Written Testimony of Arturo Bejar before the Subcommittee on Privacy, Technology, and the Law, November 7, 2023

Subcommittee Chair Blumenthal, Subcommittee Ranking Member Hawley, and members of the Subcommittee:

My name is Arturo Bejar and I am a dad with firsthand experience of a child who experienced unwanted sexual advances on Instagram and an expert with 20 years of experience working as a senior leader, including leading online security, safety, and protection at Facebook.

From 2009 to 2015, I was the senior engineering and product leader at Facebook responsible for its efforts to keep users safe and supported. I ran a group called Protect & Care. It was responsible for "Site Integrity" – stopping attacks and malicious behavior; "Security Infrastructure" – which engineered resilient systems and worked on compliance; and a group called "Care"– which developed Facebook's user-facing and internal customer care tools. I also oversaw the child-safety tools. I reported directly to CTO Mike Schroepfer, who himself reported to Mark Zuckerberg. For each of these areas I was responsible for the combined effort of engineering, product, user research, data, and design. This included doing strategic product reviews every six months with the Facebook executive team including: Mark Zuckerberg, Sheryl Sandberg, and Chris Cox.

When there was a significant engineering security, site integrity, child safety, or care issue, I was one of the people that the executive team worked with to oversee investigations, help craft accurate responses to external enquiries, and to coordinate engineering and product changes where needed in those areas. For most of that time I was also the manager for Facebook's "Product Infrastructure" team, which built key parts of the product engineering frameworks of Facebook, and developed REACT, one of the core technologies of the web today. Prior to that I had been recruited to Facebook from Yahoo! where I worked from 1998 to 2009. I was hired as Yahoo!'s first product security engineer, eventually becoming head of Information Security reporting to the CTO.

Earlier this year, I was subpoenaed to testify under oath about emails I sent Facebook's executive team as part of a government investigation and I realized that I had written these emails over two years ago and yet nothing had changed. Meta continues to publicly misrepresent the level and frequency of harm that users, especially children, experience on the platform, And they have yet to establish a goal for actually reducing those harms and protecting children. It's time that the public and parents understand the true level of harm posed by these "products" and it's time that young users have the tools to report and suppress online abuse.

While leading Facebook's Protect & Care group up until 2015, I spent much of my time working on ways to make it easier for users to tell the company when things weren't going right for them on the service. That included building tools to make it easier to report problems, as well as creating survey tools so the company could understand what people were experiencing and how they felt about the service and its features. When I left in 2015, after six years at the company, I felt good that we had built numerous systems that made using our products easier and safer.

I left in part because I wanted more time with my family. My daughter was entering her teenage years. Like most young Americans she is an avid user of social media, particularly Instagram and Snapchat. But as a result, I got an up-close look at the experience of a real young person's daily experience on these products. While there is plenty of good that comes from her time using social media, frequently she has to deal with awful problems. Since they were 14, my daughter and all of her friends have repeatedly faced unwanted sexual advances, misogyny, and harassment. This has been profoundly distressing to them (and to me). But even though social media services enabled these distressing experiences, the services provided little or no help to her in dealing with them. And their experience is far from unique. In fact it's common. So in 2019, I decided I wanted to see why my kids and their friends were having these problems, and went back to Facebook as an independent consultant. I stayed for two years, working with the team at Instagram that focused on "well-being."

It was not a good experience. Almost all of the work that I and my colleagues had done during my earlier stint at Facebook through 2015 was gone. The tools we had built for teenagers to get support when they were getting bullied or harassed were no longer available to them. People at the company had little or no memory of the lessons we had learned earlier.

The most important lesson we learned about teens before 2015 was that when one of them asks for help, the content that was involved in the incident does not matter as much as giving them support to help them in the moment. So we built a tool that asked them–using language we tested and developed in partnership with teens–what was happening, what emotion they were experiencing, and how intense it was. Based on that information, we could better offer specific help.

The process worked, at the beginning of this process, when we were giving them tools to deal with 'bullying and harassment,' 11% of 13-14 year olds completed the flow, as we called this series of questions. By the end of it, for a flow with more steps that covered a wider variety of problems, 82% were completing the flow. What that showed us is that if the steps we offer are helpful in a moment of need, teens will take them.

