International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

"I Started to Actually Develop Confidence": A Controlled Usability Study of a Personal Informatics Tool for Supporting Online News Content Discernment Among Young Adults

Prerana Khatiwada, Nabiha Syed, Luke Halko, Ashrey Mahesh, Ricky Kiamilev, Antonia Vazquez & Matthew Louis Mauriello

To cite this article: Prerana Khatiwada, Nabiha Syed, Luke Halko, Ashrey Mahesh, Ricky Kiamilev, Antonia Vazquez & Matthew Louis Mauriello (18 May 2026): "I Started to Actually Develop Confidence": A Controlled Usability Study of a Personal Informatics Tool for Supporting Online News Content Discernment Among Young Adults, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2026.2664696

To link to this article: https://doi.org/10.1080/10447318.2026.2664696

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2026.2664696

"I Started to Actually Develop Confidence": A Controlled Usability Study of a Personal Informatics Tool for Supporting Online News Content Discernment Among Young Adults

Prerana Khatiwada[a], Nabiha Syed[a], Luke Halko[a], Ashrey Mahesh[b], Ricky Kiamilev[c], Antonia Vazquez[d] and Matthew Louis Mauriello[a]

[a]Department of Computer and Information Sciences, University of Delaware, Newark, DE, USA; [b]Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA; [c]Department of Political Science and International Relations, University of Delaware, Newark, DE, USA; [d]Business and Economics, University of Delaware, Newark, DE, USA

ABSTRACT

Readers today face increasing challenges in distinguishing genuine news from misinformation and identifying biased content within algorithmically curated media environments. While prior research has focused on automated detection and adjustments to news recommendation systems, less attention has been given to user-facing interventions that shape how individuals engage with news. To address this gap, we introduce a browser-based personal informatics dashboard designed to support online news consumption by increasing awareness of bias, source quality, and credibility. In a between-subjects study ($n = 100$), our tool improved users' ability to identify false information and encouraged more active engagement with news content ($p < 0.05$). We also observed increased lateral reading behaviors, such as cross-checking sources, suggesting deeper critical engagement. Findings are based on an integrated set of intervention features and combine statistical analyses with exploratory insights into user behavior. By adopting a personal informatics perspective, this work shifts focus from passive misinformation detection toward active, user-centered media literacy and more reflective interaction with news.

KEYWORDS

Misinformation; digital media literacy; news consumption; personal informatics; collaborative and social computing

1. Introduction

In today's digital era, online media platforms have transformed how we access and engage with information (Ahlers, 2006; Galan et al., 2019; Jones & Kang, 2020; Walker & Matsa, 2021). These digital platforms provide unprecedented opportunities for diverse voices to be heard, enabling greater accessibility, real-time updates, and the democratization of content creation (Asad & Parket, 2025; Singh & Singh, 2025). Yet, the ease of online publishing, combined with the speed of virality, makes distinguishing credible journalism from misinformation increasingly difficult (Pennycook & Rand, 2021; Zhang et al., 2018). Misinformation circulates not only via social media but also through independent and hyperpartisan news sites, where users actively consume narratives shaped by sensationalism, partisanship, and selective truth-stretching (Mourão & Robertson, 2019). Reputable outlets such as BBC, CNN, and the New York Times coexist alongside false and fabricated news sources (e.g., abcnews.com.co)[1] and satirical sites like The Onion (Burgers & Brugman, 2022), creating an environment where credibility is often blurred. This complex landscape is further complicated by politically polarized content, where partisan framing shapes news presentation (Feldman, 2011; Garrett, 2009; Stroud, 2010). Even reputable outlets may distort the truth due to biases, fact-checking lapses, or the rush to cover breaking stories (Baly et al., 2018)—sometimes presenting rumors as factual before later correcting or removing them (Allcott & Gentzkow, 2017; Garimella et al., 2021; Jaeger & Taylor, 2021). Collectively, these dynamics erode public confidence in the media (Ognyanova et al., 2020), and the blending of opinion

CONTACT Prerana Khatiwada preranak@udel.edu Department of Computer and Information Sciences, University of Delaware, Newark, DE, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/10447318.2026.2664696.

with news coverage further blurs the line between objective reporting and editorial commentary, shaping readers' perceptions.

This complex information landscape has led to efforts to combat misinformation through automated filtering and downranking (Epstein et al., 2020; Hawa et al., 2021; Sallami et al., 2023), AI-assisted decision aids (Lu et al., 2022), and automated fact-checking systems (Guo et al., 2022; Hassan et al., 2017). However, these solutions typically operate independently of user engagement: while they shape what content is surfaced, they rarely invite users into the process or help them practice evaluating information themselves. Recent IJHCI work further highlights the importance of user involvement, showing that active social media engagement positively predicts fact-checking behavior, while passive engagement may reduce it (Zhou et al., 2025). In this sense, many current approaches still function as black-box algorithms that lack transparency, trust (Saeed & Omlin, 2023) and, crucially, do not foster independent skill-building in evaluating and interpreting news. Other common strategies include attaching warning labels and fact-check tags (Clayton et al., 2020), providing credibility signals such as source ratings (Yaqub et al., 2020), and surfacing transparency cues about sourcing and evidence (Bhuiyan et al., 2021a). These interventions appear as alerts, banners (Avram et al., 2020; Bhuiyan et al., 2021b), shaping how news is perceived and shared. Research shows such cues can modestly reduce belief in false content or dissuade sharing (Clayton et al., 2020; Yaqub et al., 2020). However, these approaches operate as top-down corrections and lack transparency, leaving individuals dependent on platform judgments rather than developing their own evaluation skills.

Yet, as a recent survey of user-centered misinformation interventions by Hartwig et al. (2024) shows, relatively few efforts on news platforms focus on interactive, user-facing supports that actively guide critical engagement while reading. We use the term "*user-facing*" to emphasize this latter category of interventions, and we focus on not only displaying warnings or credibility cues but also on scaffolding reflective practices, helping readers actively assess credibility for themselves. Such approaches align with the broader goals of media literacy (Eisenberg, 2008; Media Literacy Now, 2025), including strategies like lateral reading (Wineburg & McGrew, 2019). In sum, while technological solutions to misinformation have gained traction (Sharma et al., 2019), the challenge lies in bridging the gap between algorithmic detection and user-centered design grounded in behavioral theories (Konstantinou et al., 2025). A deeper understanding of how people interact with online news is essential to inform interventions that move beyond correction toward empowerment.

Drawing on prior work on news informatics (Sundar et al., 2022), we extend personal informatics (PI) concepts—traditionally applied in health, wellbeing, and self-regulation (Epstein et al., 2015b; Kocielnik & Sidorova, 2015; Li et al., 2010a)—to the context of online news, media literacy, and misinformation mitigation. Applying a PI lens to news consumption enables readers to develop ongoing, self-reflective awareness of their information practices. This perspective reframes news discernment not as a one-time correction but as a continuous process of self-monitoring and critical growth. To explore this, we introduce a browser-based PI dashboard and reading intervention that provides a pseudo-real-time experience by simulating the analysis of news articles. With our prototype, we address the following research questions: *(RQ1) How does a PI news-reading tool influence readers' interactions with online news and their confidence in credibility judgments during online news browsing? (RQ2) How does the PI tool shape reading, verification, and information-seeking behaviors? And, (RQ3) which intervention features (e.g., sentiment visualizations, bias indicators) do users find most useful for discerning issues with online news content, and why?*

To answer these questions, we conducted an in-lab study ($n = 100$) using a between-subjects design in which participants read news articles with or without the intervention, followed by debriefing interviews ($n = 56$; 28 per group). We also examined two hypotheses: (*H1*) that exposure to the intervention would increase participants' self-reported confidence in evaluating news credibility and (*H2*) that the intervention would increase engagement, such as scrolling depth, clicking behavior, and information-seeking activities. Our findings suggest that participants who used the tool engaged more actively with the news: they interacted with the articles more frequently, conducted additional web searches, and adopted more critical attitudes toward the content compared to those in the control group. Beyond behavioral engagement, the intervention group also reported greater confidence in judging the

credibility of news articles, identifying key claims, and assessing the quality of sourcing ($p < .05$). Our findings are supported by statistical comparisons where appropriate and complemented by exploratory behavioral and qualitative insights, and should be interpreted as indicative patterns rather than definitive causal effects.

This work contributes (i) a novel in-situ intervention prototype that applies PI principles to news discernment, (ii) findings from a realistic news-reading environment demonstrating its effects on engagement and confidence, and (iii) distills design recommendations that balance usability with behavioral insights to support media literacy during online browsing. While our participant pool was composed largely of individuals aged 18–34, positioning our findings most directly within the context of young adult news consumption, the implications extend more broadly to the design of user-facing interventions.

2. Related work

Here, we review prior work on misinformation, broadly defined as false or partially false information spread intentionally or unintentionally (O'Connor & Weatherall, 2019). We survey its mitigation strategies and discuss broader signals and cues that shape readers' perceptions of news content. In the article, we use the term "discernment" to describe users' ability to critically evaluate and engage with content, including identifying potential bias, misinformation, and emotional manipulation. While Pennycook et al. (2021) defines discernment more narrowly as the difference between sharing mainstream vs. misinformation sources or in accuracy judgments, our study adopts a broader perspective, focusing on critical engagement and self-reported credibility assessment. With this broader conceptualization of discernment in mind, we next situate our work within existing research on misinformation and the range of mitigation approaches proposed in prior literature.

2.1. Background: Misinformation and mitigation approaches

Rapid news cycles increasingly prioritize speed over accuracy, often leading to errors and gradually eroding public trust in the media (Arsenault & Castells, 2006). When fact-checking lags behind reporting, misinformation can spread quickly and widely, whether through deliberate dissemination by biased outlets, hasty reporting for clicks, or amplification across social media platforms (Lazer et al., 2018). National surveys indicate that over 60% of Americans report confusion caused by misinformation, and only a minority is reliably able to distinguish facts from opinions (Barthel et al., 2016; Mitchell et al., 2018). Exposure to misleading information has significant downstream consequences across democratic participation (Iizuka et al., 2022), public health (Bernecker et al., 2021), and national security (Vosoughi et al., 2018). These growing risks have motivated a wide range of approaches and mitigation strategies to address misinformation.

Beyond everyday political misinformation, crises and conflict contexts further amplify these challenges (Patel et al., 2020). During periods such as pandemics or geopolitical conflicts, audiences increasingly rely on online platforms for collective sense-making, often encountering diverse and competing interpretations that can heighten vulnerability to manipulation and information disruption (Schwaderer, 2025). Recent work in crisis informatics and user-centered countermeasures highlights how misinformation spreads rapidly when authoritative information is scarce, underscoring the need for tools that support real-time, user-facing credibility assessment rather than solely retrospective fact-checking (Reuter et al., 2025).

Prior work falls into four broad categories. First, *detection and credibility indicators* embed ratings or bias labels into interfaces (Della Vedova et al., 2018; Gupta et al., 2022). Tools such as *Media Bias/Fact Check*[2] and *Ground News*[3] surface outlet-level judgments or cross-spectrum coverage comparisons. Yet they remain limited: ratings are applied at the *source* rather than the article level. While these approaches surface judgments of content quality, their effectiveness is limited: interventions sit outside everyday news flows and remain static, one-size-fits-all solutions that often fail to build user trust or encourage deeper reflection on news consumption. Second, *cognitive and behavioral interventions*—including inoculation games, reflective nudges, and pre-bunking

strategies—encourage users to pause and evaluate content before sharing (Kozyreva et al., 2024; Pennycook et al., 2020). Related work on accuracy prompts and friction-based interventions demonstrates that brief reminders to consider accuracy or small delays before sharing can reduce misinformation spread at scale without requiring content-level classification (Jahn et al., 2025; Pennycook & Rand, 2022). Recent work further explores behaviorally informed nudges and users' perceptions of them (Konstantinou et al., 2025; Konstantinou & Karapanos, 2025b). While these approaches can improve accuracy in judgment and reduce the sharing of misleading content, their effects may attenuate over time or fail to integrate into habitual browsing contexts. These findings show the promise of lightweight, scalable behavioral nudges, but also highlight limitations in sustaining deeper engagement. Third, *platform-level moderation* strategies such as warning labels, downranking, and account suspensions attempt to curb misinformation's reach (Chowdhury et al., 2022; Myers West, 2018). Although effective at limiting spread, such measures can provoke resistance when perceived as opaque or paternalistic. Finally, *digital literacy interventions* aim to build long-term resilience by strengthening critical reasoning and source evaluation skills (Bronstein et al., 2019; Guess et al., 2020), including game-based tools like *Fakey* (Micallef et al., 2021) and corrective messaging approaches (Vraga et al., 2022).

Collectively, this body of work illustrates a rich design space, yet notable gaps remain. Many interventions operate either at a coarse platform level (platform-level suppression) or as isolated (one-time nudges). Few systems integrate credibility signals, behavioral reflection, and usage analytics within users' natural news-reading workflows. In particular, limited attention has been given to personal informatics approaches that surface patterns in users' own engagement behaviors as a scaffold for reflective evaluation. These gaps motivate the design of our PI tool, which combines credibility indicators with session-level feedback and lateral-reading prompts to support sustained, self-directed discernment.

2.2. Understanding news interaction and evaluation

Understanding how people interact with news is central to designing effective interventions, yet prior work often treats browsing behaviors and credibility judgments as separate strands of research. Within this space, studies of browsing behaviors highlight how people start sessions through search, social media, or direct navigation, and how they curate news diets across domains and topics (Bentley et al., 2019). Other work demonstrates how users manage transitions between background and foreground news consumption in their daily routines (Wirfs-Brock & Quehl, 2019) and how article features, such as sentiment and polarity, influence attention, affect, and engagement in laboratory settings (Arapakis et al., 2014). Complementing this, research has investigated how people consume and verify news on social media (Flintham et al., 2018) and how civic and motivational factors shape engagement (Omar et al., 2023). Taken together, this literature highlights the need to integrate behavioral traces with credibility judgments, since interventions are only effective if they align with how people actually interact with and reflect on news in everyday contexts.

Extending this focus to broader user behaviors, (Grinberg, 2018) explores patterns of news engagement using large-scale client-side interaction data, identifying six reading behaviors and the information gains associated with them. While informative for understanding general trends, this approach overlooks individual user needs and decision processes. One well-studied corrective strategy is *lateral reading*—stepping away from an article to cross-check information with other sources—which has been shown to improve credibility judgments (Wineburg & McGrew, 2019). Yet even here, deeper questions remain about how judgments unfold during active reading, how attention and interaction shape perception, and how to design tools that encourage mindful engagement rather than superficial scanning. Our PI approach extends this line of research by shifting focus from aggregate patterns used in ranking and recommendation toward interventions that help individuals monitor their own behaviors and practice mindful reading through self-observation and control (Lee & Suh, 2022).

