AI LABEL WARNING RESEARCH STUDY

Team Gojo | Simon & Amy

TABLE OF CONTENTS

01 INTRODUCTION

OVERARCHING GOALS

- Study the effectiveness of various AI content labeling approaches
- Gain a better understanding of how users interact with AI labels, what makes an AI label more effective than another and what methods can be replicated to create a consistent and effective AI content label.
    - Very relevant to the current social media and UI/UX fields because AI is development which massively impacts both of these fields.

RESEARCH QUESTIONS

- How do regular short video-sharing platform users perceive and interact with warning labels on AI-generated content?
- How do warning labels affect people's accuracy in detecting misinformation?
- How do warning labels affect people's engagement with AI-generated content on short video-sharing platforms?

02 LITERATURE REVIEW

CURRENT STATE OF THE FIELD

- Prior studies explore AI-generated content and misinformation (e.g. deepfakes, ChatGPT content) and how users interact with algorithmic feeds.
    - Especially research on credibility judgments often focuses on long-form or static content (e.g. news articles, Facebook posts).
- Eye-tracking and think-aloud methods have been used in UX testing, but rarely applied to AI label visibility.
- However, Few studies have examined real-time user behavior toward AI warning labels in short-form video platforms like TikTok.
- Existing work rarely explores how label placement, design, and timing affect engagement and trust at the moment of interaction.

THE NEED

- As AI content grows, platforms must balance transparency with user experience.
- Poorly designed labels risk being ignored or distrusted, while overly intrusive ones cause disengagement.
- Our research fills a UX gap: it shows how actual users respond to AI warnings and what design strategies work in practice.

03 METHODS & RESEARCH

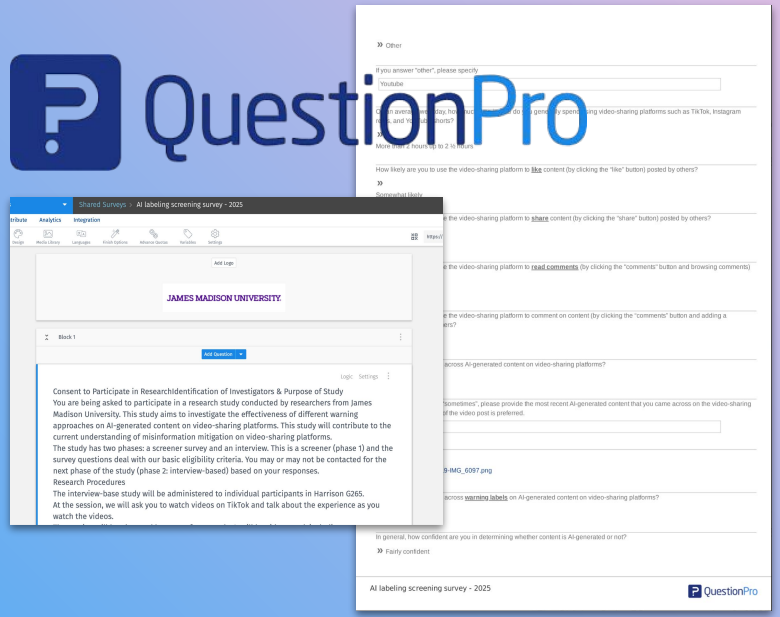
PROCEDURE

- Reserved Whisper Room and assigned roles (facilitator, observer, tech)
- Greeted participant, explained study, obtained Zoom recording consent
- Launched Expo Go app, calibrated Tobii Pro eye tracker
- Participant viewed 8 videos (AI and non-AI, various labels), think-aloud protocol
- Took notes on behavior, UI interaction, and verbal reactions
- Conducted semi-structured interview on credibility, labels, and engagement
- Saved recordings and eye-tracking data for analysis and report

03 METHODS & RESEARCH

RECRUITING

- All participants filled out a pre-screening survey on **QuestionPro** before conducting the research.
- Selected both rounds of participants by directly asking them to be involved with our study.
- All four participants who were all in their early 20s, interacted with social media regularly.
- Specifically recruited individuals whose screening survey responses indicated ***greater skepticism toward AI*** for second round of participants.



03 METHODS & RESEARCH

PHASE 1: Participants Elle and Cac.

CAC - BACKGROUND

- 21 years old, Male, Asian, Computer Information Systems major
- Frequent user of Instagram Reels, TikTok
  - more than 2 to 2.5 hours daily on video-sharing platforms.
- Extremely likely to like content and read comments, but extremely unlikely to share and somewhat unlikely to comment.
- Often comes across AI-generated content.
- Sometimes comes across warning labels.
- Fairly confident in determining if content is AI-generated.
- Slightly distrustful of AI-generated content on these platforms

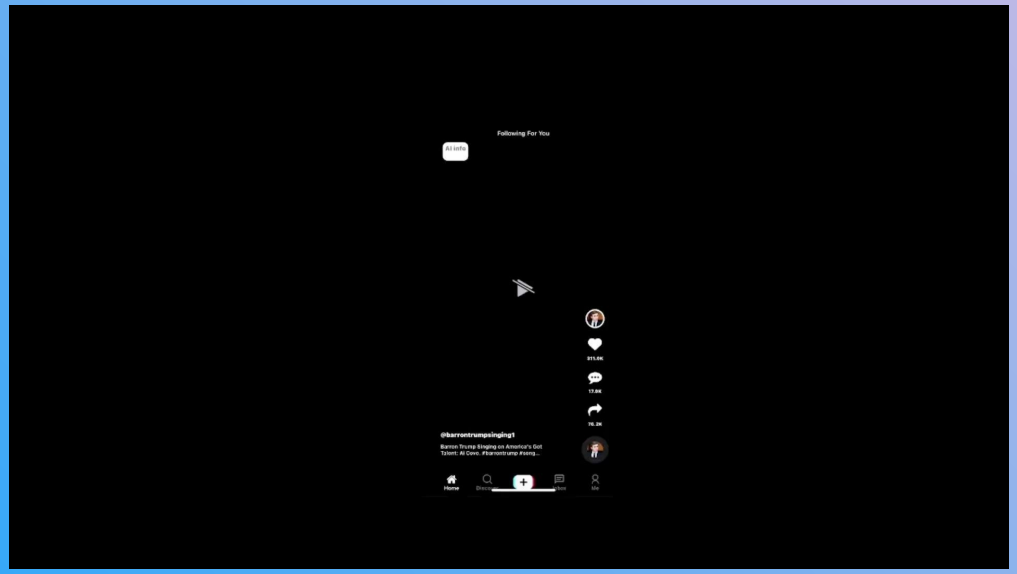
FINDINGS

- Less skeptical about AI than anticipated
- Was more open to interacting with AI and wasn't as interested in keeping track of sign of AI

EYE TRACKING

UX

Shown in this video is the Tobii eye tracking software we used to collect data.



This clip is from our think aloud session with Cac and it displays his attention moving across the screen and finally onto the highly contrasted AI label.

03 METHODS & RESEARCH

PHASE 1: Participants Elle and Cac.

ELLE - BACKGROUND

- 23 years old, Female, African American, Library and Information Sciences major, Bachelor's degree.
- Uses Instagram Reels, TikTok, and YouTube
  - spends more than 2 to 2.5 hours daily on video-sharing platforms.
- Somewhat likely to like and read comments, but somewhat unlikely to share or comment.
- Often comes across AI-generated content (most recent was Inzoi AI features).
- Hardly ever comes across warning labels. Fairly confident in determining if content is AI-generated.
- No trust at all in AI-generated content

FINDINGS

- More skeptical about AI than anticipated
- Had keen eye for visual incoherences or if visuals looked “too perfect”

03 METHODS & RESEARCH

PHASE 2: Participants Evan and Kori.

EVAN - BACKGROUND

- 21 years old, Male, White, SMAD major
- Uses TikTok, YouTube Shorts, and Twitter (X)
  - spending ½ up to 1 hour daily on video-sharing platforms.
- Extremely likely to like content, somewhat likely to read comments, and neither likely nor unlikely to share or comment.
- Sometimes comes across AI-generated content.
- Hardly ever comes across warning labels.
- Fairly confident in determining if content is AI-generated.
- No trust at all in AI-generated content.

> "Honestly, I tuned out once I realized it was AI. it was hard to focus"
> **– EVAN NETO**

FINDINGS

- More skeptical about AI
- Was very aware of visual inconsistencies in videos and was able to identify very quickly

03 METHODS & RESEARCH

PHASE 2: Participants Evan and Kori.

KORI - BACKGROUND

- 23 years old, Female, White, Philosophy major
- Uses Instagram Reels and YouTube Shorts
  - spending more than 1 to 1.5 hours daily on video-sharing platforms.
- Somewhat likely to like content, somewhat unlikely to share or read comments, and extremely unlikely to comment.
- Sometimes comes across AI-generated content.
- Never comes across warning labels.
- Slightly confident in determining if content is AI-generated.

"I fell for this for like a second... and then laughed."
– KORI HALL

FINDINGS

- Less skeptical about AI
- Was more entertained by the video and noticed AI markers more so when she took a moment to think it through

04 RESEARCH FINDINGS

Theme	ELLE	CAC	EVAN	KORI
Label Visibility	Often noticed, prefers central/top	Misses labels unless obvious	Sometimes delayed recognition	Mixed; some too subtle
Engagement	Disengaged after seeing AI label	No major change in behavior	Tended to scroll or zone out	Interest > label; little effect
Trust in Label	High when label clear	Relies more on instinct	Trusted clear ones, not vague	Neutral; not strongly swayed
Visual Cues	Judged blinking & movement	Trusted label more than visuals	Noticed syncing, facial stiffness	Flagged stiffness, “Botox face”
Suggestions	Place near caption/ username	Make cooler + less intrusive	Clearer, not gray or blend-in	More contrast, intuitive spot

04 RESEARCH FINDINGS

Stage	What Users Do	What We Observed	Design Implications
1. Initial Scroll	Scan captions, thumbnails, username	Eye-tracking shows little attention to corner tags	Place labels near caption/username area
2. First Impression	Focus on facial motion, tone of voice, realism	Users judge content based on “vibes” before noticing labels	Use visual cues (e.g., icons, subtle animation) to guide attention
3. Label Encounter	Occasionally notice tag (varies by placement/design)	Clear, centered labels more likely to be trusted and remembered	Ensure strong contrast, bold shape, center-top/side alignment
4. Credibility Judgment	Blend visual cues + label + prior knowledge	AI skepticism increased when label confirmed user hunches (but didn't always engage)	Labels should include hover/click info or a short reason indicator
5. Engagement/ Exit	Watch, scroll, skip, or zone out	Users with no trust disengaged regardless; some entertained despite knowing AI	Label design must balance disclosure with entertainment flow

DISCUSSION

KEY INSIGHTS

- AI labels are often missed or misread without clear placement or contrast.
- Even when labels are noticed, skeptical users disengage quickly.
- Label design must go beyond accuracy—it needs to align with natural user behavior.

WHY IT MATTERS

- Platforms risk losing user trust if AI content isn't clearly and effectively disclosed.
- UX teams, product designers, and moderators should rethink label presentation to support informed engagement.

LIMITATIONS

- Small sample size (n=4), all college-aged, U.S.-based.
- Simulated interface lacked real-world scrolling behavior.
- Label styles were pre-set, limiting user-driven interaction.

06 CONCLUSION

PURPOSE:
Explore how users perceive and engage with AI-generated content labels on short-form video platforms.

METHODS:
Used eye-tracking, think-aloud protocols, and semi-structured interviews with 4 college-aged participants.

MAIN FINDINGS:

- Labels are often missed unless high-contrast and centrally placed.
- Even when seen, trust doesn't guarantee engagement.
- Visual realism and instinct often outweigh label influence.

TAKEAWAY:

- "Better" labeling means being visible, intuitive, and user-aligned—not just accurate.

DESIGN RECOMMENDATIONS:

- Place labels near captions/usernames
- Use bold contrast and familiar icons
- Include expandable context to explain the AI label

FUTURE RESEARCH:

- Larger, more diverse participant pool
- Real-world testing on live platforms
- Study long-term effects on trust, engagement, and misinformation awareness

REFERENCES

- Fogg, B. J. (2003). Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann.
- Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. CHI.
- Tobii Technology. (2020). Using Eye Tracking in UX Research. Retrieved from www.tobii.com
- Meta (2023). Meta's AI-generated content labeling policy. Retrieved from [www.meta.com/policies]
- TikTok (2024). Community Guidelines and Synthetic Media Labeling. Retrieved from [www.tiktok.com/safety]
- UX Collective. (2022). Why users don't see your labels: A UX case study. [uxdesign.cc]