

FILED UNDER SEAL

EXHIBIT 24

Message

From: [REDACTED] [/O=THEFACEBOOK/OU=EXTERNAL
(FYDIBOHF25SPDLT)/CN=RECIPIENTS/CN=[REDACTED]]
Sent: 8/22/2022 2:00:25 PM
To: [REDACTED]@fb.com]; [REDACTED]@fb.com]; [REDACTED]@fb.com];
[REDACTED]@fb.com]
Subject: Message summary [{"otherUserFbId":null,"threadFbId":5381041991975158}]
Attachments: 300254897_522586736292168_3034713114060022893_n.png;
300183841_765582437893736_6164706288390047700_n.png;
300427210_625526365807210_7938773429010650533_n.png;
300127656_481995853310996_1196244511317400600_n.png;
300232705_413632950870826_7603688651882829297_n.png;
300034887_371603251657379_6954546939559685766_n.png;
300459168_481001870087416_8903395493895026899_n.png

[REDACTED] (8/22/2022 01:09:12 PDT):
>Ok friends, is there anything else we need prior to tomorrow's chat with [REDACTED] and Tom?

[REDACTED] (8/22/2022 01:10:43 PDT):
>If you haven't seen my "defence of follow signals" I write it into
<https://docs.google.com/document/d/1GQnFPnQlbeiWXLjLbnOM8hNf0CceIapiRAVfVQMvInE/edit#> along with a
detailed analysis of 20 people who would hit the "strong" flag with following. (spoilers: they were all
seeking some form of porn/sexual conversation)

[REDACTED] (8/22/2022 01:12:02 PDT):
>#shareadoc

[REDACTED] (8/22/2022 01:12:14 PDT):
>wrote*

[REDACTED] (8/22/2022 01:13:14 PDT):
>Great question.
>
>I think where we have updated the actual signal definitions we should include that in the description
(eg NFX with 'direct' to align with UnIS)

[REDACTED] (8/22/2022 01:13:55 PDT):
>Ok, and that's the case, [REDACTED] in your analysis? You used the "with at least one direct" criterion?

[REDACTED] (8/22/2022 01:15:52 PDT):
>That's the only one. right?

[REDACTED] (8/22/2022 01:17:21 PDT):
>And also, are we advocating that applies to 10+ blocks from minors as well?

[REDACTED] (8/22/2022 01:20:32 PDT):
Tom had strong views on making sure we could avoid FPs for
>(3) re other users tagging you
>- can we check if it's possible to add caveat 'and not untagged themselves / reported'
>
>(4) being followed by >X accounts (Y % of followers)
>- I think you have some examples here but let's be ready to justify the %. Could also make this based on
new followers over a given time period eg 1yr
>
>(12) message receipts of nmcrc reported content
>Is there a way we can check receptiveness eg NFX following receipt?

[REDACTED] (8/22/2022 01:22:37 PDT):
>re this - i think latest sizing is updated - [REDACTED] can confirm
>
>we need to land on a threshold so I'd say just align with the UnIS enforcement threshold?

[REDACTED] (8/22/2022 01:24:20 PDT):
>> (4) being followed by >X accounts (Y % of followers)
>- I think you have some examples here but let's be ready to justify the %. Could also make this based on
new followers over a given time period eg 1yr
>
>This % is also the number DOI have long used for "this is a significant number of your graph"

[REDACTED] (8/22/2022 01:24:38 PDT):
>It's kind of standard on investigator side. Also if we change the lookback period we'll need to change the threshold.

[REDACTED] (8/22/2022 01:25:15 PDT):
>I *also* think this isn't as clear because (unlike blocks or likes) we actually have to find and take down any accounts. This takes time so there's a detection bias that we'll need to build in.

[REDACTED] (8/22/2022 01:26:01 PDT):
>(3) re other users tagging you
>- can we check if it's possible to add caveat 'and not untagged themselves / reported'
>
>I expect we can do this. This is lower priority on the implementation front, so I think we have options.

[REDACTED] (8/22/2022 01:26:39 PDT):
>> Is there a way we can check receptiveness eg NFX following receipt?
>
>It's very, *very* unusual to report all of the incoming CSAM if you are being targeted.

[REDACTED] (8/22/2022 01:26:58 PDT):
>This is why the threshold is really high with a carveout for high-following people.

[REDACTED] (8/22/2022 01:27:01 PDT):
>Do we see [REDACTED]?

[REDACTED] (8/22/2022 01:27:31 PDT):
[REDACTED]

[REDACTED] (8/22/2022 01:27:53 PDT):
>The threshold is the guardrail. A normal person receiving 30 pieces of CAM *that we actually catch* is really, really high.

[REDACTED] (8/22/2022 01:28:29 PDT):
>Overall, these FP concerns are built into the thresholding. (tags excepted)

[REDACTED] (8/22/2022 01:29:47 PDT):
>I think it's less a worry about the threshold precision and more about the premise of taking control away from an account

[REDACTED] (8/22/2022 01:30:12 PDT):
>even if it's 99% accurate, the 1% FP is a particularly bad kind of FP (for IG Policy)

[REDACTED] (8/22/2022 01:30:17 PDT):
>Yeah ... but that's being weaponised against us. And the response is that if you aren't into these communities, you won't notice the soft actions.

[REDACTED] (8/22/2022 01:32:47 PDT):
>re message receipts can we look over a period of time

[REDACTED] (8/22/2022 01:32:58 PDT):
>eg if it's coordinated, likelier to come over short time

[REDACTED] (8/22/2022 01:33:17 PDT):
>but maybe if we lookback over a few weeks and see they received something every week

[REDACTED] (8/22/2022 01:33:29 PDT):
>Our detection isn't that strong.

[REDACTED] (8/22/2022 01:33:48 PDT):
>otherwise how valuable is this signal post [REDACTED]

[REDACTED] (8/22/2022 01:33:55 PDT):
[REDACTED]

[REDACTED] (8/22/2022 01:34:03 PDT):
[REDACTED]

[REDACTED] (8/22/2022 01:34:07 PDT):
>maybe kill it then? get macsa going now

[REDACTED] (8/22/2022 01:34:22 PDT):
>i really don't

[REDACTED] (8/22/2022 01:34:33 PDT):
>don't wanna spend effort getting things now that die with [REDACTED]

[REDACTED] (8/22/2022 01:35:08 PDT):
[REDACTED]

>So we *already* have Policy permission to use this for disable. Are we just saying ... not use this? Or just not include in the soft action section?

[REDACTED] (8/22/2022 01:35:45 PDT):
>I'd skip it for soft actions on this lama

[REDACTED] (8/22/2022 01:35:56 PDT):
>concession to Tom

[REDACTED] (8/22/2022 01:36:45 PDT):
>Ok, as long as we can use a bargaining chip.

[REDACTED] (8/22/2022 01:37:11 PDT):
>(I think [REDACTED] didn't even have this in the analysis?)

[REDACTED] (8/22/2022 01:38:26 PDT):
>Or it was really minimal. Also we don't have tags in the data either so that's also something we can punt if he wants to be more thoughtful about mitigations.

[REDACTED] (8/22/2022 01:38:35 PDT):
>So of the 3 Tom commented on:
>
>(3) re other users tagging you -> add caveat for showing engagement eg not untagged themselves / reported'
>(4) being followed by >X accounts (Y % of followers) -> let's make the case for this at current thresholds with examples
>(12) message receipts of ncmecc reported content -> we can concede this one

[REDACTED] (8/22/2022 01:39:14 PDT):
>Ok, so follow*ing* we have the examples for. I can also do follow*ed* by

[REDACTED] (8/22/2022 01:41:17 PDT):
>I think the other thing we need to emphasise is the targeted-ness of these soft actions. Again, let's imagine that some crazy thing happens and a huge group of malicious child actors follow you in a coordinated way (so far I have never seen such a thing, but I can be surprised). The only thing you'll lose is discovery of new minors. You can't get too strong (which is just losing the minor follow button) without something else.

[REDACTED] (8/22/2022 01:42:01 PDT):
>The latest sizing doc includes this scenario ☺ 5+NFX and direct

[REDACTED] (8/22/2022 01:42:59 PDT):
>Difficult to parse this conversation, what is "this" in this sentence?

[REDACTED] (8/22/2022 01:43:09 PDT):
>inbox deletes?

[REDACTED] (8/22/2022 01:43:14 PDT):
>receives 30+ ncmecc

[REDACTED] (8/22/2022 01:43:34 PDT):
shared: 300254897_522586736292168_3034713114060022893_n.png

[REDACTED] (8/22/2022 01:43:39 PDT):
>Yes it's in there as a primary, and 20+ as a secondary

[REDACTED] (8/22/2022 01:43:58 PDT):
>but it wasn't on the drivers breakdown, so I'm assuming this is tiny

[REDACTED] (8/22/2022 01:44:20 PDT):
>It's tiiny

[REDACTED] (8/22/2022 01:44:33 PDT):
>Like 1k people for secondary and 422 people for primary

[REDACTED] (8/22/2022 01:44:58 PDT):
>Ok. Actually [REDACTED], that's the discussion point.

[REDACTED] (8/22/2022 01:45:17 PDT):
><0.01% so it rounds down to 0

[REDACTED] (8/22/2022 01:45:19 PDT):
shared: 300183841_765582437893736_6164706288390047700_n.png

[REDACTED] (8/22/2022 01:45:22 PDT):

>This is really, stupid hard to get to.

[REDACTED] (8/22/2022 01:45:52 PDT):
>And that's with the creator/follower guardrail, right?

[REDACTED] (8/22/2022 03:51:22 PDT):
>(@ [REDACTED] are we suggesting 1 direct with the 10+ blocks?)

[REDACTED] (8/22/2022 03:51:47 PDT):
>I haven't done this comparison, 10+blocks is just 10+ blocks

[REDACTED] (8/22/2022 03:51:58 PDT):
>Isn't block direct?

[REDACTED] (8/22/2022 03:52:11 PDT):
>... good question

[REDACTED] (8/22/2022 03:52:14 PDT):
>Sounds like it to me :)

[REDACTED] (8/22/2022 03:52:28 PDT):
>not sure - i think you can block from the profile surface

[REDACTED] (8/22/2022 03:52:50 PDT):
>let me try this

[REDACTED] (8/22/2022 03:52:53 PDT):
>direct = in response to something the actor did to you

[REDACTED] (8/22/2022 03:52:57 PDT):
>Yes :-) just wanted to double check from the query:
>
>> primary_ncmec_recip AND NOT is_high_engagement,
>
><https://www.internalfb.com/intern/daiquery/workspace/387171320194009/2168442193337140/>
>
>FYi the query for these questions when I'm on PTO

[REDACTED] (8/22/2022 03:53:08 PDT):
>Let me check what's direct

[REDACTED] (8/22/2022 03:53:42 PDT):
>ok, you can block a user you aren't following, etc. Just tried with a random person

[REDACTED] (8/22/2022 03:53:50 PDT):

shared: 300427210_625526365807210_7938773429010650533_n.png

[REDACTED] (8/22/2022 03:53:56 PDT):
>block_messages is included as direct

[REDACTED] (8/22/2022 03:54:09 PDT):
>but not a general block

[REDACTED] (8/22/2022 03:54:28 PDT):
>I think that's right. Block is unlikely to be used in an attack, right?

[REDACTED] (8/22/2022 03:54:58 PDT):
>Like if we all get together and block [REDACTED] what happens to him ... nothing I would be aware of, right?

[REDACTED] (8/22/2022 03:55:44 PDT):
>Especially because you can't get disabled only on blocks.

[REDACTED] (8/22/2022 03:56:38 PDT):
>You'd get hard feature limited though since it's primary

[REDACTED] (8/22/2022 03:56:45 PDT):
>So you'd get composer-block type stuff

[REDACTED] (8/22/2022 03:56:51 PDT):
>(Gabe would if we blocked him)

[REDACTED] (8/22/2022 03:57:00 PDT):
>"we" being 10 minors lol

[REDACTED] (8/22/2022 03:57:14 PDT):

>Is the request: 10+blocks and 1 direct?

[REDACTED] (8/22/2022 03:57:21 PDT):

>Does that solve the problem?

[REDACTED] (8/22/2022 03:58:14 PDT):

>For IG it's only that you can't see follow buttons for minors

[REDACTED] (8/22/2022 03:58:46 PDT):

>My question is ... should we size this?

[REDACTED] (8/22/2022 03:59:07 PDT):

>Sure thing, can't commit to it today but can do it over dinner tonight and have it for you tomorrow

[REDACTED] (8/22/2022 04:00:10 PDT):

>The threshold was not sized for this, but let's see what the drop is.

[REDACTED] (8/22/2022 04:00:15 PDT):

>units is 5+ for measurement, and 7+ for enforcement right?

[REDACTED] (8/22/2022 04:00:27 PDT):

>(enforement has the 1+ direct condition)

[REDACTED] (8/22/2022 04:00:29 PDT):

>For IG

[REDACTED] (8/22/2022 04:00:31 PDT):

>yes

[REDACTED] (8/22/2022 04:02:34 PDT):

>primary blocks is only 13% of strong gating so that's an upper bound for what's going to change if we add "direct" on.

>

>it doesn't make up a lot of diables either (1.1%)

[REDACTED] (8/22/2022 04:02:36 PDT):

shared: 300127656_481995853310996_1196244511317400600_n.png

[REDACTED] (8/22/2022 04:03:05 PDT):

>So feels like this'll be fine to change, like with the data we have now 😊

[REDACTED] (8/22/2022 04:05:11 PDT):

>going to 7 however will be a significant hit to IIC T2 overlap, right?

[REDACTED] (8/22/2022 04:05:36 PDT):

>(at the Moderate level)

[REDACTED] (8/22/2022 04:06:01 PDT):

>Yes, wed go down to 5.1

[REDACTED] (8/22/2022 04:06:30 PDT):

>against measurement or enforcemet?

[REDACTED] (8/22/2022 04:06:33 PDT):

>>Overlap done:

>5+ NFX: 10%

>5+ NFX and 1 direct: 7%

>7+ NFX and 1 direct: 5%

[REDACTED] (8/22/2022 04:07:03 PDT):

>That's comparison between the moderate feature gating from MACSA (enforcement) and the IIC T2 .3 heuristic (currently used for feature gating aka enforcement)

[REDACTED] (8/22/2022 04:08:10 PDT):

>ohh gotcha

[REDACTED] (8/22/2022 04:10:55 PDT):

>so I can't run nice easy overlaps anymore since dim_fuss_ig doesn't have the 2022-07-21 segment anymore

[REDACTED] (8/22/2022 04:10:57 PDT):

shared: 300232705_413632950870826_7603688651882829297_n.png

[REDACTED] (8/22/2022 04:11:24 PDT):
>You should have features on later ds now, though.

[REDACTED] (8/22/2022 04:11:38 PDT):
>Right but that's a big overhaul for like

[REDACTED] (8/22/2022 04:11:42 PDT):
>redoing everything for a different day

[REDACTED] (8/22/2022 04:11:48 PDT):
>Like, reasonably big overhaul

[REDACTED] (8/22/2022 04:11:49 PDT):
>fair point

[REDACTED] (8/22/2022 04:12:01 PDT):
>let me see if I can just uhaul dim_fuss_ig_users for that older date

[REDACTED] (8/22/2022 04:13:22 PDT):
>Yeah, I think we should likely argue that the threshold of 5 is more correct given the overlap with IIC T2.

[REDACTED] (8/22/2022 04:13:33 PDT):
>That's my personal thought

[REDACTED] (8/22/2022 06:43:04 PDT):
shared: 300034887_371603251657379_6954546939559685766_n.png

[REDACTED] (8/22/2022 06:43:15 PDT):
>These are the stats for 5+ and 1direct for secondary
>
>And 10+ and 1 direct for primary

[REDACTED] (8/22/2022 06:43:54 PDT):
>oh ignore that

[REDACTED] (8/22/2022 06:44:17 PDT):
>doesn't look right, let me double check

[REDACTED] (8/22/2022 06:49:37 PDT):
>yeah something must have gone mangoed with these tables, not going to be able to debug during the next 10 mins while Haotian is presenting

[REDACTED] (8/22/2022 06:50:05 PDT):
>let's push to UnIS team then?

[REDACTED] (8/22/2022 06:50:42 PDT):
>Wait ... like the push the result?

[REDACTED] (8/22/2022 06:50:54 PDT):
>like, isn't this their table / problem?

[REDACTED] (8/22/2022 06:50:56 PDT):
>Or push that the tables are broken

[REDACTED] (8/22/2022 06:51:02 PDT):
>Cool your beans man

[REDACTED] (8/22/2022 06:51:09 PDT):
>I dont' know yet what is broken

[REDACTED] (8/22/2022 06:57:47 PDT):
>Coolio, it is because the unis ig table doesn't exist for that date anymore

[REDACTED] (8/22/2022 06:57:49 PDT):
>☹️

[REDACTED] (8/22/2022 06:57:51 PDT):
shared: 300459168_481001870087416_8903395493895026899_n.png

[REDACTED] (8/22/2022 06:58:06 PDT):
>I can update everything to be 2022-08-02

[REDACTED] (8/22/2022 06:58:11 PDT):

>I think all the tables exist for that date 😊

[REDACTED] (8/22/2022 06:59:20 PDT):

>Should be ok.

[REDACTED] (8/22/2022 06:59:27 PDT):

>Let me know if you're missing any features.

[REDACTED] (8/22/2022 07:00:00 PDT):

>Maybe cheeky question but should we go back to signals for the for implementation focused chat to keep this one higher signal for Alex?

[REDACTED] (8/22/2022 07:00:25 PDT):

>Yeah 😊 I should probably just chat directly with [REDACTED] I'm sure this is all very low value for Alex(!)