At this time we knew that well over 90% of the issues that people were reporting did not technically violate company policy. Most importantly and initially surprisingly, we learned that

in 50% of the issues teens were flagging as the most intense bad experiences, the content involved was discernibly benign to us, as outside observers.

And our feedback tools also showed that 90% of the time, the people who posted content that got reported by another user, said they did not intend it to be upsetting to someone else.

By the time I left, in 2015, based on experience, we had learned that to obtain the best measure of the success of a support flow, we simply needed to ask 'Did you get help with what you needed?'

When I returned in 2019, I was confounded. There were a great many motivated and talented team members working on online safety. But no one on that team was aware of the work we had done at Facebook and the lessons we had learned four years earlier. The group at Instagram and the talented internal research teams had developed some very troubling evidence that young teens were experiencing great distress and abuse on the Instagram platform. But senior management was externally reporting different data that grossly understated the frequency of harm experienced by users.

I was committed to understanding this gap and began to assemble some of the research in order to convey the problem to senior leaders. Eventually, on October 5, 2021, after having the analysis reviewed internally, I sent a detailed email to Mark Zuckerberg and the other senior leaders detailing what I had found. First, I pointed out how the reporting process grossly understated misconduct on the site. I explained that the number of people reporting to surveys that they had a negative experience on Instagram was 51% *every week* but only 1% of those reported the offending content and only 2% of those succeeded in getting the offending content taken down.

Thereafter, I detailed the staggering levels of abuse that teens aged 13-15 were experiencing every week. The initial data from the research team indicated that as many as 21.8% of 13-15 year olds said they were the target of bullying *in the past seven days*, 39.4% of 13-15 year old children said they had experienced negative comparison, *in the past seven days*, and 24.4% of 13-15 year old responded said they received unwanted advances, all in the prior seven days. Later, the research team revised the survey results to state that the likely number of 13-15 year old children receiving unwanted sexual advances in the past seven days was likely *only* 13 percent, still a shocking number. Obviously, an even higher percentage of these children are receiving unwanted sexual advances on a monthly basis.

The reaction was not constructive. Sheryl Sandberg expressed empathy for my daughter but offered no concrete ideas or action. Adam Mosseri responded with an request for a follow up meeting, Mark Zuckerberg never replied. That was unusual. It might have happened, but I don't

recall Mark ever not responding to me previously in numerous communications, either by email or by asking for an in-person meeting.

Today, most harm remains unaddressed

Most of the distress people experience online because of unwanted contact and content is not addressed today by Meta and other social media companies. I say that based on my extensive experience working to keep users safe, along with my direct knowledge of extensive research and data about what people experience on Meta's services and elsewhere online. But it's not just that the companies disregard people's distress. The way they respond to problems often makes those problems worse, because it normalizes harmful behavior and encourages unwanted contact and content. In addition, the way companies talk about these problems to regulators, policy makers, and the general public is seriously misleading.

It's been almost two years since I left Instagram, and as it stands right now, I don't believe anything is going to change. All this time there has been extensive harm happening to teenagers, and the leadership has been aware of it, but they have chosen not to investigate or address the problems. I know because I respectfully communicated this directly to the executive team in 2021, and have watched them do essentially nothing in response. One key fact I told company leaders, for example, was that based on carefully-crafted and vetted surveys, we had identified the disturbing fact that 13% of Instagram users aged 13-15 self reported having received unwanted sexual advances via the platform *within the previous seven days*. That is an awful statistic. Looked at over time, it is likely the largest-scale sexual harassment of teens to have ever happened, and one that clearly calls for action.

Instagram's stated mission is to provide a safe and supportive place. As I was preparing these documents, someone wrote to a teenager I know asking them to sell nude photographs of themselves. It is September of 2023, two years after my briefings, and there is still no way, so far as I or teenagers I know can determine, for a minor to flag a conversation in Instagram to indicate it contains unwanted sexual advances. And this is just one of several categories of meaningful harm that teenagers experience. An environment where unwanted sexual advances are normalized is hardly safe and supportive.

