AMENDED Exhibit 213

PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

Document Provided in Native Format

Emerging Themes from Instagram Well-being: From Reels Integrity to Racial Justice
CDS 9.25

Agenda

01 Intros

02 Building Product at Instagram

03 Emerging Themes from Instagram Well-being

04 Instagram - CDS Collaboration

Building Product at Instagram

(and how it affects how we approach understanding our communities)

Instagram's mission is to bring you closer to the people and things you love.

Add well-being mission?

Instagram Values

People First

We are driven by making sure we solve problems for people. We support our community by fostering kindness and closeness that bring people together.

Simplicity

We do the simple thing first. We focus on the few ideas that create the most value and we make sure something is working well before we add to it.

Craft

We are thoughtful about what we create and we sweat the details. We take deep pride in our work so people can take pride in using Instagram.

Key things to note about Instagram Product Development and doing understand work

Communities of interest

- Teens
- Creators
- Small Businesses

Focus Countries Index (FCI)

- Australia, Brazil, India, France, Germany, Japan, Korea, United Kingdom, and the United States

Multiple Accounts

It's important to remember why we have the FCI in the first place, and it's clearly articulated in the first letter of the abbreviation: Focus. The primary benefit is that this list allows us to talk about a small, consistent subset of countries across the entire organization and ensure we can maintain one of our core values: Simplicity. Therefore, one of our goals is to keep the list of markets as small as possible while still reflecting the diversity of our global community. See post: https://fb.workplace.com/100023013870040/posts/540748470035611/

Instagram Well-being:

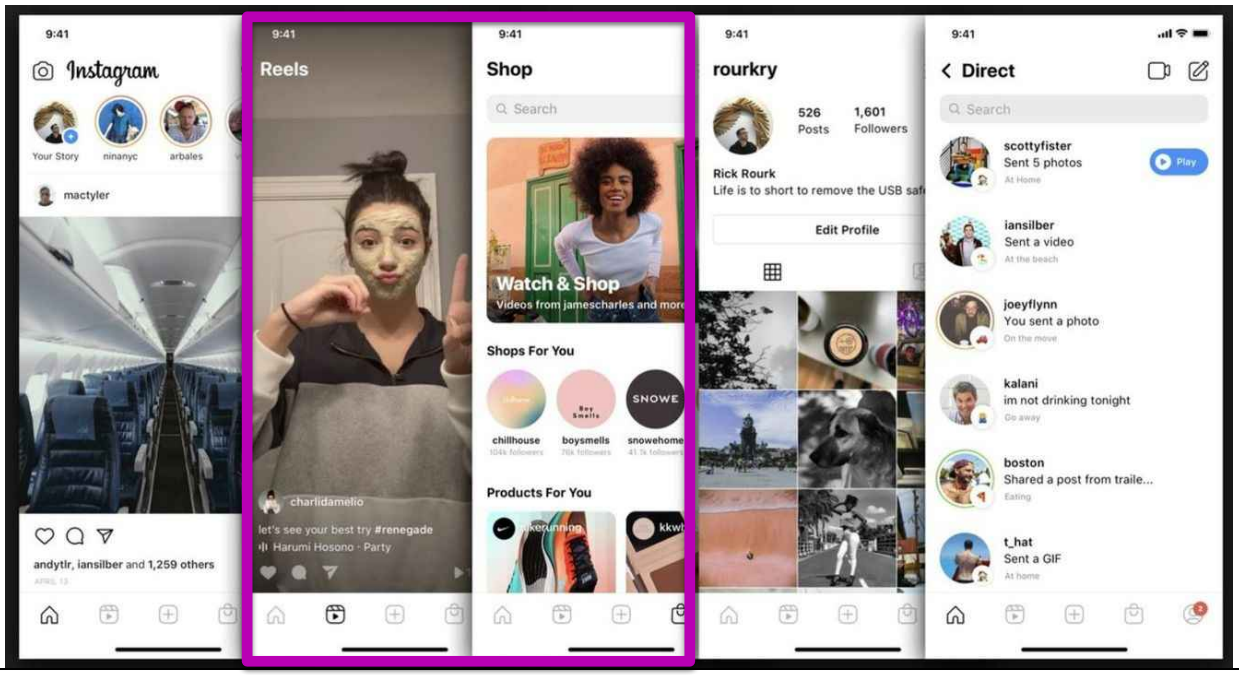
Our mission is to create the safest and most supportive global community.

Add well-being mission?

Emerging Themes from Instagram Well-being

Theme 1:
Users Are Increasingly Exposed to Unconnected Surfaces

Introducing two new “unconnected” surfaces: Reels Tab and Shopping Tab



https://fb.workplace.com/groups/fyi/permalink/3450206931703109/

On Sept. 9th, the Instagram team started testing several new looks for the home screen, creating dedicated tabs for Reels and Shop, setting up our big bets for success.

- We must continue to build for the next generation of young people and creators to ensure that Instagram is the place where culture moves forward.
- At Instagram, we focus on emergent culture, which means we are attuned to the activity of teens and creators: our core trendsetters.

Right now, we’re witnessing enormous shifts in how people create and consume culture.

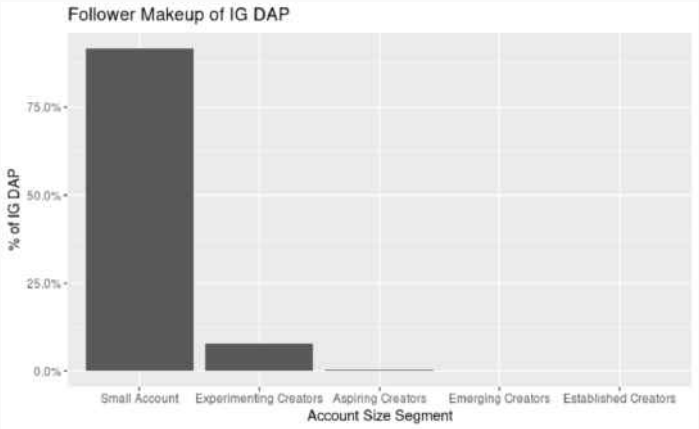
- Power has shifted from the organization to the individual, meaning that people are more interested in the creator behind the product than the brand they’re promoting.
- People want to be entertained. We’ve seen an explosion of video, which is why we’ve been focusing on IGTV and Reels.
- In the commerce space, people are shopping online in a more connected world, and millions of small businesses are adjusting, which is accelerating our work on Shopping.

Instagram’s home was originally designed for a Feed-centric world, but that’s not how people use the app today. One risk is not that we change too quickly, but that we become irrelevant. We need to lean into usage shifts and adapt.

- To that end, we’re beginning to test new navigation options, featuring dedicated tabs for Reels and Shop.

Reels is a strategic shift toward entertainment

- People prefer entertaining, short-form video content
- Creators are key to IG ecosystem
- Need to incentivize 'breaking' artists to create



https://fb.workplace.com/permalink.php?story_fbid=638923173551473&id=100023013870040

We believe the world is changing around us in terms of people's preferences for entertaining, short-form content and we need to respond.

Given that Explore is already serving this entertainment job with video chains (which account for approx. 20% of Explore TS), we are going to focus Explore to best serve this entertainment job.

We also believe the Reels format is our key ingredient to win this job with short-form, full-screen, mobile-native video, centered around performative categories like music, dance, and comedy.

A vibrant production ecosystem is the foundation of consumption in social products. In order for the content creation flywheel to be successful, we need to provide all of the right incentives for producers so that we can dramatically expand the inventory of content people find funny and entertaining.

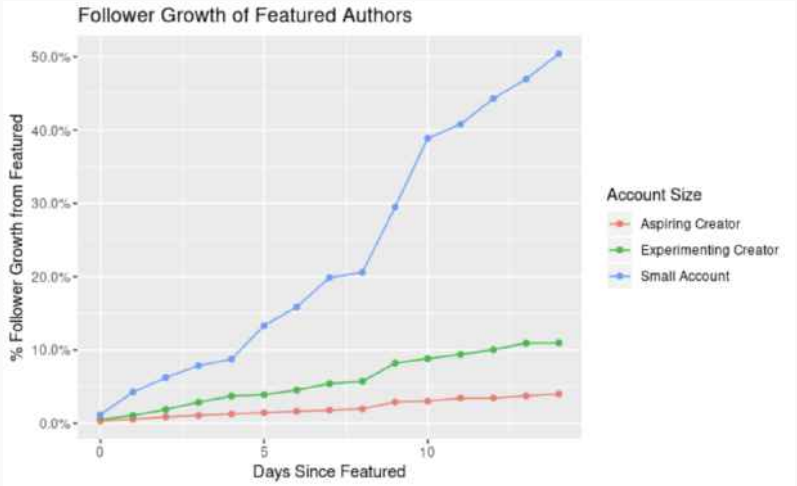
- In Reels v1, we reduced the barrier for production with creative tools but there wasn't enough of a payoff for people who put time into creating these videos.
- In Reels v2 by pulling it out of Stories and making it its own format with permanence and public feedback.
- We also need to ensure Reels has a proper stage and that we give aspiring creators a chance to "break" and quickly find an audience.
- We believe this stage needs to be a prominent part of Explore to give people confidence in their ability to get unconnected distribution and discovery.

Virality is at the core of Reels

Reels 'Breaking' Creators Goal Proposal:

- Deliver disproportionate distribution outcomes relative to author follow counts
- Example: Create moments where people get 10-100x plays relative to their following when they create great Reels

Outcomes of Virality (Source)



https://fb.workplace.com/groups/243894009320301/permalink/1186062491770110/

To make Reels successful, we need to create an accessible stage where the best content can find an audience, regardless of how many followers its author has. This means creating more viral moments for the best content with Star Search, and following that up with support from outreach and curation. If we do this well, we'll grow a new crop of talent on Instagram with Reels and deliver an entertaining experience for consumers.

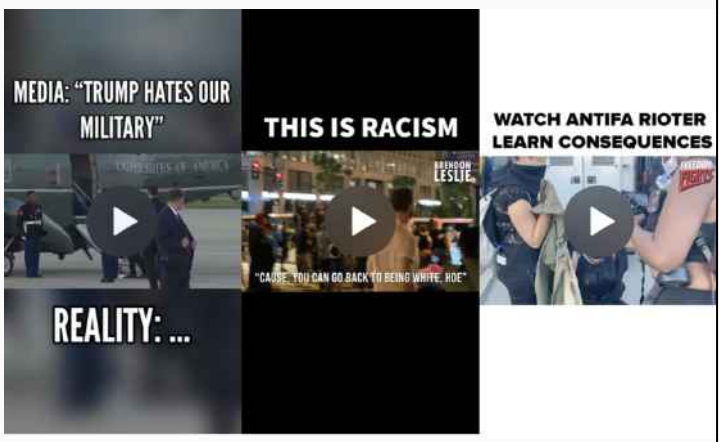
- Goal: We will goal on delivering disproportionate distribution outcomes relative to author follow count. In other words, creating moments where people get 10-100x plays relative to their following when they create great Reels. This is in service of our Producer PR goal.
- Focus: We will focus on small accounts (<1k followers), experimenting (1k-10k), and aspiring creators (10k-200k). The incentive should be to get more content from smaller creators but keep an eye on how this affects larger ones. Driving creation with them is the only way to hit our 5% producer PR goal.
- Track: We will track progression of people through the creator segmentation developed by the Creators and Discovery team. We are not goaling on follower growth. Over time we should find a way to track growth more directly.

Is it safe to amplify exposure to small accounts?

Accounts Associated with Top 10 Civic Reels Media by Unconnected VPVs (Source)

- donaldtrump_2020
- ryanafournier
- freedomfights
- nbacounty
- bathtubdubs
- donaldtrump_2020
- gopcharged
- hiphopcomeup
- freedomfights
- republicans.45

Only ~8k or ~20k followers, receiving 100k-300k imps



https://fb.workplace.com/groups/291827642072328/permalink/327590201829405/

A qualitative look at the predicted civic Reels getting the widest unconnected distribution (where the viewer does not follow the author) surfaced several disturbing trends.
In particular, this (small) sample suggests we are recommending content from users that regularly post likely misinfo, spam, or information related to conspiracy theories. The vast majority of these top predicted civic Reels were also from the right and far right, suggesting we may be amplifying certain viewpoints over others.

Types of harms included: Likely misinformation ("BLM is chanting 'death to America'"); Conspiracies (A video covering the 'Bill Gates' vaccine agenda); Spam ("Like this post for a free Trump)

We seemed to be amplifying extreme voices, and may also be disproportionately amplifying *one side* of the ideological spectrum

- Both the account and Reels samples had an *extreme* skew towards conservative or even-far right content, with ~7 out of the top 10 Reels and accounts coming from conservatives or the far-right.
- The surprising prevalence of this content could reflect disparate classifier performance on certain types of content, could reflect the overall prevalence of this type of content among Reels or other media, or, most

concerning, could reflect a bias in our recommendation engines towards inflammatory conservative content.

- In other words, our recommendation engines *could* be functioning in a manner similar to YouTube by rewarding highly-engaging but inflammatory content from the alt-right and right (see a summary of the crisis this sparked over at Youtube here). Of course, we would need to do *much* more data work to validate this particular hypothesis.

Many of the Reels getting wide unconnected distribution were from small accounts

- As expected given recent launches, many of the accounts generating the most unconnected civic impressions were from accounts with only ~8k or ~20k followers, but receiving 100k-300k or even >1m impressions on their civic Reels in a single week.

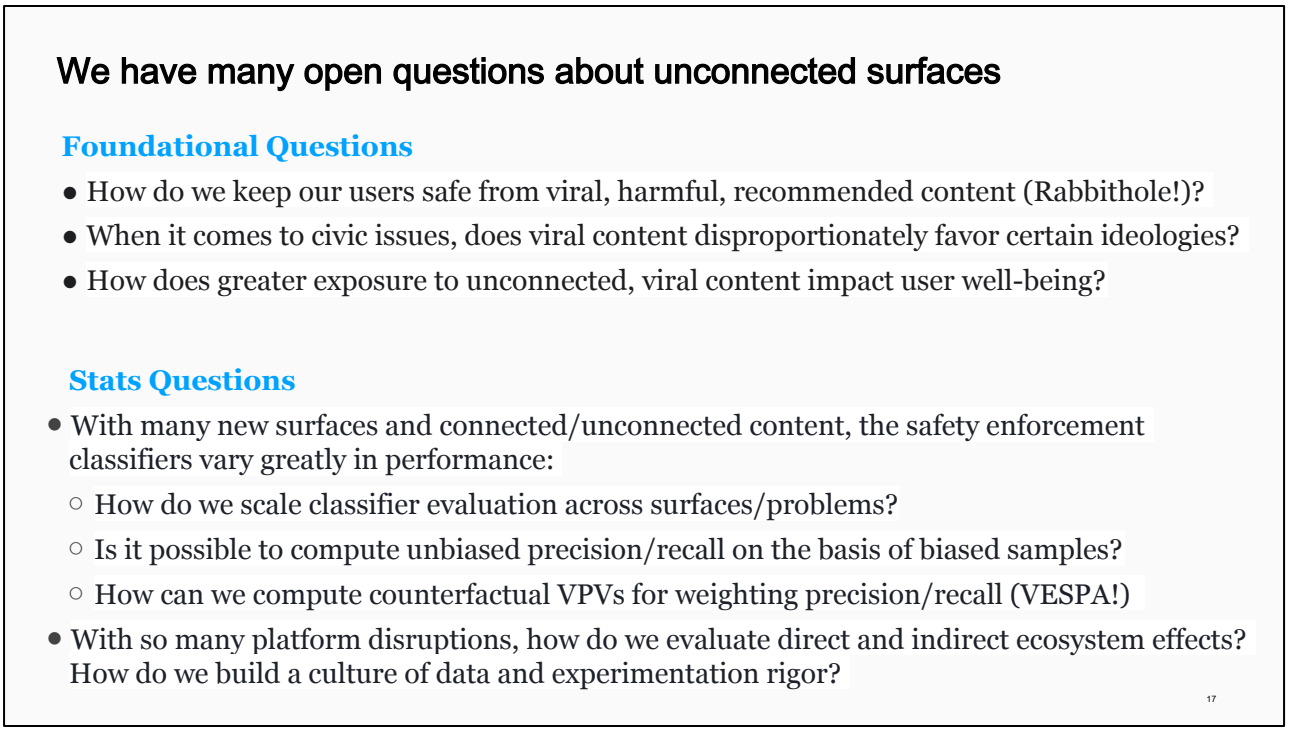
WHY THIS MATTERS

This is content that IG is *choosing* to show users that could cause real world harm or shape vote choices

Even if we don't care about our real world impact (which I of course think we should...), this *looks* really bad.

Amplifying small accounts is an extremely risky thing to do right before an election!

- In general, amplifying small accounts is more risky than amplifying big accounts.
- This is because we tend to know more about larger accounts, so we are more confident about what tier to put them in.
- If a larger (high follower count) account *is* harmful, larger accounts are also more likely than smaller accounts to have already triggered one of our existing integrity defenses, since most are dependent on extremely high-VPV 'trigger' thresholds (ID verification, username change review, HERO, or even review by 3PFC if the 3PFC opts to sort by views).
- The wide distribution given to small accounts likely compounds existing integrity vulnerabilities on Reels, because we expect most of our integrity classifiers to perform worse on this type of content (since most weren't trained on this type of media).



We have many open questions about unconnect surfaces at IG Well-Being. There are foundational questions about how to keep our users safe from viral, harmful content? I'm personally super excited by CDS's "Rabithole" project that works on understanding radicalization of recommendation systems. This is especially acute during this sensitive time, when exposure to misinformation and conspiracy theories can have consequential impacts on the election outcomes. One of the questions we are asking ourselves now is whether viral content disproportionally favors certain ideologies? Aside from just safety concerns, how does viral, unconnected content impact user well-being?

There are also a number of technical areas were we realize we need a lot of help. With many new surfaces and connected/unconnected content, we realized that the classifiers with employ to remove or demote harmful content on our platform varies greatly by problem and surface. So how do we scale the classifier evaluation across surfaces and problems? How do we navigate issues related to biases in our ground-truth labels used to train the classifiers - is it possible to compute unbiased precision and recall on the basis of biased samples? And especially with virality, one would argue precision and recall should be weighted by VPVs - and to do this right, we need counter-factual VPVs. A big shoutout to VESPA for advancing VPV predictions methods.

We have three main types of auto-enforcement: demotions, filtering from explore, and deletions. Each has their own challenges for computing counterfactual VPVs.

Two areas where we are working with VEPSA:

-Improving VPV predictions for content which is deleted immediately (and hence we have no data on realized VPVs)

-Improving predictions for demotions (so we have a VPV curve with a discontinuity where demotion was applied and want to forecast the VPVs without demotion)

- **Computing unbiased precision/recall estimates for auto-enforcement classifiers**
  - **Most of our ground-truth labels have huge sampling bias (prevalence, outcomes of reactive reports)**
  - **Can we build a scalable automated system to compute unbiased P/R using this biased data?**
- **Getting user-facing precision/recall using VPV weighting**
  - **Not all classifier mistakes are equal! Want to add VPV weighting to P/R computation**
  - **Ideally weights should be "number of VPVs prevented if classifier were to act on this content"**
  - **Working with VEPSA team to estimate counterfactual VPVs for these weights**

Theme 2:
Users Are Increasingly Exposed to Mixed Surfaces: Connected + Unconnected Content

Introducing recommendations in IG Feed

Early tests show significant increases in consumption and time spent.

