AMENDED Exhibit 183

PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

Reset.

Evaluation of Instagram's Processes for Risks to Minors

III

Evaluation of Instagram's Content Moderation Systems in Creating and Perpetuating Risks to Minors

An Evaluation of Instagram's Content Moderation Systems in Creating and Perpetuating Risks to Minors

Research questions:

1. Does Instagram's adequately moderate pro-suicide and/or self-harm material when they become aware of it?
2. Does Instagram's adequately moderate pro-restrictive eating disorder material when they become aware of it?

Methodology

The research involved five steps:

1. **Developing criteria to define harmful material.**
   - This research explored two bodies of content posing psychological and physiological risks to minors: pro-suicide and/or self-harm material, and pro-restrictive eating disorder material.
   - We used the community guidelines for each platform to develop a coding schema to classify content (see Appendix 1 for more details). This ensures that only content violating Instagram's Terms of Service was included in this research. Each piece of content, according to their guidelines, should warrant a content-moderation action from Instagram.

2. **Identifying pro-suicide and/or self-harm material.**
   Using simple searches, we identified content on Instagram that met our criteria and had not been labelled by the platform already. We consulted a clinical psychologist who assessed each piece of content that was identified, confirming that it presented a risk to young people who consume it. Material that was not deemed to be harmful by a psychologist was not included in this research.

   In total we identified:
   - Pro-suicide and self-harm content: 119 pieces
   - Pro-restrictive eating disorder content: 125 pieces

   See Appendix 2 for examples of these bodies of content.

3. **Monitoring content pre-reporting.**
   We tracked this content for two weeks noting:
   - View counts and growth rates;
   - Labelling or warning rates, to ascertain whether any of this content was labelled by Instagram during these two weeks. We considered a piece of content labelled if an age-restriction warning, sensitivity filter, or any other sort of flag was placed on it; and
   - Take down rates, to ascertain whether any of this content was taken down by Instagram during these two weeks.

4. **Reporting the content.**
   We reported each piece of content as suicide and self-harm, or restrictive eating disorder content violating the Terms of Service to the platform.

5 **Monitoring content post-reporting.**

After reporting, we tracked this content for two further weeks noting:

- View counts and growth rates;
- Labelling or warning rates to observe if any content was labelled by the platforms during these two weeks. Considered labelled if an age-restriction warning, sensitivity filter, or any other flag was placed on it;
- Take down rates, to ascertain whether any of this content was taken down by the platforms during these two weeks.

According to our analysis of the platform's community guidelines (see Appendix 1), Instagram should delete pro-suicide and/or self-harm content, and pro-eating disorder content when they become aware of it. In practice, we often see platforms label and add sensitivity filters or age filters to this body of materials; we therefore also assessed these.

Below, we describe what we found over four weeks of monitoring.

Findings

Instagram's response to pro-suicide and/or self-harm material

Instagram does not appear to adequately label or demote pro-suicide and/or pro-self-harm content.

Removal appears to be the most common response to pro-suicide and/or pro-self-harm material, but the platform's reactions to reporting are inadequate. The majority of content remained available and unlabelled, even after user-reporting.

Over two weeks monitoring	Instagram
Pre reporting removal rate. This is the % of content that was removed during the two weeks before we reported it. It may have been reported by other users, and it is often not clear why content was removed (e.g. users may have deleted the content or their accounts, moved to private, or platforms may have deleted it). However this represents the best estimate of organic removal rates.	0 %
Post reporting removal rate. This is the % of content that was removed within 2 weeks after we reported it.	29.41 %
Effect of reporting on removal rate	**+29.41 %**
Pre reporting labelling or warning rate. This is the % of content that was labelled during the two weeks before we reported it. It may have been reported by other users, but represents the best estimate of organic labelling rates.	0 %
Post reporting labelling or warning rate. This is the % of content that was labelled within 2 weeks after we reported it.	0 %
Effect of reporting on labelling rate	**No change**
Pre reporting growth rate. This is the average growth rate of content over two weeks before we reported it (week-on-week).	0.63% growth week-on week
Pre reporting growth rate. This is the average growth rate of content over two weeks after we reported it (week-on-week).	0.1 % growth week-on week
Effect of reporting on growth rate	**-0.5 %**

Instagram's response to pro-restrictive eating disorder material

Instagram does not appear to adequately label or demote pro-restrictive eating disorder content.

Removal appears to be the most common response to pro-eating disorder material; however, Instagram's reactions to reporting are inadequate. Most of the content remained available and unlabelled, even after user reporting.

Over two weeks monitoring	Instagram
Pre reporting removal rate. This is the % of content that was removed during the two weeks before we reported it. It may have been reported by other users, and it is often not clear why content was removed (e.g. users may have deleted the content or their accounts, moved to private, or platforms may have deleted it). However this represents the best estimate of organic removal rates.	0%
Post reporting removal rate. This is the % of content that was removed within 2 weeks after we reported it.	10.40%
Effect of reporting on removal rate	**+10.40%**
Pre reporting labelling or warning rate. This is the % of content that was labelled during the two weeks before we reported it. It may have been reported by other users, but represents the best estimate of organic labelling rates.	0%
Post reporting labelling or warning rate. This is the % of content that was labelled within 2 weeks after we reported it.	0%
Effect of reporting on labelling rate	**No Change**
Pre reporting growth rate. This is the average growth rate of content over two weeks before we reported it (week-on-week).	3.46% (week-on-week)
Pre reporting growth rate. This is the average growth rate of content over two weeks after we reported it (week-on-week).	3.09% (week-on-week)
Effect of reporting on growth rate	**-0.38%**

Limitations

When content is removed, the reasons for its removal can be unclear. The users could have removed it, they may have deleted their account or switched to the private mode, or the platform may have removed their content or accounts.

Therefore, the estimations for removal rates represent the highest-end estimations of removal rates by platforms.

Conclusion

- Instagram under-moderates both pro-restrictive eating disorder content, pro-suicide, and/or pro-self-harm materials.
- There is a muted response to these materials when Instagram becomes aware of them via the user-reporting system; Instagram does not appear to adequately remove, label, or demote pro-restrictive eating disorder content, nor pro-suicide and/or pro-self-harm materials.