There are plenty of things that need to change about how social media systems function in society. A number of states are already introducing laws that impose restrictions on how young people can use such systems. The U.S. Surgeon General recently released a statement underscoring his grave concern about what the services are doing to our children.

The platforms must change how they identify and measure unwanted contact

But regardless of other reforms, I have realized that in order to begin to reduce the harm they facilitate, social media companies have to be compelled to change how they identify and measure users' exposure to unwanted contact. In addition, the services must be compelled to publicly report these measurements. Doing so will inevitably encourage them to introduce features that enable users to better deal with those harms. For example, public earnings calls or other formal public processes should include a report on the percentage of teenagers who experienced unwanted sexual advances in that quarter.

Social media companies are not going to start addressing the harm they enable for teenagers on their own. They need to be compelled by regulators and policy makers to be transparent about these harms and what they are doing to address them.

In order to support regulatory and transparency efforts, I am proposing, based on my experience as a senior leader who used to manage these areas and respond to regulatory enquiries, a group of measures that I believe are pragmatic and straightforward to implement. These could help start the process of change. These measures do not require significant investments by the platforms in people to review content or in technical infrastructure. However, they do require the leadership of social media companies to prioritize. They will ensure that the companies have enough people, infrastructure, and know-how to implement these recommendations quickly. This document provides an overview. Below are links to documents with more detailed explanations, which aim to support further action around the different areas.

My goal in all this, as a father and as an engineer who has worked on these problems for many years, is to help regulators, policy-makers, academics, journalists, and the public better understand how companies think about these problems and what it would take to address them. I also want to create a meaningful increase in support for integrity and trust and safety workers at the companies. I am not selling anything, nor will I be looking for work in this field. This is my statement of retirement from the technology industry, which I will make freely available and open for comments. Any work I do in the future to support regulators or others committed to reducing harm on social media will be pro bono.

Meta's current approach to these issues only addresses a fraction of a percent of the harm people experience on the platform. In recent years, repeated examples of harm that has been enabled by Meta and other companies has come to light, through whistleblowing, outside research studies, and many stories of distressing experiences people have there. Whenever such reports emerge, Meta's response is to talk about 'prevalence', and its investment in moderation and policy, as if that was the only relevant issue. But there is a material gap between their narrow definition of prevalence and the actual distressing experiences that are enabled by Meta's products. However, managers including Meta CEO Mark Zuckerberg do not seem to seek to understand or actually address the harms being discussed. Instead, they minimize or downplay published findings, and

even sometimes the results of their own research. They also try to obfuscate the situation by quoting statistics that are irrelevant to the issues at hand.

Successful reforms must be based on data

Social media companies, and Meta in particular, manage their businesses based upon a close and ongoing analysis of data. Nothing gets changed unless it is measured, so it is critical that when it comes to unwanted and distressing content, the data gathered and metrics established be based on people's actual experiences. This is a company guided by data. Once Meta establishes metrics for anything, employees are given concrete incentives to drive those metrics in the direction the company deems useful and valuable. Metrics determine, for example, how many people work in a given department. Most of all, metrics establish the companies' priorities.

When outside critics point to harms caused by Facebook or Instagram, I have often observed Meta CEO Mark Zuckerberg and his managers try to change the conversation to the things they measure. If the problems identified are not problems that the company's systems are designed to detect and measure, managers literally have no means to understand them. Zuckerberg is unwilling to respond to criticisms of his services that he feels are not grounded in data. For Meta, a problem that is not measured is a problem that doesn't exist.

When I worked at Facebook, I helped design, introduce, and manage systems that measured important harms as well as methods to reduce those harms, for teens and others. However, many of those systems were later shut down or changed completely in ways that degraded their effectiveness. Meanwhile, the range of unwanted content that people receive has increased in scope, reach, and intensity. So now I believe it is necessary for outsiders–that means governments and regulators–to step in and require metric-based goals based on the harm that teenagers experience, as well as transparent systems of public reporting and disclosure based upon what those tools reveal. I know from experience that this is possible to do methodically and effectively.