2.3. Cognitive style, reading habits, and misinformation

Beyond patterns of browsing and engagement, prior work also highlights how individual cognitive styles and reading habits shape susceptibility to misinformation. A large body of research shows that analytical reasoning is the primary factor that explains why some individuals are better at distinguishing real from false or hyperpartisan news. For example, Pennycook et al. (2018) demonstrates the "*illusory truth effect*," showing that even a single prior exposure to a fake-news headline increases its perceived accuracy, regardless of political ideology or fact-check labels. In related work, Pennycook and Rand (2019) argue that susceptibility to fake news is better explained by "lazy" thinking than by motivated reasoning. Performance on the Cognitive Reflection Test correlates with improved ability to distinguish real from fake news—even when headlines are ideologically consistent. More recent studies extend this to hyperpartisan headlines, finding that analytic thinking generally improves discernment (Ross et al., 2021). Together, these findings suggest that interventions should focus more on fostering reflective and deliberate reasoning.

Building on this insight, Pennycook et al. (2020) tested "accuracy nudges" during the COVID-19 pandemic, finding that simple reminders to consider accuracy significantly improved participants' ability to discern truth from falsehood when choosing what to share online. These findings demonstrate that even lightweight prompts can counteract inattention to accuracy and complement longer-term efforts to build critical thinking skills. Beyond cognitive style, research has linked reading and search behaviors to online reasoning (Coiro & Dobler, 2007). Nagel et al. (2020), for example, found that students who consulted diverse, high-quality sources and reflected on the credibility of those sources demonstrated stronger reasoning. In contrast, Ulu (2019) showed that positive reading attitudes and consistent habits foster metacognitive awareness and critical thinking. Together, this work suggests that misinformation susceptibility is shaped not only by cognition but also by reading and information-seeking behaviors, highlighting the need for interventions that support more reflective news engagement.

2.4. Credibility indicators and provenance in news

Building on insights about cognitive style and reading behaviors, another line of work emphasizes the design of credibility indicators and provenance systems that help readers judge whether information can be trusted. Such cues range from bias ratings and editorial standards to provenance metadata, such as source, website reputation, author, publisher, and content or message itself, which are central to how readers assess trustworthiness (Metzger, 2007). Automated classifiers increasingly incorporate these signals, often achieving better performance when accounting for source bias or origin (Spezzano et al., 2021). Efforts such as the Coalition for Content Provenance and Authenticity (C2PA) (C2PA, 2025) aim to combat misinformation by embedding provenance directly into digital media through open technical standards for media origin, thereby promoting greater transparency. Yet design matters: users prefer explanations that clarify why content is flagged and that support learning, while poorly designed cues risk reactance or backfire (Kirchner & Reuter, 2020; Nyhan & Reifler, 2010). When thoughtfully integrated, provenance systems can enhance trust and accuracy (Feng et al., 2023; Sidnam-Mauch et al., 2022), and users often repurpose platform affordances such as filters or tags for their own evaluation tasks (Jahanbakhsh et al., 2022). These findings demonstrate the need for flexible, user-centered credibility supports.

Related systems further explore how interface-level annotations can surface problematic rhetoric directly within text. For example, systems such as PRTA (Da San Martino et al., 2020) automatically highlight spans associated with specific propaganda techniques and support comparative analysis across articles. While such systems can promote media literacy, they are primarily oriented toward post hoc analysis by researchers, journalists, or analysts at scale (Da San Martino et al., 2020). Our proposed highlight mechanism similarly surfaces localized cues, but differs in that it is embedded within a personal informatics workflow designed for in situ use by everyday readers. Rather than focusing solely on rhetorical technique detection or aggregate analysis, our approach integrates article-level signals with session-level engagement patterns, enabling users to reflect on their own reading behaviors and evaluation processes in real time. Thus, our PI tool combines credibility indicators and provenance features

with reflective supports, aligning technical detection with self-observation and usability principles (Epstein et al., 2015b; Li et al., 2010b).

To operationalize these principles, we next articulate the design rationale as a synthesis of system requirements derived from prior literature, formative insights, and theoretical grounding.

3. Design considerations

This section presents the design considerations underlying the Personal Informatics (PI) tool and explains how they informed the implementation of the dashboard and interaction features within the Chrome extension.

3.1. Design rationale for the personal informatics (PI) tool

The design of the PI tool (Figure 1) draws from media literacy theory (Adjin-Tettey, 2022) to guide user engagement with news, online news consumption patterns, and user perspectives on misinformation interventions. For example, recent work by Khatiwada et al. (2025) highlights the gap between perceived and actual news engagement, as well as users' perspectives on online interventions in news browsing. Using a mixed-methods study design involving surveys, browser-based logging, and debrief interviews, their study found that users often overestimate their news habits and rarely engage in lateral reading. In response to these findings, our tool incorporates lateral reading cues and feedback to promote critical evaluation and self-reflection. Beyond this, Konstantinou and Karapanos (2025a) emphasizes in their scoping review that effective misinformation interventions require grounding in behavioral theory, identifying 17 behavioral objectives and 24 theoretical frameworks that highlight the limitations of simple corrective warnings.

Building on these insights, our tool incorporates in-situ feedback mechanisms and a dashboard that tracks patterns such as reading time, clicks, scrolling patterns, and bias exposure. Complementary systems research highlights the value of real-time analytic tools for revealing media bias. Wang et al. (2025) introduced the Media Bias Detector, which uses large language models to examine coverage

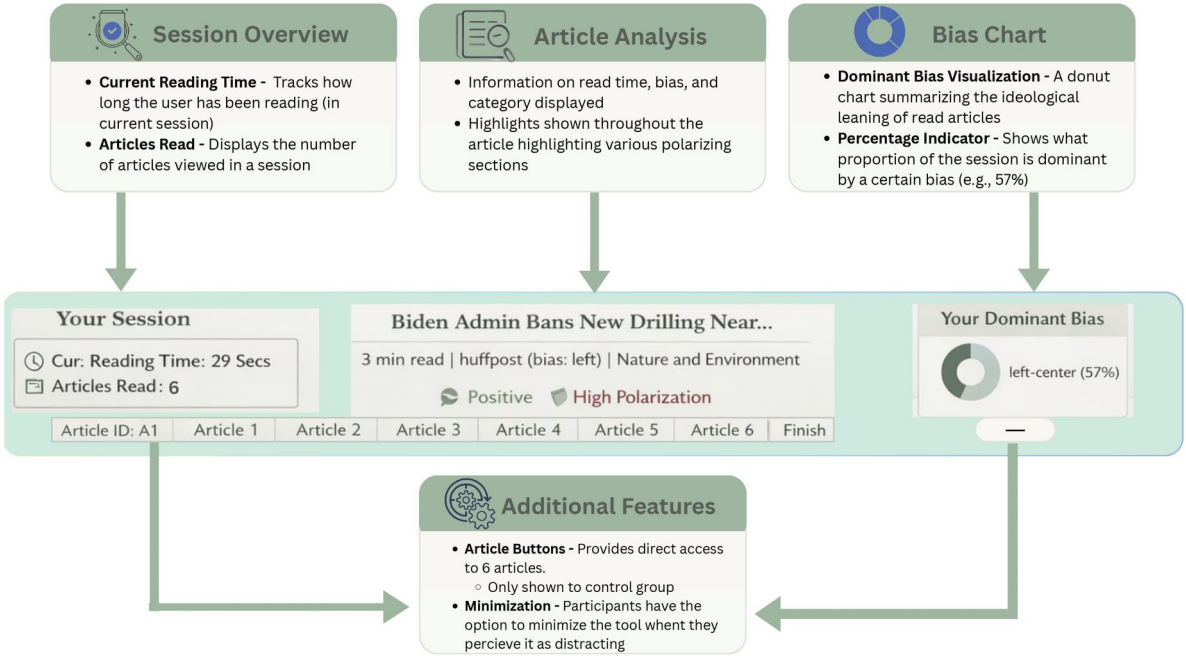


Figure 1. PI tool interface: the interface depicted in this picture represents the complete interface presented to participants in the intervention group (IG). The tool includes features and visual aids to support news consumption and decision-making. The tool is automatically injected whenever a news domain is detected and sits anchored at the top of the news page within the browser. The Control group (CG) viewed the same articles in a text-only format, allowing us to isolate the effect of the dashboard.

volume, political lean, and framing across publishers. Similarly, BIASsist (Noh et al., 2025) demonstrated that explaining and neutralizing bias can enhance users' awareness, detection accuracy, and motivation to critically evaluate news. Together, these works show the importance of integrating literacy theory, behavioral interventions, and real-time analytical affordances into PI tool design. Our tool (Figure 1) builds on these directions by offering a Chrome extension, similar to VanTol et al.'s Open Label (VanTol et al., 2022), and complementary to Trustnet's in-place (Jahanbakhsh & Karger, 2024), follower-based credibility signaling; in contrast, our design centers personal informatics—session metrics and lateral-reading prompts over importing external judgments. Additionally, the Chrome extension that powers our dashboard is inspired by prior workshop work on browser-based interventions (Khatiwada et al., 2022), which explored integrating in-situ annotations, personal informatics, and crowd-powered mechanisms to tackle misinformation online. More broadly, these design elements: dashboard-based reflection, pop-up explanations, are inspired by prior intervention paradigms such as accuracy prompts (Pennycook & Rand, 2022), friction interventions (Jahn et al., 2025), and warning labels (Martel & Rand, 2023) explored in social media and crisis informatics contexts. This suggests the potential applicability of PI-based approaches across diverse platforms, where lightweight, in-situ cues can support critical engagement without disrupting user workflows. Next, we present the PI tool's features in detail.

3.2. Dashboard features

The PI tool was implemented as a Chrome extension that provides an in-situ dashboard. Within the dashboard, participants are presented with three key information areas. The *"Current Session Information"* shows participants their current reading time and the total articles read during the ongoing session. The *"Article Information"* section provides granular data, including the article's title, source, source bias, estimated reading time, article sentiment, and political polarization. Participants can also click article buttons embedded in the dashboard, which link directly to the full stimulus news articles from their original sources. Although the analyses are preprocessed, the tool mimics a real-time experience. To support a natural interaction flow, we included progress spinners to simulate processing time, keeping the interface responsive rather than static. During the debrief, participants were informed that these spinners were UX cues rather than indicators of real-time computational processing. Tooltip icons offer quick explanations of each feature, and article source and bias information are provided by AllSides.[4] The *"Bias Chart"* section further illustrates bias by aggregating the source leanings of articles included in the intervention, thereby showing the dominant perspectives within participants' study-related content exposure. In addition to presenting information, the extension logs participants' browsing data on the backend, including URLs visited, clicks, and dwell time, allowing us to capture behavioral traces of news engagement for later analysis.

3.3. Analysis of news metadata and personalized metrics

To ensure that all metrics described in Section 3.2 operated reliably, we pre-processed and analyzed articles offline before integrating them into the tool, rather than performing computations in real time. For instance, we used the Sentiment Intensity Analyzer from the NLTK library and the VADER Lexicon[5] to evaluate article sentiment. Similarly, the dominant bias chart was derived from the bias labels of the articles users read. It tracked visits to left, left-center, right-center, right, center, or undefined sources. It identified the dominant bias as the category with the highest article count, expressed as a percentage of total articles viewed. If the counts were equal, the bias was marked as none. The reading session time was measured from the moment a user opened an article until they navigated to another, updating every second to accurately capture the total engagement duration. These pre-analyzed metrics were displayed within the dashboard and used to trigger corresponding interventions, with further details provided in Section 3.4.

3.4. News article interventions

The PI tool offers two forms of interventions: (i) pop-up alerts and (ii) passive annotations that provide context-specific feedback through highlights. Pop-up alerts present short, preconfigured messages tailored to the article's content (e.g., propaganda, sensationalism, or opinion-based framing), encouraging users to verify information and distinguish between facts and editorial tone. To enable highlighted annotations (see Figure 2), keywords within news articles are color-coded as users read and scroll through the content. To generate these annotations, we developed a context-sensitive keyword list using a structured, multi-step review process designed to support transparency and replicability. Two trained researchers independently reviewed each article and identified candidate terms that signaled subjective framing, evaluative tone, or emotionally charged phrasing.[6] To avoid over-reliance on isolated terms, coders evaluated each candidate keyword in its surrounding sentence-level context to determine whether it conveyed subjective or opinionated framing. Any discrepancies were resolved through consensus adjudication, with coders comparing interpretations and aligning on how the term functioned within its paragraph context. For example, instead of labeling text as "opinionated" based solely on terms like "by guilting," we examined how the phrase was used within the sentence or paragraph. A third researcher with domain expertise in media literacy reviewed and validated the final keyword selections to ensure conceptual consistency with framing scholarship. Given the limited number of articles and focused set of candidate keywords, formal IRR metrics were not appropriate. Overall, the development process drew on guidance from UVic Libraries' Fake News typologies[7] and linguistic characterizations of deceptive discourse in *The Language of Fake News* (Grieve & Woodfield, 2023).

Once finalized, the keyword set (summarized in Table 1) was implemented using a consistent annotation template. To ensure intervention transparency, each keyword was assigned to a predefined intervention category (e.g., Opinionated, Speculative, Unsupported Claim, Misleading Reference) based on its contextual function within the article. Categories were mutually exclusive at the assignment stage to preserve interpretive consistency. Each category was paired with a standardized message template designed to prompt reflection rather than assert correction (e.g., "this phrasing may signal evaluative framing" or "this reference lacks a clearly identifiable source"). For example, evaluative descriptors such as "notorious" or "gaffes" were mapped to the Opinionated category because they introduce value-laden framing without necessarily providing counterbalancing context. Terms such as "theory" were mapped to Speculative, signaling that the surrounding claim may reflect conjecture rather than verified reporting. Phrases like "law enforcement official" or institutional references such as "White House" were mapped to Misleading Reference when the article did not provide attribution to a specific, verifiable source (e.g., named spokesperson, official statement, or documented report), consistent with research on attribution ambiguity in misinformation and fake news (Grieve & Woodfield, 2023). In cases involving named entities (e.g., references to unfamiliar outlets such as "Dunning-Kruger-Times.com"), the

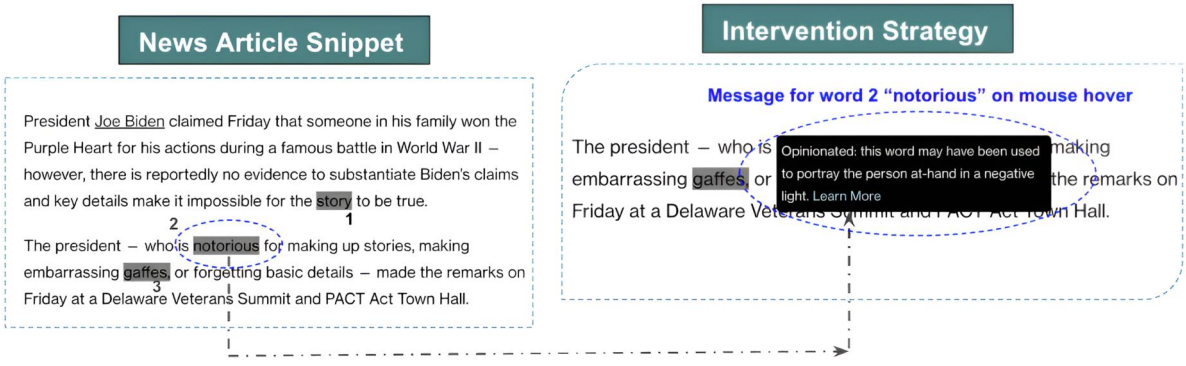


Figure 2. The figure displays a snippet of a news article on the left, with certain keywords highlighted in grey. These keywords carry strong emotional connotations or are biased and charged terms. When users hover over these keywords, a message box appears, as shown on the right side of the figure. For example, the word "notorious" is labeled as opinionated, indicating that it may have been used to portray the individual negatively. Similarly, other keywords are accompanied by corresponding messages.

