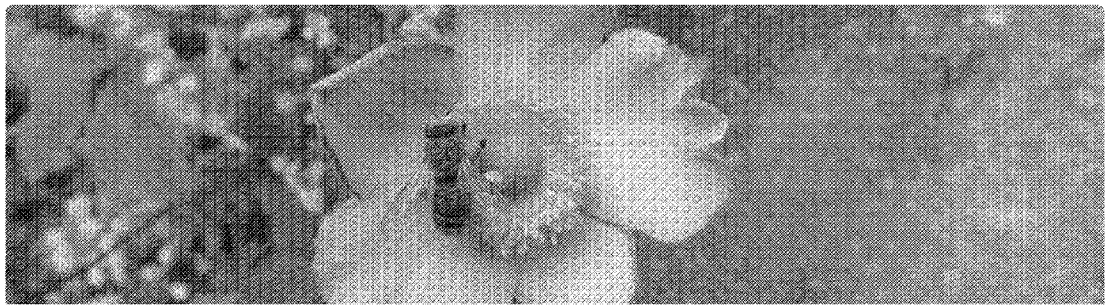
Reducing Bad Experiences for Our Users

Arturo [redacted] and 2 others · Last edited November 6, 2021 · 7 minute read

Save ···



Reducing Bad Experiences for Our Users

TL;DR: Reducing bad experiences on our platforms goes beyond black-and-white policy and requires creative approaches. How do we give people and creators the right tools to be good citizens of a safe and supported community? Through this note, I share some key questions and opportunities for us as we build new products and features for our community.

My daughter Joanna, 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. So, her only options were to delete or block.

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

For some context, I started supporting Site Integrity and Customer Care in 2009, when there were only 17 people on the team. We grew the team that became Protect and Care, where I supported engineering, product, user research, data analytics, and design for 6 years. With some great engineers, I also started Product Infrastructure and Security Infrastructure. Two years ago, I came back as a part-time independent contractor to support the Well-being team at Instagram to help tackle some of the complex issues we face as an industry. I believe figuring out ways to help people be respectful of each other online, while having different values, is one of the most important problems of our age.

There is a lot of great work being done by the Well-Being and Integrity teams around the company. And as I wrap up my contract on IG Well-being, I wanted to take this time to share 3 important opportunities I see for teams working in the space:

1. Can we take people's *experiences* as our north star?
Today, much of our work to identify and resolve bad experiences is rooted in whether or not the experience was caused by something that violated our policies. What if instead we took our users' experiences as our ground truth? More specifically:

PX-01398, Page 1 of 7

Tennessee v. Meta
Case No. 23-1364-IV
PX-01398

HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY

METATNAG-022-00000380

- Can we set goals on user experience (such as what we see in TRIPS) in addition to Prevalence? For example, goals around people who experience being targets of bullying or harassment independent of the content.
- What if we stop using "perceived" and instead use the word "experienced"? People don't perceive being harassed, they experience it.
- Finally, the key to this, is how can we better understand content or interactions that drive non-policy violating bad experiences?

2. Are we giving people the opportunity to express their bad experiences?
There is a tension between the concept of reporting and giving people support with what they are experiencing. *Reporting* is about people communicating to us content that they believe violates our community standards. *Support* is about helping people with what they need.

Should we reframe our *reporting* flow into a *support* flow that gives people the right tools to express their experiences and feel taken care of? Ideally, a *support* flow should serve 3 purposes: 1) inform us of what a user is experiencing and 2) provide the support and tools for what they are experiencing 3) continue to help us direct enforcement for policy violation.

For example: Joanna often gets unwanted sexual advances in messages, like 'friends with benefits?' from other teens and friends of friends. This is a negative experience for her, but it's not always clear how to express this on the reporting flow. Also how could we capture all the ways people make unwanted advances? When people come into someone's DMs, maybe we should consider a "my house, my rules" approach: It doesn't matter what people say---if it feels like a sexual advance in my space, it is. Additionally, if the only action Joanna can take is to block unwanted sexual contact, when she blocks, the actor who initiates is none-the-wiser. Without the context from Joanna about what happened we can't make our systems smarter about what she's experiencing, as well as provide feedback to the actor.

3. What tools can we create to improve our community over time?
Our integrity efforts primarily focus on taking down bad content, but when the goal is to stop bad things, you're always going to be surprised by the latest.