I have specific recommendations for regulators, to require any company that operates social media services for teenagers to develop certain metrics and systems. These approaches will generate extensive user experience data, which then should be regularly and routinely reported to the public, probably alongside financial data. I believe that if such systems are properly designed, we can radically improve the experience of our children on social media. The goal must be to do this without eliminating the joy and value they otherwise get from using such services.

I don't believe such reforms will significantly affect revenues or profits for Meta and its peers. These reforms are not designed to punish companies, but to help teenagers. And over time, they will create a safer environment.

Any of these reforms could have been introduced voluntarily by the companies earlier, and in fact the Protect & Care team successfully implemented measures based on these principles at Facebook as much as 10 years ago. But those were abandoned, along with the entire underlying approach we developed. Now, we are faced with the unwillingness of companies on their own to seek to understand the range of unwanted content they recommend and the unwanted contact they enable. Because of their unwillingness to develop and maintain tools and improvements in the face of material data and evidence of harms, they must be required to do so.

There are many opportunities for impactful and helpful innovation, if companies are required to set their priorities and devote resources with the goal of reducing the amount of unwanted and distressing content that teens experience. Much of that innovation might eventually spring from the companies themselves. But the product features, policies, and punishments Meta and other companies have today are necessary but not sufficient. They are clearly not able to create a safe environment for teens online. These include their existing reporting, policy, and punishment systems, as well as the inadequate tools they offer today to enable users to block or restrict other users.

What regulators and governments can do right away

Regulators should require changes in three areas:

1. Users ought to be given the ability to signal that they have experienced unwanted contact in messages or comments to their posts. They also need to be able to share with the platform the reason why that contact is unwanted. Currently, platforms make it unnecessarily difficult for people to report unwanted contact, and they often compel users to mischaracterize their experience in order to report it within the narrow categories defined by their policies. The reporting categories they provide do not match what people are experiencing. People, especially teens, need to be presented with options that are relevant, clear, and flexible. For example, the options presented to users for the reason contact may be unwanted could include "because it's gross", "it's harassing me", "it's fake", etc. When we did such work in the past, giving teens options that matched their experience led to meaningful increases in their use of the support tools.

2. People should have a voice about the content they experience. That can be achieved by giving them the ability to curate their own content "feed" to avoid unwanted or unhealthy interactions. This can be done by creating a feedback mechanism to indicate the content

that makes them uncomfortable and the reason why. Such a process would be straightforward to implement, it is already implemented in advertising for Instagram and Facebook. Such a mechanism should also be applied to products like Reels or other algorithmically-controlled recommendation systems. Again, users should be given the ability to share the reason why certain interactions or information they are exposed to makes them uncomfortable. One example of a possible response made available for teens might be "it makes me feel bad about myself." Information gathered from such responses should be used to improve recommendations, and to improve the community experience over time. Algorithms are only as good as their inputs, and without meaningful feedback and programming adjustments, they will continue recommending unwanted content.

3. Work on addressing these harms needs to be done in partnership with experts in the relevant fields. Company product managers and engineers don't have sufficient knowledge or expertise in social science and other areas of relevance. Policy team members, who often do have more knowledge, do not, however, drive product design. Issues such as self-image, addiction or problematic use, and bullying and harassment ought to be addressed by creating product features in close partnership with people who have experience and are experts in the subject, such as outside academics. Meta has sometimes said that this cannot be done without violating user privacy and security, but that is misleading. As part of my work at Facebook we did this regularly and without impinging on user privacy. The company can bring in third parties and give them access to all the necessary data, while still remaining bound by the privacy and security standards it has promised. There are also processes which appropriately anonymize data in order to share findings.

(I include more detailed thoughts about how to consider regulation below)

Those who initiate unwanted actions need to be considered, too

When it comes to the people who initiate unwanted messages or post unwanted content, mechanisms should be implemented to give them private feedback so they can understand the impact of their actions. The feedback they receive should explain the reason why that content was not desired or appropriate. Of course, safety is a key consideration in the design of these features. The principle ought to be that all users, even apparent offenders, deserve to be treated with respect and, at least initially, be given sufficient information to enable them to operate as respectful community members. Today, content creators are often surprised when they lose access to their account, or see the distribution of their content restricted, and frequently do not know why it happens. Most creators would in fact appreciate knowing the reasons why they lose

users and what they can do to help their content reach more people. Only if they repeatedly disregard such warnings should they be restricted from features or distribution.

