Tennessee v. Meta
Case No. 23-1364-IV
PX-00778

Letter draft

Dear Adam,

It's been almost two years since I started supporting the Well-being team at Instagram. During that time and through my conversations with people I have found what, in my experience, is the most important thing for the company to be working on. I know that you and the rest of the leadership team deeply care about the people we serve and the communities we are trying to nurture, yet there are many challenges. I would like to bring to your attention that today we don't understand, or have solutions that improve the community for most of the bad experiences people have on Instagram. I also believe we have a great product opportunity here to innovate by investing in tools that help develop social norms. I believe that this gap is responsible for the cycles of finding new classes of harmful content through the outside, and some of the concerns people have about our services.

51% of Instagram users say 'yes' to having experienced one or more of the TRIPS questions. 1% of those report and 2% of reports have the content taken down. Of 10,000 people who have a bad experience on Instagram, 100 report, and 2 get the content taken down.

My daughter, now 16, is what we call an experimenting creator. She posts about cars, and in a recent post someone commented 'Get back to the kitchen'. It was an awful experience for her. She tried to report it, to no avail as this is something far from violating policy, and so her only options were to delete or block.

Why does someone think it is ok to post that? I believe it is because it doesn't violate policy, and other than deleting or blocking, there is no feature that helps people know that kind of behavior is not ok.

There are many examples of this. Today we don't know what % of content people experience as harassment or racism racism is policy violating. We have done great work in driving down prevalence, and there will always be more to do, but what if policy based solutions only cover a small percentage of what is affecting people?

Policy is necessary when the content is clearly inappropriate, yet it has many limitations. It trails behavior, the interventions are heavy, and getting the border line right is extraordinarily difficult. Policy enforcement is analogous to the police, it is necessary to prevent crime, but it is not what makes a space feel safe. It is possible that over 99% of the content that causes bad experiences is not policy violating.

What makes a workplace, or a school, or a dinner table feel safe is social norms.

Today most of the areas we are investing in make bad experiences less visible: hide/delete/down rank/block/personalize/unfollow. While this is the best tool for some issues, if someone goes around telling women to 'get back to the kitchen', and the only thing that happens is their content is deleted, don't we run the risk of normalizing bad behavior? How are people to learn to be members of a safe and supportive community without visible interventions that help set the social norms for the environment?

I believe the most important thing we should be working on is understanding the bad experiences that people are having on the platform, what is the content that causes them, how intense their experience is, and how they relate to policy violations. If we learn that 90% of the content that causes bad experiences is not something for which we could write an objective policy, what would we build?

This is hand in hand with the opportunity of building visible interventions that help set the social norms for our communities. There is a great opportunity to innovate where our tools grow from making bad experiences invisible, to developing tools where there is no burden on the victim. We empower the creators and community to set the norms for their spaces, and we engage with the actors who are the source of the content that drives bad experiences. If we can separate out the people who would behave differently given feedback from those who are intentionally malicious, all of our systems become more effective.

I have been working with the Well-being team on this, and we have a few people working on understanding bad experiences, but it is the things that you name as priorities that are fully resourced: Youth, Creators, etc. I believe this process of understanding bad experiences, and building tools to create social norms, is essential for Youth. From watching my daughter when she was younger, the things she found distressing online like rumors, people being mean, were not things that would be covered by our current tools. Preliminary data in the space supports this, as well as the data from when we had different flows for people under 18 on Facebook.

I might be wrong about my assessment, and welcome feedback about any effort or data that I'm missing. I believe that it is important to get the following efforts well-funded and prioritized:

- What is the content that is causing bad experiences for our users? How intense is the experience?

- What % of that content is policy violating? (i.e. how much of TRIPS is driven by content other than what drives Prevalence?)
- What are visible product solutions that make the community better over time? e.g. actor feedback, comment covers, pinned comments, etc.

The solutions we create I believe should have the following properties:

- The person who has the negative experience should feel heard, you don't 'perceive' racism or harassment, you experience it, and you are the source of truth for that.
- We should empower creators, communities, and Instagram, in setting the social norms for the spaces they are a part of.
- Where appropriate we should give feedback to actors, in the belief that they are acting with good intention and might have caused unintentional harm. This will allow us to separate the people who would behave differently given feedback, from the ones who are intentionally causing harm.

If you would like I can give more details or specifics on this. I am appealing to you because I believe that working this way will require a culture shift. I know that you, Mark, Sheryl, and everyone in the leadership team deeply cares about the people we serve, and the communities we are trying to nurture, and I believe that this work will be of service to that.

Arturo
Contact Support

Conversation History

David Gillis opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted] Ted [redacted], Shubhro [redacted], Sam [redacted] Boo [redacted] Ashish [redacted], Fritz [redacted] Zvika [redacted] Elizabeth [redacted] Yoav Shapira, Peter [redacted] Nino [redacted] Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Oct 21, 2021 at 4:51 pm

---

Robbie [redacted] opened your conversation with Miki Rothschild, Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted] Yoav Shapira, Peter [redacted] Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Oct 14, 2021 at 2:09 pm

---

Sam [redacted] added you to a document Oct 14, 2021 at 2:08 pm

---

Sam [redacted] added Brian [redacted] and Robbie [redacted] to the thread Oct 14, 2021 at 2:14 pm

---

Sam [redacted] added you to a document Oct 14, 2021 at 2:08 pm

---

Arturo Bejar edited Sep 24, 2021 at 6:10 am

- What % of that content is policy violating? (i.e. how much of TRIPS is driven by content other than what is the relationship between TRIPS and drives Prevalence?)

Arturo Bejar added David Gillis to the thread Sep 22, 2021 at 7:17 pm

---

Arturo Bejar added you to a document Sep 22, 2021 at 7:11 pm

Andrew [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted] Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted] Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 21, 2021 at 8:59 pm

---

Sayed Otaru added Andrew [redacted] to the thread Sep 21, 2021 at 9:00 pm

---

Sayed Otaru added you to a document Sep 21, 2021 at 8:55 pm

Boo [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted] David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted] Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 21, 2021 at 7:45 pm

---

Arturo Bejar Sep 20, 2021 at 11:44 pm

Thank you Ted [redacted] - added a paragraph with more examples, not as specific, but pointing out what we don't know.

Ted [redacted] Sep 20, 2021 at 11:30 pm

+1 to everyone. Maybe you link to a list of other less violating, but still bad experience examples Arturo Bejar

---

Ted [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted] Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted] Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 20, 2021 at 11:25 pm

---

Shubhro [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted] Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 20, 2021 at 10:42 pm

---

Arturo Bejar added Shubhro [redacted] to the thread Sep 20, 2021 at 10:44 pm

---

Arturo Bejar added you to a document Sep 20, 2021 at 10:38 pm

---

Arturo Bejar Sep 20, 2021 at 10:38 pm

I think in order to do a survey of solutions, what would be needed is to be able to place them in a bad experiences framework. For example what % of user bad experiences in a particular category is addressed by soft actions for borderline content? That would help us better prioritize or assess the impact of our work.

Arturo Bejar Sep 20, 2021 at 10:07 pm

Thank you for the feedback, I think what I'm trying to communicate is that it would be good to change how we approach our measurement, as well as different solutions. I did make the choice of not making this a survey of solutions, in my experience when I used to work with m-team, there is value to when you let them know that something needs their support. My goal in this is to help get resources and funding into these areas, and a mindset that gets us faster to other kinds of features that will help create a safer community, as well as the kind of measurement that help us better measure our work.

Arturo Bejar Sep 20, 2021 at 10:00 pm

I made the language more precise, as the point here is that we don't have a good causal understanding of content, to action, to how users experience it, as well as community impact. Another example is the conversation with my daughter about unsolicited penis pictures, she tried profile reporting, but ultimately what happens between her and her friend is they block the person. who then feels it is ok to send pictures to other people. Part of the hope of this work is to look a remediation differently. where we can assess if the person got the support they needed, and where appropriate there is also action to help make the community better. Right now we don't have many features whose goals is to improve the community over time.

Arturo Bejar Sep 20, 2021 at 9:54 pm

Yes I think so - ultimately we are talking about comments, reels, posts, etc., while the action of posting them is an interaction, I think you start with content level understanding, and the fit it into behavioral context.

Arturo Bejar edited Sep 20, 2021 at 11:45 pm
51% of Instagram users say 'yes' to having experienced one or more of the TRIPS questions. 1% of those report and 2% of reports have the content taken down. Of 10,000 people who have a bad experience on Instagram, 100 report, and 2 get help the content taken down.

There are many examples of this. Today we don't know what % of content people experience as harassment or racism racism is policy violating. We have done great work in driving down prevalence, and there will always be more to do, but what if policy based solutions only cover a small percentage of what is affecting people?
Policy is necessary when the content is clearly inappropriate, yet it has many limitations. It trails behavior, the interventions are heavy, and getting the borderline border line right is extraordinarily difficult. Policy enforcement is analogous to the police, it is necessary to prevent crime, but it is not what makes a space feel safe. It is possible that over 99% of the content that causes bad experiences is not policy violating.

- Understand What is the content that is causing bad experiences for our users? How intense is the experience?

Nino [redacted] Sep 20, 2021 at 9:42 pm

I think this specific example, but also some of the others across the note point to bad experiences with interactions on IG - interactions in comments and messages. We know that this is an area of major concern for us where even policy-violating experiences are not "taken care of". I wonder if it might be more fruitful to focus on on these surfaces as an appeal for resourcing.

Nino [redacted] Sep 20, 2021 at 9:40 pm

Not sure that I agree with 2 get help. I assume self-remediation isn't part of this 2, right? Also, reporting is only 1 self-remediation option among many, and we have shown that self-remediation via blocks/restricts/etc. does result in better outcomes.

Nino [redacted] Sep 20, 2021 at 9:37 pm

Is this incremental to what is being done today or meant to be comprehensive of today's solutions as well? For example, where does self-remediation for victims and actors fall? Where does soft-actioning for borderline content fall? I think being comprehensive and highlighting what we do and don't do (or don't do well) would be helpful.

Nino [redacted] edited Sep 20, 2021 at 9:35 pm

Nino [redacted] Sep 20, 2021 at 9:31 pm

Are we using content as a catch-all bucket? I think of causes to bad experiences being content AND interactions, but there is also account-level issues (e.g., compromised accounts), etc.

Nino [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted] Ashish [redacted], Fritz [redacted] Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 20, 2021 at 9:23 pm

---

Arturo Bejar added Boo [redacted] to the thread Sep 20, 2021 at 5:00 pm

---

Arturo Bejar added you to a document Sep 20, 2021 at 4:55 pm

Arturo Bejar edited Sep 20, 2021 at 4:38 pm
It's been almost two years since I started supporting the Well-being team at Instagram. During that time and through my conversations with people I... the communities we are trying to nurture, yet there are many challenges. I would like to bring to your attention that today we don't measure, understand, or have solutions that improve the community for most of the bad experiences people have on Instagram. I also believe we have... responsible for the cycles of finding new classes of harmful content through the outside, and some of the concerns people have about our services.

Arturo Bejar Sep 20, 2021 at 4:38 pm

Thank you - I'll take out the word measure. the distinction here is that while we have a measurement of magnitude (i.e. TRIPS/BEEF), but I think it is accurate to say we don't understand - as we have no content level understanding, or accurate behavioral understanding. For example, we don't have an understanding of how ptrips relates to the content that has bad experiences, and any proxy metrics that rely on pReport likely address a small % of the content driving bad experiences, I think that is one of the reasons why most of the interventions we ship have small effects on pTrips/Trips.

Miki Rothschild Sep 20, 2021 at 2:39 pm

You may want to refine "we don't measure, understand and solve" since technically we have measurement for these (TRIPs, BEEF etc) and a whole set of workstreams to make those more operational (PIB, pTRIPs etc). Also, you cite their prevalence lower down which may confuse the reader.

Perhaps you mean to say that we don't have operational metrics for them, and the strategy being pursued (reduce) is incomplete or not the right one?

Miki Rothschild edited Sep 20, 2021 at 2:35 pm

Arturo Bejar Sep 18, 2021 at 1:18 am

Done, and done, thank you.

Arturo Bejar edited Sep 18, 2021 at 1:25 am
51% of Instagram users say 'yes' to having experienced one or more of the TRIPS questions. 1% of those report and 2% of reports have the content taken down. Of 10,000 people who have a bad experience on Instagram, 100 report, and 2 get help.

Lori [redacted] Sep 18, 2021 at 12:37 am

might be helpful to clarify - yes to having experienced...