Table 1. Highlighting intervention messages | displayed across all the false articles.

Articles keyword	Displayed message
"notorious"	Opinionated: this word may have been used to portray the person at-hand in a negative light.
"story"	Opinionated: there is no mention of counterarguments or context that could serve as potential clarifications for the statements made by the person under question.
"report"	Unrelated: this portion of the article attempts to create an association of issues that may or may not exist to portray the person at-hand in a negative light.
"theory"	Speculative: keep in mind that the following information may be conjecture.
"law enforcement official"	Misleading: a vague reference that lacks ties to a relevant named entity or agency.
"white house"	Misleading: although the source references this entity, this article is not tied to an official statement from this entity.
"Natasha"	Before Reading: "Dunning-Kruger-Times.com" is not a well-known or established news source.
"by guilting"	Opinionated: the article does not present both sides of the story or provide any input from Miss USA or official representatives.
"As they always do"	Unsupported Claim: The article does not provide evidence or sources to back up its claims.
"gaffes"	Opinionated: this word may have been used to portray the person at hand in a negative light.

Note: The keywords and messages are derived from their specific use in the three analyzed fake articles. They serve as examples, and actual user-facing messages will provide clearer context when presented alongside the full article content.

Figure 3. Participants clicked the link provided by our highlighting intervention. The (left image) showed a summary generated by our team, while the (right image) presented direct information from factcheck.org. Article title *"Biden claims his uncle won purple heart during world war II battle, but details make Biden's story impossible"*.

annotation triggered a source-level credibility cue rather than a content-level framing message. These assignments were guided by media literacy frameworks emphasizing source evaluation, attribution clarity, and evidentiary support as core verification strategies (Russo et al., 2019; Wineburg & McGrew, 2019). The intervention messages were intentionally phrased as prompts for reflection rather than definitive judgments to avoid overclaiming and to support user agency.

When participants hover over these keywords, small message boxes appear with links to fact-checking websites like FactCheck.org[8] for further verification and context. The highlighted terms were also embedded with links to sample resource pages (implemented in Google Docs) that provided concise article summaries. Figure 3 illustrates the external resources accessed through the "Learn More" option within these annotation pop-ups. This process ensured that the intervention systematically drew attention to specific words and phrases with strong emotional or subjective connotations, while also providing participants with immediate access to credible verification resources.

The PI tool displays the true articles (*verified to be true)* without any of these interventions while flagging false articles with these interventions. We identified true articles based on verification from trusted fact-checking sources such as Snopes,[9] PolitiFact.[10] It is important to note that the articles were not modified in any way to alter their accuracy. Articles classified as false were selected from the same fact-checking resources and from additional sources known for publishing misleading or fake news. Each false article was further examined against a Fake News Checklist[11] to confirm indicators of misinformation, such as unverified claims and misleading framing. Further details on reading materials are provided in Section 4.1. We now turn to the study design, outlining the experimental setup, procedures, and measures.

4. Study design and method

We conducted a between-subjects in-lab study ($n = 100$) to evaluate the effects of our PI tool. Participants were randomly assigned to one of two conditions. In the *Intervention Group (IG)*, participants used the tool with full dashboard features enabled. In the *Control Group (CG)*, participants had access to the tool but did not see any dashboard features beyond the standard browser interface and article buttons; however, their interactions were still recorded for analysis.

Participants were recruited through class announcements and social media platforms, including LinkedIn, Twitter/X, Slack, and Discord. Eligible participants were 18 years or older, U.S. residents, proficient in English, and regular readers of online news on desktop computers. The sample included undergraduate and graduate students, university staff, and a few local community members outside the institution. In total, 100 participants were recruited: 40 between July 24 and August 20, 2023 (fewer during the summer break), and 60 between September and December 2023 (during the semester). Assignment to conditions was kept roughly equal (see Section 4.2 for details on participant grouping). Although individual study sessions lasted approximately 45–60 min, data collection was spread across six months due to the need to stagger sessions around participants' availability, semester breaks, and scheduling constraints common in in-person recruitment.

To estimate an appropriate sample size, we conducted a power analysis using G*Power (Kang, 2021). For a two-tailed independent-samples t-test, the analysis indicated that a minimum of 128 participants would be needed to achieve statistical power of 0.80 at an alpha level of 0.05 with a medium effect size ($d = 0.5$). These parameter choices align with standard conventions for social and behavioral research (Cohen, 2013; Faul et al., 2007; Kang, 2021). Due to practical constraints such as time and participant availability, we recruited 100 participants. While this falls short of the ideal sample size for maximal power, it remains sufficient for exploratory research focused on identifying general trends and patterns.

Participants first completed an interest form (see Supplemental Note 7), which helped balance demographics across gender, race/ethnicity, and political identity, and then self-scheduled their sessions via Calendly. Each session lasted approximately 45–60 min, consisting of a 35-minute reading and survey task followed by an optional 10-minute debrief interview. Participants received proportional compensation ($15 for a typical session) via digital Amazon gift card or direct deposit, depending on university payroll status. This study was approved by the University of Delaware Institutional Review Board (Protocol #1871618-8).

4.1. Study materials

In Spring 2022, we conducted a pilot study with 10 participants to test the initial version of our tool. Participants responded positively, finding it useful for reflection, appreciating its transparency features, and expressing interest in real-world use. However, including 10 news articles led to fatigue and extended study time. Following recommendations for transparent pilot reporting (Oppenlaender et al., 2024), we adjusted the final study to include six diverse, recent articles drawn from outlets such as HuffPost, CBS News, The Free Beacon, The Daily Wire, NBC, and Dunning-Kruger News. The selected articles reflected the active news cycle at the time of the study. We controlled for article length and image count, though some later incorporated external videos due to ongoing updates. Of the six articles, three were classified as true and accurate, and three as fake, containing false, misleading, or unsubstantiated claims. The latter were sourced from fact-checking sites such as Snopes, as well as fake-news and satirical platforms, ensuring authentic examples of misinformation. Consistent with prior work (Ma et al., 2025), this balanced selection of verified and misleading content across diverse topics helped prevent bias and supported a realistic examination of fake-news dynamics. Table 2 summarizes six news articles and their distribution. These materials formed the basis of the lab study, supporting a controlled and realistic evaluation of participant responses.

4.2. Research execution

Participants were assigned to the IG or CG using simple random assignment at the time of enrollment. No blocking, stratification, or demographic controls were applied during allocation, although overall

Table 2. Summary of news articles used in the study.

Article ID	News Article title	News source	Type	Sentiment	Polarization	Category
A1	Biden Administration Bans New Oil Drilling Near "Irreplaceable" Tribal Cultural Site In New Mexico	HuffPost	True	Positive	Highly polarized	Nature and environment
A2	Miss USA Boycotts Miss Universe Pageant—"I am not competing against a man"	Dunning Kruger	Fake	Positive	Highly polarized	Entertainment
A3	I Forced A Bot to Write an Amy Coney Barrett Story Based on Media Reporting and This Was the Result	Free Beacon	True	Negative	Highly polarized	Humour and satire
A4	Substance found in White House confirmed to be cocaine	CBS News	Fake	Positive	Highly polarized	Politics
A5	At least 17 mass shootings mark Fourth of July holiday weekend, data shows	NBC	True	Negative	Less polarized	Current events
A6	Biden Claims His Uncle Won Purple Heart During World War II Battle, But Details Make Biden's Story Impossible	Daily Wire	Fake	Negative	Less polarized	Politics

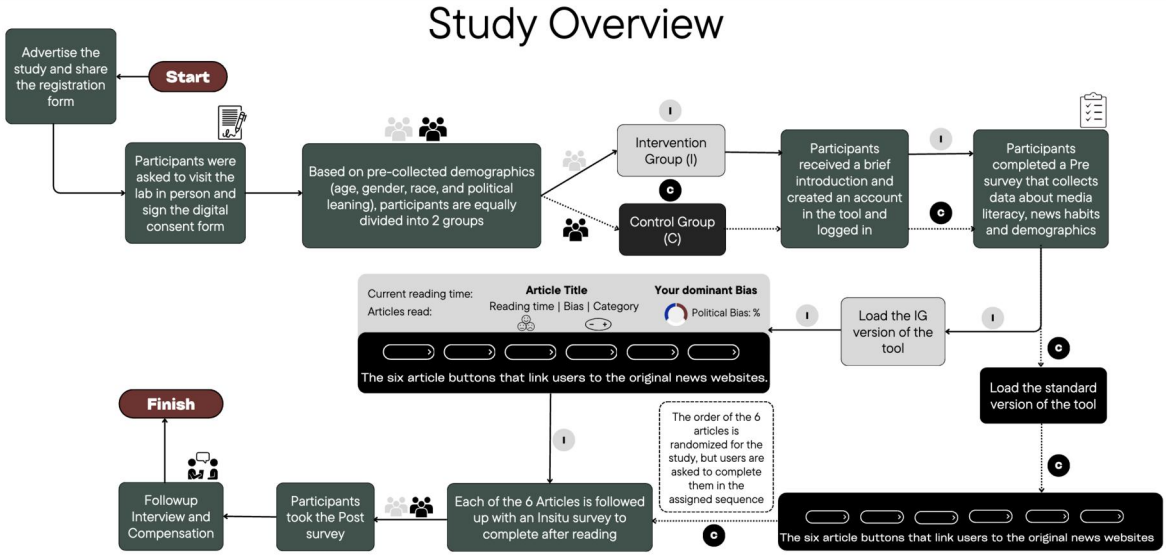


Figure 4. Overall study design and methodology. The solid-line path represents the study flow for the intervention group (IG), while the dotted-line path represents the control group (CG). All steps in the procedure were identical across groups—participants were recruited, provided consent, completed a pre-survey, read six randomized articles (each followed by an in-situ survey), and then completed a post-survey and follow-up interview. The only difference was that IG participants interacted with the intervention-enhanced version of the tool, whereas CG participants used the standard version.

condition counts were monitored to maintain approximately equal group sizes (n = 50 per condition). This study is exploratory in nature, and demographic characteristics (e.g., age, gender, race, political leaning) were broadly similar across groups, though exact distributions were not controlled. Recruitment followed a rolling, semester-based participation model, which limited opportunities for fine-grained stratification; additionally, some characteristics (e.g., political leaning) were self-reported and unevenly distributed within the participant pool. The study took place in a lab using a dual-screen setup: one for news interaction and one for surveys. Sessions lasted approximately one hour; see Figure 4 for the overall study flow.

Participants first received a brief introduction (see Moderator Guide in Supplemental Note 1) and completed a digital consent form, with the option to ask questions or withdraw at any time. To minimize demand characteristics and behavioral bias, the study's specific purpose was withheld until debriefing; participants were informed only that the session focused broadly on news-reading behavior. In line with standard ethical practices, all participants were fully debriefed afterward about the tool's actual

purpose. Although conducted in person, the study used Zoom screen sharing—with participants' consent—to record their interactions with news articles for later analysis.

All participants then completed a pre-study survey in Qualtrics[12] to assess their attitudes and opinions on news media literacy and information literacy topics using a standardized scale (Boh Podgornik et al., 2016; Vraga et al., 2015). The pre-survey also included four items assessing participants' proficiency and confidence in evaluating news articles, which were repeated in the post-survey to measure changes over time and assess the study's before-and-after effects (see Section 5.4). After the pre-survey, participants registered for our tool using their study registration email and accessed the tool's reading interface upon logging in. Both groups received similar instructions, with the IG additionally receiving an explanation of the tool. The control condition was designed to approximate standard online article reading without dashboard and prompts, thereby isolating the presence of intervention features rather than introducing an alternative interface paradigm. Both conditions used the same article content, screen setup, and reading workflow, with the primary difference being the activation of alerts, highlights, and dashboard feedback in the intervention group. While differences in interactivity may independently influence engagement, a well-documented consideration in interface and persuasive system design research (Munzner, 2025), the text-only condition reflects typical news consumption environments in which no reflective cues are embedded. We therefore interpret differences as emerging from the bundled intervention features, while acknowledging that UI-level effects cannot be fully disentangled in the present design.

To minimize bias, we maintained response anonymity and framed survey questions around participants' experiences rather than direct tool evaluations. IG instructions framed the tool as part of a general reading task, without explicitly linking it to the experiment's main goal. Similarly, for the group without the tool, we designed an equivalent reading task that did not mention the tool, ensuring they did not perceive any intentional difference between the conditions. Next, all participants were asked to read six articles and were encouraged to interact with them as they would while browsing news online normally.

The researcher stated: [….*"During the article reading task, there are no right or wrong ways to engage with the content. We want you to feel relaxed, comfortable, and at ease. Treat this like any other day when you consume news. You can read the articles in any way that feels comfortable to you, as if you were casually browsing through news stories as part of your daily routine. Read at your own pace, scroll as you like and in whatever way feels natural to you"*….].

Participants read articles in random order, randomized by the tool to minimize order effects, with interventions provided as needed. A single researcher facilitated the study, staying in the room but avoiding interaction to minimize influence and allow independent engagement. Upon completing each reading, participants completed brief in-situ surveys on credibility perception, sharing intention, and fact-checking. As an attention check, participants entered the article ID (A1–A6) shown on the tool into the survey, which also facilitated later article-level tracking and analysis. After reading, participants completed an exit survey assessing their confidence and skills gained, and provided open-ended feedback. The full questionnaire is in Supplemental Note 8.

