this context, new threats to democracy have emerged online, as a barrage of false content, sometimes produced by bots or fake identities, can cause people to lose faith in the electoral process, even when the actual impact on the integrity of voting systems may be minimal. Micro-targeted advertising is also deployed in attempts to suppress voter turnout, through nudging parts of the electorate into not voting—to the benefit of the rival party.

There have been many attempts to consider how to respond to real or potential harms identified above, sometimes with new regulation that directly concerns swathes of online content. Yet as discussed in the UNESCO/Broadband Commission report **Balancing Act: Countering Digital Disinformation while respecting Freedom of Expression**, such regulatory approaches carry significant risk. Legal efforts to combat so-called 'fake news', for example, often have the effect of disproportionately restricting speech, often for political reasons.

Transparency - the state of play

Between introducing new and potentially intrusive regulation of content, and a completely laissez-faire approach, a third way is increasingly being proposed: to focus more on issues of process, rather than content, and especially to focus on greater transparency of the processes used by the platform companies. In many countries, a degree of transparency is required by law around ownership and the legal status of corporate actors, although controls may still sometimes be hidden. In other cases, companies must provide transparency about their handling of content; for example, in Germany, companies are required to report on steps taken to counter hate speech.

Corporate transparency reporting

Beyond legal requirements for transparency, a growing number of companies have committed to self-regulatory actions, jointly or individually, such as in publishing voluntary transparency reports. Such reports (around 70 by 2018, **according to Access Now**) contain information such as the number of governmental requests received and responded to over a particular time period.

Examples of transparency reporting from leading internet companies:

Twitter has increased the amount of information it makes available in its **Transparency Center** to include information requests, removal requests, potential copyright and trademark infringements, its rules governing enforcement, and information on state-backed information operations and attempts to manipulate the platform.

Google is on its 10th year of producing **transparency reports**, which cover a wide range of topics from security (requests for user information, android security, e-mail encryption), through removals of content under different legal regimes and government request for access to user data, to specific thematic reports such as a **recent report on political advertising** in seven regions and countries.

YouTube's Community Guidelines enforcement report contains transparency data which show how many channels have been removed and why—over two million channels at the last count—with the main reason (more than 75%) being attributed to having contained false or misleading information.

TikTok publishes information about its content moderation, algorithms, and privacy and security practices. Its **transparency reports** show the volume and nature of content removed for violating TikTok's Community Guidelines or Terms of Service, and how TikTok responds to law enforcement requests for information, government requests for content removals, and copyrighted content take-down notices. The company has announced Transparency and Accountability Centers in Los Angeles and Washington, D.C.

Facebook's **transparency report** deals with content removals, government requests for user data by country, intellectual property issues, its community standards and the extent of internet disruptions. It also provides limited access to some data to accredited external actors, through its CrowdTangle subsidiary. Facebook's Oversight Board, established in 2020, has committed to transparency by publicly sharing written statements about its decisions and rationale.

Snap is an example of a much smaller internet company that still publishes a **transparency report**, recording governmental requests for Snapchatters' account information and the amount of content removed (in the most recent report, Snap say that they enforced against 3,872,218 pieces of content.)

It is evident from the above that a considerable amount of information about companies' policies and outcomes is available in the public domain. However, while there are areas of overlap, particularly regarding questions of content removal, each company reports on different issues and in different ways, making simple comparisons impossible. Information about actual practices, not only of moderation, but especially of curation, the approach to trade-offs between rights, and the role of company interests, is usually less forthcoming.

It may be that no transparency system could enable comparisons or ratings systems, given the differences in business practices and engineering. But establishing a set of high-level principles can encourage companies to report on the same issues. Further, if there were then some means of independent assessment, it could allow meaningful comparisons to be made.

At present, the levels of transparency do not generally allow for the possibility for verification of the data presented; therefore, much depends upon what the companies choose to share, and how they interpret it, which reflects largely how they wish to set the agenda of debate.

Legitimate concerns over privacy and the potential misuse of user data are often cited by companies as reasons for caution in sharing data, although systems do exist that seek to effectively anonymise and add noise to data in order to protect individual privacy. Nevertheless, this challenge cannot be avoided, even when there are legal privacy and network security implications. Further complexity arises with transparency in regard to end-to-end encrypted messaging applications such as WhatsApp and Telegram or closed groups of Facebook. However, metadata about communication networks and the behaviours of users is an area where transparency is still feasible.

Transparency therefore is not a straightforward or simple matter. Add to this, the mind-boggling amount of content curated and moderated by the platforms every day. This poses a challenge for developing and implementing any content management policy, and for assessing the inevitable reliance upon automated processes. Yet, here too, a system of transparency that allows for independent assessment of company operations could help users understand the effects that at-scale processes are having on freedom of expression issues and respect for privacy.

As indicated above, for smaller companies, detailed transparency requirements could be onerous to the point of disadvantage compared to bigger players. For this reason, it is important to distinguish between large platform companies and smaller companies in the requirements placed upon them. Adopting high-level principles is a starting point for a more nuanced applied approach to this issue.

Key international and collaborative initiatives

In addition to the transparency reports, companies are engaged with other actors in a variety of initiatives that could provide enhanced transparency. Among the many examples are:

Partnership on Artificial Intelligence to Benefit People and Society: launched in September 2016 by several major technology companies (Google, Facebook, Microsoft, Amazon and IBM), membership includes academic institutions, civil society groups and think tanks. The Partnership promotes transparency for 'black box' AI models through documenting machine learning systems.

Global Internet Forum to Counter Terrorism (GIFCT): a partnership established by Microsoft, Twitter, Facebook and YouTube to increase technical co-operation to identify terrorist and violent extremist content. It established a working group on transparency, but as of December 2020, little obvious progress had been made.

Global Partnership on Artificial Intelligence (GPAI): a multi-stakeholder initiative (in which UNESCO is an observer) which supports research into AI activities and fosters international co-operation, building on the OECD Recommendation on Artificial Intelligence. It is led by 18 governments, including 14 of the world's 20 largest economies. Among its projects is an attempt to develop audit mechanisms to evaluate AI systems for responsibility and trustworthiness, based on metrics such as safety, robustness, accountability, transparency, fairness, respect for human rights, and the promotion of equity.

In 2020, at the prompting of the so-called '5 eyes' governments of the USA, UK, Canada, Australia and New Zealand, the **Voluntary Principles to Counter Online Child Exploitation and Abuse** were adopted by several internet companies. Principle 11 requires companies to publish or share meaningful data and insights on their efforts to combat child sexual exploitation and abuse.

Digital Trust & Safety Partnership: launched in March 2021 by a range of companies including Discord, Facebook, Google, Microsoft, Pinterest, Reddit, Shopify, Twitter and Vimeo. Among its objectives are to "adopt explainable processes for product governance, including which team is responsible for creating rules, and how rules are evolved" and to "ensure that relevant trust & safety policies are published to the public, and report periodically to the public and other stakeholders regarding actions taken".[2]

The Organisation for Economic Co-operation and Development, **OECD**, is currently overseeing a process for developing a **Voluntary Transparency Reporting Framework** specifically for terrorist and violent extremist content. However, it has been hampered by difficulty in finding consensus, even on this narrow issue. Differing expectations among stakeholders has emerged as a significant challenge.

Other initiatives worth noting are the **Santa Clara Principles on Transparency and Accountability in Content Moderation**, which are currently in the process of review and update. These focus more narrowly on content moderation, but the **review process**—which involved consultation with civil society groups from across the Global South—highlighted a need for more metrics and detail on areas like algorithms, artificial intelligence and advertising.

The **Global Network Initiative**, a coalition of internet companies, human rights

---

2   There is also the Trust and Safety Professionals Association: https://www.tspa.info/

organizations and academics, carries out independent periodic assessments of its corporate members' policies, including how transparent they are, particularly on government requests for their data.

Civil society also play an important role in promoting transparency through monitoring, academic research and journalistic investigations.

Among civil society groups, **Ranking Digital Rights** analyses and compares the performance of major internet companies against a set of criteria, principally the protection of freedom of expression and privacy. It broadly concludes that while companies are more open about policies on free expression and privacy, they still do not inform users about the ways in which their information is collected and shared.

The New America Foundation's **Open Technology Institute** has created a **transparency report tracking tool** that outlines the metrics and categories of content that Facebook, Instagram, Reddit, TikTok, Twitter and YouTube use in their transparency reports related to content rules enforcement.

Access Now has developed a **transparency reporting index**, which tracks the number of technology companies that publish transparency reports by companies. It finds that the growth rate in the number of companies publishing transparency reports globally is falling.

The **Forum on Information and Democracy** has produced **recommendations** related to the transparency of social media platforms. A regional collaboration among civil society groups in Latin America, under the rubric of **Libertad de Expresióny Plataformas de Internet**, has produced a set of proposals including more transparency to increase the accountability of internet companies.

The **Centre for International Governance Innovation** has a Transparency Working Group within its **Global Platform Governance Network**, that is examining a range of transparency requirements and access. Work has also been done on the issue by the **Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression** at the Annenberg Public Policy Centre of the University of Pennsylvania. This latter argues that transparency is essential to enable evidence-based policy, and it proposes three tiers of access (the public, accredited researchers, and transparency regulators).

It is noteworthy that many of the ideas put forth by existing initiatives are, at this stage, still aspirational or limited, and are being pursued in isolation from each other. This underlines the potential value of UNESCO convening and encouraging the development of a common vision that can combine the best and most consensual elements of all of these approaches. In developing such an approach, UNESCO would logically also consider its unique interest, value-add and potential influence, including what types of partnerships or informal alliances are appropriate within the constellation of other actors in this space.

A UNESCO approach:

Illustrative high-level transparency principles that could enhance the transparency of internet platform companies

UNESCO has a mandate to promote freedom of expression, access to information, the right to privacy and the open sharing of knowledge. The mandate includes the safety of journalists as well as media development. All these areas could be strengthened if these principles below were to be further developed and made operational.

These illustrative high-level principles could apply broadly to those internet companies which provide services around finding, creating, discovering, sharing, curating, prioritising, monetising, communicating and editing content.

General:

1. Companies should explicitly recognise they have an obligation to protect human rights, and particularly freedom of expression and access to information, and the privacy of their users;

2. Companies should recognise the need for the proactive disclosure of information as well as responding to requests for information;

3. Companies should be transparent about ownership and control, including of their subsidiary company(ies);

4. Companies should indicate what kinds of commercially-sensitive data they do not wish to disclose.

Content and process transparency:

5. Companies should be transparent about any terms and standards they enforce on their own platforms, setting out the limits of what they deem to be acceptable behaviour, and how these parameters align to respect for international standards for freedom of expression;

6. Companies should be transparent about any processes they have in place to identify, remove or reduce the impact of disinformation and hate speech, including pre- and post-publication measures; and how such processes respect the free exchange of ideas and opinions;

7. Companies should disclose what percentage of content is automatically and pro-actively removed compared to the percentage that is removed as a result of complaints;

8. Companies should be transparent about any processes they have in place to identify and act against inauthentic behaviour and false identities when these are used to undermine human rights;

9. Companies should disclose whether their processes for removing content and prohibiting behaviour are periodically subject to third party assessment as to human rights compliance, carried out by a respected external independent institution or oversight body; and consider whether such assessments should themselves be transparent as well as the company's own response to any recommendations arising.

Due diligence and redress:

10. Companies should be transparent as to whether they have processes to enable people to raise concerns about content, including that which appears to violate human rights or advocates incitement to violence, hostility or discrimination, as well as inaccurate content; and they should be transparent about implementation of such processes in terms of numbers and types of complaints and actions taken;

11. Companies should be transparent about whether they conduct risk assessments for their operations, such as in contexts of upcoming elections or in countries in conflict, highlighting any serious potential threats to freedom of expression, privacy and other human rights, as well as their proposals for mitigating those threats;

12. Companies should disclose if they have risk assessments of any algorithms whose application can have the potential to discriminate against people unfairly, and if there are any proposed mitigation measures;

13. Companies should publish guidelines for how they will develop ethical AI processes which make consequential decisions that can impact on human rights.

Empowerment:

14. Companies should disclose any efforts they make which help to promote the media and information literacy competencies of those who are using their services;

15. Companies should disclose the terms and conditions for grants made in support of research, education and advocacy, as well as lobbying activities.

Transparency and commercial dimensions:

16. Companies should provide information about political advertisements, including the author and those paying for the ads, and should retain these advertisements in a publicly accessible library online;

17. Companies should reveal practices of advertising and data collection regarding children's rights;

18. Companies should enable individuals to find out on what basis they are being targeted for advertising.

Personal data gathering and use:

19. Companies should provide information that enables people to have a meaningful (i.e. concise, transparent, intelligible, reasonably comprehensive and easily accessible) understanding about what kinds of personal data are collected and how these are used;

20. Companies should provide the means for the user to check the accuracy of their personal data held by the service, and disclose how people can request amendments or deletion in line with privacy and data protection principles;

21. Companies should state how many government requests for access to personal data have been received and the legal status of those requests;

22. Companies should disclose if and how their processes for managing privacy and data protection are subject to third party assessment by a respected external independent institution (or oversight body), following an agreed standard that respects human rights;

23. Companies should disclose their use of tracking cookies, or other systems, that gather user data on their and other services across the internet, and with whom they share this data;

24. Companies should disclose data breaches and what actions are being taken to strengthen data security.

Data access:

25. Companies should, in an analogous fashion to many public statistical bodies, have a process to allow researchers access to personal data they hold, where this will advance important public interest goals such as open access and open science, while guaranteeing users' privacy through the range of necessary measures;

26. Companies should be transparent about their third-party agreements which allow access to personal data that is purchased, shared, directly harvested or held by them.

Conclusion

While transparency proposals are usually presented as part of a general approach to making internet companies more accountable, embracing transparency is also in the companies' own self-interest. Increasingly distrusted and blamed for societal problems, the positive contribution that companies make to enabling freedom of expression, knowledge diffusion and community building is increasingly overshadowed. Enhanced and systematic transparency can help correct the problems and bring more of the benefits to light.

