

FILED UNDER SEAL

EXHIBIT 18

Message

From: [REDACTED] [REDACTED]@fb.com]
Sent: 2/26/2021 5:52:39 PM
To: [REDACTED] [REDACTED]@fb.com]; [REDACTED] [REDACTED]@fb.com]
Subject: Message summary [{"otherUserFbId":100036360240939,"threadFbId":null}]
Attachments: 155396120_888306231963337_2332544339900682754_n.png;
154278033_238116781326642_8844695330812967545_n.png;
154558026_234137721781306_7211199988893414722_n.png

[REDACTED] (2/26/2021 04:33:31 PST):

>Hi [REDACTED]
>I needed a help with one account: 320772272709477

[REDACTED] (2/26/2021 04:34:00 PST):

>This user has been disabled fro grooming, however when we reviewed with job id: 320772272709477

[REDACTED] (2/26/2021 04:34:15 PST):

>we had made a decision of not_a_groomer

[REDACTED] (2/26/2021 04:34:45 PST):

>account is: 100005708838661

[REDACTED] (2/26/2021 04:35:29 PST):

>so would it be possible to know what country have been deciding factor of disabling user which we couldn't find in SRT job?

[REDACTED] (2/26/2021 04:40:55 PST):

>yep yep

[REDACTED] (2/26/2021 04:50:16 PST):

>Hi [REDACTED]
>
>So when I look at that user you're correct we took no action on grooming job
>
>But when I check in Centra - can you open 'centra 1616265641907048' admin record -> disabled by Policychain: AUTOMATIC_IIC iic autoaction disable / IICOnlineHarmOfflineClassifier : "from 243443154"
>
>This user seems to be really spammy if you look at the messaging and friending behavior, so I suspect that's why automatic disabled account. But when I look in their messages the user is reaching out to dozens of users (adults and minors) but nothing sexual, and usually just 1-2 messages in conversation.

[REDACTED] (2/26/2021 04:51:05 PST):

>[REDACTED] was going to get us a query for all users disabled by SRT review OR automation, which were then reinstated by a human (so working an escalation Task outside of SRT as we're not working appeals) to see is there a problem with auto disabled

[REDACTED] (2/26/2021 04:51:09 PST):

>*auto disables

[REDACTED] (2/26/2021 04:51:44 PST):

>got it. This is very helpful. So it seems we checkpointed the user but UFAC kicked in and disabled him may be because of multiple checkpoints

[REDACTED] (2/26/2021 04:51:47 PST):

>As we are seeing Oops Tasks 'frequently' (5-10 a week) that are automation disables and we can't figure out why they'd be classified as groomers.

[REDACTED] (2/26/2021 04:52:21 PST):

shared: 155396120_888306231963337_2332544339900682754_n.png

[REDACTED] (2/26/2021 04:52:22 PST):

>okay, are these the one which shouldn't be groomer and were wrongly marked as groomer?

[REDACTED] (2/26/2021 04:52:39 PST):

shared: 154278033_238116781326642_8844695330812967545_n.png

[REDACTED] (2/26/2021 04:52:53 PST):

>yes they will be when she gets to it, she has a few open data requests

[REDACTED] (2/26/2021 04:52:54 PST):

>got it, makes sense

[REDACTED] (2/26/2021 04:53:44 PST):

>this will be very useful

[REDACTED] (2/26/2021 04:54:09 PST):

>if we are auto-disabling user based on current classifier score, we should stop doing this, because classifier is not very good currently

[REDACTED] (2/26/2021 04:58:08 PST):

>where did this come up and I'll mention to Reva

[REDACTED] (2/26/2021 05:18:48 PST):

>i was trying to look at current grooming model to identify what is an issue (why calibration score are < 50%

[REDACTED] (2/26/2021 05:19:36 PST):

[REDACTED] (2/26/2021 05:22:35 PST):

[REDACTED] (2/26/2021 05:23:09 PST):

>[REDACTED]

[REDACTED] (2/26/2021 05:23:45 PST):

>[REDACTED]

[REDACTED] (2/26/2021 05:24:22 PST):

>[REDACTED]

[REDACTED] (2/26/2021 05:25:32 PST):

>sure :)

[REDACTED] (2/26/2021 06:48:59 PST):

>Sorry [REDACTED] I didn't mean to imply you weren't absolutely killing it last half, I just broke a lot more of [REDACTED]'s stuff! <3

[REDACTED] (2/26/2021 06:49:36 PST):

>ah no issue at all :)

[REDACTED] (2/26/2021 07:41:21 PST):

>H1

>Would you mind telling me difference between IIC T2 vs T1?

[REDACTED] (2/26/2021 07:42:32 PST):

>i mean, when we would call something is IIC T2? T1 I partially understand but not T2

[REDACTED] (2/26/2021 08:26:21 PST):

>[REDACTED]

>[REDACTED]

>[REDACTED]

[REDACTED] (2/26/2021 08:41:09 PST):

>[REDACTED]

[REDACTED] (2/26/2021 08:43:02 PST):

>[REDACTED]

>[REDACTED]

[REDACTED] (2/26/2021 08:43:27 PST):

>okay, thanks this is very useful :)

[REDACTED] (2/26/2021 08:50:59 PST):

>Final question (sorry for asking so many questions)

>What is difference between groomer and potential groomer?

[REDACTED] (2/26/2021 08:51:19 PST):

>And why are we labelling potential groomer as negative labels?

[REDACTED] (2/26/2021 08:52:36 PST):
>Potential Groomer is either of these two things

[REDACTED] (2/26/2021 08:52:38 PST):

shared: 154558026_234137721781306_7211199988893414722_n.png

[REDACTED] (2/26/2021 08:52:49 PST):
[REDACTED]

[REDACTED] (2/26/2021 08:54:43 PST):
>In that case, shouldn't that be something we can consider as positive example for training and evaluating?

[REDACTED] (2/26/2021 08:55:25 PST):
>Currently in evaluation, we only consider groomer or groomer escalation as positive and all other as negative

[REDACTED] (2/26/2021 08:57:26 PST):
>Ehhhhhhhhhhh maybe it should be after 2 x potential IIC decisions and not one particularly if it is 'adult using minor's account'

[REDACTED] (2/26/2021 08:59:34 PST):
>sorry, would you mind a quick call? (if you would prefer I can schedule one on Monday)

[REDACTED] (2/26/2021 08:59:45 PST):
>i am free at 5:30 if that's ok?

[REDACTED] (2/26/2021 08:59:58 PST):
>sure :)

[REDACTED] (2/26/2021 09:00:05 PST):
>thanks a lot [REDACTED]

[REDACTED] (2/26/2021 09:32:19 PST):
>sorry will be two min

[REDACTED] (2/26/2021 09:33:38 PST):
>np :)

[REDACTED] (2/26/2021 09:49:41 PST):
>[REDACTED]

[REDACTED] (2/26/2021 09:51:11 PST):
>English, India, Portuguese, Spanish

Lisa Reilly (2/26/2021 09:51:42 PST):
>500-1000

[REDACTED] (2/26/2021 09:52:39 PST):
>https://drive.google.com/drive/u/0/folders/157D7rtg4066h1ZcAbMghEUcZ_grbcVGL