

## META COMMUNITY FORUM: RESULTS ANALYSIS

June 2023

Report Prepared by



**Deliberative Democracy Lab at Stanford University, housed within the Center on Democracy, Development, and the Rule of Law at the Freeman Spogli Institute for International Studies**

**Report Contributors (in alphabetical order): Samuel Chang PhD, Estelle Ciesla, and Tenzin Kartsang**  
**With additional support from numerous undergraduate and graduate students at the Deliberative Democracy Lab.**

## Table of Contents

Executive Summary	4
Project Impact and Significance	7
Methodology	8
Sampling Methodology	8
Actual Recruitment Numbers	18
Weighting Methodology	23
Quantitative Results	29
General Analysis	31
Who should be responsible for monitoring social VR spaces?	33
To whom should cases of bullying and harassment be reported?	35
Should creators be made aware of reports made to platform owners?	38
What should be done about spaces where there is repeated bullying and/or harassment?	40
What tools should be used to identify bullying and/ or harassment in social VR spaces?	47
Who should review video capture and automatic speech detection alerts?	53
Should moderators be used to monitor bullying and harassment in social VR spaces?	61
Regional Analysis	68
Who should be responsible for monitoring social VR spaces?	68
To whom should cases of bullying and harassment be reported?	70
Should creators be made aware of reports made to platform owners?	74
What should be done about spaces where there is repeated bullying and/or harassment?	75
Visibility	75
Consequences for creators	78
User Notifications	83
What tools should be used to identify bullying and harassment in social VR spaces?	84
Video Capture	84
Automatic Speech Detection	90
Who should review alerts?	95
Video Capture	95
Automatic Speech Detection	98
Moderators: Visible and Invisible	103
Demographic Analysis	108
Gender	108
Age	110
Experience with Social Spaces in VR	113
Control Group	115
Qualitative Analysis	118
Session 1: What tools should be used to identify bullying and harassment in the metaverse, and	

where should they exist?	119
Session 2: Who is in the best position to look out for bullying and harassment in different kinds of spaces?	121
Session 3: What should be done in response to bullying and harassment?	124
Session 4: What should be the consequences for creators of spaces where bullying and harassment happens repeatedly?	126
Deliberative Quality Insights	129
Implications and Next Steps	133
Appendix	136
A. Assessment of Data Source: DataReportal Statistics	136
B. Demographic Results	136
Gender	136
Age	146
Metaverse Experience	156

## Executive Summary

The Community Forum was a first of its kind experiment in global deliberation. Scientific samples of the world's social media users were recruited for a weekend-long deliberation from 32 countries in nine regions around the world speaking 19 languages. A matching control group of comparable size did not deliberate but took the same questionnaires in the same time period in December 2022. The issue is a novel and important one: how to regulate bullying and harassment in virtual reality, particularly in the new private or “members only” social VR spaces that are being created in the Metaverse. More than 6,300 deliberators, representative of global social media users<sup>1</sup> were selected by 14 survey research partners who recruited respondents who deliberated in 19 languages. For more on the design and for the weighting of the sample that supports inferences to the global population of social media users, please see the methodology section.

### Deliberative Polling®<sup>2</sup>

The design for the deliberations followed the Deliberative Polling® model under the direction of the [Stanford Deliberative Democracy Lab](#) and in collaboration with the [Behavioral Insights Team \(BIT\)](#) and Meta. A distinguished Advisory Committee vetted the briefing materials for the deliberations and provided many of the experts for the plenary sessions. The process alternated small group discussions and plenary sessions where experts would answer questions agreed on in the small groups. The agenda was a series of 56 policy proposals that could be implemented by Meta or other platform owners. The proposals came not only with background materials but also with pros and cons posing trade-offs that the participants might want to consider. Video versions of the briefing materials were also provided.<sup>3</sup>

The small group discussions were conducted on the [Stanford Online Deliberation Platform](#) which moderated the video based discussions, controlled the queue for talking, nudged those who had not volunteered to talk, intervened if there was incivility, and moved the group through the agenda of policy proposals and their pros and cons. Near the end of each discussion, it also guided the groups in formulating key questions that they wished to pose to the panels of competing experts in the plenary sessions. The Stanford Online Deliberation Platform is a collaboration between the [Crowdsourced Democracy Team](#), led by Ashish Goel, and the Deliberative Democracy Lab, both at Stanford University.

The core issue posed for deliberation was the responsibility of platform owners such as Meta versus the role of individual Creators for regulating behavior in private or “members only” VR social spaces. To what extent should the platform owners stay out, since these “members-only” spaces are not public, and only joined by mutual consent and the participants who may want to establish their own norms of behavior? Or to what extent do the platform owners, such as Meta, have a responsibility to act to protect against bullying and harassment? If the platforms have a responsibility, what should they do? These are novel issues, and they amount to the beginnings of a social contract for how these new spaces in virtual reality

---

<sup>1</sup> Please refer to the methodology section for countries the project was unable to sample from.

<sup>2</sup> Deliberative Polling® is a trademark of James S. Fishkin. Any fees from the trademark are used to support research at the Stanford Deliberative Democracy Lab.

<sup>3</sup> Here are links to the English versions of the briefing videos. [Session 1](#) [Session 2](#) [Session 3](#) [Session 4](#)

should be governed. Importantly, the proposals and discussion focused on the topic of bullying and harassment only; other abusive or illegal online behaviors were not at issue.

### Key Takeaways

Here are some key takeaways among the 56 policy proposals that participants discussed in this Community Forum.

First, a full range of questions were posed for public spaces. There was less concern for privacy in public spaces so that the role of the platform in regulating and protecting against bullying and harassment was clear from the outset but also rose significantly with deliberation. Before and after questionnaire results for all the proposals, along with qualitative excerpts and analyses from the small group discussions can be found in the report which follows.

The findings showed clear guidance from participants with regard to what tools or technologies parties should apply to address bullying and harassment. Participants mainly discussed video capture<sup>4</sup> and automatic speech detection<sup>5</sup> as tools that could be used in virtual social spaces. Participants deliberated on whether “platform owners should have access to video capture in members-only spaces”. This proposal rose significantly from 59% to 71%, an increase of 12 points (means on the 0 to 10 scale rose from 6.814 to 7.253). Further there was a significant increase in support for “spaces where there is repeated bullying and/or harassment, platforms should take action against creators”. This rose about 10 points, from 57.3 to 66.9% (means increased from 6.39 to 6.901 on the 0 to 10 scale).

Now onto what actions should be taken to address bullying and harassment. Here are some proposals that experienced the most movement before and after deliberation. First, participants felt that spaces where there is repeated bullying and harassment should be made less visible to users, according to 63% of the deliberators (up from 53%, an increase of about 9.5 points). Second, such spaces should no longer be publicly discoverable, increasing from 55 to 61%) providing a real disincentive to creators who want to grow their membership but who have permitted repeated bullying and harassment. Third, participants felt that creators should be required to take a course on how to moderate the spaces they create. Support for this proposal rose from 67% to 78%, a rise of 11 points. And for the members only spaces where there has been repeated bullying and harassment, participants felt that users should receive notification of such cases when entering a space, this proposal rose more than nine points from 76% to 85%.

---

<sup>4</sup> A system built and maintained by a platform developer that continuously records what is happening in case the video needs to be referenced later. Video is overwritten or deleted after a few minutes unless someone reports misconduct or a violation. This is similar to how security recording works in the physical world.

<sup>5</sup> Technology that “listens” for a predefined set of upsetting words (for example, insults or offensive language) and flags instances of these words for moderators to review.

But the recommendations from the global sample are not punitive. The sample declined to endorse more severe punishments. For example, support for removing members-only spaces where there is repeated bullying only reached 43% (up from 39%). Support for banning creators from making additional members only spaces if there was repeated bullying and harassment only reached 45% (up from 38%). Support for banning creators of such spaces from inviting additional people to join only reached 49% (up from 43%). And, lastly, support for preventing creators of such spaces from making money off their spaces only reached 54% (up from 49%).

### Knowledge and Evaluations

A vast majority of participants in the Community Forum have not had any experience in virtual social spaces. However, despite not having any experience, participants had thoughtful and considered deliberations about bullying and harassment in virtual social spaces and what policy proposals they felt could help address their concerns. Even though some participants had never experienced virtual social spaces, they were able to use their own mental models to share how they related to the tools, technologies, and experiences being discussed. The [qualitative section](#) of this report shares many examples of mental models participants used to discuss these policy proposals.

Now, to quantitatively assess participants' understanding of the policy proposals and virtual social spaces overall, the participants were asked a series of eight knowledge questions about bullying and harassment and about ways of detecting and responding to it in virtual social spaces. These knowledge questions are important pieces of information that participants would have come across in these deliberations, in the briefing videos, or in the reading materials they received. The findings of these knowledge questions showed that there was an average gain from deliberation on these questions, about ten percentage points, a statistically significant increase. Specifically, one of the multiple questions asked participants to define members only spaces and the percentage of participants answering this question correctly increased from 62 to 75%, an 13 percentage point increase. Participants' understanding of what an avatar can do also increased from 48 to 58%, a 10 percentage point increase. Overall, the average increase for the eight knowledge questions was 8 percentage points, from 48 to 56%.

The participants were also more confident in their knowledge of all aspects of this agenda. When asked whether or not they were confident about their knowledge of the metaverse, Meta's role, the differences between public and members-only spaces, what bullying and harassment are like in the metaverse and what platforms like Meta are currently doing about it, participants expressed confidence on these questions with above 70% and some above 80% confidence on these questions.

At the end of the process, participants completed a variety of evaluation questions. More than 80% thought that the important aspects of the issues were covered in the group discussions, that the experts presented different points of view in a balanced way, that the plenary sessions facilitated a balanced discussion, and that the issue guide was balanced. 73% thought that the deliberation platform "tried to make sure that opposing arguments were considered" and 78% thought the members of their group

“participated relatively equally.” More than 60% thought that the small group discussions and the process as a whole helped them clarify their position on the issues. Reflecting on the event as a whole, 75% agreed “I learned a lot about people very different from me – about what they and their lives are like.”

The process also increased participants’ sense that they had “opinions worth listening to on these issues,” rising from 72% to 79%. A similar percentage said they were confident that they had “come to an informed judgment that Meta can consider in making decisions.” 82% said they would “recommend this event to Meta as a way to make decisions in the future.” Finally, 79% said they were “confident that Meta will take the polling results seriously in making decisions.”

### Project Impact and Significance

This first-ever global Deliberative Poll, Community Forum, is significant not only for Meta, but also for the field of deliberative democracy. Over the last 35 years, since James Fishkin pioneered the method of Deliberative Polling in 1988, there has not been a single deliberative event at this scale, let alone implemented with the scientific and social science rigor of this Community Forum. There was one [Global Climate Assembly](#) in 2021. While participants attended from various countries globally, that event was small (only 99 participants) and thus could not credibly represent the citizens of the world. Being able to conduct a globally representative deliberation is no small feat, as it must be scientifically rigorous, while, at the same time, logistically manageable, and result in findings that are impactful and implementable. The Community Forum, in our view, achieved all three.

First, on scientific rigor, the methodology of this event was a credible application of established social science methods for survey recruitment and for the design of balanced deliberation. What was unusual is that these methods were applied on a global scale. Furthermore, by honing in on social media users, the event was able to focus the recruitment efforts to populations that could more reasonably be reached with newer and more targeted recruitment methods, such as web-based panels.

Second, one of the key technologies used to make this event possible was the Stanford Online Deliberation Platform. Not only was this platform used in 19 languages, but the AI-assisted featured moderator allowed for 2,069 small groups to have the same deliberative experience. Being able to offer the same deliberative experience and eliminate variability in implementing the actual deliberative process is key to ensuring that any change (or no change) after deliberations is due to the deliberation itself, and not factors such as moderators or materials presentation. Every participant in this event heard the same introduction, briefing videos, prompts for the discussions, and interacted with the same interface. No matter how much training DDL can give human moderators, there will always be variations between moderators, because moderators are human. Over the last 35 years, DDL has held in-person and online sessions with human moderators and regardless of the amount of training, there will also be moderators that do not perform well and therefore, negatively affect participants’ experiences. The Stanford Online Deliberation Platform has eliminated the human variation and has allowed for participants to deliberate with AI-assistance successfully in 32 countries for the Community Forum.

Logistically, without this platform, this event would have needed over 500 human moderators in 19 languages and would have needed additional assistants to support over 2,000 small group sessions. This event proved that deliberation can be done at scale and moreover, the positive evaluation and experiences from the participants indicate that they enjoyed the use of the platform, learned about discussion topics and others, and changed their opinions.

Furthermore, the AI platform has produced a wealth of quantitative and qualitative data that sheds light on participants' opinions, small group dynamics, cultural differences, and many more research questions that can be answered.

Lastly, as the findings in this report will demonstrate, this event had statistically significant and impactful results, which provide clarity on what participants would like to see implemented as the guardrails for behavior in the metaverse. While the results of this event are advisory, an overwhelming majority of participants globally had confidence that Meta would use these results to make policies. Participants have signaled that they in turn feel confident about the opinions they shared and therefore, Meta should feel confident in these results as well. This statement is in fact quite significant as in many public policy surveys and even after some Deliberative Polls, participants often do not feel confident that government and/or policymakers will follow through.

The results of this Community Forum revealed what mattered to the participants, how important many of these proposals are to them, and the type of Metaverse they would want to spend time in. With the confidence that the participants have in the process, they are confident that Meta will use the results of this Community Forum.

### **Methodology**

#### **Sampling Methodology**

Meta, working in collaboration with Stanford's Deliberative Democracy Lab (DDL) and the Behavioral Insights Team (BIT), held a first-of-its-kind event--a virtual deliberation of the global public of social media users. The event brought together a representative sample of social media users in each of the main regions of the world.

This virtual event was based on the method of Deliberative Polling® that measures public opinion on core issues before and after deliberation. The project design offered both an in-depth and representative look at opinion around the world on important policy issues and demonstrates how public opinion can shift after structured and substantive dialogue.

The participants in this event were representative of social media users in each of the major regions of the world. The sampling and recruitment techniques differed based on regions and countries. For example, where available, participants were drawn from already existing online panels. And, in some



countries, participants were drawn from address-based probability sampling. Instead of aiming to sample from every country in the world which is impractical, we chose to create representative samples in consultation with regional experts.

The selection of countries within each region from which social media users were sampled was based upon regional demographics, market of social media users, market of VR users. We also considered the practicalities of language, so as to not proliferate the number of languages, which would make this event unmanageable. Post stratification weighting was implemented based on regional demographics and information about social media usage.

This project organized the deliberations based on the different time zones in different regions of the world. The deliberations took place over a weekend (and involved 12 hours of contribution) or weekend equivalent, where sessions were offered during the weekday over the course of a couple weeks.

This project anticipated about 700 participants from each region, who would engage in small group deliberations of 8-12 people from their own regions. Depending on languages available in the region and participants' preferences, participants would engage in multi-national discussions. For example, in Latin America, participants from Mexico, Chile and Colombia were Spanish-speaking and thus, the participants from these three countries were randomly assigned to rooms and deliberated together. All the group deliberations were within a given region (and proximate time zones). There were not cross regional discussions, that is, an English speaker from the Sub-Saharan Africa region did not deliberate with an English speaker in the United Kingdom. Following the small group deliberations, all participants joined plenary session panels with other participants in that region. Those deliberations were simultaneously translated into multiple languages.

### The Regions

This event defined the world into nine regions:

East Asia
South and Southeast Asia
Oceania
North America
Latin America
Middle East and North Africa
Sub-Saharan Africa
Western Europe

## Eastern Europe and Central Asia

This categorization considered geographical regions, time zones, industry market categorization, culture, and languages. Categorizations made by world organizations such as the UN, World Bank, and WHO were also used as references. As a point of comparison, WHO uses six regions, while the UN and World Bank use seven regions (with slightly different categorizations). Although these organizations divide the world into fewer regions, some of those categories mixed together different cultures and markets that make it more difficult to analyze results. As an example, the WHO combines the Americas into one single region, but this project would strive to clearly delineate the Americas and in fact, study the differences between participants in North America and Latin America. Similarly, the UN combines Europe and Northern America into one region, obscuring differences which this project intended to study. Lastly, the categorization of regions also took into account the [World Values Survey](#), which categorizes the world into cultures according to survival vs self-expression values and traditional vs secular values. The definition of regions considered some degree of shared values within regions.

### Selection Criteria

There are almost 200 countries in the world. Since this event intended to study regions, not specific countries, there was not a need to have separate samples of each country. If this event wanted to draw meaningful inferences about the social media users in each country, the samples for each country would need to be large enough to do so, at least 200-300. And, such numbers would have led to an enormous project, larger than necessary for analyses on a regional basis.

Here were the key selection criteria:

- a) **Sample Size:** Each region needed to have a sample large enough for meaningful evaluation of the representativeness of that region and the statistical significance of any changes of opinion among the participants from that region.
- b) **Regional Focus:** The aim of the sampling was to represent social media users in the region as a whole, not in each country.
- c) **Social Media and VR usage:** Generally, countries would be included if there is high usage of social media and high near-term potential for growth in virtual reality usage.
- d) **Demographic Representativeness:** The selection of countries aimed for demographic representativeness of the region as a whole, considered apart from those countries which would make participation unlikely or impossible—situations of conflict<sup>6</sup> or authoritarian countries<sup>7</sup> which prohibit Meta and/or other social media platforms.
- e) **Practical Considerations:** The number of languages, and hence the number of simultaneous translators for the **plenary** sessions, needed to be kept limited. Hence, mostly

<sup>6</sup> Removes 22 countries – Ukraine, Colombia, Venezuela, Algeria, Lebanon, Libya, Myanmar, Syria, Tunisia, Yemen, Afghanistan, Iran, Sri Lanka, Burkina Faso, Central African Republic, Chad, Mali, Mozambique, Niger, Somalia, South Sudan

<sup>7</sup> Removes 4 countries that have banned Meta products – China, Russia, Turkmenistan, North Korea.

monolingual countries that did not constitute a major market for social media and/or for Meta, were not included. For example, Japan is treated as a largely mono-lingual country, however, the level of VR usage in Japan is extremely high, making it a priority for inclusion in this event. In contrast, small islands with their own languages were excluded.<sup>8</sup> These countries constitute only a tiny percentage of the world population.

In each region, the anticipated sample size was approximately 700 with the exception of Oceania and North America. In these regions, the anticipated sample size was 500, because these regions included only 2 countries: Australia and New Zealand for Oceania and Canada and the United States for North America. Other island nations as noted above generally have their own languages and also constitute a much smaller percentage of world population. Further, island nations are generally more expensive to sample from given their geographical locations.

As such, one country from each region will have a sample size of 400 to allow for further study into subgroups in that country. The project is unable to have a sample size of 400 per country due to resources, as every additional person requires incentives for participation. The remaining 3 countries in each region will have a sample size of 100.

Region	Anticipated Sample
East Asia	700
South and Southeast Asia	700
Oceania (AUS & NZ)	500
North America	500
Latin America	700
Middle East and North Africa	700
Sub-Saharan Africa	700
Western Europe	700
Eastern Europe and Central Asia	700
<b>TOTAL</b>	<b>5900</b>

<sup>8</sup> [List of Pacific Islands](#) and [List of Caribbean countries](#)

The sample size of 700 was selected for each region despite the population differences. The reasoning is a larger sample size is actually not needed for a larger population, based on the basic principles of statistical sampling. Though it sounds counter intuitive, it is true. For reference, a [simple sample size calculator](#) can help give you a rough estimate of the sample required for the South and Southeast Asia population (2,643,114,343) and compare the Oceania population (41,499,343). This calculator will show you based on a 95% confidence level, both populations would require a sample of 384 persons. The above table aims for 500 persons because, based on previous Deliberative Polling experiences, recruitment efforts and dropout rates will differ by country.

### Sampling and Recruitment Techniques

This Community Forum recruited a population sample of global social media users aged 18 or older from the 14 survey organizations. The sampling and recruitment techniques employed in each region varied given the state of public opinion research in that country. All regions employed post stratification weighting where necessary and provided demographic comparisons to the region as a whole. Below are the list of regions and the polling firms used in each region and country.

#### **North America: United States and Canada**

**United States and Canada - YouGov:** YouGov is a global public opinion company that operates online panels in all regions of the world. In many countries, many of their panels have millions of panelists, which allow for greater ability to access a representative population sample. YouGov panels are opt-in and use proprietary weighting algorithms to ensure accurate representation. Stanford's Deliberative Democracy Lab has engaged with YouGov since the late 2000s for Deliberative Polling recruitment and has since worked together on over a dozen projects.

#### **South America: Chile, Colombia, Mexico, Brazil**

**Colombia and Mexico - YouGov:** See YouGov description above.

**Chile - Dinamica Plataforma:** Dinamica is a Chilean polling firm that partnered with our collaborator Fundacion Tribu, who led our national Deliberative Poll in Chile. Fundacion Tribu shared the probability-based sample created for the previous Deliberative Poll with Dinamica. Fundacion Tribu served as an advisor to Dinamica for this project.

**Brazil - OPUS and YouGov:** For Brazil, this project opted to use two polling firms due to the larger sample size and also the uncertainty of potential drop-off rates. DDL has conducted Deliberative Polls (and Latin America) in Brazil previously and drop off rates were quite high and unpredictable. To ensure we reached our target, we opted for two firms. OPUS is a polling firm in Brazil that operates an online panel and a concierge service that stays in touch with its respondents frequently. This service was necessary given the culture of the country. See above for YouGov description.

#### **Western Europe: United Kingdom, Spain, Germany France**

**United Kingdom, Spain, Germany France - YouGov:** See YouGov description above.

### **Eastern and Central Europe: Romania, Czech Republic, Poland, Turkey**

**Romania and Poland - YouGov:** See YouGov description above.

**Czech Republic - STEM:** STEM is a non-profit organization that has been involved in mapping public opinion, political behavior, and quality of life in Czech society since the 1990s. STEM's main focus is on topics of democracy, trust in other people and national institutions, perception of international institutions (EU, NATO), and civil participation. STEM employed web-based surveys and provided a concierge service to assist participants.

**Turkey - TEAM:** The Center for Social Impact Studies (TEAM) has been conducting studies in the field of consumer and voter profile research since 2009. TEAM is one of the leading polling firms in Turkey and rooted in rigorous and evidenced-based research. TEAM employed web-based surveys and provided a concierge service to assist participants.

### **Middle East and North Africa: Saudi Arabia, Morocco, Israel, Egypt**

**Saudi Arabia - YouGov:** See YouGov description above.

**Morocco - YouGov and Random Dynamic Resources:** See YouGov description above. Random Dynamic Resources is a leading survey organization in Sub-Saharan Africa and MENA. RDR employed address-based probability sampling in-person and by phone. While in-person and phone recruitment are more time and resource intensive, it is noted that, knowing its population and the commitment required for this event, in-person and phone recruitment was necessary to ensure high turnout.

**Egypt - YouGov and ERTC:** See YouGov description above. Egyptian Research and Training Center (ERTC) is a leading survey research firm in Egypt, with focus on rigorous academic research. ERTC employed probability-based sampling through CATI, using phone and web surveys to recruit the participants.

**Israel - Panel4All:** Panel4All was established in 2006 and was the first polling firm in Israel to perform surveys based on a panel of internet respondents. Panel4All has a panel of respondents consisting of over 30,000 active panelists. The panel and system are used by tens of thousands of research and academic entities in Israel and worldwide.

### **Oceania: Australia and New Zealand**

**Australia - YouGov:** See YouGov description above.

**New Zealand - Kantar:** Kantar is a global polling firm and one of the subsidiaries is Kantar New Zealand, is a full-service data and insights polling firm operating in New Zealand and the wider Asia Pacific region. Kantar New Zealand maintains an online panel, ConsumerLink, which regularly runs web surveys for a wide range of clients.

### **East Asia Region: Japan, South Korea, Hong Kong, and Taiwan**

**Japan - RJC Research:** RJC is based in Japan and maintains an online panel of close to 4 million panelists. DDL recruited Dr. Shinya Wakao, a longtime Japanese collaborator from the College of the Mainland in Texas, to assist with this project and liaise with RJC. DDL has previously worked with Dr. Wakao and RJC for another Deliberative Polling project, so they not only have experience working together, but working together on a Deliberative Polling project.

**South Korea - Hankook Research:** Hankook is a leading polling firm in South Korea and maintains an online panel of over 800,000 panelists. Hankook had previously conducted Deliberative Polls nationally for the South Korean government. The team is well-versed in the Deliberative Polling method and its rigor.

**Hong Kong - HKPORI:** The Hong Kong Public Opinion Research Institute is run by Dr. Robert Chung, who is the current President of the World Association for Public Opinion Research (WAPOR). Dr. Robert Chung also founded WAPOR Asia, which has created a formal setting for leading polling firms in Asia. HKPORI runs an online panel of over 85,000 panelists. HKPORI has worked with DDL since 2009 and has conducted Deliberative Polls in Hong Kong themselves.

**Taiwan - National Taiwan University (NTU):** The National Taiwan University runs their own Web Survey (NTUWS) panel that is sponsored by Taiwan's National Science and Technology Council. NTUWS has over 25,000 panelists.

### **South and Southeast Asia: Indonesia, India, Thailand, and Philippines**

**Indonesia - Alvara Strategic:** Alvara is a research and consulting company that specializes in conducting online surveys using probability-based sampling.

**India - CVOTER:** Centre for Voting Opinion & Trends in Election Research (CVOTER) is South Asia's leading agency for Public Opinion Research with almost two decades of experience. CVOTER conducts a weekly omnibus of national representative 1200+ Samples in 10 languages. CVOTER maintains a robust panel of 100,000 respondents across India who actively use the internet on a day-to-day basis.

**Thailand - CVOTER and YouGov:** For Thailand, this project engaged with two polling firms to recruit the national sample. CVOTER and YouGov both operate online panels in Thailand. See description of CVOTER above. See YouGov description above.

**Philippines - Social Weather Station, HKPORI, and YouGov:** For the Philippines, this project first commissioned an address-based sample from Social Weather Station, the leading polling firm in the Philippines. Due to cost limitations, the recruited sample was passed to HKPORI for management for the rest of the project. See YouGov description above.

### **Sub-Saharan Africa: Nigeria, Kenya, Ghana, South Africa**

**Nigeria, Kenya, Ghana, South Africa - Random Dynamic Resources:** A leading survey organization in Sub-Saharan Africa and MENA. RDR employed address-based probability sampling in-person and by

phone. While in-person and phone recruitment are more time and resource intensive, it is noted that, knowing its population and the commitment required for this event, in-person and phone recruitment was necessary to ensure high turnout.

### Target Countries by Region

Based on the above criteria, below is the list of countries selected by region, along with sample size, and language. It is important to note that the minimum sample size per country/jurisdiction is 100. A sample of this size was necessary to ensure good sample selection, anything less would make it difficult to engage in rigorous probability sampling (or the nearest equivalent from online panels).

#### **East Asia (4 languages)**

Country/Jurisdiction	Sample Size	Language
Hong Kong	~100	Cantonese
Japan	~400	Japanese
Taiwan	~100	Mandarin
South Korea	~100	Korean

#### **South/Southeast Asia (5 languages)**

Country	Sample Size	Language
Philippines	~100	Tagalog/English
India	~400	Hindi/English
Indonesia	~100	Indonesian
Thailand	~100	Thai

#### **Sub-Saharan Africa (1 language)**

Country	Sample Size	Language
Nigeria	~400	English
Kenya	~100	English
South Africa	~100	English

## COMMUNITY FORUM - RESULTS ANALYSIS

Ghana	~100	English
-------	------	---------

### Oceania (1 language)

Country	Sample Size	Language
Australia	~400	English
New Zealand	~100	English

### North America (1 language)

Country	Sample Size	Language
Canada	~100	English
United States	~400	English

### Latin America (2 languages)

Country	Sample Size	Language
Mexico	~100	Spanish
Brazil	~400	Portuguese
Colombia	~100	Spanish
Chile	~100	Spanish

### Middle East/North Africa (3 languages)

Country	Sample Size	Language
Egypt	~400	Arabic
Saudi Arabia	~100	Arabic
Morocco	~100	Arabic/French
Israel	~100	Hebrew

### Western Europe (4 languages)



## COMMUNITY FORUM - RESULTS ANALYSIS

Country	Sample Size	Language
UK	~100	English
Germany	~400	German
France	~100	French
Spain	~100	Spanish

### Eastern Europe (5 languages)

Country	Sample Size	Language
Turkey	~400	Turkish
Poland	~100	Polish
Romania	~100	Romanian
Czech	~100	Czech

In sum, the project proceeded with 32 countries/jurisdictions, 19 languages, and anticipated 5900 participants participating in this Community Forum.

As a controlled experiment<sup>9</sup>, this project had a control group of a comparable size to mirror the participant group. The control group completed the pre-deliberation and post-deliberation survey at the same time as the participants in the Community Forum. The only difference was that the control group did not participate in any deliberation activities including small group discussions, plenary sessions and did not receive any briefing materials. This control group demonstrated the effect of deliberation on the Deliberative Polling participants. And the control group showed the Deliberative Polling participants were not any different from those that were not invited to participate.

### Actual Recruitment Numbers

The numbers of deliberators in the Community Forum and in the control group exceeded this project's expectations. This project expected a total of 5,900 deliberators in the Community Forum and the 5,900 respondents in the control group. In actual turnout, the Community Forum had 6,488 deliberators and 6,357 in the control group. Compared to our expected turnout, the Community Forum had 608 more participants across the 32 countries. This cumulative difference adds together the number of participants over the targets in some countries and under the targets in other countries. In actual difference, there were several countries where this project did not meet the expected numbers, as such among deliberators, the actual difference was 558. Among the control group, the cumulative difference was 987 and the actual difference was 297.

### Expected v. Actual Turnout: Deliberators and Control Group

---

<sup>9</sup> Research Note on Sampling: It might be argued that the oversampled countries in each region (United States, Germany, Australia, Brazil, Japan, Nigeria, Turkey, Egypt) will distort the sample because the countries are not sample proportional to the population sizes. However, the essential treatment for the Deliberative Poll is the discussion in the small groups made up of 10-12 participants. Participants spend the majority of their interactions with others in the small group and there is no expectation that any given small group would represent a country, region, or the world. It is actual interaction and deliberation that is being tested in this experiment, therefore the sample size as whole for any given country would not distort the sample. DDL has conducted controlled experiments and collected other data that establish that it is the small group discussions that drive the opinion change in the Deliberative Polling design. No small group of ten can represent a large population. Hence, the over-sample will not distort the effective treatment as it is experienced by the participants.

Another angle on this is that DDL also sees little variation in the small group changes depending on their composition when the participants are randomly assigned. The design has sufficient balance built in that the groups tend overwhelmingly to move in the same direction. Hence the over-sampling of one country in the region as a whole is unlikely to affect the opinion movements in the small groups. Participants are receiving essentially the same treatment regardless of the group they are assigned to. For reference, in the 2021 National US Deliberative Poll on Climate and Energy, the project had nearly a thousand participants randomly assigned to the small groups for deliberation with the AI-assisted platform. The results showed no discernible effect from the composition of the different groups even though the composition varied widely because of random assignment.

	<b>Deliberators</b>	<b>Control</b>
<b>Expected Total</b>	<b>5900</b>	<b>5900</b>
<b>Actual Total</b>	<b>6488</b>	<b>6887</b>
Cumulative Difference	608	987
Actual Difference	-558	-297

In the Oceania region, where both countries utilized web panels for recruitment, 317 deliberators attended from Australia (expected 400) and 86 from New Zealand (expected 100). The Australian deliberators took part in weekday and weekend sessions, while the New Zealand deliberators only joined for the weekend. The New Zealand polling firm opted for weekend only to concentrate the recruitment on one weekend.

Recruitment in the South and Southeast Asia region did not meet our expectations. Of the 700 expected deliberators, only 364 deliberators attended from this region. Of the 364, 121 participants joined us from Indonesia, well above the expected number of 100. The firm used web-based surveys with concierge servicing. Unfortunately, India, Philippines, and Thailand fell short of expectations. While the Philippines, Thailand, and Indonesia held weekday and weekend sessions, India opted for weekend only sessions.

In East Asia, all four countries reached and/or went beyond the expected number of deliberators. Hong Kong, Taiwan, and South Korea were expected to have 100 deliberators each, and the actual numbers were 111, 126, and 104, respectively. In Japan, the expected number was 400 and the actual number was 398. All four countries used web panels for recruitment and only opted for the weekend sessions.

In Western Europe, the United Kingdom, Spain, and France exceeded the expectation of 100 deliberators per country as the actual numbers were 159, 119, and 111, respectively. In Germany, the expected number was 400 and the actual number was 351. All four countries participated in the weekday and weekend sessions and were recruited through YouGov panels.

In Central and Eastern Europe, Turkey and Poland exceeded expectations and Romania and Czechia came very close. Turkey was expected to have 400 deliberators and the actual number was 568. Poland was expected to have 100 deliberators and the actual number was 105. Romania and Czechia both were expected to have 100 deliberators each and the actual numbers were 85 and 90, respectively. Czechia was the only country of the four that opted for weekend only sessions. All four countries used web-based panels.

In North America, the expected number for the United States was 400, with the actual being 759. For Canada, the expected number was 100 and the actual number was 93. The almost double amount in the United States was not expected. Both countries deliberated in the weekday and weekend sessions. Both countries used web-based panels from YouGov.

In Latin America, all four countries exceeded expectations. The expected number for Colombia, Mexico, and Chile was 100 deliberators for each country, and the actual numbers were 159, 181, and 107. The expected number for Brazil was 400, and the actual number was 517. All four countries deliberated in the weekday and weekend sessions. All countries used web-based panels; Brazil's OPUS had a concierge service as well.

In Sub-Saharan Africa, three of the four countries exceeded expectations, with the fourth country coming very close. Kenya, Ghana, and South Africa were expected to have 100 deliberators per country, and the actual numbers were 109, 120, and 92, respectively. In Nigeria, the expected number was 400, and the actual number was 473. All four countries joined the weekday and weekend sessions. The SSA polling firm used phone and web-based surveys and deployed a concierge service as well.

## Deliberators in the Community Forum

Region/Country	Expected	Actual	Difference
<b>Oceania</b>			
<b>AUS</b>	400	317	-83
<b>NZ</b>	100	86	-14
<b>South/Southeast Asia</b>			
<b>Philippines</b>	100	40	-60
<b>India</b>	400	143	-257
<b>Indonesia</b>	100	121	21
<b>Thailand</b>	100	60	-40
<b>East Asia</b>			
<b>Hong Kong</b>	100	111	11
<b>Taiwan</b>	100	126	26
<b>South Korea</b>	100	104	4
<b>Japan</b>	400	398	-2
<b>West EU</b>			
<b>Germany</b>	400	351	-49
<b>UK</b>	100	159	59
<b>Spain</b>	100	119	19

## COMMUNITY FORUM - RESULTS ANALYSIS

France	100	111	11
<b>Central/Eastern Europe</b>			
Poland	100	105	5
Romania	100	85	-15
Turkey	400	568	168
Czechia	100	90	-10
<b>North America</b>			
US	400	759	359
Can	100	93	-7
<b>Latin America</b>			
Mexico	100	181	81
Chile	100	107	7
Brazil	400	517	117
Colombia	100	159	59
<b>Sub-Saharan Africa</b>			
Kenya	100	109	9
Ghana	100	120	20
Nigeria	400	473	73
South Africa	100	92	-8
<b>Middle East/North Africa</b>			
Saudi Arabia	100	58	-42
Israel	100	120	20
Morocco	100	98	-2
Egypt	400	508	108

Control Group in the Community Forum

## COMMUNITY FORUM - RESULTS ANALYSIS

Region/Country	Expected	Actual	Difference
<b>Oceania</b>			
<b>AUS</b>	400	314	-86
<b>NZ</b>	100	100	0
<b>South/Southeast Asia</b>			
<b>Philippines</b>	100	35	-65
<b>India</b>	400	347	-53
<b>Indonesia</b>	100	121	21
<b>Thailand</b>	100	105	5
<b>East Asia</b>			
<b>Hong Kong</b>	100	111	11
<b>Taiwan</b>	100	125	25
<b>South Korea</b>	100	132	32
<b>Japan</b>	400	404	4
<b>West EU</b>			
<b>Germany</b>	400	486	86
<b>UK</b>	100	135	35
<b>Spain</b>	100	142	42
<b>France</b>	100	153	53
<b>Central/Eastern Europe</b>			
<b>Poland</b>	100	209	109
<b>Romania</b>	100	178	78
<b>Turkey</b>	400	489	89
<b>Czechia</b>	100	108	8
<b>North America</b>			
<b>US</b>	400	512	112

<b>Can</b>	100	140	40
<b>Latin America</b>			
<b>Mexico</b>	100	144	44
<b>Chile</b>	100	89	-11
<b>Brazil</b>	400	731	331
<b>Colombia</b>	100	168	68
<b>Sub-Saharan Africa</b>			
<b>Kenya</b>	100	95	-5
<b>Ghana</b>	100	93	-7
<b>Nigeria</b>	400	392	-8
<b>South Africa</b>	100	94	-6
<b>Middle East/North Africa</b>			
<b>Saudi Arabia</b>	100	97	-3
<b>Israel</b>	100	50	-50
<b>Morocco</b>	100	191	91
<b>Egypt</b>	400	397	-3

## Weighting Methodology

For the weighting methodology, DDL engaged with methodological consultants to verify and assess our methodologies.<sup>10</sup>

## Methodology

<sup>10</sup> Our Methodology Consultants were from NORC at the University Chicago and from Generation Lab, a leading polling firm focused on the younger generations. From NORC at the University Chicago, DDL consulted with Norman Bradburn, Professor Emeritus and Senior Fellow at NORC at the University of Chicago and Colm O'Muircheartaigh, Professor at the University of Chicago and Senior Fellow at NORC at the University of Chicago. Both are leading experts in the survey field, Norman Bradburn was the head of NORC for decades prior to stepping down. From Generation Lab, DDL consulted with Madeleine Gates, Statistician at Generation Lab and Rebecca Oh, Chief Methodology Officer at Generation Lab. DDL has collaborated with Generation on multiple occasions over the last few years. Generation Lab served to verify the execution of our weighting methodology and provide third party assessment.

Weighting survey samples is a commonly used practice to reduce the bias from lack of coverage. In other words, weighting can help ensure that the survey data more accurately represents the demographics of the population the survey is meant to represent. In the context of this survey, key demographics thought to impact social media usage include age, gender, and region of the world.

Weighting requires reliable reference data about the target population. We use publicly available data about the population of social media users in the stipulated nine regions of the world. The calculation of weights was conducted using the R package *anesrake*<sup>11</sup>, which computes raking weights. Raking is a widely used weighting method for survey samples which ensures that weighted survey demographics match the target population demographics. It requires knowledge of population totals of demographic variables without needing to know all cross classifications. Multiple weighting schemes were calculated for different population levels of analysis, and were applied to both treatment and control groups, respectively. After weighting, we conducted difference in difference analyses between the treatment and control groups for the results.

#### Weighting A: Individual Country Weights

The goal for this level of weighting would be to look at rates of social media use within a certain country. Therefore, weights are calculated within each respective country sample using reference data corresponding to that country. Weighting was conducted with variables gender (male or female) and age category (18-24, 25-34, 35-44, 45-54, and 55+). Because cross-tabular data was used as a reference, this is cellular weighting with 10 levels. Reference data for these relative country-specific population proportions are publicly available. This analysis was only able to access Meta only figures of social media use for the country level analyses. The original reference data also included individuals under 18, so the relative proportions were normalized by the over-18 sum of proportions. All reference data for the 32 countries can be found [here](#).

#### Weighting B: Global Weights

The goal for this level of weighting is to look at rates of social media use globally. However, we realize that the countries included are not a random sample of all countries globally or even within a region. Our use of these weighting schemes attempts to ensure that overall demographic breakdowns reflect global social media users. The reference data relies on publicly available data aggregated across 8 social media platforms. All global reference data can be found [here](#).

This option applies raking weighting using relative proportions of global social media users by gender (male or female), age categories (18-24, 25-34, 35-44, 45-54, or 55+), and region (North America, Latin America, Western Europe, Central/Eastern Europe, East Asia, South/Southeast Asia, Oceania, Sub-Saharan Africa, or MENA). This option adjusts for the size of the population of social media users by region.

---

<sup>11</sup> Package details here: <https://cran.r-project.org/web/packages/anesrake/anesrake.pdf>



A weighting schematic for the treatment group and control group is available for reference at the end of this document.

### Assessment of Weighted Results

Overall, the weights A and B have minimal impact on the results. To examine the effects of the weights, comparisons were made between the unweighted results and the results of A and B. The analysis was performed on the Overall results, results by Age categories, results by Gender, and results by the 9 regions. For each comparison, the analysis examined the unweighted results at T1 (before deliberation), at T2 (after deliberation) and the difference T2-T1 to the weighted results. For example, for comparing Weight B, the analysis computed the difference between the unweighted results for a given question and B results for the same question. If the two results were the same or very similar there would be little to no difference. The same analysis was done for T2 and T2-T1 for every question.

Please see detailed comparisons for:

[Participants – Unweighted v Weighted – Overall](#)

[Participants – Unweighted v Weighted – By Age](#)

[Participants – Unweighted v Weighted – By Gender](#)

[Participants – Unweighted v Weighted – By Region](#)

The same analysis was conducted for the control group, but only for the Overall results.

Please see detailed comparison for:

[Control – Unweighted v Weighted – Overall](#)

An assessment was also made of the Difference of Differences. That is the ‘Participants’ Difference T2-T1’ and ‘Control Group’s Difference T2-T1’. This assessment examines whether the participants that deliberated had more substantial shifts in opinions than the control group that did not deliberate.

Please see detailed comparison for:

[Difference of Difference – Unweighted v Weighted - Overall](#)

The analysis revealed that there were minimal differences between the unweighted and weighted results. Almost all of the differences were close to 0%-1%.

Conclusion: Global generalizations can be made. If regional generalizations were to be made, it would be important to note any outlier countries within a particular region.<sup>12</sup>

---

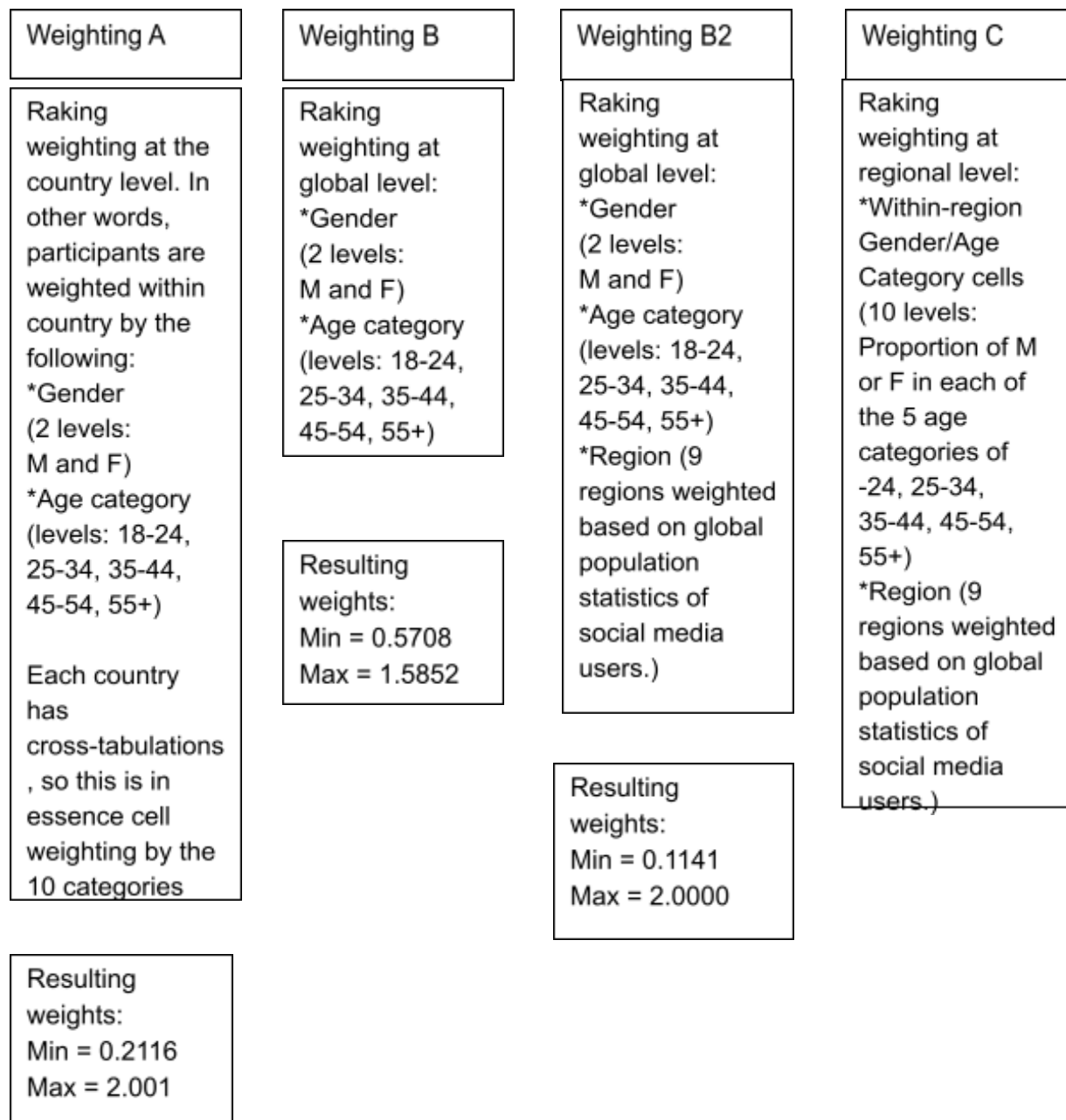
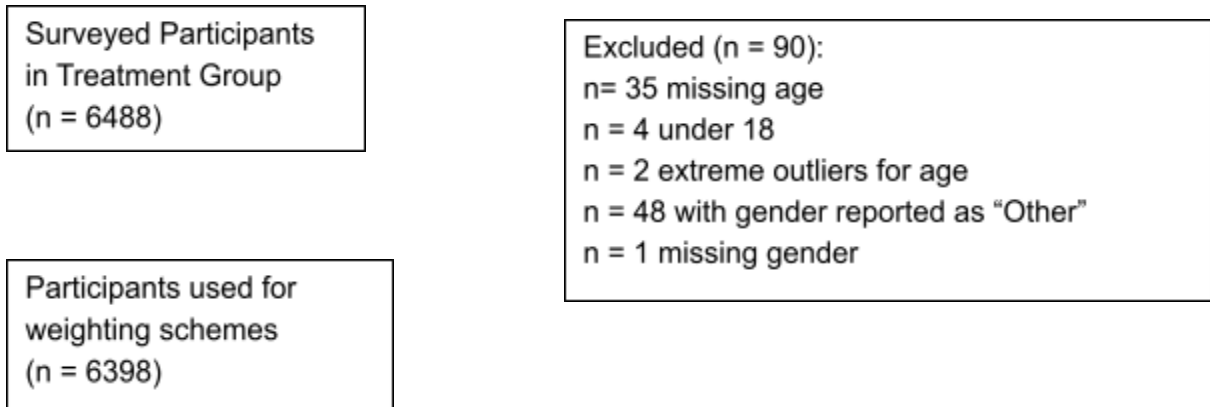
<sup>12</sup> Note on Interpretation and Generalization of Weighted Results

Prior to weighting, DDL closely examined the unweighted results by region, gender, and age categories. This examination was necessary to inform how the weights should be interpreted and generalized. The analyses found that movement of opinions for countries within regions and globally were generally in

---

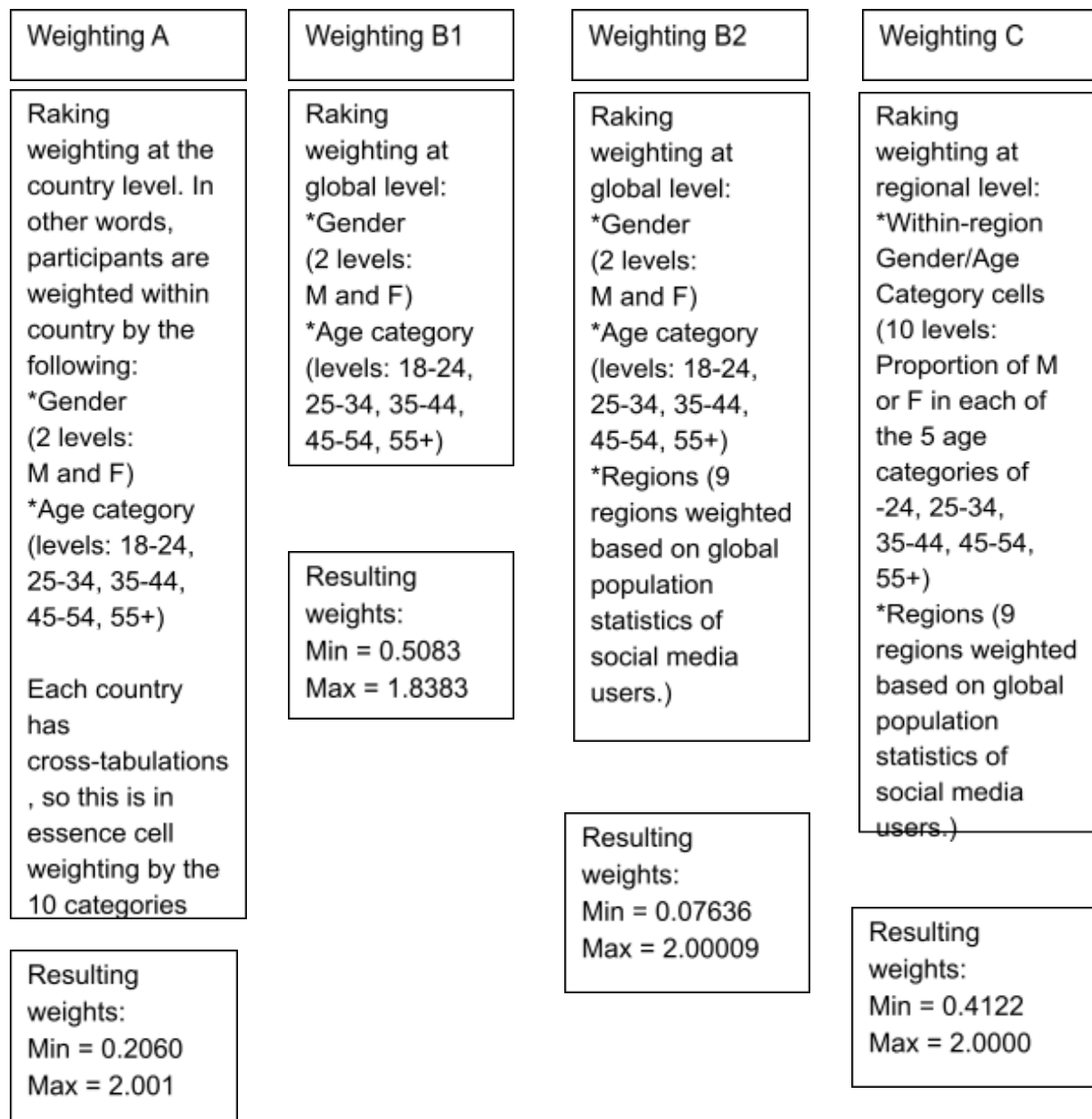
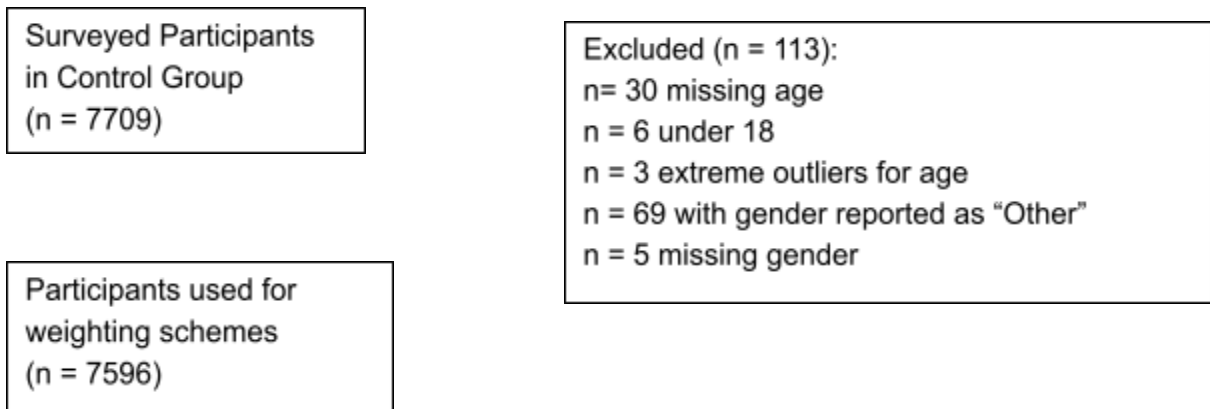
the same direction. That is, participants, no matter what country or region they were joining from, largely moved in the same direction. To be clear, the starting points at T1 and where participants' opinions arrived at in T2 may be different (and that is expected), but the movement of opinions were largely in the same direction. As such, generalizations of movements at a global and regional level are acceptable.

## Weighting schematic – Treatment Group



## COMMUNITY FORUM - RESULTS ANALYSIS

### Weighting schematic – Control Group



### Quantitative Results

The quantitative analysis starts with a review of the survey results for all participants, followed by results by region, and finally by demographics (gender, age and metaverse experience). For each section, the analysis is broken down into seven questions:

1. Who should be responsible for monitoring social VR spaces?
2. To whom should cases of bullying and harassment be reported?
3. Should creators be made aware of reports made to platform owners?
4. What should be done about social VR spaces where there is repeated bullying and/or harassment?
5. What tools should be used to identify bullying and harassment in social VR spaces?
6. Who should review video capture and automatic speech detection alerts?
7. Should moderators be used to monitor bullying and harassment in social VR spaces?

The results take two forms. Firstly, the participants' answers are averaged for each question on a 0 to 10 scale:

- 0 to 3 = strong opposition to the proposal
- 4 = moderate opposition to the proposal
- 5 = in the middle, the proposal is neither supported nor rejected
- 6 = moderate support for the proposal
- 7 to 10 = strong support for the proposal

On top of the average, the results for each question are broken down into percentages of participants that:

- Oppose the proposal = participants whose answers on the 0 to 10 scale were under 5 [0-5]
- Neither oppose nor support the proposal = participants whose answers on the 0 to 10 scale were 5 [5]
- Support the proposal = participants whose answers on the 0 to 10 scale were above 6 [6-10]
- Don't know / NA = no answer to the question

The average answer on the 0 to 10 scale and the percentage of participants are two ways to express the same results. **The average answer gives a general overview of the sentiment for a proposal, whereas the percentage of participants helps break down the distribution of participants on a given proposal.** The percentage of participants do not express as much nuance as the 0 to 10 scale does, as opposition lumps all results under 5 together, meaning a very strong opposition (0 to 3) and a moderate opposition (4) are treated equally. The same is true for support, where strong support (7 to 10) is treated the same as moderate support (6). As such, a proposal could have a large number of participants supporting it, for example 70%, but its average answer could be rather moderate, for example 6.1, because the support was moderate for most participants. **Therefore, when reviewing the results from this Community Forum, it is important to take both types of measures into consideration to get a comprehensive and accurate understanding of the participants' answers.**

When reading the results, please:

- **Consider the subtleties.** Most proposals have answers on the positive side of the scale (above 5), however an answer of 6 expresses a much lower level of support than a 7 or 8. The former expresses a substantial amount of skepticism, whereas the latter two express a much more firm and at times, almost unanimous support from participants. **This is especially important as even a small difference in the average answer could represent an important difference in the percentage of participants.**
- **Consider how the answers contrast and compare with other answers.** When participants were asked several versions of the same proposal, for example one about creators and the other about platform owners, both might have received the participants' support, but it is important to see which one received more support than the other. As such, an answer of 6.4 for one proposal might seem to indicate moderate support from participants, but if the alternative proposal had an answer of 6.8, it clearly indicates that one proposal received more support than the other, even though both proposals received moderate support.
- **Consider the big picture.** Even if one proposal received more support than the other proposal, if the answers were moderate (6) or below for both, it indicates that many participants opposed the proposals, neither supported nor opposed them, or did not answer.

Participants were surveyed twice for this Community Forum, once before deliberations and once after. When analyzing the results of this Community Forum, it is thus important to consider the changes in results before and after the deliberations. This allows us to see how the deliberations impacted participants' opinions and attitudes towards the proposals discussed.

- An increase in support for a proposal after deliberations generally suggests that the participant found the proposal more compelling or persuasive as a result of the deliberations. This could be due to a variety of factors, such as exposure to new information, addressing concerns or reservations, or a shift in perspective or priorities.
- Conversely, a decrease in support for a proposal could indicate that the participant found it less persuasive after deliberations. This could be due to factors such as new information that contradicted their initial beliefs, or concerns that were not addressed during the deliberations.

### General Analysis

This section looks at the results for all participants, across all regions and demographics. It showcases the global levels of support and opposition to each of the proposals from the 32 countries included in this Community Forum. When reading this section, it is important to remember that these are global averages, and that the answers to each proposal vary by region and demographics. For a deeper dive into the results by region and demographics, please see later sections of this paper.

### Key Takeaways

#### Monitoring

- Participants expressed support for both creators and platform owners to monitor social VR spaces. In public spaces, they supported platform owners more than creators, and in members-only spaces creators more than platform owners.

#### Reporting

- Participants expressed support for cases of bullying and harassment to be reported both to creators and platform owners. However, support for reports to be made to platform owners was higher than for reports to be made to creators, across both public and members-only spaces.
- Additionally, participants expressed strong support for severe cases of bullying and harassment that are sent to creators to be shared with platform owners for additional enforcement.
- For reports made to platform owners, participants wanted those reports to be shared with creators. When such reports would be shared with creators, participants preferred for the personal information of the users involved to not be disclosed.

#### Visibility of spaces

- Participants expressed moderate support for the proposals to make members-only spaces where there is bullying and/ or harassment less visible and not publicly discoverable. The proposal to remove such spaces from the platform was not supported by a majority of participants.

#### Consequences for creators

- Participants expressed moderate support for actions to be taken against creators who built spaces where there is repeated bullying and/ or harassment.
- The proposal to ban creators from making additional members-only spaces and the one to ban creators from inviting additional people to join a space where there is repeated bullying and/ or harassment both failed to garner the support of a majority of participants.
- Participants expressed moderate support for the proposal that creators should not be able to earn money from spaces where there is repeated bullying and harassment.
- The most popular type of action to be taken against creators of spaces where there is repeated bullying and/ or harassment was to require creators to take a course of how to moderate the

spaces they create. This proposal received significantly more support than the other proposed types of actions to be taken against creators.

### User notification

- Participants expressed strong support for users to be notified when entering spaces where there is repeated bullying and/ or harassment.

### Use of technology

- Participants expressed support for the use of video capture and automatic speech detection in public spaces and spaces where there is bullying and/ or harassment. They also supported their use in members-only spaces, but to a lesser extent than for other types of spaces.

### Review of alerts

- Participants expressed support for both creators and platform owners to have access to video capture. However, more participants supported for platform owners to access video capture than for creators.
- Participants also supported both creators and platform owners to have access to automatic speech detection alerts. However, their preferences for who should have access to those alerts depended on the type of space. For public spaces and spaces where there is repeated bullying and/ or harassment, they preferred for platform owners to review the alerts, and for members-only spaces they preferred for creators to review the alerts.

### Moderators

- The extent to which participants supported the use of moderators varied according to the type of space. In public spaces, participants strongly supported the use of moderators, both visible and invisible. In spaces where there is repeated bullying and/ or harassment moderators received moderate support. Finally, in members-only spaces moderators received even more moderate support.
- Participants expressed moderate support for moderators to enter spaces when requested by creators and users. The levels of support were almost equivalent for requests by creators and requests by users.
- Overall participants supported the use of visible moderators more than invisible moderators, regardless of the type of space or the type of request (creator or user).
- Finally, participants supported the proposal that creators should be able to select other users to be visible moderators in their spaces.

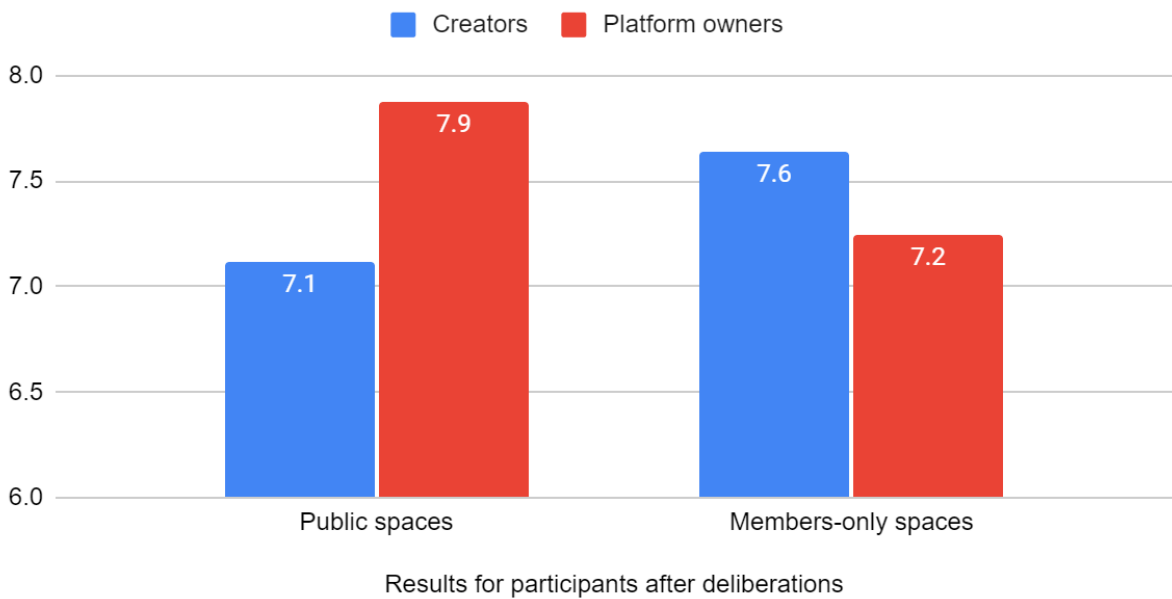
A more detailed breakdown of the findings can be found below.



Who should be responsible for monitoring social VR spaces?

Participants were asked whether creators or platform owners should be responsible for monitoring social VR spaces. Overall, participants supported social VR spaces to be monitored both by creators and platform owners. However, the extent to which participants supported platform owners and creators, and which one they supported more, depended on the type of space. For public spaces, participants supported the monitoring by platform owners more than by creators. For members-only spaces it was the opposite, with creators receiving more support than platform owners.

Who should be responsible for monitoring spaces in the metaverse?



a. Monitoring of Public spaces

Participants supported the monitoring of public VR spaces by platform owners more than by creators. In fact, the average answer for creators monitoring public spaces slightly decreased over deliberations, going from 7.2 pre-deliberations to 7.1 after. On the contrary, support for platform owners to monitor public spaces increased from 7.6 pre-deliberations to 7.9 post-deliberations. As such, after deliberations, the average answer was 0.8 points higher for platform owners than creators. That is a significant difference, as it translates to 10 percentage points more participants supporting the monitoring of public spaces by platform owners (≈80%) than creators (≈70%). Opposition to the monitoring of public spaces by creators was also double (15.4%) that of platform owners (7.6%).

	Before deliberations	After deliberations
Creators should be primarily responsible for monitoring public spaces they create.	7.2	7.1

Oppose	14.10%	15.40%
In the middle	10.40%	12.60%
Support	70.10%	69.90%
DK/NA	5.40%	2.10%
Platform owners should be primarily responsible for monitoring public spaces even if they're built by Creators.	7.6	7.9
Oppose	9.60%	7.60%
In the middle	9.60%	10.30%
Support	75.20%	79.70%
DK/NA	5.50%	2.50%

b. Monitoring of Members-only spaces

For members-only spaces, the participants supported monitoring by the creators of the spaces more than by platform owners. Although before deliberations the average answer was the same for creators and platform owners, at 7.3, after deliberations it went down to 7.2 for platform owners, when it went up to 7.6 for creators. The difference post-deliberations between creators and platform owners is not as significant for members-only spaces (0.4 points) as it is for public spaces (0.8 points), but overall participants endorsed the monitoring of members-only spaces by creators more than by platform owners. 7 percentage points more participants supported creators monitoring public spaces ( $\approx 78\%$ ) than platform owners ( $\approx 71\%$ ). Opposition to the monitoring of members-only spaces by creators went down over deliberations, going from 12.3% to 9.4% of participants, whereas it went up for platform owners, from 11.8% to 13.7%.

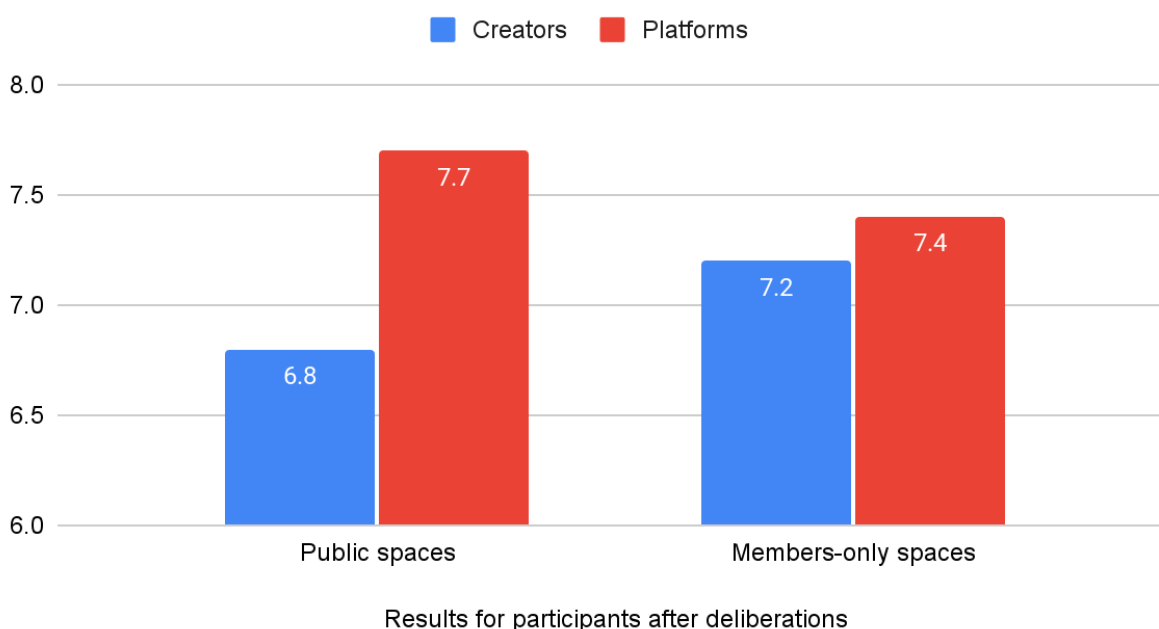
	Before deliberations	After deliberations
Creators should be primarily responsible for monitoring members-only spaces they create.	7.3	7.6
Oppose	12.30%	9.40%
In the middle	11.60%	10.20%
Support	71.00%	78.20%
DK/NA	5.20%	2.30%
Platform owners should be primarily responsible for monitoring members-only spaces even if they're built by Creators	7.3	7.2
Oppose	11.80%	13.70%
In the middle	10.90%	12.80%

Support	71.40%	70.80%
DK/NA	5.90%	2.80%

### To whom should cases of bullying and harassment be reported?

Participants supported bullying and harassment cases to be reported to both creators and platform owners, but expressed more support for platform owners than for creators, both for public spaces and members-only spaces.

### To whom should reports be made to?



#### A. Reporting of bullying and harassment for public spaces

For public spaces, the preference for reporting to platform owners was particularly pronounced, with an average pre-deliberations answer of 7.2 in support of reporting to platform owners compared to 6.7 in support of reporting to creators. After deliberations, the gap between these two widened in favor of platform owners, with an average score of 7.7 in support of reporting to platform owners compared to 6.8 in support of reporting to creators. The initial preference was already in favor of platform owners before deliberations, with a 0.5 gap between scores in support of creators and platform owners. However, after deliberations, the gap further increased to 0.9 in favor of platform owners. This was because opposition to reporting to platform owners in public spaces decreased from 12.9% pre-deliberations to 8.8% post-deliberations, while opposition to reporting to creators in public spaces increased slightly from 18.2% to 18.6%. On the other hand, support for creators to receive reports increased only from 62.6% to 65.5% over deliberations, whereas for platform owners it increased much

more, going from 70.7% to 78.1%. As such, platform owners had 8.1% more support than creators before deliberations, and 12.6% more after deliberations.

	Before deliberations	After deliberations
In public spaces built by Creators, people should report to Creators, and Creators decide the consequences.	6.7	6.8
Oppose	18.20%	18.60%
In the middle	11.90%	12.70%
Support	62.60%	65.50%
DK/NA	7.20%	3.10%
In public spaces built by Creators, people should report to platform owners, and platforms decide the consequences.	7.2	7.7
Oppose	12.90%	8.80%
In the middle	10.00%	10.30%
Support	70.70%	78.10%
DK/NA	6.40%	2.80%

#### B. Reporting of bullying and harassment for members-only spaces

For members-only spaces, participants favored reporting cases of bullying and harassment to platform owners than to creators, but to a lesser extent than for public spaces. Prior to deliberations, the average answer in support of reporting to creators was 6.9, which increased to 7.2 after deliberations. In contrast, the average answer in support of reporting to platform owners was 7.2 before deliberations and 7.4 after deliberations. Participants showed a preference for reporting to platform owners both before and after deliberations, but the gap between the two decreased during deliberations (from 0.3 to 0.2). This was because support for reporting to creators increased at a slightly higher rate than for platform owners. However, despite this increase, the level of support for reporting to creators in members-only spaces was not enough to surpass the level of support for reporting to platform owners.

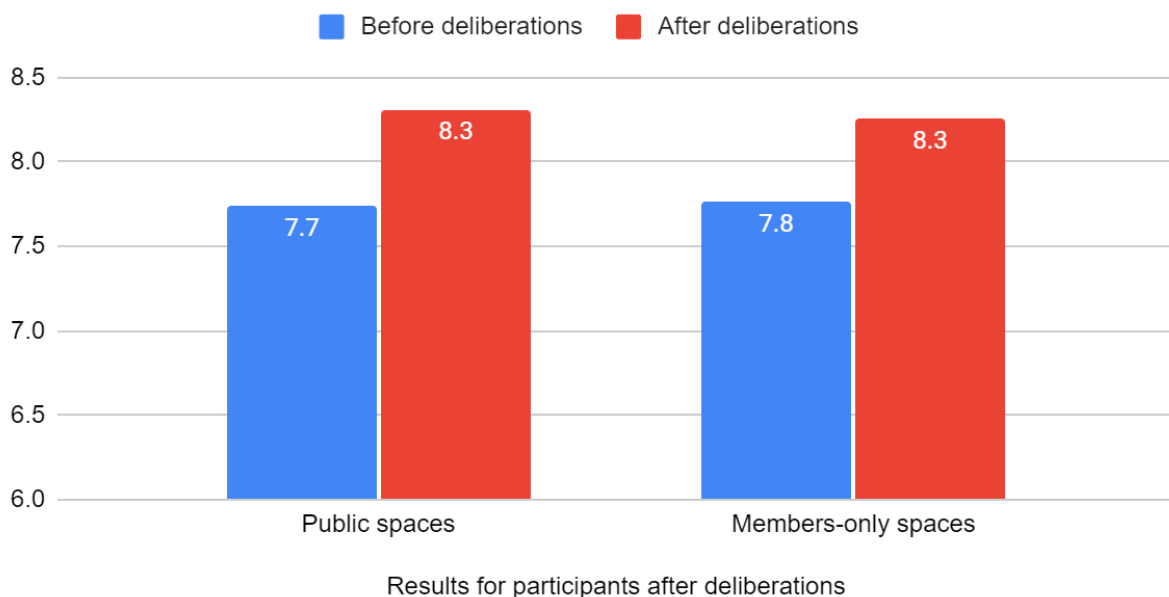
	Before deliberations	After deliberations
In members-only spaces, people should report to Creators who built them, and Creators decide the consequences.	6.9	7.2
Oppose	15.00%	13.20%
In the middle	11.00%	11.90%
Support	66.50%	71.60%

DK/NA	7.40%	3.30%
In members-only spaces, people should report to platform owners and platforms decide the consequences.	7.2	7.4
Oppose	13.20%	12.00%
In the middle	10.20%	11.00%
Support	69.80%	74.10%
DK/NA	6.80%	3.00%

### C. Severe cases of bullying and harassment

Finally, participants were asked whether creators should send severe cases of bullying and harassment to platform owners for additional enforcement. Prior to deliberations, the average answer for this proposal was 7.7 in public spaces and 7.8 in members-only spaces, which increased to 8.3 after deliberations in both cases. As such, the proposal received strong support before deliberations, and this support grew even further as a result of deliberations.

### Should creators send severe cases to platform owners for additional enforcement?



	Before deliberations	After deliberations
In public spaces built by Creators, Creators should send severe cases to platform owners for additional enforcement	7.7	8.3

Oppose	8.80%	4.90%
In the middle	7.90%	6.50%
Support	77.10%	85.90%
DK/NA	6.30%	2.70%
In members-only spaces, Creators should send severe cases to platform owners for additional enforcement.	7.8	8.3
Oppose	7.70%	4.90%
In the middle	8.10%	6.70%
Support	70.20%	78.30%
DK/NA	14.00%	10.10%

In conclusion, participants demonstrated a greater preference for reports of bullying and harassment to be made to platform owners, regardless of the type of space and the severity of the case. Nonetheless, participants also showed support for reports to be made to creators, but to a lesser extent than for platform owners.

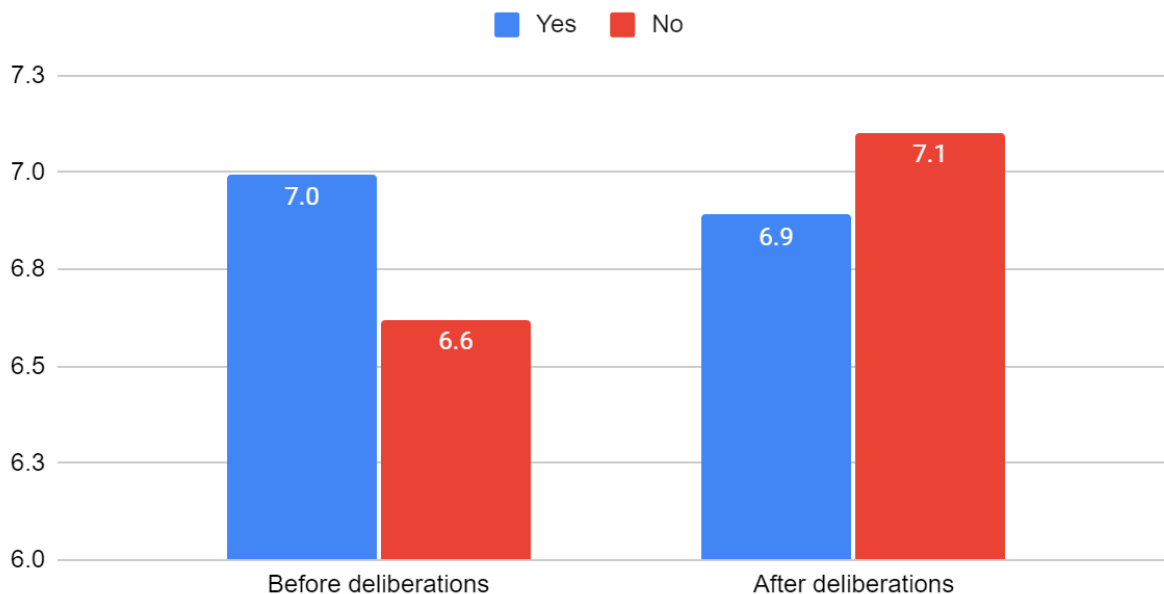
## Should creators be made aware of reports made to platform owners?

As discussed in the previous section, more participants supported reports of bullying and harassment to be made to platform owners than to creators. In addition, participants were asked if reports made to platform owners should be shared with creators. Overall, participants were opposed to platform owners keeping the reports to themselves, with an average answer of 4.4 before and after deliberations. This low score was due to a large opposition to the proposal, with 47.7% of participants opposing it before deliberations and 50% after deliberations. 10% of participants remained neutral. Only a little over a third of the participants, or 35%, were in favor of the proposal. This was one of the most disapproved proposals from the deliberations. This suggests that participants believe that creators have a right to know about reports made in spaces they created and that transparency is important in addressing issues of bullying and harassment in social VR spaces.

	Before deliberations	After deliberations
Reports are made to platform owners 1) Platform owners should not notify the creator of the space where it happened.	4.4	4.4
Oppose	47.70%	50.00%
In the middle	10.20%	9.60%
Support	35.30%	35.90%
DK/NA	6.80%	4.60%

Participants were also asked if the personal information of users involved in reported incidents should be disclosed to creators, in the event the reports were shared with them. Before deliberations, the average answer in favor of disclosure was 7, while the average answer for non-disclosure was 6.6. However, after deliberations, the trend was reversed, with non-disclosure taking the lead with an average answer of 7.1, compared to 6.9 for disclosure. Over deliberations, the option to disclose personal information lost support, with the average answer decreasing from 7 to 6.9. On the other hand, the proposal for non-disclosure gained significant support, with the average answer increasing from 6.6 to 7.1, a 0.5 increase. Even though both options, to share and not share personal information, received support from participants, the option to not disclose information received slightly more support after deliberations than the option to disclose information.

When reports are shared with creators, should personal information be included?



	Before deliberations	After deliberations
Reports are made to platform owners 2) Platform owners should notify the creator of the space where it happened, but not include personal information of the users involved.	6.6	7.1
Oppose	19.30%	14.80%
In the middle	12.30%	11.90%
Support	62.10%	68.90%
DK/NA	6.30%	4.40%

Reports are made to platform owners 3) Platform owners should notify the creator of the space where it happened with details about the event and the users involved.	7.0	6.9
Oppose	16.40%	17.80%
In the middle	11.50%	12.00%
Support	65.30%	65.50%
DK/NA	6.80%	4.70%

In conclusion, even though participants preferred reports of bullying and harassment to be made to platform owners, they still wanted those reports to be shared with the creators of the concerned spaces. The proposal to share with creators the personal information of users in reported incidents was initially supported before deliberations, but lost support after deliberations, with participants favoring non-disclosure after deliberations. These findings highlight the importance of protecting the privacy of users involved in reported incidents.

## **What should be done about spaces where there is repeated bullying and/or harassment?**

Participants were asked what should be done about spaces where there is repeated bullying and/ or harassment.

In terms of visibility of these spaces, the proposal that received the most support was the one to make members-only spaces less visible to users. The proposal to make members-only spaces not publicly discoverable received slightly less support. Finally, the proposal to remove such members-only spaces from the platform was not supported by a majority of participants.

Participants expressed moderate support for creators of spaces with repeated bullying and/ or harassment to face consequences. The proposals to ban such creators from making additional members-only spaces or from inviting more people to such spaces were not supported by a majority of participants. The proposal for creators to not be able to make money off of such spaces received slightly more support, but overall support was very moderate. The only consequence for creators that received significant support from participants was the proposal to have creators take a course on how to moderate the spaces they create.

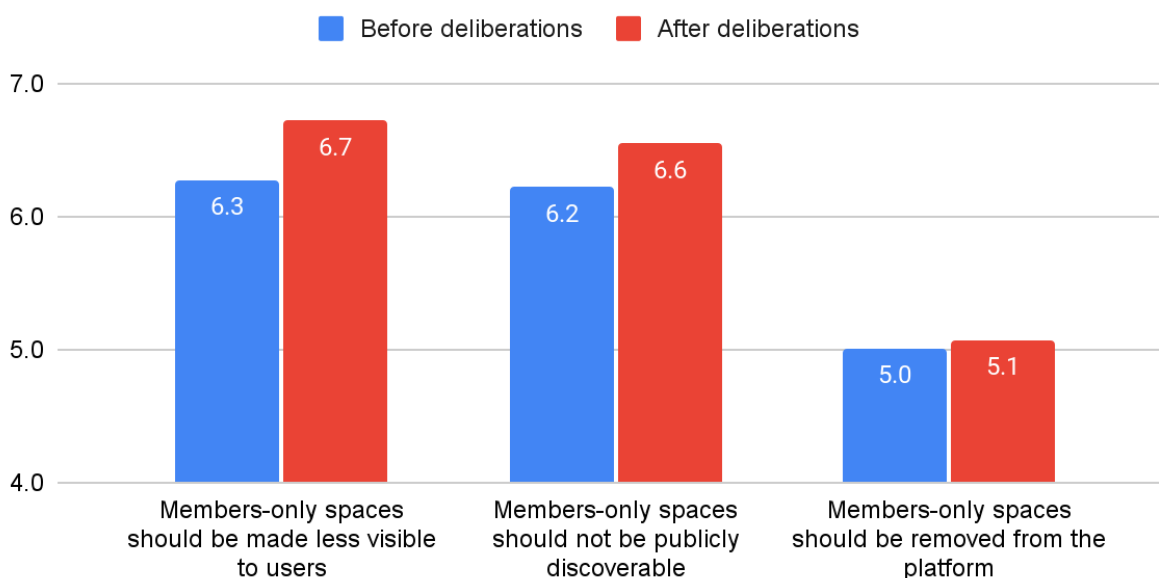
Finally, participants expressed enthusiastic support for the proposal to notify users when they enter a space where there is repeated bullying and/ or harassment. This was the only proposal, apart from the one to have creators take a moderation course, which had an average answer above 7 before deliberations and above 8 after deliberations. All the other proposals received support between 5 and 6, which means that a large proportion of participants did not support them. As such, participants favored proposals which were more corrective (making members-only spaces less visible or not publicly



discoverable, educating creators, or telling users about spaces where there is bullying and/ or harassment), rather than punitive (banning spaces or banning creators from doing something).

## A. Visibility of members-only spaces where there is repeated bullying and/ or harassment

### What should be done about the visibility of spaces where there is repeated bullying and/ or harassment?



#### a. Should members-only spaces be made less visible to users?

Participants expressed mild support for members-only spaces where there is repeated bullying and harassment to be made less visible to users. The average answer to the proposal increased from 6.3 prior to deliberations to 6.7 after. This was due to a decrease in opposition and from those who were neutral, and an increase in support.

	Before deliberations	After deliberations
Members-only spaces should be made less visible to users.	6.3	6.7
Oppose	18.70%	16.40%
In the middle	16.00%	14.40%
Support	53.90%	63.10%
DK/NA	11.50%	6.00%

#### a. Should members-only spaces not be publicly discoverable?

This proposal received slightly less support from the participants than the previous proposal, but still above 60% of support. The average answer to the proposal went from 6.2 prior to deliberations to 6.5 after. Opposition to this proposal was greater than the previous proposal, with 21.5% opposing it before deliberations and 20% after deliberations, compared to 18.7% and 16.4% respectively for the previous proposal.

	Before deliberations	After deliberations
Members-only spaces should not be publicly discoverable.	6.2	6.6
Oppose	21.50%	20.00%
In the middle	14.50%	14.00%
Support	54.50%	60.80%
DK/NA	9.50%	5.20%

## b. Should members-only spaces be removed from the platform?

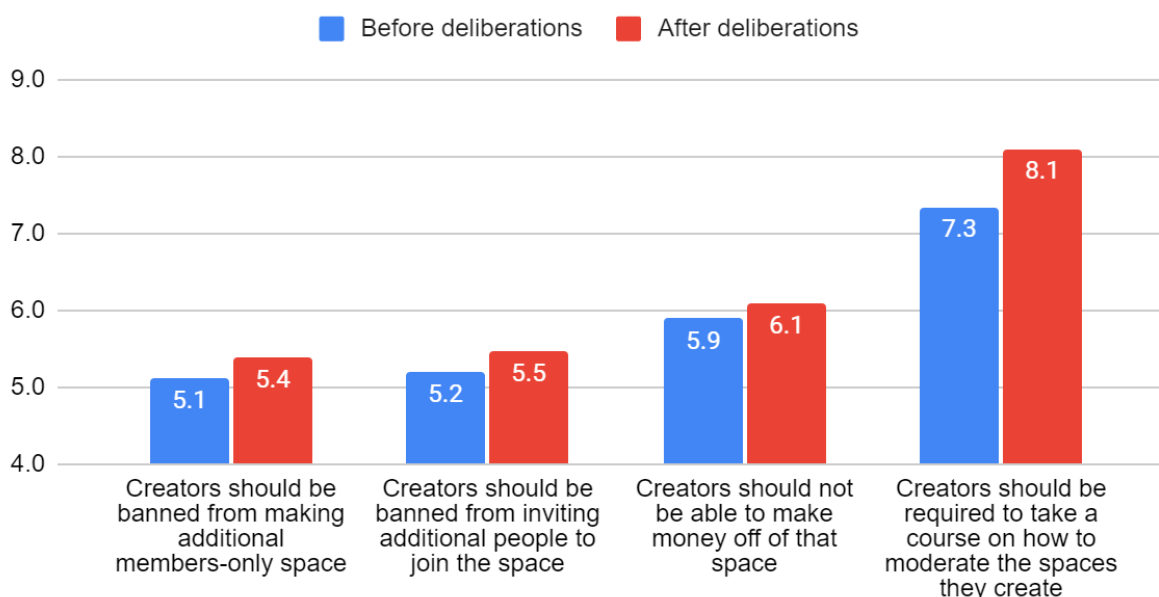
This was the least favored proposal among the three discussing changes in visibility to spaces with repeated bullying and/or harassment. The average answer remained at  $\approx 5$  both before and after deliberations, as an almost equal number of participants supported and opposed the proposal. This proposal failed to gain majority support from participants, with only 38.8% supporting it before deliberations, and 42.8% after deliberations. Even though support grew slightly over deliberations, opposition also grew slightly, going from 34.3% before deliberations to 36.3% after deliberations.

The substantial opposition to this proposal and the lack of majority support indicate that participants preferred corrective measures (making spaces less visible or not publicly discoverable) over punitive measures such as this one. This highlights the importance of taking a nuanced approach to addressing bullying and harassment in social VR spaces.

	Before deliberations	After deliberations
Members-only spaces should be removed from the platform	5.01	5.07
Oppose	34.30%	36.30%
In the middle	14.40%	14.80%
Support	38.80%	42.80%
DK/NA	12.60%	6.10%

## B. Consequences for creators

## What type of actions should be taken against creators of spaces where there is repeated bullying and/or harassment?



### a. Should platforms take action against creators?

Participants were asked whether platforms should take action against creators who built spaces, both public and members-only, in which there is repeated bullying and/or harassment. Participants expressed moderate support for those proposals. Prior to deliberations, the average response was 6.4 for both public and members-only spaces. After deliberations, it went up to 6.7 for public spaces and 6.9 for members-only spaces. The increase in average responses for both public and members-only spaces after deliberations suggests that the discussion had a positive impact on participants' opinions regarding the need for action against creators of online spaces where there is repeated bullying and harassment.

Support for actions being taken against creators in public spaces grew from 56.7% to 62.2%. Support for actions being taken against creators of members-only spaces grew even more, rising from 57.5% before deliberations to 66.7% after. Around a fifth of participants disagreed with action being taken against the creators of public spaces, both before and after deliberations. In contrast, the opposition to taking action against creators of members-only spaces decreased from 20.1% before deliberations to 16.3% after. The decrease in opposition to taking action against creators of members-only spaces suggests that the deliberations may have helped to address concerns or objections that some participants had initially.

Nevertheless, despite the increase in support and slight decrease in opposition for members-only spaces, support levels remained moderate. A substantial number of participants were either opposed to this proposal or undecided.

	Before deliberations	After deliberations
In public spaces Creators built, platforms should take action against Creators.	6.4	6.7
Oppose	20.90%	19.80%
In the middle	12.40%	12.70%
Support	56.70%	62.20%
DK/NA	10.10%	5.30%
In members-only spaces, platforms should take action against Creators.	6.4	6.9
Oppose	20.10%	16.30%
In the middle	12.30%	12.10%
Support	57.50%	66.70%
DK/NA	10.10%	4.90%

b. Should creators be banned from making additional members-only spaces?

After asking whether they should take actions against the creators of spaces where there is repeated bullying and harassment, participants were asked about the types of consequences that should be taken against creators. The first option offered was for creators to be banned from making additional members-only spaces. This proposal did not receive a lot of support from participants, with the average answer being 5.1 before deliberations and 5.4 after deliberations.

Opposition to the proposal remained around 30% both before and after deliberations. The number of participants neither for nor against also remained around 15%. Although support for the proposal did increase from 38.2% to 44.5% over deliberations, it failed to convince a majority of participants. Notably, the percentage of participants who did not know how to reply remained high over the course of deliberations, going from 14.5% before deliberations to 10.2% after deliberations. The high proportion of participants who did not know how to reply indicates that this was a complex issue with many factors to consider. It is possible that participants felt that they did not have enough information or knowledge to make an informed decision about the proposal, or that they were unsure about the potential consequences of banning creators from making additional members-only spaces.

This suggests that participants believed that simply banning creators making additional members-only spaces would not necessarily solve the issue of repeated bullying and harassment. Overall, the lack of strong support for the proposal suggests that participants were looking for more nuanced solutions to the issue of bullying and harassment in members-only spaces, and that a punitive approach was not seen as the most effective way to address the problem.

	Before deliberations	After deliberations
Creators should be banned from making additional members-only space	5.1	5.4
Oppose	32.70%	30.60%
In the middle	14.60%	14.60%
Support	38.20%	44.50%
DK/NA	14.50%	10.20%

c. Should creators be banned from inviting more people to join the space

The participants were also presented with the option of banning creators from inviting more people to join spaces with repeated instances of bullying and harassment. Like the first proposal, this option did not gain much support from participants. Before deliberations, the average answer was 5.2, going up to 5.5 after. Opposition to the proposal was substantial, with 34.3% of participants opposing it before deliberations, and 31.6% after deliberations. Although support for the proposal increased slightly during deliberations, it did not reach a majority of participants, rising from 42.7% to 48.9%.

Overall, this proposal garnered slightly more support than the previous one, but not significantly. As such, both proposals failed to receive majority support from the participants. This indicates once again that participants were hesitant to impose severe consequences on creators and were more interested in finding ways to address the issue of bullying and harassment in members-only spaces that do not involve harsh punishments for creators.

	Before deliberations	After deliberations
Creators should be banned from inviting additional people to join the space.	5.2	5.5
Oppose	34.30%	31.60%
In the middle	13.20%	14.00%
Support	42.70%	48.90%
DK/NA	9.90%	5.40%

d. Should creators be able to make money off of that space?

The participants were then presented with the proposal that creators of spaces with repeated bullying and harassment should not be allowed to profit from those spaces. This proposal received slightly more support than the previous ones, with an average answer of 5.9 before deliberations and 6.1 after. A fourth of participants opposed this proposal, and while before deliberations only 49.1% of participants

supported that proposal, after deliberations a slim majority of participants, 54%, supported it. Despite this increase in support, the level of endorsement remained low compared to other proposals discussed in this Community Forum. This suggests that there may have still been some reservations among participants about imposing severe consequences on creators.

	Before deliberations	After deliberations
Creators should not be able to make money off of that space.	5.9	6.1
Oppose	26.90%	25.90%
In the middle	14.40%	14.40%
Support	49.10%	54.00%
DK/NA	9.60%	5.80%

e. Should creators take a course on how to moderate the spaces they create?

Finally, participants were asked whether creators should be required to take a course on how to moderate the spaces they create. This was, by far, the most popular proposal for the type of action to be taken against creators. Before deliberations, the average answer was 7.3, growing to 8.1 after deliberations. Not only was this the most popular type of consequence both before and after deliberations, but it is also the one that gained the most support over the course of deliberations, with a +0.8 difference in answers. This important growth in support over the course of deliberations is due to opposition to this proposal decreasing, going from 12% before deliberations to 7.4% after deliberations. On the other hand, support grew from 67.2% to 77.4%.

The fact that this proposal was the most popular one and gained the most support over the course of deliberations indicates that participants may believe that providing creators with the knowledge and skills to moderate their spaces effectively could be an effective way to reduce bullying and harassment. It also suggests that participants may view education as a more constructive and positive approach to addressing the issue, rather than simply imposing punitive measures such as banning or restricting creators.

	Before deliberations	After deliberations
Creators should be required to take a course on how to moderate the spaces they create.	7.3	8.1
Oppose	12.00%	7.40%
In the middle	10.00%	8.10%
Support	67.20%	77.40%
DK/NA	10.80%	7.20%

C. Should users be notified when entering such a space?

The participants were asked whether platforms should notify users when entering spaces where there is repeated bullying and/ or harassment. This proposal received strong support from participants, with an average answer of 7.7 before deliberations and 8.3 after deliberations. The proposal was highly favored by the majority of participants, with 75.9% supporting it before deliberations and 84.6% after. Furthermore, the opposition to this proposal was relatively low, with only 9.1% opposing it before deliberations and 5.3% after. This proposal was the most supported among all the actions suggested for tackling bullying and harassment in spaces, reflecting the importance of transparency in promoting a safer online environment.

	Before deliberations	After deliberations
Users should receive notification of such cases when entering a space.	7.7	8.3
Oppose	9.10%	5.30%
In the middle	8.50%	7.20%
Support	75.90%	84.60%
DK/NA	6.50%	2.90%

**What tools should be used to identify bullying and/ or harassment in social VR spaces?**

Participants were asked whether video capture and automatic speech detection should be used in social VR spaces to help with the identification of bullying and/ or harassment.

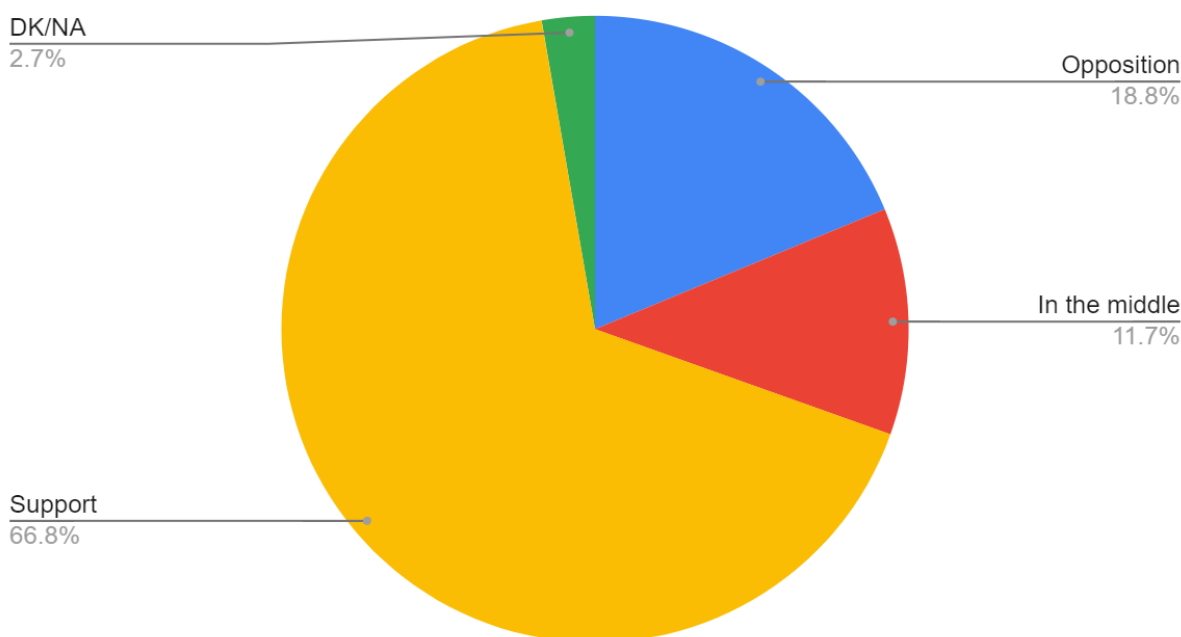
Video capture is a system built and maintained by a platform developer that continuously records what is happening in case the video needs to be referenced later. Video is overwritten or deleted after a few minutes unless someone reports misconduct or a violation. This is similar to how security recording works in the physical world. Overall, participants were approving of the use of video capture, but the degree of approval depended on the type of space. The use of video capture was most supported for public spaces, followed by spaces with repeated bullying and/ or harassment. The only type of space where video capture received moderate levels of support was members-only spaces.

Automatic speech detection is technology that “listens” for a set of upsetting words (for example, insults or offensive language) and flags instances of these words for moderators to review. The participant’s support for the use of automatic speech detection followed that of video capture. Automatic speech detection was most supported in public spaces, followed by spaces with repeated bullying and/ or harassment, and members-only spaces received the most moderate levels of support.

## A. Video Capture

Average support for the use of video capture across all types of social VR spaces increased by 0.4 after deliberations, going from 6.5 before deliberations to 6.9 after. Most of this increase is attributable to changes in opinions from participants who were previously undecided. Before deliberations, 6.75% of participants indicated that they did not know whether they supported the use of video capture in social VR spaces, compared to only 2.73% after deliberations. The same is true for those who were in the middle, going from 13.55% of participants before deliberations to 11.70% after deliberations. The number of participants opposing the use of video capture remained almost the same before and after deliberations, going from 20% to 19%. The deliberations thus enabled participants who were previously neither for nor against the use of video capture or undecided to take a stronger stance. Overall, after deliberations, 67% of participants supported the use of video capture in social VR spaces, and 19% opposed it.

Average participant answer for video capture after deliberations



Before deliberations, the use of video capture was most supported for spaces where repeated bullying and/ or harassment occur, irrespective of the type of space (public or members-only). Support for the use of video capture in these spaces grew even more after deliberations, with the average answer going from 6.8 to 7. Although support for the use of video capture in spaces where repeated instances of bullying and harassment had occurred was high overall, they were slightly higher for public spaces, both before and after deliberations, with 6.9 and 7.1 respectively, compared to 6.7 and 6.9 for members-only spaces.

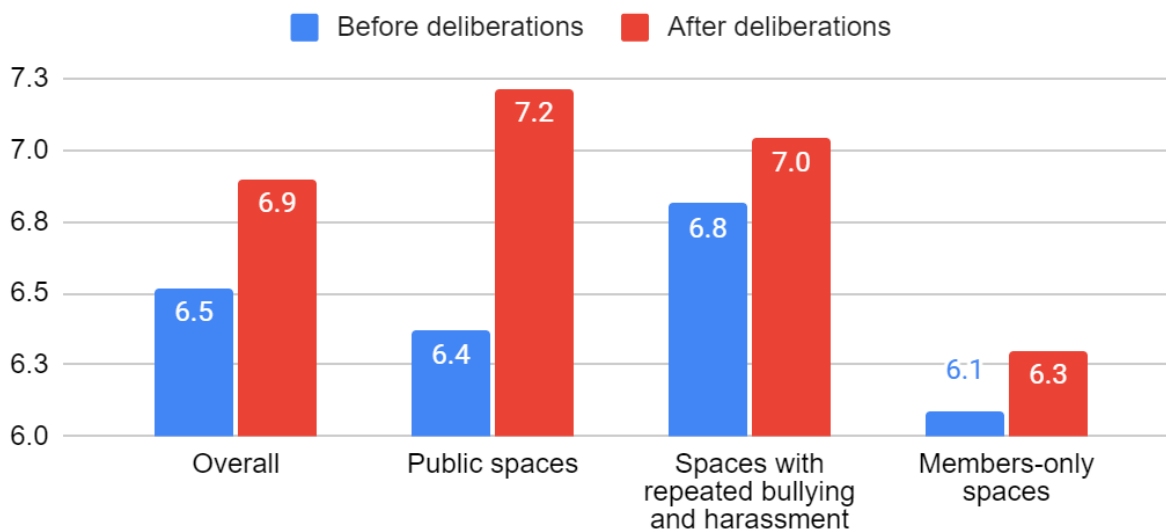


Even though support for video capture in spaces with repeated instances of bullying and harassment was high after deliberations, with the average answer at 7, support was even higher for public spaces, with an average answer of 7.2. The use of video capture in public spaces went from the second most favored option before deliberations, with 6.4, to the most favored option after deliberations, with 7.2, a 0.8 increase, the single biggest increase across all spaces. This means that the proposal went from being moderately supported by participants to strongly supported. As such, the deliberations led more people to support the use of video capture in public spaces than in any other type of space.

Support for the use of video capture in member-only spaces was lower than for other types of spaces previously discussed. Participants expressed moderate support, with 6.1 before deliberations and 6.3 after deliberations. The use of video capture in members-only spaces had the lowest levels of support among participants, with 52.4% before deliberations and 58% after deliberations. As such, after deliberations, 11% less of participants supported the use of video capture in members-only spaces than in spaces with repeated bullying and/ or harassment, and 13% less than in public spaces. The use of video capture in members-only spaces also had the most opposition, with around a quarter of participants opposing it.

	Overall		Public spaces		Spaces with repeated bullying and harassment		Members-only spaces	
	T1	T2	T1	T2	T1	T2	T1	T2
<b>Average answer</b>	<b>6.5</b>	<b>6.9</b>	<b>6.4</b>	<b>7.2</b>	<b>6.8</b>	<b>7</b>	<b>6.1</b>	<b>6.3</b>
<b>Oppose</b>	20%	18.8%	21.2%	15%	17.4%	18%	23.9%	24.1%
<b>In the middle</b>	13.6%	11.7%	14.9%	11.6%	12.1%	10.3%	15.2%	14.6%
<b>Support</b>	59.8%	66.8%	57.8%	71%	64.5%	69.2%	52.4%	58%
<b>DK/NA</b>	6.8%	2.7%	6.1%	2.4%	6.2%	2.6%	8.5%	3.3%

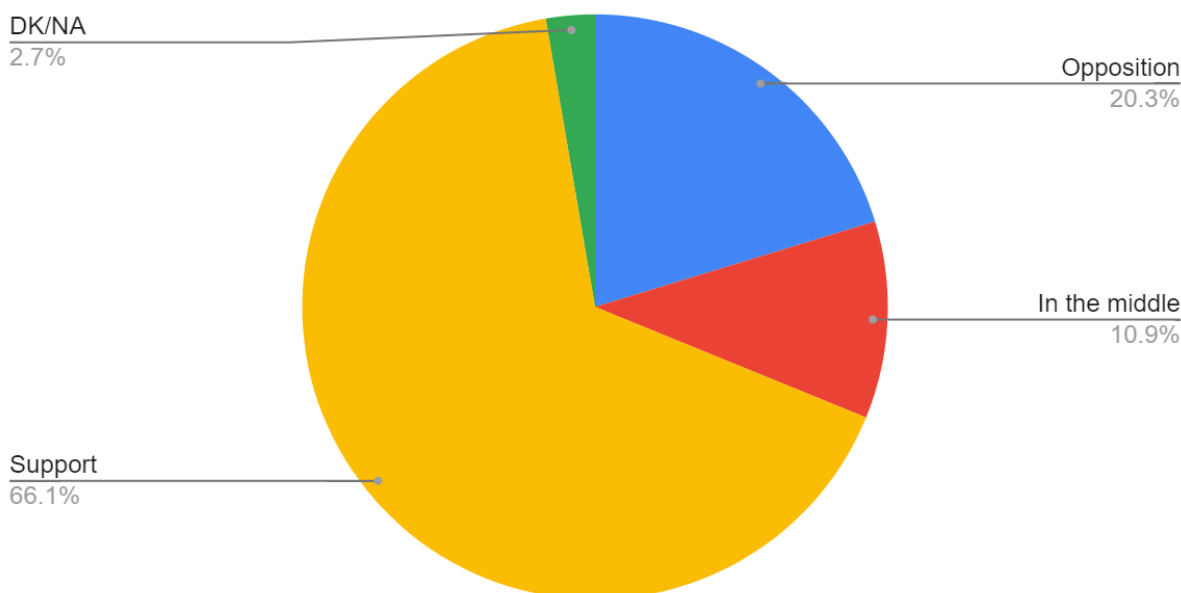
## Average participant answer for the use of video capture in social VR spaces



### B. Automatic Speech Detection

The use of automatic speech detection in social VR spaces had similar levels of support as video capture. Before deliberations, the average answer was 6.6 (6.5 for video capture), and 6.8 after (6.9 for video capture).

### Average participant answer for automatic speech detection after deliberations



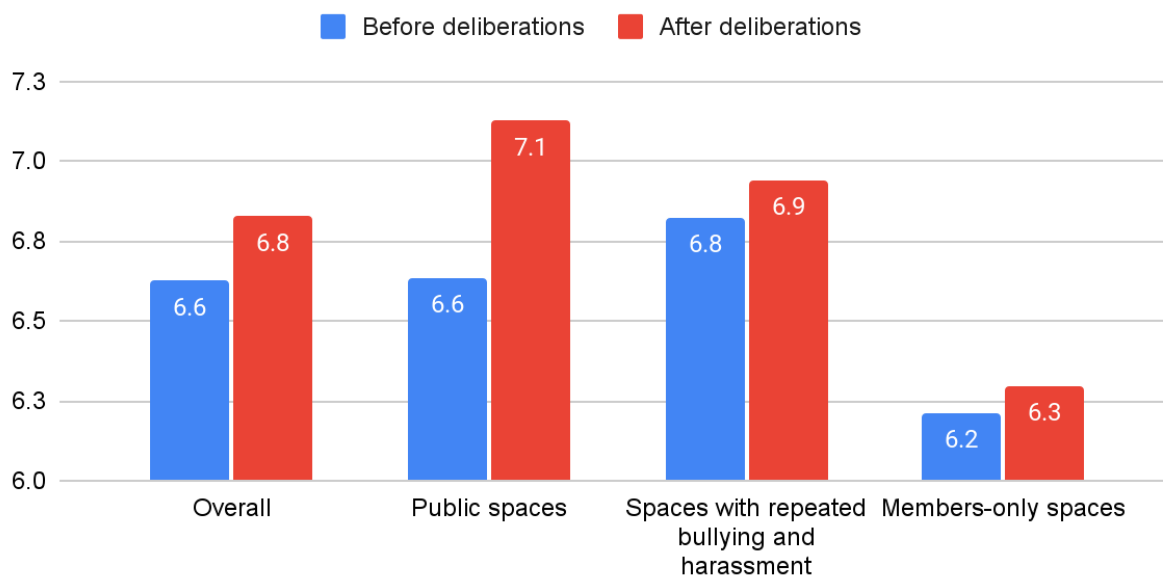
Just as for video capture, the use of automatic speech detection was most supported pre-deliberations for spaces where there is repeated bullying and/ or harassment, with the average answer being 6.8. Post-deliberations, the support grew slightly to 6.9. Among spaces with repeated bullying and harassment, support for the use of speech detection was slightly higher for public spaces, with an average of 6.9 pre-deliberations, and 7.1 post-deliberations, compared to members-only spaces, with an average of 6.8 both pre- and post-deliberations.

Post-deliberations, automatic speech detection was most supported in public spaces, with the average of answers being 7.1. Opposition to the use of automatic speech detection in public spaces decreased after deliberations, going from 21.5% to 17.8%, while support increased from 62.4% to 69.7%.

Automatic speech detection received the least amount of support in members-only spaces, with an average answer of 6.2 pre-deliberations and 6.3 post-deliberations, which expresses lukewarm support from participants. Opposition to the use of automatic speech detection in members-only spaces grew slightly after deliberations, going from 24.3% to 25%. The number of participants in the middle grew similarly, from 12.9% to 13.3%. Support also grew slightly, going from 56.1% to 58.5%, but to a much lesser extent than support grew for public spaces (from 62.4% to 69.7%) and for spaces with repeated bullying and/ or harassment (from 65.1% to 68.3%). As such, automatic speech detection received only moderate support for members-only spaces, and significantly less support than for other types of spaces (around 10% less).

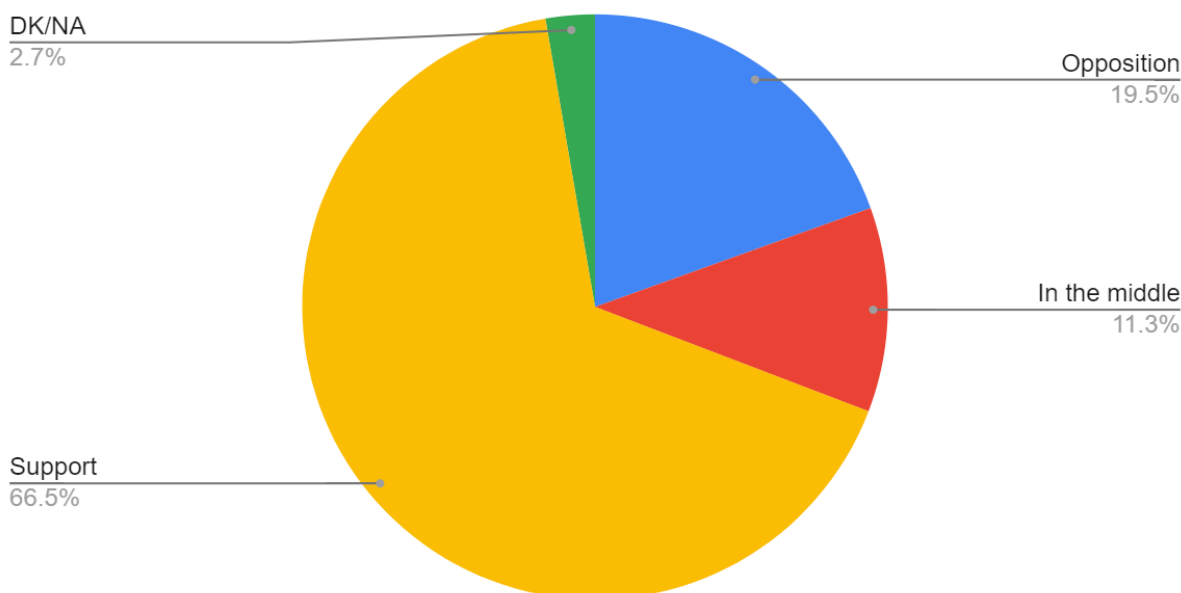
	Overall		Public spaces		Spaces with repeated bullying and harassment		Members-only spaces	
	T1	T2	T1	T2	T1	T2	T1	T2
Average answer	6.6	6.8	6.6	7.1	6.8	6.9	6.2	6.3
Oppose	20.7%	20.3%	21.5%	17.8%	18.4%	19.1%	24.3%	25.1%
In the middle	11.3%	10.9%	10.5%	10%	10.9%	10.2%	12.9%	13.3%
Support	62.2%	66.2%	62.4%	69.7%	65.1%	68.3%	56.1%	58.5%
DK/NA	6%	2.7%	5.7%	2.5%	5.7%	2.6%	6.7%	3.1%

### Average participant answer for the use of automatic speech detection in social VR spaces



C. Conclusion: What tools should be used to identify bullying and harassment in social VR spaces?

### Average participant answer for automatic speech detection and video capture after deliberations



In conclusion, the majority of participants in the study supported the use of video capture and automatic speech detection technologies in social VR spaces, with support increasing after deliberations. Support was strongest for public spaces, closely followed by spaces in which repeated instances of bullying and/or harassment happened. Overall, around a fifth of participants opposed the use of these technologies in social VR spaces. Additionally, a tenth of participants neither opposed nor supported the use of these technologies. Altogether, a little less than a third of participants either opposed or were undecided about the use of video capture and automatic speech detection in social VR spaces. This indicates that while there is significant support for the use of video capture and automatic speech detection in social VR spaces, there may be concerns among some users about the potential implications of such technologies for privacy and freedom of expression in virtual spaces.

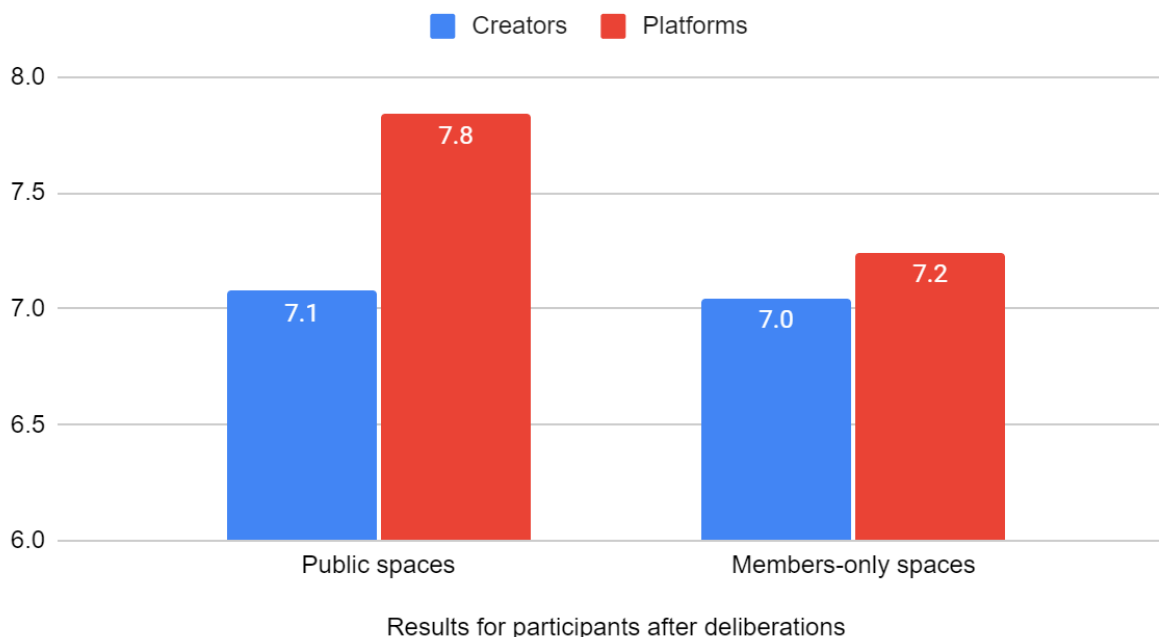
#### Who should review video capture and automatic speech detection alerts?

Participants were asked who between creators and platform owners should review video capture and automatic speech detection alerts. For video capture, while both creators and platform owners received support to review alerts, platform owners received more support than creators. For automatic speech detection alerts, participants preferred for them to be reviewed by platform owners in public spaces and spaces where there is repeated bullying and/or harassment, but had a slight preference for creators to review alerts in members-only spaces.

## A. Video capture review

Participants supported the review of video capture by both creators and platform owners, but support was overall stronger for platform owners than for creators.

### Who should have access to video capture?



		Creators	
Public spaces	Before deliberations	7	7.2
	After deliberations	7.1	7.8
Members-only spaces	Before deliberations	6.9	6.8
	After deliberations	7	7.2

Participants expressed the highest level of support for the review of video capture by platform owners in public spaces, both before and after deliberations. The average answer for platform owners to review video capture in public spaces was 7.15 before deliberations, increasing to 7.84 after deliberations, representing the largest increase in support for the review of video capture. This was due to a decrease in opposition to platform owners reviewing video capture in public spaces, which went from 13.7% before deliberations to 8.3% after deliberations. Conversely, support for the proposal increased significantly, from 64.2% before deliberations to 80.1% after deliberations, a 25% increase in support. These findings suggest that participants strongly supported for platform owners to review video capture in public spaces and that deliberations led to even greater approval of this proposal.

The review of video capture in public spaces by creators received less support compared to platform owners. Prior to deliberations, the average answer for creators reviewing video capture was 7, slightly lower than the answer for platform owners (7.15). After deliberations, the average support for creators increased to 7.1, which was significantly lower than the increase in support for platform owners (7.84). The reason for this difference is that the levels of opposition to the review of video capture in public spaces by creators increased during deliberations, from 15.7% to 17.9%. Meanwhile, support for the proposal increased only from 62.2% to 68.8%. In contrast, the levels of opposition to the review of video capture in public spaces by platform owners decreased during deliberations, from 13.7% to 8.3%, while support increased significantly, from 64.2% to 80.1%. As such, participants, although generally favorable to the review of video capture alerts by both platform owners and creators, supported the review by platform owners  $\approx 11\%$  more than by creators after deliberations.

	Before deliberations	After deliberations
Creators should have access to video capture in public spaces they built.	7.01	7.08
Oppose	15.70%	17.90%
In the middle	9.40%	10.50%
Support	62.20%	68.80%
DK/NA	12.80%	2.90%
Platform owners should have access to video capture in public spaces Creators built.	7.15	7.84
Oppose	13.70%	8.30%
In the middle	8.90%	8.90%
Support	64.20%	80.10%
DK/NA	13.10%	2.60%

For members-only spaces, prior to deliberations, participants favored the review of video capture by creators, with an average answer of 6.9 for creators, compared to 6.8 for platform owners. However, after deliberations, the preference was reversed, with an average answer of 7.2 for platform owners compared to 7 for creators. Although support for the review of video capture grew for both creators and platform owners, it grew at a faster rate for platform owners. This is due to the fact that, as with public spaces, opposition to the review of video capture by creators in members-only spaces increased during the deliberations, while it decreased for platform owners. Overall, the results suggest that the deliberations led participants to be more in favor of the review of video capture by platform owners. Nevertheless, it is important to stress that this preference was a lot less marked for members-only spaces (+0.2 in favor of platform owners compared to creators), than for public spaces (+0.7 in favor of

platform owners compared to creators). Levels of support and opposition were very close for creators and platform owners for members-only spaces (see table below for details).

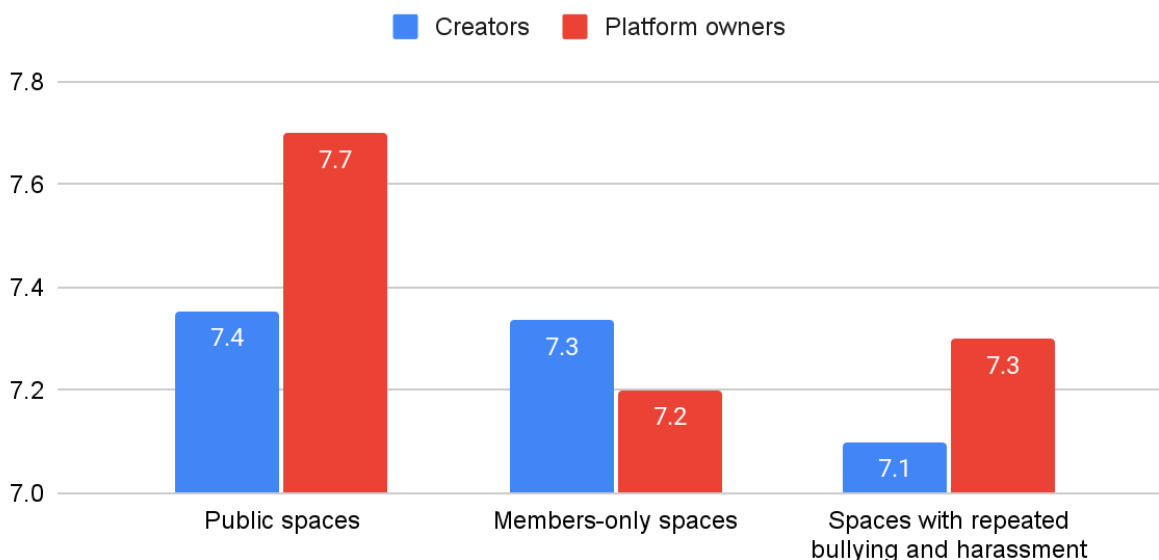
	Before deliberations	After deliberations
Creators should have access to video capture in members-only spaces they built.	6.9	7.0
Oppose	16.10%	16.20%
In the middle	9.60%	11.30%
Support	61.30%	69.90%
DK/NA	13.00%	2.60%
Platform owners should have access to video capture in members-only spaces Creators built.	6.8	7.2
Oppose	16.50%	14.10%
In the middle	9.70%	10.30%
Support	59.50%	71.20%
DK/NA	14.30%	4.30%

## B. Automatic speech detection alerts

Just like for video capture, support for the review of automatic speech detection alerts was overall slightly higher for platform owners than for creators. However, unlike for video capture, participants had a slight preference for the review of automatic speech detection in members-only spaces to be done by creators rather than by platform owners.



## Who should review automatic speech detection alerts?



Results for participants after deliberations

The review of alerts from automatic speech detection had the most endorsement for public spaces, both for creators and platform owners. Following deliberations, support for the review of alerts in public spaces increased for both creators (from 7.3 to 7.4) and platform owners (from 7.5 to 7.7). Even though participants were favorable to the review of automatic speech detection alerts by both creators and platform owners, more participants supported their review by platform owners.

	Before deliberations	After deliberations
Creators should review alerts from automated speech detection in public spaces they create.	7.3	7.4
Oppose	14.00%	14.30%
In the middle	9.40%	9.60%
Support	70.50%	73.60%
DK/NA	6.10%	2.50%
Platform owners should review alerts from automated speech detection in public spaces even if they're built by Creators.	7.5	7.7
Oppose	10.60%	9.90%
In the middle	9.00%	9.30%
Support	74.50%	78.00%

DK/NA	5.90%	2.80%
-------	-------	-------

Creators and platform owners received similar amounts of support for reviewing alerts from automatic speech detection in members-only spaces. Support for the review of alerts by creators grew from 7.1 to 7.3 over deliberations, while it remained at 7.2 for platform owners. Even though before deliberations participants had a slight preference for automatic speech detection alerts to be reviewed by platform owners, after deliberations participants changed their preference in favor of creators. This is due to opposition to creators reviewing alerts going down over deliberations, from 14.3% to 13%, and support going up from 69.3% to 73.8%. On the other hand, opposition for platform owners to review alerts went up over deliberations, from 13.8% to 15.7%, while support stagnated at ≈70%.

	Before deliberations	After deliberations
Creators should review alerts from automated speech detection in their members-only spaces.	7.1	7.3
Oppose	14.30%	13.00%
In the middle	10.00%	10.60%
Support	69.30%	73.80%
DK/NA	6.40%	2.70%
Platform owners should review alerts from automated speech detection in members-only spaces.	7.2	7.2
Oppose	13.80%	15.70%
In the middle	10.00%	10.40%
Support	70.10%	70.90%
DK/NA	6.10%	2.90%

For public spaces with repeated instances of bullying and harassment, creators and platform owners received before deliberations equal amounts of support to review automatic speech detection alerts (7 for both). However, after deliberations, support for the review of alerts increased much more for platform owners, with an average score of 7.4, compared to creators who had an average score of 7.1. As such, while participants before deliberations supported the review of alerts in public spaces where there is repeated bullying and/ or harassment by both creators and platform owners, they supported after deliberations platform owners more than creators, just like for regular public spaces.

	Before deliberations	After deliberations
Creators should review alerts from automated speech detection in	7.0	7.1

public spaces they create but only when bullying and harassment has occurred there repeatedly.		
Oppose	16.60%	16.30%
In the middle	10.10%	9.80%
Support	67.50%	71.30%
DK/NA	5.80%	2.60%
Platform owners should review alerts from automated speech detection in public spaces even if they're built by Creators, but only when bullying and harassment has occurred there repeatedly.	7.0	7.4
Oppose	15.90%	13.60%
In the middle	11.10%	8.70%
Support	67.60%	74.70%
DK/NA	5.40%	3.00%

For members-only spaces where there is repeated bullying and/ or harassment, participants supported in equal proportions the review of alerts by creators and platform owners (in both instances the average answers were  $\approx 7$  pre-deliberations and 7.1 post-deliberations). Even though creators and platform owners received equal amounts of support for the review of alerts prior to deliberations, it is notable that support for the review of alerts by platform owners grew much more after deliberations for public spaces with repeated bullying and/ or harassment. This means that while participants supported the review of alerts by platform owners for members-only spaces with repeated bullying and harassment, they supported it less than for public spaces with repeated bullying and harassment.

	Before deliberations	After deliberations
Creators should review alerts from automated speech detection in their members-only spaces but only when bullying and harassment has occurred there repeatedly	6.93	7.13
Oppose	16.70%	15.90%
In the middle	10.30%	9.80%
Support	66.30%	68.90%
DK/NA	6.80%	5.50%
Platform owners should review alerts from automated speech detection in members-only spaces, but only when bullying and harassment has occurred there repeatedly.	6.97	7.12
Oppose	16.50%	16.20%

In the middle	10.40%	9.70%
Support	67.40%	71.00%
DK/NA	5.70%	3.10%

In conclusion, the participants' preferences for who should review automated speech detection alerts were consistent with their preferences for who should be responsible for addressing bullying and harassment in social VR spaces overall. In public spaces and spaces with repeated instances of bullying and/ or harassment, participants preferred for platform owners to be responsible and review speech detection alerts, whereas for members-only spaces they had a slight preference for creators to be responsible and review alerts. However, support was high for both platform owners and creators, with average answers over 7 on the 0 to 10 scale.

		<b>Creators</b>	<b>Platform owners</b>
<b>Overall</b>	Before deliberations	7.1	7.2
	After deliberations	7.2	7.3
	Average	7.2	7.3
<b>Public spaces</b>	Before deliberations	7.3	7.5
	After deliberations	7.4	7.7
	Average	7.3	7.6
<b>Members-only spaces</b>	Before deliberations	7.1	7.2
	After deliberations	7.3	7.2
	Average	7.2	7.2
<b>Spaces with repeated bullying and harassment</b>	Before deliberations	7	7
	After deliberations	7.1	7.3
	Average	7.1	7.2

## C. Conclusion - who should review video capture and automatic speech detection alerts?

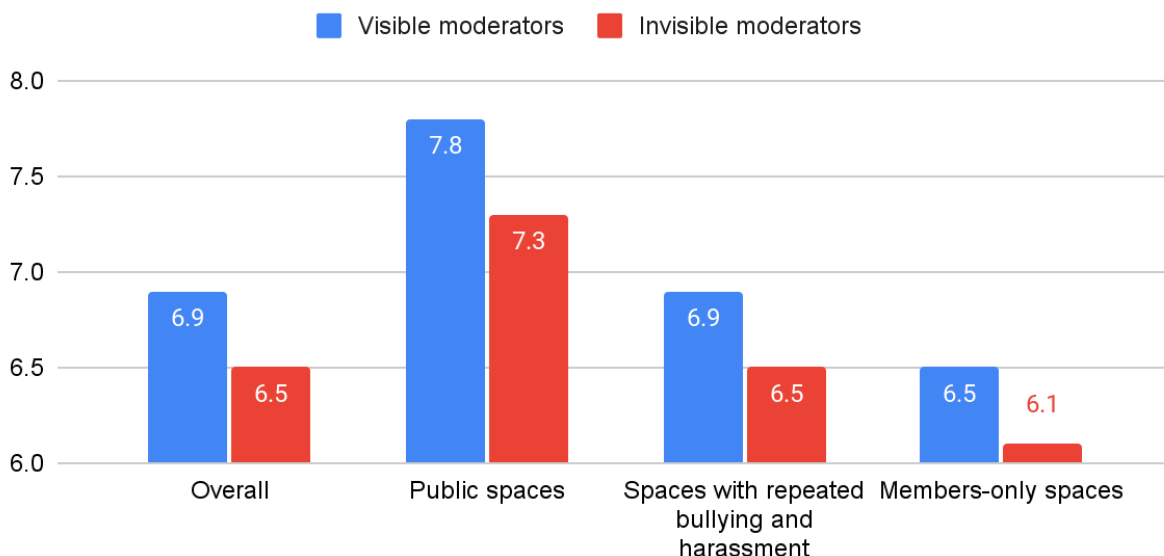
Overall, the results suggest that participants expressed slightly higher levels of support for platform owners to review both video capture and automatic speech detection alerts, compared to creators. This preference was particularly strong in public spaces and spaces with repeated instances of bullying and

harassment, where participants had a marked preference for platform owners over creators for reviewing both types of alerts. In members-only spaces, participants had a slight preference for platform owners over creators for the review of video capture, but for the review of automatic speech detection alerts they preferred creators over platform owners. The findings suggest that participants' preferences for who should review alerts from video capture and automatic speech detection are for the most part consistent with their overall preferences for who should be responsible for addressing bullying and harassment in social VR spaces.

## Should moderators be used to monitor bullying and harassment in social VR spaces?

Participants expressed greater approval for the use of visible moderators in social VR spaces compared to invisible ones. Following deliberations, the level of approval for both types of moderators increased slightly, with visible moderators receiving an average score of 6.9 (up from 6.8) and invisible moderators receiving an average score of 6.5 (up from 6.4). Despite an increase in support for invisible moderators during deliberations, rising from 59% to 62%, there was also a slight uptick in opposition from 22% to 23%. This change came from participants who prior to deliberations had not expressed their preferences for the use of invisible moderators in social VR spaces. On the contrary, opposition to the use of visible moderators slightly decreased following deliberations, going from 18% pre-deliberations to 17%. However, most of the proposals for moderators received average levels of support (6 on the 0 to 10 scale).

## Should moderators be used?



Results for participants after deliberations

A. Should moderators be used in public spaces?

The only type of space where participants expressed strong support for the use of moderators, whether visible or invisible, was public spaces. The average approval rating for visible moderators was 7.5 before deliberations, which increased to 7.8 after deliberations. For invisible moderators, the average approval rating was 7 before deliberations and 7.3 after deliberations. While support for both types of moderators increased during deliberations, more participants supported visible moderators (72.9% pre-deliberations and 79.1% post-deliberations) than invisible moderators (65.5% pre-deliberations and 71% post-deliberations). Although both visible and invisible moderators were supported by a majority of participants, almost twice as many participants opposed the use of invisible moderators (16.8%) than visible moderators (8.5%) for public spaces.

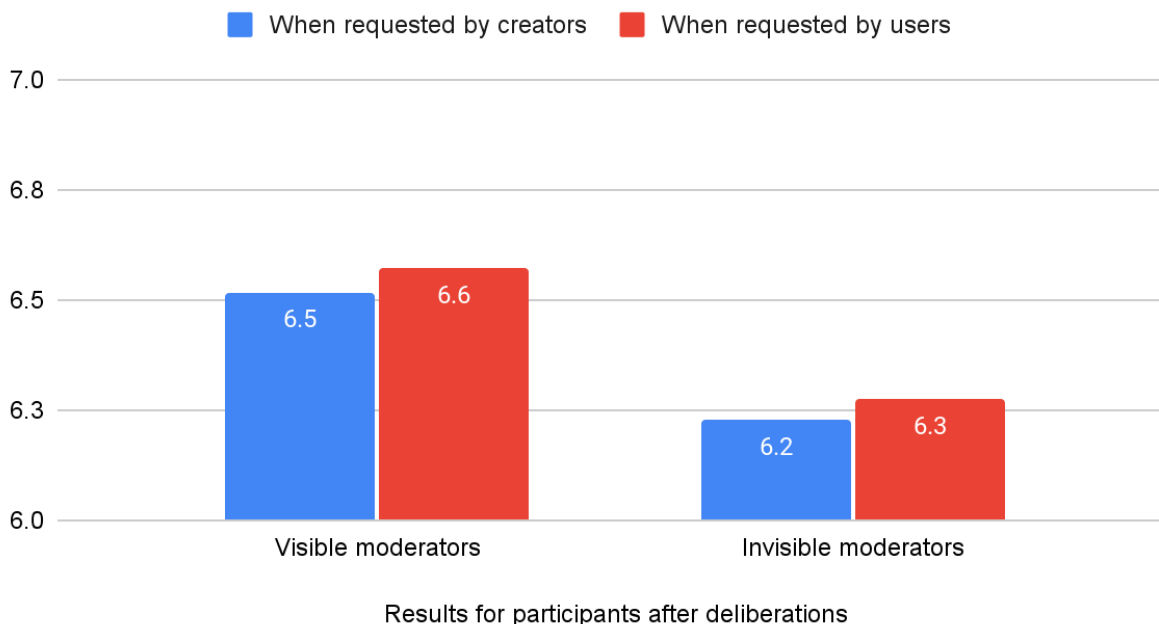
	Before deliberations	After deliberations
Visible moderators should be able to enter public spaces, even if they're built by Creators.	7.5	7.8
Oppose	11.20%	8.50%
In the middle	9.90%	9.80%
Support	72.90%	79.10%
DK/NA	6.00%	2.60%
Invisible moderators should be able to enter all public spaces.	7.0	7.3
Oppose	17.90%	16.80%
In the middle	10.20%	9.70%
Support	65.50%	71.00%
DK/NA	6.40%	2.60%

## B. When should moderators enter public spaces?

In addition to asking in which spaces should moderators be used, and which type of moderators (visible or invisible) should be used, participants were also asked under what conditions should moderators be allowed to enter public spaces, whether when asked by users or by creators of those spaces.

		When requested by creators	When requested by users
<b>Visible moderators</b>	Before deliberations	6.4	6.2
	After deliberations	6.5	6.6
<b>Invisible moderators</b>	Before deliberations	6.23	6.25
	After deliberations	6.23	6.27

## When should moderators enter public spaces?



Participants expressed moderate approval for visible moderators to enter spaces, whether at the request of users (6.2 pre-deliberations and 6.6 post-deliberations) or creators (6.4 pre-deliberations and 6.5 post-deliberations). Over the course of deliberations, participants went from preferring for moderators to enter public spaces when requested by creators to preferring for moderators to enter public spaces when requested by users. The increase in support for moderators being requested by users was greater (0.4) than for moderators being requested by creators (0.1). That is because while support increased for both over the course of deliberations (from 58.4% to 62.9% for requests by creators and from 55.7% to 63.5% for requests by users), opposition went down for requests by users (from 24.8% to 21.1%) while it went up for requests by creators (21.7% to 22.1%). Despite the slight preference for requests by users, both options received lukewarm support from participants and had slightly more than a fifth of participants opposing them.

	Before deliberations	After deliberations
Visible moderators should only enter public spaces when requested by Creators.	6.4	6.5
Oppose	21.70%	22.10%
In the middle	12.80%	11.60%
Support	58.40%	62.90%
DK/NA	7.20%	3.40%

Visible moderators should only enter public spaces when requested by users.	6.2	6.6
Oppose	24.80%	21.10%
In the middle	13.70%	12.20%
Support	55.70%	63.50%
DK/NA	5.80%	3.20%

Participants showed slightly less support for invisible moderators to enter public spaces, whether requested by creators or users, compared to visible moderators. The levels of support for invisible moderators to enter spaces when requested by creators or by users were almost identical. Overall the levels of support were moderate, with less support and more opposition than for visible moderators.

Interestingly, the average answers for when invisible moderators should enter public spaces remained the same before and after deliberations. On the contrary, support for visible moderators to enter spaces increased over the course of deliberations, as discussed in the previous paragraph. That is because for invisible moderators, while support grew a little bit over deliberations, opposition also grew to the same extent. Therefore, the average score did not change.

	Before deliberations	After deliberations
Invisible moderators should only enter public spaces when requested by Creators.	6.23	6.23
Oppose	23.80%	25.00%
In the middle	12.50%	13.10%
Support	56.70%	58.30%
DK/NA	7.00%	3.60%
Invisible moderators should only enter public spaces when requested by users	6.25	6.27
Oppose	23.10%	23.90%
In the middle	12.10%	12.90%
Support	57.60%	59.40%
DK/NA	7.20%	3.80%

C. Should moderators be used in spaces where there is repeated bullying and/ or harassment?

After public spaces, spaces with repeated instances of bullying and harassment (whether public or members-only), received the most approval for the use of moderators, both visible and invisible. As with



public spaces, participants showed a greater level of approval for visible moderators (6.7 pre-deliberations and 6.9 post-deliberations) compared to invisible moderators (6.4 pre-deliberations and 6.5 post-deliberations). However, the difference in approval between visible and invisible moderators was less significant for spaces with repeated bullying and harassment (0.3 pre-deliberations and 0.4 post-deliberations) than for public spaces (0.5 both before and after deliberations). This indicates that while more participants preferred the use of visible moderators over invisible moderators in both public spaces and spaces with repeated bullying and harassment, the preference for visible moderators was less pronounced in spaces with repeated bullying and harassment.

It is also important to note that while the levels of support for visible and invisible moderators were similar for both public and members-only spaces where there is repeated bullying and /or harassment, it was slightly higher for public spaces, than for members-only spaces. Overall levels of support remained moderate, no matter the type of moderator or the type of space (between 6 and 7).

Spaces where there is repeated bullying and/ or harassment		Visible moderators	Invisible moderators
Public spaces	Before deliberations	6.7	6.5
	After deliberations	7.0	6.6
Members-only spaces	Before deliberations	6.7	6.4
	After deliberations	6.9	6.5
Public + members only spaces	Before deliberations	6.7	6.4
	After deliberations	6.9	6.5

## D. Should moderators be used in members-only spaces?

In comparison to public spaces and spaces with repeated incidents of bullying and harassment, members-only spaces received the least amount of approval for the use of moderators, both visible and invisible. Although a majority of participants supported the use of moderators in these spaces, the average approval ratings were lower than for any other type of space. Prior to deliberations, the average approval rating was 6.8 for visible moderators and 6.4 for invisible moderators. These levels of support were already lower than for any other type of space, and after deliberations, they decreased even further, which was the only type of space to experience a decline in approval as a result of deliberations. Post-deliberations, the average approval rating for visible moderators was 6.5 (-0.3 from pre-deliberations), and for invisible moderators, it was 6.1 (-0.3 too).

These results suggest that although a majority of participants approved of the use of moderators in members-only spaces, significantly fewer participants supported this measure for members-only spaces in comparison to other types of spaces. For example, 79.1% of participants supported the use of visible moderators in public spaces post-deliberations, compared to 61.7% for members-only spaces, i.e. 17.4% fewer participants. The amount of participants opposing the use of moderators in members-only spaces

were also much greater (21.2% for visible moderators and 27.9% for invisible moderators) than for public spaces (8.5% for visible moderators and 16.8% for invisible moderators). As such, participants were less enthusiastic and more skeptical about the use of moderators in members-only spaces than in public spaces, and to a lesser extent in spaces with repeated bullying and/ or harassment.

	Before deliberations	After deliberations
Visible moderators should be able to enter members-only spaces.	6.8	6.5
Oppose	17.50%	21.20%
In the middle	12.00%	13.90%
Support	63.40%	61.70%
DK/NA	7.00%	3.10%
Invisible moderators should be able to enter all members-only spaces.	6.4	6.1
Oppose	23.10%	27.90%
In the middle	11.70%	11.70%
Support	58.20%	57.40%
DK/NA	7.00%	3.00%

E. Should creators be able to select other users to be visible moderators in their spaces?

Finally, participants were asked whether creators should have the option to select other users to be visible moderators in their spaces. This proposal was supported by participants, with an average rating of 7.2 before deliberations and 7.3 after. The percentage of participants opposing the proposal did not change before and after deliberations, remaining at 12%, while support increased by ≈3%, from ≈68% to ≈71%. The majority of the increase in support was due to participants who did not express their preferences prior to deliberations. These results indicate that most participants favored the idea of creators being able to select visible moderators in their spaces, and there was a slight increase in support after deliberations.

	Before deliberations	After deliberations
Creators should be able to select other users to be visible moderators in their spaces.	7.2	7.3
Oppose	12.00%	12.00%
In the middle	11.70%	11.90%
Support	67.80%	71.10%

## COMMUNITY FORUM - RESULTS ANALYSIS

DK/NA	8.40%	5.10%
-------	-------	-------

## Regional Analysis

### Key Takeaways

For public spaces, most regions favored reporting to platform owners rather than to creators, except for South-East Asia which preferred reporting to creators. The preference of who to report to in member-only spaces was more varied. Oceania, Sub-Saharan Africa, South-East Asia, and North America preferred reporting to creators, while CEE, Western Europe, East Asia, and Latin America preferred reporting to platform owners. MENA had no preference and supported reporting to both in equal measures. The proposal to require creators to take a course was highly supported as an action for platform owners to take action against creators. The proposal gained support in all regions during the deliberations, raising the global average from 7.3 pre-deliberations to 8.1 post-deliberations. After deliberations, the average response for this proposal was around or above 8 for all regions.

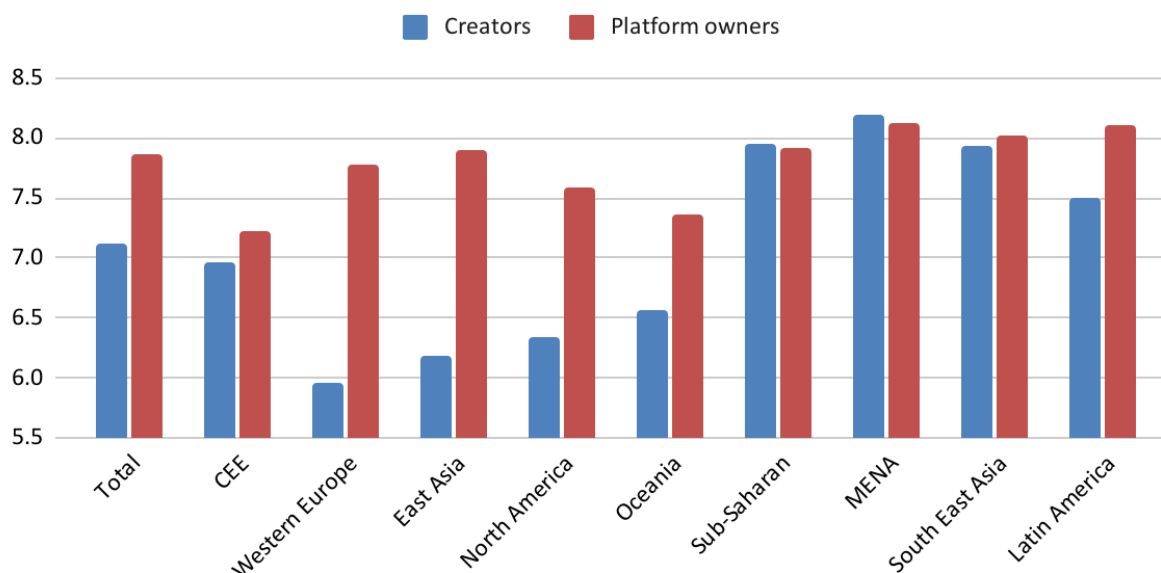
### Who should be responsible for monitoring social VR spaces?

#### Public spaces

Most regions of the world expressed a preference for public spaces to be monitored by platform owners rather than by creators. That preference was most marked after deliberations for Western Europe (7.8 for platform owners vs 6 for creators), East Asia (7.9 for platform owners vs 6.2 for creators), North America (7.6 for platform owners vs 6.3 for creators), and Oceania (7.4 for platform owners vs 6.6 for creators). CEE, Southeast Asia, and Latin America also expressed greater approval for the monitoring of public spaces by platform owners over creators, but to a lesser extent than the regions previously listed. East Asia is the region for which deliberations deepened the participants' preference for public spaces to be monitored by platform owners over creators the most, going from 6.8 for creators and 7.2 for platform owners before deliberations, to 6.2 for creators and 7.9 for platform owners after deliberations.

On the other hand, Sub-Saharan Africa supported for public spaces to be monitored by platform owners as much as creators (7.9 for both after deliberations). Finally, MENA expressed slightly more support for creators to monitor public spaces rather than platform owners (8.2 vs 8.1 respectively after deliberations).

## Who should be responsible for monitoring public spaces?



Participants' average answers by region after deliberations

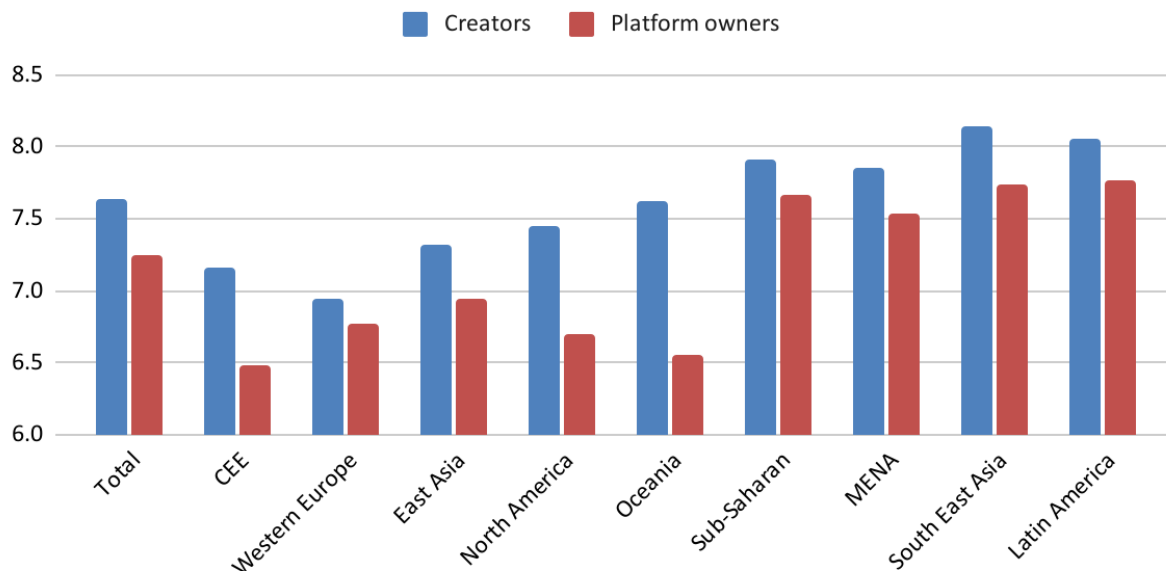
		Total	CEE	Western Europe	East Asia	North America	Oceania	Sub-Saharan Africa	MENA	Southeast Asia	Latin America
Creators	Before	7.2	6.4	6.3	6.8	6.6	6.7	8.0	8.0	7.8	7.9
	After	7.1	7.0	6.0	6.2	6.3	6.6	7.95	8.2	7.9	7.5
Platform owners	Before	7.6	6.4	7.6	7.2	7.5	7.3	7.7	7.7	7.8	8.0
	After	7.9	7.2	7.8	7.9	7.6	7.4	7.93	8.1	8.0	8.1

### Members-only spaces

Before deliberations, three regions, CEE, East Asia, and North America, had no preference between members-only spaces being monitored by creators or platform owners. Western Europe is the only region that had a marked preference before deliberations for members-only spaces to be monitored by platform owners (7.3) over creators (6.5). The five other regions, Oceania, Sub-Saharan Africa, MENA, South-East Asia, and Latin America all expressed before deliberations a moderate preference for creators to monitor members-only spaces over platform owners.

After deliberations however, all regions preferred for creators to monitor members-only spaces. As such, the deliberations led participants to prefer members-only spaces to be monitored by creators rather than platform owners, even in regions where originally the opposite was true.

## Who should be responsible for monitoring members-only spaces?



Participants' average answers by region after deliberations

		Total	CEE	Western Europe	East Asia	North America	Oceania	Sub-Saharan Africa	MENA	Southeast Asia	Latin America
Creators	Before	7.3	6.2	6.5	7.0	7.1	7.3	7.5	7.4	7.8	8.1
	After	7.6	7.2	6.95	7.3	7.5	7.6	7.9	7.9	8.1	8.1
Platform owners	Before	7.3	6.2	7.3	7.0	7.1	7.0	7.4	7.3	7.6	8.0
	After	7.2	6.5	6.77	7.0	6.7	6.6	7.7	7.5	7.7	7.8

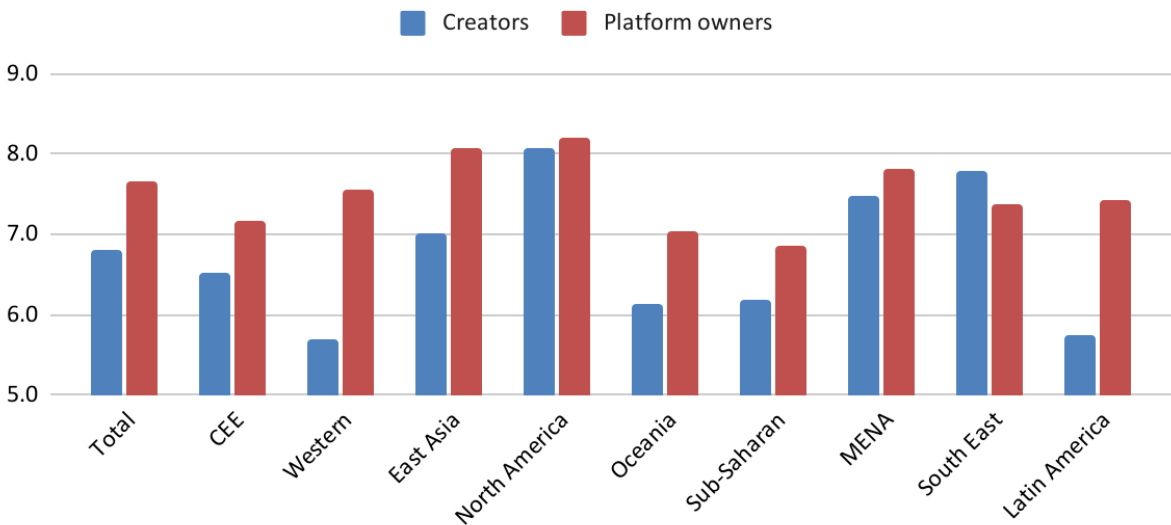
## To whom should cases of bullying and harassment be reported?

Public spaces

For public spaces, after deliberations all regions favored for reports to be made to platform owners more than to creators, except for South-East Asia, which preferred reports to be made to creators (7.7 pre-deliberations and 7.8 post-deliberations) rather than to platform owners (7.3 pre-deliberations and 7.4 post-deliberations). The range of approval for reports to be made to platform owners ranged after deliberations from 6.9 for Sub-Saharan Africa to 8.1 for East Asia. On the other hand, the range of approval for reports to be made to creators was much larger, going from 5.7 for Western Europe up to 8.1 for North America.

Deliberations made the majority of regions more in favor of reports in public spaces being made to platform owners rather than to creators. The biggest change in favor of reports being made to platform owners rather than creators was recorded in Western Europe.

To whom should cases of bullying and harassment in public spaces be reported?



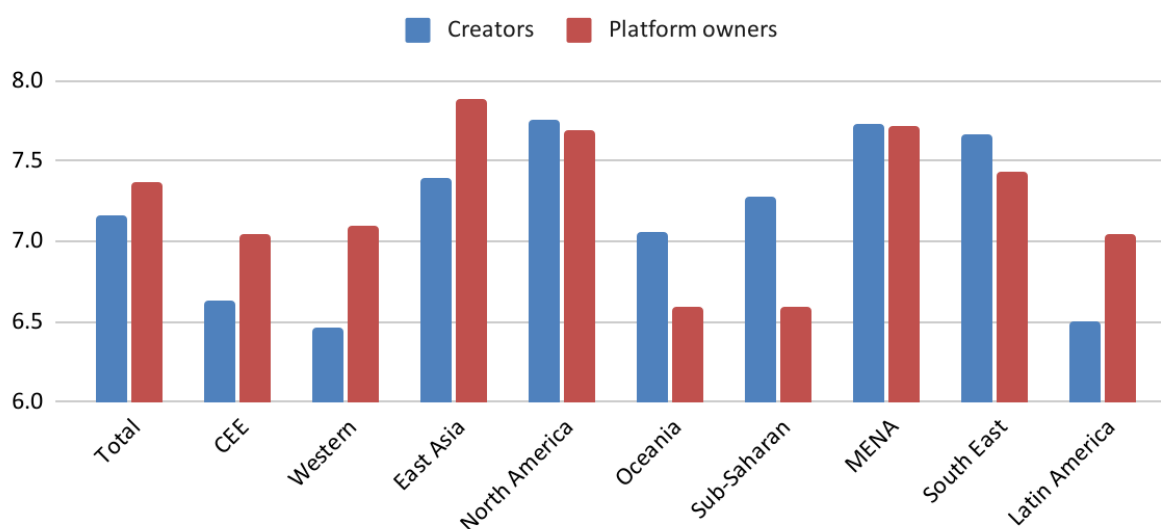
Participants' average answers by region after deliberations

		Total	CEE	Western Europe	East Asia	North America	Oceania	Sub-Saharan Africa	MENA	Southeast Asia	Latin America
Creators	Before	6.7	6.2	6.2	7.3	7.3	5.9	6.0	7.4	7.7	5.9
	After	6.8	6.5	5.7	7.0	8.1	6.1	6.2	7.5	7.8	5.7
Platform owners	Before	7.2	6.2	6.9	7.8	7.5	6.9	7.0	7.5	7.3	7.1
	After	7.7	7.2	7.5	8.1	8.2	7.1	6.9	7.8	7.4	7.4

## Members-only spaces

Whether reports in member-only spaces should be made to creators or platform owners was more contentious. Several regions expressed a preference for reports to be made to creators rather than platform owners, notably Oceania and Sub-Saharan Africa, but also South-East Asia and North America to lesser extents. On the contrary, CEE, Western Europe, East Asia, and Latin America had a preference for reports from members-only spaces to be made to platform owners. MENA was the only region to express no preference one way or another, supporting reports being made to creators and platform owners equally.

## To whom should cases of bullying and harassment in members-only spaces be reported?



Participants' average answers by region after deliberations

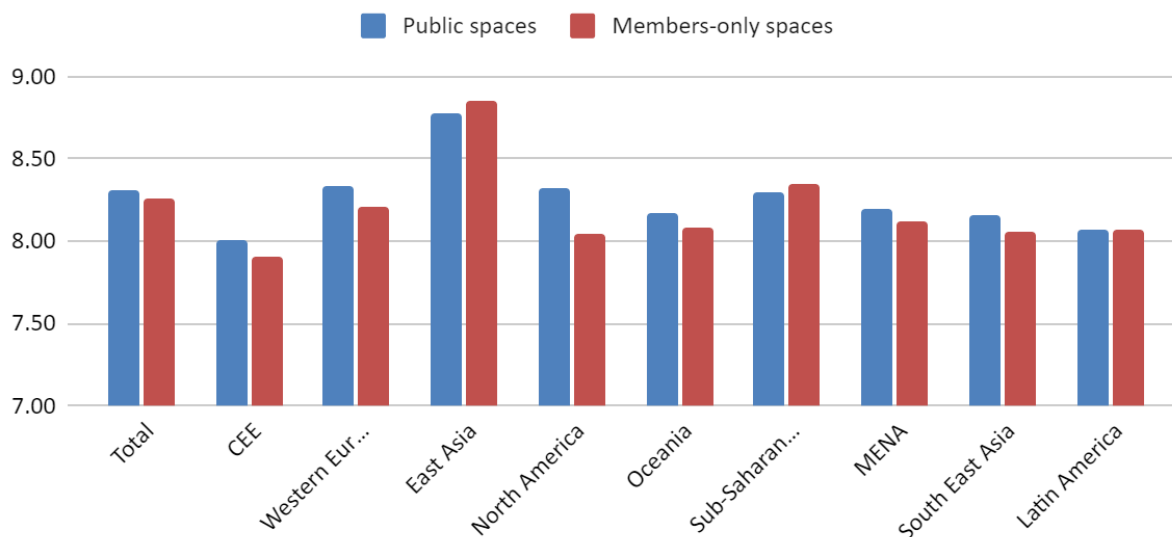
		Total	CEE	Western Europe	East Asia	North America	Oceania	Sub-Saharan Africa	MENA	Southeast Asia	Latin America
Creators	Before	6.9	6.3	6.4	7.4	7.2	6.5	6.7	7.5	7.3	6.4
	After	7.2	6.6	6.46	7.4	7.8	7.1	7.3	7.7	7.7	6.5
Platform owners	Before	7.2	6.4	6.8	7.7	7.2	6.9	6.9	7.4	7.3	7.0
	After	7.4	7.1	7.10	7.9	7.7	6.6	6.6	7.7	7.4	7.0



## Severe cases of bullying and/or harassment

In cases of severe bullying and/or harassment, regions supported for severe cases to be sent to platform owners almost equally in public spaces as in members-only spaces. The only exception was North America, where public spaces received more support than members-only spaces. Participants from East Asia expressed much more enthusiasm for these proposals than those of other regional, with approval reaching 8.8 for creators after deliberations, and 8.9 for platform owners.

## Should creators send severe cases to platform owners for additional enforcement?



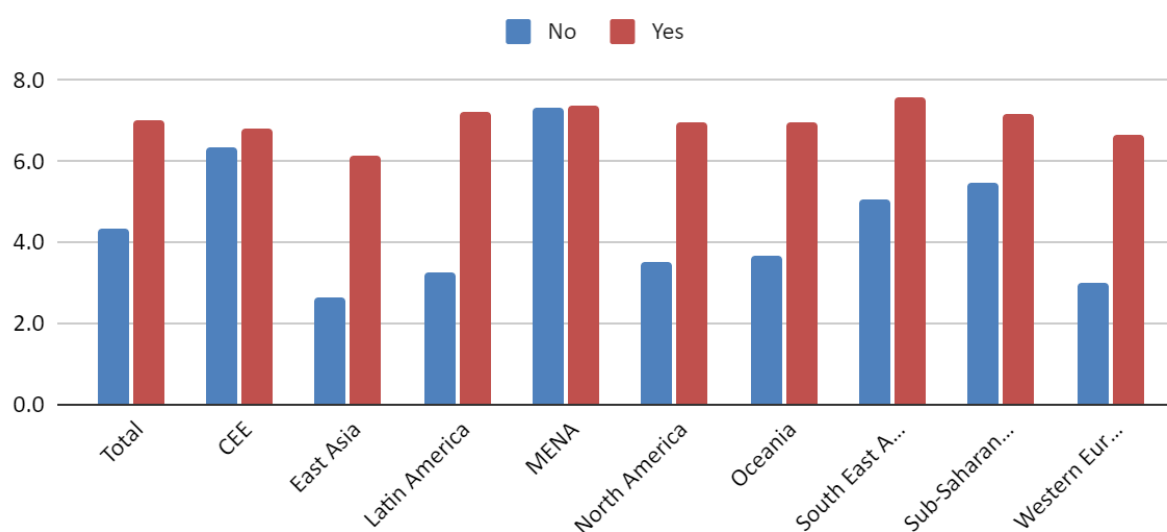
Participants' average answers by region after deliberations

		Total	CEE	Western Europe	East Asia	North America	Oceania	Sub-Saharan Africa	MENA	Southeast Asia	Latin America
Public spaces	Before	7.74	6.6	7.5	8.4	7.7	7.7	8.0	7.6	7.9	7.8
	After	8.31	8.0	8.33	8.8	8.3	8.2	8.29	8.2	8.2	8.1
Members-only spaces	Before	7.77	6.6	7.4	8.5	7.9	7.7	7.9	7.8	7.8	7.6
	After	8.26	7.9	8.21	8.9	8.0	8.1	8.35	8.1	8.1	8.1

## Should creators be made aware of reports made to platform owners?

Participants were asked, in instances when reports of bullying and harassment are made to platform owners, should those reports be shared with the creators of the spaces concerned. Most regions were against these reports not being shared with creators. Two regions, South-East Asia and Sub-Saharan Africa were neither for nor against (5.1 and 5.5 respectively after deliberations). Only two regions expressed support for reports to not be shared with creators, CEE (6.3 after deliberations) and MENA (7.3 after deliberations). However, even in those regions where the proposals to not share the reports with creators were supported, the proposals to share the reports had more support than that to not share the reports. As such, in all regions the proposal to share reports with creators was higher than that to not share the reports.

## Should creators be made aware of reports made to platform owners?



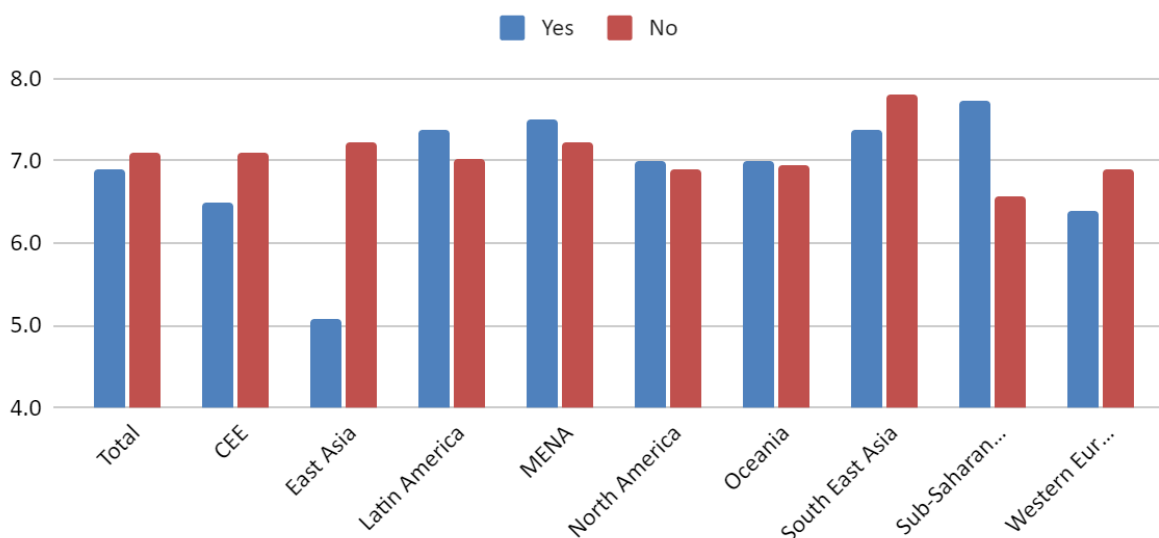
Participants' average answers by region after deliberations

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	4.4	5.6	3.0	3.6	6.6	3.8	3.5	5.3	5.7	3.4
After	4.4	6.3	2.6	3.3	7.3	3.5	3.7	5.1	5.5	3.0

Then participants were asked whether the reports that are shared with creators should include personal information about the users involved and more, or not. Regions were divided on this question. After deliberations, CEE, Western Europe, Southeast Asia, and East Asia all preferred for personal information to not be shared with creators. On the other hand, Latin America, MENA, and Sub-Saharan Africa

preferred personal information to be shared with creators. North America and Oceania also expressed a preference for personal information to be shared with creators, but to a lesser degree than previous regions listed.

In the reports shared with creators, should personal information be included (such as users involved etc)?



Participants' average answers by region after deliberations

		Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Yes	Before	7.0	6.6	5.7	7.6	7.0	7.1	7.2	7.4	7.4	6.6
	After	6.9	6.5	5.1	7.4	7.5	7.0	7.0	7.4	7.7	6.4
No	Before	6.6	6.0	6.7	6.6	6.6	6.3	6.8	7.3	6.9	6.6
	After	7.1	7.1	7.2	7.0	7.2	6.9	7.0	7.8	6.6	6.9

**What should be done about spaces where there is repeated bullying and/or harassment?**

Visibility

*Members-only spaces where there is repeated bullying and/or harassment should be made less visible to users*

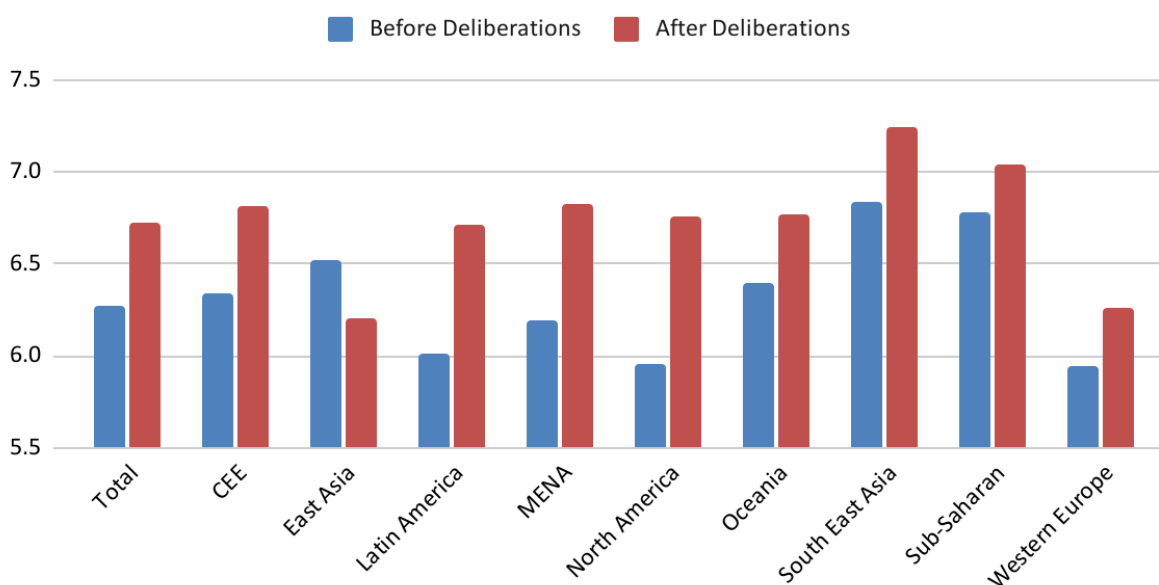
Regions expressed mild support for making private spaces less visible to users, with the lowest answer pre-deliberations being 5.9 from Western Europe and the highest being 6.8 from both Sub-Saharan Africa and Southeast Asia. Post deliberations, the lowest answer rose to 6.2 in East Asia and the highest to 7.2 in Southeast Asia.

Overall, Southeast Asia was the region the most supportive of this proposal, followed by Sub-Saharan Africa. These two regions were the only ones where support for this proposal reached the 7 point mark after deliberations. All other regions expressed only moderate support for the proposal, even after deliberations. On the other side, East Asia and Western Europe were the two regions which expressed the least amount of support for this proposal.

Finally, support for this proposal grew the most over the course of deliberations in North America, followed by Latin America. In most other regions, support also grew over deliberations, but to a lesser extent, except for East Asia, where support for the proposal decreased as a result of deliberations.

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	6.3	6.3	6.5	6.0	6.2	6.0	6.4	6.8	6.8	5.9
After	6.7	6.8	6.2	6.7	6.8	6.8	6.8	7.2	7.0	6.3

Should members-only spaces where there is repeated bullying and/or harassment be made less visible to users?



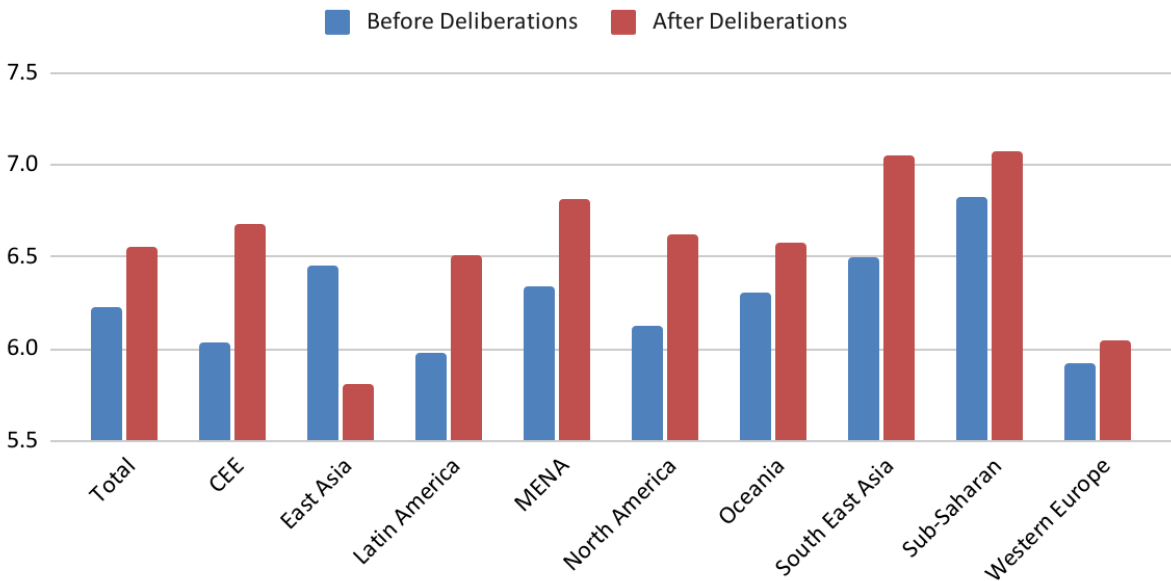
*Members-only spaces where there is repeated bullying and/ or harassment should not be publicly discoverable*

There was mild support for members-only spaces not being publicly discoverable. Just like for the previous proposal, only Southeast Asia and Sub-Saharan expressed more enthusiastic support for this proposal, going over the 7 point mark post-deliberations. East Asia and Western Europe were once again the regions the least supportive of this proposal.

Just like for the previous proposal, all countries but East Asia increased their support for the proposal over deliberations. In East Asia, support for this proposal decreased more over deliberations, than it did for the previous proposal (0.3 drop for the previous proposal vs 0.6 for this proposal).

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	6.2	6.0	6.4	6.0	6.3	6.1	6.3	6.5	6.8	5.9
After	6.6	6.7	5.8	6.5	6.8	6.6	6.6	7.1	7.1	6.1

Should members-only spaces where there is repeated bullying and/ or harassment not be publicly discoverable?



*Members-only spaces where there is repeated bullying and/ or harassment should be removed from the platform*

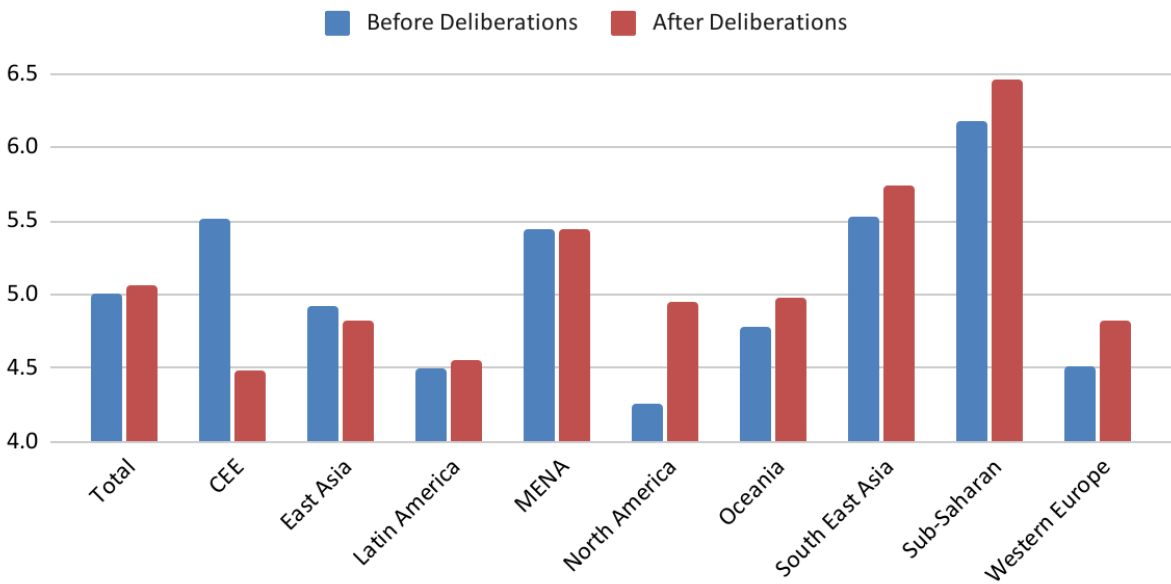
This proposal was the least favored of all three options on visibility reduction for spaces where there is repeated bullying and/ or harassment. The answers remained largely consistent pre and post

deliberations on a global scale, but two outliers emerged. CEE and North America saw a -1.0 decrease and +0.7 increase respectively post deliberations.

CEE, East Asia, Latin America, and Western Europe all had scores under or equal to 5 after deliberations, indicating that the proposal failed to gain majority support from the participants of these regions. On the other hand, Sub-Saharan Africa expressed the highest amount of approval for the removal of members-only spaces where there is repeated bullying and/ or harassment, with an average answer of 6.5 post-deliberations. Nonetheless, even the region with the highest amount of support only expressed moderate support, and less than for the two previous proposals discussed. As such, this was the least favored of all three options for reducing the visibility of spaces where there is repeated bullying and/ or harassment, with many regions not expressing majority support for this proposal.

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	5.0	5.5	4.9	4.5	5.4	4.3	4.8	5.5	6.2	4.5
After	5.1	4.5	4.8	4.6	5.4	5.0	5.0	5.7	6.5	4.8

Should members-only spaces where there is repeated bullying and/ or harassment be removed from the platform?

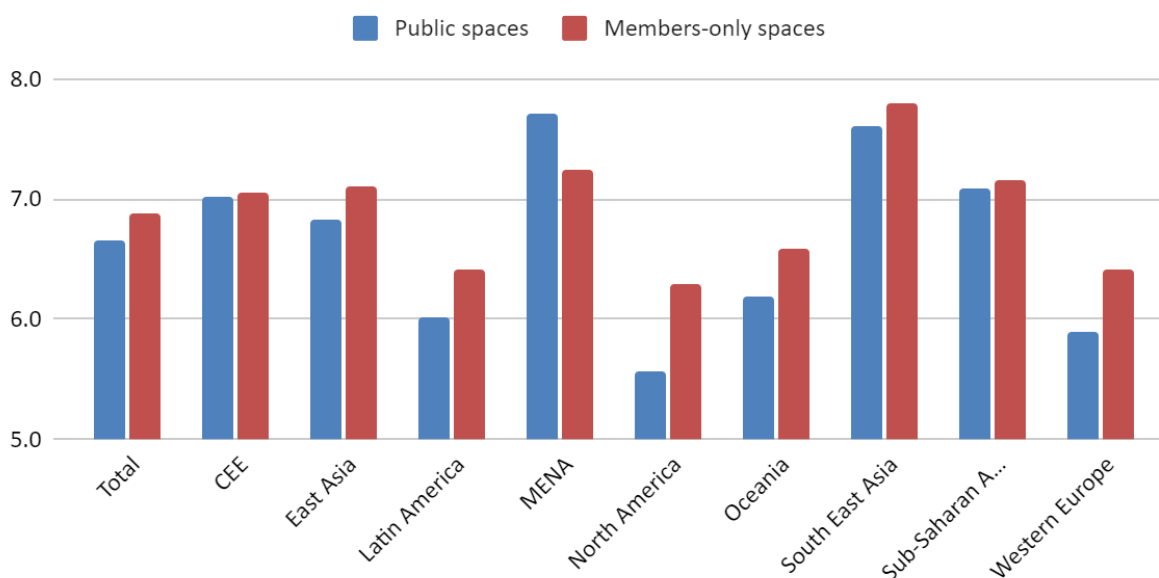


Consequences for creators

*Actions should be taken against creators of spaces where there is repeated bullying and/ or harassment*

Latin America, North America, Oceania and Western Europe expressed moderate support after deliberations for actions to be taken against creators. In CEE, East Asia and Sub-Saharan Africa support was slightly higher, around 7. MENA and South East Asia were the two regions with the highest support for this proposal. For most regions there was more support for actions to be taken against creators in members-only spaces than in public spaces, except for MENA where it was the reverse.

## Should actions be taken against creators?



Participants' average answers by region after deliberations

		Total	CEE	East Asia	Latin America	MENA	North America	Oceania	South East Asia	Sub-Saharan Africa	Western Europe
Public spaces	Before	6.4	6.3	6.8	5.8	6.9	5.4	6.2	7.3	7.3	5.8
	After	6.7	7.0	6.8	6.0	7.7	5.6	6.2	7.6	7.1	5.9
Members-only spaces	Before	6.4	6.2	6.8	5.9	6.6	5.9	6.3	7.1	6.9	6.2
	After	6.9	7.1	7.1	6.4	7.2	6.3	6.6	7.8	7.1	6.4

*Creators should be banned from making additional members-only spaces*

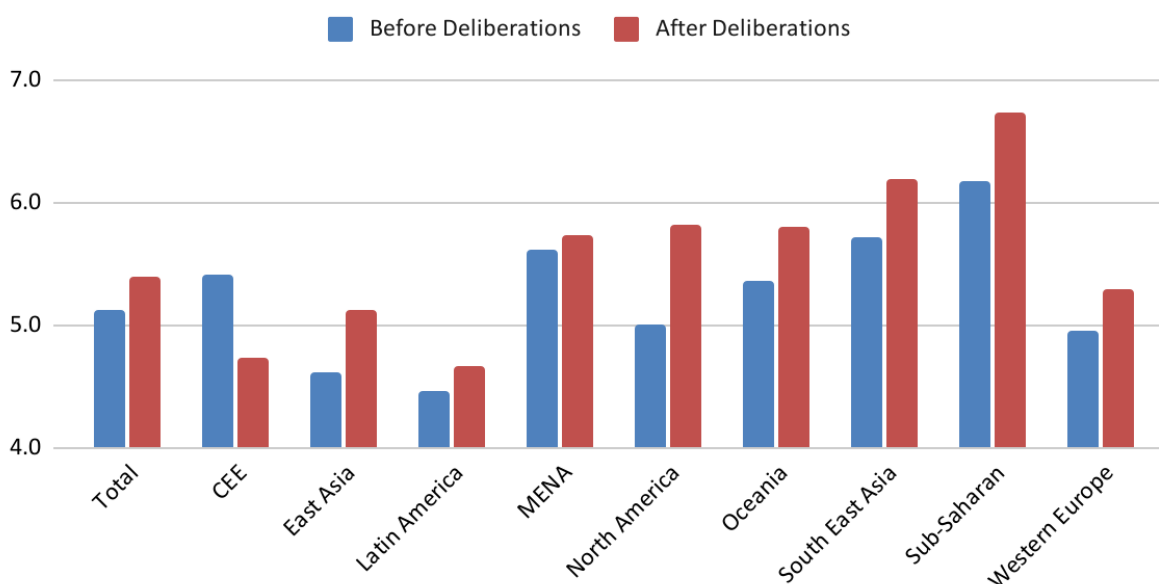
Regions expressed mild support for banning creators from making additional members-only spaces where there is repeated bullying and/or harassment. Over the course of deliberations, support for this

proposal increased in all but one region, with the greatest increases taking place in Sub-Saharan Africa, Southeast Asia, and East Asia. The region that decreased was CEE, with a -0.7 decrease.

Post-deliberations, two regions failed to pass the 5-points mark, CEE, and Latin America, meaning this proposal did not garner majority support from the participants of those regions. In the other regions, support was for the most part between 5 and 6, indicating neither support nor opposition from the participants of those regions. Only South-East Asia and Sub-Saharan Africa garnered average answers post-deliberations above 6. This means that across all regions, support was at best moderate for the proposal to ban creators from making additional members-only spaces.

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	5.1	5.4	4.6	4.5	5.6	5.0	5.4	5.7	6.2	5.0
After	5.4	4.7	5.1	4.7	5.7	5.8	5.8	6.2	6.7	5.3

Should creators of spaces with repeated bullying and harassment be banned from making additional members-only spaces?



*Creators should be banned from inviting more people to join a space where there is repeated bullying and/ or harassment*

This proposal received slightly higher levels of support than the previous one, but the overall levels of support remained low, with a global average of 5.3 post-deliberations. Latin America was the only region to not reach the 5-point mark after deliberations. East Asia, MENA, Oceania and North America all had

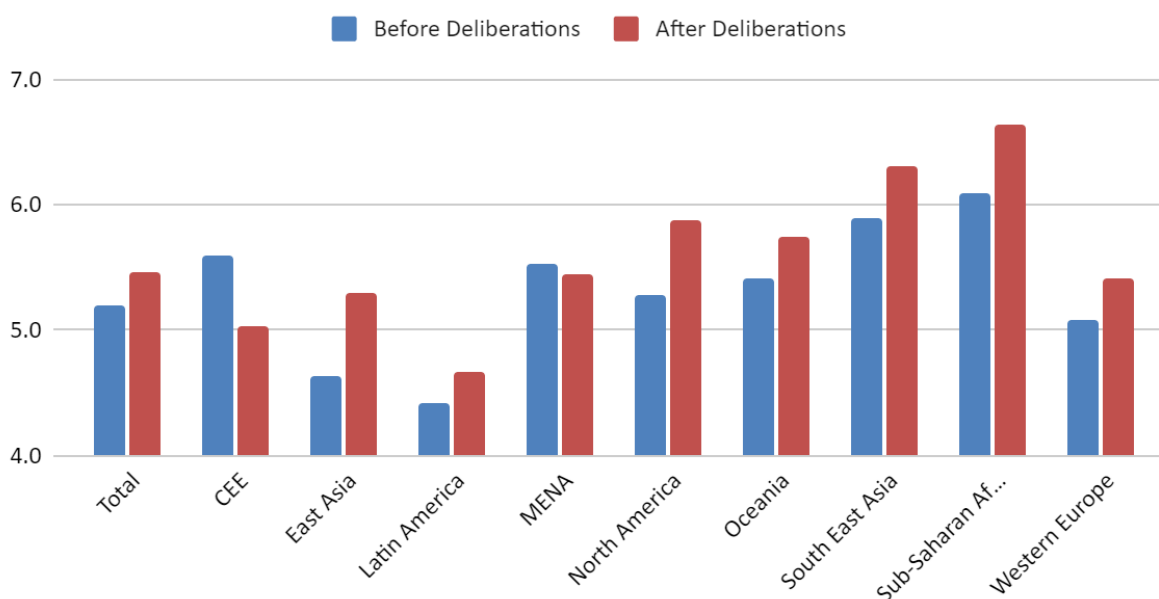


scores between 5 and 6, meaning that participants were neither for nor against the proposal. Only in Southeast Asia and Sub-Saharan Africa did the proposal get moderate support from participants, with average answers above 6.

Mild to moderate increases in support for this proposal were observed with East Asia by 0.7 and North America by 0.6. However, losses also occurred, most significantly by -0.5 in CEE.

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	5.2	5.6	4.6	4.4	5.5	5.3	5.4	5.9	6.1	5.1
After	5.5	5.0	5.3	4.7	5.5	5.9	5.7	6.3	6.6	5.4

Should creators be banned from inviting more people to join spaces where there is repeated bullying and/ or harassment?



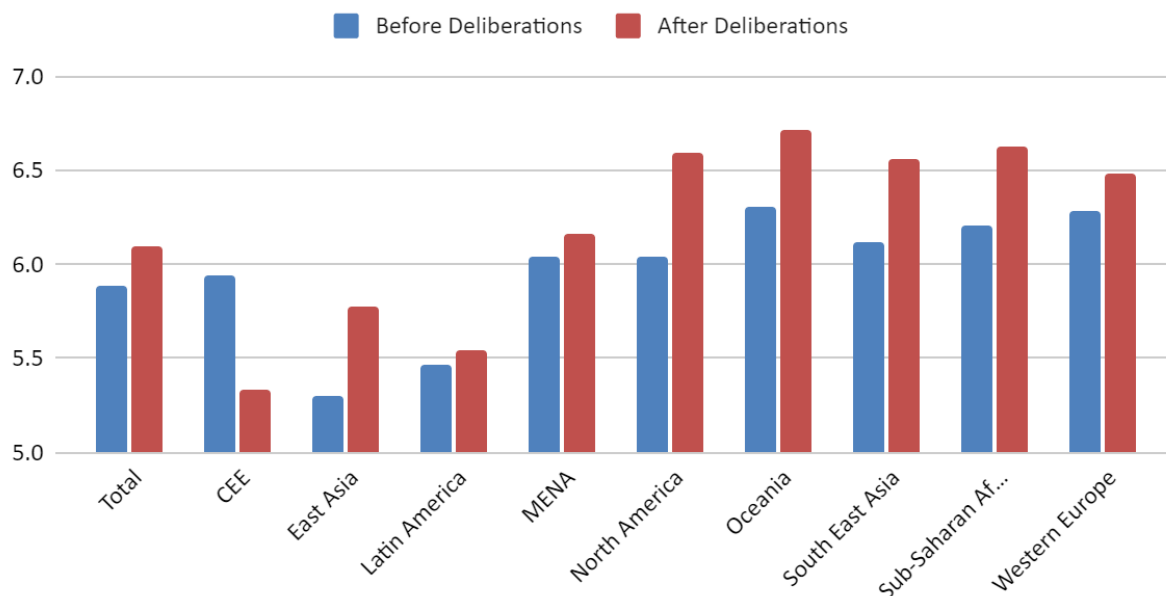
*Creators should not be able to make money off of spaces where there is repeated bullying and/ or harassment*

Regions expressed mild support for creators to not be able to make money off of spaces with repeated bullying. There were modest increases in support for this measure across all regions but CEE, the region's approval decreasing from 5.9 to 5.3 after deliberations. This proposal was more supported than the prior two measures. The lowest level of support for this proposal after deliberations were in CEE and Latin America. All the other regions had answers between 6 and 7, indicating moderate support.

## COMMUNITY FORUM - RESULTS ANALYSIS

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	5.9	5.9	5.3	5.5	6.0	6.0	6.3	6.1	6.2	6.3
After	6.1	5.3	5.8	5.5	6.2	6.6	6.7	6.6	6.6	6.5

Should creators not be able to make money off of spaces where there is repeated bullying and/ or harassment?



*Creators should take a course on how to moderate the spaces they create*

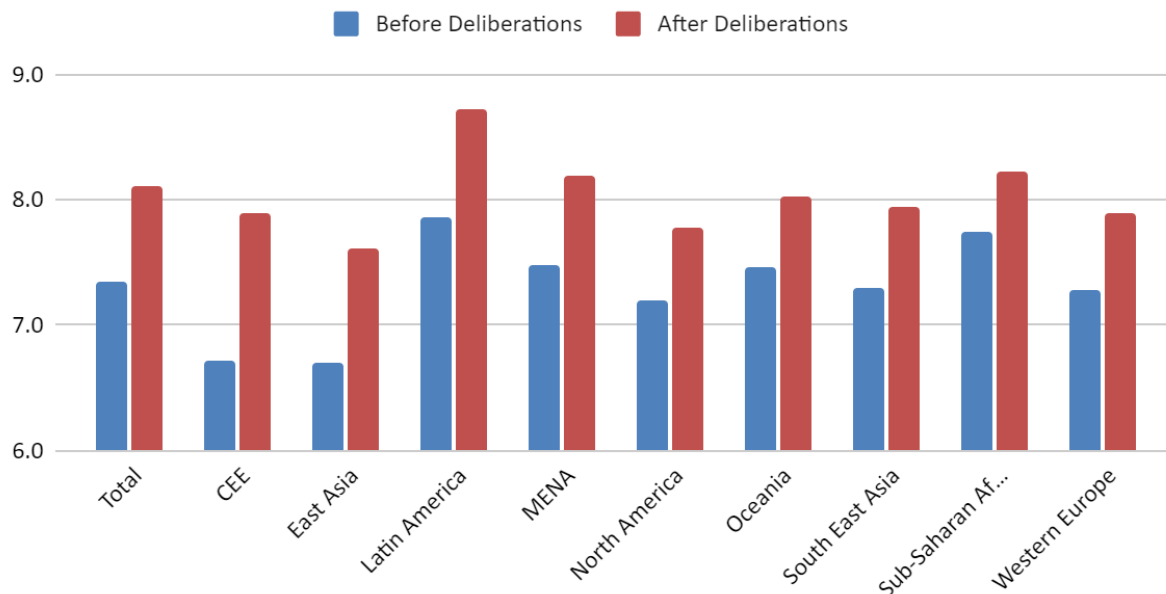
This was the most supported proposal among actions to be taken against creators. Support for this proposal increased in all regions over the course of deliberations, raising the global average from 7.3 pre-deliberations to 8.1 post-deliberations. Post-deliberations, the average answer to this proposal was close to or above 8 for all regions.

Moderate to large increases can be highlighted. CEE had a +1.2 increase pre and post deliberations, with East Asia and Latin America increasing by 0.9. Following was a 0.6 increase in four regions and the lowest increase of 0.5 hailing from Sub-Saharan Africa, which was already the second highest in its support pre deliberations at 7.7 to 8.2 post-deliberations.

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	7.3	6.7	6.7	7.9	7.5	7.2	7.5	7.3	7.7	7.3

After	8.1	7.9	7.6	8.7	8.2	7.8	8.0	7.9	8.2	7.9
-------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Should creators take a course on how to moderate the spaces they create?

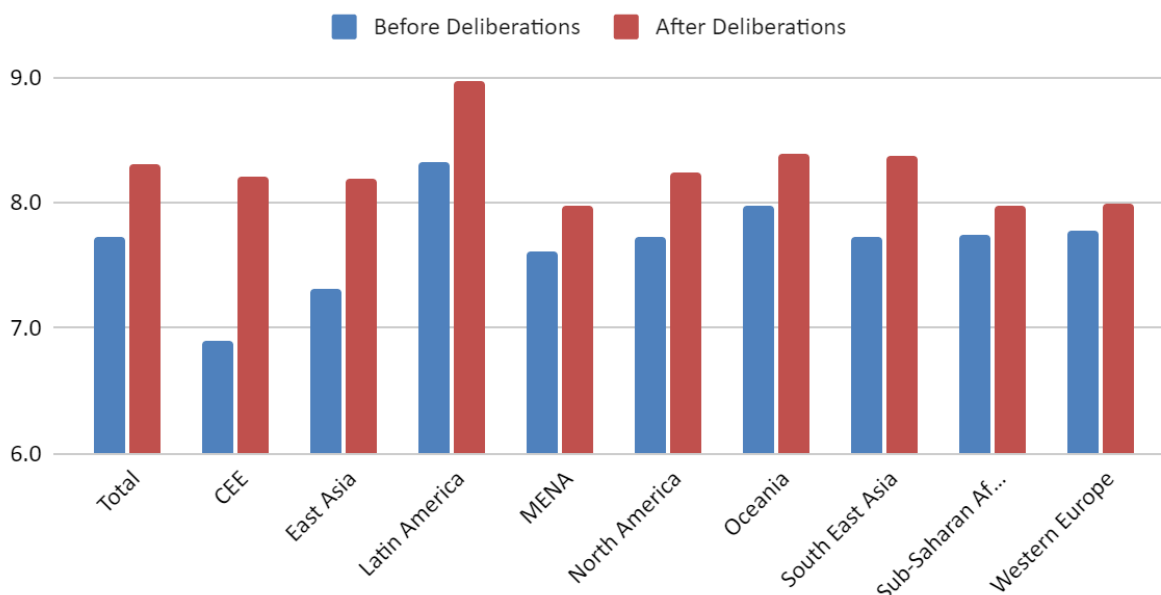


## User Notifications

After deliberations, all regions supported for users to be notified when entering a space where there is repeated bullying and/ or harassment. Support for this proposal increased for all regions over the course of deliberations. This was one of the most supported actions to be taken against spaces where there is repeated bullying and/ or harassment. CEE and East Asia expressed slightly less support for the proposal than other regions pre-deliberations, but post-deliberations they had similar levels of support to other regions.

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Before	7.7	6.9	7.3	8.3	7.6	7.7	8.0	7.7	7.7	7.8
After	8.3	8.2	8.2	9.0	8.0	8.2	8.4	8.4	8.0	8.0

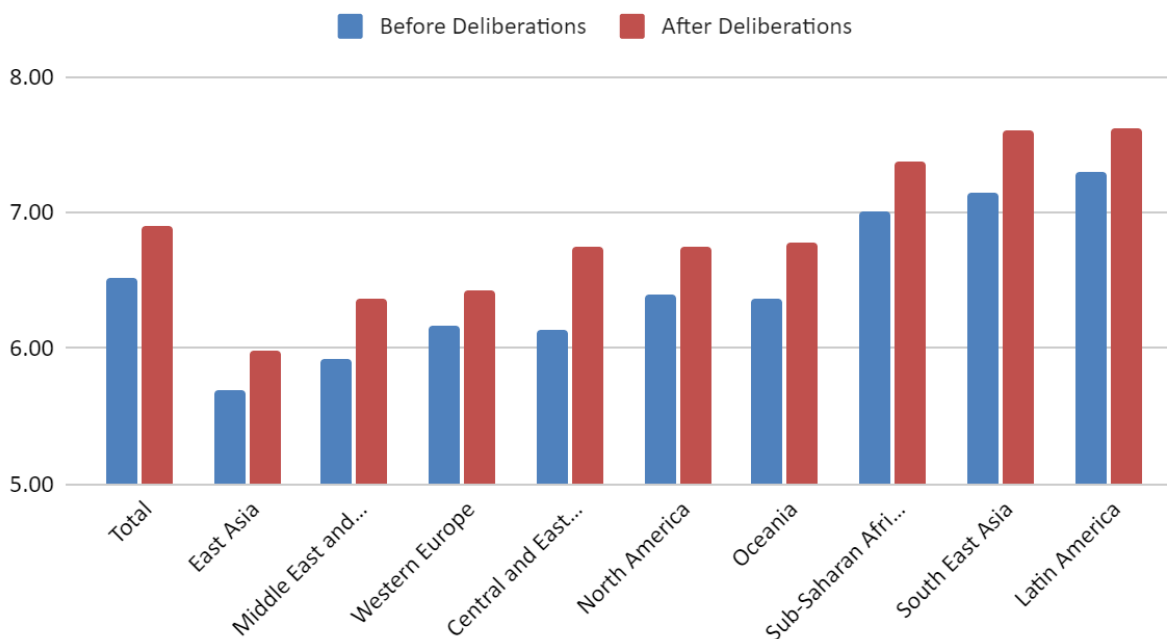
Should users be notified when entering a space where there is repeated bullying and/ or harassment?



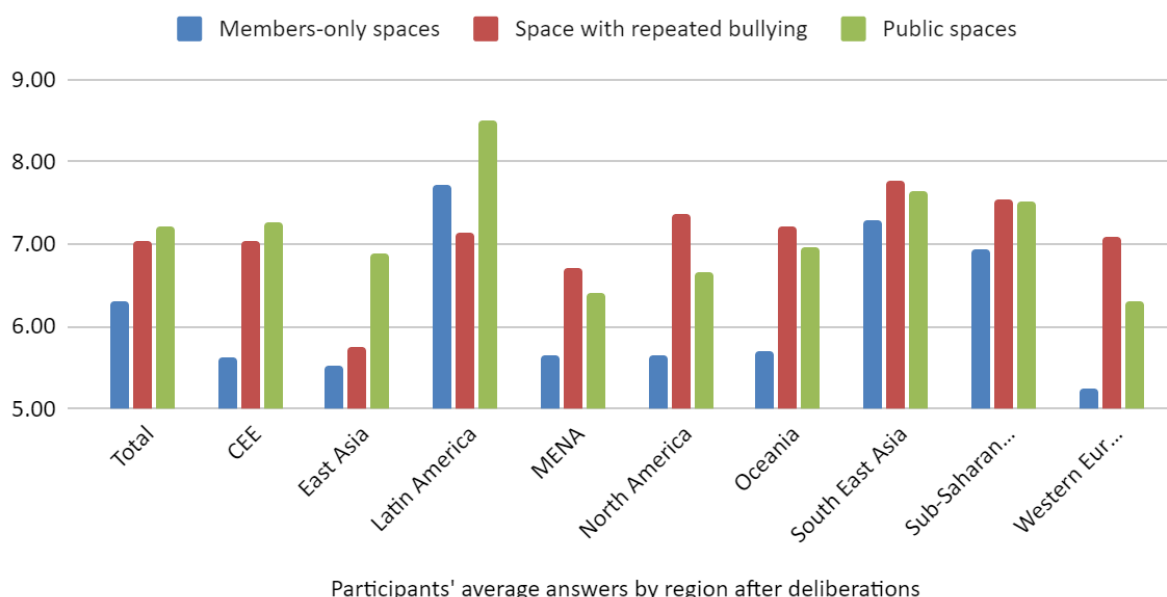
What tools should be used to identify bullying and harassment in social VR spaces?

#### Video Capture

Should video capture be used in social VR spaces?



Should video capture be used in public spaces, members-only spaces, and spaces with repeated bullying and/ or harassment?



	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Members-only spaces	6.30	5.61	5.52	7.72	5.65	5.64	5.71	7.28	6.94	5.25
Space with repeated bullying	7.04	7.05	5.75	7.14	6.71	7.36	7.22	7.76	7.53	7.09
Public spaces	7.22	7.26	6.88	8.51	6.40	6.65	6.96	7.65	7.51	6.30

### East Asia

The region of East Asia exhibited the lowest level of support for the use of video capture in social VR spaces, both prior to and following deliberations.

Prior to deliberations, the average answer across all spaces for the use of video capture was 5.69, which increased to 5.98 after deliberations. Despite this increase in support, East Asia retained its position as the least enthusiastic region towards the use of this technology. In comparison, the global average for the use of video capture was 6.5 before deliberations and 6.9 after deliberations.

In East Asia, support was the lowest for the use of video capture in members-only spaces, with an average answer of 5.18 before deliberations, and 5.52 after deliberations. Around 30% of East Asian participants opposed the use of video capture in members-only spaces, both before and after deliberations. However, there was a notable increase in the proportion of East Asian participants supporting the use of video capture in members-only spaces, from 39.2% prior to deliberations to a slim majority of 51.8% after deliberations.

Support for the use of video capture in public spaces was initially moderate, with an average rating of 5.45 prior to deliberations. However, this shifted towards enthusiastic approval after deliberations, with the average rating soaring to 6.88. Notably, opposition to the use of video capture in public spaces in East Asia decreased by 34% over the course of deliberations, dropping from 25.9% prior to deliberations to 17.1% after deliberations. Furthermore, support for the technology saw a significant increase, rising from 43.4% before deliberations to 71.6% after deliberations, representing a 65% surge in support. Finally, the proportion of East Asian participants who were uncertain about their stance on video capture in public spaces decreased by a striking 84% throughout the deliberations, falling from 11% to just 1.8%.

In East Asia, support for the use of video capture in social VR spaces with repeated cases of bullying and harassment was the strongest before deliberations compared to members-only spaces and public spaces. The average answer for these spaces was 6.05 before deliberations, but decreased to 5.75 after deliberations, trailing behind public spaces with an average answer of 6.88 after deliberations. This drop in the average answer was due to a significant increase in opposition towards the use of video capture in spaces with repeated bullying and harassment, which rose from 21.1% prior to deliberations to 30.8% after deliberations, representing a 46% growth in opposition. Conversely, support remained stable at 53% both before and after deliberations. It is worth noting that members-only spaces and public spaces with repeated cases of bullying and harassment were treated as a single category due to the similarity in their results.

In general, there were important changes in attitudes towards video capture in East Asia as a result of deliberations, particularly in public spaces, where there was a significant increase in support and decrease in opposition. While East Asian participants initially demonstrated the highest level of support for the use of video capture in spaces with repeated cases of bullying and harassment, after deliberations, public spaces garnered the highest level of support. This shift in preference was due to the fact that the deliberations prompted East Asian participants to reduce their support for the use of video capture in spaces with repeated cases of bullying and harassment, while their support for members-only spaces and public spaces increased. Finally, East Asian participants showed the least support for the use of video capture in members-only spaces.

### Middle East and North Africa

Following East Asia, the Middle East, and North Africa (MENA) region displayed the least support for the use of video capture in social VR spaces. The average answer for the use of video capture across all types of spaces was 5.92 before deliberations, which increased to 6.37 after deliberations.

The MENA region exhibited the lowest support for the use of video capture in members-only spaces, with an average answer of 5.38 before deliberations, increasing to 5.65 after deliberations. The proportion of MENA participants opposing the use of video capture in members-only spaces remained steady at around 30% both before and after deliberations, while support slightly increased from 45.8% to 52.6%, representing a 15% increase.

Support for the use of video capture in public spaces had the highest increase, going from an average answer of 5.7 before deliberations to 6.4 after deliberations, a 0.7 increase in support. This increase is due to a decrease in opposition by 28%, going from 29% before deliberations to 21% after deliberations, and an increase in support of 20%, going from 52% before deliberations to 62% after deliberations.

MENA participants expressed the highest level of endorsement for the use of video capture in spaces where bullying and/or harassment occur frequently. The degree of support was marginally higher for public spaces with repeated incidents of bullying and harassment (6.4 pre-deliberations and 6.8 post-deliberations) than for members-only spaces (6.2 pre-deliberations and 6.6 post-deliberations). This slight discrepancy can be attributed to a slightly greater proportion of participants who opposed the use of this technology in members-only spaces with repeated bullying and/or harassment, as opposed to public spaces. Furthermore, support for the use of video capture in public spaces with repeated bullying and/or harassment was also marginally greater than that for members-only spaces.

Overall, while there was an overall increase in support for video capture across all types of spaces after deliberations, the MENA region still exhibited low levels of support of video capture in comparison to other regions. However, there were notable changes in attitudes towards the technology after deliberations, particularly in public spaces where support increased by 20% and opposition decreased by 28%. These findings suggest that while the MENA region may have a lower level of initial support for video capture, the deliberations and discussions can still have a positive impact on attitudes towards them.

### Western Europe

After the East Asian and MENA regions, Western Europe displayed the lowest levels of support for the use of video capture. The average support for the use of video capture in Western Europe was 6.17 before deliberations and 6.43 after deliberations. Although Western Europe's average answers are getting closer to the global average of 6.5 before deliberations and 6.9 after deliberations, they still remain lower.

Like for the two regions previously discussed, Western European participants supported the least the use of video capture in members-only spaces. The average level of support remained unchanged at 5.2 before and after deliberations. This can be attributed to a simultaneous increase of 10% in support (from 40% to 44%) and 16% in opposition (from 31% to 36%). Despite the slight increase in support levels, video capture failed to garner majority support from Western European participants.

Similarly, to East Asia and MENA, the deliberations had the biggest impact in Western Europe for public spaces, where the average answer rose from 5.7 to 6.3, a 0.6 increase. Opposition to the use of video

capture in public spaces went down from 30.6% before deliberations to 26.2% after deliberations, while support rose from 49.8% before deliberations to 59% after deliberations.

Finally, the use of video capture was most supported in Western Europe for spaces with repeated bullying and/ or harassment. The degree of support was slightly higher for public spaces with repeated incidents of bullying and harassment (7 pre-deliberations and 7.2 post-deliberations) than for members-only spaces (6.8 pre-deliberations and 6.9 post-deliberations). Despite the difference, for both types of spaces a majority of Western European participants supported the use of video capture.

### Central and Eastern Europe

The Central and Eastern European (CEE) region had similar results to Western Europe before deliberations, 6.14 and 6.17 respectively, but higher ones after deliberations, 6.74 and 6.43 respectively. This means that support for the use of video capture grew at a higher rate during deliberations for the CEE region than for Western Europe.

CEE participants least supported the use of video capture for members-only spaces. In fact, the levels of support reduced as a result of deliberations, going from 5.8 before deliberations to 5.6 after deliberations. This decrease is due to the level of support decreasing from 53% to 44%, and the proportion of participants neither for nor against the proposal doubling, from 12% to 24%.

After members-only spaces, the order of preference of spaces where to use video capture was different in the CEE regions than other regions previously discussed. While for other regions, public spaces came second, in CEE it was spaces with repeated instances of bullying and/ or harassment. Before deliberations, the average answer for the use of video capture in spaces where there is repeated bullying and/ or harassment was 6.17, which increased to 7.1 after deliberations, a 0.9 difference. Support was higher for public spaces where there is repeated bullying and/ or harassment (6.3 pre-deliberations and 7.4 post-deliberations) than for members-only spaces where there is repeated bullying and/ or harassment (6.1 pre-deliberations and 6.7 post-deliberations).

The CEE region most supported the use of video capture in public spaces, with an average answer of 6.4 before deliberations and 7.3 after deliberations, a 0.9 increase. This increase was largely driven by a significant decrease in opposition to the use of video capture in public spaces, which dropped from 23.3% to 10%. Notably, while the level of support did not change significantly (61% before deliberations and 65.6% after deliberations), there was a substantial increase in the proportion of participants who were undecided, rising from 11.1% before deliberations to 20.7% after deliberations.

### North America

In North America, support for video capture as a means of identifying bullying and harassment in social VR spaces was unusually low. North American participants' support for turning on video capture in all public spaces rose from 5.8 before deliberations to 6.6 points after deliberations, while the control group's support rose just 0.4 points. Even after this post-deliberations surge in support for this proposal; however, North Americans' support still fell far short of the global average: 7.2 points. Similarly, North



Americans had mixed opinions about turning video capture on in all members-only spaces, awarding this proposal 5.6 points both before and after deliberations. As before, the level of support for this proposal in North America fell far below the global average, which was 6.3 points. North American participants expressed greater acceptance of video capture in spaces where repeated bullying and harassment have occurred. Regional support for video capture in such spaces after deliberations was 7.4 and 7.3 points for public and members-only spaces, respectively.

### Oceania

Participants from Oceania expressed an average degree of support for video capture in public spaces and an unusually low degree of support for video capture in members-only spaces. Support among participants from Oceania for turning on video capture in all public spaces rose from 6.0 points before deliberations to 7.0 points after deliberations. Meanwhile, support dwindled for video capture in members-only spaces, with this proposal receiving 5.6 points before deliberations and 5.7 points after deliberations. Participants from Oceania diverged from their East Asian counterparts in their support of video capture in members-only spaces where repeated bullying and harassment have occurred. More specifically, participants from Oceania favor video capture in members-only spaces that have witnessed repeated bullying and harassment about as much as they favor video capture in a typical public space: 7.1 points.

While participants from some of the above regions exhibited heightened skepticism of video capture in spaces without repeated incidents of bullying and harassment or in members-only spaces, participants from some other regions had fewer such concerns.

### Sub-Saharan Africa

Participants from Sub-Saharan Africa expressed unusually high support for video capture in social VR spaces. Diverging from their Southeast Asian or Latin American counterparts, however, participants from Sub-Saharan Africa expressed an identical degree of support for proposals concerning video capture in all public spaces, public spaces where repeated bullying and harassment have occurred, and members-only spaces where repeated bullying and harassment have occurred: 7.5 points. In all three cases, this rating represented a moderate 0.2-to-0.4-point increase in support from participants' ratings before deliberations. Mirroring participants from other regions, however, participants from Sub-Saharan Africa showed moderated support for video capture in all members-only spaces. Regional support for this proposal reached 6.9 points after deliberations. This level of support exceeds the global average of 6.3 points but is lower than regional support for video capture in any other type of space.

### South and Southeast Asia

In Southeast Asia, participants were broadly supportive of video capture in social VR spaces. Support for turning on video capture in all public spaces was 7.0 points before deliberations and rose to 7.6 points after deliberations, exceeding the global average of 7.2 points. Much of this increase in support may be attributable to the deliberations process, as the control group exhibited no change in opinion on this

proposal. Southeast Asian participants' support for video capture in public spaces was even higher when repeated bullying and harassment had occurred, increasing from 7.5 points before deliberations to 7.8 points after deliberations. And support for video capture did not dwindle much on proposals concerning members-only spaces. Southeast Asian participants rated the proposal to turn on video capture in all members-only spaces 7.3 points, far above the global average of 6.3 points. As with public spaces, participants' support for video capture in members-only spaces was even higher when repeated bullying and harassment had occurred: 7.7 points after deliberations.

### Latin America

Latin American participants showed even greater support for video capture than their Southeast Asian counterparts. Support for turning on video capture in all public spaces narrowly exceeded the global average of 7.2 points before deliberations and rose to 8.5 points after deliberations. Unlike participants from any other region, participants from Latin America supported video capture in public spaces to a lesser degree where repeated bullying and harassment have occurred. Regionally, this proposal received 7.3 points before deliberations and 7.1 points after deliberations. Mirroring Southeast Asia, Latin American support for video capture in members-only spaces was also high. The proposal to turn on video capture in all members-only spaces increased from 7.3 points before deliberations to 7.7 points after deliberations, far above the global average of 6.3 points. That stated, it is unclear how much of this increase in support results from the deliberative process: the control group also witnessed a 0.3-point increase in support for this proposal. As with public spaces, Latin American participants diverged from participants from other regions in supporting video capture in members-only spaces to a lesser degree where repeated bullying and harassment have occurred. This proposal received 7.1 points before deliberations and 7.2 points after deliberations.

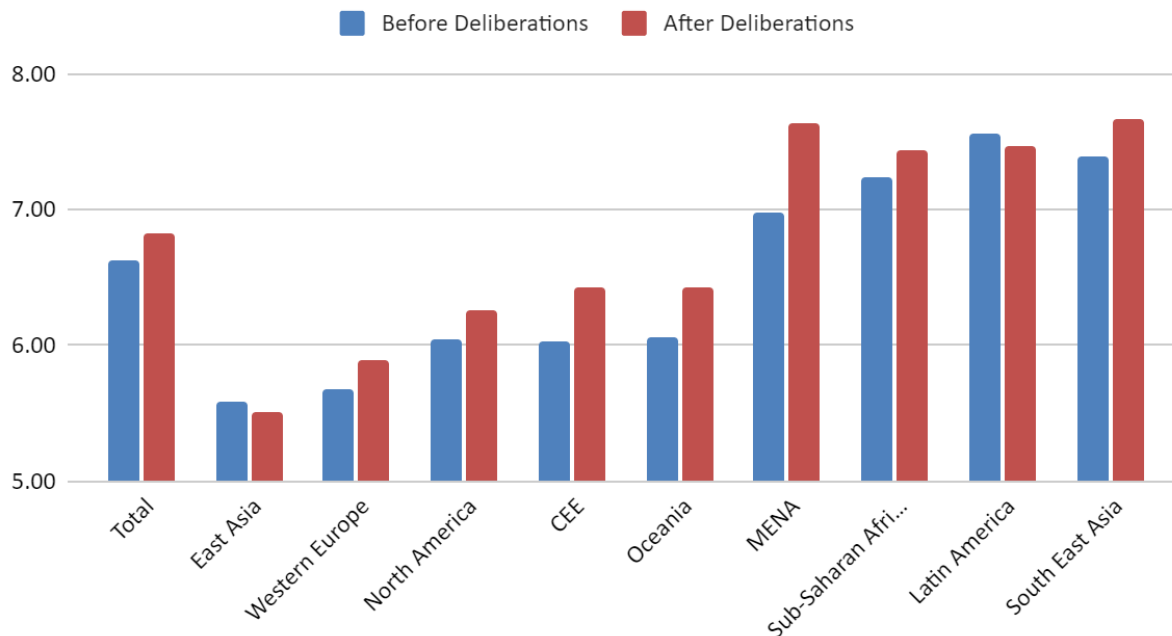
### Automatic Speech Detection

This section examines how participants viewed automatic speech detection differently in the nine regions. Overall, globally, participants felt that automatic speech detection should be used. However, when examining the nine regions specifically, there is some great contrast, in particular, there is much lower support in East Asia and Western Europe and rather high support in Middle East/North Africa and South and Southeast Asia. Across most regions, there was increased support for automatic speech detection after deliberations, with the exception of East Asia.

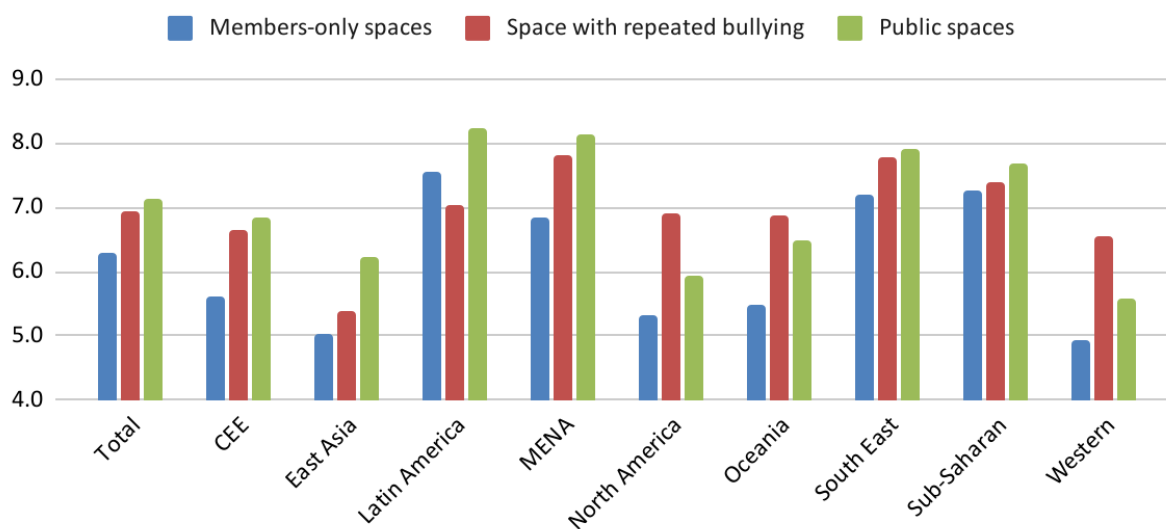
In evaluating the three types of spaces where automatic speech detection could be used: public, members-only, and spaces where there is repeated bullying and harassment, participants in Latin America and MENA rated having speech detection in public spaces the highest, with over a mean of 8, on a 10-point scale. In Western Europe, where participants were concerned with the use of speech detection, they rated using speech detection in spaces where there is repeated bullying and harassment the highest, followed by public spaces, and then members only spaces. For East Asian participants, they

preferred having speech detection in public spaces, then spaces where there is repeated bullying and harassment, and then members only spaces.

### Should automatic speech detection be used in social VR spaces?



### Should automatic speech detection be used in public spaces, members-only spaces and spaces where there is repeated bullying?



Participants' average answers by region after deliberations

	Total	CEE	East Asia	Latin America	MENA	North America	Oceania	Southeast Asia	Sub-Saharan Africa	Western Europe
Members-only spaces	6.3	5.6	5.0	7.6	6.8	5.3	5.5	7.2	7.3	4.9
Space with repeated bullying	6.9	6.6	5.4	7.0	7.8	6.9	6.9	7.8	7.4	6.5
Public spaces	7.1	6.8	6.2	8.2	8.1	5.9	6.5	7.9	7.7	5.6

## Latin America

Latin American participants had unusually high degrees of support for automatic speech detection across the board. While the average global participant rated a proposal to turn on automatic speech detection in public spaces 7.1 points, Latin American participants rated this proposal 7.9 points before deliberations and 8.2 points after deliberations. Similarly, the average global participant rated a proposal to turn on automatic speech detection in members-only spaces 6.3 points, while Latin American participants rated this proposal 7.6 points both before and after deliberations. Also unlike participants in other regions, Latin American participants were less willing to support automatic speech detection in spaces that have witnessed repeated bullying and harassment. Regional support for a proposal to turn on speech detection in public spaces with repeated incidents of bullying and harassment was 7.1 points, matching the global average. And regional support for a proposal to turn on speech detection in members-only spaces with repeated incidents of bullying and harassment was 7.0 points, just above the global average of 6.8 points.

## MENA

Similarly, participants from the Middle East and North Africa voiced notably high support for automatic speech detection in public and members-only spaces, rating these proposals 8.1 points and 6.8 points, respectively. Unlike their Latin American counterparts, Middle Eastern and North African participants' support for automatic speech detection in spaces that have witnessed repeated bullying and harassment matched or exceeded their support for speech detection in typical spaces. More specifically, regional support for a proposal to turn on speech detection in public spaces with repeated incidents of bullying and harassment was 8.1 points after deliberations, while regional support for a proposal to turn on speech detection in members-only spaces with repeated incidents of bullying and harassment was 7.5 points after deliberations.

## South and Southeast Asia

Southeast Asian participants, like Latin American and Middle East and North African participants, were broadly supportive of automatic speech detection. Southeast Asian participants rated the proposal to

turn on automatic speech detection in public spaces 7.5 points before deliberations and 7.9 points after deliberations, outpacing the global average of 7.1 points. Correspondingly, Southeast Asian participants rated the proposal to turn on automatic speech detection in members-only spaces 7.1 points before deliberations and 7.2 points after deliberations, far above the global average of 6.3 points. Uniquely, Southeast Asian participants' support for automatic speech detection in public spaces that have witnessed repeated bullying and harassment was nearly the same as their support for automatic speech detection in typical public spaces—7.8 points after deliberations. That said, regional support for automatic speech detection in members-only spaces where repeated bullying and harassment have occurred was 7.8 points after deliberations—1 point higher than the global average for this proposal, and 0.6 points higher than regional support for speech detection in more typical members-only spaces.

### Sub-Saharan Africa

Participants from Sub-Saharan Africa also expressed widespread support for automatic speech detection in social VR spaces. Regional support for the proposal to turn on automatic speech detection in public spaces was 7.6 points before deliberations and 7.7 points after deliberations. Unlike participants from other regions, Sub-Saharan African participants had about the same degree of support for automatic speech detection in public versus members-only spaces. Regional support for the proposal to turn on automatic speech detection in all members-only spaces was 7.3 points after deliberations. Similarly, whether a space had witnessed repeated bullying and harassment did not significantly alter Sub-Saharan Africans' opinions about whether that space should have automatic speech detection turned on. Regional support for the remaining two proposals about turning on speech detection in public and members-only spaces with repeated bullying and harassment attained 7.3 and 7.4 points after deliberations, respectively.

Participants from some other regions expressed more skepticism about automatic speech detection in social VR spaces.

### East Asia

Notably, participants from East Asia rated the proposal to turn on automatic speech detection in all public spaces 5.5 points before deliberations and 6.2 points after deliberations, falling beneath the global average of 7.1 points. On a similar note, East Asian participants rated the proposal to turn on automatic speech detection in all members-only spaces 5.1 points before deliberations and 5.0 points after deliberations, again falling beneath the global average of 6.3 points. Regional support for automatic speech detection in spaces that have witnessed repeated incidents of bullying and harassment remained low. East Asian support for the remaining two proposals about turning on speech detection in public and members-only spaces with repeated bullying and harassment attained 5.5 and 5.3 points after deliberations, respectively. Regional support for these two proposals actually decreased 0.4 and 0.6 points after deliberations, while the control group's support for both of these proposals increased by 0.4 points.

### Central and Eastern Europe

Central and Eastern European participants shared East Asian participants' relative opposition towards automatic voice detection in members-only spaces, though their opinions on automatic voice detection in public spaces closely resembled the global average. Regarding members-only spaces, Central and Eastern Europeans rated the proposal to turn on automatic voice detection 5.7 points after deliberations and 5.6 points before deliberations. Central and Eastern European participants had slightly heightened support for automatic voice detection in members-only spaces with repeated bullying and harassment—6.4 points after deliberations—but still fell short of the global average for this proposal of 6.8 points. As stated above, regional support for automatic voice detection in public spaces was similar to the global average. The proposal to turn on automatic voice detection in all public spaces received 6.8 points after deliberations, resembling the global average of 7.1 points. And the proposal to turn on automatic voice detection in public spaces where repeated bullying and harassment have occurred received 6.9 points after deliberations, falling just short of the global average of 7.1 points.

Participants from North America, Oceania, and Western Europe were unusually disinclined to support automatic voice detection in public and members-only spaces alike, though their opinions on automatic voice detection in spaces where repeated bullying and harassment have occurred closely resembled the global average.

### North America

North American participants rated the proposal to turn on automatic voice detection in all public spaces 5.6 points before deliberations and 5.9 points after deliberations, falling short of the global average of 7.1 points. Similarly, North American participants rated the proposal to turn on automatic voice detection in all members-only spaces 5.3 points both before and after deliberations, again falling short of the global average of 6.3 points. Regional support jumped regarding spaces where repeated bullying and harassment have occurred. The proposal surrounding automatic voice detection in public spaces with repeated bullying and harassment received 7.0 points after deliberations in North America, closely resembling the global average of 7.1 points. And the proposal surrounding automatic voice detection in members-only spaces with repeated bullying and harassment received 6.8 points after deliberations in North America, matching the global average exactly.

### Oceania

Mirroring their North American counterparts, participants from Oceania rated the proposal to turn on automatic voice detection in all public spaces 5.7 points before deliberations and 6.5 points after deliberations. This leap in support may be attributable to the deliberations process, as the control group's support for this proposal rose just 0.1 points. That said, even after deliberations, regional support for this proposal was far beneath the global average of 7.1 points. Similarly, participants from Oceania rated the proposal to turn on automatic voice detection in all members-only spaces 5.3 points before deliberations and 5.5 points after deliberations, again falling short of the global average of 6.3 points. As in North America, regional support for automatic voice detection was much higher regarding spaces where repeated bullying and harassment have occurred. The proposal surrounding automatic

voice detection in public spaces with repeated bullying and harassment received 7.1 points after deliberations in Oceania, matching the global average, while the proposal surrounding automatic voice detection in members-only spaces with repeated bullying and harassment received 6.6 points after deliberations in Oceania, falling just 0.2 points short of the global average.

### Western Europe

Like participants from North America and Oceania, Western Europeans expressed skepticism towards automatic speech detection that moderated concerning spaces where repeated bullying and harassment have occurred. Participants from Western Europe rated the proposal to turn on automatic voice detection in all public spaces 5.1 points before deliberations and 5.6 points after deliberations, falling far beneath the global average of 7.1 points. Further, participants from Western Europe rated the proposal to turn on automatic voice detection in all members-only spaces just 4.9 points after deliberations, a global low. As before, regional support for automatic voice detection more closely resembled the global average regarding spaces where repeated bullying and harassment have occurred. Regional support for the proposals surrounding automatic voice detection in spaces with repeated bullying and harassment received 6.8 and 6.3 points after deliberations for public and members-only spaces, respectively.

### Who should review alerts?

#### Video Capture

#### North America

Participants from North America aligned closely with global averages in their opinions on who should review alerts from video capture. In North America, participants expressed the highest level of support for the review of video capture by platform owners in public spaces, rating this proposal a 7.1 before deliberations and a 7.7 after deliberations. The global average score for this proposal was 7.8 points. Regional support was lower for review of video capture in public spaces by creators, with this proposal receiving 6.7 points before deliberations and 6.8 points after deliberations. Again, this regional post-deliberations number closely resembled the global average of 7.1 points. Regarding members-only spaces, North American participants' degree of support for review of video capture alerts by creators and by platform owners were similar. North Americans rated creators' review of video capture in members-only spaces 7.2 points after deliberations, and they rated platform owners' review of video capture in members-only spaces 7.1 points after deliberations. The average global participant rated these proposals 7.0 and 7.2 points, respectively.

#### Oceania

Participants from Oceania expressed opinions similar to those of their North American counterparts, and thus resembled the global average. As with North America, participants from Oceania expressed the highest level of support for the review of video capture by platform owners in public spaces, rating this proposal a 6.9 before deliberations and a 7.7 after deliberations. This increase in support may be attributable to the deliberations process, as the control group increased their support for this proposal

by just 0.1 points. Also paralleling North America, Oceanic regional support was lower for review of video capture in public spaces by creators, with this proposal receiving 6.4 points before deliberations and 6.6 points after deliberations. While participants from Oceania otherwise resembled the global average, their support for this proposal fell slightly short of the global average of 7.1 points. Regarding members-only spaces, participants from Oceania displayed the same degree of support for review of video capture alerts by creators and by platform owners. Participants from Oceania rated both of these proposals 6.9 points after deliberations, just below the global average.

### East Asia

Participants from East Asia exhibited more opposition to granting platform owners and creators access to video capture. Regional support for the proposal to grant platform owners access to video capture in public spaces was 6.3 points before deliberations and 7.1 points after deliberations. As in other regions, it appears that deliberations increased participant support of this proposal—the control group did not change their rating of the proposal. However, even after deliberations, East Asian support of this proposal fell beneath the global average of 7.8 points. Participants from East Asia grew even more skeptical of granting creators access to video capture in public spaces after deliberations, lowering their rating from 6.5 points to 5.7 points. This degree of regional support is meaningfully lower than the global average of 7.1 points. East Asian participants also displayed an unusually low degree of support for review of video capture in members-only spaces. Like participants from other regions, they expressed similar degrees of support for granting creators and platform owners access to video capture in members-only spaces, rating these proposals 6.0 and 6.2 points, respectively.

### Western Europe

Like participants from East Asia, participants from Western Europe expressed broadly lower-than-average support for review of video capture, though unlike participants from East Asia, Western European participants approached the global average degree of support for review of video capture by platform owners. Regional support for the proposal to grant platform owners access to video capture in public spaces was 6.8 points before deliberations and 7.6 points after deliberations, nearing the global average of 7.8 points. As in other regions, it appears that deliberations increased participant support of this proposal—the control group increased their rating of this proposal by just 0.1. However, regional support for the proposal to grant creators access to video capture in public spaces was much lower, decreasing from 6.1 points before deliberations to 5.9 points after deliberations, while the global average stood at 7.1 points. Western European participants drew a similar distinction between platform owners' and creators' review of video capture regarding members-only spaces. Regional support for the proposal to grant platform owners access to video capture in members-only spaces was 6.7 points after deliberations, falling just beneath the global average of 7.2 points. Meanwhile, regional support for the proposal to grant creators access to video capture in members-only spaces was 6.3 points after deliberations, far below the global average of 7.0 points.

### Central and Eastern Europe



Central and Eastern European participants closely resembled their Western European counterparts. Regional support for the proposal to grant platform owners access to video capture in public spaces was 6.4 points before deliberations and 7.5 points after deliberations, nearing the global average of 7.8 points. Also resembling Western Europe, regional support for the proposal to grant creators access to video capture in public spaces was much lower. Support increased moderately from 6.3 points before deliberations to 6.4 points after deliberations, while the global average stood at 7.1 points. Regarding members-only spaces, Central and Eastern European participants drew a similar distinction between platform owners' and creators' review of video capture. Regional support for the proposal to grant platform owners access to video capture in members-only spaces was 6.5 points after deliberations, while the global average was 7.2 points. A much larger gap emerged between regional support for granting creators access to video capture in members-only spaces—6.1 points after deliberations—and the global average of 7.0 points.

### MENA

While participants from Western, Central, and Eastern Europe appeared to distinguish strongly between review by platform owners and by creators, participants from the Middle East and North America appeared to draw a more meaningful distinction between review in public spaces and in members-only spaces. Middle Eastern and North African participants rated the proposals to grant platform owners and creators access to video capture in public spaces were both 7.2 points after deliberations, resembling global averages. However, Middle Eastern and North African participants were unusually skeptical about proposals regarding review of video capture in members-only spaces. Regional ratings for the proposals to grant platform owners and creators access to video capture in members-only spaces were both 6.2 points after deliberations, falling far below global averages.

### Latin America

Across the board, participants from Latin America were unusually supportive of video capture review. Regarding public spaces, Latin American participants gave the proposal to grant platform owners access to video capture 8.2 points before deliberations and 8.8 points after deliberations, far exceeding the global average of 7.8 points. As in other regions, much of this increase can be attributed to the deliberations process, as the control group's rating of this proposal increased by just 0.2 points. Similarly, Latin American support for the proposal to grant creators access to video capture in public spaces was 8.0 points after deliberations, above the global average of 7.1 points. Regional support for review of video capture in members-only spaces was also high. After deliberations, Latin American participants rated the proposals to grant platform owners and creators access to video capture in members-only spaces 8.5 points and 8.1 points, respectively.

### Sub-Saharan Africa

Participants from Sub-Saharan Africa also expressed unusually high degrees of support for review of video capture by creators, though they expressed more average degrees of support for review of video capture by platform owners. Regarding public spaces, Sub-Saharan African participants rated the

proposal to grant platform owners access to video capture in public spaces 7.8 points after deliberations, matching the global average exactly. Meanwhile, regional support for granting creators access to video capture in public spaces was 7.6 points before deliberations and 8.0 points after deliberations, while the global average for this proposal stood at 7.1 points. Sub-Saharan African support for review of video capture in members-only spaces showed a similar split. While the average global participant was more supportive of granting platform owners access to video capture in members-only spaces, Sub-Saharan African participants awarded the proposals to grant platform owners and creators access to video capture in members-only spaces the same score after deliberations: 7.6 points.

### South and Southeast Asia

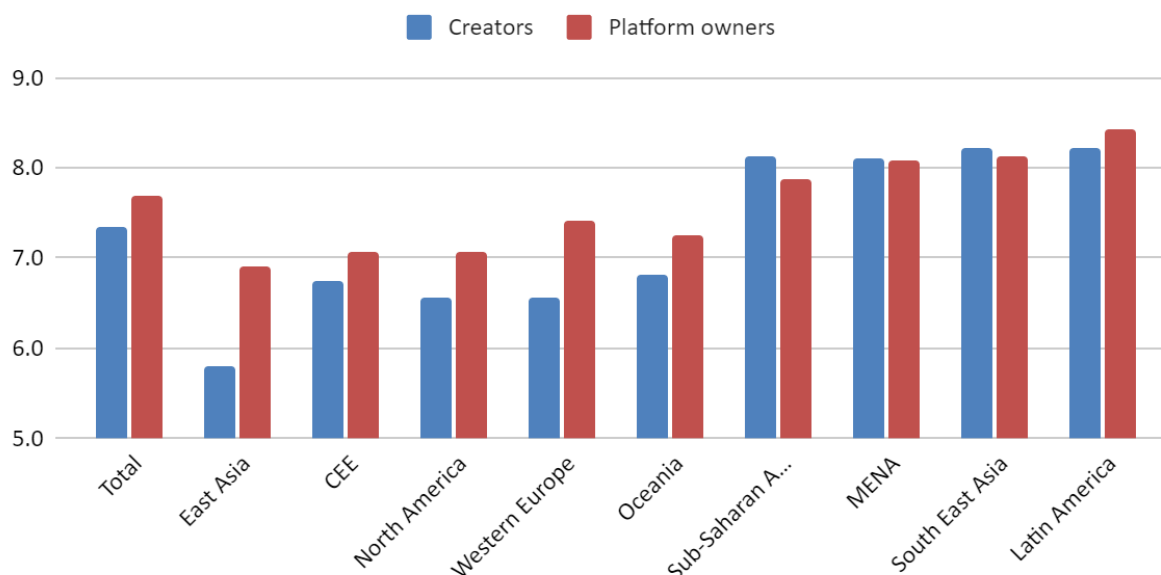
This divide in opinion between video capture review by platform owners versus by creators was even more stark in Southeast Asia. Regional ratings for the proposals granting platform owners access to video capture in public spaces and in members-only spaces both increased to 7.8 points after deliberations. Meanwhile, regional ratings for the proposals granting creators access to video capture in public spaces and in members-only spaces both increased to 7.6 points after deliberations. As a result, Southeast Asian participants barely exceeded the global average in their support for video capture review in public spaces and more meaningfully exceeded the global average in their support for video capture review in members-only spaces.

### Automatic Speech Detection

#### *Who should review speech detection alerts in public spaces?*

Participants across the global differing views on who should review alerts from automatic speech detection. Participants from Latin America, Southeast Asia, MENA, and Sub-Saharan Africa rated both platform owners and creators highly, where the means were 8 (or close to 8) on a 10-point scale. That is, participants from these regions felt both creators and platform owners should review alerts for automatic speech detection. The rest of the regions hovered around 7 on the 10-point scale for having platform owners review alerts from automatic speech detections and the means for creators to review alerts was between 5.5 and 6.5. For these regions, participants preferred platform owners to review the alerts versus the creators.

## Who should review speech detection alerts in public spaces?



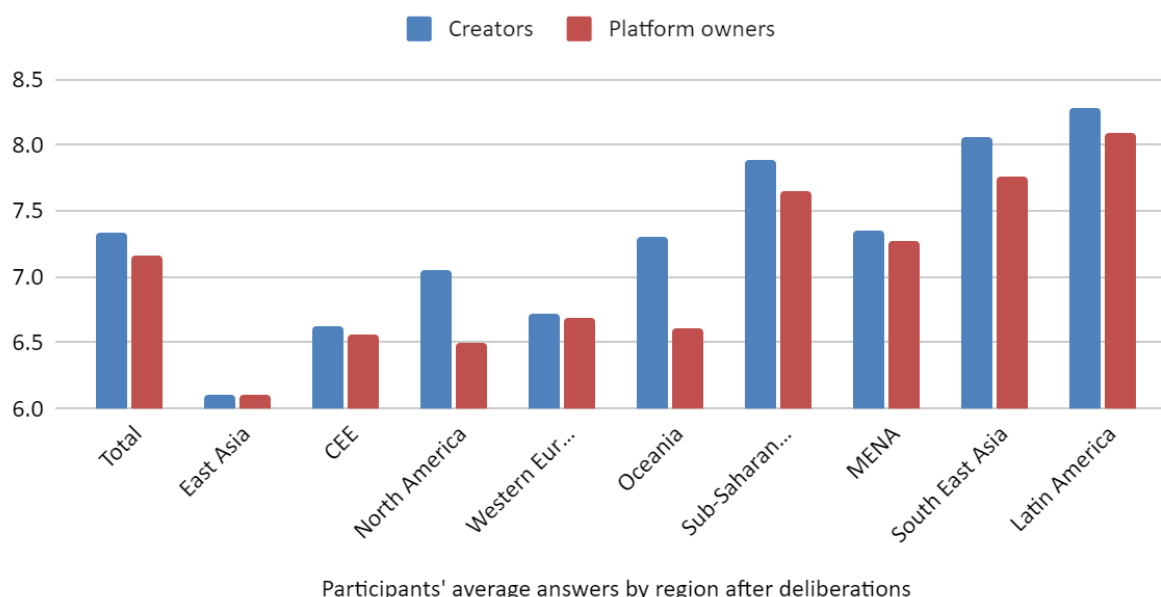
Participants' average answers by region after deliberations

		Total	East Asia	CEE	North America	Western Europe	Oceania	Sub-Saharan Africa	MENA	Southeast Asia	Latin America
Creators	Before	7.3	6.2	6.2	6.8	6.5	7.0	8.0	7.7	8.0	8.3
	After	7.4	5.8	6.7	6.6	6.6	6.8	8.1	8.1	8.2	8.2
platform owners	Before	7.5	6.6	6.3	7.4	7.5	7.4	7.7	7.8	7.8	8.4
	After	7.7	6.9	7.1	7.1	7.4	7.3	7.9	8.1	8.1	8.4

### Who should review speech detection alerts in members-only spaces?

In members-only spaces, there was greater variation in preferences among participants. The most striking in East Asia is that participants rated both creators and platform owners as close to 6, on the 10-point scale. While they are on the 'support' side of the scale, their support for either creators or platform owners was not very strong. In contrast, participants from Latin America supported creators and platform owners to review alerts with a mean of above 8, on the 10-point scale, and they had a slight preference for creators. Actually, participants in all regions, except for East Asia, preferred creators to review alerts from automatic speech detection. Oceania, in particular, preferred creators (7.5) over platform owners (6.5).

## Who should review automatic speech detection alerts in members-only spaces?



		Total	East Asia	CEE	North America	Western Europe	Oceania	Sub-Saharan Africa	MENA	Southeast Asia	Latin America
Creators	Before	7.1	6.3	6.1	7.0	6.5	7.0	7.4	6.9	7.6	8.2
	After	7.3	6.1	6.63	7.1	6.7	7.3	7.9	7.3	8.1	8.3
Platform owners	Before	7.2	6.4	6.2	6.8	6.8	6.8	7.4	7.1	7.7	8.2
	After	7.2	6.1	6.56	6.5	6.7	6.6	7.7	7.3	7.8	8.1

### Notable Regions

#### East Asia

East Asia was the region which demonstrated the least amount of approval for speech detection to be used in social VR spaces so, naturally, it also had the lowest levels of approbation for speech detection alerts to be reviewed, either by creators (6.2 before deliberations and 6 after), or by platform owners (6.3 both before and after deliberations). Those answers were a lot lower than the global average for creators (7.1 before deliberations and 7.2 after) and for platform owners (7.2 before deliberations and 7.3 after). Notably, while originally creators and platform owners had relatively similar levels of support for the review of speech detection alerts in social VR spaces before deliberations (6.2 and 6.3 respectively), after deliberations the level of support went down for creators while it remained the same

for platform owners. As such, East Asian participants had an overall slight preference for speech detection alerts to be reviewed by platform owners. The levels of support, however, varied depending on the public versus members only spaces.

The preference of East Asian participants for platform owners to review speech detection alerts was stronger for public spaces. The average answer in favor of creators reviewing speech detection alerts in public spaces was 6.2 before deliberations and 5.8 after. This decrease in the average answer is due to opposition to the review of speech detection by creators in public spaces growing over the course of deliberations, going from 21.5% before deliberations to 30% after deliberations, a 39% increase in opposition. On the other hand, support for the review of speech detection alerts by creators in public spaces went down over deliberations, from 59.4% before deliberations to 53.5% after deliberations, a 10% decrease in support.

The average answer in favor of platform owners reviewing speech detection alerts was 6.6 before deliberations and 6.9 after deliberations. Opposition for platform owners to review speech detection alerts was lower than for creators throughout the deliberations, at around 16% both before and after deliberations. Support for platform owners to review speech detection was higher than for creators throughout the deliberations, with 64.5% before deliberations and 69.6% after deliberations. As such, after deliberations the review of speech detection by platform owners had the support of 16% more participants (69.6%) than for creators to review speech detection alerts (53.5%).

For spaces with repeated bullying and/ or harassment, both public and members-only, the preference for platform owners to review speech detection alerts over creators was less important. While both creators and platform owners started with 6.2 before deliberations, after deliberations the average answer in favor of creators went down to 6.1 while it remained at 6.2 for platform owners. This slight difference in results after deliberations is because while opposition for the review of alerts grew over deliberations both for creators (from 22.4% to 28%) and platform owners (from 21% to 27.5%), support for creators to review alerts stagnated over deliberations at 58%. For the review of alerts by platform owners, support went up over deliberations for public spaces with repeated instances of bullying and harassment (from 6.2 to 6.3), while it went down for members-only spaces with repeated bullying and harassment (from 6.2 to 6.1). This means that while East Asian participants did have a slight preference for platform owners to review speech detection alerts in public spaces with repeated bullying and harassment, for members-only spaces with repeated bullying and harassment they had no preference between creators and platform owners.

Finally, for members-only spaces, East Asian participants had no preference for speed detection alerts to be reviewed by creators or platform owners, similarly to how it was for members-only spaces with repeated bullying and harassment. While before deliberations participants had a slight preference for the review of speech detection alerts to be done by platform owners (6.4 average answer) than by creators (6.3), after deliberations the average answers for both creators and platform owners went down to 6.1. This means that over the course of deliberations, support for the review of speech detection alerts went down for both creators and platform owners, but it went down more for platform owners. That is because opposition to the review of speech detection alerts went up much more for platform

owners (from 19.2% to 27.6%, a 44% increase) than it did for creators (from 20.4% to 24.3%, a 19% increase). In the end, while the difference is not reflected in the average answer, East Asian participants had a very slim preference towards the review of alerts in members-only spaces by creators (24.3% opposition and 60% support) than by platform owners (27.6% opposition and 58.7% support). The reason why the average answer does not reflect this slight preference is because the proportion of participants that were in the middle was slightly higher for creators (13%) than for platform owners (10.2%).

### Central and Eastern Europe

CEE had the second lowest scores for the review of speech detection alerts by both creators (6.1 pre-deliberations and 6.7 post-deliberations) and platform owners (6.2 pre-deliberations and 6.9 post-deliberations). While the overall results point towards a slight preference by CEE participants for the review of speech detection alerts by platform owners, the preferences depend on the type of space.

For public spaces, CEE participants preferred for speech detection alerts to be reviewed by platform owners (6.3 pre-deliberations and 7.1 post-deliberations) than by creators (6.2 pre-deliberations and 6.7 post-deliberations). While the proportion of support was actually equivalent between creators (56.8% before deliberations and 65% after deliberations) and platform owners (57.7% before deliberations and 65% after), the level of opposition was greater for creators (26.8% before deliberations and 17.4% after deliberations) than for platform owners (24.7% before deliberations and 13.3% after).

For spaces with repeated bullying and harassment, CEE participants preferred for platform owners to review speech detections alerts, both in public and members-only spaces. For public spaces with repeated bullying and/ or harassment, the average answer for creators was 6.2 before deliberations and 6.7 after, compared to 6 before deliberations and 7.1 after deliberations for platform owners. While CEE participants originally favored the review of speech detection alerts in public spaces with repeated bullying and harassment by creators, over the course of deliberations platform owners gained a lot more support than creators. For members-only spaces where there is repeated bullying and harassment, CEE participants preferred for speech detection alerts to be reviewed by platform owners (6.2 before deliberations and 6.7 after) than by creators (6.04 before deliberations and 6.6 after). The difference is slim but is due to slightly more participants opposing for creators to review speech detection alerts after deliberations (19.1% for creators compared to 17.8% for platform owners); and to more participants supporting for speech detection alerts to be reviewed by platform owners (57.4% pre-deliberations and 62.2% post-deliberations) than by creators (55.6% pre-deliberations and 60.6% post-deliberations).

Finally for members-only spaces, while CEE participants originally favored for speech detection alerts to be reviewed by platform owners (6.2 average answer) than by creators (6.1 average answer), after deliberations the trend was reversed in favor of creators (6.63 average answer vs 6.56 for platform owners). The deliberations led more CEE participants to stop opposing the review of speech detection alerts by creators (from 25.5% before deliberations to 17.3% after deliberations, a 32% decrease in opposition), than by platform owners (from 23.5% before deliberations to 19.3% after deliberations, a 18% decrease). Support for the review by creators was overall slightly higher (59.1% before deliberations

and 62.8% after deliberations) than by platform owners (57.9% before deliberations and 58.7% after deliberations).

### North America

North American participants expressed almost equal amounts of overall support for the review of speech detection alerts by creators (6.90 before deliberations and 6.87 after deliberations) than by platform owners (6.92 before deliberations and 6.94 after). But as for the previous part, the preferences in favor of the review by creators or platform owners varied a lot according to the type of space.

For public spaces, North American participants had a clear preference for the review of speech detection alerts by platform owners (7.4 pre-deliberations and 7.1 post-deliberations) than by creators (6.8 pre-deliberations and 6.6 post-deliberations). Interestingly, the average answer for the review of speech detection went down over the course of deliberations both for creators and platform owners. That is because deliberations made North American participants more opposed to the review of speech detection alerts both by creators (from 18.6% pre-deliberations to 22.6% post-deliberations, a 22% increase in opposition) and by platform owners (from 12.8% pre-deliberations to 16.7% post-deliberations, a 30% increase). On the other hand, support remained the same throughout deliberations both for creators (around 61%) and platform owners (around 67%).

For members-only spaces, North American participants preferred for speed detection alerts to be reviewed by creators (7 pre-deliberations and 7.1 post-deliberations) than by platform owners (6.8 pre-deliberations and 6.5 post-deliberations). As such, while the overall answer in favor of the review of alerts went up slightly for creators, it went down significantly for platform owners, reinforcing the gap between them after deliberations. While opposition to the review of speech detection alerts grew both for creators (from 14.3% to 17.8%) and platform owners (from 18.2 to 23.8%), it grew more for platform owners. Support for the review of speech detection alerts by platform owners also stagnated at 60% over deliberations, when it went up for creators from 66% to 70%. After deliberations, 10% more North American participants supported the review of speech detection alerts in members-only spaces by creators than by platform owners.

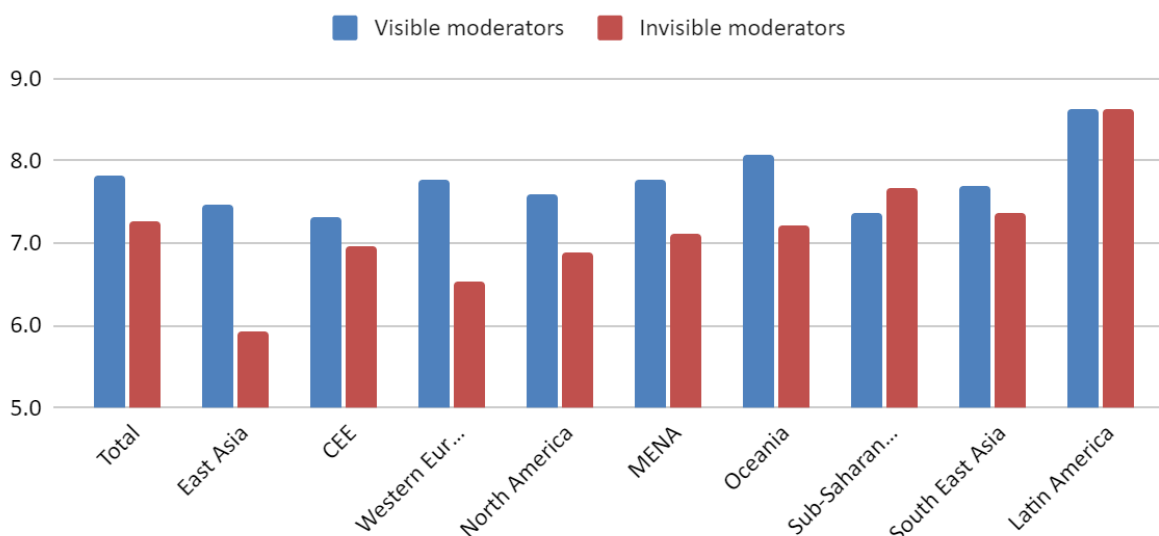
Spaces with repeated bullying and/ or harassment followed the same trend as spaces without repeated bullying and harassment. North Americans preferred for speech detection alerts in public spaces where there is repeated bullying and/ or harassment to be reviewed by platform owners (6.7 before deliberations and 7.2 after deliberations) than by creators (6.9 both before and after deliberations). For members-only spaces where there is repeated bullying and/ or harassment, North Americans preferred for speech detection alerts to be reviewed by creators (6.9 before deliberations and 7 after deliberations) than by platform owners (6.8 before deliberations and 6.9 after).

### Moderators: Visible and Invisible

#### 1. Public Spaces

Across the board, visible moderators were favored over invisible moderators, except for Latin America where visible and invisible moderators received equal amounts of support. The level of support for visible or invisible moderators varied depending on the region.

## Should moderators be used to monitor bullying and harassment in public spaces?



Participants' average answers by region after deliberations

		Total	East Asia	CEE	Western Europe	North America	MENA	Oceania	Sub-Saharan Africa	Southeast Asia	Latin America
Visible moderators	Before	7.5	6.8	6.5	7.7	7.3	7.2	7.6	7.6	7.5	8.2
	After	7.8	7.5	7.3	7.8	7.6	7.8	8.1	7.4	7.7	8.6
Invisible moderators	Before	7.0	6.4	6.3	6.5	6.6	6.7	6.7	7.6	7.1	8.0
	After	7.3	5.9	7.0	6.5	6.9	7.1	7.2	7.7	7.4	8.6

### b. When should moderators enter public spaces? (When requested by users v creators)

Overall, there was less support for invisible moderators to enter public spaces than for visible moderators, regardless of whether they were requested by users or creators.



When requested by users, visible moderators were supported mild to moderately with eight regions above 6, and of those regions three regions above 7. Only two regions reported a 5.9. Invisible moderators were less supported. After deliberations, the lowest answer was a 5.4 from Western Europe and the highest a 7.7 from MENA.

When requested by creators, visible moderators were supported mild to moderately with eight regions with an average answer above 6. The lowest average answer post-deliberations was Western Europe with a 5.5 and the highest MENA with a 7.9. Invisible moderators received slightly lower levels of support, with four regions having average answers between 5 and 6 after deliberations, with the lowest being Western Europe again, with a 5.2. The highest average answer for invisible moderators to enter spaces when requested by creators was 7.7 for MENA.

Deliberations led to an increase in approval for visible moderators to enter spaces when requested by users. Despite this increase, post-deliberations, visible moderators were supported to enter spaces when requested by creators or users almost equally. The only exceptions are CEE and Western Europe, where participants had a stronger preference for moderators to enter spaces when requested by users than when requested by creators

Invisible moderators were supported to enter spaces when requested by creators or users in similar proportions too. The only exceptions are Western Europe where they preferred for moderators to enter spaces when requested by users over creators, and Oceania, where they preferred for moderators to enter spaces when requested by creators over users.

		Total	East Asia	CEE	Western Europe	North America	MENA	Oceania	Sub-Saharan Africa	Southeast Asia	Latin America
When requested by creators	Visible	6.5	6.0	6.1	5.5	6.4	7.9	6.6	7.1	7.6	6.0
	Invisible	6.2	5.4	5.9	5.2	5.8	7.6	6.3	7.0	7.3	6.0
When requested by users	Visible	6.6	6.1	6.4	5.9	6.5	7.9	6.7	7.2	7.5	5.9
	Invisible	6.3	5.6	6.1	5.4	5.8	7.7	6.0	6.9	7.3	5.8

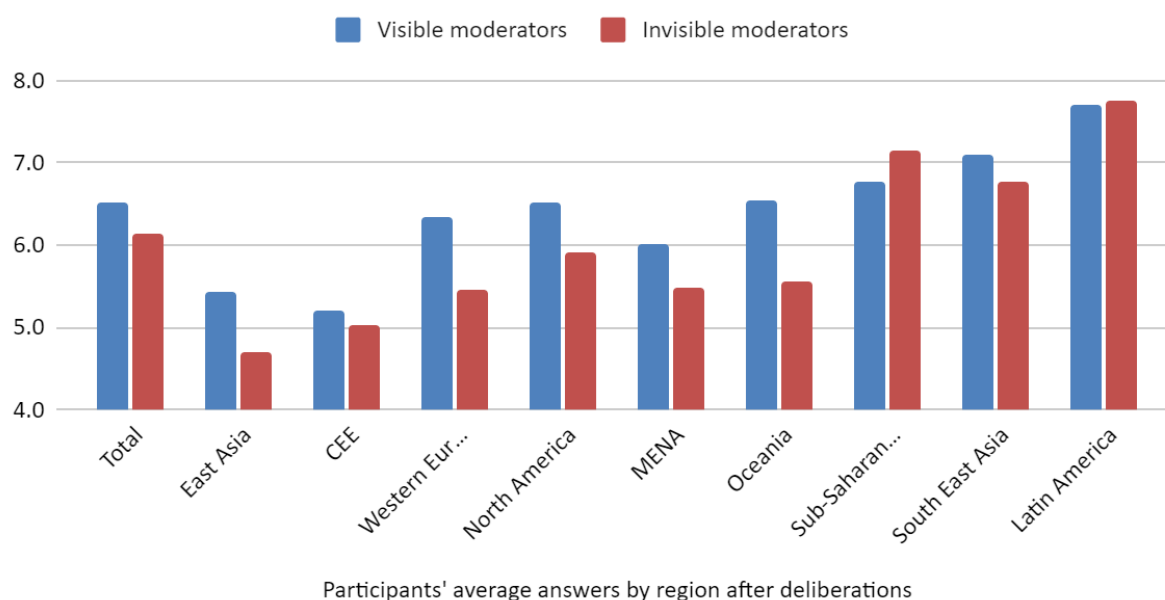
### c. Spaces with repeated bullying and harassment

Support was moderate to strong for moderators to enter areas with repeated bullying and harassment. Every country before and polled near or above 7.0 in scoring besides the lowest of 6.6 for North America. Invisible moderators only garnered mild support, with the lowest a score of 5.6 from CEE. Every other region reported values past six, with Sub-Saharan Africa and Southeast Asia at the high of 7.2 after deliberations.

## d. Members-only spaces

Invisible moderators were more likely to produce a higher negative value in both spaces. All but one region reported a zero or negative preference for invisible over invisible moderators after deliberations: Sub-Saharan Africa at 0.4. It is apparent that regions reserve little support for invisible moderators in members-only spaces when compared to visible moderators. Overall, visible moderators in members-only spaces garnered mild to moderate support with eight regions scoring a six or higher, and two regions, CEE, and Western Europe, scoring 5.7. Invisible moderators had mild support with the lowest score of 4.7 after deliberations from East Asia, four countries scoring above a six, and the rest scoring a five including a 5.9 score from North America.

## Should moderators be used to monitor bullying and harassment in members-only spaces?

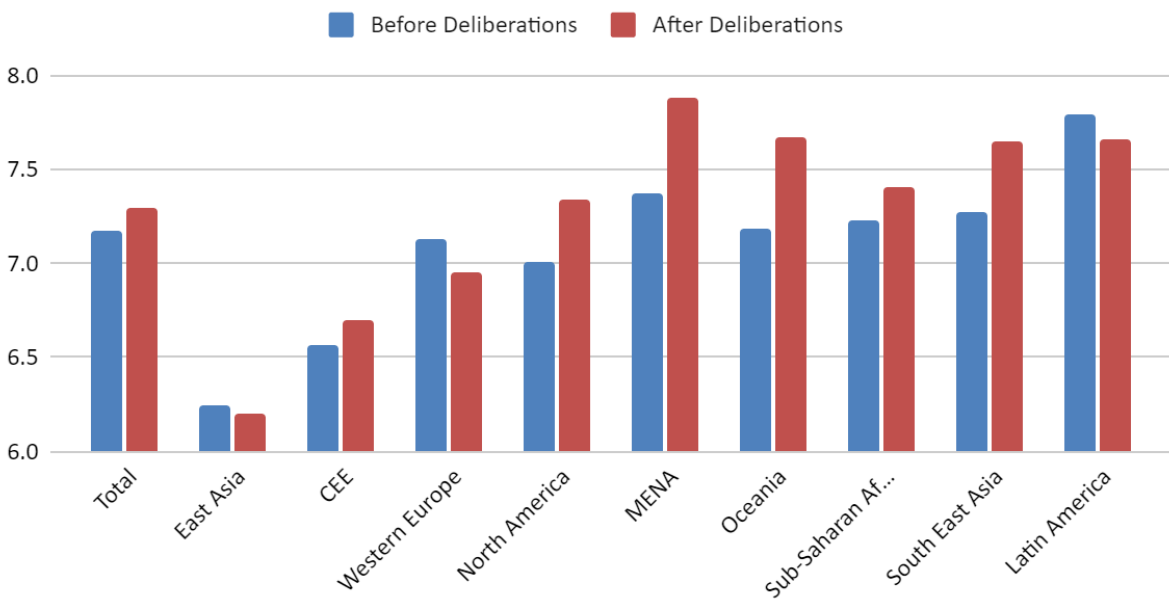


		Total	East Asia	CEE	Western Europe	North America	MENA	Oceania	Sub-Saharan Africa	Southeast Asia	Latin America
Visible moderators	Before	6.8	5.9	6.2	6.6	6.8	5.9	6.6	6.8	7.3	7.8
	After	6.5	5.4	5.2	6.3	6.5	6.0	6.5	6.8	7.1	7.7
Invisible moderators	Before	6.4	5.6	5.7	6.0	5.8	5.5	5.9	6.7	6.9	7.7
	After	6.1	4.7	5.0	5.5	5.9	5.5	5.6	7.1	6.8	7.8

e. Should creators be able to select other users to be visible moderators in their spaces?

Mild to moderate support was garnered when answering if creators should be able to select other users to be visible moderators in their spaces. Some regions were very supportive, with a high post-deliberations of 7.9 from MENA and 7.7 from Latin America. The lowest score, however, was from East Asia with a 6.2.

Should creators be able to select other users to be visible moderators in their spaces?



	Total	East Asia	CEE	Western Europe	North America	MENA	Oceania	Sub-Saharan Africa	Southeast Asia	Latin America
Before Deliberations	7.2	6.2	6.6	7.1	7.0	7.4	7.2	7.2	7.3	7.8
After Deliberations	7.3	6.2	6.7	6.9	7.3	7.9	7.7	7.4	7.6	7.7

### Demographic Analysis

Analysis of demographic factors suggests that differences in gender, age, and experience with social spaces in VR did not lead to differences in ratings of proposals for countering bullying and harassment. Instead, mean proposal ratings were highly correlated (.89 and above) between females and males, across age groups, and among people with different levels of experience with social spaces in VR. When participants' opinions shifted significantly from T1 to T2, they tended to shift in the same direction by a similar magnitude. When they did not, the differences were substantively small (less than 1 point on a 0 to 10 scale). Collectively, these results suggest that these demographic factors did not lead to differences in proposal ratings.

Before and after the Community Forum, participants rated a series of proposals for countering bullying and harassment in social spaces in VR addressing the following main questions:

- 1) Who should be responsible for addressing bullying and harassment in virtual spaces?
- 2) To whom should cases of bullying and harassment be reported?
- 3) Should creators be made aware of reports made to platform owners?
- 4) What should be done about spaces where there is repeated bullying and/or harassment?
- 5) What tools should be used to identify bullying and harassment in social spaces in VR?
- 6) Who should review alerts?
- 7) Should moderators be used to monitor bullying and harassment in virtual spaces?

A separate control group that did not participate in the Community Forum and did not receive any briefing materials rated the same proposals before and after the event. The proposals were rated on a scale of 0 to 10 (0 = strongly oppose, 5 = middle, 10 = strongly support).

Participant and control groups respondents had highly consistent ratings of proposals across the demographic characteristics of gender, age, and whether they had experience with social spaces in VR.

### Gender

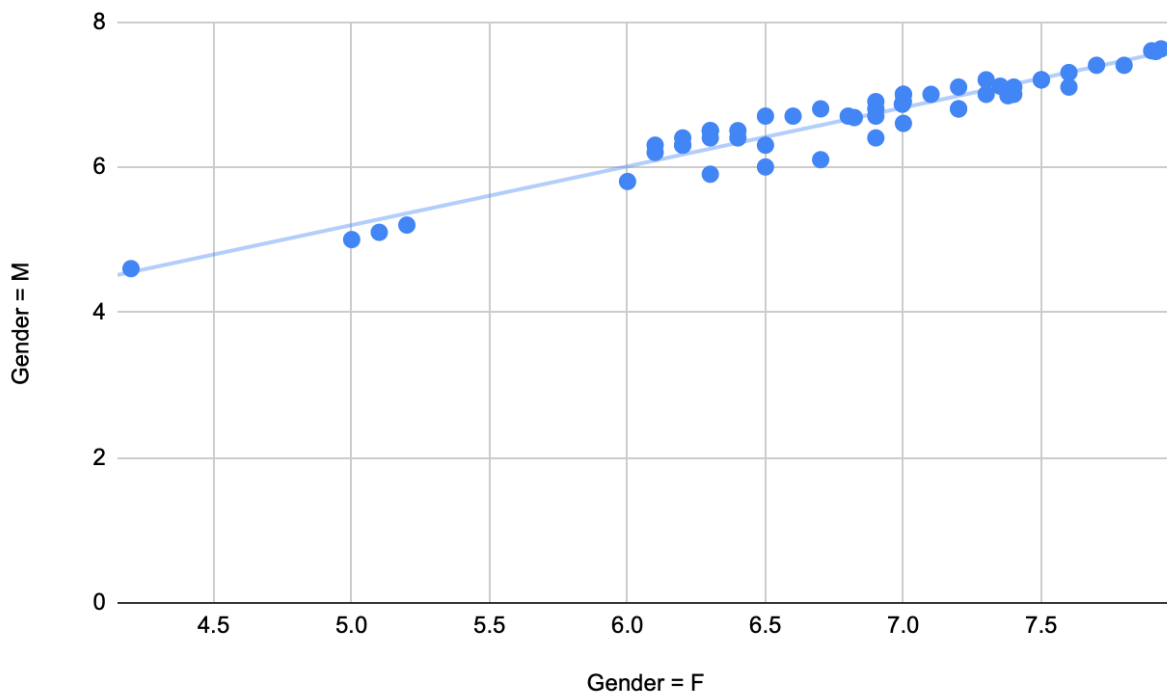
Females and males provided very similar mean proposal ratings both at T1 and T2. When there were differences in significant changes in mean ratings from T1 to T2 between genders, they were substantively small (0.3 or less on a 0 to 10 point scale).

The participant group included 3,397 women and 3,001 men (total  $n = 6,398$ ).  
The control group consisted of 3,291 women and 3,066 men (total  $n = 6,357$ ).

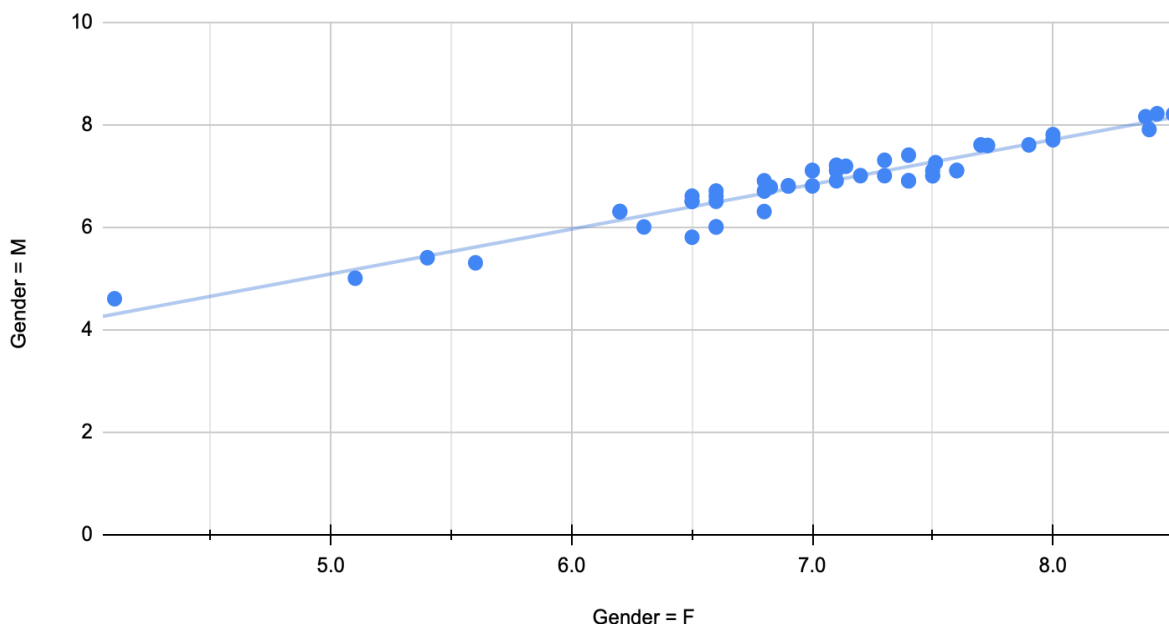
The charts below show a scatterplot of the mean participant responses of females vs. males. The trendlines highlight high correlations between female and male responses both before (T1,  $r = .96$ ) and

after (T2,  $r = .95$ ) the Community Forum. For example, in the first chart, the leftmost point shows that at T1, female participants gave a mean rating of 4.2 to a proposal (Platform owners should not notify the creator of the space where it happened), while male participants provided a mean rating of 4.6. In this case, both females and males rated this approach slightly unfavorably. Other responses showed similar ratings between females and males before and after the event.

Mean Participant Responses by Gender at T1



Mean Participant Responses by Gender at T2



When proposal ratings shifted from before to after the event, they tended to shift in the same direction and magnitude. For instance, on the question, “What should be done about spaces with repeated bullying and/or harassment?” Females gave an initial rating of 6.2 to the proposal that “Members-only spaces should be made less visible to users,” while males gave it an initial favorable rating of 6.4 at T1. After the event, females increased their rating to 6.8, while males increased their rating to 6.7. The difference in increases was 0.3, which was substantively small – less than half a point on the 0 to 10 rating scale.

The control group showed a similar pattern of high correlations between mean female and male responses both at T1 ( $r = .98$ ) and T2 ( $r = .97$ ). Collectively, these results suggest that gender did not play a determinative role in participants’ ratings of the proposals overall.

### Age

Like gender, mean ratings of proposals were highly correlated across age groups both at T1 and T2. When there were differences in changes of mean ratings from T1 to T2 among age groups, they were substantively small (less than 1 point).

For the participant group, the distribution by age group was:

## COMMUNITY FORUM - RESULTS ANALYSIS

1	2	3	4	5	Total
18-24	25-34	35-44	45-54	Over 54	
n = 1218	n = 1691	n = 1233	n = 1065	n = 1191	n = 6398

The control group had the following age distribution:

1	2	3	4	5	Total
18-24	25-34	35-44	45-54	Over 54	
n = 923	n = 1523	n = 1304	n = 1058	n = 1549	n = 6357

The correlation matrices below show high correlations among the five age groups with the lowest (but still very high) correlations between the 18-24 and Over 54 groups (T1,  $r = .91$ ) and (T2,  $r = .89$ ).

<b>Correlation Matrix of Mean Participant Responses by Age Group at T1</b>					
Age Group	1	2	3	4	5
1	1.00				
2	0.98	1.00			
3	0.96	0.98	1.00		
4	0.96	0.99	0.99	1.00	
5	0.91	0.95	0.97	0.97	1.00

<b>Correlation Matrix of Mean Participant Responses by Age Group at T2</b>					
Age Group	1	2	3	4	5
1	1.00				
2	0.97	1.00			
3	0.96	0.98	1.00		
4	0.95	0.97	0.99	1.00	
5	0.89	0.94	0.96	0.97	1.00

In other words, participants across different age groups tended to provide similar ratings for proposals. Changes from before to after the Community Event also tended to be of a similar magnitude and direction.

When there were differences, they tended to be substantively small. For example, the largest difference in changes of ratings was on the question of whether creators of social spaces in VR should be informed of reports of bullying and harassment made to platform owners. The 18-24 age group gave the proposal, “Platform owners should not notify the creator of the space where it happened” a mean rating of 4.9 at T1 and increased it to 5.1 at T2. The 45 to 54 age group lowered its unfavorable mean rating of 4.2 at T1 to 3.6 at T2. The difference in their change of ratings was about 0.8, less than 1 point on a 0 to 10 scale.

All other differences in changes among age groups were even smaller. For instance, the 18-24 age group rated the proposal, “Platform owners should notify the creator of the space where it happened with details about the event and the users involved” favorably at a mean of 6.7 at T1 and did not significantly change its rating. The Over 54 group gave a 7.2 mean rating at T1 and lowered it to 6.7 at T2. Similarly, for the proposal, “Creators should be primarily responsible for monitoring members-only spaces they create,” the Over 54 group did not significantly change its mean rating of 7.2 from T1 to T2, while the 18-24 group increased its mean rating from 7.1 at T1 to 7.7 at T2.

The correlation matrices below show that the control group’s responses were also highly correlated across age groups. At T1, the lowest (though still very high) correlation was between the 18-24 and Over 54 groups ( $r = .92$ ), while at T2, the lowest correlation was between the 35-44 and Over 54 groups.

<b>Correlation Matrix of Mean Control Group Responses by Age Group at T1</b>					
Age Group	1	2	3	4	5
1	1.00				
2	0.97	1.00			
3	0.97	0.98	1.00		
4	0.96	0.98	0.97	1.00	
5	0.92	0.94	0.95	0.96	1.00

<b>Correlation Matrix of Mean Control Group Responses by Age Group at T2</b>					
Age Group	1	2	3	4	5
1	1.00				
2	0.97	1.00			



3	0.95	0.97	1.00		
4	0.95	0.97	0.96	1.00	
5	0.92	0.93	0.91	0.97	1.00

On average, age did not appear to be a determinative factor in people's rating of proposals, despite some substantively small divergences in opinion changes between age groups.

### Experience with Social Spaces in VR

As with gender and age, mean proposal ratings were highly correlated across experience with social spaces in VR at both T1 and T2. Any differences in participants' mean responses were substantively small (within 0.4 on a 0 to 10 scale).

Participants in the deliberation had the following distribution by experience with social spaces in VR.

No Experience	Unsure	Experience	Total
n = 3,931	n = 531	n = 1,934	n = 6,398

The control group had the following distribution by experience with social spaces in VR.

No Experience	Unsure	Experience	Total
n = 3,878	n = 807	n = 1,671	n = 6,357

About 8.3% of participants were unsure whether they had experience with social spaces in VR, while about 12.7% of the control group was unsure about this point. Nevertheless, as the matrices below show, there were high correlations among the mean ratings of the proposals for each experience group. As with gender and age, when there were differences in significant changes of mean ratings from T1 to T2, they tended to be substantively small. The largest difference was related to the question of what should be done about spaces with repeated instances of bullying and harassment. For the proposal, "Members-only spaces should not be publicly discoverable," participants without experience with social spaces in VR gave a mean rating of 6.1 at T1 and 6.4 at T2. Participants who were unsure about their VR experience also gave an initial mean rating of 6.1 at T1 but raised it to 6.8 at T2. Both groups still rated the proposal positively at T2 and had ratings within 0.4 of each other. The difference in the changes of their ratings from T1 to T2 was also substantively small at 0.4.

Other differences in changes of mean ratings between experience groups from T1 to T2 were less than 0.4. For example, on the question of who should be responsible for addressing bullying and harassment in virtual spaces, participants without experience in VR social spaces gave an initial mean rating of 5.0 at T1 and increased it to 5.3 at T2. For people unsure about their VR experience, their mean rating was 5.4 at T1 and 6.0 at T2. The increase in their rating was only 0.3 more than that of people without VR experience.

Correlation Matrix of Mean Participant Responses by VR Social Space Experience at T1			
VR Social Space Experience?	<i>No</i>	<i>Unsure</i>	<i>Yes</i>
<i>No</i>	1.00		
<i>Unsure</i>	0.97	1.00	
<i>Yes</i>	0.98	0.97	1.00

Correlation Matrix of Mean Participant Responses by VR Social Space Experience at T2			
VR Social Space Experience?	<i>No</i>	<i>Unsure</i>	<i>Yes</i>
<i>No</i>	1.00		
<i>Unsure</i>	0.96	1.00	
<i>Yes</i>	0.98	0.97	1.00

Control group responses were also highly correlated across experience in VR social spaces, as the matrices below show.

Correlation Matrix of Mean Control Group Responses by VR Social Space Experience at T1			
VR Social Space Experience?	<i>No</i>	<i>Unsure</i>	<i>Yes</i>
<i>No</i>	1.00		

<i>Unsure</i>	0.95	1.00	
<i>Yes</i>	0.97	0.96	1.00

Correlation Matrix of Mean Control Group Responses by VR Social Space Experience at T2			
VR Social Space Experience?	<i>No</i>	<i>Unsure</i>	<i>Yes</i>
<i>No</i>	1.00		
<i>Unsure</i>	0.95	1.00	
<i>Yes</i>	0.97	0.95	1.00

On average, having experience with social spaces in VR did not appear to be a determinative factor in people’s rating of proposals, despite some substantively small divergences in opinion changes among groups.

Taken together, it appears that neither gender, age, nor experience with VR were determinative of participant or control group respondents’ ratings of the proposals for countering bullying and harassment.

More information about the proposals and ratings can be found in Appendix \*\*\*.

## Control Group

### Key Takeaways

There were minimal differences between the control group and the deliberators before deliberations, as expected. All polling firms recruited one representative sample and then respondents were randomly assigned to either the control group or deliberation group. As such, each respondent had an equal chance of being placed in the control group or deliberation group. With this random assignment, we expect there to be minimal differences between the two groups. Only three questions had notable differences: should creators be banned from making additional members-only spaces, should automatic speech detection be turned on in all members-only spaces and should visible moderators only enter public spaces when requested by users. More important, while these are notable differences, the deliberators and the control group remained on the same side of the scale. That is, both groups had the

same opinions of the proposal, the only difference was that one group had slightly stronger opinions than the other.

## I. Difference before and after deliberations

In the Deliberative Poll, a total of 12,755 individuals were divided into two groups: 6,357 in the control group who did not participate in deliberations, and 6,398 in the participants group who did. Given that the control group did not engage in the deliberations, their survey responses before and after the poll were largely consistent, with differences ranging from 0 (indicating no change) to 0.37. Among all questions surveyed, only three showed a variation in responses exceeding 0.3 points between both surveys for the control group.

- 0.32 difference - Should creators be banned from making additional members-only spaces?
- 0.32 difference - Should automatic speech detection be turned on in all members-only spaces?
- 0.37 difference - Should visible moderators only enter public spaces when requested by users?

These differences are small, and only represent a small proportion of all questions, around 5%. Another 9 questions had a difference between 0.2 and 0.3 between both surveys. In comparison, for participants, 31 questions (55% of total) had a difference above 0.2, and 17 (30% of total) above 0.3. This shows that participants changed their opinions much more over the course of deliberations than the control group did. On average, participants' answers had a difference of 0.23 between the pre- and post-poll surveys, which was more than double the difference observed in the control group, 0.11.

	Control group	Participants	Difference	Difference in %
Average difference in answers before and after deliberations	0.11	0.23	0.12	112.4% more for participants
Max difference in answers between before and after deliberations	0.37	0.85	0.47	127.8% more for participants
Percentage of questions with a difference in answers above 0.3 before and after deliberations	5%	30%	25%	500% more for participants

### II. Before deliberations

The control group's answers before deliberations were very similar to that of participants. The range of differences went from 0, signifying no difference in answers between the control group and participants, up to 0.58. Only five questions had a variation in answers above 0.2 points between both groups:

- 0.58 difference - Should members-only spaces where there is repeated bullying and/ or harassment be removed from the platform?
- 0.57 difference - Should creators be banned from inviting additional people to join spaces where there is repeated bullying and/ or harassment?
- 0.33 difference - Should creators not be able to make money off of spaces where there is repeated bullying and/ or harassment?
- 0.24 difference - Should members-only spaces where there is repeated bullying and/ or harassment not be publicly discoverable?
- 0.23 difference - Should creators of spaces where there is repeated bullying and/ or harassment be banned from making additional members-only spaces?

Interestingly, all 5 questions which had a difference above 0.2 between the control group and participants were related to the last session and questions of measures to take against spaces where there is repeated bullying and/ or harassment. For all 5 questions, the control group's answers were higher than that of participants.

There were 4 more questions for which the control group's answers were lower than that of participants by more than -0.3 points:

- -0.32 difference - Should automatic speech detection be turned on in all public spaces?
- -0.32 difference - Should invisible moderators be able to enter all public spaces?
- -0.36 difference - Should creators send severe cases of bullying and harassment in members-only spaces to platform owners?
- -0.39 difference - Should visible moderators be able to enter public spaces?

Overall, only 12.5% of questions had a difference above 0.3 or under -0.3 between the control group and participants.

### III. After deliberations

After deliberations, the difference in answers between the control group and participants became more important, as expected. 36% of questions had a difference above 0.3 or under -0.3 between the control group and participants. There were many more questions where the control group's answers were lower than that of participants (29%), than questions where the control group's answers were higher than that

of participants (7%). That is because the participants' answers varied a lot more before and after deliberations than the control group.

### Qualitative Analysis

This section presents qualitative analysis of the Community Forum on bullying and harassment in the Metaverse. Across 32 countries and 19 languages, this Community Forum had 2,069 small group sessions. The breakdown of countries, languages, weekday sessions, and weekend sessions is available [here](#). With the limited time available, this section analyzed a random subset of transcripts, selecting transcripts from each region of the world. DDL will continue analyzing all 2,069 transcripts to understand more comprehensively what all deliberators had to say.

This section is organized by sessions from the Community Forum and highlights how participants related to social virtual spaces, similarities and differences in arguments from participants globally, and presents a subset of quantitative results that demonstrate the numbers of arguments participants made and how many pros and cons arguments they made in discussing these topics.

### Key Takeaways

In Session 1, where participants discussed tools used in social virtual spaces, participants commonly made the metaphor of a "camera" in the physical world throughout the deliberations justification of surveillance these spaces; participants found cameras as something reasonable and prevalent in the physical world, so it shouldn't be unreasonable to have them these spaces. However, automatic speech detection could fearfully be used to collect data and formulate curated advertisements for every user. Furthermore, automatic speech detection might not be able to understand banter, slang, etc., and it might flag people who don't deserve to be flagged.

In Session 2, participants discussed the presence of moderators in social virtual spaces and where and when moderators should be present. Participants often compared invisible and visible moderators to police officers in the real world, who can enact consequences to those who act outrageous and uncivil. Participants also felt Metaverse creators may not be competent (or even 'bullies') and will need someone to regulate them (ie, platform owners). However, on the other hand, it may be impractical to have platform owners solely responsible for addressing bullying and harassment; having a communicative dynamic between creators and platform owners may be ideal to address bullying and harassment issues.

In Session 3, participants discussed what should be done when bullying and harassment occurs. Participants often compared the process of addressing bullies/harassers as a process similar to addressing bullying/harassment by a HR department. They felt creators should be responsible for addressing bullying/harassment incidents, and platform owners should create long-term plans to

address harassment/bullying. And, participants discussed that for extreme situations, there should be a transparent and effective triangular relationship between creators, platform owners, and law enforcement agencies.

In Session 4, participants discussed consequences for creators in spaces where bullying and harassment occurs repeatedly. Participants drew comparisons to existing features on Facebook, where some participants suggested using preemptive measures such as applications to filter potential users by asking them why they want to join, which is similar to the process used by many private Facebook groups to restrict entry to a specific group of people. On specific actions, participants felt Metaverse users should be warned prior to entering a space about whether or not reports of bullying have been made in the past (a “notice” perhaps) and if a space cannot be regulated or surveillanced enough by creators or platform owners, it should not exist in the first place.

Session 1: What tools should be used to identify bullying and harassment in the metaverse, and where should they exist?

Identifying what counts as bullying and harassment can be challenging, both in the physical world and in social VR spaces, like those that exist in the Metaverse. In order for Meta to investigate reports of bullying and harassment in digital space, we might look at evidence like victim reports or security camera footage in the real world to understand what has been happening. Therefore, this session asked participants to deliberate about the various tools within the Metaverse to tackle bullying and harassment. 8 proposals were discussed in Section 1.

How did participants relate to video capture and automatic speech detection?

Across the various countries involved in this deliberation, it is evident that participants are able to conceptualize Automatic Speech Detection (ASD) and video capture tools as surveillance devices and programs both in the real world and already existing digital platforms. The metaphor of a “camera” in the physical world was referred to throughout these deliberations, especially in public spaces in the Metaverse. North American participants noted that consequences for uncivil behavior in the real world and virtual platforms exist in the form of monitoring unruly actions and behavior both in physical spaces like a store and in virtual platforms like Zoom meeting rooms. One participant noted that having a camera in a store can catch criminals who break into the store and steal, and another participant noted that in Zoom meetings, meeting goers are warned that they are in a recorded and monitored space and that if they do not consent, they will get kicked out of the call and chat. Taiwanese and Japanese deliberators noted that cameras and surveillance in the virtual world shouldn’t be some novel or provocative idea, as we have “surveillance cameras in our daily lives” that prevent people from acting outrageous or uncivil; likewise, these tools will allow Meta to prevent outrageous actions in the Metaverse.

“Have you seen the movie Ready Player One...you have a complete open world. Everybody can create whatever they want, but in your private rooms, I think those would be set up more for work, or whatever that group wants to do. But I think it would be great, like, [in] the movies if you can go in and just completely interact with your surrounding[s in the Metaverse].” - 17331, comparing the Metaverse to the virtual reality universe in the 2018 movie Ready Player One (20221203-meta-enna-1yc-b\_transcript)

### Similar and Different Views in Session 1:

Overall, it is understood that these tools could reasonably tackle bullying and harassment for the benefit of Metaverse users; this is based on the concept that Metaverse (and social VR spaces in general) are part of the future of daily life for an average person regardless of their age, nationality, gender, etc. Many see surveillance in the Metaverse, for the sake of countering bullying and harassment, as a mundane, reasonable, and beneficial form of regulation. Participants across many Asian nations note that the risk of bullying and harassment by not implementing some mode of surveillance is not sensible, at the very least in public spaces.

Participants in different countries and continents find surveillance as a daily part of physical (and now virtual) life, and that at the very least Meta can and should use surveillance tools to counter bullying and harassment. However, many were critical of similar forms of surveillance in member-only spaces. What is the point of having differential spaces if both are to be equally monitored? An Oceanian participant provided a compromise concept, where they believe surveillance devices can be “on and off, both private and public”; it can be extrapolated that the active utilization of surveillance devices doesn’t need to be either completely present nor never present, but perhaps formulated to be active at certain times random only or under a particular algorithm in private and public Metaverse spaces.

However, it was clear that some feared that with the intent of safety, these proposals for ASD and video capture could be utilized by platforms like Meta to market personalized advertisements. One participant asked whether or not the information gathered by Metaverse’s surveillance tools can gather “information [to] be sold to companies to monetize our opinions and eventually sell things to us.” In the same North American deliberation, a different participant noted this duality of safety and the extent of information monetization, saying they “want to be safe” but “don’t want people selling me stuff based on me having conversations with other people.” Even if personal data about Metaverse users weren’t monetized, some participants were cynical about their data being collected in the first place. Whether it is to happen or not, the possible reality of information collection and monetization under surveillance devices brought up cynicism amongst participants who overwhelmingly seemed supportive of surveillance tools in the Metaverse.

Additionally, the potentially consequential inability of ASD to sense banter and playfulness between Metaverse users was of concern to potential users; discourse on this topic was prevalent amongst deliberators from Asian countries. Participants fear the possibility of users falsely accused of bullying and harassment. What allows ASD or video capture to differentiate between an ill-spirited insult, gesture, slur, etc. and a vulgar but unintentionally offensive banter? Would Meta allow a user to use vulgar words if they were said with endearment? In a Taiwanese discussion, a participant noted that playful banter



between friends is reasonable, even if offensive or harmful words were present; if all parties involved in the banter don't feel insulted or harmed, then ASD should not flag it or bring about consequences. An Indian participant feared how often ASD would fail because of an inability to make this distinction. In the same group, another Indian participant noted how "slang" words exist in different languages, and that offensive terms might also be used as terms of endearment. In a Korean deliberation, a participant noted how participants who may speak informally or use niche Korean slang can avoid getting flagged; they questioned whether a "voice kanji system" would need to be implemented to understand slang/non-formal Korean phrases. A possible understanding of private, member-only spaces might be that they are places where perhaps offensive banter can exist. One Korean participant saw member-only spaces in the Metaverse as a space that "is for my close friends to joke with each other" and be more comfortable compared to perhaps restraining themselves in a surveillance public space.

Some participants noted that a consequence of utilizing Meta's surveillance tools could be that it might flag participants who objectively did nothing offensive or outrageous. In an Oceanian deliberation, a participant noted that a Metaverse user "could be doing not anything illegal, not anything rude or mean" but "they're just weird and they [are] just expressing themselves but in a way that they would never do in public"; by flagging and punishing that behavior, this participant concluded that they would be limiting their behavior, even when there's nothing worth surveilling. While the conversation conceded that it is important to record what Metaverse users say and do, we might be wary of inaccurate detections and unintended consequences for users and Meta itself.

"[My wife and I] have an [Amazon] Alexa and, you know, that's listening because the next day, if you talked about an item, that would show up on my wife's [Facebook] or, you know, on...pop-up ads, it's creepy." - Participant 17331, voicing concern over tailored advertisements due to surveillance (20221203-meta-enna-1yc-b\_transcript)

Session 2: Who is in the best position to look out for bullying and harassment in different kinds of spaces?

In this session, participants deliberated on who should be responsible for looking out for bullying and harassment in the Metaverse, including using the tools we discussed in Session 1. Participants discussed the validity of allowing both visible and invisible Metaverse moderators to enhance surveillance of bullying and harassment. Participants also discussed the roles and responsibilities, if any, platform owners and content creators have to monitor bullying and harassment. 25 proposals were discussed in Session 2. Some questions addressed in Session 2 were:

1. Who is primarily responsible for addressing bullying and harassment?
2. Who should use automatic speech detection?
3. Should visible moderators be allowed?
4. Should invisible moderators be allowed?

How did participants relate to video capture, automatic speech detection, creators, platform owners, and moderators?

Platform owners and creators were seen as owners of a virtual space, which might exist for a particular reason (classroom, conference room, etc.). Participants noted that those responsible for a Metaverse room can be legal owners just like those who own a physical space. One participant in England noted how a creator or platform owner is like a landlord of a British pub; such a person is “responsible for making sure fights don't break... he doesn't monitor every conversation. He just sits in the corner, but at the end of the day the brewery is responsible. In a sense, a platform owner or creator, like a pub owner, cannot monitor every single action or discussion occurring in their space but should execute major decisions if needed. In a Taiwanese deliberation, one participant noted how a private space in the Metaverse is like their “own club, [their own leisurely space] where it is up to them to “decide the standards in this space,” similar to how a nightclub can enforce standards on who enters the space (i.e., entrance fee, appropriate clothing, etc.)

Moderators, both invisible and visible, were conceptualized as police officers in the real world. Moderators whether AI or as a person could analyze offenses Metaverse users made, collect evidence, and enact disciplinary punishment, like how an officer can ticket/fine people in the real world.

In one deliberation, a participant feared the unintended consequences of false flags noted by ASD. Can the AI of ASD note when it makes a mistake, and how is that mistake rectified; does a “morally questionable gray zone” exist for AI to ponder? A North American participant alluded in their deliberation to an incident where an older friend was removed from an organization, they were in for 40+ years because they used a word which was not accepted in that organization. Both in reality and virtual reality, and in different languages and cultures, words and phrases mean different things to people and in different situations. Similarly, this participant feared that people’s lives could be changed because of a misunderstanding of a word/phrase in the Metaverse by ASD.

“There was an instance where a friend of mine actually had a specific word used against him and they were actually thrown out of an organization that they were in for like 40 years without discussion. One way or another, because he was an older [man], gently put it that way, and he hadn't gotten the playbook yet, okay? But it wasn't a word that was really bad.”-Participant 21880, wary of how ASD can lead to major consequences if people don’t know what words and phrases are and are not acceptable in the Metaverse (20221208-meta-enna-2yc-o\_transcript)

Similar and Different Views in Session 2:

An overarching view throughout Session 2 was that there was a need for a transparent and clear relationship between platform owners and creators; this is especially true in public spaces in the Metaverse. Their responsibilities intersected, as noted by various participants; the proposals regarding responsibility were seen as false binary choices for many deliberation participants. However, many

participants concede that, ultimately, it is the platform owner who holds the most responsibility in this dynamic when major problems arise, and drastic decisions must be made. One British participant noted the dynamic as a descending staircase, a “main responsibility and then on a staircase that is downwards.” A Japanese participant held an overlapping view, asserting that “creators have a responsibility in that narrow, small area...but the platform owner...should be responsible for the whole [Metaverse space.]”. Arguments that support such a dynamic include when creators are failing to address issues, as they may unfortunately be “incompetent and not in a position to oversee their [Metaverse space]” as one Oceania participant noted; creators may also be bullies, and they may need monitoring in their own spaces by a higher power for the sake of other users.

Some did voice concern over platform owners being absolutely responsible. The sheer size and foreseeable growth of the Metaverse makes it practically unfeasible for one party to note every issue. One Korean participant still found the Metaverse as a novel idea, and started that “Realistically, it's not an extremely old and advanced platform yet to manage all sessions on the platform... not enough staff can't catch... realistically troubled sessions.” This participant finds the fundamental idea of platform owners themselves to be responsible as impractical and unwarranted.

Discussions about the presence of Metaverse moderators seemed to revolve around the oddness of an invisible moderator monitoring a user and their violation of personal space. Participants felt that the presence of an invisible moderator, some who you cannot see but they can see you, made them uneasy (compared to a visible moderator) and were thus critical of the concept. One North American participant asserted that “it feels creepy invisible moderators” being able to watch their every move in the Metaverse.

Another participant builds on the criticism of the presence of invisible moderators. What was their purpose? Something different than a visible moderator or ASD? In the same North American deliberation, participants pondered what it was that an invisible moderator would counter. An American participant noted that if something problematic was occurring in the Metaverse (they used the dramatic example of terrorism), invisible moderators wouldn't be needed; they noted that those responsible for addressing dramatic issues are “called the FBI and the CIA and... Homeland Security. We have stuff for that, so we don't need to have invisible moderators to do that kind of thing.” While this is an extreme example, it serves to highlight the view that invisible moderators at the very least seem like an unnecessary and unscrupulous layer of surveillance by Meta. It is possible that harmful and illegal activity on online spaces can have national security ramifications; this is perhaps best exemplified by the April arrest of a 21-year old member of the Massachusetts Air National Guard who leaked classified U.S. national intelligence information on a Discord server dedicated to the video game “*Minecraft*.” A British participant overlaps with this view but utilizes this argument to defend visible moderators by crafting a similarly extreme example. This participant defends the presence of visible moderators irrespective of a public or member-only space; they argue the hypothetical that a Metaverse space may be a breeding ground for “a group making bombs or something like that” and asks fellow participants if they think “their platform should be hosting illegal activity.” Their argument explores the role in which a visible

moderator, while perhaps powerless on its own, has the agency and information to notify proper external channels and law enforcement organizations to prevent extremism in the Metaverse.

“I think that bullying and harassment are actually very difficult to define. Maybe it's about repetition, maybe it's about three or more times, and maybe being tagged or reported could also be a factor.”- Participant 23927, addressing concern and confusion about what Meta defines as bullying and harassment (2yc-a)

Session 3: What should be done in response to bullying and harassment?

In this session, participants discussed who should decide whether reported conduct qualifies as bullying and harassment after a report is made, and what (if any) action should be taken. Since not every single instance of bullying and harassment can be detected and recorded by tools like ASD and video capture, it is important to see what is needed in a report for consequences to be assigned and what the consequences should be too. 13 proposals were discussed in Session 3.

Some questions addressed in Session 3 were:

1. How should people report bullying and harassment in Metaverse spaces?
2. How should bullying and harassment reports be reviewed, and by whom?
3. What should platform owners do when they receive reports of bullying and harassment?

How did participants relate to video capture, automatic speech detection, creators, platform owners, and moderators?

A British deliberation compared the process of notifying and rectifying bullying and harassment in the Metaverse similar to that of an HR complaint in the real-world workplace. At least six British participants used an analogy to real-life management of problems in the workplace—bring it up with the perpetrator, and if that doesn't work, escalate to workplace superiors. One participant, a former trade union representative, noted that when a workplace accident occurs, “the person was asked to take it up with the individual...and see [the] response and if the response was negative then they could sort of kick it upstairs,” as in bringing the issue up to a manager or workplace superior. Likewise, another British participant saw the rectification of bullying as something “similar to the way it works in real life”; the first action to rectify an issue is “speak [to] the person that's causing the problem and say I'm not comfortable with that, please stop and then you can report it [and there] should be an appeal process for both people.”

Similar and Different Views in Session 3:

Like sentiment present in Session 2, participants across the globe believe a working relationship must exist between platform owners and creators when responding to bullying and harassment in the Metaverse. There is a sense that both sides must be aware of bullying and keeping information from

other parties is irresponsible and distrustful. As a Japanese participant noted, “it is necessary to deliver that information [of bullying and harassment] to both sides, not just one or the other.” However, there is an overwhelming sense that platform owners have an outsized role in addressing bullying and harassment compared to creators. In North America, participants assert that platform owners should take the lead on managing bullying but concede that creators should be kept updated and consulted upon. Many Asian participants believed that creators should be the first and main party responsible for addressing bullying. Korean deliberators, in the spirit of collaboration, noted that it should ultimately be platform owners who handle cases of bullying and harassment and laying groundwork for which creators can prevent such actions in their Metaverse spaces in the future. In Taiwan, a participant noted that the “main responsibility should be taken by the owner of the platform” to formulate a “long-term and creative plan” to address bullying and harassment.

Similarly, other participants defend the outsized role of platform owners by highlighting the impracticality of letting creators have too much responsibility. British participants worried that “giving the creators too much power and especially...if they haven't had any training...would be inconsistent” when addressing bullying; another argued that creators might not be actively able to address every instance of alleged bullying compared to well-resourced and well-staffed platform owners. North American deliberators note the amount of information and evidence needed to be investigated when conducting these cases, and that it is unrealistic for a single creator to do all of the work.

However, an African participant highlights a reverse argument regarding impartiality. They argued that creators, more so than platform owners, should handle reports of bullying because platform owners may either have to deal with large amounts of complaints or have other issues to deal with. Creators will have more time to address these issues, and users may feel more comfortable knowing creators are addressing these issues rather than an overseeing platform owner or owners. German deliberators also felt the view that creators should be the first and main party responsible for addressing bullying and harassment. One participant felt that a creator is the first point of contact for a user in the Metaverse and that contacting platform owners would be a drastic “last resort” to address issues. Similarly, another German participant felt that responsibility should be aimed towards creators who can decide who and how to address issues, and then having platform owners come in if creators cannot resolve them.

“Actually, maintaining a really big public space is a really big job, and I think it might be too big of a job for an individual creator, sometimes. So, the platform ultimately being responsible, I think makes the most sense.”- Participant 17708, highlighting the impracticality of individual creators addressing bullying and harassment (20221204-meta-enna-3nc-a\_transcript)

There seems to be misunderstanding, and consequently confusion, about what behavior is expected from Metaverse users. There needs to be a standardized framework or guidelines that all users must agree upon so that once bullying and harassment reports are brought to creators or platform owners, there are criteria in which these reports must meet. There must be, as a Taiwanese deliberator noted, a “unified approach” to make addressing bullying and harassment “more fair and effective.” Various British deliberators agree with a sense of standardization and clear definition of what bullying and harassment

is and isn't. One participant argued that there must be clear and transparent rules that users must follow in both member-only and public spaces; another argued that even before becoming a Metaverse user, one must understand and agree to what is classified as bullying and not, similar to how one agrees to terms and conditions before making an account for a website or organization.

There were questions about the laws of a given country and how law enforcement can protect users when they're grossly violated. The German deliberations, for example, produced an interesting "triangular relationship" to address extreme and illegal forms of bullying and harassment, with creators, platform owners and law enforcement as interworking parties. Certain activities, like sexual harassment, will need to have law enforcement and legal authorities to be involved. One participant passionately made such a case: "it doesn't have to be difficult, ultimately, its criminal prosecution, a criminal offense...I'm of the opinion that the creator doesn't actually have to be familiar with the criminal code, that's [for] the public prosecutor's office. I think he should perhaps know the basic data protection regulation and no more." It could be the responsibility of both creators and platform owners to understand what should and shouldn't be worth notifying to law enforcement on their platforms.

Furthermore, a Japanese deliberation in this session pondered over what counts as "Personal Identification" in the Metaverse, and how laws meant to protect one's identity are implemented in the Metaverse. How can Meta explain to users who is who when people have alias or profile handle names? In the real world, it was noted that a Japanese citizen is protected under Japan's *Act on the Protection of Personal Information (APPI) Law*. How does this law, which is meant to protect someone's individual identity, compare to online activity in the Metaverse if someone's "handle name is linked to 5 [different] people?"

Session 4: What should be the consequences for creators of spaces where bullying and harassment happens repeatedly?

It is possible that first time offenders of bullying and harassment in the Metaverse will continue to act egregiously on the platform. Even with best efforts by platform owners and creators, repeat offenses can happen in public and private spaces. Therefore, participants in this session deliberated on what the consequences should be for creators of spaces where bullying and harassment happens repeatedly. There were 10 proposals discussed in Session 4.

How did participants relate to video capture, automatic speech detection, creators, platform owners, and moderators?

Several comparisons were drawn to existing features for Facebook creators looking to moderate public and private groups. Some participants in Oceania pointed to preemptive measures like applications to help filter potential users by requiring why they wanted to join. This was seen as similar to the process many private Facebook groups utilize when looking to ensure that only their intended specific group of people can enter. Compared to other sessions, there were very few allusions to parallel, current measures to prevent bullying and harassment on the part of the creator, or actions that should be taken

against a creator if their group continues to be in violation of regulation. Perhaps this was because the question was phrased more as a what should happen, rather than a review of specific features.

### Similar and Different Views in Session 4:

Participants emphasized the importance of implementing features to aid users in understanding if a space has harmful language and conduct, and to what extent. The usage of a warning prior to entering a space was suggested in various deliberations. By warning users, they would be aware of behavior expectations and regulations by platform owners and creators. In North America, one participant presented the idea of an informational warning message that would be displayed before the user were even to join a group, stating “it's pretty much like a disclaimer, like you're letting the users know what's going on ahead of time...so they have the option to opt in or out.” A UK participant offered that adding something like a color-code system could be used as an indicator.

Three deliberators in Japan made the same recommendation, calling it a “notice” to be given to users before joining that bullying had been reported in that space. One participant went as far to say that users should be actively notified if incidents had transpired while in the group: “yes, yes, I think so too, and I think users should receive notifications...even if there is, on the contrary, in the sense that it's good to know something like this, well, I think it's the right to know, the right to know the meaning.” This could be in addition to the incident counter suggested by a participant in Sub-Saharan Africa. It would display the total number of bullying and harassment incidents, with the counter updated with each subsequent report.

The person creating the group would be also responsible for “attaching” each member to the rules and regulations as soon as possible after entering a group, another participant from Sub-Saharan Africa advocated.

In an Indian deliberation, a star rating system was another specific feature recommended. This could help someone gauge how frequent or severe bullying or harassment might be based on reviews from users with experience in the group, a participant from India explained as their rationale. All three participants agreed that there should be some way to know before joining a group and to be notified while in the group if harassment, especially sexual harassment, had taken place. One deliberator said this is also a preventative feature. Rather than just a rating system based on the reports filed from the group's inception, that same deliberator in India suggested that the number and nature of new reports filed should be published every weekend.

Deliberators across all countries surveyed agreed users should be empowered to make informed choices about whether to join a group or not. More than simply reacting to harm, participants deeply valued preventing harm. Overwhelmingly, the creator was seen as having hefty responsibility for regulation. Many small groups emphasized the role of regulation falling primarily onto the creator, and that it is not solely for individual users to have to handle issues of harassment and bullying between themselves and Meta. As a participant in Sub-Saharan Africa noted, creators should be aware that if they “violate the rules...[their space] will be brought down as a result of bullying and harassment under their supervision.



This participant highlights the reoccurring sentiment that responsibility includes preventing their groups from becoming unmanageable, from as mentioned above screening potential new members to removing those that had refused to abide by regulations.

This can be seen by a participant in Japan saying they “also think it depends on the degree...it is repeated over and over again, I think there are various problems with the space, so I think it is necessary to take measures against the creator.” Moreover, those asked supported taking actions against creators allowing bullying and harassment on their platform. They also “think that strict measures should be gradually taken for creators who don't take action against them even after giving various warnings, but I think that it should be changed according to the level.” A standardized, yet incremental, approach was preferred. There was consensus that the actions taken against a creator need to be gradual and in proportion to the severity of the infractions occurring within their group.

All regions had participants that stressed that creators are ultimately responsible for their groups. While people from a variety of geographic backgrounds could agree on this point, in what scenarios 1) if there were extenuating circumstances that should lessen or waive reprimand and 2) if a group should be taken down entirely were more contentious. Some argued that while ultimately responsible, more tolerance of unregulated spaces should be considered for why they lacked regulation, in the case of someone trying but failing to keep up the volume of people and incidents causing harm. “The consequences of allowing bullying should depend on the reason why it's happened, so if it's just that someone is negligent...and they don't have policies and procedures in place the consequences should be more severe for example [than if] they're struggling for resources,” offered one participant from the UK. This quote supports the idea that punishment should not only be proportional to the infractions, but also with a creator's capacity to respond.

A similar line of reasoning, some expressed hesitation as to if an entire group should be taken down if it were to break a certain threshold or rating for a space deemed too harmful. A participant from Oceania expressed that “I don't know closing the space is really fair to the people that might be otherwise actually enjoying themselves...you might be there with some great friends, forming friendships and socializing and then all of a sudden you know because of something that someone else then they been punished and it's you know. it's a little bit unfair in that way.” Policy could inadvertently harm meaningful connections between groups members that were not involved with bullying or harassment.

This concern seems to have a similar nuance in how to address groups with members acting poorly but seems contradictory to the participant in Japan who argues it is more of a community responsibility than a creator-based responsibility. If one were to target creators with reprimand, then it can put blame on them for actions they are actively working, and struggling, to curb violations. However, if the policy was to target groups rather than a creator of a group, then perhaps this leads to adverse consequences for those not committing regulation violations. Even more complicated still, far more participants, as illustrated by their consensus, believe that if a creator did not regulate a group well enough, then it is still their responsibility when it becomes overrun by those violating regulations. They would argue that if the space cannot be managed, to paraphrase the participant from Sub-Saharan Africa, it should not exist.



This exposes a conundrum that is not easily ameliorated without enough moderators and resources to properly target reprimands at those causing the harm, rather than a group creator or group as a whole. A potential incremental scale of reprimand could be implemented to first target individual users, then the group creator, and then the group as a whole.

### Deliberative Quality Insights

The quality of deliberation is an important component to the success of any Deliberative Poll. In deliberative theory, many critics of deliberation have raised concerns about the quality of deliberation being impacted by those who may be more privileged (generally, those that are white, educated and/or male) and the quality of arguments offered by participants. For this Community Forum, the process of analyzing deliberative qualitative is extensive as it requires examining all 2,069 transcripts and given the limited time, DDL has only processed a random subset of the transcripts. Therefore, in this section, DDL focuses on insights regarding the number of statements and the number of pros and cons offered in the small groups for a random subset of transcripts from a few countries.

Over the last few years, DDL, in collaboration with Fileread.ai, designed an AI-text analysis platform called Argument Miner. This platform is designed to identify the number of statements made by each individual, identify the pro and con arguments within each statement, and summarizes the discussions of the individuals and the groups for us to better understand and more quickly assess the quality of deliberation in each of the groups. The platform is trained on transcripts from past Deliberative Polls, while building on the leading industry technologies.

Even with Argument Miner, DDL was only able to process a random subset of the over 2,000 transcripts from across the globe. For non-English transcripts, DDL worked with native speakers to use chatGPT to translate the transcripts into English. Native speakers cross checked the chatGPT-translated transcripts to ensure accuracy. The translated transcripts were then placed into Argument Miner for analysis. The outputs from Argument Miner were then examined by native speakers to ensure accuracy.

The analysis below shows a random subset of small groups within North America, Oceania, United Kingdom, and Taiwan. The analysis shows their total number of statements and arguments made per small groups in each region/country's deliberation, along with the average number of statements and arguments made per small group. The assessment of pro and con statements is assessed based on the following proposal questions.

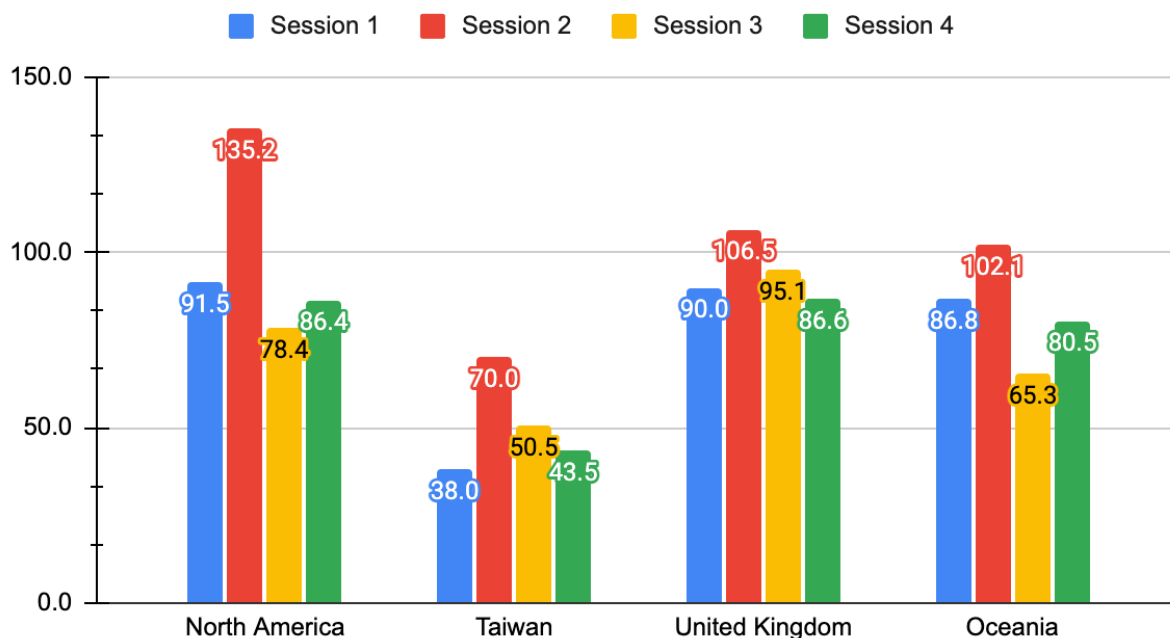
1. Should video capture be used in members-only spaces or public spaces?
2. Should automatic speech detection be used in members-only spaces or public spaces?
3. Who should be responsible for bad behavior in members-only, creators or platform owners?
4. Who should be responsible for bad behavior in public spaces, creators, or platform owners?
5. What kind of punishments should there be for bad behavior?

The analysis did not use the exact policy proposal from the deliberations because there was overlap in the deliberation proposals and these questions captured the key proposals.

First, Argument Miner identified the number of statements offered in each of these countries by session. The chart below illustrates the average number of statements made in the respective countries and regions by the four sessions. Overall, participants from North America, United Kingdom and Oceania made relatively similar amounts of statements, whereas participants from Taiwan made substantially less statements across the four sessions. For North America, the second session about who is in the best position to look out for bullying and harassment yielded the most statements among the four sessions. For the United Kingdom and Oceania, session two was also the highest number of average statements, but the other three sessions were not too far behind. While Taiwan participants made substantially less statements, their average number for session two was also the highest amongst the four sessions.

A further analysis that DDL is working on is to overlay various demographics when examining the average number of statements. In past research, DDL has examined whether more privileged participants made more or less statements. Past findings suggest that participants regardless of gender, age, race, and other demographics differ very little in the number of statements contributed to the discussion. More important to the number of statements is actually which statements drive opinion change, which will also be the next step in this qualitative research.

### Average Number of Statements by Country/Region



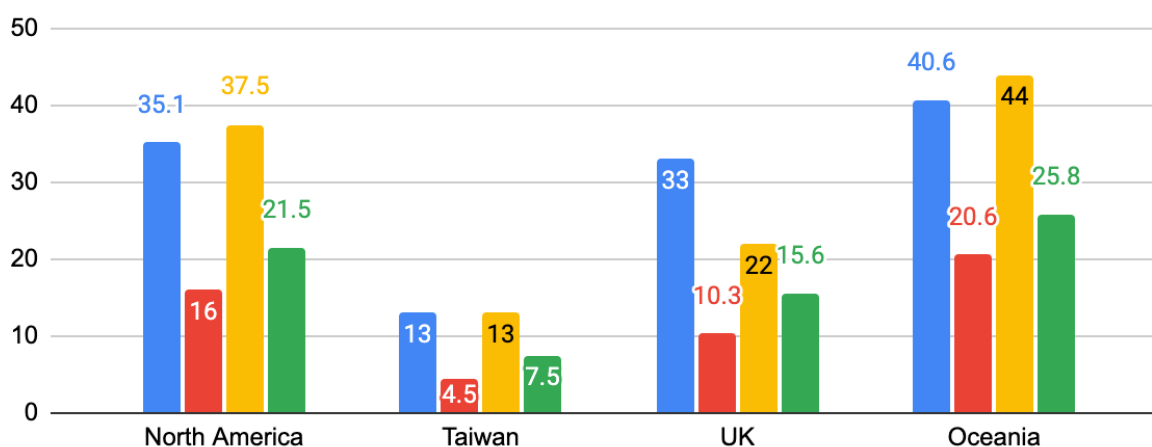
Now, for each of the sessions, Argument Miner identified the number of pro and con arguments made by each individual. Here, in the charts below, this section summarizes on average how many pro and con

arguments were made in each session. In further analysis, each individual's qualitative contribution to the discussion is quantified and can be integrated into the quantitative dataset to further examine the effect of the pro and con arguments on opinion change.

For now, this section provides a look at the number of pro and con arguments offered by each country and region. On the topic of what types of tools should be used, participants discussed the tools that could be used in these social virtual spaces. There were more arguments for than against, when discussing video capture and automatic speech detection. For some countries, participants made double the amount of pro arguments than con arguments. For arguments against these tools, participants from Oceania and North American roughly made about 20-25 arguments in each group. Participants in the UK made about 15 con arguments per group and 22 for con arguments on automatic speech detection. Examples of these arguments were used in the qualitative analysis above in providing examples and quotes from participants in [Session 1](#).

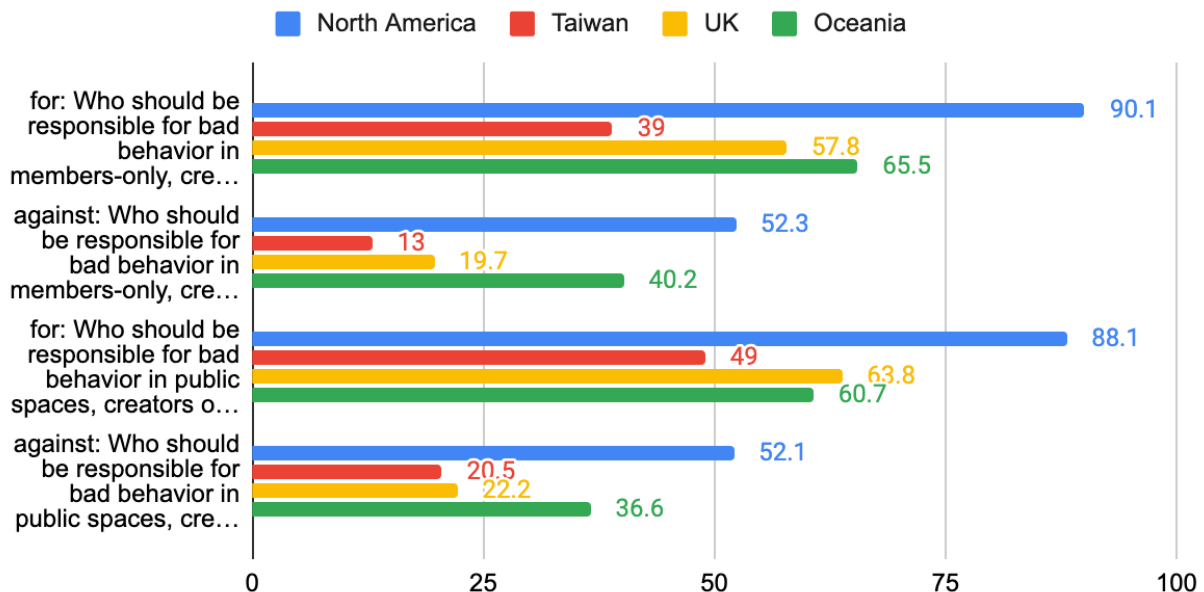
## Average Number of Pro and Con Arguments Offered - What tools to use?

- for: Should video capture be used in members-only spaces or public spaces?
- against: Should video capture be used in members-only spaces or public spaces?
- for: Should automatic speech detection be used in members-only spaces or public spaces?
- against: Should automatic speech detection be used in members-only spaces or public spaces?



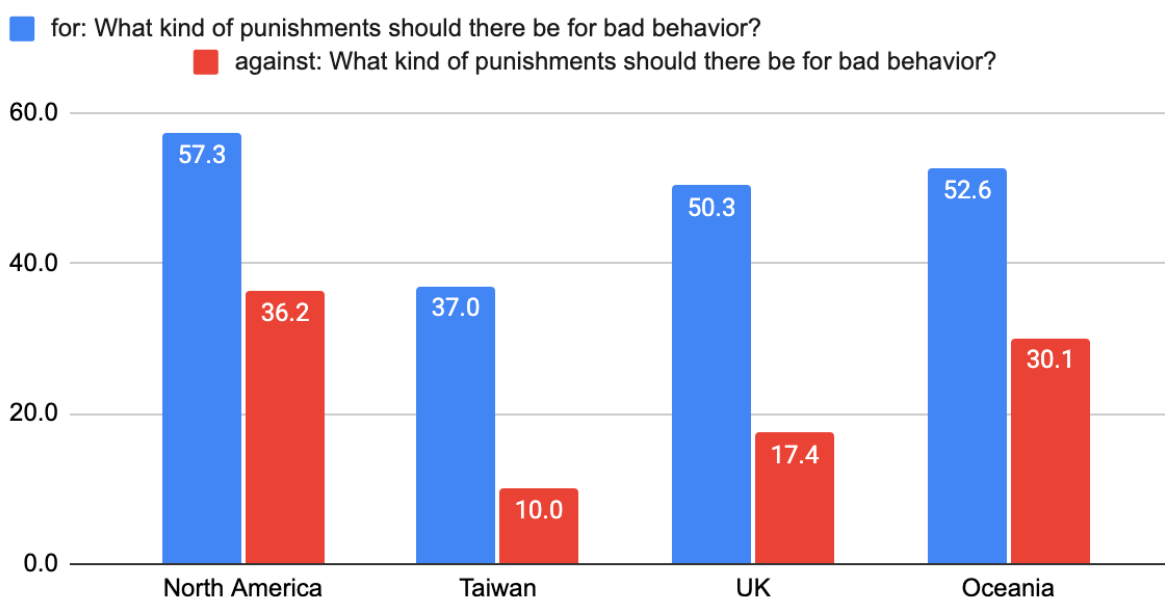
With respect to who should be primarily responsible for bad behavior, there were similar patterns as above with some differences. In terms of the number of pro and cons statements, the North American participants consistently contributed close to 90 pro arguments per group and 50 con arguments per group. This pattern of contributing more arguments for than against was also present among the other countries. For Taiwanese participants, they appeared to have more pro arguments to share regarding 'who should be responsible for bad behavior in public spaces. And, in general, Taiwanese participants had more to say about topics regarding 'who should be responsible', as opposed to whether and where to use automatic speech detection or video capture. Examples of these arguments were used in the qualitative analysis above in providing examples and quotes from participants in [Session 2](#).

## Average Number of Pro and Con Arguments Offers - Who should be responsible



On the issue of types of punishments for bad behavior, participants from North America, UK, and Oceania contributed roughly the same number of arguments for each of the policy issues, between 50 and 60 per group. The number of con arguments for each of the policy issues varied by country. North American participants contributed about 35 arguments against per group, UK participants contributed about 18 participants per group, and Oceania participants about 25 arguments against per group. Taiwanese participants varied for the average number of arguments against, with a range of 7 to 22 arguments. Taiwanese participants had the most arguments against the issue of automatic speech detection for this session. Examples of these arguments were used in the qualitative analysis above in providing examples and quotes from participants in [Session 4](#).

## Average Pro and Con Statements for What kind of punishments should there be for bad behavior?



As a note, Argument Miner is continually updated, and it is possible that the platform may make mistakes in what it considers to be a relevant argument. DDL regularly performs comparison with human coders on a random subset of transcripts to evaluate consistency. In future iterations of this analysis, DDL plans to examine all 2,069 transcripts and provide more thorough analysis of participants' arguments and why they are in support or opposition of these policy issues.

Raw Data:

[Argument Miner Results - Random Subset](#)  
[Total Number of Metaverse Small Groups](#)

### Implications and Next Steps

The findings in this report demonstrated that the Community Forum produced statistically significant opinion changes and that the opinions, irrespective of change, were considered and thoughtful. Around the world and across demographics, while there were some proposals that participants may not have been in complete agreement on, on the whole, participants were generally on the same side of the scale, showing agreement and understanding on the options put forth to them. Some of the results had clear directives, globally, on what participants preferred, such as not being punitive, while for other results, participants had more varied results globally, such as views towards applications of the tools. More

important, since decisions are to be made globally about the Metaverse, it is clear, overall, that participants viewed the policy proposals in a favorable fashion.

While this report presented detailed quantitative and qualitative results from the Community Forum, DDL will continue to study and explain *why* participants held certain opinions and better understand the impacts of group processes (if any) that were present in the deliberations. For example, the quality of arguments offered may be a predictor in whether and how much participants shifted their opinions. After processing all 2,069 transcripts using DDL's AI text analysis platform, the results can determine the types of arguments offered and how they correlate with the opinion changes. In addition, a deeper level of study into the regional differences, while it may not affect global policy, could shed light on whether certain regions and/or countries would have stronger reactions or opinions about certain policy proposals. Being able to understand their specific concerns and arguments for or against these proposals, would allow for better anticipation of any issues that may arise.

And, to think beyond this deliberative event, in future iterations, whether smaller or at an even larger scale, it is important to consider the lessons learned from this event. If there were an opportunity to run another global event, DDL would offer three lessons learned. First, after working with 14 different polling firms for 32 countries and 19 languages, the importance of having polling firms that offer a 'concierge service' is highly recommended for this process. Especially in countries that culturally rely more on in-person or person-to-person relationships, the polling firms that did well went above and beyond to ensure participants felt comfortable and were ready to attend. Of course, a concierge service in any country requires more resources, but it would be worth the discussion next time. Second, having plenary sessions with fewer countries. While it was remarkable for some of the sessions to have had over 1000 participants, it also meant very few participants' questions were able to be asked. DDL has found that while participants always enjoy and rate the small group sessions more highly, the plenary sessions actually serve as a way for participants' voices to be heard. When participants hear their own or their groups' questions being addressed, there is a sense of pride and acknowledgement that participants gain. This is a critical part of the deliberative process for making sure participants feel they are being listened to and taken seriously. Lastly, having more time. This event was accomplished from beginning to end in about six months, including conception of the design, development of briefing materials, development of survey, recruitment to execution of the event. While timing was of the essence for this project, perhaps there are ways to consider helping Meta think about building these deliberative processes into some key decision-making processes, which will allow for more time and potentially more resources.

The Community Forum was the first-of-its-kind for Meta, DDL, and the entire deliberative democracy field. This deliberative event demonstrated how the entire world can come together practically to deliberate on important policy issues and the results certainly speak to the importance of this event to the over 6,000 participants from around the world. DDL is certain that many around the world anticipate the release of the results from the Community Forum.



## Appendix

### A. Assessment of Data Source: DataReportal Statistics

#### 1. What is DataReportal?

[DataReportal](#) is a website that offers reports on digital trends. The reports are produced mainly by Kepios, a for-profit advisory firm which aims to help organizations and companies better understand the digital world.

#### 2. Why should we use DataReportal?

DataReportal offers the most comprehensive reports on digital trends. They have the most countries and measures compared to any other reports. They are the leaders in their field and are widely used all around the internet. In the case of social media data, most websites source their data from DataReportal.

#### 3. Why should we trust DataReportal?

DataReportal clearly indicates their sources for each data. For social media, their data are for the most part sourced directly from social media companies. DataReportal simply combines together the data of individual social media companies.

#### 4. Conclusion

DataReportal offers the most accessible and comprehensive data on the use of social media worldwide.

### B. Demographic Results

#### Gender

The participant group included 3.397 women and 3.001 men (total n = 6.398).

The control group consisted of 3.291 women and 3.066 men (total n = 6.357).

#### I. Who should be responsible for addressing bullying and harassment in social VR spaces?

Responsibility for Addressing Bullying and Harassment in social VR spaces? (Mean Participant Responses)									
		Gender			Gender			Gender	
Proposal	T1	F	M	T2	F	M	T2-T1	F	M



## COMMUNITY FORUM - RESULTS ANALYSIS

1) Creators should be primarily responsible for monitoring public spaces they create.	7.2	7.4	7.1	7.1	7.3	7.0	-0.132 (0.002)	-0.086 (0.141)	-0.171 (0.005)
2) Creators should be primarily responsible for monitoring members-only spaces they create.	7.3	7.3	7.2	7.6	7.7	7.6	0.369 (0.000)	0.394 (0.000)	0.348 (0.000)
3) Platform owners should be primarily responsible for monitoring public spaces even if they're built by Creators.	7.6	7.7	7.4	7.9	8.0	7.8	0.293 (0.000)	0.243 (0.000)	0.334 (0.000)
4) Platform owners should be primarily responsible for monitoring members-only spaces even if they're built by Creators	7.3	7.5	7.2	7.2	7.5	7.1	-0.081 (0.050)	-0.031 (0.585)	-0.122 (0.043)

In the control group, men and women had overall positive ratings for these approaches. Women significantly decreased their support for Proposal 1 (7.2 to 7.0,  $p = 0.00$ ). Men significantly increased their support for Proposal 3 (7.2 to 7.3,  $p = .03$ ).

### II. To whom should cases of bullying and harassment be reported?

To Whom Should Bullying and Harassment Be Reported? (Mean Participant Responses)									
		Gender			Gender			Gender	
Proposal	T1	F	M	T2	F	M	T2-T1	F	M

## COMMUNITY FORUM - RESULTS ANALYSIS

1) In public spaces built by Creators, people should report to Creators, and Creators decide the consequences.	6.743	6.822	6.68	6.798	6.826	6.776	0.055 (0.220)	0.004 (0.944)	0.096 (0.136)
2) In public spaces built by Creators, people should report to platform owners, and platforms decide the consequences.	7.22	7.351	7.112	7.653	7.729	7.589	0.433 (0.000)	0.378 (0.000)	0.477 (0.000)
3) In public spaces built by Creators, Creators should send severe cases to platform owners for additional enforcement	7.737	7.915	7.59	8.31	8.433	8.208	0.573 (0.000)	0.518 (0.000)	0.618 (0.000)
4) In members-only spaces, people should report to Creators who built them, and Creators decide the consequences.	6.923	6.996	6.864	7.163	7.139	7.183	0.240 (0.000)	0.144 (0.017)	0.319 (0.000)
5) In members-only spaces, people should report to platform owners and platforms decide the consequences.	7.16	7.379	6.981	7.369	7.512	7.253	0.209 (0.000)	0.133 (0.018)	0.272 (0.000)
6) In members-only spaces, Creators should send severe cases to platform owners for additional enforcement.	7.765	7.934	7.626	8.255	8.385	8.149	0.490 (0.000)	0.451 (0.000)	0.523 (0.000)

Women and men in the control group had different changes in their responses from T1 to T2. Women did not significantly change any of their responses, on average, while men significantly increased their average support for Proposals 2, 5, and 6.

III. Should creators be made aware of reports made to platform owners?

Should Creators Be Notified of Reports to Platform Owners? (Mean Participant Responses)									
		Gender			Gender			Gender	
Proposal	T1	F	M	T2	F	M	T2-T1	F	M
1) Platform owners should not notify the creator of the space where it happened.	4.4	4.2	4.6	4.4	4.1	4.6	-0.080 (0.116)	-0.079 (0.262)	-0.080 (0.273)
2) Platform owners should notify the creator of the space where it happened, but not include personal information of the users involved.	6.6	6.5	6.7	7.1	7.1	7.1	0.483 (0.000)	0.564 (0.000)	0.415 (0.000)
3) Platform owners should notify the creator of the space where it happened with details about the event and the users involved.	7.0	7.2	6.8	6.9	7.0	6.8	-0.103 (0.030)	-0.146 (0.028)	-0.068 (0.317)

For the control group, women significantly increased their average support for Proposal 1 (4.4 to 4.6,  $p = 0.00$ ), while men significantly increased their average support for Proposal 1 (4.6 to 4.9,  $p = 0.00$ ) and Proposal 3 (6.8 to 7.0,  $p = 0.00$ ).

IV. What should be done about spaces where there is repeated bullying and/or harassment?

What Should Be Done About Spaces with Repeated Bullying and/or Harassment? (Mean Participant Responses)						
		Gender			Gender	

## COMMUNITY FORUM - RESULTS ANALYSIS

Proposal	T1	F	M	T2	F	M	T2-T1	F	M
1) In public spaces Creators built, platforms should take action against Creators.	6.4	6.3	6.5	6.7	6.6	6.7	0.277 (0.000)	0.327 (0.000)	0.238 (0.000)
2) In members-only spaces, platforms should take action against Creators.	6.4	6.3	6.4	6.9	6.8	6.9	0.493 (0.000)	0.533 (0.000)	0.461 (0.000)
3) Members-only spaces should be made less visible to users.	6.3	6.2	6.4	6.7	6.8	6.7	0.445 (0.000)	0.618 (0.000)	0.307 (0.000)
4) Members-only spaces should not be publicly discoverable.	6.2	6.1	6.3	6.6	6.6	6.5	0.327 (0.000)	0.458 (0.000)	0.223 (0.002)
5) Members-only spaces should be removed from the platform.	5.0	5.0	5.0	5.1	5.1	5.0	0.061 (0.266)	0.123 (0.111)	0.012 (0.875)
6) Creators should be banned from making additional members-only space	5.1	5.1	5.1	5.4	5.4	5.4	0.271 (0.000)	0.283 (0.000)	0.263 (0.001)
7) Creators should be banned from inviting additional people to join the space.	5.2	5.2	5.2	5.5	5.6	5.3	0.263 (0.000)	0.399 (0.000)	0.155 (0.049)
8) Creators should not be able to make money off of that space.	5.9	6.0	5.8	6.1	6.3	6.0	0.200 (0.000)	0.286 (0.000)	0.132 (0.089)

## COMMUNITY FORUM - RESULTS ANALYSIS

9) Creators should be required to take a course on how to moderate the spaces they create.	7.3	7.5	7.2	8.1	8.4	7.9	0.761 (0.000)	0.816 (0.000)	0.718 (0.000)
10) Users should receive notification of such cases when entering a space.	7.7	7.9	7.6	8.3	8.5	8.2	0.589 (0.000)	0.604 (0.000)	0.577 (0.000)

In the control group, women significantly increased their average support for Proposals 2, 3, and 5-8. Men significantly increased their average support for Proposals 1-6 and 9.

### V. What tools should be used to identify bullying and harassment in social VR spaces?

#### Video Capture

Should Video Capture Be Used to Identify Bullying and Harassment? (Mean Participant Responses)									
		Gender			Gender			Gender	
Proposal	T1	F	M	T2	F	M	T2-T1	F	M
1) Should be turned on in all public spaces.	6.4	6.5	6.3	7.2	7.5	7.0	0.845 (0.000)	0.984 (0.000)	0.731 (0.000)
2) Should be turned on in public spaces only where repeated bullying and harassment has occurred.	6.9	6.9	6.9	7.1	7.1	7.2	0.226 (0.000)	0.130 (0.053)	0.306 (0.000)
3) Should be turned on all members-only spaces.	6.1	6.3	5.9	6.3	6.6	6.0	0.218 (0.000)	0.304 (0.000)	0.148 (0.028)

## COMMUNITY FORUM - RESULTS ANALYSIS

4) Should be turned on in members-only spaces only where repeated bullying and harassment has occurred.	6.7	6.8	6.7	7.0	7.1	6.9	0.233 (0.000)	0.308 (0.000)	0.171 (0.010)
---	-----	-----	-----	-----	-----	-----	------------------	------------------	------------------

In the control group, women and men supported video capture, on average. Women significantly increased their average support for Proposals 1 and 3, while men significantly increased their average support for Proposals 1, 3, and 4.

### Automatic Speech Detection

Should Automatic Speech Detection Be Used to Identify Bullying and Harassment? (Mean Participant Responses)									
	Gender			Gender			Gender		
Proposal	T1	F	M	T2	F	M	T2-T1	F	M
1) Should be turned on in all public spaces.	6.6	6.9	6.4	7.1	7.4	6.9	0.491 (0.000)	0.502 (0.000)	0.482 (0.000)
2) Should be turned on in public spaces only where repeated bullying and/or harassment has occurred	6.9	6.9	6.8	7.1	7.0	7.1	0.196 (0.000)	0.019 (0.770)	0.342 (0.000)
3) Should be turned on in all members-only spaces.	6.2	6.5	6.0	6.3	6.6	6.0	0.083 (0.073)	0.112 (0.080)	0.059 (0.376)

4) Should be turned on in members-only spaces only where repeated bullying and/or harassment has occurred.	6.8	6.9	6.7	6.8	6.9	6.8	0.035 (0.463)	-0.015 (0.830)	0.076 (0.258)
--	-----	-----	-----	-----	-----	-----	------------------	-------------------	------------------

Males in the control group significantly increased their support for all Proposals, while women in the control group significantly increased their support for automatic speech detection in public and members-only spaces, Proposal 1 (6.5 to 6.7,  $p = 0.00$ ) and Proposal 3 (6.4 to 6.8,  $p = 0.00$ ).

#### VI. Who should review alerts?

Who Should Review Alerts from Automatic Speech Detection? (Mean Participant Responses)									
		Gender			Gender			Gender	
Proposal	T1	F	M	T2	F	M	T2-T1	F	M
1) Creators should review alerts from automated speech detection in public spaces they create.	7.3	7.6	7.1	7.4	7.6	7.1	0.061 (0.127)	0.038 (0.479)	0.080 (0.175)
2) Creators should review alerts from automated speech detection in public spaces they create but only when bullying and harassment has occurred there repeatedly.	7.0	7.1	7.0	7.1	7.1	7.2	0.132 (0.003)	0.054 (0.382)	0.197 (0.002)
3) Creators should review alerts from automated speech detection in their members-only spaces.	7.1	7.3	7.0	7.3	7.6	7.1	0.226 (0.000)	0.307 (0.000)	0.159 (0.010)

4) Creators should review alerts from automated speech detection in their members-only spaces but only when bullying and harassment has occurred there repeatedly	6.9	7.0	6.9	7.1	7.1	7.1	0.208 (0.000)	0.150 (0.015)	0.253 (0.000)
5) Platform owners should review alerts from automated speech detection in public spaces even if they're built by Creators.	7.5	7.8	7.4	7.7	7.9	7.6	0.157 (0.000)	0.113 (0.029)	0.194 (0.001)
6) Platform owners should review alerts from automated speech detection in public spaces even if they're built by Creators, but only when bullying and harassment has occurred there repeatedly.	7.0	7.0	7.0	7.4	7.4	7.4	0.403 (0.000)	0.399 (0.000)	0.407 (0.000)
7) Platform owners should review alerts from automated speech detection in members-only spaces.	7.2	7.4	7.0	7.2	7.4	6.9	-0.029 (0.471)	0.061 (0.276)	-0.105 (0.081)
8) Platform owners should review alerts from automated speech detection in members-only spaces, but only when bullying and harassment has occurred there repeatedly.	7.0	7.0	7.0	7.1	7.2	7.0	0.147 (0.001)	0.241 (0.000)	0.070 (0.269)

In the control group, women significantly increased their average support for Proposal 6 (6.8 to 7.0,  $p = 0.00$ ), while men significantly increased their average support for Proposals 1, 3, 6, and 8.

**VII. Should moderators be used to monitor bullying and harassment in social VR spaces?**

- Public Spaces
- When should moderators enter public spaces?
- Spaces with repeated bullying and harassment
- Members-only spaces
- Should creators be able to select other users to be visible moderators in their spaces?



## Visible Moderators

Should Visible Moderators Be Used to Monitor Bullying and Harassment? (Mean Participant Responses)									
		Gender			Gender			Gender	
Proposal	T1	F	M	T2	F	M	T2-T1	F	M
1) Visible moderators should be able to enter public spaces, even if they're built by Creators.	7.5	7.6	7.3	7.8	8.0	7.7	0.360 (0.000)	0.351 (0.000)	0.367 (0.000)
2) Visible moderators should be able to enter public spaces only when bullying and harassment has occurred repeatedly.	6.7	6.7	6.8	7.0	7.0	7.1	0.286 (0.000)	0.291 (0.000)	0.282 (0.000)
3) Visible moderators should be able to enter members-only spaces.	6.8	7.0	6.6	6.5	6.8	6.3	-0.255 (0.000)	-0.220 (0.000)	-0.284 (0.000)
4) Visible moderators should be able to enter members-only spaces only when bullying and harassment has occurred repeatedly.	6.7	6.6	6.7	6.9	6.9	6.8	0.196 (0.000)	0.285 (0.000)	0.123 (0.063)
5) Visible moderators should only enter public spaces when requested by Creators.	6.4	6.3	6.5	6.5	6.5	6.5	0.110 (0.026)	0.141 (0.043)	0.084 (0.228)
6) Visible moderators should only enter public spaces when requested by users.	6.2	6.1	6.2	6.6	6.5	6.6	0.397 (0.000)	0.423 (0.000)	0.376 (0.000)
7) Creators should be able to select other users to be visible moderators in their spaces.	7.2	7.2	7.1	7.3	7.3	7.3	0.125 (0.004)	0.030 (0.619)	0.200 (0.001)

In the control group, women significantly increased their average support for Proposal 6 (6.2 to 6.5,  $p = 0.00$ ), while men significantly increased their average support for Proposals 2, 5, and 6.

## Invisible Moderators

Should Invisible Moderators Be Used to Monitor Bullying and Harassment?  
(Mean Participant Responses)

## COMMUNITY FORUM - RESULTS ANALYSIS

		Gender			Gender			Gender	
Proposal	T1	F	M	T2	F	M	T2-T1	F	M
1) Invisible moderators should be able to enter all public spaces.	7.0	7.2	6.8	7.3	7.5	7.0	0.287 (0.000)	0.326 (0.000)	0.254 (0.000)
2) Invisible moderators should be able to enter all members-only spaces.	6.4	6.7	6.1	6.1	6.5	5.8	-0.215 (0.000)	-0.132 (0.048)	-0.283 (0.000)
3) Invisible moderators should be able to enter public spaces only when bullying and harassment has occurred repeatedly.	6.5	6.4	6.5	6.6	6.6	6.6	0.142 (0.004)	0.125 (0.063)	0.155 (0.029)
4) Invisible moderators should be able to enter members-only spaces only when bullying and harassment has occurred repeatedly.	6.4	6.4	6.4	6.5	6.5	6.5	0.075 (0.123)	0.079 (0.238)	0.071 (0.310)
5) Invisible moderators should only enter public spaces when requested by Creators.	6.2	6.2	6.3	6.2	6.2	6.3	0.003 (0.950)	0.026 (0.709)	-0.015 (0.831)
6) Invisible moderators should only enter public spaces when requested by users	6.3	6.2	6.3	6.3	6.2	6.3	0.021 (0.673)	0.037 (0.582)	0.008 (0.916)

In the control group, women significantly increased their average support for Proposals 2 and 4-6. Men significantly increased their average level of support for all approaches to using invisible moderators.

### Age

For the deliberation group, the distribution of participants by age group was:

1	2	3	4	5	Total
18-24	25-34	35-44	45-54	Over 54	
n = 1218	n = 1691	n = 1233	n = 1065	n = 1191	n = 6398

The control group had the following age distribution:

## COMMUNITY FORUM - RESULTS ANALYSIS

1	2	3	4	5	Total
18-24	25-34	35-44	45-54	Over 54	
n = 923	n = 1523	n = 1304	n = 1058	n = 1549	n = 6357

### I. Who should be responsible for addressing bullying and harassment in social VR spaces?

Who Should Be Responsible for Addressing Bullying and Harassment in social VR spaces? (Mean Participant Responses)																		
	Age Group						Age Group						Age Group					
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) Creators should be primarily responsible for monitoring public spaces they create.	7.2	7.1	7.3	7.4	7.4	7.2	7.1	7.2	7.1	7.2	7.2	6.8	-0.132 (0.002)	0.115 (0.207)	-0.229 (0.006)	-0.121 (0.190)	-0.194 (0.054)	-0.406 (0.000)
2) Creators should be primarily responsible for monitoring members-only spaces they create.	7.3	7.1	7.2	7.5	7.4	7.2	7.6	7.7	7.7	7.6	7.8	7.3	0.369 (0.000)	0.624 (0.000)	0.430 (0.000)	0.121 (0.178)	0.363 (0.000)	0.023 (0.832)
3) Platform owners should be primarily responsible for monitoring public spaces even if they're built by Creators.	7.6	7.4	7.5	7.7	7.7	7.8	7.9	7.8	7.9	7.9	8.0	7.9	0.293 (0.000)	0.413 (0.000)	0.361 (0.000)	0.164 (0.047)	0.290 (0.001)	0.048 (0.597)
4) Platform owners should be primarily responsible for monitoring members-only spaces even if they're built by Creators	7.3	7.1	7.2	7.5	7.5	7.6	7.2	7.1	7.2	7.2	7.5	7.3	-0.081 (0.050)	-0.004 (0.962)	0.018 (0.830)	-0.274 (0.003)	-0.049 (0.634)	-0.238 (0.015)

For the control group, people aged 25-34 significantly increased their average support for Proposals 1 and 4. In the 35-44 bucket, average support for Proposal 1 increased significantly. For people over 54, average support for Proposal 2 decreased significantly.

## II. To whom should cases of bullying and harassment be reported to?

Who Should Be Responsible for Addressing Bullying and Harassment in social VR spaces? (Mean Participant Responses)																		
	Age Group						Age Group						Age Group					
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) In public spaces built by Creators, people should report to Creators, and Creators decide the consequences.	6.7	6.7	6.8	6.8	6.8	6.3	6.8	7.0	6.8	6.9	6.9	6.2	0.055 (0.220)	0.231 (0.020)	-0.073 (0.419)	0.051 (0.596)	0.107 (0.311)	-0.069 (0.521)
2) In public spaces built by Creators, people should report to platform owners, and platforms decide the consequences.	7.2	7.1	7.2	7.4	7.3	7.1	7.7	7.6	7.8	7.7	7.5	7.5	0.433 (0.000)	0.493 (0.000)	0.574 (0.000)	0.245 (0.005)	0.259 (0.009)	0.403 (0.000)
3) In public spaces built by Creators, Creators should send severe cases to platform owners for additional enforcement	7.7	7.6	7.7	7.8	7.9	7.8	8.3	8.3	8.2	8.4	8.4	8.3	0.573 (0.000)	0.618 (0.000)	0.560 (0.000)	0.605 (0.000)	0.552 (0.000)	0.470 (0.000)
4) In members-only spaces, people should report to Creators who built them, and Creators decide the consequences.	6.9	7.0	6.9	7.1	6.9	6.5	7.2	7.2	7.2	7.3	7.3	6.7	0.240 (0.000)	0.262 (0.006)	0.223 (0.009)	0.210 (0.026)	0.311 (0.003)	0.212 (0.049)
5) In members-only spaces, people should report to platform owners and platforms decide the consequences.	7.2	7.0	7.2	7.3	7.3	7.1	7.4	7.3	7.4	7.5	7.4	7.3	0.209 (0.000)	0.285 (0.002)	0.231 (0.004)	0.199 (0.025)	0.075 (0.464)	0.132 (0.205)

## COMMUNITY FORUM - RESULTS ANALYSIS

6) In members-only spaces, Creators should send severe cases to platform owners for additional enforcement.	7.8	7.5	7.8	7.9	7.9	7.9	8.3	8.1	8.2	8.4	8.4	8.3	0.490 (0.000)	0.598 (0.000)	0.420 (0.000)	0.507 (0.000)	0.484 (0.000)	0.421 (0.000)
---	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------------------	------------------	------------------	------------------	------------------	------------------

In the control group, there were significant average increases in support for Proposal 2 for the 18-24 (7.0 to 7.3,  $p = .01$ ) and 25-34 (7.1 to 7.2,  $p = .05$ ) groups. There were no other significant changes by age group from the first to second control group survey.

### III. Should creators be made aware of reports made to platform owners?

Should Creators Be Made Aware of Reports Made to Platform Owners? (Mean Participant Responses)																		
	Age Group						Age Group						Age Group					
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) Platform owners should not notify the creator of the space where it happened.	4.4	4.9	4.6	4.3	4.2	3.3	4.4	5.1	4.6	4.1	3.6	3.1	-0.080 (0.116)	0.216 (0.043)	0.000 (0.998)	-0.219 (0.067)	-0.610 (0.000)	-0.197 (0.110)
2) Platform owners should notify the creator of the space where it happened, but not include personal information of the users involved.	6.6	6.6	6.7	6.7	6.7	6.3	7.1	7.3	7.1	7.2	6.9	6.7	0.483 (0.000)	0.690 (0.000)	0.406 (0.000)	0.465 (0.000)	0.285 (0.017)	0.440 (0.000)
3) Platform owners should notify the creator of the space where it happened with details about the event and the users involved.	7.0	6.7	6.9	7.3	7.2	7.2	6.9	6.8	6.8	7.1	7.2	6.7	-0.103 (0.030)	0.050 (0.638)	-0.124 (0.194)	-0.158 (0.108)	-0.003 (0.979)	-0.422 (0.000)

In the control group, average support for Proposal 1 increased significantly for participants in the 18-24 and 25-34 age groups. Average support for Proposal 2 decreased significantly for participants in the over 54 age group, and average support for Proposal 3 fell significantly for participants in the 25-34 age group.

IV. What should be done about spaces where there is repeated bullying and/or harassment?

What Should Be Done About Spaces with Repeated Bullying and/or Harassment (Mean Participant Responses)																		
	Age Group						Age Group							Age Group				
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) In public spaces Creators built, platforms should take action against Creators.	6.4	6.3	6.4	6.5	6.2	6.2	6.7	6.7	6.7	6.7	6.5	6.4	0.277 (0.000)	0.357 (0.001)	0.276 (0.003)	0.210 (0.043)	0.307 (0.009)	0.168 (0.154)
2) In members-only spaces, platforms should take action against Creators.	6.4	6.3	6.4	6.6	6.5	6.2	6.9	6.9	7.0	7.0	6.8	6.5	0.493 (0.000)	0.615 (0.000)	0.566 (0.000)	0.413 (0.000)	0.361 (0.001)	0.271 (0.019)
3) Members-only spaces should be made less visible to users.	6.3	6.4	6.3	6.3	6.2	5.9	6.7	7.0	6.8	6.7	6.4	6.2	0.445 (0.000)	0.596 (0.000)	0.478 (0.000)	0.419 (0.000)	0.169 (0.167)	0.318 (0.013)
4) Members-only spaces should not be publicly discoverable.	6.2	6.1	6.3	6.3	6.3	6.1	6.6	6.7	6.5	6.7	6.3	6.3	0.327 (0.000)	0.548 (0.000)	0.280 (0.005)	0.400 (0.000)	-0.044 (0.723)	0.181 (0.156)
5) Members-only spaces should be removed from the platform.	5.0	4.9	5.1	5.3	5.0	4.4	5.1	5.2	5.1	5.2	4.8	5.0	0.061 (0.266)	0.240 (0.045)	-0.055 (0.587)	-0.129 (0.316)	-0.154 (0.259)	0.513 (0.000)
6) Creators should be banned from making additional members-only space	5.1	5.0	5.1	5.3	5.1	5.0	5.4	5.4	5.4	5.5	5.2	5.4	0.271 (0.000)	0.410 (0.001)	0.298 (0.004)	0.155 (0.228)	0.041 (0.765)	0.317 (0.017)
7) Creators should be banned from inviting additional people to join the space.	5.2	5.1	5.2	5.4	5.3	5.2	5.5	5.3	5.5	5.6	5.2	5.5	0.263 (0.000)	0.233 (0.050)	0.389 (0.000)	0.251 (0.047)	-0.051 (0.702)	0.343 (0.010)
8) Creators should not be able to make money off of that space.	5.9	5.6	5.9	6.1	6.0	6.2	6.1	6.0	6.1	6.1	6.0	6.4	0.200 (0.000)	0.382 (0.001)	0.250 (0.017)	0.032 (0.800)	-0.062 (0.637)	0.192 (0.143)
9) Creators should be required to take a course on how to moderate the	7.3	7.2	7.2	7.3	7.5	7.8	8.1	8.0	8.1	8.1	8.3	8.2	0.761 (0.000)	0.778 (0.000)	0.836 (0.000)	0.793 (0.000)	0.792 (0.000)	0.442 (0.000)

## COMMUNITY FORUM - RESULTS ANALYSIS

spaces they create.																		
10) Users should receive notification of such cases when entering a space.	7.7	7.5	7.5	7.9	7.9	8.2	8.3	8.2	8.3	8.4	8.4	8.4	0.589 (0.000)	0.720 (0.000)	0.711 (0.000)	0.532 (0.000)	0.465 (0.000)	0.198 (0.018)

In the control group, among respondents aged 18-24, the average rating increased significantly for all Proposals except 5, 7, 8 from T1 to T2. For the 25-34 age group, the average rating increased significantly for Proposals 1, 2, 5, and 6. For the 35-44 age group, the average rating for Proposals 5, 6, and 10 increased significantly. For the 45-54 age group, the average rating increased significantly only for Proposal 3. For the over 54 age group, the average rating increased significantly only for Proposal 1.

### V. What tools should be used to identify bullying and harassment in social VR spaces?

#### Video Capture

Should Video Capture Be Used to Identify Bullying and Harassment? (Mean Participant Responses)																		
	Age Group						Age Group							Age Group				
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) Should be turned on in all public spaces.	6.4	6.3	6.3	6.6	6.5	6.2	7.2	7.3	7.2	7.3	7.3	7.0	0.845 (0.000)	0.931 (0.000)	0.879 (0.000)	0.702 (0.000)	0.828 (0.000)	0.804 (0.000)
2) Should be turned on in public spaces only where repeated bullying and harassment has occurred.	6.9	6.7	6.9	7.1	7.0	6.9	7.1	7.2	6.9	7.2	7.3	7.2	0.226 (0.000)	0.465 (0.000)	0.044 (0.633)	0.132 (0.214)	0.308 (0.006)	0.224 (0.053)
3) Should be turned on all members-only spaces.	6.1	5.9	6.1	6.3	6.2	6.0	6.3	6.2	6.3	6.5	6.3	6.2	0.218 (0.000)	0.263 (0.016)	0.205 (0.025)	0.234 (0.020)	0.152 (0.182)	0.188 (0.101)
4) Should be turned on in members-only spaces only where repeated bullying and harassment has occurred.	6.7	6.6	6.7	7.0	6.9	6.9	7.0	6.9	6.8	7.1	7.1	7.2	0.233 (0.000)	0.361 (0.001)	0.173 (0.058)	0.112 (0.300)	0.274 (0.019)	0.251 (0.035)

In the control group, respondents in the 18-24 age group significantly increased their average support for all Proposals. For the 25-34 age group, Proposals 1 and 3 saw significant average increases in support. For the 35-44 age group, respondents significantly increased their average support for Proposals 1, 3, and 4. In the 45-54 age group, respondents significantly decreased their average support for Proposal 2.

For respondents over 54, there was a significant average increase in support for Proposal 1 and a significant average decrease in support for Proposals 2 and 4.

## Automated Speech Detection

Should Automated Speech Detection Be Used to Identify Bullying and Harassment? (Mean Participant Responses)																		
	Age Group						Age Group							Age Group				
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) Should be turned on in all public spaces.	6.6	6.8	6.7	6.6	6.5	6.2	7.1	7.3	7.2	7.2	7.1	6.6	0.491 (0.000)	0.462 (0.000)	0.456 (0.000)	0.554 (0.000)	0.654 (0.000)	0.381 (0.000)
2) Should be turned on in public spaces only where repeated bullying and/or harassment has occurred	6.9	6.9	6.9	7.0	6.9	6.7	7.1	7.2	6.9	7.2	7.2	6.8	0.196 (0.000)	0.283 (0.007)	0.029 (0.757)	0.249 (0.015)	0.369 (0.001)	0.173 (0.145)
3) Should be turned on in all members-only spaces.	6.2	6.2	6.2	6.4	6.2	5.9	6.3	6.2	6.4	6.4	6.2	5.9	0.083 (0.073)	0.024 (0.819)	0.196 (0.027)	0.078 (0.450)	0.036 (0.763)	-0.031 (0.772)
4) Should be turned on in members-only spaces only where repeated bullying and/or harassment has occurred.	6.8	6.8	6.7	6.9	6.8	6.7	6.8	6.9	6.7	7.0	6.9	6.8	0.035 (0.463)	0.073 (0.492)	-0.060 (0.522)	0.081 (0.443)	0.096 (0.408)	0.062 (0.611)

For the control group, among respondents aged 18-24, average support for automatic speech detection increased significantly from T2 to T1 for all proposals. For respondents aged 25-34 and 35-44, average support increased significantly for Proposals 1 and 3. For respondents over 54, average support increased significantly for Proposal 1. There were no significant changes for respondents aged 45-54.

## VI. Who should review alerts?

Who Should Review Alerts from Automated Speech Detection? (Mean Participant Responses)																		
	Age Group						Age Group							Age Group				
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) Creators should review alerts from automated speech detection in public spaces they create.	7.3	7.2	7.3	7.5	7.2	7.2	7.4	7.4	7.3	7.5	7.4	7.1	0.061 (0.127)	0.159 (0.069)	0.041 (0.587)	-0.034 (0.705)	0.181 (0.078)	-0.088 (0.372)



## COMMUNITY FORUM - RESULTS ANALYSIS

2) Creators should review alerts from automated speech detection in public spaces they create but only when bullying and harassment has occurred there repeatedly.	7.0	6.9	6.9	7.3	7.0	7.0	7.1	7.2	7.0	7.4	7.2	6.9	0.132 (0.003)	0.284 (0.003)	0.062 (0.478)	0.163 (0.087)	0.114 (0.286)	-0.067 (0.550)
3) Creators should review alerts from automated speech detection in their members-only spaces.	7.1	7.0	7.1	7.3	7.2	7.1	7.3	7.3	7.3	7.5	7.4	7.0	0.226 (0.000)	0.295 (0.002)	0.293 (0.000)	0.189 (0.035)	0.215 (0.046)	-0.039 (0.705)
4) Creators should review alerts from automated speech detection in their members-only spaces but only when bullying and harassment has occurred there repeatedly	6.9	6.9	6.8	7.2	6.9	7.0	7.1	7.2	7.0	7.4	7.1	6.9	0.208 (0.000)	0.349 (0.000)	0.184 (0.039)	0.247 (0.008)	0.169 (0.110)	-0.076 (0.500)
5) Platform owners should review alerts from automated speech detection in public spaces even if they're built by Creators.	7.5	7.4	7.4	7.9	7.7	7.7	7.7	7.6	7.7	7.8	7.8	7.6	0.157 (0.000)	0.235 (0.006)	0.342 (0.000)	-0.056 (0.484)	0.064 (0.498)	-0.066 (0.479)
6) Platform owners should review alerts from automated speech detection in public spaces even if they're built by Creators, but only when bullying and harassment has occurred there repeatedly.	7.0	6.9	6.9	7.3	7.0	7.1	7.4	7.5	7.3	7.5	7.5	7.2	0.403 (0.000)	0.660 (0.000)	0.389 (0.000)	0.193 (0.038)	0.434 (0.000)	0.164 (0.151)
7) Platform owners should review alerts from automated speech detection in members-only spaces.	7.2	7.0	7.1	7.4	7.3	7.3	7.2	7.0	7.2	7.3	7.2	7.1	-0.029 (0.471)	0.033 (0.720)	0.047 (0.554)	-0.115 (0.201)	-0.050 (0.642)	-0.215 (0.030)

## COMMUNITY FORUM - RESULTS ANALYSIS

8) Platform owners should review alerts from automated speech detection in members-only spaces, but only when bullying and harassment has occurred there repeatedly.	7.0	6.9	6.9	7.2	7.1	7.0	7.1	7.1	7.1	7.3	7.2	7.0	0.147 (0.001)	0.198 (0.043)	0.196 (0.026)	0.078 (0.428)	0.153 (0.140)	0.005 (0.962)
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------------------	------------------	------------------	------------------	------------------	------------------

For the control group, respondents aged 18-24 showed significant average increases in their ratings for Proposals 6 and 8. Respondents in the 25-34 and 35-44 age groups had significant average rating increases for Proposal 6. The over 54 age group had a significant average rating increase for Proposal 7.

### VII. Should moderators be used to monitor bullying and harassment in social VR spaces?

Should Visible Moderators Be Used to Monitor Bullying and Harassment? (Mean Participant Responses)																		
	Age Group						Age Group							Age Group				
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) Visible moderators should be able to enter public spaces, even if they're built by Creators.	7.5	7.3	7.4	7.6	7.6	7.6	7.8	7.7	7.8	7.9	8.0	7.8	0.360 (0.000)	0.364 (0.000)	0.482 (0.000)	0.250 (0.004)	0.371 (0.000)	0.189 (0.036)
2) Visible moderators should be able to enter public spaces only when bullying and harassment has occurred repeatedly.	6.7	6.7	6.6	6.9	6.8	6.7	7.0	7.2	7.0	7.2	7.0	6.6	0.286 (0.000)	0.455 (0.000)	0.314 (0.000)	0.287 (0.005)	0.208 (0.078)	-0.099 (0.384)
3) Visible moderators should be able to enter members-only spaces.	6.8	6.5	6.8	6.9	6.9	7.0	6.5	6.3	6.6	6.6	6.7	6.7	-0.255 (0.000)	-0.248 (0.018)	-0.210 (0.016)	-0.329 (0.002)	-0.225 (0.035)	-0.309 (0.004)
4) Visible moderators should be able to enter members-only spaces only when bullying and harassment has occurred repeatedly.	6.7	6.6	6.6	6.9	6.8	6.7	6.9	6.9	6.8	7.1	6.8	6.8	0.196 (0.000)	0.282 (0.007)	0.222 (0.015)	0.187 (0.062)	0.047 (0.685)	0.099 (0.377)

## COMMUNITY FORUM - RESULTS ANALYSIS

5) Visible moderators should only enter public spaces when requested by Creators.	6.4	6.4	6.4	6.6	6.5	5.9	6.5	6.7	6.6	6.6	6.3	5.9	0.110 (0.026)	0.339 (0.002)	0.157 (0.104)	-0.013 (0.904)	-0.165 (0.172)	-0.067 (0.599)
6) Visible moderators should only enter public spaces when requested by users.	6.2	6.4	6.1	6.3	6.1	5.7	6.6	6.9	6.6	6.5	6.4	6.1	0.397 (0.000)	0.499 (0.000)	0.434 (0.000)	0.225 (0.036)	0.337 (0.007)	0.409 (0.001)
7) Creators should be able to select other users to be visible moderators in their spaces.	7.2	7.1	7.1	7.5	7.2	6.9	7.3	7.3	7.3	7.5	7.2	6.9	0.125 (0.004)	0.222 (0.018)	0.190 (0.033)	0.041 (0.650)	-0.006 (0.955)	-0.012 (0.921)

In the control group, respondents aged 18-24 significantly increased their average ratings for Proposals 2, 5, and 6. Respondents in the 25-34 and 45-54 age groups had significant increases in their average ratings for Proposal 6. Respondents in the 35-44 age group had significant increases in their average ratings for Proposals 5-7. Respondents over 54 increased their average rating significantly for Proposals 1 and 6 from T1 to T2.

### Invisible Moderators

Should Invisible Moderators Be Used to Monitor Bullying and Harassment? (Mean Participant Responses)																		
	Age Group						Age Group						Age Group					
Proposal	T1	1	2	3	4	5	T2	1	2	3	4	5	T2-T1	1	2	3	4	5
1) Invisible moderators should be able to enter all public spaces.	7.0	6.9	7.0	7.1	6.9	6.8	7.3	7.3	7.2	7.3	7.3	7.1	0.287 (0.000)	0.356 (0.000)	0.271 (0.002)	0.159 (0.127)	0.417 (0.001)	0.237 (0.040)
2) Invisible moderators should be able to enter all members-only spaces.	6.4	6.2	6.5	6.5	6.3	6.3	6.1	6.1	6.1	6.2	6.3	6.1	-0.215 (0.000)	-0.150 (0.156)	-0.325 (0.000)	-0.271 (0.012)	-0.052 (0.679)	-0.152 (0.196)
3) Invisible moderators should be able to enter public spaces only when bullying and harassment has occurred repeatedly.	6.5	6.5	6.4	6.6	6.3	6.4	6.6	6.8	6.5	6.8	6.5	6.3	0.142 (0.004)	0.283 (0.008)	0.051 (0.596)	0.205 (0.064)	0.127 (0.284)	-0.033 (0.779)
4) Invisible moderators	6.4	6.5	6.3	6.5	6.4	6.4	6.5	6.5	6.5	6.6	6.4	6.3	0.075	0.045	0.197	0.103	-0.020	-0.128

## COMMUNITY FORUM - RESULTS ANALYSIS

should be able to enter members-only spaces only when bullying and harassment has occurred repeatedly.													(0.123)	(0.674)	(0.037)	(0.345)	(0.860)	(0.281)
5) Invisible moderators should only enter public spaces when requested by Creators.	6.2	6.4	6.3	6.2	6.2	5.5	6.2	6.6	6.3	6.3	5.8	5.5	0.003 (0.950)	0.151 (0.179)	0.016 (0.863)	0.039 (0.734)	-0.370 (0.002)	-0.048 (0.705)
6) Invisible moderators should only enter public spaces when requested by users	6.3	6.5	6.3	6.2	6.2	5.7	6.3	6.6	6.3	6.3	5.9	5.7	0.021 (0.673)	0.142 (0.194)	0.030 (0.750)	0.054 (0.635)	-0.307 (0.010)	0.003 (0.981)

The control group also saw significant average increases in ratings for some proposals for using invisible moderators. In the 18-24 range, it was Proposals 1, 2, and 5. For the 25-34 age range, it was Proposals 4 and 6. For the 35-44 age range, it was 2 and 4-6. For the over 54 range, it was Proposals 1 and 2. Other combinations of age group and proposals did not have significant average changes in ratings.

### Metaverse Experience

Participants in the deliberations had the following distribution by experience in the Metaverse.

No Experience	Unsure	Experience	Total
n = 3,931	n = 531	n = 1,934	n = 6,398

The control group had the following distribution by experience in the Metaverse.

No Experience	Unsure	Experience	Total
n = 3,878	n = 807	n = 1,671	n = 6,357

### I. Who should be responsible for addressing bullying and harassment in social VR spaces?

Who should be responsible for addressing bullying and harassment in social VR spaces? (Mean Participant Responses)														
		Metaverse Exp?					Metaverse Exp?					Metaverse Exp?		
Proposal	T1	N	Unsure	Y	T2	N	Unsure	Y	T2-T1	N	Unsure	Y		
1) Creators should be primarily responsible for monitoring public spaces they create.	7.2	7.3	7.1	7.2	7.1	7.1	7.0	7.2	-0.132 (0.002)	-0.222 (0.000)	-0.051 (0.713)	0.009 (0.900)		
2) Creators should be primarily responsible for monitoring members-only spaces they create.	7.3	7.2	6.9	7.4	7.6	7.6	7.4	7.7	0.369 (0.000)	0.364 (0.000)	0.433 (0.003)	0.365 (0.000)		
3) Platform owners should be primarily responsible for monitoring public spaces even if they're built by Creators.	7.6	7.6	7.4	7.6	7.9	7.9	7.8	7.9	0.293 (0.000)	0.294 (0.000)	0.363 (0.008)	0.275 (0.000)		
4) Platform owners should be primarily responsible for monitoring members-only spaces even if they're built by Creators	7.3	7.3	7.0	7.5	7.2	7.1	7.2	7.4	-0.081 (0.050)	-0.144 (0.007)	0.186 (0.229)	-0.025 (0.729)		

In the control group, respondents without Metaverse experience significantly decreased their support for having creators be primarily responsible for monitoring their public spaces from T1 to T2. Respondents who were unsure about their Metaverse experience significantly increased their support for having creators primarily responsible for their members-only spaces. Respondents with Metaverse experience significantly decreased their support for this approach and for having platform owners primarily responsible for monitoring members-only spaces even if they are built by creators.

## II. To whom should cases of bullying and harassment be reported?

To Whom Should Cases of Bullying and Harassment Be Reported? (Mean Participant Responses)												
		Metaverse Exp?				Metaverse Exp?				Metaverse Exp?		
Proposal	T1	N	U	Y	T2	N	U	Y	T2-T1	N	U	Y
1) In public spaces built by Creators, people should	6.7	6.7	6.6	6.9	6.8	6.7	6.7	6.9	0.055 (0.220)	0.057 (0.329)	0.092 (0.554)	0.044 (0.570)

## COMMUNITY FORUM - RESULTS ANALYSIS

report to Creators, and Creators decide the consequences.												
2) In public spaces built by Creators, people should report to platform owners, and platforms decide the consequences.	7.2	7.2	6.9	7.4	7.7	7.7	7.2	7.7	0.433 (0.000)	0.495 (0.000)	0.309 (0.043)	0.348 (0.000)
3) In public spaces built by Creators, Creators should send severe cases to platform owners for additional enforcement	7.7	7.7	7.4	7.9	8.3	8.3	8.0	8.3	0.573 (0.000)	0.652 (0.000)	0.613 (0.000)	0.426 (0.000)
4) In members-only spaces, people should report to Creators who built them, and Creators decide the consequences.	6.9	6.8	6.7	7.2	7.2	7.1	7.1	7.3	0.240 (0.000)	0.300 (0.000)	0.442 (0.001)	0.094 (0.200)
5) In members-only spaces, people should report to platform owners and platforms decide the consequences.	7.2	7.1	7.0	7.3	7.4	7.3	7.3	7.5	0.209 (0.000)	0.224 (0.000)	0.385 (0.006)	0.148 (0.037)
6) In members-only spaces, Creators should send severe cases to platform owners for additional enforcement.	7.8	7.7	7.5	8.0	8.3	8.3	7.9	8.3	0.490 (0.000)	0.568 (0.000)	0.405 (0.002)	0.379 (0.000)

In the control group, respondents who were unsure about their Metaverse experience significantly increased their average positive ratings for Proposals 5 and 6. Other ratings did not change significantly from T1 to T2.

### III. Should creators be made aware of reports made to platform owners?

Should Platform Owners Inform Creators of Incidents?(Mean Participant Responses)						
		Metaverse Exp?		Metaverse Exp?		Metaverse Exp?

Proposal	T1	N	U	Y	T2	N	U	Y	T2-T1	N	U	Y
1) Platform owners should not notify the creator of the space where it happened.	4.4	4.2	4.6	4.7	4.4	4.3	4.6	4.4	-0.080 (0.116)	0.036 (0.581)	0.040 (0.819)	-0.301 (0.001)
2) Platform owners should notify the creator of the space where it happened, but not include personal information of the users involved.	6.6	6.5	6.8	6.8	7.1	7.1	6.9	7.2	0.483 (0.000)	0.563 (0.000)	0.138 (0.423)	0.411 (0.000)
3) Platform owners should notify the creator of the space where it happened with details about the event and the users involved.	7.0	6.9	7.0	7.2	6.9	6.7	6.9	7.1	-0.103 (0.030)	-0.154 (0.014)	-0.114 (0.500)	-0.013 (0.871)

In the control group, respondents with no Metaverse experience, unsure about their experience, and with experience all significantly increased their average support for Proposal 1. Those with Metaverse experience also significantly increased their average rating for Proposal 3.

#### IV. What should be done about spaces where there is repeated bullying and/or harassment?

What Consequences Should There Be for Spaces with Repeated Bullying and Harassment (Mean Participant Responses)												
		Metaverse Exp?				Metaverse Exp?				Metaverse Exp?		
Proposal	T1	N	U	Y	T2	N	U	Y	T2-T1	N	U	Y
1) In public spaces Creators built, platforms should take action against Creators.	6.4	6.3	6.4	6.5	6.7	6.6	6.8	6.8	0.277 (0.000)	0.291 (0.000)	0.378 (0.018)	0.234 (0.005)

## COMMUNITY FORUM - RESULTS ANALYSIS

2) In members-only spaces, platforms should take action against Creators.	6.4	6.3	6.4	6.5	6.9	6.8	7.0	7.0	0.493 (0.000)	0.495 (0.000)	0.558 (0.000)	0.477 (0.000)
3) Members-only spaces should be made less visible to users.	6.3	6.1	6.2	6.6	6.7	6.6	6.8	6.8	0.445 (0.000)	0.529 (0.000)	0.647 (0.000)	0.265 (0.002)
4) Members-only spaces should not be publicly discoverable.	6.2	6.1	6.1	6.4	6.6	6.4	6.8	6.7	0.327 (0.000)	0.334 (0.000)	0.716 (0.000)	0.237 (0.008)
5) Members-only spaces should be removed from the platform.	5.0	4.8	5.4	5.3	5.1	5.0	5.5	5.2	0.061 (0.266)	0.172 (0.015)	0.107 (0.559)	-0.134 (0.165)
6) Creators should be banned from making additional members-only space	5.1	5.0	5.4	5.3	5.4	5.3	5.6	5.5	0.271 (0.000)	0.327 (0.000)	0.200 (0.302)	0.187 (0.066)
7) Creators should be banned from inviting additional people to join the space.	5.2	5.0	5.4	5.4	5.5	5.3	6.0	5.6	0.263 (0.000)	0.301 (0.000)	0.607 (0.001)	0.131 (0.181)
8) Creators should not be able to make money off of that space.	5.9	5.9	6.1	5.9	6.1	6.1	6.3	6.1	0.200 (0.000)	0.205 (0.003)	0.265 (0.164)	0.178 (0.059)
9) Creators should be required to take a course on how to moderate the spaces they create.	7.3	7.3	7.2	7.4	8.1	8.1	8.0	8.1	0.761 (0.000)	0.805 (0.000)	0.739 (0.000)	0.687 (0.000)
10) Users should receive notification of such cases when entering a space.	7.7	7.8	7.5	7.7	8.3	8.3	7.9	8.4	0.589 (0.000)	0.551 (0.000)	0.414 (0.001)	0.691 (0.000)

In the control group, respondents without Metaverse experience significantly increased their average support for Proposals 1-3, 5-7, and 9. Respondents who were unsure about their Metaverse experience



increased their average support significantly for Proposals 3 and 4. For respondents with Metaverse experience, their average support for Proposals 2, 3, 6, and 10 increased significantly from T1 to T2.

## V. What tools should be used to identify bullying and harassment in the metaverse?

### Video Capture

Where should video capture be used to identify bullying and harassment in the metaverse?(Mean Participant Responses)												
Proposal	T1	Metaverse Exp?			T2	Metaverse Exp?			T2-T1	Metaverse Exp?		
		N	U	Y		N	U	Y		N	U	Y
1) Should be turned on in all public spaces.	6.4	6.1	6.4	6.8	7.2	7.1	7.4	7.5	0.845 (0.000)	0.921 (0.000)	0.999 (0.000)	0.679 (0.000)
2) Should be turned on in public spaces only where repeated bullying and harassment has occurred.	6.9	6.7	7.0	7.2	7.1	7.0	7.2	7.3	0.226 (0.000)	0.311 (0.000)	0.166 (0.282)	0.089 (0.272)
3) Should be turned on all members-only spaces.	6.1	5.9	6.2	6.5	6.3	6.1	6.7	6.6	0.218 (0.000)	0.260 (0.000)	0.501 (0.003)	0.087 (0.276)
4) Should be turned on in members-only spaces only where repeated bullying and harassment has occurred.	6.7	6.6	6.8	7.0	7.0	6.9	7.2	7.1	0.233 (0.000)	0.327 (0.000)	0.375 (0.021)	0.038 (0.635)

In the control group, respondents without experience in the Metaverse significantly increased their average support for video capture in all public spaces and in all members-only spaces. Respondents unsure about their Metaverse experience significantly increased their average support for video capture in all members-only spaces. Those with Metaverse experience significantly increased their average support for video capture in all public spaces.

### Automatic speech detection

Where should automated speech detection be used to identify bullying and harassment in the metaverse? (Mean Participant Responses)				
	Metaverse Exp?		Metaverse Exp?	

## COMMUNITY FORUM - RESULTS ANALYSIS

Proposal	T1	N	U	Y	T2	N	U	Y	T2-T1	N	U	Y
1) Should be turned on in all public spaces.	6.6	6.5	6.6	7.0	7.1	7.0	7.3	7.3	0.491 (0.000)	0.568 (0.000)	0.641 (0.000)	0.325 (0.000)
2) Should be turned on in public spaces only where repeated bullying and/or harassment has occurred	6.9	6.8	6.8	7.1	7.1	7.0	7.0	7.2	0.196 (0.000)	0.249 (0.000)	0.205 (0.218)	0.101 (0.224)
3) Should be turned on in all members-only spaces.	6.2	6.0	6.4	6.5	6.3	6.1	6.5	6.6	0.083 (0.073)	0.120 (0.050)	0.070 (0.649)	0.021 (0.789)
4) Should be turned on in members-only spaces only where repeated bullying and/or harassment has occurred.	6.8	6.7	6.8	7.0	6.8	6.8	6.8	6.9	0.035 (0.463)	0.081 (0.198)	0.048 (0.775)	-0.047 (0.567)

In the control group, support for using automatic speech detection increased significantly across all levels of Metaverse experience for all proposals, with the exception of Proposals 1 and 2 for respondents unsure about their Metaverse experience and Proposal 2 for respondents with Metaverse experience.

### V. Who should review alerts?

Speech detection

Who Should Review Automated Speech Detection Alerts? (Mean Participant Responses)												
		Metaverse Exp?				Metaverse Exp?				Metaverse Exp?		
Proposal	T1	N	U	Y	T2	N	U	Y	T2-T1	N	U	Y
1) Creators, public spaces.	7.3	7.2	7.2	7.4	7.4	7.2	7.4	7.6	0.061 (0.127)	-0.011 (0.828)	0.226 (0.113)	0.155 (0.021)
2) Creators, public spaces if repeated bullying/harassment.	7.0	7.0	6.8	7.1	7.1	7.1	7.1	7.3	0.132 (0.003)	0.122 (0.031)	0.245 (0.117)	0.128 (0.103)
3) Creators, members-only spaces.	7.1	7.0	6.9	7.3	7.3	7.2	7.4	7.6	0.226 (0.000)	0.141 (0.009)	0.554 (0.000)	0.309 (0.000)

## COMMUNITY FORUM - RESULTS ANALYSIS

4)Creators, members-only spaces if repeated bullying and harassment.	6.9	6.9	6.7	7.1	7.1	7.1	7.2	7.3	0.208 (0.000)	0.197 (0.001)	0.433 (0.005)	0.181 (0.022)
5)Platform owner, public spaces even if built by Creators.	7.5	7.5	7.3	7.7	7.7	7.6	7.6	7.8	0.157 (0.000)	0.137 (0.005)	0.371 (0.014)	0.148 (0.027)
6)Platform owner, public spaces even if built by Creators, only when repeated bullying and harassment.	7.0	6.9	6.9	7.2	7.4	7.4	7.3	7.5	0.403 (0.000)	0.432 (0.000)	0.401 (0.007)	0.352 (0.000)
7)Platform owner members-only spaces.	7.2	7.1	7.0	7.4	7.2	7.1	7.2	7.3	-0.029 (0.471)	-0.038 (0.479)	0.166 (0.286)	-0.054 (0.434)
8)Platform owner members-only spaces, but only when repeated bullying and harassment.	7.0	6.9	7.0	7.1	7.1	7.1	7.1	7.2	0.147 (0.001)	0.194 (0.001)	0.161 (0.320)	0.062 (0.425)

In the control group, respondents across all levels of Metaverse experience significantly changed their average positive ratings for Proposal 3. Respondents without and unsure about their Metaverse experience increased their support, while those with Metaverse experience decreased their support. Respondents in all three groups significantly increased their average support for Proposal 6. People who were unsure about their Metaverse experience also significantly increased their average ratings for Proposals 5 and 8.

V. Should moderators be used to monitor bullying and harassment in social VR spaces?

Should Moderators Be Used to Monitor Bullying and Harassment? (Mean Participant Responses)												
		Metaverse Exp?				Metaverse Exp?				Metaverse Exp?		
Proposal	T1	N	U	Y	T2	N	U	Y	T2-T1	N	U	Y
1) Visible moderators should be able to enter public spaces, even if they're built by Creators.	7.5	7.4	7.4	7.6	7.8	7.8	7.6	7.9	0.360 (0.000)	0.429 (0.000)	0.173 (0.233)	0.276 (0.000)
2) Visible moderators should be able to enter public spaces only when bullying and harassment has occurred repeatedly.	6.7	6.7	6.6	6.9	7.0	7.0	6.7	7.2	0.286 (0.000)	0.326 (0.000)	0.055 (0.749)	0.261 (0.001)
3) Visible moderators should be able to enter members-only spaces.	6.8	6.6	6.9	7.1	6.5	6.4	6.7	6.8	-0.255 (0.000)	-0.212 (0.000)	-0.221 (0.187)	-0.338 (0.000)
4) Visible moderators should be able to enter members-only spaces only when bullying and harassment has occurred repeatedly.	6.7	6.6	6.7	6.9	6.9	6.8	6.8	6.9	0.196 (0.000)	0.273 (0.000)	0.135 (0.424)	0.072 (0.378)
5) Visible moderators should only enter public spaces when requested by Creators.	6.4	6.2	6.3	6.7	6.5	6.4	6.4	6.7	0.110 (0.026)	0.206 (0.001)	0.139 (0.447)	-0.064 (0.454)
6) Visible moderators should only enter public spaces when requested by users.	6.2	6.1	6.0	6.4	6.6	6.6	6.4	6.6	0.397 (0.000)	0.528 (0.000)	0.421 (0.016)	0.165 (0.052)
7) Creators should be able to select other users to be visible	7.2	7.0	6.8	7.5	7.3	7.2	7.2	7.4	0.125 (0.004)	0.200 (0.001)	0.346 (0.023)	-0.049 (0.492)

## COMMUNITY FORUM - RESULTS ANALYSIS

moderators in their spaces.												
1) Invisible moderators should be able to enter all public spaces.	7.0	6.8	6.9	7.2	7.3	7.2	7.3	7.4	0.287 (0.000)	0.361 (0.000)	0.388 (0.028)	0.137 (0.084)
2) Invisible moderators should be able to enter all members-only spaces.	6.4	6.1	6.5	6.8	6.1	5.9	6.5	6.5	-0.215 (0.000)	-0.205 (0.001)	-0.027 (0.881)	-0.268 (0.001)
3) Invisible moderators should be able to enter public spaces only when bullying and harassment has occurred repeatedly.	6.5	6.4	6.4	6.6	6.6	6.6	6.5	6.6	0.142 (0.004)	0.198 (0.001)	0.163 (0.368)	0.038 (0.671)
4) Invisible moderators should be able to enter members-only spaces only when bullying and harassment has occurred repeatedly.	6.4	6.3	6.6	6.6	6.5	6.4	6.6	6.6	0.075 (0.123)	0.128 (0.037)	0.016 (0.926)	-0.007 (0.934)
5) Invisible moderators should only enter public spaces when requested by Creators.	6.2	6.1	6.0	6.6	6.2	6.1	6.3	6.4	0.003 (0.950)	0.040 (0.545)	0.290 (0.094)	-0.117 (0.168)
6) Invisible moderators should only enter public spaces when requested by users	6.3	6.2	6.0	6.5	6.3	6.2	6.3	6.4	0.021 (0.673)	0.063 (0.319)	0.301 (0.099)	-0.110 (0.209)

In the control group, for visible moderators, all three respondent groups significantly increased their average positive ratings for Proposal 6. Respondents who were unsure about their Metaverse experience also significantly increased their average positive ratings for Proposals 2-5 from T1 to T2. On the topic of invisible moderators, respondents without Metaverse experience increased their average positive ratings for all proposals. Those unsure about their Metaverse experience increased them only for Proposals 1, 3, and 4, and those with Metaverse experience increased them only for Proposals 5 and 6.