If there was a way in the product for people to express their bad experience, we could identify patterns and give the actors feedback---and the opportunity to learn from their behavior. Research shows that most people who post reported content did so with positive intentions (thinking it was important, funny, or helpful, etc). Just as we are with our co-workers, we must first assume good intent and give them feedback to learn and change. If the person doesn't change their behavior after the feedback, then depending on the context, you can have different measures, such as feature limits.

This is why it's important to understand groups of users who cause bad experiences:

- Good intent: people who would behave differently if they got feedback.
- Provocateurs: people who intentionally seek attention.
- Bad actors: people who set out to troll or intentionally to harm.

Understanding the difference between these 3 groups is important because they require different approaches. If our goal is to create a community that is safe and supportive, then we must **give people and creators more feedback and tools to be good citizens of this safe and supported community.**

I believe the best thing we can do is to develop more features that can create and maintain social norms. Social norms are what make a workplace, park, school, or other social settings feel safe---they also signal to people what's ok or not ok in a certain setting. This does not mean limiting

PX-01398, Page 2 of 7
HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY
METATNAG-022-00000381

voice; on the contrary, it protects it. It means helping people understand how their voice affects others. To do this well, we need a varied toolbox, such as:

- Visible norm-setting features: creators' pinned comments, custom comment covers that set the tone, gentle comment nudges, or after-the-fact nudges for actors
- Personalization, with transparency for creators, so that people can define for themselves what they find objectionable (such as more ways to personalize suggestions in Reels or Explore), and creators can get feedback from the community about what can be affecting their reach
- Policy and punitive measures for provocateurs or trolls who won't change, building on existing tools that we have.

Looking Ahead

The Well-being team at Instagram is working on features that, over time, will make a big difference:

- Giving creators better controls over their spaces, including the ability to pin comments
- Allowing people, after they are blocked, to tell us about their experience and use our reporting tools
- Giving actors respectful feedback when their content might have been hurtful to someone else
- Proactive social norm experiments

Everyone in the industry has the same problems right now. Prevalence-based policy measures, while necessary, don't create a safe and supportive community alone; you're always behind the latest harmful thing.

There is a great product opportunity here. Based on understanding of what people experience, what are visible features that help create a more respectful community? What classifiers and features would we build if we had better feedback from people about what they are experiencing?

It is possible to work these issues in partnership with industry leaders and academics. At the center of many of these issues is human behavior, and there is much we can learn from the people who study it. It is possible to create public conversation where we move forward understanding of these problems, with the knowledge that it will take time.

Getting these features right will take a long time, and each version will have many lessons. But if we start setting our goals on what users experience—if we help them express to us in our product what they experience in a way that makes them feel best supported—and if we create features that help develop and maintain social norms, we will lay a foundation that will make our communities safer and more supported over time.