In addition, transparency reporting allows companies to demonstrate their fulfilment of corporate social responsibility, over and above what is required by law. In this sense, the application of a set of global principles for transparency by a company could be a competitive advantage, as well as provide for more consistent standards internationally.

It may be noted that the term 'transparency' is used frequently, but rarely defined, which can present challenges for developing practical applications. Complicated questions emerge: how is an 'item' of data defined? How is a URL containing thousands of illegal images counted? Will the adoption of rules for disclosing content moderation make companies adopt simpler rules that do not take account of nuance? What does transparency mean for AI algorithms that continuously evolve based on machine learning? How to balance the benefits of transparency against the risk of actors who violate human rights exploiting insight to intensify their activity?

Against this background, working for transparency by developing high-level principles can set out goals in ways that can inform subsequent attempts to develop more granular indicators. The illustrative principles in this brief provide a basis for further multistakeholder work on all levels.

The approach set out here builds upon existing initiatives to identify areas where there is a consensus rather than starting from zero. This includes recognising the principles of the historic Windhoek Declaration, endorsed by UNESCO Member States in 1991, which set out the role of states in ensuring freedom of expression, media pluralism and media independence. In this Declaration, the role of the state is both to refrain from adversely limiting expression, as well as to take active measures to ensure pluralism by ending monopolies. When it comes to regulation of the internet, states similarly have a duty to refrain from certain regulation, and yet to intervene where appropriate in order to enable the internet to be human rights-respecting, open, accessible and governed in a multi-stakeholder way.

Transparency steps may be an appropriate response. Nevertheless, the possibility also exists for voluntary principles. A consensus-based approach, such as that favoured by the OECD, can also attain results if the barriers to reaching agreement can be overcome and companies step up to the challenge. It is when self-regulation lacks real commitment and/or functional institutional mechanisms for accountability, that states are impelled to act. To achieve transparency, a combination of voluntary and statutory regulation may be an optimum mix, with the latter setting a broad minimum standard (nuanced as per its applicability to company size).

The recommended approach therefore is to focus on the high-level transparency principles that address UNESCO's key policy concerns, including freedom of expression, safety of journalists, viability of news media, privacy and combating hate speech and disinformation. In the first instance, the goal is to secure consensus among stakeholders, including the major internet companies, as regards these broad objectives.

Recommendations

UNESCO should promote discussion on these illustrative high-level principles as a first step in developing a more detailed approach to transparency. Subject to future consideration, this could involve producing a more detailed commentary referencing existing standards and taxonomies and then developing indicators that can assess the effectiveness of the companies' response to the transparency requirements and means of verification. UNESCO should also bring these insights to the attention of policy makers and regulators worldwide, and advocate for a broad consistent approach to potential regulatory transparency requirements and for alignment to international human rights standards.

Companies should discuss these principles and commit to working with UNESCO to develop them further into a framework for transparency across the platform industry.

Regulators should reflect on these principles as the basis for incorporating transparency provisions in future regulatory initiatives, and as a benefit for developing evidence-based policies related to the internet and AI. Appropriate regulators could also engage in collaborative dialogue with other regulators that deal with internet companies (such as competition, data protection and privacy), to provide agreed and consistent expectations of transparency.

Governments, in considering whether or how to regulate internet companies, should examine how transparency principles like those proposed in this paper can contribute to achieving public policy objectives in place of more intrusive legislation that can pose risks to freedom of expression.

Civil society should encourage elaborated transparency objectives in their advocacy for greater accountability among internet companies, and raise these issues with national regulators and policy makers.

Academia and the technical community should give more attention to the issue of transparency in their research priorities as well as in further development of the internet architecture.

About this brief

This brief comes as part of the UNESCO series World Trends in Freedom of Expression and Media Development. It discusses how greater transparency in the operations of internet companies could strengthen freedom of expression and other issues central to UNESCO's work, and it outlines existing mechanisms and initiatives. The brief sets out a preliminary selection of illustrative high-level principles, which could serve as a basis for future discussions towards a framework for transparency to guide companies, policy makers and regulators.

The font used in the title is "Facts", a font specially developed for the 2020 celebration of World Press Freedom Day. Download it here

For more resources, visit: World Trends in Freedom of Expression and Media Development.

About the author

Andrew Puddephatt is an internet policy expert who focuses on the intersection of internet policy and human rights. A long term consultant to UNESCO, He is the founder Director of Global Partners & Associates Ltd, a social purpose company that promotes democracy and human rights online. He is Founder Director of Adapt, a social purpose company that helps organizations manage their users' data ethically. He also chairs the Internet Watch Foundation, a charity registered in the UK that works with the internet industry to remove child sex abuse images online, and International Media Support, a Danish based NGO that promotes independent media in conflict zones. He is the Deputy Chair of the philanthropic foundation the Sigrid Rausing Trust. He was awarded an OBE for services to human rights.