

# Response to the ‘Call for Submissions’ on The Santa Clara Principles on Transparency and Accountability

**Authored by:** Torsha Sarkar and Suhan S  
**With editorial inputs:** Gurshabad Grover

30 June, 2020

**1. Currently the Santa Clara Principles focus on the need for numbers, notice, and appeals around content moderation. This set of questions will address whether these categories should be expanded, fleshed out further, or revisited.**

- a. *The first category sets the standard that companies should publish the numbers of posts removed and accounts permanently or temporarily suspended due to violations of their content guidelines. Please indicate any specific recommendations or components of this category that should be revisited or expanded.*

While the Principles provide a robust framework for content moderation practices carried out by the companies itself, we believe that the framework could be expanded significantly to include more detailed metrics on government requests for content takedown, as well as for third-party requests. For government requests, this information should include the number of takedown requests received, the number of requests granted (and the nature of compliance - including full, partial or none), the number of items identified in these requests for takedown, and the branch of the government that the request originated from (either from an executive agency or court-sanctioned).

Information regarding account restrictions, with similar levels of granularity, must also form a part of this vertical. These numbers must be backed with further details on the reasons ascertained by the government for demanding takedowns, i.e. the broad category under which content was flagged. For third party requests, similar metrics should be applied wherever appropriate.

Additionally, for companies owning multiple platforms, information regarding both internal content moderation and moderation at the behest of external requests (either by the state or third-parties), must be broken down platform-wise. Alternatively, they should publish separate transparency reports for each platform they own.

- b. *The second category sets the standard that companies should provide notice to each user whose content is taken down or account is suspended about the reason for the removal or suspension. Please indicate any specific recommendations or components of this category that should be revisited or expanded.*

While this category envisages companies to provide notice to its users across removals related to all categories of content, additional research reveals that oftentimes, [companies](#) create further categorization of 'exceptional circumstances', where it may hold the discretion for not sending a notice, including for CSAM or threats to life. While the intent behind such categorization might be understandable, we believe that any list of exceptional circumstances should not be ideally left to company discretions, and must be prepared in a collaborative fashion. Accordingly, we recommend that the Principles be expanded to identify a limited set of exceptional circumstances, where not sending a notice

to a user would be permissible, and would not count as a violation of the Principles.

Additionally, while the current framework provides requirements for granular details in the notice in case of content flagged by the company's internal moderation standards, we believe a similar model should also be emulated for content removals at the behest of the state. When a piece of content has been identified as illegal by a government takedown request, then the notice issued by the company to the user should be as granular as possible, within the permissible limits of the law under which the takedown request was issued in the first place. Such granularity must include, among other things, the exact legal provision under which the content has been flagged, and the reasons that the government has given in implementing this flagging.

- c. *The third category sets the standard that companies should provide a meaningful opportunity for timely appeal of any content removal or account suspension. Please indicate any specific recommendations or components of this category that should be revisited or expanded.*

Currently, the category of 'appeals' in the Santa Clara Principles is focussed on having accountability processes in places, and emphasize on the need of having meaningful review. The framework of the Principles also currently envisage only internal review processes carried out by the company. However, in light of Facebook unveiling its plans for an [Oversight Board](#), a structurally independent body, which would arbitrate select appeal cases of content moderation, these pre-existing principles might need revisiting.

While the Oversight Board is a relatively novel concept, given the important precedence it sets, setting certain fundamental principles of transparent disclosures and accountable conduct around it, might allow researchers and regulators alike to gauge the efficacy of this initiative. Accordingly, the Principles should consider some base-level disclosures that the company must make when it is referring a select category of cases for independent external review. This might include a statement of reasons explaining why certain cases were prioritized for independent review, and in the instance that the decision hinges on a public interest question, then the proceedings of the independent review might also be required to be made public (with due recourse paid to security issues and the confidentiality of the parties involved).

**2. Do you think the Santa Clara Principles should be expanded or amended to include specific recommendations for transparency around the use of automated tools and decision-making (including, for example, the context in which such tools are used, and the extent to which decisions are made with or without a human in the loop), in any of the following areas:**

**Content moderation (the use of artificial intelligence to review content and accounts and determine whether to remove the content or accounts; processes used to conduct reviews when content is flagged by users or others)**

Companies have begun to rely on a variety of automated tools to aid their content removal processes, across a variety of content, including [revenge porn](#), [terrorist content](#) and [CSAM](#). Research however, has shown that the tools deployed often have their limitations, which include [over-removal](#), and [censorship](#) of perfectly [legitimate speech](#).

We recommend that the Principles should accordingly be expanded to [include](#) content removed by automatic flagging, the error rates encountered by the tools, and the rate at which wrongly taken down content is being reinstated. There should also be a qualitative aspect to the information presented by these companies, and therefore, there should be a clearer disclosure of the kind of automated tools they use. Such disclosure must, of course, be [balanced](#) against interests of the security of the platform and the necessity to ensure that information disclosed is not used by malicious third-party actors to circumvent legitimate moderation.

Additionally, with specific reference to ‘extremist content’, several online companies have collaborated to form the Global Internet Forum to Counter Terrorism (GIFCT), with the intent of facilitating better moderation. The GIFCT uses a hash-based technology of a shared database of ‘terrorist’ content for filtering content on their platforms. However, as it has already been [noted](#), this initiative provides very little information regarding how it functions, and operates without any collaboration with civil society or human rights groups, and without any law enforcement oversight.

Such similar collaborative measures going forward, for deployment of varied forms of automated tools to filter out various forms of content, without any transparency or accountability, can be problematic, since it makes information regarding the efficacy of these tools scarce, research into the processes difficult, and ultimately, any reformative suggestions impossible.

Accordingly, the Principles must emphasize that collaborative efforts to the effect of using automated tools in content moderation must be done with sufficient consideration to the basic principles of transparency and accountability. This might include sharing information about processes with a select list of civil society and human rights groups, and in the transparency reports, separately presenting information about the accuracy rates of the tools.

**Content ranking and downranking (the use of artificial intelligence to promote certain content over others such as in search result rankings, and to downrank certain content such as misinformation or clickbait)**

Ranking and downranking algorithms have been deployed by companies for various purposes and across different services they offer. For the purposes of our discussion, we

would restrict ourselves to two chief use-cases of these processes: search engines and internet platforms.

### Search engines

The algorithms that have been developed to find accurate results for query are oftentimes not perfect, and they have been accused of being [biased](#), including being politically non-partisan and [burying](#) certain ideologies. Similarly, in the case of automated systems to downrank misinformation, accuracy is not guaranteed as such systems can identify accurate information as misinformation. Since the algorithm is constantly learning and [updating](#), it becomes difficult to know exactly why certain content may be made less visible.

As case-studies of several search engines [indicate](#), a company's ranking processes often use a combination of algorithms and human moderators. Requirement for transparency therefore, can mandate disclosure of the training materials for these human moderators. For instance, Google has a scheme of '[Search Quality Raters](#)', which comprises a group of third-party individuals responsible for giving feedback regarding search results. The guidelines on which their feedback is based on, are [publicly available](#). The Principles can therefore call for similar disclosure of other companies that deploy human help for their ranking processes.

### Internet platforms

For social media platforms, ranking algorithms are utilized for [curation](#) of news-feeds: dashboards showing content to the user that the algorithm thinks are relevant. The algorithm makes these decisions based on different signals that it is trained with. Information around these algorithms is hard to come by, and even if it is, the algorithms are often blackboxes, with their decisions not [explainable](#).

There are however, ways by which transparency around these algorithms can be improved without compromising the security and integrity of the platform. This might include companies [informing](#) users, in an accessible manner, "*(i) how they rank, organize and present user generated content.*", and [updating](#) the data in a timely manner, allowing researchers and regulators the appropriate opportunity to utilize this information while it is still relevant.

Companies should also have an easy-to-access policy that [outlines](#) how it plans to manage the human rights risks arising out of the system(s) it deploys. The human rights impacts assessment must additionally consider the broad social contexts within which the algorithm system is used.

**Ad targeting and delivery (the use of artificial intelligence to segment and target specific groups of users and deliver ads to them)**

Companies such as Facebook and Google collect a wide variety of data from its audience, using a variety of data points (including age, location, race) which is used to deliver personalised advertisements by the advertisers affiliated with the company. Methods like [activity tracking and browser-fingerprinting](#) are employed to track users, with or without explicit notice. Since a user's [privacy](#) is greatly affected by such tracking, more transparency is needed where user data is collected by companies and where they are processed using the company's algorithms to target and deliver ads. Additionally, targeted advertising, especially in the context of political advertising, result in segmenting groups of people and subjecting them to advertising campaigns. This, in turn may have drastic [consequences](#), since they seem to deepen divisiveness over critical issues.

### Notice

The Principles should identify metrics of a meaningful notice that companies must give users when their data is collected for delivering advertisements. Among others, such notice should specify all kinds of data the company is collecting regarding the user, and the categories across which they have been segmented or categorized for advertising.

### Disclosure

Companies should also strive to disclose how data is collected and processed, specifically to segment users and deliver advertisements, in detail. This might include disclosing all the [categories](#) made available to advertisers by the company, and the names and identities of third parties (both advertisers and data-brokers) with whom such data is shared. CNBC, for instance, in 2019 [reported](#) that Facebook selectively shared user data with select partners while denying rival companies from accessing the data. Additionally, companies that allow users to opt out of their data being wholly or partly should disclose this option and make it easy to access. For Example, Facebook lets users turn off data being used for advertising in [three different categories](#). Facebook Ad Preferences menu hidden in a user's settings is detailed. However, barring a [public post](#) that attempts to explain how and why users see certain ads on Facebook, which has one line at the end that directs users to their Ad Preference settings to "View and use" their controls, the company does not have any public document explaining users their choices. Amazon, on the other hand allows users to turn off personalized ads completely and has a dedicated [page](#) that explains how a user's data is used for personalizing advertisements and options to disable it.

### **Content recommendations and auto-complete (the use of artificial intelligence to recommend content such as videos, posts, and keywords to users based on their user profiles and past behavior)**

Algorithms and recommendation systems are designed to suggest content that a user is likely to interact with, on the basis of their browsing behaviour and interaction on the platform. These algorithms are constantly updated to be more accurate. Popular examples include [Instagram](#) and [YouTube](#). It is interesting to note that these systems

have been documented to often suggest [radical content](#) to users, and upon user-interaction with such content, continuously amplify them. YouTube's algorithm, for instance, has been previously accused of pushing users towards [extremist or inflammatory](#) ideologies.

Studying how recommendation algorithms function however, and [why](#) certain extremist content are being recommended to users, have been difficult, due to *one*, the complexity of the current [information ecosystem](#), and *two*, because of the lack of information around these algorithms. The Santa Clara Principles can, by way of an expansion of scope, look to address the second difficulty, by urging companies to be more transparent with their internal processes.

### Sharing of data or open-sourcing algorithms

With due recourse paid to the security and integrity of the platform, we recommend that the code for the algorithm used for recommendations should be open-source and publicly available online. [Reddit](#), for instance, publishes its code for curation of news feeds in an open-source format.

Another way of doing this, as has been [studied](#), is to consider a two-pronged method of sharing data. In the first count, datasets identified as 'sensitive', are shared in partnerships with certain institutions, under non-disclosure agreements. In the second count, more non-sensitive data is shared in an anonymized format publicly, and made available for any researcher to access.

This idea, however, must be taken with a few caveats. *One*, sharing of datasets may not always fulfill the public-facing model of transparency and accountability that the Santa Clara Principles envisage. *Two*, this might be a particularly onerous obligation for smaller and medium enterprises, and without sufficient economic data, it might be difficult to implement this. And *three*, any framework adopting this must consider the privacy aspect of such sharing. At this juncture, therefore, we do not recommend this as a compulsory binding obligation that any company adopting the Principles must abide by. Rather, we hope and encourage for more conversations to be held around this concept, so that the aforementioned competing interests are accommodated optimally.

### Qualitative transparency

The other mode of ensuring more clarity into the recommendation system should be by asking companies to publish user-facing, clearly accessible policies and explainers that outline how the company uses algorithms to recommend content to users. This can also include [creation](#) of a visible list of topics, which the company has chosen 'not to amplify' (for instance, topics such as self-harm, eating disorders), and updated regularly.

**3. Do you feel that the current Santa Clara Principles provide the correct framework for or could be applied to intermediate restrictions (such as age-gating, adding warnings to content, and adding qualifying information to content). If not, should we seek to include**



## **these categories in a revision of the principles or would a separate set of principles to cover these issues be better?**

The Santa Clara Principles, as they had been originally envisaged, adhered to the commonly adopted binary of take down/leave up in content moderation, where a piece of unlawful, or problematic content (or an account), was either censored from public view or allowed to continue. However, since then, platforms dealing with user-generated content have resorted to a variety of novel and intermediate techniques to moderate and regulate speech which fall outside the aforementioned binary. With adoption of such steps therefore, it is also important for the Principles to evolve and take into consideration the expanded scope of content moderation. In light of that, we recommend the following steps to be taken in the intermediate areas of regulation:

### Adding warnings, qualifying information to content

As mentioned above, in recent past, online intermediaries have resorted to more intermediate restrictions to deal with 'harmful' content online. These measures have seen an added boost in light of the Covid-19 outbreak, where there has been a massive [increase](#) in misleading information and conspiracy theories online. These measures have included, among others, [connecting](#) users who have interacted with misinformation to verified, debunked information and introducing a [spectrum](#) of actions based on the degree of harm posed by the content, which includes adding labels, warning, and finally, removal. Such intermediate measures currently are not accommodated within the framework of the Santa Clara Principles, for reasons enumerated above, and going forward, it may become important for the Principles to look at the learnings from these measures and adopt them, wherever appropriate, into the framework.

Additionally, as conversations around the instance of Twitter adding a fact-check to Donald Trump's tweet show, the application of these intermediate measures are often [ad-hoc](#), since there is often no explanation why certain items receive the moderation treatment, while other, similarly misleading content from same [sources](#), continue to stay online. Accordingly, it is difficult to ascertain the exact reasoning process behind these steps. Therefore, adoption of principles related to measures of adding labels or warnings to information online must also require companies to be transparent with their decision-making processes.

### Fact-checking

In recent years, with the proliferation of misinformation on online platforms, several companies have either begun to [collaborate](#) with fact-checkers, or deploy their own in-house [teams](#). While these initiatives should be appreciated, it should also be noted that the term 'fact checking' assumes a partisan meaning in certain circumstances, including when sources of [misinformation](#) themselves offer this service. Accordingly, it becomes important that the fact-checking initiatives adopted by companies adhere to some standards of international best practices, and the decisions made are not riddled with biases, either political or ideological.



The Santa Clara Principles are useful to ascertain the transparency of any fact-checking initiatives, and can be applied across both collaborations between companies and fact-checkers, as well as for in-house fact checking initiatives.

For any manner of collaborations, companies must disclose, in clear terms, the names and identities of the fact-checking organizations that they are teaming up with (this [example](#) from Facebook divides this list of names country-wise) and the nature of this collaboration, which must include details of whether the organization stands to any monetary gains, and what is the level of access to the platform and its dashboards given by the company to the fact-checking organization.

For in-house initiatives, the Santa Clara Principles must require companies to disclose information regarding any training programs carried out and the background of the fact-checkers, and this might also include a statement regarding the objectivity and non-partisanship of the initiative.

Lastly, comprehensive information about fact-checking must be presented in a clearly accessible format in the company's regular transparency reports, which should include data on how many pieces of content got fact-checked in the reporting period, the nature of the content (text, photos, videos, multimedia), the nature of misinformation that was being perpetuated (health, communal etc.), and the number of times the said piece of content was shared before it could be fact-checked.

### Age-gating

The Digital Economy Act of 2017, proposed by the UK Government (and since dropped in 2019) serves as an early model of the legislature around the world to regulate the process of putting in place age-restrictions. By the application of that [law](#), any websites offering pornography would have to show a landing page to any user with an UK IP address, which would not go away till the user is able to show that they are over the age of eighteen years. However, the government had left the exact technical method of implementing the age-gate upto the website, which meant that websites were free to adopt any methods they deem fit for verifying age, which might also include [facial recognition](#).

However, learnings from the UK Model, and several other models of attempted age-gating have shown that there are often easy [methods](#) of [circumvention](#) and the information collected in lieu of implementation of these methods goes on to raise [privacy concerns](#). It is our understanding that the regulation of age-restrictions is currently in a flux, and setting principled guidelines at this stage may not be completely evidence-based. In such light, it is our recommendation that the Santa Clara Principles should not be expanded to include age-gates. Separate consultations and discussions on the merits of the various forms of age-gating should precede any principles in this subject.

**4. How have you used the Santa Clara Principles as an advocacy tool or resource in the past? In what ways? If you are comfortable with sharing, please include links to any resources or examples you may have.**

In 2019, we developed specific methodologies to analyse information relating to government requests for content takedown and user information, from transparency reports made available by online companies for India. For creating our methodology for government requests for content takedown, we relied significantly on some of the metrics of the Santa Clara Principles, and utilized them to expand our scope of analysis. Our methodology comprised of the following metrics adopted from the Principles:

- **Numbers:** We utilized this metric, and further clarified that the numbers should include a numerical breakdown of the requests received under different laws on content takedown.
- **Sources:** The Santa Clara Principles recommend that the intermediary identify the source of the flagging. Under the intermediary liability regime in India, content takedown requests can be sent by the executive, the courts, or third parties. We accordingly argued that transparency reports must classify the received requests into these three categories.
- **Notice:** We also utilized this metric for our methodology.

The full version of our methodology and the results from our analysis can be found [here](#).

**5. How can the Santa Clara Principles be more useful in your advocacy around these issues going forward?**

We intend to apply this methodology for future editions of the report as well, and build up a considerable body of work on transparency reporting practices in the Indian context.

**6. Do you think that the Santa Clara Principles should apply to the moderation of advertisements, in addition to the moderation of unpaid user-generated content? If so, do you think that all or only some of them should apply?**

Moderation of advertisements in the recent years have become an interesting point of contention, be it advertisements that [violate](#) the companies policies on disruptive ads policies, or advertisements with more nefarious undertones, including [racist language](#) and associations to [Nazi symbols](#).

Several companies already have various [moderation policies](#) for these kinds of harmful advertisements and other content that advertisers can promote, and these are often public. Based on this, we think that the Santa Clara Principles can be expanded to include the moderation of advertisements, and the metrics contained within would be applicable across this vertical, wherever appropriate.

**7. Is there any part of the Santa Clara Principles which you find unclear or hard to understand?**

N/A.

**8. Are there any specific risks to human rights which the Santa Clara Principles could better help mitigate by encouraging companies to provide specific additional types of data? (For example, is there a particular type of malicious flagging campaign which would not be visible in the data currently called for by the SCPs, but would be visible were the data to include an additional column.)**

N/A.

**9. Are there any regional, national, or cultural considerations that are not currently reflected in the Santa Clara Principles, but should be?**

While utilizing the Principles for the purposes of our research, we found that the nature of information that some of these online companies make available for users residing in the USA, is very different from the information they make available for users residing in other countries, including in India. For instance, Amazon's transparency reports regarding government requests for content removal, till the first half of [2018](#), was restricted only to the US, despite the company having a considerably large presence in India (during our [research](#), Alexa Rank showed Amazon.com to be the 14th most visited website in India).

A public commitment to uphold Santa Clara Principles (as several companies have undertaken, see EFF's recent Who Has Your Back? [report](#)) would mean nothing if these commitments do not extend to all the markets in which the company is operating. Accordingly, we believe that it must be emphasized that the adoption of these Principles into the transparency reporting practices of the company must be consistent across markets, and the information made available should be as uniform as it is legally permissible.

**10. Are there considerations for small and medium enterprises that are not currently reflected in the Santa Clara Principles, but should be?**

Our understanding at this current juncture is that not enough data exists around the economic costs of setting up the transparency and accountability structures. Accordingly, at the end of this Consultation period, should the Principles be expanded to include more intermediate restrictions and develop accountability structures around algorithmic use, we recommend that a separate consultation be held with small and medium enterprises to identify a) whether or not there would be any economic costs of adoption and how best the Principles can accommodate them, and b) what are the basic minimum guidelines that these enterprises would be able to adopt as a starting point.

**11. What recommendations do you have to ensure that the Santa Clara Principles remain viable, feasible, and relevant in the long term?**

Given the dynamic nature of developments in the realm of content moderation, periodical consultations, in the vein of the current one, would ensure that the stakeholders are able to raise novel issues at the end of each period, allow the Principles to take stock of the same, and incorporate changes to that effect. We believe that this would allow for the Principles to continue to be aware of the realities of content moderation, and allow for evidence-based policy-making.

**12. Who would you recommend to take part in further consultation about the Santa Clara Principles? If possible, please share their names and email addresses.**

N/A.

**13. If the Santa Clara Principles were to call for a disclosure about the training or cultural background of the content moderators employed by a platform, what would you want the platforms to say in that disclosure? (For example: Disclosing what percentage of the moderators had passed a language test for the language(s) they were moderating or disclosing that all moderators had gone through a specific type of training.)**

By now, there have been well documented [accounts of](#) human moderators, by independent investigations or admissions by companies. For instance, this [blogpost](#) authored in 2018 by Mark Zuckerberg documented the percentage of human moderators who were trained in the Burmese language, in reference to moderating content on the platform in Myanmar. Comprehensive information about linguistic and cultural backgrounds of human moderators is a useful tool to contextualize the decisions made by the platform, and also useful in pushing more effective reforms.

Additionally, it has also been seen that a company's public facing moderation norms often differ from its internal guidelines, which are shared with its team of human moderators. For instance, TikTok's internal norms had [asked](#) its moderators to 'suppress' content from users perceived to be 'poor' and 'ugly'. The gaps in these norms means that there are surreptitious forms of censorship behind-the-scenes, and it is difficult to ascertain the reasonableness and appropriateness of these decisions.

We would also like to emphasize more stringent disclosure requirements from companies regarding the nature of engagement with which they employ their human moderators. As [investigations](#) have [revealed](#), the task of human moderation is often outsourced by these companies to third-party firms, and the working conditions in which the moderators make their decisions are inhospitable. Additionally, more often than not, there are no publicly available methods to ascertain whether the company in question is doing enough to ensure the well-being and safety of these moderators.

Therefore, alongside disclosure regarding the nature of training given to the human moderators and their internal moderation norms, we also recommend that the Principles recognize certain fundamental ethical guidelines with relation to their human moderators that companies must adopt. This might include providing identifying information of the

third-party firms to which the company outsources its moderation and assurances of sufficient number of counsellors for the moderators.

#### **14. Do you have any additional suggestions?**

While the Santa Clara Principles provide a granular and robust framework of reporting, currently it stands to only cover aspects of quantitative transparency - concerning numbers and items. As we have indicated throughout this submission, and in our previous research, there are also need for companies to adhere to more norms focussing on qualitative transparency - in the form of material disclosure of the policies, processes and structures they associate with, or make use of. Aside from the suggestions in the previous sections, in this section we highlight two additional recommendations that we think can help achieve this.

##### Material regarding local laws

One of our preliminary findings regarding the way these intermediaries report data for other regions (including India) has been that most of the time, the information is incomplete, especially with regards to material regarding the local laws. Compared to the US, for which most of these companies dedicate separate sections, other regions feature relatively fewer times in their reports. Each country in which the company functions, there would be various laws governing content removal, different authorities empowered to issue orders, and varied procedural and substantive requirements of a valid request. For the empowerment of users, we believe that the exact metrics and requirements of these laws must be presented by the intermediaries, in a clear and readable format.

##### Accessibility of policies

On the topic of empowerment of users, we also believe that the basic information and policies regarding these requests should be placed at one place, for maximum accessibility by users. During our research, we discovered that the disclosures made in lieu of the Principles were spread over different policies, some of which were not easily accessible. While it is not possible at this juncture to predict a comprehensively objective way of making all this information accessible, we believe it would be a useful step if the basic information regarding the intermediary's transparency reporting policies were presented in the same manner as the company's Terms and Services and Privacy Policy. Additionally, we believe that these disclosures should be translated into major languages in which the company operates, for further accessibility.

#### **15. Have current events like COVID-19 increased your awareness of specific transparency and accountability needs, or of shortcomings of the Santa Clara Principles?**

The Covid-19 pandemic proves to be a watershed moment for the history of the internet, inasmuch in the manner of proliferation of various forms of misinformation and conspiracy theories, as well as the way in which companies have stepped up to remove said content from their platforms. This has included companies like [Google](#), [Twitter](#) and

[Facebook](#), who have sought to increasingly rely on automated tools for rapid moderation of harmful content related to the pandemic.

These practices reaffirm the need for having strong requirements for transparency disclosures, both qualitative and quantitative, especially around the use of automated tools for content takedown. This is because of two main reasons.

One, the speed of removal would never tell us [anything](#) about the accuracy of the measure. A platform can say that in one reporting period, it took down 1000 pieces of content; this would not mean that its actions were always accurate, or fair or reasonable, since there is no publicly available information to ascertain so. This phenomenon, aggregated with the heightened [pressure](#) to remove misinformation related to the pandemic, may contribute to *firstly*, erroneous removals (as YouTube has [warned](#) in blogs), and *secondly*, towards deepening the information asymmetry regarding accurate data around removals.

Two, given the novel and diverse forms of misleading information related to the pandemic, this offers a critical time to [study](#) the relation between online information and the outcomes of a public health crisis. However, these efforts would be thwarted if reliable information around removals relating to the pandemic continue to be unavailable.