Tennessee v. Meta
Case No. 23-1364-IV
PX-00895

Random meeting notes

10/16 - chat with Arturo about prioritizing work on user perceived problems

How do we convince people that reducing negative experiences (even if not policy violating) is worth trading off a bit with growth and/or free speech concerns? Make the case w data -- this is the size of the problem, this is how bad it makes people feel, this is how many people leave because of it

Scoping the problem

people have a bad experience (measured by TRIPS)

--> they feel agency to do something about it (measured by negative actions)

--> resolution (we take down the content, SR action solves their problem)

What can we do to decrease the gaps between these?

- Between 1 and 2
  - understand why people don't use available tools -- e.g. no tool fits their motivations well, they think we won't take action anyway, diffusion of responsibility, not really bothered that much, don't know where tools are
  - potentially... create new tools, surface tools better, help users feel they can make a difference
- Between 2 and 3
  - improve reporting flow so people complete the flow, e.g. better language
  - improve reporting flow so people get a resolution
    - branching to triage the problem to appropriate actions (e.g. sub-problem selection that includes non-violating categories)
    - help people protect the community, even for non-violating problems
      - nudge authors to create less bad content
        - direct person-to-person feedback
        - indirect/aggregate signals to author like "X people said this made them feel bad"
      - use aggregate negative user signals in soft actions, and tell reporters that we're doing so
    - help people see less stuff they won't like, e.g. personalization

01/21 - Survey metrics for supportiveness, transparency, and fairness w/ bryan [redacted] and mingliang

- Past and future projects
  - Actor/reporter appeal
  - Reporting flow optimization
  - Account appeal: smoother user experience
  - Support inbox: centralized place for all reporting/appeal notifications and status
  - Reporting flow optimization: better education screens
- Current metrics
  - Volume of actioned reports, volume of actioned appeals (quantity)
  - Action rate of reports, action rate of appeals (quality)
  - Report outcome view rate (do people see the outcome) -- goal metric for FB team, this is the best predictor of feeling supported
- Bryan -- they came up with 3 survey tools that capture
  - Based on procedural justice. Regardless of whether people get the outcome they want, do they feel good about the experience
  - Supportiveness: breaks down into transparency, fairness, voice (they cover 60% of supportiveness)
  - These should be similar for IG, but consider differences in content type and who is reporting
  - 2 tools
    - post-support survey - asks about reporting experience

        - complete reporter experience survey (3 day delay, so they should have received outcome response by then) -- asks about both reporting experience and outcome
    - Difficulty is in understanding how to weigh impact of moving survey metrics -- e.g. if supportiveness goes up, how do we trade that off with other metrics?

08/15 -- WB research Chat with Arturo Bejar

- for the company, reporting flows are meant to identify content that violates policies, so the categories are based around policy.
    - if you have a reporting flow and the action rate is low, you have a bad reporting flow
    - reducing reporting is misguided, there's no such thing as a bad report. their experience/emotion is real.
    - reporting should reflect categories and language that reflect users' experience
        - example: "hate speech" doesn't mean anything in India -- landed on "makes fun of something that's important to me"
        - this raises engagement with the reporting flow, but also leads to lots of abandoned reports
        - "were your needs met?" -- people who dropped out of the flow often said yes, because they just wanted to express what happened to them. This should be the metric, not finishing the flow.
    - example: many nudity reports had no nudity, but did include themself in the photo.
        - had an open-ended thing you can send to the person who posted it, but low completion rate
        - added "it's embarrassing" as a prefilled message you can send -- completion rate skyrocketed
        - 90% of people who post content that's reported have good intentions, but it's hard to give them feedback (either public, or depends on other users taking initiative)
    - designed reporting flows that lead to different things -- e.g. high intensity / fear-based bullying leads to an immediate support number, low-key harassment leads to peer-to-peer resolution advice
        - what's the goal of the reporting flow? you'll think of different solutions if you think it's removing bad content from the platform vs. helping people be better members of the community they're part of
        - ask about emotion, provide different options based on it. this provides different solutions for different cohorts of people.
- to get leadership attention, ask the questions we're afraid to ask -- e.g. mental health questions, are you lonely, are you using FB instead of hanging out with your friends, are you really upset
    - integrity needs to be woven into the DNA of the company, the way that growth or engagement is
    - the way to do this is to have a metric that they're compelled to move
- social norms -- there's no other social space where you can say such flagrantly offensive things, and it's not because of policing -- it's regular people disapproving and providing that feedback
    - there will always be some bad actors, but 90% of people would behave differently if given feedback
    - most people don't like to look at divisive stuff (this is what people complain about re: fb), but they don't have any way of expressing that
    - how do you motivate this work? ask people what the worst stuff on fb is. what do you hate most about fb. measure reach of those things. show m-team that people hate those things, that they have high reach, but that they're allowed within policy.
        - don't fight about where the policy line is -- e.g. antivax. that's not close to this.
- what do you do with all the non-policy-violating stuff?
    - social feedback loops
        - reporting flow -- one of the options for teens was "it was annoying". if they pick that, suggest that they tell their friend that it's annoying (provide default language to use that they can edit)
        - did research on the language they should use (I statements, nonviolent communication, etc) -- range of politeness
        - humans want to help people in distress, as long as we don't feel threatened. people who receive these messages are highly receptive. it doesn't always lead to people taking down the content, but it starts a dialogue -- (why is this embarrassing, I didn't even notice, etc). but once this starts, people think twice when they post next time.
    - does this work for everyone? what about cultures that don't like to talk about feelings?
        - the data suggested it still worked in those countries (people filled out the flow at the same rate, people receiving the message were just as satisfied with it)

- why aren't we using these flows anymore?
    - the teams got broken up, different teams are responsible for different parts of them now, institutional knowledge gets lost
- Look up this Compassion research -- Jennifer Guadagno will know about it.

06/06 - Bryan [redacted] re: Developing sentiment metrics

- Team needed a measure that could help with ship/no-ship for quickly-iterated product experiments, and for quarterly goaling
- Sentiment measurement for product changes is probably a bad idea (post from Curtiss) - wide error bars, insensitive to incremental changes, leads to demoralization
- Sentiment should be related to your specific product experience (i.e. CAU is too broad), and is probably not just one specific measure (one number isn't granular enough to capture what you're looking for, and not specific enough to be able to move)
- Phrase 1: develop framework. A year full of qual research to ID themes about what matters to reporters, which aligned well with the model of procedural justice, so ran a quant test to validate (supportiveness = transparency + voice + fairness)
    - Quant test: asked questions about transparency, voice, fairness, and then several outcome measures including supportiveness (experience, resolution, CAU). The facets correlate well with all the outcome measures, but best with supportiveness. Supportiveness was also least correlated with actual problem resolution, which is good.
    - More quant and qual testing (links in deck) to understand what each facet actually means to people, and to pick questions
- Phase 2: testing framework. Understand if this framework can actually be used to measure sentiment for goaling and experimentation.
    - Good goals: quantify your mission, tied to something inherently valuable, easy to understand, easy to measure, hard to game, help you learn something, can you move it.
    - Easy to understand: cognitive testing. ID language issues, validate that a yes/no question could convey respondent intent, help pare down similar questions
    - Easy to measure: quant testing in 5 countries. Exploratory factor analysis and IRT on massive set of potential questions. It takes a long time.
    - Can you move it. Used v1 of survey tools to attempt to ID differences in sentiemnt on 9 product experiments that we hypothesized should improve sentiment.
- I had to leave halfway -- get notes from Hitomi!

03/22 - elena davis re: wb measurement on the fb side

- There's a spectrum of changeable vs matters in the big picture of someone's life
    - If we're talking about big-picture overall WB, big things (death in the family, etc) make an impact, but feature changes aren't going to register
- Big-picture overall WB is not going to be useful for product changes, but because we inherently care about it, it's definitely useful for:
    - Detecting natural variation, by demographic variables or usage variables (posting on FB is related to higher WB)
    - Longitudinal -- e.g. Posting on FB is related to higher WB, and increasing posting is related to increased WB -- lets you hold demographic & other variables constant since they're the same people. This is something that's never been done on instagram but would be useful.
- Ryan has done work showing that recent WB on FB is related to general WB, but it's actually more related to brand sentiment. They haven't looked at facets of WB, only overall. https://our.internmc.facebook.com/intern/anp/view/?id=611068
    - Their current assessment of this is that WB-on-FB could still be a useful measure, but digging into it more now.
- Elena is currently doing a proof of concept survey experiment measuring WB at 3 levels: overall, in the moment, and in the moment on facebook. All of these are overall WB, not facets. They're comparing people in the Memories holdout (on this day). Even for a major, obviously feel-good product, they assume effect sizes will be small enough that they're surveying literally 1m people.
- Another approach is to look for meaningful behaviors that are correlated with (or cause) WB. We could get hypotheses from FB knowledge (e.g. active use), academic literature, etc. Then we'd try to drive up those behaviors and assume it's leading to better WB, even if we can't detect that change.

- Elena is also advocating focusing on specific facets -- we know that they're related to overall WB from research, so we don't have to verify every time. We can focus on ones we have some expectation of changing as a way of narrowing in and gaining power. To make sure it's not affecting other areas, you could still include them (e.g. everyone gets the facet of interest + 1 random other one as a countermetric).
    - For facets, Facebook is concerned about: social comparison, loneliness, problematic use/control
- Compassion -- nobody is focusing on this specifically, it's all being wrapped up as predictors of WB (along with demographics, etc). But she thinks that sucks and that we should totally work on it.
    - Separate into things that happen on our platform (integrity problems, social comparison) vs. things that happen off-platform but that we might exacerbate
- Elena has a deck and is currently getting buy-in, but she's the right POC for all this work and I should check in with her in a few weeks

03/04 - wendy re: safety survey

- She's currently running a survey about people's perception of safety on IG. The goal is to have data that helps with a messaging campaign about IG taking safety seriously. They want to ID focus areas in each country (issues that people think are prevalent, that are IG's responsibility, but where IG isn't doing enough), and also pick messaging that works for each focus area.
    - Reputation - have you heard of this happening on IG?
    - Responsibility - do you think IG should do something about it?
    - Action - do you think IG is currently doing enough?
    - Messaging -- will probably end up something like "Instagram is working hard to keep me safe"
- Running this survey both in-app (mobile only) and email -- nobody has done this before! The goal is to gain a general understanding of which approach is better and/or how we can expect data to vary by surface
- Also starting a survey on pressure to be perfect (working with policy -- they ideally want data to back up a campaign about posting "real" content, e.g. X% of people feel better after seeing authentic content)
    - How often do you feel better/worse after comparing yourself to someone else on IG
    - How much content do you see that's meant to look perfect vs. real
    - How does perfect vs. real content make you feel
    - Breakdown by friend content vs influencer
- Setting up biweekly syncs

01/17 - Kyle re: Nielson study

- From his work on feed, one of the major pain points is people feeling like they're missing posts or that they won't see things on time. They've found that people who are anxious about this are the ones who tend to get upset about ranking changes, etc. The "you're all caught up" feature is a reaction to that.
- Can we add some questions measuring "feed anxiety" to the baseline questionnaire? Then slice questions about well-being/fomo/feeling caught up (especially anxiety-related) by feed anxiety.
    - He's going to send me his questions!

01/17 - Lauren & Frank re: TRIPS & IG integration

- They're wondering whether we're ready/eager to begin testing/rolling out their version of the CI tracker to Instagram. Basically, this would mean testing their questions / flow and comparing results to a) their results, and b) our current results. We assume that questions will have to be changed because of differences in the product (e.g. follows, explore), differences in how people use and talk about the product (e.g. following interests and influencer accounts), and different populations. So, we'd be writing question variants and comparing them to ground truth, cognitive testing translated versions, etc.
- From their side, Guy and others in leadership have asked to be able to compare FB to IG (though that conversation last happened 6 months ago). Is there the same appetite for comparison on IG? Is there appetite for an ongoing (daily) tracker?
    - They're going to check in with CI leads today and see if comparing to IG is still top of mind.
- It'll be easier this time, since I can start with what they've already learned. The challenge will be knowing what to do if we find inconsistent results with what they found.
- If there's no current time pressure from our side, that would actually make things better in some ways because we can work on it slowly without someone breathing down my neck. Lauren estimated 1 day a week. Ana would be working closely with me.
- An open question is how we would split the work, and whether IG or FB would own it.

01/14 - Devon re: perceived reach & intensity

- Definitions of different measurements
  - DAP exposed = % of DAP who had an impression of known violating content (i.e. content deleted in the 7 days following)
    - Content is deleted when caught by a classifier (can be automatically deleted for drugs & N&P, but otherwise a CO person judges), or when reported (and judged violating by CO)
  - Prevalence = % of impressions that have violating content, as measured by human raters
    - Content from 1% of impressions is run through a classifier, a sample of that is sent to human raters (with higher-likelihood-to-violate content upweighted), prevalence is calculated taking into account the weighting
- How can we compare DAP exposed (people active today who saw violating content today) and perceived prevalence in the survey (people active today who had negative experiences in the last week)
- Getting Devon's data
  - Devon is going to query and send the data for the same time period (November), including all violation types
  - Ask him to group together N&P + suggestive, hate + threats, commercial spam + copyright infringement, bullying + harassment
  - Ask him to limit to top 10 countries
- Tables:
  - IG data dex: https://our.internmc.facebook.com/intern/wiki/Instagram-data/core-tables/
  - Time spent: `ig_user_ts_datelist looks like the right one, maybe double-check with Tanya`
  - Likelihood of churning / L7 / L28 (ask Alec [redacted] about churn is defined)
  - Content production/Sharing (do you become less active, but still continue using passively): posts created, comments created, stories -- probably `fct_ig_media_creation`

Contact Support

Conversation History

Shilpa Mody edited Oct 16, 2020 at 6:31 pm

10/16 - chat with Arturo about prioritizing work on user perceived problems

How do we convince people that reducing negative experiences (even if not policy violating) is worth trading off a bit with growth and/or free speech concerns? Make the case w data -- this is the size of the problem, this is how bad it makes people feel, this is how many people leave because of it

Scoping the problem

people have a bad experience (measured by TRIPS)

--> they feel agency to do something about it (measured by negative actions)

--> resolution (we take down the content, SR action solves their problem)

What can we do to decrease the gaps between these?

- Between 1 and 2
  - understand why people don't use available tools -- e.g. no tool fits their motivations well, they think we won't take action anyway, diffusion of responsibility, not really bothered that much, don't know where tools are
  - potentially... create new tools, surface tools better, help users feel they can make a difference
- Between 2 and 3
  - improve reporting flow so people complete the flow, e.g. better language
  - improve reporting flow so people get a resolution
    - branching to triage the problem to appropriate actions (e.g. sub-problem selection that includes non-violating categories)

- help people protect the community, even for non-violating problems
    - nudge authors to create less bad content
        - direct person-to-person feedback
        - indirect/aggregate signals to author like "X people said this made them feel bad"
    - use aggregate negative user signals in soft actions, and tell reporters that we're doing so
- help people see less stuff they won't like, e.g. personalization

---

Shilpa Mody edited Jan 21, 2020 at 5:35 pm

01/21 - Survey metrics for supportiveness, transparency, and fairness w/ bryan [redacted] and mingliang

- Past and future projects
    - Actor/reporter appeal
    - Reporting flow optimization
    - Account appeal: smoother user experience
    - Support inbox: centralized place for all reporting/appeal notifications and status
    - Reporting flow optimization: better education screens
- Current metrics
    - Volume of actioned reports, volume of actioned appeals (quantity)
    - Action rate of reports, action rate of appeals (quality)
    - Report outcome view rate (do people see the outcome) -- goal metric for FB team, this is the best predictor of feeling supported
- Bryan -- they came up with 3 survey tools that capture
    - Based on procedural justice. Regardless of whether people get the outcome they want, do they feel good about the experience
    - Supportiveness: breaks down into transparency, fairness, voice (they cover 60% of supportiveness)
    - These should be similar for IG, but consider differences in content type and who is reporting
    - 2 tools
        - post-support survey - asks about reporting experience
        - complete reporter experience survey (3 day delay, so they should have received outcome response by then) -- asks about both reporting experience and outcome
    - Difficulty is in understanding how to weigh impact of moving survey metrics -- e.g. if supportiveness goes up, how do we trade that off with other metrics?

Shilpa Mody enabled link sharing · View/Edit Aug 16, 2019 at 5:16 pm

---

Shilpa Mody edited Aug 16, 2019 at 5:07 pm

Meeting Random meeting notes

08/15 -- WB research Chat with Arturo Bejar

- for the company, reporting flows are meant to identify content that violates policies, so the categories are based around policy.
    - if you have a reporting flow and the action rate is low, you have a bad reporting flow
    - reducing reporting is misguided, there's no such thing as a bad report. their experience/emotion is real.

    - reporting should reflect categories and language that reflect users' experience
        - example: "hate speech" doesn't mean anything in India -- landed on "makes fun of something that's important to me"
        - this raises engagement with the reporting flow, but also leads to lots of abandoned reports
        - "were your needs met?" -- people who dropped out of the flow often said yes, because they just wanted to express what happened to them. This should be the metric, not finishing the flow.
    - example: many nudity reports had no nudity, but did include themself in the photo.
        - had an open-ended thing you can send to the person who posted it, but low completion rate
        - added "it's embarrassing" as a prefilled message you can send -- completion rate skyrocketed
        - 90% of people who post content that's reported have good intentions, but it's hard to give them feedback (either public, or depends on other users taking initiative)
    - designed reporting flows that lead to different things -- e.g. high...
        - line is -- e.g. antivax. that's not close to this.
- what do you do with all the non-policy-violating stuff?
    - social feedback loops
        - reporting flow -- one of the options for teens was "it was annoying". if they pick that, suggest that they tell their friend that it's annoying (provide default language to use that they can edit)
        - did research on the language they should use (I statements, nonviolent communication, etc) -- range of politeness
        - humans want to help people in distress, as long as we don't feel threatened. people who receive these messages are highly receptive. it doesn't always lead to people taking down the content, but it starts a dialogue -- (why is this embarrassing, I didn't even notice, etc). but once this starts, people think twice when they post next time.
    - does this work for everyone? what about cultures that don't like to talk about feelings?
        - the data suggested it still worked in those countries (people filled out the flow at the same rate, people receiving the message were just as satisfied with it)
- why aren't we using these flows anymore?
    - the teams got broken up, different teams are responsible for different parts of them now, institutional knowledge gets lost
- Look up this Compassion research -- Jennifer Guadagno will know about it.

Shilpa Mody edited Aug 13, 2019 at 6:44 pm

Shilpa Mody edited Jun 6, 2019 at 9:27 pm

06/06 - Bryan [redacted] re: Developing sentiment metrics

- Team needed a measure that could help with ship/no-ship for quickly-iterated product experiments, and for quarterly goaling
- Sentiment measurement for product changes is probably a bad idea (post from Curtiss) - wide error bars, insensitive to incremental changes, leads to demoralization
- Sentiment should be related to your specific product experience (i.e. CAU is too broad), and is probably not just one specific measure (one number isn't granular enough to capture what you're looking for, and not specific enough to be able to move)
- Phrase 1: develop framework. A year full of qual research to ID themes about what matters to reporters, which aligned well with the model of procedural justice, so ran a quant test to validate (supportiveness = transparency + voice + fairness)
    - Quant test: asked questions about transparency, voice, fairness, and then several outcome measures including supportiveness (experience, resolution, CAU). The facets correlate well with all the outcome measures, but best with supportiveness. Supportiveness was also least correlated with actual problem resolution, which is good.

- More quant and qual testing (links in deck) to understand what each facet actually means to people, and to pick questions
- Phase 2: testing framework. Understand if this framework can actually be used to measure sentiment for goaling and experimentation.
  - Good goals: quantify your mission, tied to something inherently valuable, easy to understand, easy to measure, hard to game, help you learn something, can you move it.
  - Easy to understand: cognitive testing. ID language issues, validate that a yes/no question could convey respondent intent, help pare down similar questions
  - Easy to measure: quant testing in 5 countries. Exploratory factor analysis and IRT on massive set of potential questions. It takes a long time.
  - Can you move it. Used v1 of survey tools to attempt to ID differences in sentiemnt on 9 product experiments that we hypothesized should improve sentiment.
- I had to leave halfway -- get notes from Hitomi!

Shilpa Mody edited Mar 22, 2019 at 10:48 pm

03/22 - elena davis re: wb measurement on the fb side

- There's a spectrum of changeable vs matters in the big picture of someone's life
  - If we're talking about big-picture overall WB, big things (death in the family, etc) make an impact, but feature changes aren't going to register
- Big-picture overall WB is not going to be useful for product changes, but because we inherently care about it, it's definitely useful for:
  - Detecting natural variation, by demographic variables or usage variables (posting on FB is related to higher WB)
  - Longitudinal -- e.g. Posting on FB is related to higher WB, and increasing posting is related to increased WB -- lets you hold demographic & other variables constant since they're the same people. This is something that's never been done on instagram but would be useful.
- Ryan has done work showing that recent WB on FB is related to general WB, but it's actually more related to brand sentiment. They haven't looked at facets of WB, only overall.
  https://our.internmc.facebook.com/intern/anp/view/?id=61068
  - Their current assessment of this is that WB-on-FB could still be a useful measure, but digging into it more now.
- Elena is currently doing a proof of concept survey experiment measuring WB at 3 levels: overall, in the moment, and in the moment on facebook. All of these are overall WB,..
- look for meaningful behaviors that are correlated with (or cause) WB. We could get hypotheses from FB knowledge (e.g. active use), academic literature, etc. Then we'd try to drive up those behaviors and assume it's leading to better WB, even if we can't detect that change.
- Elena is also advocating focusing on specific facets -- we know that they're related to overall WB from research, so we don't have to verify every time. We can focus on ones we have some expectation of changing as a way of narrowing in and gaining power. To make sure it's not affecting other areas, you could still include them (e.g. everyone gets the facet of interest + 1 random other one as a countermetric).
  - For facets, Facebook is concerned about: social comparison, loneliness, problematic use/control
- Compassion -- nobody is focusing on this specifically, it's all being wrapped up as predictors of WB (along with demographics, etc). But she thinks that sucks and that we should totally work on it.
  - Separate into things that happen on our platform (integrity problems, social comparison) vs. things that happen off-platform but that we might exacerbate
- Elena has a deck and is currently getting buy-in, but she's the right POC for all this work and I should check in with her in a few weeks

---

Shilpa Mody edited Mar 4, 2019 at 8:43 pm

03/04 - wendy re: safety survey

- She's currently running a survey about people's perception of safety on IG. The goal is to have data that helps with a messaging campaign about IG taking safety seriously. They want to ID focus areas in each country (issues that people think are prevalent, that are IG's responsibility, but where IG isn't doing enough), and also pick messaging that works for each focus area.
    - Reputation - have you heard of this happening on IG?
    - Responsibility - do you think IG should do something about it?
    - Action - do you think IG is currently doing enough?
    - Messaging -- will probably end up something like "Instagram is working hard to keep me safe"
- Running this survey both in-app (mobile only) and email -- nobody has done this before! The goal is to gain a general understanding of which approach is better and/or how we can expect data to vary by surface
- Also starting a survey on pressure to be perfect (working with policy -- they ideally want data to back up a campaign about posting "real" content, e.g. X% of people feel better after seeing authentic content)
    - How often do you feel better/worse after comparing yourself to someone else on IG
    - How much content do you see that's meant to look perfect vs. real
    - How does perfect vs. real content make you feel
    - Breakdown by friend content vs influencer
- Setting up biweekly syncs

Shilpa Mody enabled the outline for the document Jan 17, 2019 at 8:28 pm

---

Shilpa Mody edited Jan 17, 2019 at 8:23 pm

01/17 - Kyle re: Nielson study

- From his work on feed, one of the major pain points is people feeling like they're missing posts or that they won't see things on time. They've found that people who are anxious about this are the ones who tend to get upset about ranking changes, etc. The "you're all caught up" feature is a reaction to that.
- Can we add some questions measuring "feed anxiety" to the baseline questionnaire? Then slice questions about well-being/fomo/feeling caught up (especially anxiety-related) by feed anxiety.
    - He's going to send me his questions!

01/17 - Lauren & Frank re: TRIPS & IG integration

- They're wondering whether we're ready/eager to begin testing/rolling out their version of the CI tracker to Instagram. Basically, this would mean testing their questions / flow and comparing results to a) their results, and b) our current results. We assume that questions will have to be changed because of differences in the product (e.g. follows, explore), differences in how people use and talk about the product (e.g. following interests and influencer accounts), and different populations. So, we'd be writing question variants and comparing them to ground truth, cognitive testing translated versions, etc.
- From their side, Guy and others in leadership have asked to be able to compare FB to IG (though that conversation last happened 6 months ago). Is there the same appetite for comparison on IG? Is there appetite for an ongoing (daily) tracker?
    - They're going to check in with CI leads today and see if comparing to IG is still top of mind.
- It'll be easier this time, since I can start with what they've already learned. The challenge will be knowing what to do if we find inconsistent results with what they found.
- If there's no current time pressure from our side, that would actually make things better in some ways because we can work on it slowly without someone breathing down my neck. Lauren estimated 1 day a week. Ana would be working closely with me.
- An open question is how we would split the work, and whether IG or FB would own it.

01/14/2019 01/14 - Devon re: perceived reach & intensity

Shilpa Mody edited Jan 14, 2019 at 8:11 pm

Meeting notes

01/14/2019 - Devon re: perceived reach & intensity

- Definitions of different measurements
  - DAP exposed = % of DAP who had an impression of known violating content (i.e. content deleted in the 7 days following)
    - Content is deleted when caught by a classifier (can be automatically deleted for drugs & N&P, but otherwise a CO person judges), or when reported (and judged violating by CO)
  - Prevalence = % of impressions that have violating content, as measured by human raters
    - Content from 1% of impressions is run through a classifier, a sample of that is sent to human raters (with higher-likelihood-to-violate content upweighted), prevalence is calculated taking into account the weighting
- How can we compare DAP exposed (people active today who saw violating content today) and perceived prevalence in the survey (people active today who had negative experiences in the last week)
- Getting Devon's data
  - Devon is going to query and send the data for the same time period (November), including all violation types
  - Ask him to group together N&P + suggestive, hate + threats, commercial spam + copyright infringement, bullying + harassment
  - Ask him to limit to top 10 countries
- Tables:
  - IG data dex: https://our.intern.facebook.com/intern/wiki/Instagram-data/core-tables/
  - Time spent: `ig_user_ts_datelist` looks like the right one, maybe double-check with Tanya
  - Likelihood of churning / L7 / L28 (ask Alec [redacted] about churn is defined)
  - Content production/Sharing (do you become less active, but still continue using passively): posts created, comments created, stories – probably `fct_ig_media_creation`

Shilpa Mody created the document Jan 14, 2019 at 6:54 pm