89 23 comments

Like Comment Share

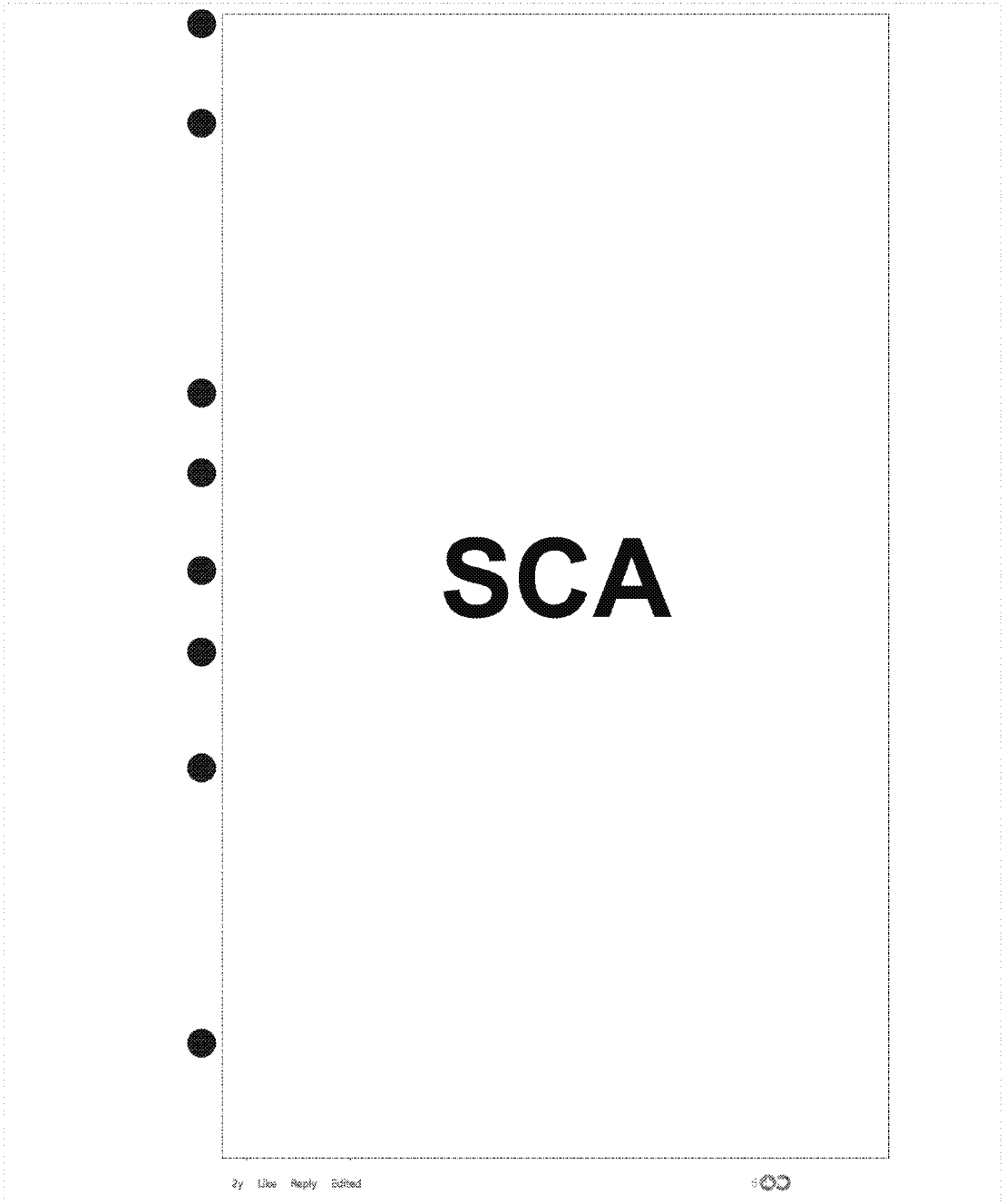
SCA

2y Like Reply Edited

PX-01398, Page 3 of 7

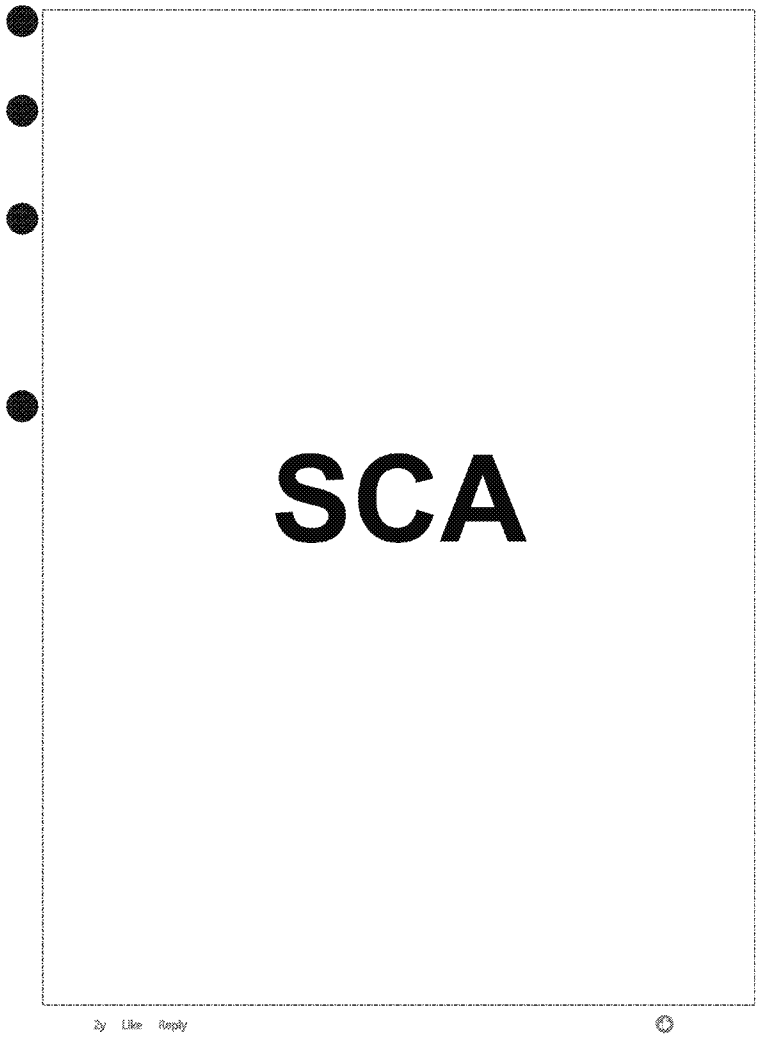
HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY

METATNAG-022-00000382



PX-01398, Page 4 of 7

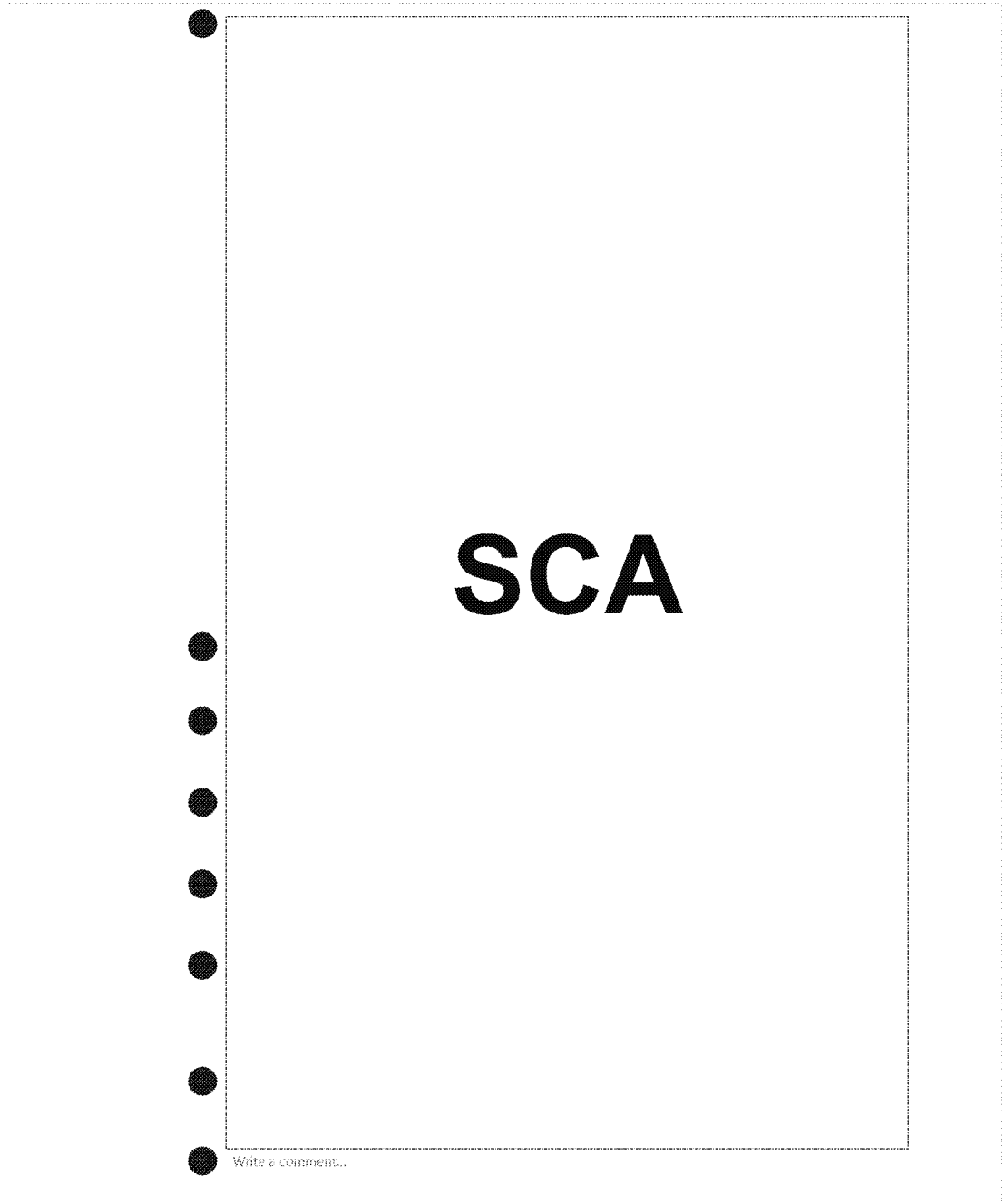
HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY



HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY

METATNAG-022-00000384

PX-01398, Page 5 of 7



HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY



PX-01398, Page 7 of 7

HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY