Motivation

- In the modern times, information and news travels extremely fast in a heavily interconnected world. With personalized recommender algorithms in social media applications and worldwide internet penetration rising, there has been a rise of misinformation sharing and acceptance, that reinforces existing beliefs and polarises the population.
- In this context, this sharing and acceptance of misinformation can be curbed by prompting the viewer in specific manners that might cause them to re-evaluate their biases and views and might lead to a more cohesive understanding of the world, which leads to social unity
- My motivation was to identify mechanisms and techniques that reduce the spread of misinformation, that might assist in mitigating societal polarization, and foster a more cohesive community where diverse perspectives can be shared and discussed—rather than leaving people divided by divergent, inaccurate beliefs, wherein the chasm is widened by rapid misinformation spread.

Introduction

- This study investigates how four types of prompts (no prompt, generic, source-based,moral/value-based) affect participants' judgments about headlines related to **polarized** (reservation policy, vaccine safety) and **non-polarized** (stock market myths, health/hygiene myths) topics.
- Generic prompts refer to a blanket statement such as "Check the accuracy of each headline" while the other two are tailored prompting techniques that are contextual to the topic/issue being talked about / discussed in the headline. One is prompted around the source that is being used in the headlines, and the other is centered on the morality aspect of the headlines
- Before the main task, participants completed a questionnaire measuring demographics, baseline skepticism, trust in media, and conservative–progressive orientations that roughly maps onto their political ideologies
- During the task, each participant is assigned to **1 of 4** prompt conditions; in a **30–45 min** online task they rate **24** mixed-veracity headlines (6 per topic set, 2 sets are polarised topics, 2 are non-polarised) on belief-accuracy, sharing-intention and confidence. These headlines are a random mix of true and false headlines. Post-task, participants provide feedback on whether the prompt was noticed, how influential it was, and how they perceived it.
- We compare outcomes across conditions, topics, and individual differences (ideology, skepticism, trust) to see which prompts are most effective—and for whom.

Approach

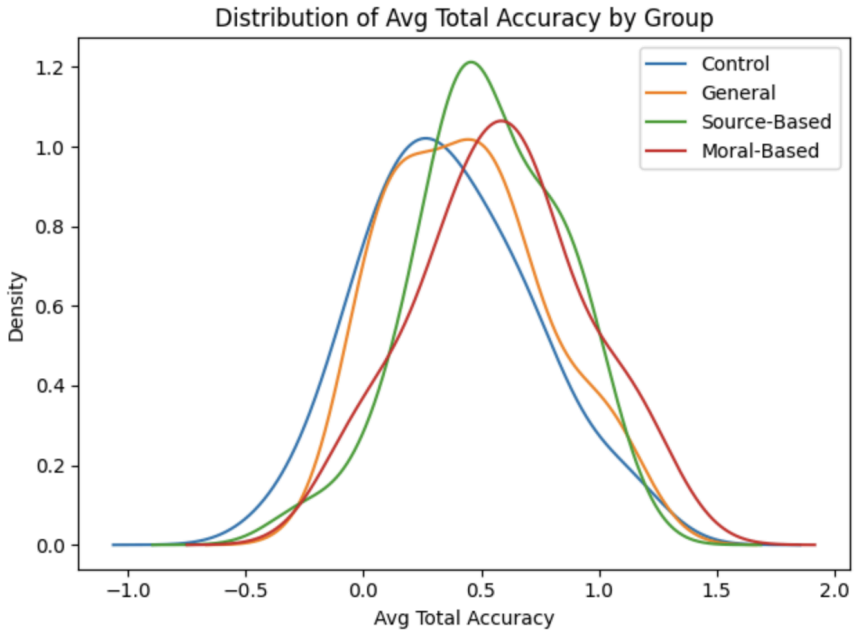
- A detailed pre-task survey and task was designed and created, and the study was conducted with 117 participants, with ~29 participants per group.
- The raw data that was generated from this survey and task was processed and verified statistically whether there was baseline equivalence in pre-task indicators (media trust, skepticism, etc.) among groups
- We also defined total accuracy, true and false accuracy (accuracy measured when only the true/false headlines are considered), polarised/non-polarised accuracy (when only polarised /non-polarised sets are considered), and the means in all the categories.

Analysis & Conclusion

- Skepticism, media trust, and the overall ideology composite all show high p-values (> .05) on both tests—hence, those distributions do not differ across the four conditions (groups) - and can be used as indicators in future analyses
- **Overall accuracy**: Moral-Based prompts shifted the mean from **0.363** (Control) to **0.582** ($t = -2.057$, $p = 0.045$), showing that only moral framing meaningfully improves general headline discernment.
- **Average false accuracy**: Source-Based (mean +0.251 over Control, $t = 2.097$, $p = 0.040$) and Moral-Based (+0.320, $t = 2.884$, $p = 0.006$) both significantly boost people's ability to spot false headlines, whereas General prompts do not ($p = 0.348$).
- **Polarised false accuracy**: Source-Based (0.558 vs. 0.072 Control, $t = 3.956$, $p < 0.001$) and Moral-Based (0.414, $t = 3.288$, $p = 0.002$) deliver large gains on politically charged items which are liable to extreme polarisation, confirming that prompting techniques that simply flag an article's source is the most reliable way to curb misinformation—and moral priming helps too, but generic warnings do not.

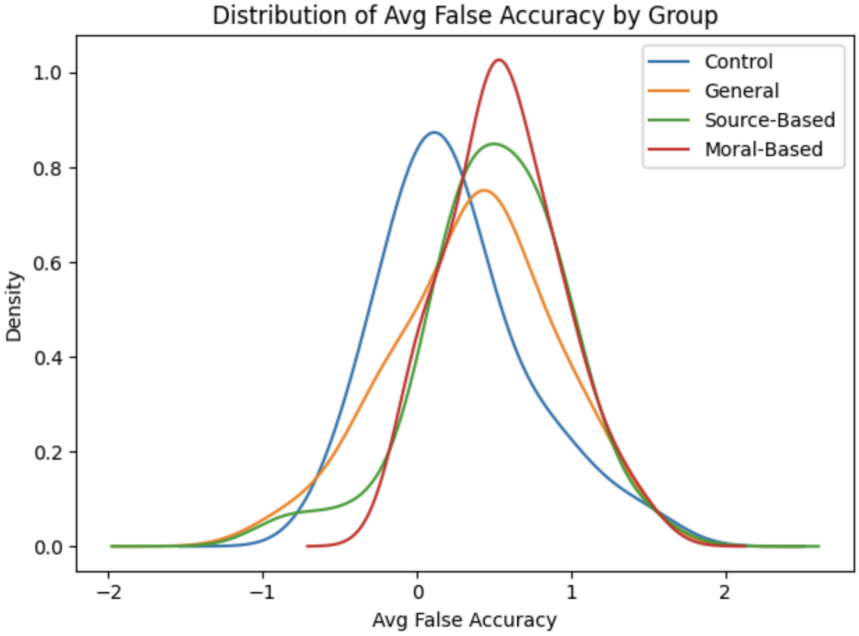
In sum, this study demonstrates that targeted prompts—especially those that flag an article's source or invoke its moral dimensions—can substantially reduce acceptance of false headlines in highly polarized content. Because these interventions are simple, low-cost, and platform-agnostic, newsrooms, social networks, and messaging apps can immediately integrate source- and morality-based prompts to help users spot misinformation before it takes hold.

Also, by showing that tailored nudges work best where ideological stakes are highest, our findings offer a practical blueprint for boosting media literacy, strengthening social resilience to disinformation, and creating safer online communities.



avg_total_accuracy: Levene's W = 0.116, p = 0.734
avg_total_accuracy: Welch's t = -2.057, p = 0.045
→ avg_total_accuracy: significant difference vs. Control (p < 0.05). Pairwise tests:
- General vs Control: t = 0.857, p = 0.395 → General mean is higher by 0.075
- Source-Based vs Control: t = 1.932, p = 0.058 → Source-Based mean is higher by 0.164
- Moral-Based vs Control: t = 2.361, p = 0.022 → Moral-Based mean is higher by 0.219

→ Highest avg_total_accuracy among prompts: Moral-Based (0.582)



avg_false_accuracy: Levene's W = 0.006, p = 0.938
avg_false_accuracy: Welch's t = -2.304, p = 0.026
→ avg_false_accuracy: significant difference vs. Control (p < 0.05). Pairwise tests:
- General vs Control: t = 0.946, p = 0.348 → General mean is higher by 0.120
- Source-Based vs Control: t = 2.097, p = 0.040 → Source-Based mean is higher by 0.251
- Moral-Based vs Control: t = 2.884, p = 0.006 → Moral-Based mean is higher by 0.320

→ Highest avg_false_accuracy among prompts: Moral-Based (0.567)