

AMENDED Exhibit 961

PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

Case No.: 4:22-md-03047-YGR

MDL No. 3047

In Re: Social Media Adolescent Addiction/Personal Injury Products Liability Litigation

A/C Privileged: Proactive vs. Reactive Defenses

October 2020 (updated April 2021) | Safety Strategy & Operations

Overview: Define “proactive vs. reactive” defenses as we work to include it in the H1'20 Transparency Report (TR) (to be published in the fall of 2021).

Proactive Detections & Enforcements for Listed Harms: ^[R1]

Listed Harm	Do we proactively review & enforce content / accounts?	Currently included in the TR?	Notes
Sexual Content	✓	✓	Autobot / rapid rules are mapped 1:1 w/ this harm.
Regulated Goods	✓	✗	Some autobot / rapid rules are mapped 1:1 w/ this harm (see appendix)
Hate Speech	✗	✗	Autobot / rapid rules are not mapped 1:1 w/ this harm.
Threatening / Violence / Harm	✓	✗	Some autobot / rapid rules are mapped 1:1 w/ this harm (see appendix)
Impersonation	✓	✗	Autobot / rapid rules are not mapped 1:1 w/ this harm. We have detection/enforcement tools in spam and abuse but need to improve our data tracking tools.
Harassment	✗	✗	Autobot / rapid rules are not mapped 1:1 w/ this harm.
Spam	✓	✗	Autobot / rapid rules are not mapped 1:1 w/ this harm. We have detection/enforcement tools in spam and abuse but need to improve our data tracking tools.
CSAM	✓	✓	PhotoDNA + rapid rules proactively detect & enforce.
Terrorism	✗	✗	Autobot / rapid rules are not mapped 1:1 w/ this harm.
False Information	✓	✗	Some autobot / rapid rules are mapped 1:1 w/ this harm (see false info doc for details)

Proactive Detections & Enforcements for each product/feature:

Product / Feature	Do we proactively review & enforce content / accounts?	^[A7] Currently included in TR?	Notes
Snaps	✓	✓	Autobot / rapid rules implemented for proactive detection & enforcement.
Chats (1:1 & Group)	✓	✓	Autobot / rapid rules implemented for proactive detection & enforcement.
My Story	✓	✓	Autobot / rapid rules implemented for proactive detection & enforcement.

Page 1 Comments

R1 @ [REDACTED] am thinking we can move these 2 charts to the other proactive doc to have it all captured there. and this one can just be our notes, decision items (ie the appendix)

[REDACTED], 6/9/2021 09:37 AM

A2 @ [REDACTED] can you explain what you mean by proactive vs. reactive? as of last week, all lenses are reviewed before they go live, so I would consider that proactive.

Althea Tupper, 5/3/2021 11:11 AM

N3 Proactive = anything that is reviewed (by human or automatically) before a user sees or before it is reported

[REDACTED] 5/3/2021 11:51 AM

N4 so in the case of lenses, based on last week's update, it would be proactive!

[REDACTED], 5/3/2021 11:51 AM

K5 Should we note the date when we stopped the velocity model and all lenses started to be reviewed?

[REDACTED] 5/3/2021 01:06 PM

N6 I think it [REDACTED] to know for ourselves internally; can add it to the notes column [REDACTED]

[REDACTED] 5/3 PM

A7 Technically though all Lenses are live first and then reviewed within an hour.

Althea Tupper, 5/3/2021 01:57 PM

Lenses	✓	✗	All lenses are proactively reviewed before they go live (4/2021). ALD is used in-app to filter search results. Automation not mapped 1:1 w/ listed harms for lenses, as the reporting menu does not capture listed harms.
Maps (Venues)	✓	✗	Autobot / rapid rules implemented for proactive detection & enforcement. Not mapped 1:1 w/ listed harms, as the reporting menu does not capture listed harms.
Our Story	✓	✓	Autobot / rapid rules implemented for proactive detection & enforcement.
Snap Minis + Games	✗	✗	We review partnerships proactively, but we don't take responsibility for content on other people's apps since we can't detect when the app developer changes things post-submission. No proactive detection; only reactive detection for Bitmoji Paint.
Publisher Tiles	✓	✗	We do near time review of Publisher Tiles. Publisher Tiles go live and then are moderated shortly after for compliance. We also do proactive checks via ALD and other methods. Approved publishers can post tiles & stories w/o additional review on Discover.
Publisher Tiles: Stories	✗	✗	No proactive detection, approved publishers can post tiles & stories w/o additional review on Discover.
Public Profiles	✓	✗	Public profiles go through an approval process w/ partnerships team beforehand. Therefore, they are not mapped 1:1 w/ listed harms, as they are detected / enforced by another team. However, our proactive detections & enforcements still apply for CSAM + sexual content.
Web Stories (ext. shared stories)	✓	✗	Story extensions that can be shared and opened in browser; subject to same detection & enforcement as My/Our story
Snap Star Tiles & Stories	✗	✓	All reviewed similar to stories, but no specific detections in place for Snap Stars.
Ads	✓	✗	Tagging & tools are implemented to catch fraud + other ad violations.
Spotlight	✓	✓	Autobot / rapid rules implemented for proactive detection & enforcement.

Appendix

April 2021 Update: In the H1'20 TR, we determined that we would aggregate proactive and reactive detections & enforcements when reporting our listed harms in the TR. At the time, we only had proactive detection & enforcement data for Sexual Content & CSAM. We have updated this document with charts delineating which products, features, and listed harms we proactively review & enforce.

Moving forward, there a few things we should consider doing:

1. Amend the H1'20 TR: Either we remove the proactive detections from the listed harms tables & country sub-pages, or the asterisk under the listed harms table, “*The Content Reports reflect alleged violations via our in-app and support inquiries” should be edited to reflect that we included proactive detections & enforcements.
 - a. **Decision:** We will not include proactive detections in the upcoming TR
2. No longer include proactive detections that expired. These detections, by nature, may be misidentifications as we refine the autobot's ML programs.
 - a. **Decision:** We will not include proactive detections in the upcoming TR

Initial Proactive enforcement for H2 2020 (not incorporated into H2 2020 TR): [REDACTED], [REDACTED] and [REDACTED] are working to determine the total number of proactive enforcements across each listed harm and feature. Currently, we have identified **203,066 proactive enforcements in H2 2020**. See appendix for details.

H2'20 Proactive detection & enforcement rules (autobot)

Current Queue	Total Tasks	Content or User Enforced	tasks expired (unreviewed)	pct expired	comment	Possible Listed Harm
Spotlight Experiment - Proactive Weed+Smoking	348,742	0	348,729	100%	New	Regulated Goods
Rapid Rules - Voter Suppression	180,162	419	168,596	94%	New	
Rapid Rules - False Information	109,162	4,982	78,994	72%	New - but sounds like we will include it	
Spotlight Experiment - Proactive Guns	86,075	33	85,625	99%	New	Regulated Goods
Rapid Rules - Review & Refine	67,318	18	67,029	100%	New - represents all kinds of abuses, has enforcements	
Rapid Rules - Potential Incitement	29,516	4,229	20,170	68%	New	Threatening / Violence / Harm
Spotlight Experiment - Proactive Weed	13,702	1,732	6,460	47%	New	Regulated Goods
Proactive Public Story - Weed	12,003	8,011	2,757	23%	New	Regulated Goods
Content Immune - Uncategorized	8,355	17	3,984	48%	New	
Proactive - Trafficking	4,920		4,920	100%	New	
Spotlight Experiment - Proactive Drugs	1,054	0	1,044	99%	New	Regulated Goods
Proactive Public Story - Guns	1,046	611	31	3%	New	Regulated Goods
Rapid Rules - Potential Incitement w/ Guns	27	2	7	26%	New	Regulated Goods
Total	862,102	20,654	788,348	91%		

Initial Proactive enforcement for H2 2020 (not incorporated into H2 2020 TR):

Listed Harm	Content and Accounts Enforced
Sexual Content	148,915
Regulated Goods	10,389
Hate Speech	--
Threatening / Violence / Harm	4,229
Impersonation*	--
Harassment	--
Spam*	--
False Information	4,964

CSAM	34,551
------	--------

*There are proactive enforcement tools for mitigating spam and impersonation, but our ability to track data in the spam & abuse team for these particular categories need to be fleshed out (improvements by EOY).

October 2020 Research & Considerations

Definitions:

- Proactive - We proactively found the content or user (by human or automatically) without the need of a Snapchatter reporting it.
 - In H1, it was solely Public UGC (discover, map, our story, my story w/ visibility = public). We also scanned Photo DNA (PDNA) private stories & PDNA private chats
 - In H2, we are going a bit further and scanning private content, specifically private stories for voter suppressions & "rapid rules"
 - "Rapid rules" = detections based on rules the t&s team creates (eg. media caption contains "covid is a hoax")
 - We're also adding new technology that can match CSAM videos (launch is public story and small test w/ private 1:1); [quip](#)
 - The discover content moderation team uses the proactive definition to mean before the publish time, and reactive means after the publish time.
- Reactive - We reacted to a Snapchatter's report of content or user.

Open Questions

- In H1: What listed harms categories (e.g. harassment & bullying, spam, etc.) are captured by proactive and reactive detection? Which categories fall under proactive detection and which fall under reactive detection?
 - A: all listed harms categories are captured by reactive.
 - In H1, sexual content is captured by proactive detection.
 - In H2, it will be misinformation, voter suppression, and possibly more
- Is Photo DNA considered proactive detection when reporting on it for the TR? Yes

Reactive Detections

Listed Harm Reason	Content or Account Flagged ¹	Content or Account Enforced ²	Report Expired ³
Sexual Content	6,063,523	33.6%	46.0%
Regulated Goods	520,426	33.4%	45.1%
Hate Speech	229,375	55.4%	13.5%
Threatening/Violence/Harm	1,062,895	42.9%	17.3%
Impersonation	1,459,480	0.1%	1.5%
Harassment	857,334	26.8%	20.5%
Spam	552,731	11.0%	18.9%
	10,745,764	32.9%	28.7%

Proactive Detections

Sexual Content	2,459,057	13.5%	86.3%
----------------	-----------	-------	-------

¹ Unique pieces of content or number of accounts flagged by reactive or proactive means

² Unique pieces of content or number of accounts determined violating community guidelines

³ Unique pieces of content or number of accounts found not reviewed before media expiring

Presentation of the data:

- Option #1: Include a high-level, total breakdown between proactive and reactive detection, combining all listed harms categories. Additionally, break out proactive vs reactive detection per each listed harms category.
- Option #2: Only include a high-level, total breakdown between proactive and reactive detection.
- Option #3: Do not include it at all.

Understanding product proactive tools

- Step #1 - what is the number
- Step #2 - can we map those to a category
- Step #3 - what will it take to do so to map it

Reference Documents

[ALD & Proactive enforcement metrics](#)

[Product & feature reporting flows](#)