



Instagram Bad Experiences

What are Bad Experiences and why we should care

Earlier this year, we met with Sam and Arturo who have been spending a lot of time thinking about bad experiences on Instagram, and how we can make the experience better for users. Since then, we've been helping to create a narrative on this work, including what bad experiences are, why they are important, and recommendations for how to better address them. Today we want to share that narrative with you, along with our recommendations and next steps for your feedback.

Caveat, that this is still pretty early stage. We have shared versions of this with several people in IGWB and partnership teams for feedback and input, but we are very open to all feedback and candor.

PLTF EX

PLTF EX-00767

Meeting Objectives

Goals

1. Share how we currently understand and address Bad Experiences
2. Discuss next steps and proposed commitments

Non-Goals

1. Product strategy for Bad Experiences
2. Deep dive into existing integrity/policy interventions

2

Goals

Share what Bad Experiences look like and how they manifest
Share how we currently understand and address Bad Experiences
Get feedback on a proposed approach and next steps

Non-Goals

Product strategy for bad experiences
Comprehensive overview of integrity/policy interventions

Selena Gomez On The Comments That Drive Her To Delete The Instagram App Once A Week

'Star Wars: The Last Jedi' actress Kelly Marie Tran deleted all her Instagram posts after months of harassment

Trolls force Billie Eilish to quit social media



Khloé Kardashian Disabled Her Instagram Comments After Racist Trolls 5-Month-Old Daughter True

Alia Bhatt's mother Soni Razdan lashes out at Instagram for not protecting her daughters from abuse

Soni Razdan, mother of Shahen and Alia Bhatt, has slammed Instagram for being seen to protect themselves from abuse. Shahen had also called out the platform for failing to curb harassment.

[BOLLYWOOD] | Updated Jan 14, 2020 2:11 PM

1st Entertainment Desk
Hindustan Times



The actress who plays AT&T's Lily is facing waves of online sexual harassment, including manipulated images and objectifying memes

The Fall of 'Terrace House'

Japan's heartwarming reality TV hit ended in tragedy when the participant Hana Kimura killed herself after a wave of online abuse. But "Terrace House" was always more complicated than it appeared.

We're no strangers to headlines like this, and can be hard to know how much to trust headlines given our current press environment. But what these headlines, partnership teams, and the data tells us, is that public figures and creators are increasingly dissatisfied with IG around their safety on the platform. And that the negative experiences they have on IG - from mass harassment and bullying to impersonation and misinfo and more - can lead to real emotional impact.

When we dig deeper, it's clear that these headlines are not isolated incidents.

These claims aren't isolated

56%

Aspiring Creators who experienced bullying in the last 30 days¹

1/4

1 in 4 people think Instagram makes social comparison worse²

36%

US creators satisfied with their well-being on Instagram³

¹ "Bullying norms on Instagram", Hitomi Hayashi-Branson, Jan. 2021

² Social Comparison on Instagram

³ Wave 3 Creator Sentiment Survey

56% aspiring creators exp bullying. And those witnessing bullying, equally high numbers.

1 in 4 people think IG makes their experience of social comparison worse - particularly true for teen girls + non-binary (37% said they "Often" or "Very often" felt worse about their appearance or body because of other people's IG posts). Worse than our competitors.

36% of creators satisfied with their well being. About lower numbers in major markets around the world, and in SR 12%

And what's worse is that users don't think IG cares... For instance:

Only 21-55% of users believe that Instagram disapproves of bullying.

7/10 people who think they've been shadowbanned believe that IG doesn't care about them

Despite all the tools and policies we currently have, this is still coming across loud and clear from users...

Notes

Satisfaction w well being even lower in major markets (UK 31%, AU 30%, Germany 25%, KR 12%) - q: how well do you feel protected from harmful experiences?

124% Increase in mass B&H cases through Early Response Esc in 2020 YoY

https://docs.google.com/presentation/d/1JGwxSmCLYTe3ZDROMQ-s6-U3mh0UsQcngHnC6C06fe0/edit#slide=id.gc717be7cff_0_13

U3mh0UsQcngHnC6C06fe0/edit#slide=id.gc717be7cff_0_13

Users feel that IG is the worst app for social comparison compared to our competitors (Snap, TikTok)

https://docs.google.com/presentation/d/1PHGJS6j0w-uCKT8J_u-Zz-FpeHKyTdvQ2vrGceI8cgI/edit#slide=id.gd62ca52962_6_50

And we're only scratching the surface

.02%

of bad experiences are
estimated to lead to
action¹

1. TRIPS and FRX data, 2021.

According to TRIPS data, we roughly estimate that only 1% of people who have a negative experience on IG report it. And of those reports, about 2% lead to action.

We have a *sense* of why this is happening, (people might think reporting won't do anything, reports don't fit their needs, may be cultural factors at play), But we don't *fully* know. What we do know is that there's a lot of opportunity in this area.

On the reporting side, we know that:

Reporting mechanisms don't support what users want to express

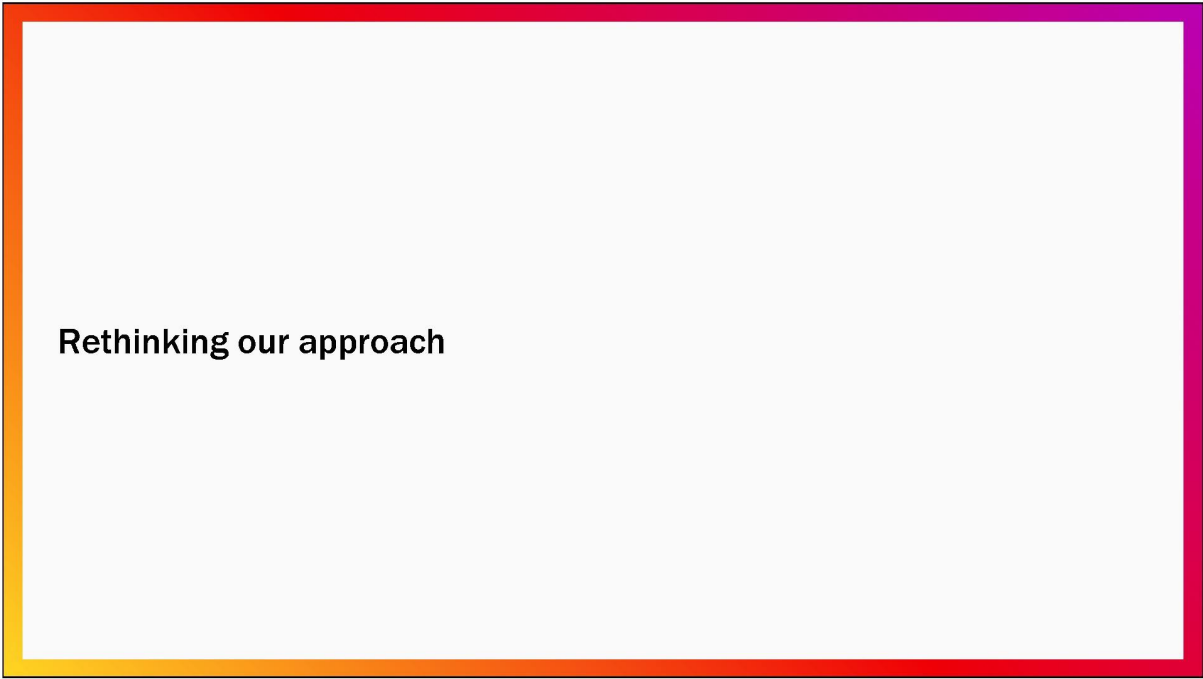
Users have low confidence in follow-up actions

And for our younger users, for examples, there are offline social concerns around impacting relationship with their peers

And on the action side:

Policy definitions which inform enforcement that don't always match how users define their experience

Individual negative experiences are anchored in context that policies can't always account for



Rethinking our approach

So one of the questions we can ask ourselves, looking at these stories and numbers, is whether there might be an opportunity to rethink our approach and move towards better support for our users when they are faced with these experiences?

How we define Bad Experiences

Any Interaction with content or an account that a user considers harmful or personally objectionable.

They can include anything that makes a user feel unsafe, unwelcome, or uncomfortable

They can vary depending on contextual dimensions (language, culture, and connection) and can be result of cumulative exposure.

For these reasons, they go beyond policy definitions of harms.

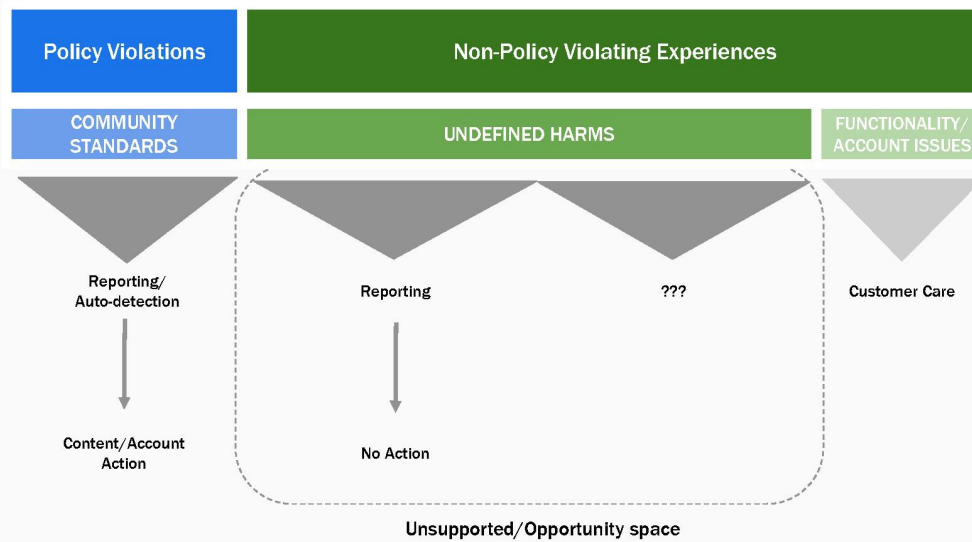
7

But first take a step back and set the context of how we define a bad experience.

For all these reasons they can go beyond our policy definitions because they center the truth of how that experience is felt by the user.

And from here we should put in perspective our current approach to supporting these experiences.

How we currently address Bad Experiences



Right now, we have 2 top-level buckets of bad experiences:

Policy Violations = categories of harms that are clearly & systematically defined in our Community Standards, and support would generally follow the flow of user reporting or auto-detection that would result in a content or account-level action

Non-policy violating experiences =

a subset of those are issues related to functionality & account setups, and we have managed partnerships in place and setting up a Customer Care organization to support those

But then, we have this middle bucket of experiences and potential harms we don't have defined solutions for. Reflecting back on the .02% number we mentioned earlier, some of these get reported but lead to no action (because they don't meet policy definitions) and for the remainder we simply don't have a good view of what happens next for users.

What this demonstrates is that while Policy and associated integrity-based solutions are a vital foundation for tackling bad experiences, limiting ourselves to a "policy-first approach" exposes us to gaps in the support that we could be providing our users. There is an opportunity space of this broader universe of bad experiences that we may not be seeing, defining and measuring, but which are still painful for our users (and in recent incidents, those have come in the form of mass harassment, non-violent threats, and body image issues)

Limits of our "policy-first" approach

Policy is effective, but ...

