

AMENDED Exhibit 882

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

AEME Design Proposal

Overview

Background and Motivation

“AEME” stands for Auto-Moderation → Exploration → (Human) Moderation → Exploration. In today’s Spotlight framework, to be feed eligible, all snaps must pass through [1] a gamut of pre-moderation filters (e.g. watermarks, low media quality, etc.), [2] auto-moderation checks (e.g. ML models predicting likelihood of approval/rejection), and finally [3] human moderation. While [3] minimizes risk, this requirement has become the primary bottleneck in accepting and ultimately exploring newly submitted content. Given the rate at which submission volume is outpacing moderation throughput, this has resulted in an increasing backlog which ultimately results in a poor submitter experience.

The AEME framework aims to increase overall moderation throughput by allowing submissions to become Spotlight feed eligible based solely on [1] prefilters and [2] auto-moderation. In particular, Spotlight submissions that pass these two checks, will be eligible for “Tier 0” exploration. After achieving a certain number of views (exact threshold TBD), we’ll kick the snap back to moderation for human review. The snap must then pass human moderation before it is eligible for subsequent distribution (e.g. Tier 1/2 exploration, exploitation, etc.).

The goal of this [auto-moderation + initial exploration] approach is that we’ll be able to more easily scale exploration proportionate to engagement. Or stated another way, assuming a sufficient^[D1] volume of submissions, we’ll always have ample candidates to explore to fully utilize our exploration budget¹.

High-Level Exploration Candidate Flow

- Spotlight submissions ingested by Moderation service
- Submissions sent through pre-filters (*same as [1] from above*) – items below the filter score thresholds are rejected (*note: these filters are applied sequentially*)
- Candidates that pass all pre-filter checks are then evaluated by our auto-moderation ML models (*same as [2] from above*) – candidates with p(rejection) score below a certain threshold (e.g. our model believes there’s a low likelihood of human rejection) are eligible for AEME and thus sent to the [DaGamaIndexer](#) via SnapJoiner → JaguarBT (*note: by design, this an entirely separate index from the existing Magellan index*)
- These DaGama exploration snaps (“Tier 0”) are included in the Mixer and eligible for serving, alongside Magellan (Tier 1+) and exploitation content
 - The DaGama exploration index will be prioritized based on pre-moderation signals, rather than trying to get through all content.
- Once a DaGama exploration candidate achieves the Tier 0 criteria, the snap will not be published to the DaGama index (re: won’t be made available for serving) and the DaGamaIndexer will update SnapJoiner indicating this submission has hit this “milestone”
- Upon receiving this updated status, SnapJoiner will spin up a corresponding moderation task for SnapGraph
 - “DaGama graduate” moderation tasks will be onto themselves such that we’ll have queuing flexibility
 - Within “DaGama graduates” awaiting human moderation, we’ll (eventually) apply similar Moderation task ranking methodology to help further prioritize the candidates we want for

¹ Current upper bounded at ~20% of total Spotlight story views



Page 1 Comments

D1 DAVID TO DO – include current ingress/egress stats, highlight how goal of this effort isn't to explore all submissions but rather to help capitalize on our total exploration budget (e.g. exploration budget constrained)

David Lue, 11/19/2020 02:04 PM

Last Updated: 11-23-20 @ 10:00am

human review²

- "DaGamma graduates" that pass human moderation (SnapJoin status again updated) are then eligible for Tier 1+ exploration
- All subsequent view distribution adheres to our existing Magellan framework

Exploration Capacity Utilization

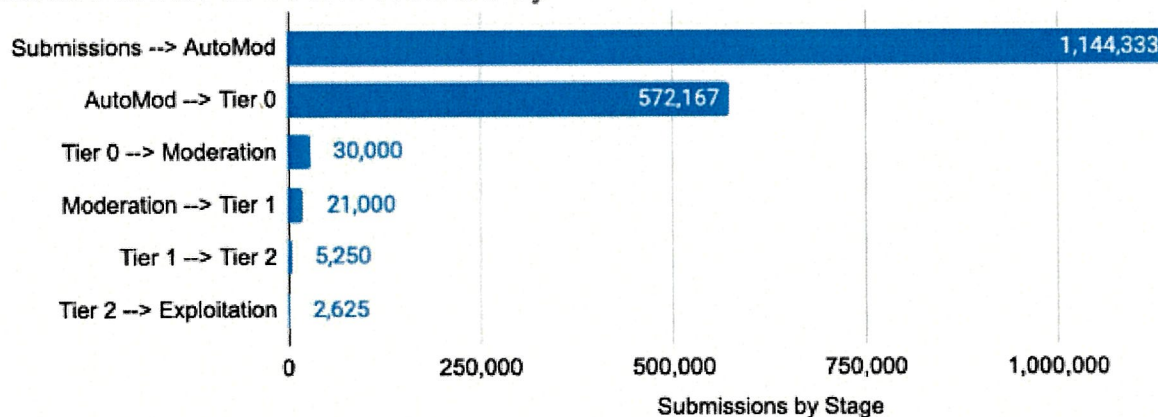
[\[model\]](#)

Configurations & Assumptions:

Prefilter/AutoMod % Accepted	50%
Tier 0: Exploration Budget	30
T0 → Mod Promotion Rate	5.2%
Moderation Acceptance Rate	70%
Tier 1: Exploration Budget	60
T1 → T2 Promotion Rate	25%
Tier2: Exploration Budget	300
Tier2 → Exploit Promotion Rate	50%
Max Human Moderation	30,000
Max Exploration Views	20%

Resulting funnel, assuming 100M views/day

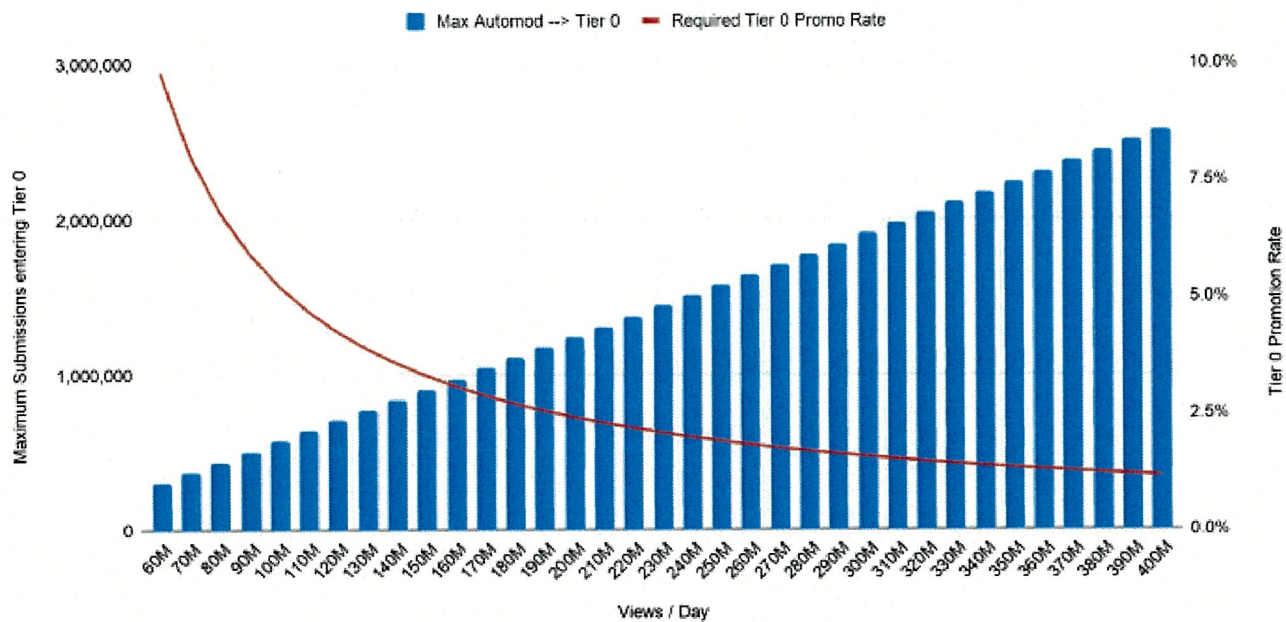
AEME Funnel at 100M Views/Day



If we constrain ourselves on only passing 30K items/day from DaGamma to moderation and always use 20% of views for exploration, then our Tier 0 promotion rate depends on the number of views. It could be anywhere from 10% (low view scenario) to just a percentage point or two.

² Even in this AEME world, it is possible we won't be able to manually review everything)

Last Updated: 11-23-20 @ 10:00am



Of course, we are not required to use 20% of views for exploration. Due to the limited number of views per submission in Tier 0 and the resulting randomness, it may make more sense to set a lower bound for our promotion rate, not use up the full 20% view budget, and improve the viewer experience.

Tier 0 Exploration Stories will be blended with Magellan exploration using the following logic [\[spreadsheet\]](#). By limiting Tier 0 views to be deeper in the feed, we also limit the number of views we can provide.

Start Pos	10	(first item is position 0)
Step	5	distance between explores
Step Delta	0	decline in distance between explores after each explore
Min Step	3	lower bound on distance between explores
Max Step	8	upper bound on distance between explores
T0 Start Pos	15	highest position for T0 story to be delivered
T0 Insertion Prob	0.5	Probability of picking T0 stories over T1/T2 stories

TO DO – areas to cover:

- Exploration Parameters (new and old)

Parameter	Value
Tier 0 View Target (New)	20

Last Updated: 11-23-20 @ 10:00am

Tier 0 Promotion Rate	20%
Estimated Human Approval Rate	60%
Tier 1 View Target	50
Tier 1 Promotion Rate	20%
Tier 2 View Target	300

- Back of envelope math for how much budget we can potentially allocate to AEME candidates (e.g. total exploration budget (20% of story views) - exploration views consumed by Magellan candidates (# of Tier 1 Magellan candidates can be upper bounded by human moderation throughput))
- Exploration views CDF by position? Thinking here is, if we wanted to insert AEME stories lower in the feed (to try and mitigate risk, note: not a guarantee given the story may appear higher in the feed on subsequent sessions due to client caching), how much does that reduce our AEME exploration capacity
- General insertion logic/interplay between Magellan and DaGama Index

[W5]

Pre-Filters, Auto-Moderation Modeling, and Thresholds

- Current state of the world (living document and details here: [Spotlight Moderation Content Filters & Model Features](#))
 - Filter out Snaps Content Quality score < 0.3
 - Filter out low media quality Snaps
 - Motion: 0.25
 - Sharpness: 0.05
 - Luminance: 0.2
 - FPS: 23
 - Aspect ratio Width >= Height
 - Filter out Snaps via driving/car classifier
 - Threshold: 0.85
 - Filter out horizontal Snaps
 - Threshold: 0.8
 - Filter out snaps from risky accounts (Account Risk Score)
 - Threshold: 0.5
 - Filter out snaps from minors
 - Threshold: (inferred age < 13) or (reported age <= 15 in FR/DE/IE)
 - Filter out Snaps with TT logo
 - Filter out snaps with NSFW score > 0.8
 - Filter out images
 - Filter out all non-vertical snaps
 - Filter out short videos
 - Threshold: < 3s
 - Filter out snaps with a high speed (proxy for driving):
 - > 10m/s
 - Filter out Snaps with any URL links

Page 4 Comments

- D2 @ [REDACTED] – will need y'all's help on this :)
David Lue, 11/19/2020 05:10 PM
- D3 Also @ [REDACTED], [REDACTED], [REDACTED]
[REDACTED] on any additional thinking of how we could potentially further leverage our pre-filters to try and send "better stuff" to AEME
David Lue, 11/20/2020 09:41 AM
- G4 yes! most CQ signals are numerical values that can be used for ranking after passing prefilter. Simple example would be that high FPS, bitrate, resolution can generally be ranked higher. Further, we can also do some analysis to find correlation between "boosted"/"starred" content with CQ signals. Not pre-filter specific, but Content team has also highlighted use of tools/time spent editing as another indicator of better quality. Happy to dive more into other ideas!
[REDACTED] 11/20/2020 10:03 AM
- W5 I suppose the features generated by CUP can be consumed by the auto-moderation or the ranking model after the snap passes the pre-filters.
- Visual Embeddings
 - Content Quality Score
 - Media Quality Metrics
 - NSFW
 - TikTok
 - Driving
 - Horizontal
 - SCC
 - YT8M
- [REDACTED] 11/20/2020 10:26 AM