Lori [redacted] Sep 18, 2021 at 12:36 am

he probably knows TRIPS but just in case we can link to the dash https://fburl.com/unidash...

Lori [redacted] edited Sep 18, 2021 at 12:31 am

Lori [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 18, 2021 at 12:29 am

---

Arturo Bejar added Ted [redacted] to the thread Sep 17, 2021 at 11:21 pm

---

Arturo Bejar added you to a document Sep 17, 2021 at 11:16 pm

---

Arturo Bejar Sep 17, 2021 at 11:07 pm

Yes, I agree, thank you, I put IG here instead of Facebook, and will edit to Facebook once it goes beyond IG.

Arturo Bejar edited Sep 17, 2021 at 11:06 pm
Today most of the areas we are investing in make bad experiences invisible less visible: hide/delete/down rank/block/personalize/unfollow. While this is the best tool for some issues, if someone goes around telling women to 'get back to the... are people to learn to be members of a safe and supportive community without visible interventions that help set the social norms for the environment?

- We should empower creators, communities, and the community Instagram, in setting the social norms for the spaces they are a part of.

Arturo Bejar Sep 17, 2021 at 11:04 pm

Great change, thank you.

Elizabeth [redacted] Sep 17, 2021 at 10:30 pm

Since there are many overlapping communities on Instagram, I wonder if we might include Facebook among the list of actors here with an opportunity and responsibility to contribute to the setting of healthy social norms. Though we mention Facebook's role in giving actors feedback for negative interactions in the next point, we may have more levers to pull in norm-setting that include modeling and increasing neutral to positive interactions.

Elizabeth [redacted] Sep 17, 2021 at 10:18 pm

You could consider "less visible" instead of "invisible," since of course despite all efforts, negative interactions are all too evident across the app.

Elizabeth [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 17, 2021 at 10:08 pm

---

Peter [redacted] Sep 17, 2021 at 9:36 pm

"...empower creators and communities in setting..." or "...creators and their communities..."? The IG community pillar recently changed their name to IG Communities to reflect that it's not just 1 IG Community.

Peter [redacted] edited Sep 17, 2021 at 9:35 pm

Yoav Shapira opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 17, 2021 at 8:05 pm

---

Samir [redacted] edited Sep 17, 2021 at 7:51 pm
Today most of the areas we are investing in make bad experiences invisible: hide/delete/down rank/block/personalize/unfollow. While this is the best tool for some issues, if someones someone goes around telling women to 'get back to the kitchen', and the only thing that happens is their content is deleted, don't we... are people to learn to be members of a safe and supportive community without visible interventions that help set the social norms for the environment?

Samir [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted] and Arturo Bejar Sep 17, 2021 at 7:50 pm

---

Arturo Bejar added you to a document Sep 17, 2021 at 7:31 pm

Arturo Bejar added you to a document Sep 17, 2021 at 7:31 pm

Arturo Bejar added you to a document Sep 17, 2021 at 7:31 pm

Arturo Bejar added Elizabeth [redacted], Lori [redacted], Nino [redacted] and Yoav Shapira to the thread Sep 17, 2021 at 7:36 pm

---

Arturo Bejar added you to a document Sep 17, 2021 at 7:31 pm

Arturo Bejar edited Sep 17, 2021 at 7:21 pm

It's been almost two years since I started supporting the Well-being team in at Instagram, during . During that time and through my conversations with people I have found what, in my experience, is the most important thing for the company to be working on. I know that you, and the rest of the leadership team deeply care about the people we serve, and the communities we are trying to nurture, yet there are many challenges. I want would like to bring to your attention that today we don't measure, understand, or have solutions that improve the community for most of the bad experiences people have on Instagram. I also believe we have a great product opportunity here to innovate by investing in tools that help develop social norms. I believe that this gap is responsible for the cycles of finding new classes of harmful content through the outside, and some of the concerns people have about our services.

51% of Instagram users say 'yes' to one or more of the TRIPS questions. 1% of those report, and 2% of reports have the content taken down. Of 10,000 people who have a bad experience on Instagram, 100 report, and 2 get help.

Policy is necessary when the content is clearly inappropriate, yet it has many limitations. It trails behavior, the interventions are heavy,... not what makes a space feel safe. It is possible that over 99% of the content that causes bad experiences is not policy violating. If you are the target most of the

Today most of our tools the areas we are investing in make bad experiences invisible: hide/delete/down rank/block/personalize/unfollow. If While this is the best tool for some issues, if someones goes around telling women to 'get back to the kitchen', and the only thing that happens is their content is deleted, don't... are people to learn to be members of a safe and supportive community without visible interventions that help set the social norms for the environment?

This is hand in hand with the opportunity of building visible interventions that help set the social norms for our communities. There is a great of room opportunity to innovate where our tools grow from making bad experiences invisible, to develop developing tools where there is no burden on the victim, we . We empower the creators and community to set the norms for their spaces, and we engage with the actors who are the source of the content that drives bad experiences, to try and . If we can separate out the people who would behave differently given feedback from those who are intentionally malicious, all of our systems become more effective.

I have been working with the Well-being team on this, and we have a few people working on understanding bad experiences, but it... that you name as priorities that are fully resourced: Youth, Creators, etc. I believe this process of understanding bad experiences, and building tools to create social norms, is essential for Youth. From watching my daughter when she was younger, the things she found distressing... Preliminary data in the space supports this, as well as the data from when we had different flows for people under 18 on Facebook.

- Understand the content that is causing bad experiences for our users? How intense is the experience?
- What % of that content is policy violating? (i... ?)
- What are visible product solutions that make the community better over time? e.g. actor feedback, comment covers, pinned comments, etc.
- We should empower creators, and the community, in setting the social norms for the spaces they are a part of (needs work/better phrasing).

If you would like I can give more details or specifics on this. I am appealing to you because I believe that working this... Sheryl, and everyone in the leadership team deeply cares about the people we serve, and the communities we are trying to nurture, and I believe that this work will be of service to that.

Peter [redacted] edited Sep 17, 2021 at 6:37 pm

Miki Rothschild opened your conversation with Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted] Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted] Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 17, 2021 at 5:16 pm

---

Arturo Bejar added Miki Rothschild to the thread Sep 17, 2021 at 5:20 pm

Arturo Bejar added you to a document Sep 17, 2021 at 5:15 pm

Ashish [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted] Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 17, 2021 at 5:12 pm

Arturo Bejar added Ashish [redacted] to the thread Sep 17, 2021 at 5:06 pm

Arturo Bejar added you to a document Sep 17, 2021 at 5:01 pm

Peter [redacted] Sep 17, 2021 at 4:52 pm

Kyle Andrew's recent presentation on Bad Experiences on the IG showed that youth on the platform express having these negative experiences at higher levels than older demographics. I'll grab that slide and share it here.

Arturo Bejar edited Sep 17, 2021 at 5:22 pm
Policy is necessary when the content is clearly inappropriate, yet it has many limitations. It trails behavior, the interventions are heavy,... not what makes a space feel safe. It is possible that over 99% of the content that causes bad experiences is not policy violating. If you are the target most of the
I might be wrong about my assessment, and welcome feedback about any effort or data that I'm missing. I believe that in order it is important to get the following efforts should get funded well-funded and prioritized:

- We should empower creators, and the community, in setting the social norms for the spaces they are a part of (needs work/better phrasing).

Peter [redacted] edited Sep 17, 2021 at 4:50 pm

Peter [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 17, 2021 at 4:04 pm

Arturo Bejar added Peter [redacted] to the thread Sep 17, 2021 at 4:09 pm

Arturo Bejar added you to a document Sep 17, 2021 at 4:04 pm

Arturo Bejar Sep 17, 2021 at 3:45 pm

Yeah, I think you're right. I don't want this to feel like I'm shaming, at the core of this is the belief that the leadership team deeply cares, and what is needed is to move our metrics frame or reference, and to start investing in other kinds of solutions.

Arturo Bejar Sep 17, 2021 at 3:43 pm

Added a 'the', rather than types or or kinds, because while we will get there what's most needed right now is understanding specific pieces of content, and how they relate to policy.

Arturo Bejar edited Sep 17, 2021 at 4:25 pm
It's been almost two years since I started supporting the Well-being team in Instagram, during that time and my conversations with people I have found what in my experience is the most important thing for the company to be working on. I know that you, and the rest of the leadership team deeply care about the people we serve. and the communities we are trying to nurture, yet there are many challenges. I want to bring to your attention that today we don't measure, understand, or have solutions that improve the community for most of the bad experiences people have on Instagram. I believe this gap is responsible for the cycles of finding new classes of harmful content through the outside, and some of the concerns people have about our services.
My daughter, now 16, is what we call an experimenting creator. She posts about cars, and in a recent post someone commented 'Get back to the kitchen'. It was an awful experience for her. She tried to report it, to no avail as this is something far from violating policy, and so her only options were to delete or block.
Policy is necessary when the content is clearly inappropriate, yet it has many limitations. It trails behavior, the interventions are heavy,... Policy enforcement is analogous to the police, it is necessary to prevent crime, but it is not what makes a space feel safe. It is possible that over 99% of the content that causes bad experiences is not policy violating.
Today most of our tools making make bad experiences invisible: hide/delete/down rank/block/personalize/unfollow. If someones goes around telling women to 'get back to the kitchen', and the only thing that... are people to learn to be members of a safe and supportive community without visible interventions that help set the social norms for the environment?

This is hand in hand with the opportunity of building visible interventions that help set the social norms for our communities. There is great of room to innovate where our tools grow from making bad experiences invisible, to develop tools where there is no burden on the victim, we empower the creators and community, and we engage with the actors who are the source of the content that drives bad experiences, to try and separate out the people who would behave differently given feedback from those who are intentionally malicious.
I have been working with the Well-being team on this, and we have a few people working on understanding bad experiences, but it... that you name as priorities that are fully resourced: Youth, Creators, etc. I believe this process of understanding bad experiences, and helping tools to create social norms, is essential for Youth. From watching my daughter when she was younger, the things she found distressing online like... Preliminary data in the space supports this, as well as the data from when we had different flows for people under 18 on Facebook.

- Understand the content is causing bad experiences for our users? How intense is the experience?

Fritz [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted] Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted] Ashish [redacted], Zvika [redacted] David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted] Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 17, 2021 at 1:08 pm

---

Sayed Otaru Sep 17, 2021 at 4:51 am

It might be helpful - if known - to add some more context about how she may have tried to find resolution in the app (but didn't have viable options?)

Sayed Otaru Sep 17, 2021 at 4:50 am

It might be helpful to bridge this sentence with the next sentence with a qualifier about the opportunity to provide people more tools to inform us when they have had a bad experience knowing that they will get a satisfactory outcome from their expression.

Sayed Otaru Sep 17, 2021 at 4:40 am

I think this is an incomplete sentence.

Sayed Otaru edited Sep 17, 2021 at 4:50 am

Sayed Otaru opened your conversation with Miki Rothschild, Robbie [redacted], Kyle Andrews, Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted], Nino [redacted], Brian [redacted], Arturo Bejar and Samir [redacted] Sep 17, 2021 at 2:08 am

Arturo Bejar added you to a document Sep 16, 2021 at 9:57 pm

Arturo Bejar added Fritz [redacted] and Sayed Otaru to the thread Sep 16, 2021 at 10:02 pm

Arturo Bejar added you to a document Sep 16, 2021 at 9:57 pm

Kyle Andrews opened your conversation with Miki Rothschild, Robbie [redacted], Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 16, 2021 at 9:09 pm

Arturo Bejar added Kyle Andrews to the thread Sep 16, 2021 at 8:15 pm

Arturo Bejar added you to a document Sep 16, 2021 at 8:10 pm

Arturo Bejar Sep 16, 2021 at 4:46 pm

Fixed, thank you.

Arturo Bejar Sep 16, 2021 at 4:44 pm

Great, thank you.

Arturo Bejar Sep 16, 2021 at 4:44 pm

Fixed.

Arturo Bejar edited Sep 16, 2021 at 4:46 pm
Today most of our tools are making bad experiences invisible: hide/delete/down rank/block/personalize/unfollow. If someones goes around telling women to 'get back to the kitchen', and the only thing that happens is... are people to learn to be members of a safe and supportive community without visible interventions that help set the social norms for the environment?
I believe the most important thing we should be working on is understanding the bad experiences that people are having on the platform, what... the content that causes them, how intense their experience is, and how they relate to policy violations. If we learn that 90% of the content people experience as that causes bad experiences is not something for which we could write an objective policy, what would we build?
I have been working with the Well-being team on this, and we have a few people working on understanding bad experiences, but it... , the things she found distressing online like rumors, people being mean, were not things that would be covered by our current tools. Preliminary data in the space supports this, as well as the data from when we had different flows for people under 18 on Facebook.