Test results:
https://fb.workplace.com/groups/243894009320301/permalink/1150637191979307/

Legos means more unconnected content in feed
https://fb.workplace.com/groups/fyi/permalink/3450206931703109/

Are we making problematic usage worse on Instagram?

"The move feels like a major reversal from Instagram's stance two years ago on screen time. When it launched the end of feed notification, the company's former CEO and co-founder, Kevin Systrom, positioned it as a decision made with users' mental health in mind. He said at the time that the company was working on "tools that will help the Instagram community know more about the time they spent on Instagram," adding that "any time should be positive and intentional." But suggested posts could easily keep people on the app for longer." (The Verge, 2020)

- Early survey results suggest that Instagrammers who use the app for to keep up with friends and family feel less in control of their feed and believe suggested posts make IG worse (Still, 2020)

How will this shift affect Instagrammers' well-being?

- How will the increase in interest recommendation feed inventory affect problematic use?
  - Teens report that it's easy to get distracted and mindlessly scroll in the app without getting a lot of meaning/utility (Jimenez, 2020; Weber, 2020; Kang 2018; Kang & Mastin, 2018)
- How will the increase in unconnected content affect feelings of connectedness with friends, family, community?
- How will the increase in unconnected content affect negative social comparison?

Alternate definition: Some dimensions of IG experience can prompt consuming similar type of content for a period of time, making teens who already feel bad, feel worse

Why based on recent teen girls body image and social comparison and social comparison relevant slides from this work

Huge shoutout to CDS for helping complete critical studies on social comparison:

- **Country differences in social comparison**. Our paper on understanding country differences in social comparison on Facebook was accepted to CSCW. By combining survey,

interview, and behavioral data, we describe how the associations between social comparison and demographics (e.g., age and gender) and on-platform behaviors differ by country, and we discuss possible reasons as to why. Notably, Likes are more strongly associated with comparison in the US and India, suggesting that interventions such as Project Daisy may be more beneficial in certain countries.)

- **How does seeing content from celebrities on Instagram affect social comparison?** In our continuing series on social comparison on Instagram, we examine how seeing posts from top accounts on Instagram is associated with feeling more or less social comparison. Followers of some celebrities (e.g., Will Smith) tend to feel better (e.g., more inspired), while followers of other celebrities (e.g., Ariana Grande) tend to feel worse about themselves in comparison to others.
- **Selfies, filters, and social comparison on Instagram**. In this note, we find that selfie-taking can make people feel worse about themselves. Sharing more selfies was associated with feeling more negative social comparison. In Stories, seeing or sharing more selfies that had filters or effects applied made people feel even worse. Still, camera effects may vary in their impact on social comparison. Some effects (e.g., heart eyes) increase negative comparison, while other effects increase positive comparison (e.g., sunglasses). Filters that turn one's focus to a novel addition (e.g., the sunglasses) may be more inspirational and less harmful.

Theme 3:
Increasing Tension Between Safety and Free Speech

ROLE OF HASHTAGS IN INSTAGRAM

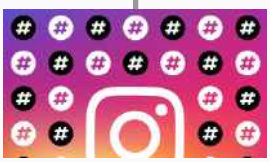
Hashtags are an intentional tool that enables people to go deep on their interests.

Consumers

- Explore their interests
- Follow updates via Home - Feed
- Curate their interests
- Entertainment (with reels!)
- And much more...

Creators

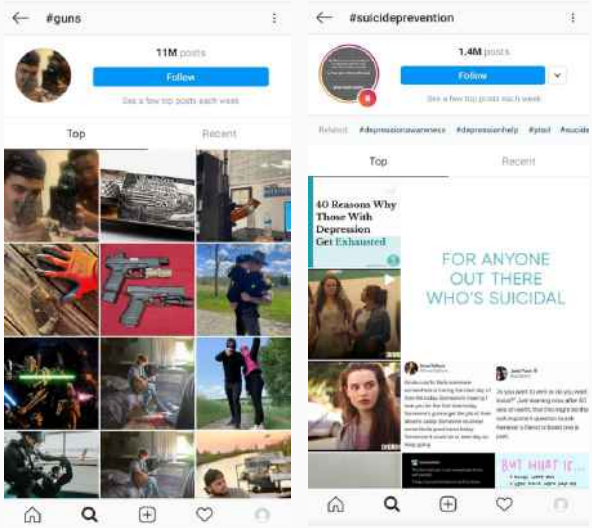
- Build audiences
- Verify if they are shadow-banned
- Curate their posts
- Engage in challenges
- And much more...



24

Hashtags are an important tool on instagram to bring you closer to people and things you love. Users rely on hashtags to explore or curate their interest, follow updates, get entertained and much more. Creators use hashtags to build their audiences, check if their content actually shows up (not shadow banned), curate their posts, and much more.

Integrity policy: Non-rec or Community Standards?



Do we stop allowing #guns, or SSI Remediation communities to thrive on Instagram?

The general guideline for recommendation surfaces is to **FILTER** non-rec content (like guns).

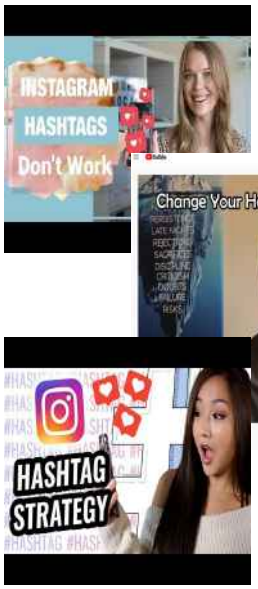
Team's opinion: No

Hashtags are built by the community and the community has a voice. If we start filtering media which is non-rec, we are suppressing voice. We should NOT filter non-recommendable content from Hashtag pages.

25

However, we are finding that hashtages are increasingly a complex entity that combined connected and unconnected content, and creating a lot of confusion about what role Instagram should play in protecting the safety of users while also giving them free expression. Are hashtags a recommendation surface? If yes, then we should be removing non-rec content, for example, guns or SSI. But hashtags are built by the community and the community expects to see them. Obviously, we would like to keep our community safe, so we should suppress content that violates our community guidelines. But often, it is not clear whether the hashtag itself is violating, since content might be mixed; or hashtag falls in the gray territory of non-recommendable, but not necessarily violating.

Treating Hashtags as a Rec surface removes transparency levers for users



Creator Transparency Lever

Hashtags are created and aggregated by the community. Anyone can create a hashtag (very popular when creating wedding hashtags or topical hashtags)

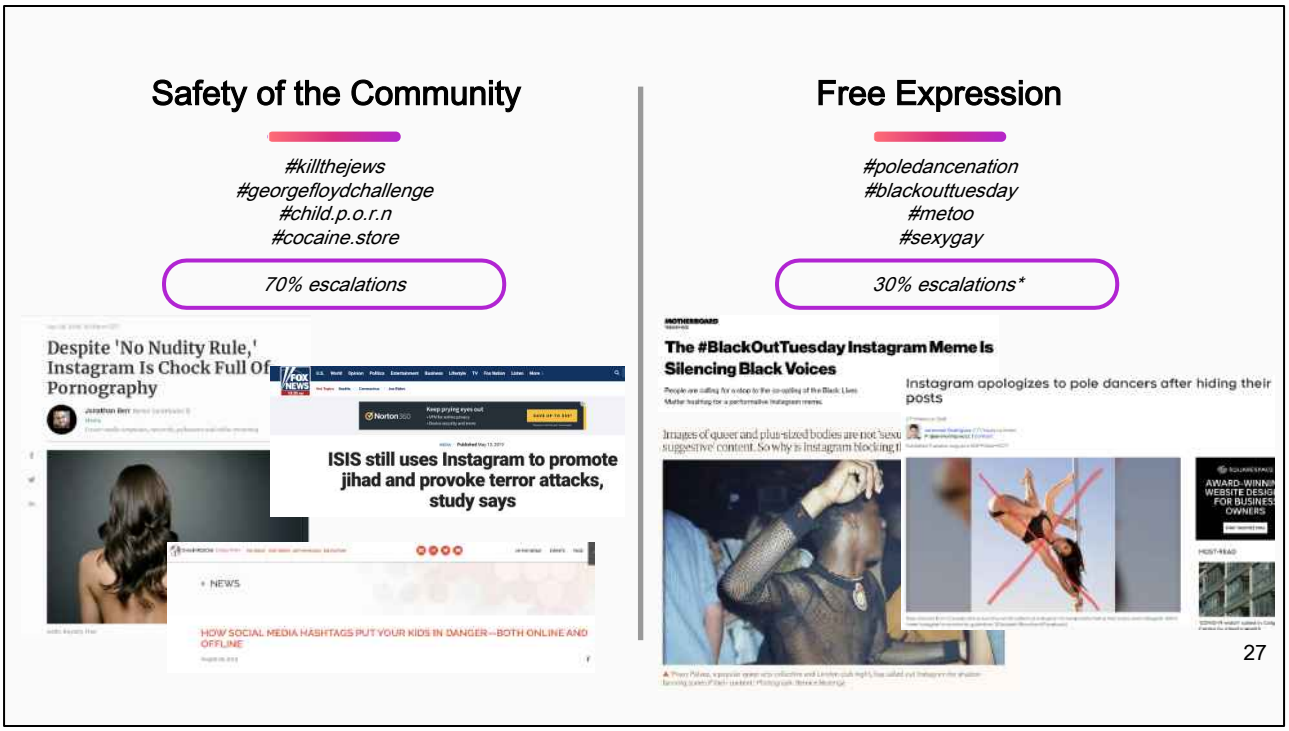
Users believe that when they tag a media using a hashtag, they SHOULD see that piece of media on the hashtag page. If they don't, the inference is that IG has shadow-banned that account/media/hashtag. i.e IG has filtered that media using a low-confidence classifier.

By making hashtags a rec surface, we will remove this transparency lever.

Voice suppression like Shadow-banning & Censorship find their origin in hashtag pages where our integrity treatments confuses users

26

Hashtags are very common for personal occasions - e.g., creating wedding or baby hashtags to curate and share personal content. But hashtags are also a key transparency layer. If you tag media by a certain hashtag, you expect that media to show up if a hashtag is called. If it doesn't show up, users perceive the content to be shadow banned, leading to escalations on censorship, and reputational effects.



Hashtags are a source of many escalations, since the policy on hashtags is not well understood, and frankly is confusing even to us.
Unfortunately, there are forces that use Hashtags in a harmful way and we do not have a consistent, well-socialized way to approach Integrity challenges on Hashtags. Being a **complex object**, hashtags has **many touchpoints** that across different surfaces and products and IG that each have their own nuances. Hence, it is all the more important to have a consistent, well-socialized policy/protocol/enforcement strategy for each of these touch-points to **holistically tackle** integrity challenges on Hashtags.

Freedom of expression challenges:

https://www.theguardian.com/commentisfree/2019/nov/08/instagram-shadow-bans-marginalised-communities-queer-plus-sized-bodies-sexually-suggestive
https://www.usatoday.com/story/opinion/2019/06/08/how-instagram-censors-could-affect-lives-everyday-women/1329016001/
https://www.nytimes.com/2019/11/22/arts/design/instagram-free-the-nipple.html
https://www.theartnewspaper.com/blog/social-media-crack-down-how-instagram-protects-us-from-art

Integrity challenges: