

Legitimacy Note 1: Developing LEGIT Metric I

Legitimacy measurement part 1: Understanding users and stakeholders' perceptions of our integrity efforts

· THURSDAY, SEPTEMBER 3, 2020 · READING TIME: 10 MINUTES 

The mission of the CI Legitimacy Team team is to help shape the overarching strategy for increasing legitimacy of our integrity efforts and develop the measurement framework for tracking legitimacy with people and stakeholders.

This series of three notes will present the legitimacy measurement strategy. I Chats

Legitimacy measurement part 1: Understanding users and stakeholders' perceptions of our integrity efforts

· THURSDAY, SEPTEMBER 3, 2020 · READING TIME: 10 MINUTES 

The mission of the [CI Legitimacy Team](#) team is to help shape the overarching strategy for increasing legitimacy of our integrity efforts and develop the measurement framework for tracking legitimacy with people and stakeholders.

This series of three notes will present the legitimacy measurement strategy. In order to measure perceived legitimacy we need to first understand how people currently view our integrity efforts. This first note presents research that investigated (1) users' perceptions of our integrity efforts and (2) external stakeholders' perceptions of our integrity efforts.

The second note will present proposed principles for building legitimacy with people based on external research, internal research, and product strategy. The third note will present the development of the LEGIT survey to measure progress with users and the general population as well as how we are tracking public sentiment with stakeholders.

TL;DR

1. Users don't trust us to do the right thing because they believe we prioritize revenue and growth over safety and society
2. Users don't perceive our content regulation system as legitimate because

[Chats](#)

TL;DR

1. Users don't trust us to do the right thing because they believe we prioritize revenue and growth over safety and society
2. Users don't perceive our content regulation system as legitimate because they don't trust our motivations, perceive our system to be ineffective and biased toward minority groups, and believe that we don't understand or listen to our users
3. There are opportunities to increase legitimacy with users by building trust in the brand through product experiences, ensuring we aren't biased toward certain groups, and empowering users to control their own experience
4. In addition to disagreeing and being concerned about our approach to integrity, many external stakeholders either aren't aware of our efforts or don't understand parts of our systems
5. There are opportunities to increase legitimacy with stakeholders by developing scaled education programs, telling a more cohesive and understandable narrative that is easily discoverable and shareable, and building integrity products that are defensible

1. Users' perceptions of our integrity efforts

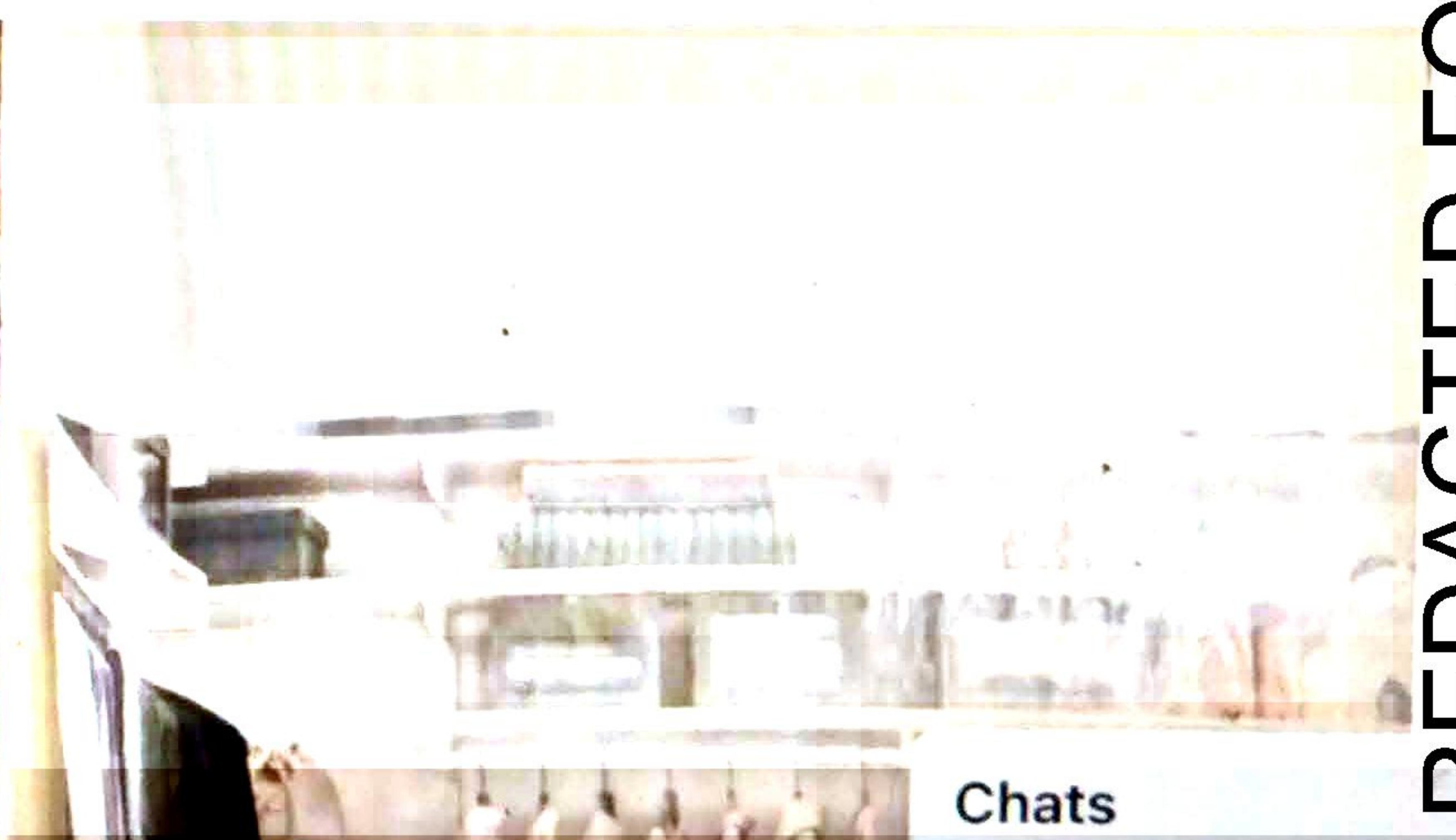
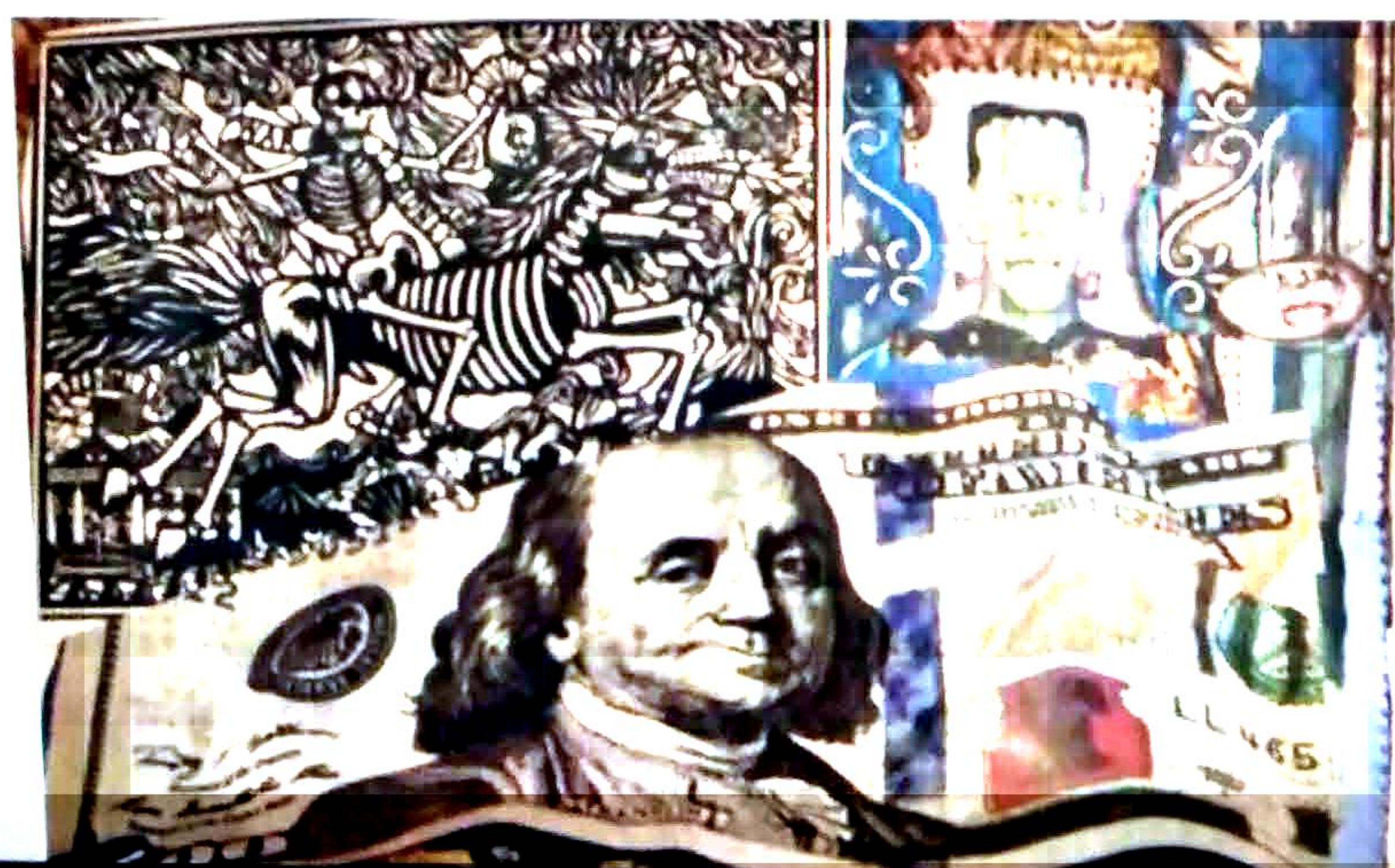
Three, three-hour research sessions were conducted in Seattle with groups of actors ($n = 6$), reporters ($n = 9$) and general users ($n = 11$). These sessions investigated what users are currently experiencing, perceiving, or hearing about that form their perceptions of our integrity efforts. Participants were diverse with regard to gender, race, and age.

Chats

Three, three-hour research sessions were conducted in Seattle with groups of actors (n = 6), reporters (n = 9) and general users (n = 11). These sessions investigated what users are currently experiencing, perceiving, or hearing about that form their perceptions of our integrity efforts. Participants were diverse with regard to gender, race, and age.

Perceptions of Facebook's integrity efforts tend to be negative

Dear Facebook, You've become a breeding ground for hate, disinformation, and scams. This is dangerous and you're turning a blind eye to it— and most likely profiting at the same time



REDACTED FOR CONGRESS



Users visualized the current state of content regulation as dark, messy, and money-driven. Their ideal state was depicted as bright, positive, representative of all people, and organized. All participants indicated FB has become more negative, hostile, and violent.

Chats

Users visualized the current state of content regulation as dark, messy, and money-driven. Their ideal state was depicted as bright, positive, representative of all people, and organized. All participants indicated FB has become more negative, hostile, and violent since they joined the platform. This was a signal to them that our content regulation system is not working or effective.

The following four themes emerged explaining why users' have a negative perception of our integrity efforts.

1 | Users don't trust us to do the right thing because they believe we prioritize revenue and growth over safety and a good user experience

*Dear Facebook, I can't take it any more.
You keep feeding me and billions of others
a whole bunch of lies and crap that is
designed to influence people and spread
conspiracies. You get paid very well
to do this and because of that have
not taken the necessary measures
to prevent it. You welcome it.*

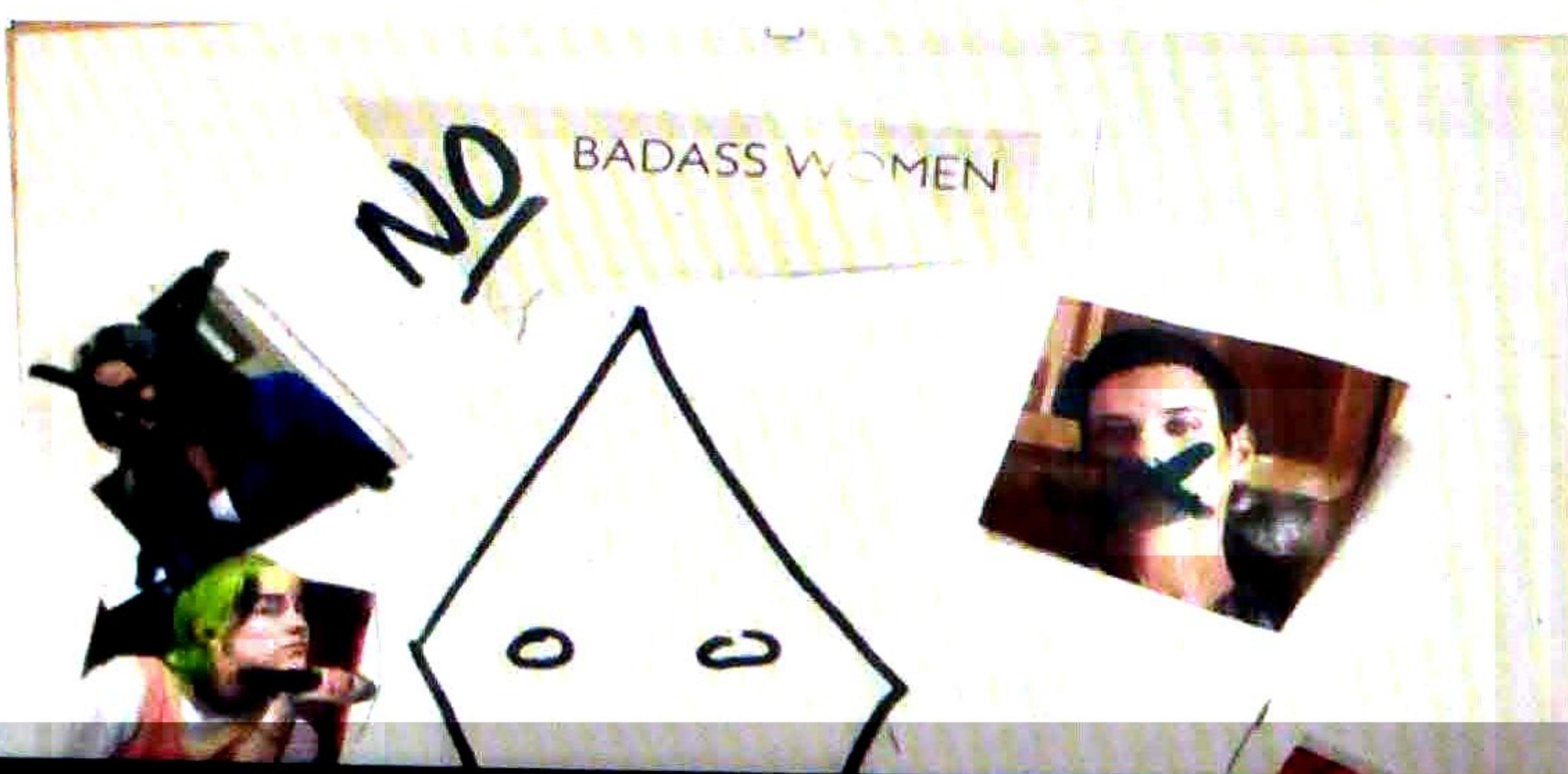
Chats

Because users don't trust FB due to past incidents, they don't believe we have good intentions or motivations when it comes to integrity efforts. Many users don't think we are actually trying to remove bad content and accounts because we care more about increasing engagement and growth than protecting people and society. ***This belief harms the legitimacy of our integrity efforts immensely.***

Recommendations:

- Continue to invest in restoring trust in the FB brand. Efforts to increase legitimacy of content regulation may need to include efforts to also increase trust in FB as a whole.
- Build trust by ensuring what we ship shows care and is defensible to regulators, auditors, and our users
- Put greater prioritization on reducing harm and bad experiences
- Invest in ways to actively demonstrate to users and stakeholders that we are committed to reducing harm both in-product (e.g., make it easier for users to report) and externally (e.g., highlight the good work we do to prevent CEI across tech)

2 | Users believe we are biased toward and silencing minority voices due to our moderation practices



“Between your clearly racist and misogynistic ‘moderation’ to your laughable appeals there is a reason we Chats

2 | Users believe we are biased toward and silencing minority voices due to our moderation practices



“Between your clearly racist and misogynistic ‘moderation’ to your laughable appeals process, there is a reason you are called RACEBOOK. **Clearly I am not included in the community you claim to protect.**”

Users who self-described as being LGBTQ or Black believed that FB is censoring and over-enforcing on minority groups. Participants described being banned for speaking out to their communities about their lived experiences and condemning hate.

POC and marginalized groups are barely able to stay afloat in an environment that doesn't understand the minority experience.

They attributed this to our policies and systems being designed by White upper/middle class people who don't understand minority experiences. Many participants acknowledged much of this enforcement is done by automation and algorithms but believe that the people who have built the algorithms are at best naive and at worst racist.

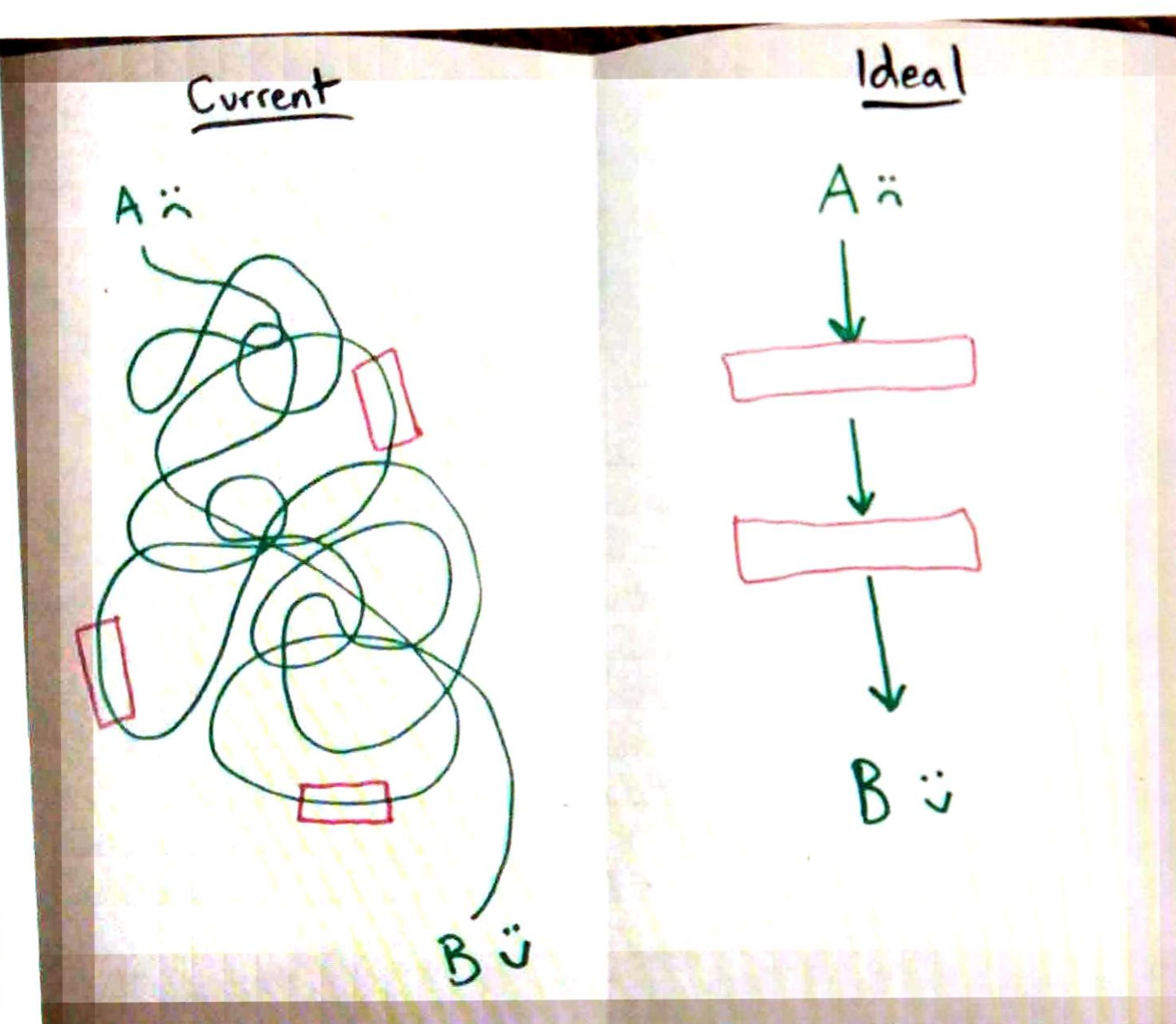
Recommendations:

- Conduct audits of actions taken on content posted by people of color, to verify if we are disproportionately actioning against one demographic's content
- Increase the representation of people from minority groups in our building process
- Ensure that any system we ship is not biased against certain groups (e.g., SAIL efforts).

3 | Users want more control over their experience and a way to help improve the system



3 | Users want more control over their experience and a way to help improve the system



“FB should empower us to choose instead of labeling what kind of audience they think we are.”

Our global user base is never going to agree with all of our policies and enforcement decisions. Instead, we should empower users to improve their own experience by giving them more control over what they see (and don't see). Users agree we should take a stance and remove what is universally thought of as bad or illegal (e.g., CEI, terrorism, violence, harassment) but leave it up to them to decide the rest. This would take some of the strain off of creating a content regulation system that is agreed upon by everyone.

Users feel like FB doesn't listen to its community because they have no way to give us feedback and express their point of view. For example, appeals are generally **Chats**

Our global user base is never going to agree with all of our policies and enforcement decisions. Instead, we should empower users to improve their own experience by giving them more control over what they see (and don't see). Users agree we should take a stance and remove what is universally thought of as bad or illegal (e.g., CEI, terrorism, violence, harassment) but leave it up to them to decide the rest. This would take some of the strain off of creating a content regulation system that is agreed upon by everyone.

Users feel like FB doesn't listen to its community because they have no way to give us feedback and express their point of view. For example, appeals are generally thought of as a way to express disagreement with a decision by an institution. However, users don't view our appeal system as legitimate because they perceive it as just clicking a button that goes nowhere.

Recommendations:

- Give users more control over what they see (and don't see). For example, users wanted one newsfeed tab for "lighter" organic content from friends and family and another tab for shared news articles and non-organic content. One user didn't want to see any borderline graphic violence close to when she was going to bed. Other internal research also provides evidence for user controls such as content covers. Other research indicates users want the ability to dial up misinfo control (e.g., choose to have misinfo removed from their feed) but not dial it down (i.e., remove misinfo labels).
- Investigate ways we can incorporate user feedback at scale such as Oversight Board case advocacy and community feedback
- Understand how rank and respond might further threaten the perceived [REDACTED]
~~our appeals system and investigate how to increase trust and acceptance~~ Chats

users want the ability to dial up misinfo control (e.g., choose to have misinfo removed from their feed) but not dial it down (i.e., remove misinfo labels).

- Investigate ways we can incorporate user feedback at scale such as Oversight Board case advocacy and community feedback
- Understand how rank and respond might further threaten the perceived legitimacy of our appeals system and investigate how to increase trust and acceptance of automation

4 | Users aren't buying that content regulation is a hard and complex problem

You created this platform and wanted this growth. You are responsible for moderating it, no matter what time, tools, and expertise it takes. FB has the expertise and money why can't they do it?

Though some participants acknowledged content moderation is difficult at scale, many viewed moderation as a cost of doing business. Because FB is a multi-billion dollar company, many users believe we *could* solve the problems with content moderation we just *choose* not to (see theme one).

Though some participants acknowledged content moderation is difficult at scale, many viewed moderation as a cost of doing business. Because FB is a multi-billion dollar company, many users believe we *could* solve the problems with content moderation we just *choose* not to (see theme one).

Recommendations:

- The narrative that content regulation is difficult and complex might not land well with users. Instead we should understand if focusing on highlighting what we *are* doing to address problems would be more effective. For example, users tend to feel positively toward content warning screens and misinfo labels because it concretely *shows* the user FB is doing something.

We will continue to understand where we are falling short with people and how problem and foundations teams can ensure they are building legitimacy into their products and processes. Additionally, the IX team (formally actor and reporter experiences) has a long history of and continues to understand and build experiences that increase supportiveness and procedural fairness among actors and reporters.

2. Stakeholders' perceptions of our integrity efforts

Investigating external stakeholders' (i.e., policy, civil society, academics, and the media) perceptions of our integrity efforts is more challenging as it is difficult to reach these audiences directly. In H2 2019 we took a pulse on where these audiences generally stand on key integrity topics by aggregating insights from our internal policy and comi
the full table here. Chats

2. Stakeholders' perceptions of our integrity efforts

Investigating external stakeholders' (i.e., policy, civil society, academics, and the media) perceptions of our integrity efforts is more challenging as it is difficult to reach these audiences directly. In H2 2019 we took a pulse on where these audiences generally stand on key integrity topics by aggregating insights from our internal policy and comms partners. See the full table [here](#).

Stakeholders' top areas of concerns include offline harms, how/if we incorporate feedback, trust in the company, soft actions, procedural justice, US-centric policies, and misinfo. In addition to disagreeing and being concerned about our approach to integrity, many stakeholders either aren't aware of our efforts (e.g., that we publicly release our enforcement rates in the CSER; that we work with law enforcement) or don't understand parts of our systems (e.g., how we use machine learning to quickly remove content; how our policies are formed).

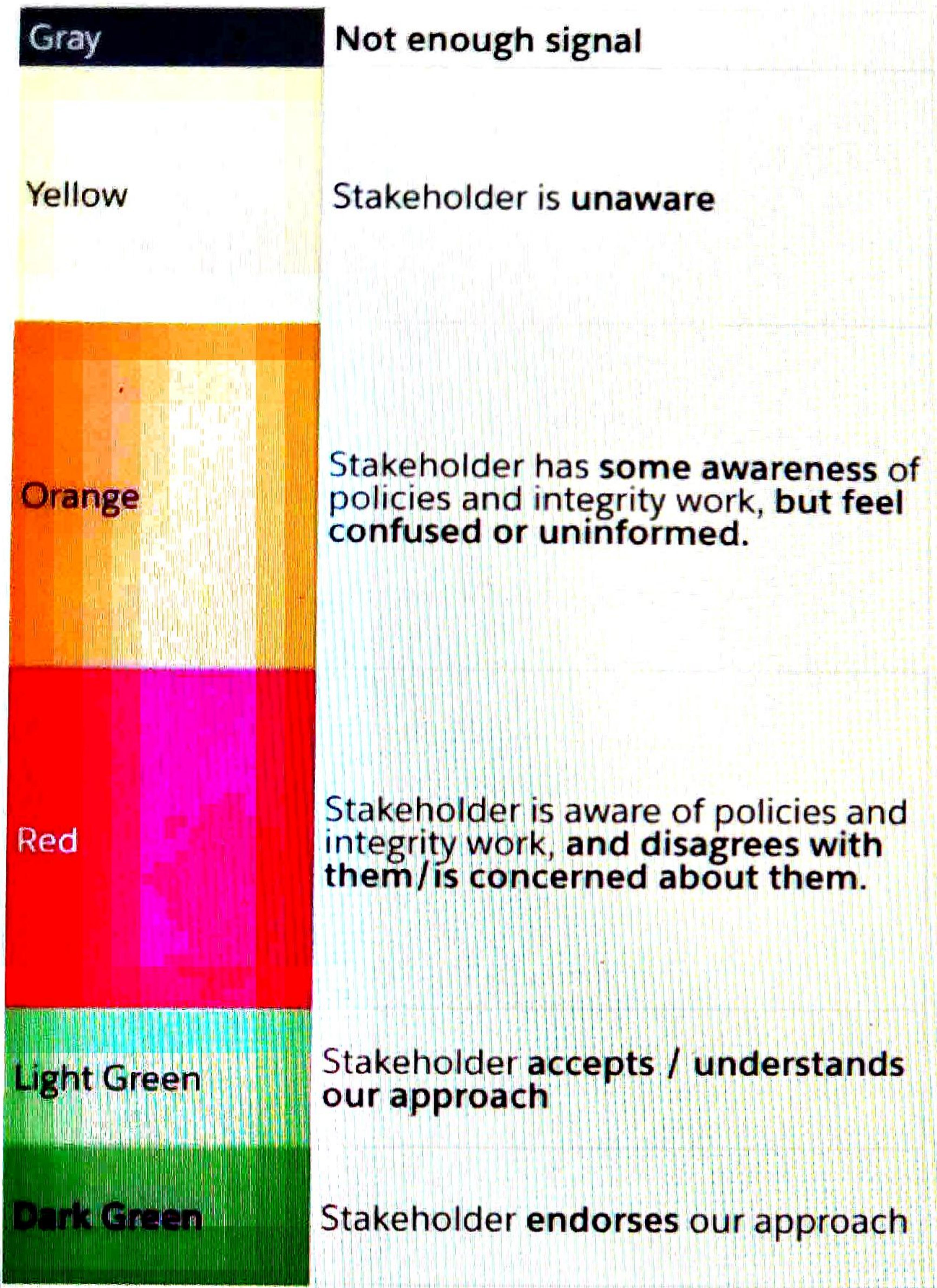
These findings highlight the need for us to develop a scaled program to educate stakeholders about our content moderation practices to help support the excellent work by our stakeholder engagement, policy, and comms partners.

The Transparency Hub product vision is an initial way we could build legitimacy with stakeholders by increasing awareness, decreasing confusion, and presenting a transparent holistic narrative about our integrity efforts.

Gray

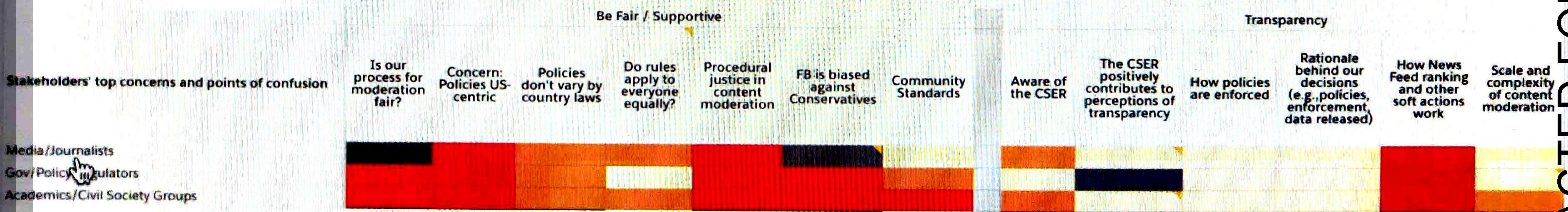
Not enough signal

Chats



Opportunities to Increase Legitimacy with Stakeholders:

- Increase *awareness* of the good work we do and how we do it
- Decrease *confusion* resulting in inaccurate conclusions and folk theories
- Reduce disagreement/concern and build advocates by including stakeholders in the development of our products and policies

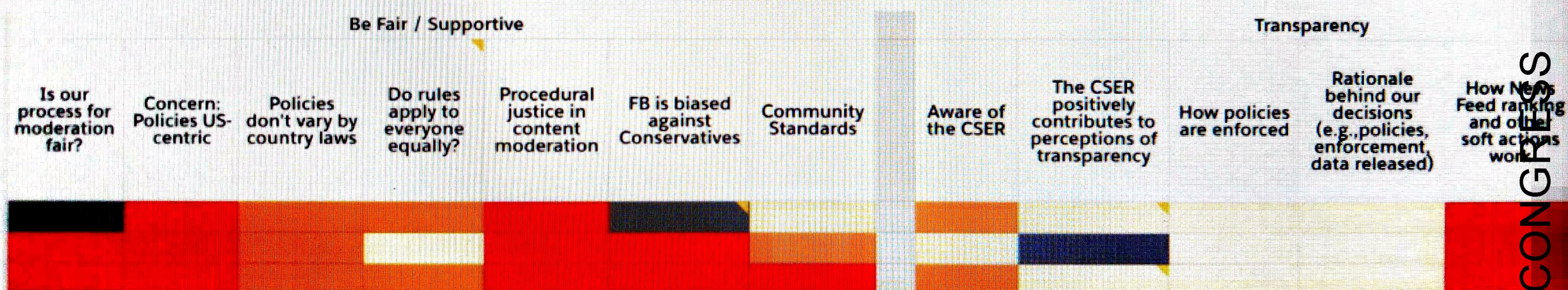


Stakeholder is aware of policies and integrity work, **and disagrees with them/is concerned about them.**

Stakeholder accepts / understands our approach

Stakeholder endorses our approach

- Decrease *confusion* resulting in inaccurate conclusions and folk theories
- Reduce disagreement/concern and build advocates by including stakeholders in the development of our products and policies



ing to conduct additional research with our policy and comms partners as well as others to further understand, synthesize, and prioritize needs and areas of concern for stakeholder groups. We also need to understand what types of content will be the most

We are planning to conduct additional research with our policy and comms partners as well as stakeholders to further understand, synthesize, and prioritize needs and areas of concern across stakeholder groups. We also need to understand what types of content will be the most useful, how to contextualize and organize this information, and how to encourage more informed stakeholders to share information with others.

The next note in this series will present proposed principles for building legitimacy with people based on external research, internal research, and product strategy. The third note will present the development of the LEGIT survey to measure progress with users and the general population as well as how we are tracking public sentiment with stakeholders.

Huge thanks to [REDACTED] for helping design and run the Seattle research and to [REDACTED] and [REDACTED] for thoughtful feedback on this note and being my copy editors!



59

29 Comments 15 Shares

 Like

 Comment

 Share

 Save

Like · Reply · 1y



[REDACTED] one word: powerful



Like · Reply · 1y

❸ *Cover photo is a participant's depiction of the current state of content regulation on FB. According to the participant, the dinosaur represents how angry users are and the person at the computer is FB's moderators completely overwhelmed.



Like · Reply · 1y

[REDACTED] thanks for the write up, [REDACTED]! for the stakeholder understanding, and the learning about lack of awareness. with transparency hub vision, are we starting to look at how to distribute the info/trainings on more of a push model, finding forums that best work for them? Consolidating all info in the central transparency portal is a great first step, but we may need something more proactive to take it further



Like · Reply · 1y

[REDACTED] ❸ Totally agree, [REDACTED]! Having a centralized, understandable, and transparent (but passive) artifact is necessary but not sufficient. Currently, there is no "one stop shop" for any interested parties to understand our approach to content regulation (e.g., CSs live on one site, advertiser rules on another, and Page/publisher rules on another; important information about our efforts living in blog posts). If stakeholders even discover these pieces we force them to put together their own narrative.

We definitely plan to work with our internal policy and comms friends to understand the best push model and how to maximize and scale impact. We also hope to build in different engagement models (e.g., maybe draft policy commentary opportunities, maybe a part of the portal that's only accessible to trusted stakeholders who can refer additional stakeholders to us) to bring more people into the process.

Like · Reply · 1y

[REDACTED] Incredibly helpful write up! Regarding the table of stakeholders, can you point me to any resource that lists [REDACTED] the academic/ civil society groups we are referencing + how we prioritized them?

Like · Reply · 1y

[REDACTED] --We are hoping to build out a more robust process to regularly gather and aggregate insights at a more precise stakeholder level via employees who own these relationships. Aggregating the qualitative feedback employees across the company are gathering from stakeholders could be super valuable to understand where stakeholders stand and provide direction and strategy for integrity teams.

For this quick pulse we relied on insights from our policy and comms partners who summarized the sentiment, top concerns, and points of confusion from the stakeholders their teams engage with. This included people from civil society groups, policy folks, journalists/media, and academics but can't distill out exactly wh

Chats

Like · Reply · 1y

[REDACTED] Incredibly helpful write up! Regarding the table of stakeholders, can you point me to any resource that lists out the academic/ civil society groups we are referencing + how we prioritized them?  1

Like · Reply · 1y

[REDACTED] --We are hoping to build out a more robust process to regularly gather and aggregate insights at a more precise stakeholder level via employees who own these relationships. Aggregating the qualitative feedback employees across the company are gathering from stakeholders could be super valuable to understand where stakeholders stand and provide direction and strategy for integrity teams.

For this quick pulse we relied on insights from our policy and comms partners who summarized the sentiment, top concerns, and points of confusion from the stakeholders their teams engage with. This included people from civil society groups, policy folks, journalists/media, and academics but can't distill out exactly which groups they are referencing.  2

Like · Reply · 1y · Edited

[REDACTED] - thanks so much! Long term, it would be really helpful to understand 1) the types of Civil Society Groups and academics that are the most vocal for each subtopic (ie: protecting ppl when we move to encryption) + 2) where we as a company are most interested in moving the needle on perception. This info would help us prioritize how we build solutions across problem types.

IE: if we knew that sex workers rights advocates and people of color are two key stakeholders that are most concerned about fairness in community standards + we know want to shift their perception on this, that would help us more effectively prioritize projects/programming around human trafficking moving forward.

Happy to chat more about this if you have q's!  2

Like · Reply · 1y

[REDACTED]  Definitely agree being able to understand these themes and target our work more effectively is super important! [REDACTED] whose teams manage these stakeholder relationships would have the most insight into your points (1) and (2)!  1

Like · Reply · 1y

[REDACTED] Write a reply... 

Like · Reply · 1y

[REDACTED] for R&R + Overall Support research  1

Like · Reply · 1y

[REDACTED] Thank you [REDACTED] for this note! Really helpful for our work on the Oversight Board  1

[REDACTED] Convenient timing from PEW: <https://www.pewresearch.org/.../few-americans-are..../>

Chats

Write a reply...

[REDACTED] for R&R + Overall Support research



1

Like · Reply · 1y

[REDACTED] Thank you [REDACTED] for this note! Really helpful for our work on the Oversight Board



1

Like · Reply · 1y

[REDACTED] Convenient timing from PEW: [https://www.pewresearch.org/.../few-americans-are.../...](https://www.pewresearch.org/.../few-americans-are.../)



PEWRESEARCH.ORG

Few Americans are confident in tech companies to prevent misuse of their platforms in the 2020 election



Like · Reply · 1y

Like · Reply · 1y

[REDACTED] > Our global user base is never going to agree with all of our policies and enforcement decisions. Instead, we should empower users to improve their own experience by giving them more control over what they see (and don't see). Users agree we should take a stance and remove what is universally thought of as bad or illegal (e.g., CEI, terrorism, violence, harassment) but leave it up to them to decide the rest. This would take some of the strain off of creating a content regulation system that is agreed upon by everyone.

this is quite reasonable. do you know if there is any appetite within the company for an approach like this?

Like · Reply · 1y · Edited

[REDACTED] I think there's awareness of it (IX team has research on user controls) but I'm not sure if anything is being built at the moment. There's also warning screens to cover borderline graphic violence and misinfo labels.

Like · Reply · 1y

[REDACTED] ok, so there's awareness. what about making it so that we take a stance and remove only what is universally thought of as bad or illegal, while leaving the rest of moderation up to the users to decide?

Like · Reply · 1y

[REDACTED] FYI measurement focused on integrity/content



3

Like · Reply · 1y

[REDACTED] US survey data from Pew weighted to be representative of the population strong

Chats

Like · Reply · 1y

[REDACTED] ok, so there's awareness. what about making it so that we take a stance and remove only what is universally thought of as bad or illegal, while leaving the rest of moderation up to the users to decide?

Like · Reply · 1y

[REDACTED] FYI measurement focused on integrity/content

1 3

Like · Reply · 1y

[REDACTED] US survey data from Pew weighted to be representative of the population strongly reinforce some of the qualitative research findings in the note

- 74% express little or no confidence in tech companies like Facebook, Twitter and Google to prevent the misuse of their platforms to influence the 2020 presidential election
- 78% say these companies have a responsibility to prevent misuse

See More



PEWRESEARCH.ORG

Few Americans are confident in tech companies to prevent misuse of their platforms in the 2020 election

1 2

Like · Reply · 1y · Formatted

[REDACTED] this is such interesting research, thank you 😊 It would be amazing to have a redux (summary or presentation) of this in one of our CI all hands or perhaps a roadshow of it - selfishly - for integrity teams. While many of us intuit the difficult relationship between access to data (to enable integrity) and legitimacy (the perception this is ok, ok for us), I don't think many of us can delineate it in the context of the existing mistrust.

1 2

Like · Reply · 1y

[REDACTED] heads up-- you tagged the wrong Emily 😊

1

Like · Reply · 1y

[REDACTED] Like · Reply · 1y

[REDACTED] happy to share more! Let me know how I can best plug in.

Like · Reply · 1y

Like · Reply · 1y

[REDACTED] happy to share more! Let me know how I can best plug in.

Like · Reply · 1y

Write a reply...



Like · Reply · 1y

[REDACTED] Is there a definition of legitimacy somewhere?



2

Like · Reply · 1y

[REDACTED] note 2 (coming soon!) will cover this more in-depth but TL;DR:

- 1) From a policing/government perspective, legitimacy is the belief that an entity should be able to make and enforce rules and that people should follow these rules
- 2) From an organizational perspective, legitimacy is the belief that an organization's actions are appropriate and desirable within society's norms and values
 - We of course want people to believe (1) and want people to follow our Community Standards but we know from survey data and qualitative research insights most users already agree "yeah, you're a private company you can make your own rules/do what you want" so measuring (1) isn't super informative or valuable.
 - However, we know from research people don't tend to agree with our rules, how we enforce them, how we treat users, that what we do is effective etc. This is in line with definition (2) and thus, measuring these perceptions is more impactful AND we can actually do something about these things.

Like · Reply · 1y · Formatted



1

Like · Reply · 1y

[REDACTED] such powerful insights on how users perceive the legitimacy of our integrity systems



1

Like · Reply · 1y

[REDACTED] this series has great inspiration for a campaign educating people on feed



2

Like · Reply · 1y

[REDACTED]