The Disclosure Dilemma: How AI Attribution Affects Reactions to Public Health Messages

Jacob A. Long, Tabitha Oyewole, Maryam Goli, Jacqueline M. Keisler, Saud Alyaqout, Michael D. Rodgers, and Arielle N'Diaye

University of South Carolina

Background

- Public communicators are increasingly using AI to create messages for efficiency and scale.
- The public is often skeptical of AI, viewing it as less empathetic and trustworthy than human experts in some cases.
- If AI is to be used, how should it be disclosed, if at all?
  - Immediate disclosure risks undermining message credibility
  - Concealed usage risks brand/source reputation

Hypotheses and Research Questions

- Disclosure of AI usage will *reduce* perceptions of message **credibility**, source **trust**, and source **expertise**.
  - Up-front disclosure will *increase* perceived **transparency**
  - Delayed disclosure will *reduce* perceived **transparency**
- Does type of disclosure ("created" vs. "edited") matter?
- Does late disclosure hurt more than early?
- Does disclosure affect **learning** and **information-seeking**?

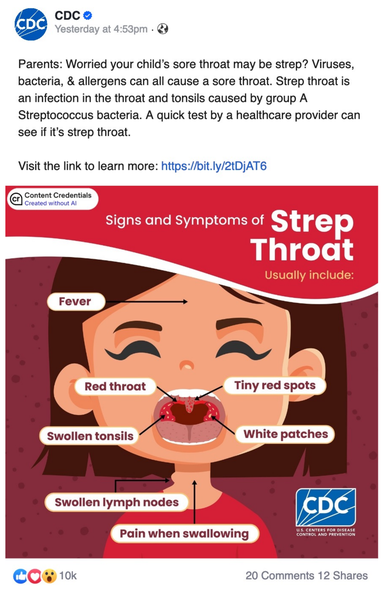
Takeaways

- Ethical disclosure of AI usage undermines message effectiveness:
  - Message is perceived as less credible
  - Recipients learn less from the message (important in public health)
- Later disclosure of AI usage (e.g., via news reporting) hurts the source:
  - Seen as less trustworthy
  - Seen as less transparent than up-front disclosing source
- If AI is used, health communicators must decide whether they will accept reduced effectiveness by disclosing use or risk reputational damage to the institution by concealing it

Methods

- **Design**: An online experiment (N = 1,500 U.S. adults) where participants view four real CDC social media posts, with random assignment of disclosure about AI usage (or not).
- **Conditions**:
  - No Disclosure: (Image with original caption)
  - Denial: "Not created with AI" badge + caption text.
  - AI Generated: "Generated by AI" badge + caption text.
  - Late Disclosure: Participants were told post-exposure that the CDC had been caught using AI without disclosure.
  - If disclosure, manipulate "generated" or "edited" with AI.

Stimuli Examples



Quick Summary

- Does the type, timing, and presence of a *disclosure of AI usage* affect perceptions of public health messages?
- **AI disclosure reduces perceived message credibility, source trustworthiness, and source expertise.**
  - Late disclosure (not included with message) hurts credibility more than disclosure with message.
- Up-front disclosure reduces learning from message
- Trade-off between credibility preservation and effectively transmitting information.

Key Results

Figure 1: AI Disclosure Harms the Message

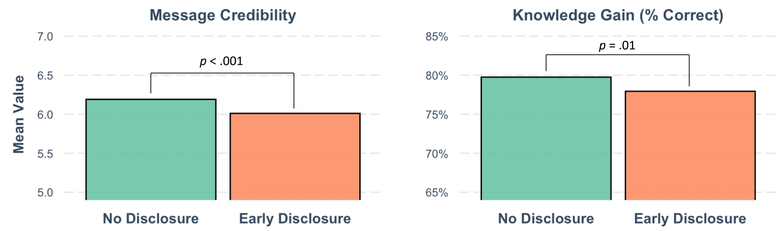
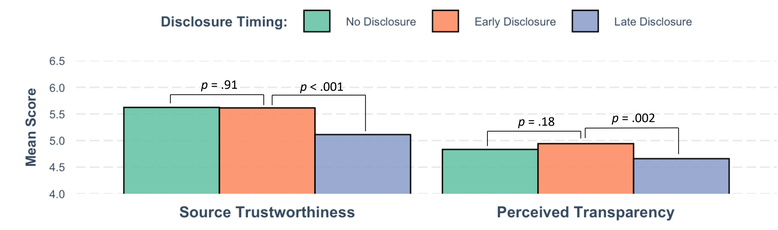


Figure 2: Late Disclosure is Costly to the Source



Key Results

- No difference between "generated" and "edited" wording
- No credibility difference between denial of AI usage vs. no mention (p = .23)
- No effects on *information-seeking* dependent variable

Pre-Registration

Data + Code

Full Paper

UNIVERSITY OF South Carolina