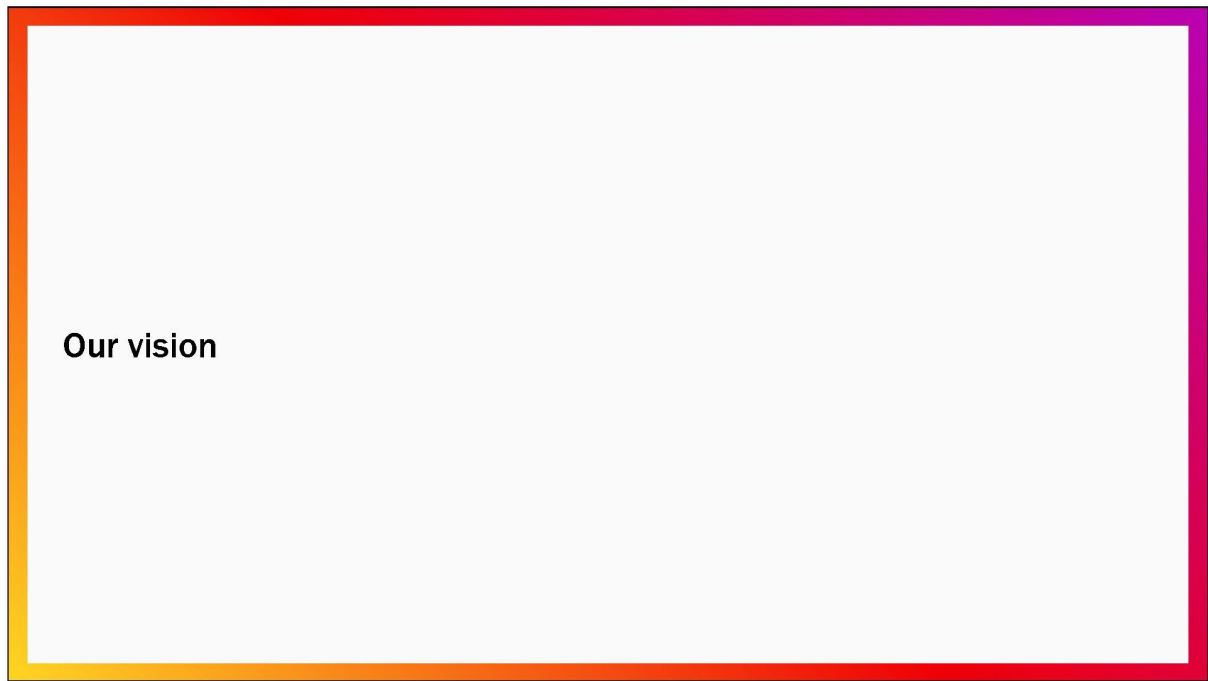
1. It's constrained to specific definitions
2. It cannot keep pace with dynamic user experiences
3. It limits our detection of "grey area" or emerging issues
4. It doesn't capture the contextual nature of bad experiences

Policy isn't the right tool when ...

1. we want to center the truth of our users' experiences
2. "bad" is impossible to determine with any consistency
3. support calls for an alternative approach to the violator/non-violator binary model

9

There is no questioning the effectiveness of our policies, but we want to recognize the impact of some inherent limitations
Is constrained to specific definitions that do not always capture the experiences our users are facing
Policies must be reliable, operationalized, & enforced systematically at scale - which requires that we carefully consider and evaluate any potential change before implementation
Limits dictation of grey area issues or emerging threats
Bad experiences are anchored in contextual dimensions such as language, culture, connection, or history. These changing nuances cannot be captured in our policies and often aren't visible to our enforcement mechanisms.
What we can acknowledge from there is that Policy may not be the right tool when:
We want to be people-first
Bad is impossible to determine with consistency
Support calls for alt approach to the violator/non-violator binary model.



What's our ideal state? How do we approach this from a people-centric perspective?

Instagram is a place where, everyone - including young people, creators, and marginalized communities - feels safe, feels supported, and feels cared for both by the platform and by the broader community.

Approaching this from a more people-centric view, our vision is:

Instagram is a place where, everyone - including young people, creators, and marginalized communities - feels safe, feels supported, and feels cared for both by the platform and by the broader community.

Policy and enforcement is crucial, but it's not enough to try and protect. We also want users to feel cared for-- and not just by IG alone, but how to we foster a safe space where the community cares for each other?

To get there, let's take a look at an example from the real world.



Bullying in schools:

All 50 U.S. states require schools to have a bullying prevention policy. But despite that, bullying remains a huge problem. (35% increase from 2015-18). What a lot of the most recent data and research tells us when it comes to bullying, is that policies are very important, but they are not enough. There are several reasons why: programs that rely on punishment have not been shown to be effective in the U.S.; and they often disproportionately target students of color. Authority figures have limited influence. Programs like that place responsibility on the children to work out conflicts can increase bullying. (Adult victims of abuse are never asked to “work it out” with their tormentor.) Bystander intervention, even among adults, only works for some people—extroverts, empaths, and people with higher social status and moral engagement.

What works: a 2019 paper from the Yale Center for Emotional Intelligence and Yale Child Study Center, Social norms: Schools with a positive climate foster healthy development, along with Social and emotional learning (SEL) i.e. teaching kids skills of self-awareness, self-management, social awareness works. While a negative school climate is associated with higher rates of student bullying, aggression, victimization, and feeling unsafe.

2016 study 30% drop in bullying - recreating social norms. Found popular kids/ influencers in schools to condemn bullying alongside campaigns (events, hashtags, t-shirts). Princeton, Yale, Rutgers

Why this is an interesting example/ what can we learn from this:

What new mental models can we use to reframe how we’ve traditionally approached bad experiences?

Policy is table stakes, but what if we need something bigger to make a lasting impact?

If we continue on our path today, do we run the risk of normalizing bad behavior, in the same way bullying has been normalized (as part of growing up) for years?

Teaching our users to be part of a community can also help separate out the bad actors from people who have good intentions but don’t see what they’re doing as wrong (or those who think it’s expected to do XYZ bad thing)

https://greatergood.berkeley.edu/article/item/what_are_the_best_ways_to_prevent_bullying_in_schools

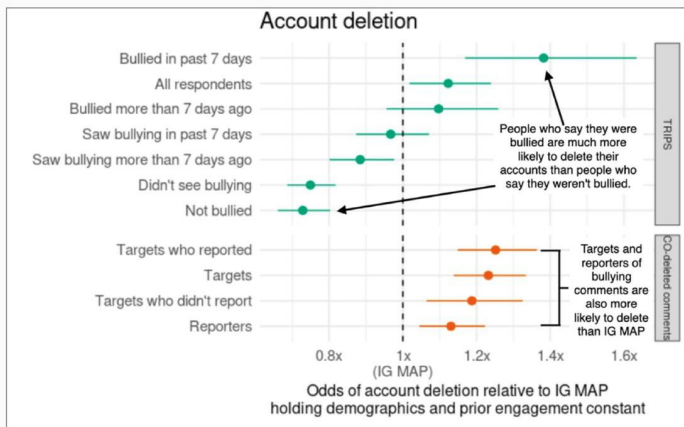
<https://www.princeton.edu/news/2016/01/04/students-influence-over-peers-reduce-school-bullying-30-percent>

What do we risk if we don't hold ourselves accountable?



Building off this example, an important question to ask ourselves is : what do we risk if we don't hold ourselves accountable to fulfilling the vision we described?

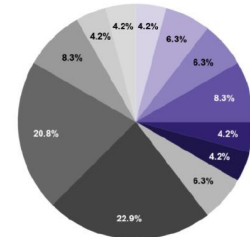
Impact on engagement



Appearance comparisons are more likely to result in visits to other apps for distraction. Users mostly choose TikTok and Snapchat.

Negative Appearance Comparison Go-To Apps

- Reddit
- Internet Browser
- Music app
- Facebook
- Pinterest
- Facebook Messenger
- Twitter
- Snapchat
- TikTok
- I did not visit another
- Photo editing app (e.g.,
- Text Messages



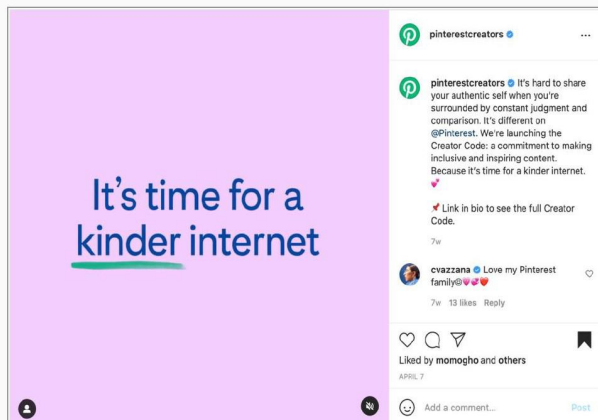
We've discussed some of real emotional impacts that these experiences can have on people, such decreased sense of safety and control,

But there are also indications that bad experiences impact the way people decide to engage with the platform and actually push them towards disengagement. Examples:

We have data showing that people who have been bullied are more likely to deactivate, or delete their accounts and leave the platform entirely (Odds of deactivation are 1.9-2.2x compared to IG MAP, and odds of deletion are 1.2-1.4x) - and this usually happens because they haven't found a possibility of recourse or do not have faith in their ability to get support

Another example is data that we've collected around appearance and social comparison that shows that these issues are more likely to be experiences on Instagram, and lead users to regularly jump onto competitor's platform to disengage and feel better. Not about having to take a break from social media per se, but specifically from IG because of the intensity of the experience.

Competitive approaches



SOMEWHERE GOOD

"Somewhere Good helps users facilitate a *better* type of online interaction."
"an opportunity to actualize the futures we envision as individuals and as a collective."¹



Our mission is to give children a safe and positive community where they can find joy, develop healthy online relationships, and aspire to their greatest potential.

1. www.somewheregood.com/

This inconsistencies in engagement metrics seem to be informing some competitive approaches and pushing smaller platforms like Pinterest to center their messaging around a more supportive product experience, or "kinder" products; and newer platforms and using this as a competitive advantage or selling point-- marketing their platform as being inclusive and kind compared to big players, building their narratives around "better" futures, healthier relationships and positive & supportive communities.

<https://www.instagram.com/p/CNXndGmh7qD/?igshid=hzn3v6vnoghg>

Imagine a place where young people and creators

Identity Development

Mass Influence

define themselves and the future.

Societal Impact

Our new vision for IG is focused on young people: tweens and teens who are mentally, socially and culturally developing into adults, learning norms about how to interact with others and forging their identities on our platform. It also focuses on creators, who are the influencers, drivers of mass behavior and shapers of norms.

Supporting these groups in defining the future means that we need to pay particular attention to the type of experiences we may be normalizing on the platform: left unsupported, non-policy violating bad experiences become part of a day-to-day negative "normal" against which young people are shaping their future-selves, and which creators are amplifying as a standard. The questions become:

what is our responsibility as a platform to not only "imagine a place where young people and creators define themselves and the future" but to also ensure that this place helps them develop into responsible digital users and in turn bring positive value to society?

How do we hold ourselves accountable to proactively and intentionally normalize more supportive experiences and productive behaviors through our products?

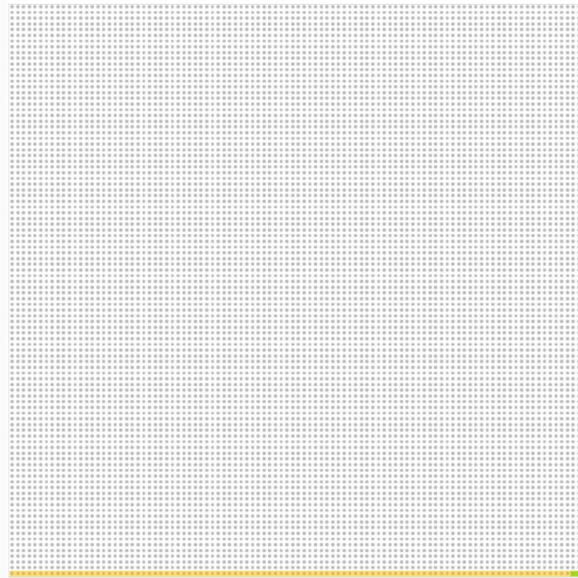
Uncovering the unknown unknowns

What % of these bad experiences are related to policy violating content?

What do people do once they've had one of these experiences?

What actions do they take outside of reporting channels?

How does this impact their immediate and long-term behavior on the platform?

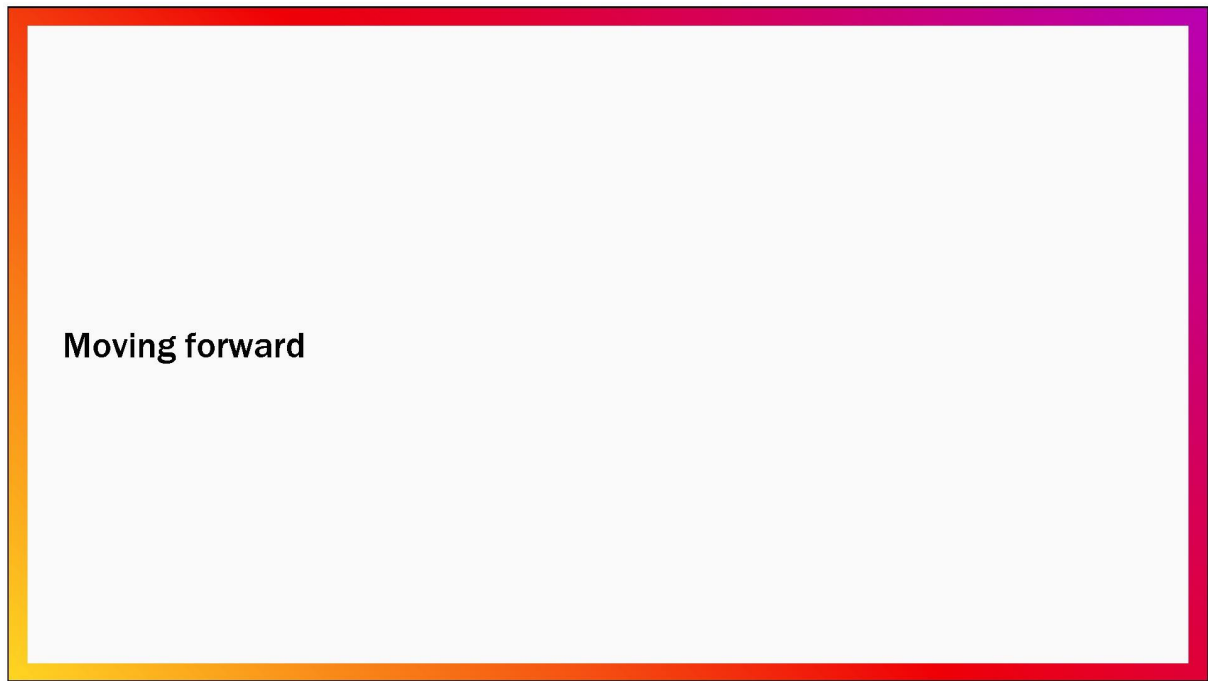


One way is to first discover what we DON'T know about those experiences that may lead us to negative societal outcomes. Looking at this graph with the dots, the yellow line represents the bad experiences that we fully acknowledge today (which mostly track with our policy violations) and the green is what we address.

Today, we have a sense of what Bad Experiences look like, but we don't fully understand the universe that they cover and the impact they are having on our users and on our platform: for example, what actions are users taking outside of our reporting channels? Are they aware of the impact of their actions on others? How would such awareness affect their behavior?

What we need a stronger way to detect and measure these experiences so that we can better identify and design user-centric support solutions for them. And what we are saying here is that in order to a product that positively impacts people's experience, we need to understand what happens in the gray area of this chart.

Passing over to Arturo to talk about the path forward.



How do we move forward?

Our recommendation

Make Bad Experiences the ground truth that we use to support people.



This means:

The ground truth across the company, in addition to prevalence

Establishing Social Norms

Product interventions - IGWB as not seeing this as a separate effort from IG vision/priorities

Ongoing measurement and understand work

Support and care integration

Our ask

1. Measure + understand what we don't know: get better understanding of people's behavior underlying and reacting to bad experiences
1. Action:
 - a. PM working group to summarize the state of bad experiences measurement on IG relative to FB App/CI
 - b. Relevant integrity teams commit to specific reduction goals around bad experiences



Including H2 goals around Bad Experiences for IGWB
Take up this topic with leadership (Chris? Adam?)
Support from VPs, including socialization around this being a priority
Need their support in order to move this forward
Exploring social norms with external and diverse perspectives
Expanding product goals beyond IGWB for H1 2022

Questions