

AMENDED Exhibit 514

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

[2023-06] Launch of Negative Affect Dispersion Strategy Global Roll-out [Wave4&5] and update&simplify dispersion strategy globally

Last Updated	June. 2023
Drafted by	[REDACTED]
Consulted with	@Ryn Linthicum (they/them)@[REDACTED]
Informed	

Executive Summary

bulb

This document serves as a pre-launch notice of the intention **launch of the dispersion strategy in wave 4&5 countries and update&simplify the dispersion strategy globally.**

Wave 4&5 Launch

- * Regions and languages impacted:
 - Languages: 12 Languages(AR,TH, MS, UR,RU,RO,NL,PL,SV,HE,ZH,EL)
 - **Region:**
'SA','OM','KW','MA','LY','MR','IQ','JO','BH','DZ','YE','QA','DJ','PS','LB','EG','TN','AE','KM','SO','MD','AM','GE','KZ','UZ','BY','KG','RU','TJ','UA','RO','TH','NL','BE','PL','SE','SG','HK','TW','MO','CN','CY','GR','IL','PK','MY'
- * A/B testing supports the launch, meeting target metrics agreed upon previously by product policy leadership for launch standards:
 - The model works as expected to significantly decrease the density of

EXHIBIT

24

TIKTOK-LINTHICUM

negative affect content TikTok users are recommended via the FYF.

- There is no evidence of increased risk of harms in the content presented to users in the dispersed feed.
- Stay duration, violation vv%, and unapproved vv metrics within standards for launch

Update&simplify the dispersion strategy globally(Wave 1-3)

In consideration of algo strategy consistency and simplification, we plan to replace the online strategy with a uniform negative affect weighted model globally(wave 1-3). We validated that:

- * The model works as expected to significantly decrease the density of negative affect content as before.
- * Stay duration, violation vv%, and unapproved vv metrics within standards for launch

Metrics for this model meet the launch standards previously sign-off on by product and policy leadership in 2022. Here we outline the key metrics informing the launch recommendation, to provide transparency to impacted policy personnel.

Please let us know ASAP if you have any feedback or concerns regarding the expected strategy launch. If we do not hear of any concerns, launch will proceed as outlined above.

Global Roll-out Progress Highlight



1. Background

* **Project Background:** In July 2021, WSJ flagged TikTok's algorithms revealed the risk of users entering "rabbit holes" (i.e. Filter bubbles) where users may get stuck in a loop of content featuring negative affect.

* **Algo Strategy Recap:** To date, we have achieved the following progress in global roll-out:

- US filter bubble user detection and dispersion strategy: launched in Jun,2022 [HYPERLINK ];

- EN/ES/FR general dispersion strategy for all users: launched in Sep, 2022 [[HYPERLINK ]]

- In the [HYPERLINK

[REDACTED] Policy Team and Product team aligned that, for each additional language, movement between each phase of testing and launch would be considered pre-approved by issue policy leadership (not require formal sign-off) as long as testing result meets success metrics prior to algo strategy launch. This is also aligned with the nimble sign-off approach taken by other dispersion projects (e.g., [HYPERLINK

- The Wave 2&3 model roll-out was launched in February 2023, to expand coverage to more languages and markets (as specified above)

- ✦ **Wave 4 &5 Roll-out:**

- We created a multilingual model to cover 12 additional languages (~[REDACTED] DAU) as wave 4&5, and opened AB testing in Feb 2023.

- In this doc, we introduce the validation result of wave 4 and 5 countries.

- ✦ **Update&simplify the dispersion strategy globally(Wave 1-3)**

- We merged all wave 1-5 regions into a unified weighted fusion model&comment model and aim to strategy consistency and simplification. We re-validated the top priority market in wave 1-3(US, GB, MX, FR,ID, BR) to make sure dispersion strategies are still valid online.

2. A/B Testing

2.1 Success Metrics

- ✦ Long-term goal: Keep bubble user DAU%(predicted by model score) < [REDACTED] (long-term goal: as close to 0% as possible) and target bubble density $\leq$ [REDACTED] per user as a safeguard threshold [[HYPERLINK [REDACTED]

- ✦ Short-term goal: we propose the first version of strategy can launch in market as long as it significantly decreases the bubble density in users' feed and improve users' experience, with no increase of higher-severity content after dispersion

- Terminology Borderline High (abbreviated as BH below) is used to categorize the violation videos under policy 4.0, e.g self harm and suicide, [HYPERLINK

[REDACTED], and related potentially harmful mental health content exempted from this dispersion strategy.

- ✦ Stay Duration top limit for filter bubble project $\leq$ [REDACTED]

2.2 Validation Methodology

2.2.1 Overall Principle

- ✦ For wave 4&5, we validated **representative countries by each language** in accordance with [[HYPERLINK](#)]

2.2.2 User Sampling Method

- * The representative countries are selected based on DAU. Both internal accounts and real users will be included in the validation journey.
- * [HYPERLINK ██████████] was applied to detect potential bubble users; we split them randomly into control and treatment groups to observe the change of negative affect bubble density;
- * According to the sample size [HYPERLINK ██████████], it is required to sample at least 30 users per region to get results at a significant level.

2.2.3 Languages & Country List

[HYPERLINK

3. Wave 4&5 AB Test Result

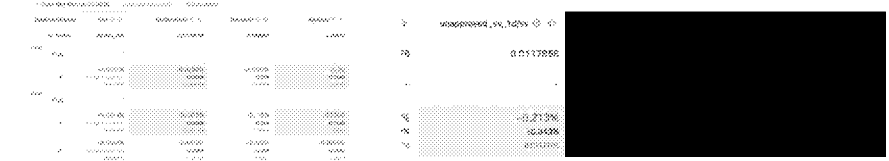
3.1 Summary

- Real users: bubble density change ██████ -> ██████; reduction ██████
- Internal accounts: bubble density change ██████ -> ██████;
- Key Metrics: stay duration ██████ (sig), Violation v ██████ (sig), unapproved vv-
██████ (sig)

3.2 Language Breakdown

Validation details: [HYPERLINK](#)

- AB metrics: 2023.4.20 -2023.6.4[[HYPERLINK](#)]

Group	Stay Duration	Unapproved_vv_ 1d/vv	Violation_vv_1 d/vv
	-0.055%(sig)	-0.213%(sig)	-0.564%(sig)
V2			

4. Globally update&simplify dispersion strategy

4.1 Background

We have merged all regions from Wave 1-5 into a unified weighted fusion model and comment model, aiming for strategy consistency and simplification. In order to expedite the dispersion strategy, we trained separate models for each country's launch. For instance, the US/EN/FR/ES markets utilized Fusion v3 and comment v1 models. Wave2&3 used Fusion v4 and comment v2, while wave4&5 employed Fusion v5 and comment v3 models. However, maintaining multiple models in the long run incurs significant costs, including ensuring the effectiveness of each model and allocating server resources for storing model data.

Considering these factors, we now intend to use a unified model to cover our global strategy and replace the previous dispersion strategy. To ensure the effectiveness of the new strategy in Wave1-3, we conducted an internal validation method in the top priority markets during Wave 1-3, including the US, GB, MX, FR, ID, and BR.

Please refer to previous launch[[HYPERLINK](#)]

4.2 Language Breakdown

Details: [[HYPERLINK](#)]

4.3 Strategy Comparison

Due to variations in performance across different countries for the new model, compared to the previous online strategy, we have standardized the recall of 90% of negative affect content for the Wave1-5 region and implemented the dispersion strategy. The following lists the changes in recall for the new and old dispersion strategies:

[REDACTED]

5. Post-Launch: Long-term monitoring

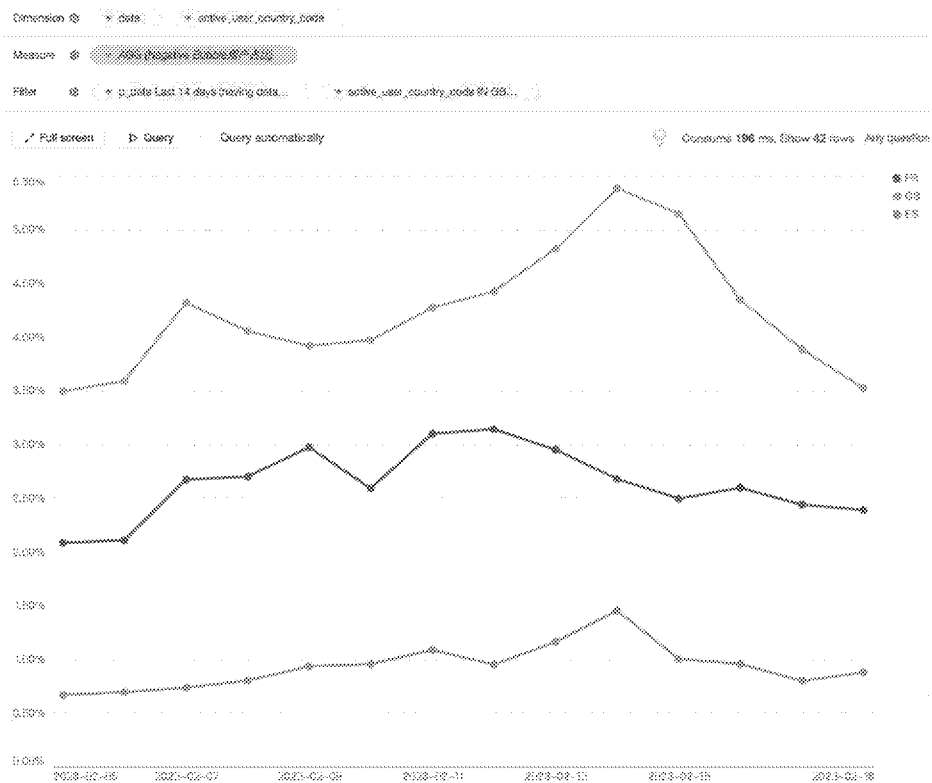
The validation data clearly shows improved feed quality and user safety resulting from the launch of the dispersion strategy. However, we note that the results display lower impact than expected when compared to the Wave 1 testing (where testing results exceeded expectations). As such, we plan on undertaking the following strategy to address any potential inequities stemming from differences in language model performance across markets:

- * We will further consult with CPMs from impacted regions, to understand if iterations on the global labeling guidelines need to occur to better localize labeling instructions to current market trends and cultural norms.
- * We will cross-compare our results with the societal rates of related mental health issues, such as depression and anxiety, to understand whether the differences across languages and countries in rates of content identified by the model follow expected patterns of symptoms across cultural context
- * Support our model on photo model to better control the dispersion strategy

Appendix

Long-term monitoring

- * We will partner with the Responsible Innovation team, to examine if the models launched exhibit unintentional bias or inequity against specific users, especially those with mental health conditions or representing specific markets or languages. Based on this testing, we will work to refine the models in future iterations.
- * Data Scientist team [REDACTED] has built the model-based filter bubble user dashboard to observe the filter bubble user DAU% metrics after the algo strategy launch[
HYPERLINK [REDACTED]
[REDACTED]
- * The dashboard is for the purpose of abnormal monitoring and can not represent the ground truth of filter bubble user DAU% in FYP, the number of which will be much smaller than model-score filter bubble users DAU%.



Example: GB/FR/ES region filter bubble user DAU% based on model score.

Algo Archive

1. Model performance

[HYPERLINK]

2. Evaluation recording

[HYPERLINK]

[HYPERLINK]

3. Why do we select the reduction of 50% as the short-term goal?

Credit to [HYPERLINK]

This is a calculator by using [[HYPERLINK \[REDACTED\]](#)]. If you provide before density, current model performance and algo strategy design (e.g 1/16 dispersion), the simulation calculator will tell you the expected after density. In negative affect scenario, we will choose reduction of ~50% as the short-term goal.

Simulation Result

Before density	Expected after density
30%	9.2494%
40%	11.8039%
50%	15.0292%
60%	19.2195%
70%	25.2939%
80%	34.9001%
90%	53.08%

DOCUMENT METADATA

Bates Number: TIKTOK3047MDL-060-01157747-TIKTOK3047MDL-060-01157754

BEGATTACH: TIKTOK3047MDL-060-01157525

ENDATTACH: TIKTOK3047MDL-060-01157793

Attachment Count: 0

PRODVOL: TIKTOK3047MDL-060

Custodian: RYN LINTHICUM; TIKTOK

File Path: /TIKTOK3047MDL-060-01157747.PDF

Confidentiality Designation: CONFIDENTIAL

HASHVALUE: 80F5C54C3AE05B3985F4EAA9AF92D2EA

DocType:

Author:

Create Date: 12/31/9999 7:00 PM

Last Modified Date: 12/31/9999 7:00 PM

LAST MODIFIED BY:

TRACK CHANGES:

COMMENTS:

HASHIDDENATA:

Filename: TIKTOK3047MDL-060-01157747.PDF

Title:

DOCEXT:

Email From:

Email To:

Email CC:

Email BCC:

Received Date: 12/31/9999 7:00 PM

Sent Date: 12/31/9999 7:00 PM

TIMEZONE:

Subject:

Case 4:22-md-03047-YGR Document 2875-63 Filed 03/25/26 Page 11 of 11

THREADID:

REDACTIONS: N

REDACTION TYPE: