

SNSにおける正確ではないと認識した情報の共有低減策に関する研究

A Study on Interventions to Reduce the Sharing of Information Perceived as Inaccurate on Social Media
 芦田 高穂・ネットワーク分科会・情報セキュリティ大学院大学

This study aims to evaluate an intervention to reduce the sharing of information that social media users recognize as inaccurate. Previous interventions against disinformation have primarily targeted factors such as the belief that fake news is true or inattention to information accuracy. However, interventions specifically addressing the sharing of information that users explicitly recognize as inaccurate have not yet been proposed. This study first focused on the existing explanatory model of such sharing mechanisms and confirmed that users may sometimes share information aligned with their partisanship even when they recognize inaccuracies in the details, as long as they perceive the core claim to be accurate. Subsequently, two interventions were developed by referencing models of psychological mechanisms that justify morally undesirable behavior, and their effectiveness was verified by comparing them with an existing intervention. The results demonstrated that one of these interventions significantly reduced the intention to share information recognized as inaccurate, whereas the existing intervention showed no significant effect. In conclusion, this study demonstrates that interventions addressing the psychological justifications for morally undesirable sharing are effective for suppressing the spread of information that the user recognizes as inaccurate.

1. 研究の背景

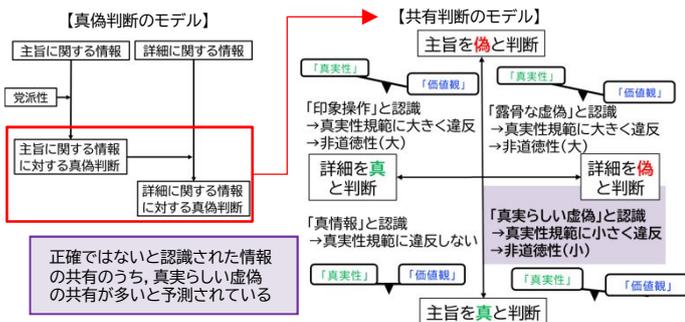
- 偽情報の共有は世界的な問題であると指摘されている。また、社会的に望ましくない行為として認識している人が多い。
- それにもかかわらず、SNSユーザが正確ではないと認識している情報を共有する事例が報告されている。
- しかし、こうした共有の要因を検証する研究やメカニズムに関するモデルを提案する研究はあるものの、介入はまだ検証されていない。

2. 先行研究

- 人は「道徳的に振る舞いたい」という意志を持つものの、従う道徳的規範は一つだけではない(Joynerら,2023)。
- 道徳的に振る舞いたいという意志を持つ人であっても、正確ではないと認識した情報を共有をする。
- 真実性規範違反が常に許容され、価値観に関する規範が常に優先されるわけではない(Langdonら,2024)。



- 正確ではないと認識した情報の共有メカニズムのモデル(Langdonら,2024)



3. 本研究の目的

- 真実らしい虚偽への介入手法を考案し、その有効性を明らかにすること

4. 真実性規範の違反許容を低減する介入の作成

- 先行研究で効果が確認されている、Xのタイムライン上の広告として表示される形態を踏襲
- 道徳的規範違反の許容に関するメカニズムに対する介入手法から、本研究の介入メッセージに応用できるものを選定。

結果の無視に関する介入

比較対象: 正確性プロンプト

「共有しようとしている投稿は正確に感じますか?」

結果の歪曲に関する介入

比較対象: 統制群

「ビジネスチャンスをつかむためのリサーチを依頼します。詳しく見る」

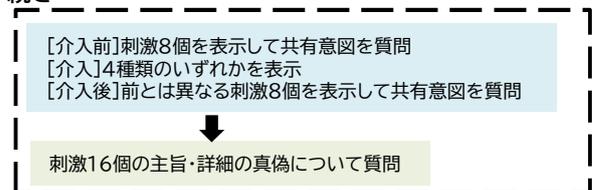
5. 介入の評価

方法

- 800名を対象にWebアンケート調査を実施
- 保守派とリベラル派各400名を4つの介入群に割当て(作成した介入2種, 既存介入, 統制群)
- 刺激:Xに投稿された情報を模した内容とし、「主旨」と「詳細」を設定。刺激例:
 奨学金で苦労している日本人大学生もいるのに、政府は外国人留学生の学費免除を優先しているらしいよ。留学生は申請するだけで学費が免除されて、しかも生活費として35万円以上支給されるみたい。

刺激の政治的傾向	主旨の客観的な真偽	詳細の客観的な真偽
保守/リベラル向け	真	真
	真	偽
	偽	真
	偽	偽

手続き



結果・考察

主旨・詳細の真偽判断による4分類の回答ごと、また、介入群ごとに、介入前後での共有回答数を比較。

[介入前後の共有比較]

条件	結果の無視に関する介入	結果の歪曲に関する介入	正確性プロンプト	統制群
真情報	有意差なし	有意差なし	有意差なし	有意差なし
真実らしい虚偽	前に比べて0.478倍共有	有意差なし	有意差なし	有意差なし
主旨が偽	前に比べて0.564倍共有	有意差なし	前に比べて0.718倍共有	有意差なし

結果の無視に関する介入のみ真実らしい虚偽に効果があった

介入の意図: 真実性規範の重み付けが相対的に低くなり、違反を許容することを良いことだとは思っていない人に、そうした共有について自問自答させる。

→ 真実らしい虚偽の共有および明らかに正確ではないと認識されている「主旨が偽」の共有を抑制させる効果が確認された。

正確性プロンプト

- 明らかに正確ではないと認識される「主旨が偽」の共有抑制には効果がみられた。
- 「真実らしい虚偽」の共有抑制には効果は確認できなかった。

6. まとめ

- 道徳的逸脱行為を低減する介入手法を採用した介入が、真実らしい虚偽と認識した情報の共有を抑制する効果を示した。