If you give people respectful private feedback, I learned in my work there that many if not most of them will adjust their behavior. Most people who initiate interactions that others don't want do so with what they themselves consider to be positive intentions. Several sociological studies have come to this conclusion, and it corresponds to what I learned at Facebook, where research studies that were conducted multiple times found that, for content other users had reported as problematic, the intent of the content creator was, in their own mind, positive or neutral 90% of the time. Of course, there were 6% of people who posted problematic content who did so despite knowing it could be upsetting, and 4% of them did it with a deliberate intent to provoke. There are trolls on these services, for sure. But if you put yourself in the seat of the people who share content you've seen that is upsetting, you will often find that they have shared it because it was important to them, because they thought it was funny, or maybe because they are afraid of something happening to their friends. Most people who have unintentionally done something that is distressing to someone else would like to know about it, as long as the feedback is delivered privately and respectfully.

Importantly, the process of giving feedback to people who take unwanted actions helps a service then identify those who nonetheless ignore that feedback. People who repeatedly engage in behavior that is distressing to others definitely require further response and restrictions. These measures are intended to separate well-meaning community members from trolls.

I oppose censorship

It is essential to note that in all of these approaches I'm not advocating for censorship. I believe in the importance of free expression. However, free expression should not allow an individual to send someone direct messages (DMs) that harass them, or to make misogynistic or hateful comments on a teenager's posts. There should be no right to harass.

Underlying this approach is the belief that in order to reduce distressing experiences for people, the most important area social media companies should work on is social norms. The current approach, based on setting legalistic definitions within policies and reactively removing content, is not sufficient. More importantly, it does not address the majority of the distressing experiences people face. What must guide the design of features to make people feel safe with each other in social media should be the actual experience of users.

Social media is unlike most environments where people spend time, because in general it does not have sufficient accepted and maintained social norms. What makes a workplace, school, or park feel safe, by contrast, is mostly not the policing, but rather how people just know how to

behave. We would never tolerate routine sexual advances to teens at our local supermarket. However, on Instagram, Meta, and other social services at the moment, we completely tolerate that happening.

Real World Standards Should Apply to Social Media

It has been implicitly argued by many that we must accept unconscionable levels of misogyny, sexually explicit content and unwanted advances, depression, bullying and other harms as an unavoidable cost of having social media. Many at Meta and elsewhere would falsely argue that we must accept it because it is merely a reflection of the "real world." This is wholly false. First, as we have seen, dangerous and harmful experiences occur on Instagram at a rate exponentially greater than the real world. Second, the reverse is actually true. The tools and algorithms that can be applied to our social media mean that Instagram and Facebook should be far safer than the real world. Meta can make these platforms far safer for our children if they were motivated to do so.

Mandatory Measurement and Transparency Will Drive Accountability

I have spent my career working in some of the most successful corporations. One thing that is common among any successful company is a discipline of setting quantifiable goals and then holding people accountable for achieving those goals. It works. What works even better is public accountability. Every public company reports its financial results every quarter and the public markets impose swift consequences for companies and executives that fall short of the quantifiable goals. Because of that accountability, the leaders are sharply focused on those results.

The same principle can apply to online safety. Establishing consistent measurements of key harms and then requiring publication of those results every quarter will ensure parents and the public can hold companies accountable. And embarrassing or eroding safety measurements will be something that leaders will want to avoid. Public and measurable accountability is a key element of making these platforms safe for children.

More Detailed Thoughts about Regulatory Approaches and Processes

The most effective way to regulate social media companies is to require them to develop metrics that will allow both the company and outsiders to evaluate and track instances of harm, as experienced by users. This plays to the strengths of what these companies can do, because data for them is everything. If something cannot be evaluated by data analysis, it is generally very difficult for Meta and other such companies to understand the problem or take action.

Process-based or policy-based regulations are essential for security and privacy. In order to effectively regulate the safety of a social media environment, the focus should be on metrics based on user experience.

It is critically important that users be given tools inside the product to communicate to the company about their experience of unwanted experiences and harm in the product. These should take the form of a statistically significant rolling survey. Companies have well developed methodologies to create surveys that accurately represent the different populations that use their product. As well as tools to communicate when something goes wrong. And when they use such tools, not only should they get a useful response from the company, but their experience should contribute to anonymized aggregated data that the companies are required to compile and publicly report.

My years of work in engineering, product, security, and compliance have convinced me that the most effective way to regulate the harms that social media enables is to require platforms to use and report different metrics than those they have historically used. (In the past the companies have not sufficiently tracked negative experiences of users.) If the proper metrics are transparently reported, a societal dialogue can develop which will eventually lead to our teenagers operating in safer online spaces.

We hear often in the press of new kinds of harms being experienced on these services. We also hear about existing harms being handled inappropriately. Whenever that happens, it represents a failure in the design or implementation of the service's tools for reporting problems..

Below, I include a variety of recommendations for specific ways that regulators can require data to be regularly published by social media platforms.

A few notes on the recommendations that follow:

- Social media companies have robust tools and sophisticated processes to do "quantitative surveys". These are surveys that gather enough data to be representative of a population. what's known as "statistically-significant." This surveying approach is sustainable because it reduces the number of people that need to be surveyed, and applies different well-known techniques that can reduce the impact of bias in survey responses.
- I say users ought to be able to "flag" content when it causes them discomfort. By this I mean giving users a simple way to indicate to the service that an interaction or piece of content is unwanted. It could be done by swiping, hitting an 'X' button, or by some other method. Flagging should always be followed by the option to give more context. If the

user does not want to supply this information, that option should be able to be easily ignored by scrolling away. Such interactions should feel rewarding and easy to use.

- The process to follow in order to develop the right options in user dialogues about harms they experience can be based on the document [Lessons learned while working on online bullying](#).
- "Support tools" help people with an issue they encounter, regardless of whether it would properly be "reported" for content review. Depending on the nature of the issue, the support tools might include the option for the user to submit a report.
- "Report flows" are the steps people take to submit content for content moderation review.

Any social media service that is used extensively by teenagers should be required to gather and publicly report the following sorts of data:.

1. Create and deploy a survey process that asks questions of teenagers who have experienced harm. This should be done through a quantitative survey, based upon a statistically-significant population. The survey should be carefully designed in order to understand the experience of harms encountered by teenagers across the service, and be reflective of the demographics of the full population (for example age, location). Results should be disclosed based upon metrics like age, gender, geography, race, etc.

   Example questions:
   - In the last 7 days, did you receive any unwanted sexual advances?
   - In the last 7 days, has someone insulted or disrespected you, spread rumors about you, threatened you, or excluded you?
   - In the last 7 days, have you seen someone saying something discriminatory against someone else because of their body?
   - In the last 7 days, have you felt worse about yourself because of someone else's posts?

   Data should include:
   - The area or feature of the service where the harms have been experienced. Was it in direct messages? In the comment on a post or photograph? In a photograph itself? Each of these categories should be reported.
   - A breakdown of such experiences by the age of the user.
   - What did the person do after they experienced the harm?

2. In order to ensure that teens are always able to report problems or ask for help when they need to, the following quantitative survey should be run routinely, and the results ought to be publicly reported every quarter. It should be a stated goal that 90% of teenagers should answer yes to both of the following questions. If that percentage is not achieved,

penalties should be levied against the company. The goal ought to be to make sure that there is something users can easily do when they have a bad experience, and that users know it. They also should feel the company's response helps them deal with the issue at hand.

- If something bad happens to me or someone else:
    - I know what to do, and it is helpful.
    - The actions I take help make the community safer.

3. Users ought to be able to flag any content, using tools that give the company information about why that piece of content made them uncomfortable. The results of such interactions ought to be tracked and quantified and reported publicly. Examples include:

- **Unwanted contact in messages.**
  As part of the design of any direct-messaging feature, the user should have an easily-findable "unwanted contact" button. They should be given the option after they click it to provide context. Users should be given choices such as the following, to indicate why the contact was unwanted:
    - It's gross.
    - It's fake.
    - It is harassing me.
- **Unwanted actions on my content.**
  People should be able to flag unwanted comments, regardless of whether the content violates policy. Options should include:
    - It's attacking someone else.
    - It's gross.
    - It is disrespectful.
- **Unwanted content in my feed or in other recommendation areas.**
  The algorithms social media services use to create recommendations often serve up inappropriate suggestions. People should thus be able to flag unwanted content that is recommended to them, and provide a reason why. For example:
    - It's gross.
    - Someone is going to get hurt.
    - It is an inappropriate depiction of body image.
- The options that a user is given to block, restrict, or take other actions on content created by others should be accompanied by secondary options that easily give the person the chance to indicate why they took that action. For example, they may block someone merely to mute an annoying friend for a day, or they may be seeking to keep away a predator. Without having such context, it is very difficult

for a platform to use the information that someone was blocked as a means to make the community safer.

4. Teens should be provided with **support tools** that help them deal with the issue they are experiencing, regardless of whether they submit a report. As part of the support tools, it should be clear how to **submit a report for review**. Support tools should be measured by subsequent surveys.
   a. A significant majority of users should reply "yes" to the following question: "I got help with the issue I am experiencing." The percentage ought to be publicly reported.
   b. There should be an option to submit feedback even if the person chooses to not use the tool. Some sample questions may include:
      ○ There is no option for my issue.
      ○ I don't think reporting this issue will help me.

5. Every quarter, companies should be required to transparently report a variety of statistics regarding the usage of **support and reporting tools**. The following are some of the kinds of information that should be required to be disclosed:
   ● The number of people who enter the report flow (see definition above).
   ● The percentage of users who abandon the flow at each step.
   ● Out of the people who entered the flow, the percentage of those who submitted content for review.
   ● The percentage of report requests that resulted in content or an account being removed, banned, etc.

In making these recommendations I am not necessarily arguing for an increase in the number of people reviewing content. My own experience in following some of these steps during Facebook's Compassion work was that most content sent to review does not violate policy. When we added options that reflected user experience, for example by helping teenagers say that someone else was being annoying, the volume of reports that needed to be reviewed went down. This freed up critical resources and people to focus on the reports for which they could take the most useful actions.

I believe that giving teenagers a voice around the things they experience will, over time, make a safer environment. These recommendations are not a solution. They are a necessary start to the work that is needed in order to create an online environment where teens can express themselves and learn to be respectful to each other. This should be an environment where appropriate teasing and playfulness is welcome, but where unwanted contact and harassment is not. I believe the only way this can happen is through regulation imposed from the outside. Meta has consistently demonstrated that it will not address these issues on its own.

Biography of Arturo Bejar, Former Director of Engineering for Protect and Care, Facebook.

From 2009 to 2015, Arturo was the senior engineering and product leader at Facebook responsible for its efforts to keep users safe and supported, reporting to Mike Schroepfer, the CTO. Arturo was responsible for "Site Integrity" – stopping attacks and malicious behavior; "Security Infrastructure" – which engineered resilient systems and worked on compliance; and a group called "Care"– which developed Facebook's user-facing and internal customer care tools, as well as child safety tools. Arturo was responsible for the combined effort of engineering, product, user research, data, and design. This included regularly doing strategic product reviews with the Facebook executive team. Arturo was also the engineering manager for Facebook's "Product Infrastructure" team, which built key parts of the product engineering frameworks of Facebook, and developed REACT, one of the core technologies of the web today

From 2019 to 2021, Arturo returned to Facebook to work as a part-time independent consultant and industry expert for the Well-being team at Instagram.

In 2022, Arturo was a Technical Advisor, for the Facebook Oversight Board.

Prior to that Arturo was recruited to Facebook from Yahoo! where he worked from 1998 to 2009 . Arturo was hired as Yahoo!'s first security engineer, eventually becoming the head of Information Security reporting to the CTO.

Arturo started working for IBM in Mexico City when he was 15, was able to study Mathematics at King's College London thanks to the support of Steve Wozniak, and first started working on security and social systems in Silicon Valley in 1994 as part of a startup called Electric Communities.