Sam [redacted] Sep 16, 2021 at 2:16 pm

And also signal boosted to the point that there's a broader culture shift not just for the specialized teams working specifically here, but also more broadly across the org?

Sam [redacted] Sep 16, 2021 at 2:13 pm

It could be powerful to summarize the thesis up top.

Riff:

'Many people are having bad experiences on Instagram and our current solutions of hiding some instances aren't addressing the source of issues so most people's bad experiences are unaddressed.'

Sam [redacted] edited Sep 16, 2021 at 2:14 pm

Sam [redacted] Sep 16, 2021 at 1:21 am

"are about making an attack invisible"?

Sam [redacted] edited Sep 16, 2021 at 1:21 am

Zvika [redacted] Sep 16, 2021 at 12:41 am

Isnt this the stat that you mentioned at the top? If not, maybe clarify what that was and what still needs to be measured?

Zvika [redacted] Sep 16, 2021 at 12:40 am

"What types of content"?

Zvika [redacted] Sep 16, 2021 at 12:39 am

Maybe its obvious, but do you think it's worth saying here or in the intro that this is not just about protecting people/keeping them safe (aka social responsibility) -- this is about users having a positive experience on the platform, which gets at growth/retention and core metrics.

Sam [redacted] Sep 16, 2021 at 12:36 am

Adam may not have the internal context that this is reacting to? (I think this anchors on the review deck sent to Schrep that framed bad experiences as "perception".)

This point could land by introducing how the current internal terminology affects our internal mindset

and how that translates to what happens for people.

Zvika [redacted] Sep 16, 2021 at 12:35 am

I think its important to add that our preliminary data/signals back this up (so it doesnt seem like youre making the case for a company priority based on your daughter's experience, which I know youre not :-)

Zvika [redacted] Sep 16, 2021 at 12:34 am

harmful content? negative content?

Zvika [redacted] edited Sep 16, 2021 at 12:42 am

Sam [redacted] Sep 16, 2021 at 12:32 am

It might be effective for Adam as your audience to remind that we do give actors feedback with "comment warning" and we were early to innovate with nudges in comments.

Yet these are implemented based on policy-based classifiers.

So the gap this is asking to prioritize is about taking an approach similar to the existing nudges and extending beyond a policy implementation.

Zvika [redacted] opened your conversation with Miki Rothschild, Robbie [redacted] Kyle Andrews, Lori [redacted], Andrew [redacted], Tec[redacted], Shubhro [redacted], Sam [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 16, 2021 at 12:30 am

---

Sam [redacted] Sep 16, 2021 at 12:27 am

extra word?

Sam [redacted] edited Sep 16, 2021 at 12:34 am

Sam [redacted] opened your conversation with Miki Rothschild, Robbie [redacted], Kyle [redacted], Lori [redacted], Andrew [redacted], Ted [redacted], Shubhro [redacted], Boo [redacted], Ashish [redacted], Fritz [redacted], Zvika [redacted], David Gillis, Elizabeth [redacted], Yoav Shapira, Peter [redacted], Nino [redacted], Sayed Otaru, Brian [redacted], Arturo Bejar and Samir [redacted] Sep 15, 2021 at 11:28 pm

---

Arturo Bejar added you to a document Sep 15, 2021 at 11:15 pm

---

Arturo Bejar added you to a document Sep 15, 2021 at 11:15 pm

---

Arturo Bejar added Sam [redacted], Samir [redacted] and Zvika [redacted] to the thread Sep 15, 2021 at 11:20 pm

---

Arturo Bejar added you to a document Sep 15, 2021 at 11:15 pm

---

Arturo Bejar renamed Letter to Letter draft Sep 15, 2021 at 11:20 pm

---

Arturo Bejar edited Sep 15, 2021 at 11:15 pm

Arturo Bejar edited Sep 15, 2021 at 3:53 am

Arturo Bejar edited Sep 14, 2021 at 2:27 pm

Arturo Bejar created the document Sep 14, 2021 at 2:15 pm
