

Message

From: James [REDACTED] [/O=THEFACEBOOK/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=[REDACTED]]
Sent: 1/28/2022 12:59:47 AM
To: Udi [REDACTED] [REDACTED]@fb.com]; Liangyu [REDACTED] [REDACTED]@fb.com]; Randy [REDACTED] [REDACTED]@fb.com]; David [REDACTED] [REDACTED]@fb.com]; Rishi [REDACTED] [REDACTED]@fb.com]; Kripa [REDACTED] [REDACTED]@fb.com]; Shankar [REDACTED] [REDACTED]@fb.com]; Daniel [REDACTED] [REDACTED]@fb.com]; Nalin [REDACTED] [REDACTED]@fb.com]; Steve [REDACTED] [REDACTED]@fb.com]; Alison Lee [REDACTED] [REDACTED]@fb.com]; Haoyang [REDACTED] [REDACTED]@fb.com]; Gerardo [REDACTED] [REDACTED]@fb.com]; Sarah [REDACTED] [REDACTED]@fb.com]; Rei [REDACTED] [REDACTED]@fb.com]; Ece [REDACTED] [REDACTED]@fb.com]; Glenn [REDACTED] [REDACTED]@fb.com]; Rong [REDACTED] [REDACTED]@fb.com]; Santhosh [REDACTED] [REDACTED]@fb.com]; Mingliang [REDACTED] [REDACTED]@fb.com]; Ashish [REDACTED] [REDACTED]@fb.com]; Peter [REDACTED] [REDACTED]@fb.com]; Meagan [REDACTED] [REDACTED]@fb.com]; Dennis [REDACTED] [REDACTED]@fb.com]; Sindhu [REDACTED] [REDACTED]@fb.com]; Sayed Otaru [REDACTED] [REDACTED]@fb.com]; Chenyang [REDACTED] [REDACTED]@fb.com]; Kathryn [REDACTED] [REDACTED]@fb.com]; James [REDACTED] [REDACTED]@fb.com]; Meghana [REDACTED] [REDACTED]@fb.com]; Abhro [REDACTED] [REDACTED]@fb.com]; Ming [REDACTED] [REDACTED]@fb.com]; Hua-Wei [REDACTED] [REDACTED]@fb.com]
Subject: Message summary [{"otherUserFbId":null,"threadFbId":4837972866290214}]

Ashish [REDACTED] (1/27/2022 11:20:25 PST):
>Perfect. Good start

Alison Lee (1/27/2022 11:24:38 PST):
>resharing and repinning tracker: <https://docs.google.com/document/d/1ZCrjO-97tLDy0GaIpBkJ0vFEoZdCVcvBYnaNTcxITM/edit#>

Alison Lee (1/27/2022 11:24:55 PST):
>#shareadoc

Ashish [REDACTED] (1/27/2022 11:25:50 PST):
>Let's also have a kickoff meeting next week (Thu/Fri) after PSC so managers can also join.

Shankar [REDACTED] (1/27/2022 11:26:44 PST):
>sorry I dropped the ball on that one, scheduling something right now and will send an invite link here for folks to join.

Shankar [REDACTED] (1/27/2022 11:34:41 PST):
>this is the meeting invite for next Thurs 9.30a PT/12.30p ET: <https://fburl.com/meeting/e305nsym>

James [REDACTED] (1/27/2022 11:35:08 PST):
>#til

Santhosh Vinod [REDACTED] (1/27/2022 14:41:42 PST):
>@Shruti - Are you still leading the Integrity effort for AYMF?
>
>Context : This is about Preference amplification via suggested accounts and could introduce new Enforcements on AYMF services.

Shruti [REDACTED] (1/27/2022 14:44:14 PST):
>I'm no longer involved here but I'll add in the TL/PM for context

Santhosh Vinod [REDACTED] (1/27/2022 14:52:52 PST):
>@Dennis - We're looking for surface integrity representation for AYMF to reduce preference amplification (other surface teams are also in loop). We have an initial kickoff this week.
>
>would you be able to help represent on behalf of AYMF (and growth overall?). Please feel free to loop in anyone else that might need to be involved. Thanks!

Dennis [REDACTED] (1/27/2022 14:54:17 PST):
>Hi Santhosh! Sorry noob question, what is preference amplification? Happy to be looped in. Incidentally we're also starting to look more into integrity efforts on Graph this half.

Santhosh Vinod [REDACTED] (1/27/2022 15:14:29 PST):
>Not a noob question at all!
>In fact, one of the first action items for the group is to clearly define (or scope down) Preference amplification. The broad definition
>"Recommender systems can 'rabbit-hole' users into a certain type of content. In the world of integrity, this could specifically lead to recommending unsafe content that might lead to high engagement, however, could result in negative experiences"
>

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY'S EYES ONLY)

META3047MDL-040-00261600
METANMAG-003-00588224
PLTF EX-01195_0001

PLTF EX

PLTF EX-01195

>Here is a paper that dives into this topic:
><https://research.facebook.com/publications/preference-amplification-in-recommender-systems/>
>
>We can define Preference amplification from multiple angles:
>a. Existing users slowly drifting into policy violating content after engaging with borderline content
>b. New users who join Instagram are sucked into an echo chamber due to a cold start problem
>c. Regulators creating fake insta accounts to emulate teens to showcase Instagram's inactions in integrity enforcement
>and more..
>
>The first step for this group would be to help scope this down, identify goals/metrics and align on possible enforcements across the different recommender systems across the app.

Rong [REDACTED] (1/27/2022 15:36:26 PST):
>We tried to define 'signal amplification' as: the content recommendation system may show a kind of content more over time to a given user (or a set of users), in response to their prior engagement signals on similar media.
>
>- our team focuses on content recommendation, but the definition could be generalized to account recommendation
>- the engagement signals that get 'amplified' may not be users' actual 'preferences', so we tried to avoid the term 'preference' and called it 'signal amplification'
>- we want to define it as a neutral phenomenon for all recommendation systems, no matter the user is new or old, though users with few historical signals may be more vulnerable.
>
>Happy to get feedback too!
>
><https://fb.workplace.com/groups/485465535579115/permalink/1094871364638526/>

Shankar [REDACTED] (1/27/2022 16:59:47 PST):
>can't add much to @Santhosh's and @Rong's excellent expositions, but I did want to say a few things:
>
>* When we used the term preference amplification in the paper above, we distinguished user's inherent/true preferences from those learned by the recommender system. We are referring to this latter concept for the most part when talking of amplification; however, we shouldn't rule out the effect the recommender system may have (over time, admittedly) on the user's true preferences. E.g., perhaps a user starts out with an initial preference of cats (0.2) vs. dogs (0.8) but maybe over time, both exogenously as well as through engagement with our RS swamping them with pics of chihuahuas, their true preference may start to _drift_ to cats (0.5) and dogs (0.5). Ultimately, all these are borne out through explicit signals the user gives and so both terms are generally compatible with one another in spirit.
>* I don't think it's controversial to claim that the most pressing integrity risks right now are what users perceive on connected content (not to mention downstream effects on Explore using a user's graph for sourcing unconnected content). To that end, I think it's important for us to look beyond just the surfaces RI owns and measure this amplification with respect to any of the growth surfaces. It is in this context that we'd love to engage with teams owning account recommendations.
>* Drebbel had success last half in implementing seed filtering based on "gateway groups" for GYSJ; the basic idea is to filter out gateway groups as being considered for fanout/candidate generation (so we're not directly enforcing on the gateways themselves, but just reducing the likelihood of overlap between groups similar to gateways and those that are non-rec). We think this is a fairly light touch lever that we should be able to apply easily to AYM and/or other such growth recommendation units.