



Navigating the Gray Interventions: The Impact of Soft Moderation Techniques Across Social Media Platforms



Content warning: Sensitive content

The Tweet author flagged this Tweet as showing sensitive content.

ⓘ This tweet may be misleading. Find out more here.

1:27PM · Oct 4 2022 · Twitter for iPhone

1.3k Reply Share



Methodology and Data

@User_Study_Research

User Study via MTurk participants (**N ≈ 160**)

- **Age groups:** evenly split (18–24, 25–34, 35–44, 45+)
- **Gender split:** 76.3% male, 23.7% female
- **Platforms:** TikTok, Instagram, Threads, X
- **Tools:** Figma, Useberry, Python
- **Method:** Pre- and Post-engagement surveys

1:27PM · Oct 4 2022 · Twitter for iPhone



Results and Analysis

@Standardized_Evaluation_Metrics

👁 Engagement Metrics (Look)

- Younger users prefer contextual (less intrusive)
- Older users engage longer to understand risk of information
- Warning labels work better in Instagram than TikTok (likely due to platform fast-pace)

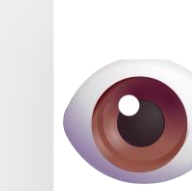
🧠 Perception Metrics (Perceive)

- Younger users are skeptical of labels, think it is less credible
- Both labels are effective for mid/older users
- X/Threads: contextual labels works better
- TikTok/Instagram: warning labels more effective

🔄 Behavior Metrics (Act)

- Mid-age group shares more unverified content
- Warning labels ignored: “not interested in additional info.”
- Contextual labels skipped: “trust in content” and “no time”

1:27PM · Oct 4 2022 · Twitter for iPhone



Would you engage with this content?



How credible you think this moderation is?



How you will act on this content?

Takeaway



Effectiveness of soft moderation techniques, varies significantly across demographics and social media platforms.

[Learn more from our paper.](#)

Recommendation



→ Hire more professional fact-checkers.

→ Tailored moderation techniques for older and younger Users.

→ Need transparency in moderation.

Namit Khurana[†],
Mohammad Majid Akhtar[†],
Rahat Masood[†],
Salil S. Kanheret[†],
Benjamin Turnbull[†]
([†]**UNSW Australia**)



UNSW
SYDNEY