4.2.1. Interview procedure

We conducted 56 debrief interviews (28 per condition), representing approximately half of the total sample. Participants who had indicated "yes" for interview participation on the Register/Intake form were invited, and their willingness was reconfirmed at the end of the reading task during the lab session; almost all who initially indicated interest agreed to participate. Thus, interviews were voluntary and conducted immediately following the lab session. Each 15–20-minute session invited participants to reflect on their experience and envision future systems to support news discernment, helping contextualize the quantitative findings. Interviews were recorded, transcribed via Otter.ai,[13] and backed up on Zoom, with the researcher also taking field notes.

Recruiting a substantial subset from both conditions allowed us to capture a range of perspectives while remaining feasible within scheduling constraints. However, because participation was opt-in, the qualitative sample may reflect self-selection bias (Collier & Mahoney, 1996), as individuals with

stronger reactions or higher engagement may have been more likely to volunteer. Additionally, interview responses may be shaped by interpersonal dynamics and reflection time inherent to qualitative settings (Downey, 2015), which can influence how participants articulate their experiences.

Inspired by Braun and Clarke's thematic analysis approach (Braun & Clarke, 2012), three raters collaboratively developed the codebook. Inter-rater reliability (IRR) was assessed using Cohen's kappa (Hallgren, 2012; Skillman et al., 2019). After multiple rounds of discussion and refinement to clarify code definitions and resolve discrepancies, the final IRR reached $\kappa = 0.79$, indicating near-perfect agreement (Campbell et al., 2013). Building on this consensus, the final themes are presented in Sections 5.10 and 5.11. Interview questionnaires and the codebook are provided in Supplemental Notes 2 and 3. Further details on our analysis approach across data sources are provided in the following section.

4.3. Analysis approach

Our analyses draw on three data sources, each aligned with specific research questions. First, pre- and post-survey responses were analyzed to assess changes in participants' confidence and credibility judgments (RQ1/H1). Second, log data from the browsing task (e.g., reading time, scrolling depth, clicks, searches) were examined to capture behavioral differences across conditions (RQ2/H2). Third, interview transcripts were thematically coded to surface user perceptions of the intervention, desired features, and design considerations (RQ3). Given the exploratory nature of the study, we interpret results as directional rather than confirmatory. Analyses aligned with RQ1–RQ3 (credibility judgments: Section 5.3, confidence: Section 5.4, behavioral engagement metrics: Section 5.5, and qualitative findings from interviews: Section 5.10 and Section 5.11) are treated as primary outcomes, while additional comparisons (e.g., tool satisfaction: Section 5.7, video analysis: Section 5.8, and additional articles visited: Section 5.9) are considered secondary outcomes that further complement the primary insights.

5. Results

Our results are structured as five key components: (i) pre-survey analysis of participants' news habits, (ii) in-situ survey analysis of article perceptions, (iii) post-survey comparisons of confidence in article evaluation, (iv) analysis of behavioral data and news interaction metrics, and (v) follow-up interviews from both groups providing qualitative feedback on the intervention strategies. Before presenting these findings, we first report the demographic results of our sample.

5.1. Participants demographics

Participants' ages ranged from 18–70 years, with a median of 22.5; IG: $M = 24.74$, $SD = 6.93$, CG: $M = 24.1$, $SD = 7.93$. While the articles primarily focused on U.S. politics and news, participant recruitment was determined by the study's location in the United States rather than by the articles' content. See Tables 3 and 4, as well as Supplementary Note 4, for additional details such as employment and household income.

5.2. Pre-survey

Pre-survey results provide context on participants' baseline news consumption habits and perceptions. We found that 42% of participants consumed news multiple times a day, with 36% spending 15–30 min on it. Social media and messaging apps (21%) and online news sites (19%) were the main sources, with the New York Times most frequently visited (20%). About 31% reported encountering biased or false news. Credibility perceptions were shaped primarily by trust in the warning source (21%), while few (6%) avoided sites with warnings. Both groups showed similar levels of media literacy (CG: $M = 6.08$, $SD = 0.43$; IG: $M = 6.14$, $SD = 0.48$) and information literacy (CG: $M = 0.53$, $SD = 0.22$; IG: $M = 0.57$, $SD = 0.23$).

Table 3. Participant demographics by age, gender, and race (counts and within-group percentages).

Age	CG	IG	Gender	CG	IG	Race	CG	IG
18–24	32 (64%)	29 (58%)	Female	31 (62%)	27 (54%)	Asian	17 (34%)	19 (38%)
25–34	17 (34%)	17 (34%)	Male	18 (36%)	19 (38%)	White	24 (48%)	21 (42%)
35–44	0 (0%)	3 (6%)	Non-binary	1 (2%)	3 (6%)	Black/African American	4 (8%)	4 (8%)
45–54	0 (0%)	1 (2%)	Prefer not to say	0 (0%)	1 (2%)	Hispanic/Latino	1 (2%)	3 (6%)
55+	1 (2%)	0 (0%)				Other/prefer not to say	4 (8%)	3 (6%)

Note: CG = Control group; IG = Intervention group; $n = 50$ per group.

Table 4. Participant demographics by education and political orientation (counts and within-group percentages).

Education	CG	IG	Political leaning	CG	IG
High school	8 (16%)	9 (18%)	Democrat	25 (50%)	25 (50%)
Some college	15 (30%)	15 (30%)	Independent	20 (40%)	21 (42%)
Bachelor's	14 (28%)	8 (16%)	Republican	3 (6%)	2 (4%)
Master's	12 (24%)	15 (30%)	Other	2 (4%)	2 (4%)
Doctoral/professional	1 (2%)	3 (6%)			

Note: CG = Control group; IG = Intervention group; $n = 50$ per group.

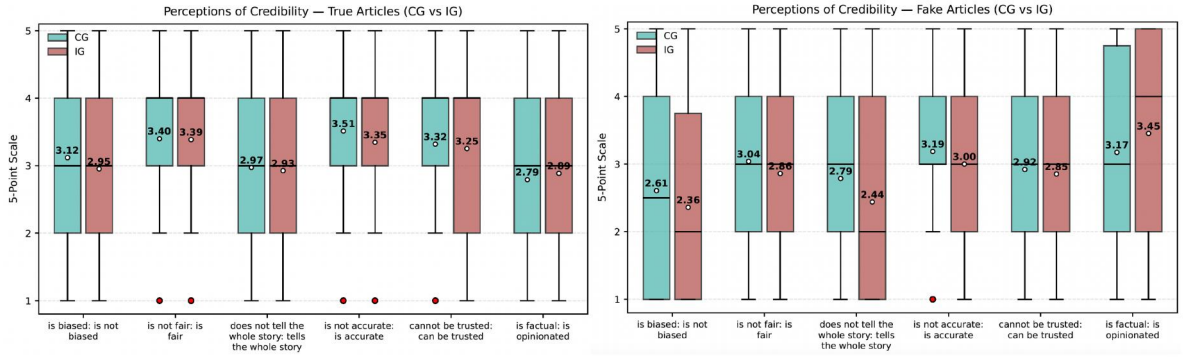


Figure 5. Perceived credibility of true and fake articles across six dimensions (bias, fairness, completeness, accuracy, trustworthiness, factuality). Boxplots show response distributions (1–5 Likert scale) for control (CG, teal) and intervention (IG, red) groups, with mean values labeled. For true articles, IG ratings were generally similar to CG with small shifts in fairness and accuracy. For fake articles, IG participants tended to give slightly lower credibility scores on bias and completeness but higher ratings on factuality, suggesting that the intervention influenced how participants judged deceptive content.

5.3. In situ surveys

In situ surveys, administered as brief check-ins after each article help us understand how the study influenced participants' engagement, actions (e.g., fact-checking), and credibility perceptions in both groups.

5.3.1. Perception on credibility

To assess evaluative behavior, participants rated articles using the six-item Gaziano's News Source Credibility Scale (Gaziano & McGrath, 1986). These ratings reliably measure how the tool may influence perceptions of news credibility and the suitability of an article for consumption or reference. Overall, the control group (CG) rated all six articles slightly higher in credibility ($M = 3.07$, $SD = 1.18$) than the intervention group (IG; $M = 2.92$, $SD = 1.20$). This pattern was also observed for false articles, where CG assigned higher credibility scores than IG ($M_{CG} = 2.95$, $SD_{CG} = 1.19$ vs. $M_{IG} = 2.68$, $SD_{IG} = 1.16$). While participants in this group may have some inherent ability to recognize truth, they lack the skills or resources necessary to detect falsehoods effectively. Thus, CG participants have difficulty discerning falsehoods. Within IG, false articles were rated significantly lower than true articles ($M_{false} = 2.68$, $SD = 1.16$ vs. $M_{true} = 3.16$, $SD = 1.18$, $p < .001$), suggesting improved discrimination between true and false content. Figure 5 illustrates these differences.

To examine whether credibility ratings varied across dimensions and differed by group, we also conducted a mixed-model ANOVA with credibility dimension (six-item scale; repeated measures) and

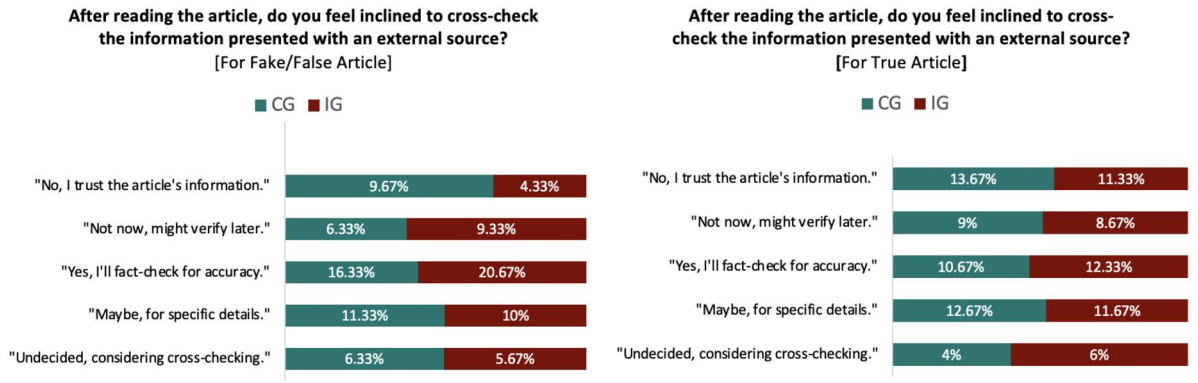


Figure 6. Exploring the impact of the tool on fact checking. The plots compare control and intervention groups across fake and true articles, showing differences in participants' inclination to cross-check information with external sources.

group (CG vs. IG). For true articles, a significant main effect of credibility dimension was observed, $p < .001, \eta_p^2 = 0.04$, indicating that participants differentiated across credibility attributes. Bonferroni-corrected post-hoc tests showed significant differences across several dimensions (all $p < .01$). For false articles, a significant main effect of credibility dimension was observed, $p < .001, \eta_p^2 = 0.06$. A small but significant interaction between group and credibility dimension was also found, $p = .03, \eta_p^2 = 0.01$, suggesting slight differences in how groups evaluated credibility attributes. Post-hoc tests (Bonferroni-corrected) indicated significant differences across dimensions (all $p < .01$).

5.3.2. Cross-checking habits

We then examined whether participants were inclined to cross-check the information after reading and whether there were any discernible group differences in this regard due to the additional information provided by the tool. We associated these responses with observed behavior using video footage and any external links recorded to cross-check (Section 5.8), excluding the six main visited articles. Participants in both CG and IG trusted "true articles," but IG showed slightly lower trust (11.33% vs. 13.67%). IG participants were more likely to fact-check both true (12.33% vs. 10.67%) and fake articles (20.67% vs. 16.33%). Notably, fewer IG participants trusted "fake articles" (4.33% vs. 9.67% in CG). Thus, participants exposed to the tool were less likely to accept the information in fake articles at face value and were more inclined to verify its accuracy. See Figure 6 for the percentage of responses distribution. To further validate these findings, a χ^2 test was performed between "Study Group" and the response to the variable "*After reading the article, do you feel inclined to cross-check the information presented with any external sources?*" We found participants' inclination to cross-check information after reading true articles did not differ significantly between CG and IG, ($X^2(4) = 2.36$, $p = .67$, *Cramer's* $V = 0.09$). However, for false articles, there is a significant difference in the inclination to cross-check information, ($X^2(4) = 9.7$, $p = .046$, *Cramér's* $V = 0.18$), with IG demonstrating a stronger inclination toward cross-checking.

Lastly, all participants reported their subjective experiences and intentions while engaging with news, answering the question: *"On a scale of 1 to 5, please indicate your level of agreement."* Figure 7 indicates, IG showed slightly higher scores than CG for information recall ($M_{IG} = 3.39$ vs. $M_{CG} = 3.23$) and engagement ($M_{IG} = 3.33$ vs. $M_{CG} = 3.25$), though these differences were not statistically significant (recall: $p = .126$; engagement: $p = .393$) However, IG reported significantly greater intention to seek further information ($M_{IG} = 3.02$, $SD = 1.34$) compared to CG ($M_{CG} = 2.79$, $SD = 1.33$), $p = .033$, $d = 0.17$, 95% CI $[-0.45, -0.02]$, indicating a small effect. Results further confirmed a significant main effect of group on information-seeking intentions ($p = .033$), with no effect of truthfulness and no interaction. In this case, seeking further information includes a broader range of investigative activities beyond just the immediate verification of facts during a reading session.

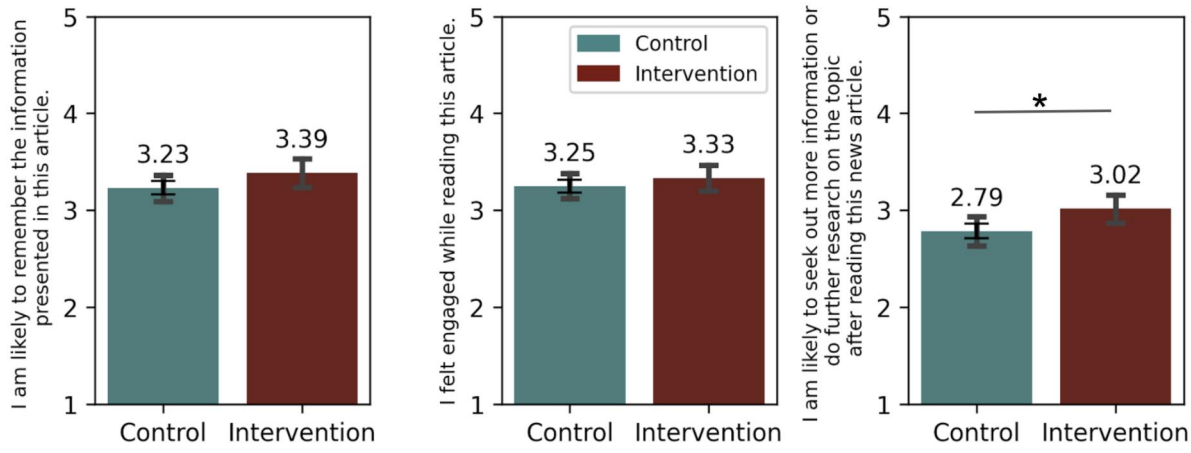


Figure 7. Self-reported Overall engagement with six news articles in the study.

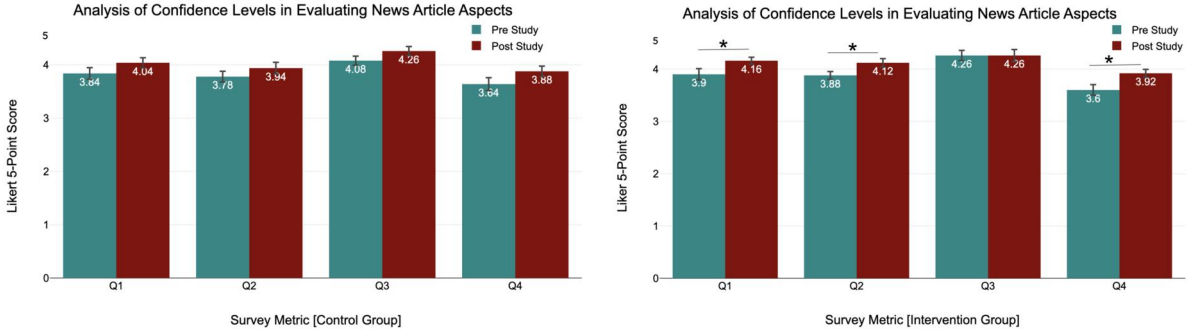


Figure 8. Comparison of self-evaluation in confidence levels: Pre- and post-survey ratings of confidence in news literacy skills. Participants rated their confidence on four statements: Q1: I am confident in my ability to evaluate the credibility and accuracy of news articles; Q2: I am confident in my ability to distinguish claims within a news article; Q3: I feel confident in my ability to perceive potential bias and identify inflammatory language in a news article; Q4: I am confident in my ability to assess the quality of sourcing in a news article.

5.4. Post survey

Here, participants reevaluated their confidence in news literacy skills on a 5-point scale using the same set of statements from the pre-survey (see Figure 8 for these dimensions and metrics). Together with the Shapiro-Wilk test values, skewness, kurtosis (Hatem et al., 2022), and the qqplot, we found the data to be approximately normally distributed. Levene's test was used to assess variance equality to determine the appropriate t-value (Gastwirth et al., 2009). We treated the four metrics as separate dependent variables and ran individual t-tests for each to provide a focused, precise analysis of their effects. Where appropriate, we additionally conducted post-hoc comparisons with multiple-comparisons correction to account for related tests.

5.4.1. Results for control group

Results indicate Pre-study values were lower than post-study for all dependent variables (*Q1–Q4*)—see Figure 8. Still, differences in CG were not statistically significant ($p > 0.05$), as evidenced by the paired t-test. Thus, our study did not significantly affect CG who were not exposed to the dashboard (see Table 5 for detailed statistical reporting of metrics).

5.4.2. Results for intervention group

IG showed significant improvements in Q1, Q2, and Q4 ($p < .05$), with small-to-moderate effect sizes (Q1: $d = 0.41$, 95% CI $[-0.51, -0.01]$; Q2: $d = 0.44$, 95% CI $[-0.45, -0.03]$; Q4: $d = 0.50$, 95% CI $[-0.57, -0.07]$; see Figure 8). Bonferroni-corrected comparisons confirmed a significant overall

Table 5. Descriptive statistics and t-test results.

Var.	Group	Pre-confidence score (M, SD)	Post-confidence score (M, SD)	t	p	95% CI	Effect size
Q1	CG	3.84, 0.77	4.04, 0.67	−1.39	0.167	[−0.49, 0.09]	0.28
Q2	CG	3.78, 0.71	3.94, 0.77	−1.08	0.281	[−0.45, 0.13]	0.22
Q3	CG	4.08, 0.6	4.26, 0.6	−1.5	0.137	[−0.42, 0.06]	0.3
Q4	CG	3.64, 0.85	3.88, 0.69	−1.55	0.125	[−0.55, 0.07]	0.31
Q1	IG	3.9, 0.76	4.16, 0.47	−2.05	0.043*	[−0.51, −0.01]	0.41
Q2	IG	3.88, 0.52	4.12, 0.56	−2.22	0.029*	[−0.45, −0.03]	0.44
Q3	IG	4.26, 0.69	4.26, 0.8	0	1	[−0.3, 0.3]	0
Q4	IG	3.6, 0.73	3.92, 0.53	−2.51	0.014*	[−0.57, −0.07]	0.5

*statistical significance at the 0.05 level.

Table 6. Comparison of metrics between CG and IG groups.

Group	Dependent variable	t-value (df)	*p*-value	95% CI	Effect size (Cohen's d)
CG vs IG	Avg reading time for true articles	−1.47 (98)	0.146	[−1.45, 0.22]	0.29
	Avg reading time for fake articles	−2.29 (78.29)	0.025*	[−1.55, −0.11]	0.46
	Scrolls for true articles	−2.88 (98)	0.005*	[−1088.63, −200.17]	0.58
	Scrolls for fake articles	−1.44 (98)	0.152	[−705.67, 111.15]	0.29
CG vs IG	Dependent variable	U	*p*-value	n1/n2	Effect size (r)
	Clicks for true articles	938	0.027*	50/50	0.22
	Clicks for fake articles	1021	0.116	50/50	0.17

*statistical significance at the 0.05 level.

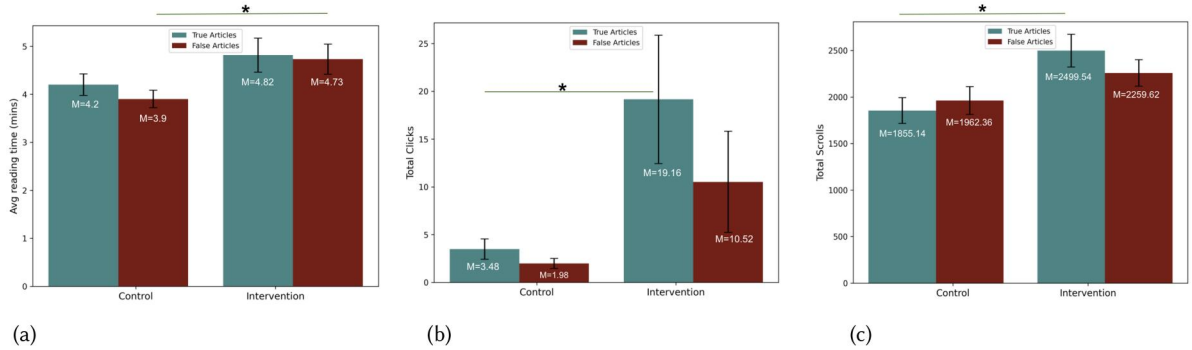
pre–post improvement ($p = .002$), consistent with item-level results. However, for the measure regarding *"confidence in the ability to perceive potential bias and identify inflammatory language in a news article (Q3),"* no significant difference was found between pre-study and post-study scores. While the tool did not improve this particular skill, its impact on other aspects of self-confidence in news evaluation led to rejecting $H1$, which stated that the intervention would have no significant effect on participants' self-confidence after the study. See Table 5 for detailed statistical reporting of metrics.

5.5. User engagement and interaction metrics

Our tool tracked participants' reading time, clicks, and scrolls on the webpage, excluding interactions within the tool dashboard. Since the tool had minimal clickable elements (mainly hover-over metric definitions), recorded actions primarily reflected engagement with webpage links and external sources. To accurately measure webpage navigation, we updated timing after each scroll. We analyzed 50 video sessions to assess dashboard interaction and noted a decline in tool use as users progressed, suggesting early tool engagement influenced reading behavior. Reading time was refined by excluding load-to-scroll and initial scroll times, capturing only active reading and enabling clearer group comparisons.

Across all six articles, significant differences were observed between the study groups. CG participants had lower average reading times (in mins) ($M = 4.05$, $SD = 1.27$) compared to IG ($M = 4.77$, $SD = 1.97$), $p = .032$, $d = 0.44$, 95% CI $[-1.38, -0.06]$. Similarly, CG had fewer total scrolls ($M = 3817.50$, $SD = 1885.53$) than IG ($M = 4759.16$, $SD = 2096.20$), $p = .02$, $d = 0.47$, 95% CI $[-1732.93, -150.39]$. For clicks, CG also showed lower interaction ($Mdn = 1.5$) compared to IG ($Mdn = 5.5$), $p = .015$, $r = 0.25$ indicating a small effect. Bonferroni-corrected comparisons confirmed a significant overall group effect ($p = .018$), with differences remaining consistent across engagement measures. Thus, we reject H_0^2 and conclude that engagement and interaction behaviors differed between groups when exposed to the tool.

We further analyzed logged data at the article level (see Table 6) using t-tests for normally distributed data and Mann-Whitney U tests otherwise, to determine whether *"true"* or *"fake"* articles attracted more interest. CG had lower average reading times for both true and fake articles (Figure 9a) compared to IG (True: $M_{CG} = 4.20$ min, $SD = 1.57$ vs. $M_{IG} = 4.82$ min, $SD = 2.51$; Fake: $M_{CG} = 3.90$ min, $SD = 1.28$ vs. $M_{IG} = 4.73$ min, $SD = 2.22$). For fake articles, IG spent significantly more time reading than CG, $p = .025$, $d = 0.46$, 95% CI $[-1.55, -0.11]$, indicating a small-to-moderate effect. No significant difference was observed for true articles.



(a) (b) (c)

Figure 9. Comparison of user engagement metrics across conditions, including reading time, interaction clicks, and scrolling behavior. These metrics highlight user interactions with news content and how they respond to our intervention. (a) Average reading time across conditions. (b) Number of user interactions (clicks). (c) Comparison of scrolling behavior.

Similarly, CG had fewer clicks (Figure 9b) and scrolls (Figure 9c) than IG for both true and fake articles. For scrolls, IG showed significantly higher values than CG for true articles ($p = .005$, $d = 0.58$, 95% CI $[-1088.63, -200.17]$), indicating a moderate effect, but not for fake articles ($p = .152$). Descriptively, for true articles, CG averaged 1855.14 scrolls ($SD = 977.08$) compared to IG's 2499.54 ($SD = 1245.19$), and for fake articles, CG had 1962.36 ($SD = 1049.54$) vs. IG's 2259.62 ($SD = 1007.95$). For clicks, IG showed significantly higher interaction than CG for true articles ($p = .027$, $r = 0.22$), but not for fake articles ($p = .116$). Descriptively, click counts were lower in CG than IG for both article types: for true articles, CG averaged 3.48 ($SD = 7.52$) compared to IG's 19.16 ($SD = 47.55$), and for fake articles, CG averaged 1.98 ($SD = 3.71$) versus IG's 10.52 ($SD = 37.49$).

Bonferroni-corrected comparisons confirmed significant effects for both interaction and reading behavior, including group differences in scrolls ($p = .034$), clicks ($p = .011$), and reading time ($p = .024$), as well as differences between true and fake articles ($p < .001$), with no interaction effects across article types. Overall, IG had significantly higher engagement, with longer reading times, increased scrolling, and more clicks. Because article lengths were matched across types, these differences reflect behavioral variation rather than differences in text size. This effect was more pronounced for true articles, supporting our hypothesis ($H2$) that the intervention would encourage deeper interaction with content.

5.6. Self-perceived overall knowledge gain

At the study's end, we assessed participants' perceived gains in critical thinking, media bias, and fake news detection using 3 Likert-scale items. Items were adapted from prior research by vraga2015multi (Vraga et al., 2015), with adjustments made to align with our study goals. As shown in Figure 10, IG reported higher scores than CG for critical thinking ($M_{IG} = 3.46$, $SD = 0.76$ vs. $M_{CG} = 3.10$, $SD = 0.93$) and media bias awareness ($M_{IG} = 3.62$, $SD = 0.92$ vs. $M_{CG} = 3.22$, $SD = 1.00$). Independent samples t-tests indicated that these differences were statistically significant (critical thinking: $p = .037$, $d = 0.42$, 95% CI $[-0.70, -0.02]$; media bias awareness: $p = .04$, $d = 0.42$, 95% CI $[-0.78, -0.02]$), reflecting small-to-moderate effects. In contrast, no significant difference was observed for fake news awareness and bias identification ($M_{CG} = 3.40$ vs. $M_{IG} = 3.68$, $p = .171$). Bonferroni-corrected post-hoc tests showed that fake news awareness was rated significantly higher than critical thinking ($p = .003$, 95% CI $[-0.41, -0.11]$).

5.7. IG tool satisfaction

We assessed IG participants' satisfaction using the Net Promoter Score (NPS) (Net Promoter Score, 2018) for its simplicity and direct focus on overall user sentiment. About 48% of participants were promoters, indicating generally positive perceptions. Overall, participants agreed that the tool enhanced their understanding of the material (see Figure 11), as supported by our qualitative feedback.

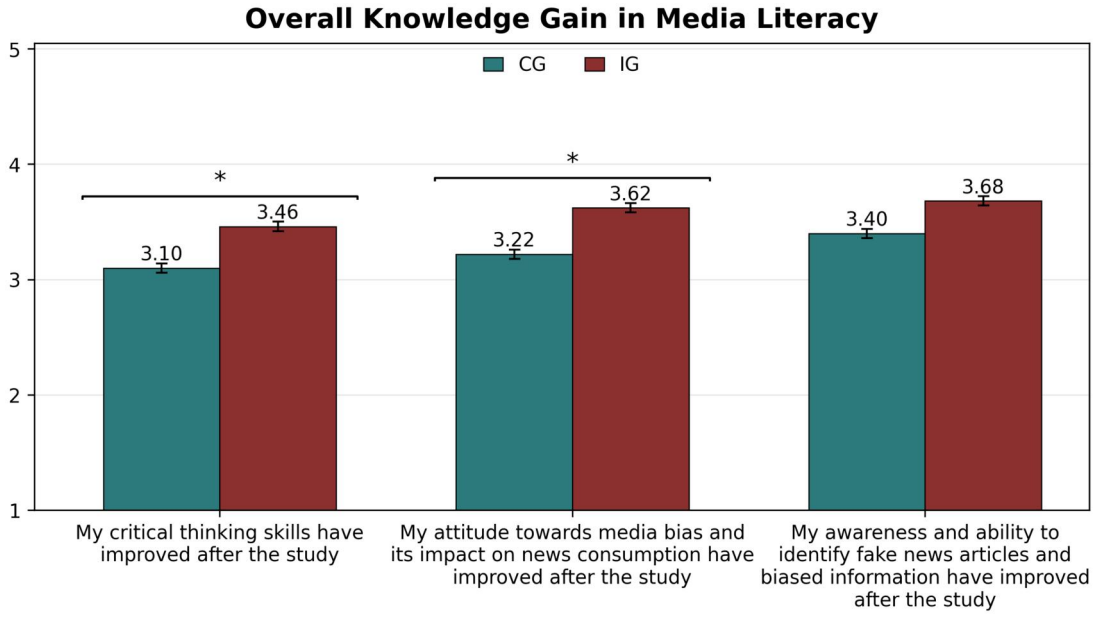


Figure 10. Comparison of knowledge gain.

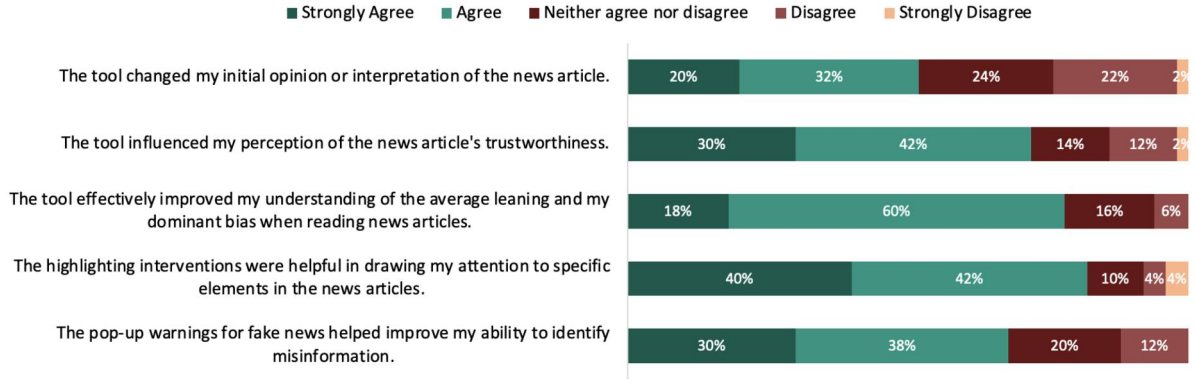


Figure 11. Impact of the tool on perception and interaction with news articles.

To better contextualize and interpret our above logging and behavioral data, we relate these findings to participants' in-study experiences, drawing on video observations to understand what was happening during their interactions.

5.8. Video-based analysis

This section analyzes participants' actions in session videos alongside recorded metrics. Of 100 sessions (∼ 90 hours), 92 high-quality recordings were retained. The research team collaboratively developed a video coding checklist (see Table 7) before analysis, which three researchers then used to systematically review behaviors and identify patterns aligned with the study objectives.

In general, participants exposed to the PI tool had more interactions with both true and false articles, as seen in Table 7. Specifically, they were more likely to conduct Google searches related to the article's content, open new tabs for fact-checking, explore the credibility of news sources, hover over and click on the author's name for credentials or background information. In general, other unlisted actions included checking word definitions in news articles by double-clicking in the Mac, which was a common practice; following text with hovering over alerts; and minimizing and unminimizing the tool. One notable example is that participants often showed heightened scrutiny when engaging with fake news, actively searching for "*Biden-related articles*" and going into Wikipedia for additional context.

Table 7. Actions taken by participants in both intervention (IG) and control (CG) groups.

	Intervention (IG)		Control (CG)	
Actions	True articles	False articles	True articles	False articles
Google search related to article content?	28	26	20	17
Opened new tabs for fact-checking articles?	29	29	13	17
Explored credibility of news source?	21	7	2	7
Highlighted specific terms while reading?	33	16	6	2
Hovered over author for credentials/background?	11	9	4	5
Watched videos while reading?	3	1	10	8
Scrolled back up and re-read article?	16	7	12	3
Clicked on embedded links in article?	7	2	3	2
Engaged with pop-up/ads in article?	4	1	4	2
Other actions?	0	1	0	0
Total actions count	152	99	74	63

For example, after the article with "*Ms America,*" participants became more aware of the tool and of the information it provided about the article. Interestingly, this search behavior mirrored their actions when reading trustworthy news, suggesting a consistent approach to information-seeking regardless of article credibility. Conversely, they focused on surface-level engagement when skimming through fake news, with minimal additional actions beyond reading. CG behavior and actions on both true and false articles were nearly evenly split (*74 vs. 63* actions) and showed balanced engagement. Overall, IG took more of an information-seeking approach.

5.9. Additional news articles

Beyond video analysis, we compared the additional articles visited and found slightly higher consumption in IG than in CG. In total, participants read 92 extra articles—51 by IG and 41 by CG. IG averaged 4 articles ($SD = 3.60$), CG 3 ($SD = 2.58$). We also extracted domain names from unique URLs to compare reliable vs. unreliable source engagement. While both groups visited partisan sources like FreeBeacon.com, with slightly more visits on IG (6 vs. 5), IG also showed increased engagement with credible sites like AP News (3 visits vs. 2) and Yahoo News (1 visit). Ballotpedia.org, a factual political site, was visited only by CG. Notably, visits to Dunning-Kruger-Times.com, a known satire/misinformation site, were higher in IG (2 vs. 1). While the intervention did not significantly reduce visits to unreliable sources, it encouraged critical evaluation and broader exploration with different perspectives rather than users passively consuming familiar content. Additionally, visiting misinformation-prone or partisan sites (e.g., Freebeacon.com) does not necessarily imply belief—users may do so out of curiosity or to verify claims—supporting our survey findings in Section 5.3.2. Having presented the survey and logging results, we now turn to the qualitative findings to provide deeper insight and context behind these patterns.

5.10. Qualitative findings: Control group

While both groups discussed their reading experiences and perceptions of the article, IG provided detailed feedback on the tool. Without tool exposure, CG speculated broadly on news discernment challenges and interest in tools like ours, whereas IG centered more on post-tool usage experiences. We present the CG findings first, followed by the IG findings.

5.10.1. Challenges in identifying fake news

CG participants (21/28) highlighted three key challenges in identifying fake articles during the study, including the objective presentation of content, limited time for verification, and limited access to alternative sources. For example, P8 noted:

> *"I sometimes think if I'm trying to reach an article and get blocked by like having to have access, or paying for an account, or signing up for advertising, and then sometimes I think, 'Oh, it would be good to know that information. But is it worth trying to get through all of this now?' And maybe that's just because we tend to be a little bit more fast-paced with information." (P8)*

Others (7/28) mentioned challenges in discerning reliability when articles seem less biased or a website seems credible. For example, P10 said, "*For inaccuracies, I guess if something is presented objectively, it's challenging because it's as if there are no opinions, and it doesn't raise red flags, but those facts can still be biased if they all support one specific argument.*"

5.10.2. News verification methods

Nearly one-third (11/28) mentioned using Google to verify news. For example, P16 stated, "*I generally Google stuff and read more…* " Wikipedia and university libraries were also cited as resources for learning more about an issue or topic. Over half (15) identified red flags for biased articles, such as inflammatory language, one-sided narratives, excessive ads or pop-ups, and missing information or errors. For example, P50 said, "*very strong or absolute language*." Another, P8, mentioned getting a feel for whether an article would be biased based on the title and asking, "*… if it's taking [some] kind of an ultra-negative or an ultra-positive side.*"

5.10.3. Interest in a news intervention tool

Most participants (20/28) expressed interest in using a news intervention tool. Others (2) had a conditional interest. For example, P14 stated that they would be more likely to use it if it covered topics they are already interested in, saying, "*I would have to be kind of invested in it to want to do it.*" Another, P94, expressed interest as long as the tool "*was reliable, itself not biased, and not super intrusive*." However, P41 strongly opposed using such a tool, "*I don't want to consume incorrect information. I want to have the facts and form my own opinion independent of others*." Several participants listed "*saving time*" as a potential benefit. For example, P80 said,

> "*Probably, like… as an extension or something of the sort. It could be pretty useful, you know, as long as it's not too much of a nuisance to actually set up or use in practice. I would definitely consider using it, but, at the same time, I'd want to know a bit more about, like, the potential biases (AI, race, politics, etc.) that create the tool. I guess, overall, it's like… 'helpful, but I'd want to trust it first." (P80)*

5.10.4. Desired features

Participants stressed the importance of accessing diverse sources directly through the tool. For example, P10 said, "*I would want it to provide a contrasting piece of information against the article I am reading, to see both sides of an issue*." P8 recommended features like "*access to factual backgrounds without all the politics* to empower users to form their own opinions. Another participant, P9, mentioned the importance of full quotes, saying, "*So I would like something that provides full quotes… so the quote checker would first tell you everything the person's quoted, not just a few words of it, and tell you whether the person actually said it*."

About one-fourth (8/28) also advocated for more nuanced approaches to measuring article reliability. P13 suggested, "*It kind of, you know, gets some of the most important keywords of the news and then gives you, like, a number in a range to indicate the reliability level rather than a binary assignment.*"

Some participants also suggested incorporating highlighters or language flagging to identify bias. Another, P1, proposed "*adding interactivity or community features*" to allow users to engage with others. A final suggestion was for the tool to explain how it rates bias or accuracy, acknowledging that its own conclusions may not be infallible. As P18 said,

> "*… the tool should almost be self-aware of own potential limits, or bias or unreliability. So it's like, a tool that cross-generates the article, and then also cross-checks itself somehow or mentions, like, this is how we did it, this is the process behind it, and, you know, it might not be perfect or fully accurate." (P18)*

Some features were implemented, as they are not just theoretical but genuinely needed and valued by users in real-world scenarios, validating the design choices.

5.11. Qualitative findings: Intervention group

Having discussed the findings from the CG, we now turn to the findings from the IG to understand how the introduction of the tool influenced outcomes.

5.11.1. Overall impression and usability of the tool

Most participants (19/28) found the tool intuitive, though some wanted better placement and more interactivity; a few (6/28) found highlighting and bias reports confusing. For example, P6 mentioned,

> "*The highlights were good but distracting in some cases. Maybe a self-bias check as well because some of the more right-leaning articles were flagged for being inflammatory, but there are also many left-wing inflammatory articles out there that could be used in this… (P6)*"

Some (9) found the tool non-distracting as they had the option to minimize it, while others (8) found it somewhat distracting, though several became accustomed to it over time.

When asked about the tool's interactivity, participants had mixed reactions, with more than half (17/28) being positive. Among these, nine appreciated the tool's balanced approach. For example, P23 said, *"It wasn't too in your face, hence not overly intrusive, but it also wasn't hidden so that I forget it."* The most popular feature was the keyword highlighting and fact-checking messages. Ten users found sentiment and bias analysis helpful. For example, P7 said,

> *"I liked the dominant bias feature because many people tend to be extreme but think they are more neutral, so I think that's amazing… It made me more conscious of potential biases in the articles. I spent extra time analyzing the language and sources used to get a balanced perspective." (P7)*

However, some (8/28) desired more engagement and interactivity, while the rest (3) found it overwhelming. P46 mentioned, *"The tool can be too much for an average reader like me. I personally prefer reading news on my mobile phone, and I would love to see how the tool works in a mobile environment."* Another expressed a more focused need, such as participant, P35, *"No, I do not need details like the time I spent reading articles; rather, I will focus on whether it is fake or real."*

5.11.2. Tool's impact on fact-checking

To confirm certain observed trends in participants' news consumption, we were interested in how our tool influenced participants' choice to engage more with news articles—a majority (19/28) acknowledged that the tool positively influenced their behavior. For example, P100 mentioned feeling more inclined to cross-check, *"When the tool spotted possible bias, I thought harder about what I was reading…. and the tool put the idea in my mind that a lot of what I was reading could be fake*. Similarly, P19 mentioned,

> *"The tool made me want to fact-check more myself. When I saw those highlighted words with pop-ups. Clicking on the links seemed like the natural thing to do, and it made me curious about the backgrounds of news sites I wasn't familiar with. The tool had this way of nudging me to seek more information and verify what I was reading." (P19)*

Among them, beyond encouraging fact-checking, 8/28 explicitly noted that it increased their critical thinking, *"I'm inclined to think more about the biases… which I couldn't have done earlier (P20)."* However, the tool's impact was not universal as few (4/28) reported feeling unaffected or relying on their prior and existing skills, *"I think it would have influenced me more if I wasn't in a journalism class because we have been taught these skills already (P45)."* A small subset (5/28) of participants also expressed concern that the tool made them generally more skeptical toward all news, rather than helping them discern fake news specifically. For example, some reported becoming hyper-aware of bias across all articles, which occasionally led them to dismiss content prematurely rather than analyze it objectively.

5.11.3. Tool's impact on reading time

From both logged metrics and participants' responses, it became evident that reading habits were not uniform across all six articles. The majority (24/28) acknowledged spending unequal amounts of time on different articles, citing factors such as article complexity, length, and personal interest. One participant, P88, said, *"I definitely didn't, and even though I believed [an article] to be true, I spent more time reading it."* Another said,

> *"It probably influenced me to spend less time reading fake news. I more quickly left the fake news site – either to abandon the article [knowing it was fake] or to look for real facts (P99)."*

Table 8. Overarching feature-specific findings on instances participants discussed key tool features.

Feature	Count	Example & impact	Participant quotes
Highlighting key terms & phrases	12	Users reported that **highlighted words** flagged for **potential bias** encouraged them to reflect on language used in the article.	*"Seeing certain words flagged made me question whether the article was pushing an agenda. Although it took me a while to realize what the highlighted words meant, once I did, I paid more attention to language cues. (P19)"*
Pop-up Warnings & Explanations	18	Users found the **pop-up alerts useful** in identifying **misleading headlines** and **potential misinformation** before reading the article.	*"I liked the pop-ups that said things like 'this headline is sensational.' It made me second-guess my immediate reaction to the article. (P100)"*
Source assessment & bias indicators	10	The **credibility and bias indicators** of news sources encouraged users to **cross-check claims**.	*"Seeing the tool mark a news site as biased prompted me to look it up, and it turned out to be true. I didn't always agree with where it flagged bias, but it made me curious enough to fact-check sources more than usual, just enough to be sure. (P60)"*
Sentiment analysis	8	Showed article *polarization*, helping users recognize hidden bias.	*"I really liked how it flagged whether the news was neutral, negative, or positive—it gave me a quick heads-up before I even started reading. (P5)"*
Cognitive nudges (tool presence)	7	The **mere presence** of the tool acted as a **reminder** to approach articles **critically**, even when users didn't actively engage with it.	*"I don't think I changed how I read, but it definitely put it on my mind that a lot of this could be fake, so I fact-checked more than usual. (P35)"*

Conversely, 4/28 participants were particularly drawn to the articles they perceived as false, spending extra time analyzing them—often out of curiosity or even amusement. While some participants disengaged from suspicious articles, others engaged more deeply with the content in an attempt to dissect misinformation or understand its framing, which could explain the differences in increased interaction metrics observed in IG. P32 noted how the tool's sentiment analysis encouraged deeper engagement with emotional aspects of news articles,

> *"… the sentiment analysis added a layer of empathy to my reading. I was drawn to sections with contrasting emotional tones, as I wanted to understand the human elements behind the words. This emotional connection naturally extended my reading sessions (P32)."*

Overall, the time spent on the articles was influenced by the information provided by the tool (e.g., bias level, personal interest, perceived credibility, and engagement with the tool's features), with different participants' interests being piqued and spending different lengths of time depending on whether they believed the article to be false or true.

To better understand which aspects of the PI tool aid content discernment and critical thinking, we examined participants' responses for feature-driven effects (see Table 8). Pop-up Warnings & Explanations (Figure 2, Right) were the most effective in fostering critical thinking (18 mentions), followed by Highlighted Keywords (Figure 2, Left, 12 mentions).

5.11.4. Likelihood of future use

Almost half of the participants (14/28) expressed clear willingness to use the tool in the future. For example, P5 said, *"I would love to because it tracks my reading habits."* Approximately one-fifth of participants (5/28) expressed more conditional interest, such as P88, who mentioned, *"If such a tool is available, but with more advanced features and is esthetically pleasing so it doesn't block anything from my view and such, I think I would use it."* Only three participants were uninterested, mainly due to limited news consumption or missing relevant content, and six participants were cautious about data

tracking and privacy concerns (e.g., P27: *"I care about my privacy, so I needed to know exactly what data the tool was collecting and how it would be used."*).

5.11.5. Tool recommendations

Participants mostly suggested personalized improvements for our current design, with one-fourth (7/28) recommending customized features. As such, P35 desired fact-checking tailored to *"crime and entertainment news."* Additionally, a participant suggested informing users about the alignment of news with their personal political biases,

> *"Maybe pointing out that it aligns with your beliefs so you're more aware of the unconscious bias, because you might think something is less biased if it supports what you believe. Because I feel like if something doesn't align with your beliefs, you're like, 'Oh, yeah, that's biased.' Maybe that unconscious bias being pointed out might be helpful to people who really rely on it (P25)."*

One participant, P21, suggested incorporating more key performance indicators (KPIs) in the dashboard (for helping users to understand the average time spent on an article) on the visual display, and monthly or weekly stats to serve as a goal or activity tracker. Another recommended a feature tracking users' time spent on specific news domains, *"It might help to know how long I've spent on different sites, like the New York Times, versus a different news site (P88)."* When asked about other general features, participants suggested using highlights to summarize articles, not just flag misinformation. As one participant explained,

> *"I think the premise of this tool would have been more like, if it is highlighted, that indicates this might be an unreliable article, and then not highlighted articles would be an implication like this is probably reliable, but I think it would just be better to have consistency (P23)."*

Furthermore, participants emphasized the need to spot emotionally charged language, e.g., *"Point out inflammatory words to indicate bias"* (P25). Building on our key findings, we now outline the broader implications in the discussion, particularly for the design and use of such tools.

6. Discussion

In this study, we designed and evaluated a PI-based dashboard that integrates user-facing interventions directly into the news-reading process. Our goal was to understand not only whether such a tool fosters more critical and reflective engagement, but also how it shapes readers' behaviors, perceptions, confidence, and feature acceptance. Inspired by the broad potential of PI in HCI (as highlighted by Epstein et al., 2015a, 2015b; Li et al., 2012; Loerakker et al., 2024), our PI dashboard captures client-side interactions and browsing behaviors, enabling users to make sense of their reading data and assess news quality within their natural reading flow. Unlike static detection tools, our approach embeds interventions into the reading process. We find that these targeted interventions positively influence news engagement: survey and interaction data show increases in fact-checking and more critical—but not uniformly distrustful—evaluation in the intervention group. Video observations further showed active cross-referencing of sources and deeper engagement with content, findings echoed in qualitative responses (See Table 8). This suggests that participants' verification behaviors were driven by a desire to confirm accuracy rather than to dismiss content. These outcomes align with cognitive reflection research by Pennycook and Rand (2019, 2020), showing that analytic strategies help readers detect inconsistencies and potential misinformation. Overall, our findings deepen understanding of how PI-based interventions shape news-reading strategies and highlight their promise for cultivating long-term critical thinking skills, while remaining exploratory and descriptive in nature and providing directional insights into how such interventions may influence user behavior.

To further explore how such tools influence user behavior, we build on previous work that examines influential aspects of online interactions (Arapakis et al., 2014), quantifying actual user engagement with "true" and "false" content. We provided statistical evidence and explained what appeared to be the most promising based on our qualitative findings. However, rather than isolating specific feature impacts in detail and quantitatively, we focused on understanding users' perceived value, empowerment, and behavioral changes. Our findings emphasize the importance of transparency, interactivity,

and user empowerment in real-time misinformation tools, rather than relying solely on flagging or prescriptive judgments of news credibility. We note that the intervention combined content- and interface-level changes; as such, the observed effects reflect the overall system experience rather than the isolated contributions of individual components. In the following sections, we address each research question in turn, outlining how our findings contribute to answering them.

6.1. Perceived discernment gains and interaction behaviors

Our analysis for RQ1 suggests that the PI tool shape both how participants interacted with articles and how confident they felt about evaluating news (Sections 5.4 and 5.6). Pre–post survey comparisons indicate some improvements in the IG across multiple evaluation items (Section 5.4.2), whereas the CG showed no comparable shifts. However, bias/inflammatory-language confidence (Q3) did not improve. This is not surprising, as perceiving and identifying bias is a nuanced skill (Pennycook & Rand, 2021). These findings suggest that procedural supports alone may be insufficient; targeted, in-flow exemplars and tunable sensitivity mechanisms may be needed to better scaffold this specific form of discernment. Additionally, participants' proficiency in identifying bias likely varies; some may be naturally more adept, while others may continue to struggle even with support, a factor not fully captured in this study. Importantly, our claims about discernment refer to *perceived discernment*, operationalized through self-reported confidence and evaluative judgments. We did not include objective performance-based assessments (e.g., calibration tasks or validated accuracy benchmarks) to independently examine whether participants' evaluations became more accurate or whether increased confidence corresponded to demonstrated skill. Accordingly, the observed changes should be interpreted as shifts in self-efficacy and perceived evaluative ability rather than as evidence of improved detection accuracy. This distinction is an important limitation of the study and narrows the strength of the claims we can make.

Similarly, article-level ratings reflect participants' interpretations rather than an established ground truth. Future work should therefore pair confidence measures with behavioral accuracy tasks and calibration metrics, while also employing larger samples to more robustly estimate effect sizes. Such work would enable clearer differentiation between perceived and actual discernment and more definitively evaluate how interventions like ours influence real-world news literacy performance (see Section 5.3.1).

Interaction logs and interviews (Sections 5.5 and 5.11) suggest that IG participants did more than spend longer on pages; they appeared to engage more reflectively. Readers paused to inspect authors and sources, opened new tabs to cross-check claims, and revisited passages; importantly, these behaviors occurred on the article pages themselves, not just within the dashboard. This pattern may indicate that the tool functioned as an indirect scaffold, surfacing habits for self-monitoring without explicitly dictating workflow. The observed behaviors are consistent with prior work suggesting that sustained attention can support discernment (Allcott & Gentzkow, 2017) and align with research linking systematic interaction patterns to higher information judgment (Pointon et al., 2023). They also echo findings from the lateral-reading literature, potentially by reducing the effort required for verification and supporting reflection as an in-situ, low-cost action (Wineburg & McGrew, 2019).

Consistent with these patterns, IG applied stricter credibility thresholds than CG, reflecting accuracy-driven skepticism rather than indiscriminate rejection. This outcome is double-edged: while skepticism can protect against overconfidence—a known driver of misinformation spread (Lyons et al., 2021), it may also contribute to general distrust and deception bias (Lewandowsky et al., 2017; Van Der Meer et al., 2023). This tension is especially salient in high-stakes contexts, such as the COVID-19 pandemic, where trust in accurate information is critical. Similarly, over-skepticism may lead users to discount accurate information, while repeated exposure to flagged or misleading content may create false reassurance (e.g., assuming unmarked content is reliable) or inadvertently amplify attention to such claims. These dynamics are consistent with prior findings that exposure to inaccurate information can increase confusion and undermine accurate understanding, and that corrective interventions may fail or even reinforce misperceptions (Nyhan & Reifler, 2010; Rapp & Salovich, 2018). Such risks are particularly concerning in crisis contexts, where heightened attention can accelerate the spread of rumors. Importantly, the motivation behind skepticism matters: accuracy-driven skepticism fosters deeper engagement, whereas identity-driven skepticism can hinder it (Li, 2025).

By encouraging extended reading and fact-checking, the PI tool positions skepticism as a constructive force for informed news consumption. Future iterations could incorporate adaptive intervention logic, such as prioritizing source credibility cues, limiting repeated exposure to misleading claims, or calibrating alert frequency to better balance critical engagement with trust in verified information.

6.2. Verification and lateral reading behaviors

Our analysis for RQ2 reveals that the PI tool promoted verification-oriented behaviors and facilitated deeper interaction with news content. Behavioral logs and video coding revealed that IG participants conducted more searches, opened additional fact-checking tabs, explored source credibility, and hovered over or clicked author information (see Table 7). These actions suggest an active verification mindset rather than passive consumption. In particular, cross-checking intent was significantly higher for false articles—potentially with a strong and lasting effect (Guess et al., 2020)—but there was no group difference for true articles. This asymmetry suggests the tool primed users to question content more carefully when inaccuracy cues were present.

Qualitative findings reinforce this pattern. Unlike credibility indicators in prior studies (Lu et al., 2022), which sometimes misled users into following incorrect AI predictions, our tool promoted active scrutiny by providing transparency on bias, source quality, and explanations without prescribing judgments. CG. Participants judged false articles more critically, rating them lower than CG, and showed gradual shifts in behavior, including more visits to partisan and misinformation sources (Section 5.9). These behaviors indicate that behavioral change in news consumption is gradual, and the intervention may have begun to shape awareness rather than immediately alter habits. Longitudinal analysis would help assess whether users slowly diversify their sources and reduce engagement with unreliable content. Our findings, particularly the increased skepticism in credibility ratings, heightened fact-checking behavior, and greater engagement among IG users, suggest that providing users with interpretative tools rather than prescriptive judgments fosters a more discerning approach to news consumption, reinforces user agency and healthy skepticism, while minimizing unintended intervention effects. Beyond the assigned articles, IG participants demonstrated modest but noticeable exploration, including more lateral reading and fact-checking. However, source exploration remained limited. This echoes Wang et al. (2025) work, which shows that cross-publisher comparisons and alternative narratives can broaden perspectives. While our tool fostered accuracy-driven skepticism, adding such comparisons could deepen critical engagement. Compared with top-down detection and static credibility labels, our PI approach centers on explanation and self-regulation. Brief "why" rationales alongside provenance/sentiment cues, lightweight session metrics that externalize reading habits, and optional one-click verification affordances appear to shift readers from passive consumption to reflective scrutiny.

Despite these promising behaviors, one caveat tempers the interpretation. It is important to note that our tool exclusively logged articles from news domains covered by the AllSides news database. Hence, our logging strategy underestimates the full scope of verification: IG participants may have read or fact-checked beyond the monitored domains, which are not included in our tracking. Overall, the findings highlight design tradeoffs: interventions must balance encouraging vigilance with sustaining trust, and future work should test whether such effects persist and scale in real-world, long-term use. Taken together, these observations motivate a closer examination of how such interventions may operate beyond controlled settings, particularly in crisis and high-stakes information environments.

6.2.1. Implications for crisis and high-stakes information contexts

While our findings suggest that reflective cues may foster scrutiny and user agency in controlled reading settings (see Table 8), their implications in crisis or high-stakes information environments warrant careful consideration (Radinsky & Tabak, 2022). Research shows that rapidly evolving, often conflicting information sources create dynamic sociotechnical ecosystems in which rumors can spread rapidly and exhibit distinct transmission patterns across content types (Palen & Liu, 2007; Vosoughi et al., 2018; Zeng et al., 2016). In such contexts, interventions that encourage skepticism may carry tradeoffs. Over-skepticism could delay trust in verified emergency guidance. At

the same time, transparency cues might provide false reassurance if users overinterpret heuristic signals (e.g., sentiment or source ratings) as definitive indicators of credibility. Additionally, highlighting potentially misleading content may increase its familiarity or salience, a dynamic observed in research on the continued influence of misinformation and the challenges of corrective interventions (Lewandowsky et al., 2012; Nyhan & Reifler, 2010).

Our observed increase in scrutiny and modest growth in lateral reading (see Section 5.9) suggest the tool can nudge users toward verification behaviors; however, in time-pressured scenarios, this additional friction could either support more careful evaluation or impede rapid decision-making. Crisis informatics research demonstrates that alternative narratives often propagate through interconnected media ecosystems, where similar claims are recycled across multiple domains, creating a false perception of source diversity and reinforcing conspiratorial interpretations (Starbird, 2017). Moreover, disinformation campaigns frequently operate as participatory and collaborative phenomena, blending orchestrated efforts with organic crowd activity, which can make misleading narratives self-sustaining and difficult to distinguish from authentic discourse (Kate et al., 2019). In such environments, reflective cues may help counter superficial credibility signals, but they also introduce design challenges related to timing, repetition, and user attention. Future adaptations for crisis contexts may therefore require tiered or urgency-sensitive interventions, clearer provenance indicators tied to verifiable institutional sources, and safeguards against repeated exposure to recycled misinformation. Designing for adversarial and rapidly evolving environments will require balancing epistemic vigilance with timely decision support.

6.3. Feature utility and acceptance

Answering RQ3, participants consistently valued transparent and contextual features—such as keyword highlights, pop-up explanations, and bias or sentiment cues (see Table 8)—that explained why content might be questionable, rather than issuing blanket credibility labels. These features appeared to support more interpretable and user-driven sensemaking by making underlying signals visible and open to scrutiny, rather than obscuring them behind opaque judgments. This preference challenges the dominant "flagging" approach in misinformation interventions, which has been shown to risk overreliance or miscalibrated trust in automated cues. Instead, it points toward the importance of explanation-rich, context-sensitive designs. At the same time, reports of distraction, feature overload, and heightened skepticism toward all news suggest design tradeoffs: supports that nudge reflection can also foster alert fatigue or excessive distrust, echoing concerns about "deception bias" in the post-truth era (Lewandowsky et al., 2017; Van Der Meer et al., 2023). While session metrics and reflective nudges align with personal informatics principles (Epstein et al., 2015a), our findings indicate they remain underutilized unless made actionable and personally meaningful. Overall, the evidence suggests that effective PI-based misinformation tools must move beyond static cues toward interventions that balance interpretability and user agency with safeguards against over-skepticism—a design challenge that remains unresolved in the broader literature.

6.4. Design insights and ethical considerations for future deployment

Our mixed-methods results suggest that effective misinformation interventions must increase scrutiny *without* inducing blanket cynicism, reduce the friction of verification, and scaffold skills that current tools do not reliably build (e.g., recognizing bias/inflammatory rhetoric). Notably, CG participants mentioned features already in our tool, indicating the design aligns with user expectations—a promising sign of intuitive appeal. We translate these insights into actionable design principles below. We use identifiers D1–D5 to present the design recommendations emerging from this study.

6.4.1. D1: Balance engagement with calibrated skepticism

Our findings show that IG readers spent more time on articles, scrolled and clicked more (Section 5.5), and engaged in more verification behaviors (Section 5.8)—particularly when content was false—yet this deeper engagement also risked fostering over-skepticism. Tools should therefore encourage critical

reading while avoiding blanket distrust. Interventions can draw attention through clear visual cues and brief explain-why rationales, while also highlighting signals of reliability (e.g., "independent sources present," "on-the-record quotes"). At the same time, lowering the cost of verification is essential: just-in-time "recipes" such as pre-filled search queries, one-click source background, fact-check lookups, or snapshots of what reputable outlets report can efficiently convert intent into action. At the same time, support should be adaptive to context. Stronger nudges may be useful when users encounter unfamiliar topics or exhibit low source diversity, while assistance can taper when consistent verification habits emerge. A personalized skepticism index could help calibrate this balance, though poorly tuned adaptations risk either reinforcing bias (over-alerting) or enabling complacency (under-alerting). Following the principle of progressive autonomy (Truszkowski et al., 2005), interventions should shift gradually from guided prompts toward independent reflection so that the tool scaffolds rather than replaces judgment. Ultimately, prioritizing accuracy-driven skepticism and reflective engagement can build long-term resilience to misinformation without undermining trust in credible information.

6.4.2. D2: Tunable detection of bias and strong language

Participants valued the highlights of loaded wording (Section 5.11.1), although tolerance for alerts varied—some found them helpful, while others considered them intrusive. Future tools can offer user-adjustable sensitivity and depth (from subtle underlines to hover examples) to help address the gap observed in Q3: *"confidence in identifying potential bias and inflammatory language in news articles."* This could be done through inline micro-exemplars (e.g., contrastive rewrites, standard frames). Defaults should remain gentle, with richer explanations on demand. Prior work, including our findings on how sentiment and framing shape perceived bias and emotional responses (Khatiwada et al., 2026), supports the need for such adaptive, user-facing interventions. Noh et al. (2025) BIASsist also demonstrates that combining identification, explanation, and neutralization can significantly improve bias detection and foster deeper engagement.

6.4.3. D3: Support autonomy through simplicity (stage, then taper)

Interventions should remain inline, lightweight, and dismissible—favoring tooltips or hover cues over intrusive modals—to balance support with user autonomy. Video analysis showed strong early use followed by a decline, suggesting the need to front-load guidance on initial items and then taper, with options to snooze or disable cues. This approach aligns with HCI principles of intuitive design (Norman, 2013) and Cognitive Load Theory (Oviatt, 2006), which stress that simplified, flexible interfaces reduce friction in complex decision-making. Our findings confirm this: while integrated features could have added distraction, participants reported the opposite, noting that minimizing the tool preserved flow without sacrificing access per Section 5.11.1 (e.g., P23: *"It wasn't too distracting… I could easily minimize it"*). Although some initial skepticism about accuracy was evident, trust grew as users engaged more critically with the tool's features. Future iterations should build on this by minimizing cognitive friction further and ensuring interactions remain user-driven, lightweight, and autonomy-preserving.

6.4.4. D4: Close the personal informatics (PI) loop with reflective metrics and gentle goals

To encourage sustained user engagement and more informed decision-making in news consumption, future tools should adopt PI approaches that surface concise, private, and actionable summaries relevant to discernment—such as session time, source diversity, lateral reads, and reliance on low-credibility sites. Our findings showed that IG participants engaged more deeply through longer reading (Section 5.11.3), more verification, and broader exploration, suggesting clear value in reinforcing such habits through thoughtfully designed feedback mechanisms. These may include optional, low-pressure goals (e.g., "two perspectives per topic," "three lateral reads this week"), periodic snapshots, and lightweight progress indicators that support reflection without creating pressure or fatigue. Real-time, personalized feedback and adaptive interfaces can gently nudge reflection without dictating behavior, enabling users to develop their own awareness and strategies over time. Importantly, such designs should remain flexible and user-controlled, avoiding one-size-fits-all models while supporting diverse

reading practices and preferences. This approach aligns with personal informatics principles by promoting mindful, user-centered engagement, strengthening critical evaluation, and supporting long-term habit formation without imposing rigid prescriptions.

6.4.5. D5: Ethical instrumentation and provenance transparency

Participants' trust depended on understanding what the tool does and why. Interventions should clearly and proactively communicate what data is logged, how it is processed, and provide users with accessible options to review, export, or delete their data. When surfacing bias or sentiment signals, tools should also surface uncertainty and methodological context (e.g., "based on AllSides ratings," "sentiment approximated; may not capture sarcasm"), helping users interpret these signals more critically. Such algorithmic humility prevents over-reliance on system judgments, encourages informed user engagement, and positions transparency as a foundational design principle rather than an optional feature. This recommendation is further supported by prior work on human–AI collaboration, which shows that AI-generated explanations can significantly increase user confidence and influence decision-making, with detailed explanations being particularly persuasive (Wang et al., 2026). This raises concerns about overreliance, reinforcing the need for transparent instrumentation and clear communication of uncertainty to support more critical engagement.

6.5. Ethical considerations and data governance

Ethical considerations are key in developing tools that shape news consumption habits. A critical aspect of is providing clear, accessible explanations of how interventions are selected and applied impartially across diverse news content. Further, while not all participants expressed concerns about data privacy, it is essential to acknowledge the perspectives of those who did. Ignoring these concerns risks undermining user confidence and engagement. Intervention designers should be upfront about how these tools may influence users' reading habits, what information they provide, and any potential risks associated with data-collection practices (Lee et al., 2022). Without this clarity, users might feel their freedom or autonomy is restricted. Preserving user autonomy is paramount; interventions should inform, not coerce, offering objective, journalistic-quality measures and guidance to enhance users' critical thinking skills. Moreover, if not carefully designed, user-centered interventions can be ineffective, misused, or have unintended consequences as certain user groups may manipulate or exploit algorithmic interventions for harmful purposes (Warner et al., 2025).

In addition to transparency around intervention logic, we implemented safeguards for backend logging. The system recorded only interaction-level metrics necessary for analysis (e.g., scroll events, clicks, reading time, and article IDs). It did not store full browsing histories or unrelated URLs outside the study interface. Data were linked to study-specific participant IDs rather than personal identifiers, and survey responses were stored separately from interaction logs. All data were stored on secure, access-controlled university servers and accessible only to authorized research team members. Participants were informed during consent that interaction data would be logged for research purposes. While the prototype did not include user-facing data deletion controls, in the future, we will implement explicit retention and deletion policies (e.g., automatic deletion of raw logs after 30 days), participant-facing data controls (e.g., export and delete-on-request), encryption at rest and in transit, and fine-grained access controls with audit logging to strengthen data privacy.

6.6. Limitations and future work

While the study has limitations, these also point to important avenues for future research and refinement. Our sample was primarily 18–34 year-olds with limited representation of republicans, constraining the generalizability of findings beyond young adult populations and across political perspectives. Our qualitative findings may reflect self-selection bias, as only volunteers participated in debrief interviews. Thus, the Hawthorne effect (McCambridge et al., 2014) and social desirability bias remain possible, as participants may have altered behavior under observation, an expected tradeoff in controlled, time-intensive studies. Ecological validity is another consideration: the study was conducted on a

desktop, whereas much news browsing occurs on mobile devices, and occasional website restrictions limited access to some article links.

In terms of sample and platform fit, the current workflow is best suited to contexts that involve deliberate, article-level engagement rather than rapid, feed-based browsing. For example, it may fit naturally in classroom settings, media literacy curricula, research or academic reading, policy analysis, investigative journalism, and other situations where users intentionally open and evaluate full-length articles. The intervention prototype was designed around sustained interaction with a selected article, which differs from the fragmented, scroll-driven consumption patterns common in mobile-first social media feeds. While this limits direct generalizability to everyday feed-based news use, it reflects scenarios where deeper verification and reflective reading are both possible and expected. For broader real-world adoption, particularly in mobile and platform-mediated environments, the interaction model would need to be adapted. This may include lightweight, feed-integrated cues, reduced cognitive load, gesture-friendly overlays, and compatibility with platform constraints, as seen in emerging plugin- and platform-level interventions. Future work should examine how the core intervention principles can be translated into mobile and feed-based ecosystems without disrupting natural browsing behavior.

Technical scalability also remains a challenge, as the prototype relied on pre-processed articles and lightweight intervention rules that do not yet support real-time processing. Our dashboard relies on AllSides bias ratings and VADER sentiment scores, which carry known limitations (e.g., coarse political categorizations and sensitivity to context), so we present these signals as indicative rather than authoritative. Scaling the tool beyond the lab will require robust infrastructure for article parsing, natural language processing, and latency-sensitive delivery of intervention messages. Finally, because the study was slightly underpowered and involved multiple comparisons, the quantitative findings should be interpreted as exploratory rather than definitive. Claims about discernment are based primarily on self-reported measures rather than objective performance; future work should incorporate behavioral or accuracy-based metrics and replicate the study with a larger, adequately powered, and stratified sample to support stronger causal claims about discernment.

Beyond these limitations, several directions remain for future studies in this space. In particular, the intervention condition presented alerts, highlights, and dashboard feedback together, so the effects observed here reflect the combined bundle of features rather than any single component. Future studies should experimentally isolate individual elements through ablation studies (e.g., alerts-only vs. highlights-only conditions, or factorial designs) to identify which mechanisms most strongly influence trust, confidence, and comprehension. Building on prior work in news informatics, research could also examine how interactivity, personalization, and real-time feedback shape user empowerment and decision-making. Conducting studies in more naturalistic, "in-the-wild" settings will be crucial for assessing ecological validity and real-world adoption. Developing stronger empirical measures of empowerment can further clarify how such interventions support users' agency when navigating online information.

7. Conclusion

In summary, our between-subjects study ($n = 100$) explored the effectiveness of an interactive browser-based PI tool designed to help readers discern false information in news articles and better understand their online experiences with news consumption. Findings suggest that participants in the intervention condition exhibited higher levels of behavioral engagement (e.g., reading time, clicks, and scrolling) and reported greater confidence in evaluating credibility, sourcing quality, and bias. Visualizing article characteristics and conducting sentiment analysis appeared to provide participants with additional perspective on their reading habits and preferences. Because the intervention was delivered as an integrated bundle of features, these results reflect the combined effects of alerts, highlights, and dashboard feedback rather than any isolated component. While the study was exploratory and slightly underpowered, and not powered for strong confirmatory claims, the patterns observed provide directional evidence on how intervention features shape engagement and discernment. These findings indicate the potential value of complementing algorithmic and fact-checking approaches with user-centered, real-time

reflective cues. More broadly, this work illustrates how personal informatics principles may be extended to news contexts to support more deliberate, reflective engagement with information, emphasizing the human-centered design and evaluation of interactive systems in real-world contexts such as online media and information environments.

Notes

1. https://en.wikipedia.org/wiki/ABCnews.com.co.
2. https://mediabiasfactcheck.com/.
3. https://ground.news/.
4. AllSides is a website that provides balanced news coverage by presenting multiple perspectives on current events. https://www.allsides.com.
5. https://www.nltk.org. VADER is lightweight, fast, and optimized for text sentiment without extensive training.
6. Unlike sentiment analysis, which assigns affective valence (positive, negative, neutral), this approach attends to expressions of subjectivity and opinion-based framing regardless of sentiment.
7. https://libguides.uvic.ca/fakenews/types.
8. https://www.factcheck.org/.
9. https://www.snopes.com/.
10. https://www.politifact.com/.
11. https://library.csi.cuny.edu/c.php?g=619342&p=4376665.
12. https://www.qualtrics.com/.
13. https://otter.ai.

Acknowledgements

We thank Dileep Nimma, Haritha Varkala, and Yongho Cho for their early engineering contributions to the Chrome extension. We are grateful to Dr. Aakash Gautam for his time and insightful manuscript reviews, and to Dr. Leila Barmaki for her guidance during the early conceptualization of this work through her course. We also appreciate the feedback and support from the dissertation committee members of Prerana Khatiwada—Dr. Kathleen McCoy, Dr. Amo Tong, and Dr. Benjamin E. Bagozzi, whose insights helped strengthen this work. Finally, we thank the members of the Sensify Lab for their valuable feedback and discussions throughout the development of this project.

Ethical approval

All procedures involving human participants in this study were reviewed and approved by the Institutional Review Board (IRB) at the University of Delaware (Protocol #1871618-8). Informed consent was obtained from all individual participants included in this study. This approval falls under the project titled "Community Communications: Understanding Online News Media Consumption Behaviors, the Influence of Accuracy Prompts, and Similar Interventions," led by Principal Investigator Dr. Matthew Louis Mauriello.

Authors' contributions

CRediT: **Prerana Khatiwada**: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing; **Nabiha Syed**: Formal analysis, Software, Writing – review & editing; **Luke Halko**: Conceptualization, Software, Writing – review & editing; **Ashrey Mahesh**: Data curation, Formal analysis, Software, Validation, Writing – review & editing; **Ricky Kiamilev**: Formal analysis, Validation, Writing – review & editing; **Antonia Vazquez**: Formal analysis, Validation, Writing – review & editing; **Matthew Louis Mauriello**: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors confirm that they have read and agree to comply with the Taylor & Francis policy on the use of Artificial Intelligence (AI)-assisted technologies. We used ChatGPT, Grammarly, and Overleaf's autocorrect to refine the text, improve grammar and clarity, and suggest improvements to the article's flow, organization, and formatting. All outputs were reviewed and revised by the authors, who take full responsibility for the final content.

Funding

This work was supported by the Artificial Intelligence Center of Excellence (AICoE) at the University of Delaware, which provided partial support for Nabiha Syed's summer research internship through its seed grant program.

ORCID

Prerana Khatiwada http://orcid.org/0009-0008-6965-9504
Nabiha Syed http://orcid.org/0009-0001-1171-2586
Luke Halko http://orcid.org/0009-0001-0150-7706
Ashrey Mahesh http://orcid.org/0009-0007-5652-3694
Ricky Kiamilev http://orcid.org/0009-0005-0256-6213
Antonia Vazquez http://orcid.org/0009-0002-9586-3281
Matthew Louis Mauriello http://orcid.org/0000-0001-5359-6520

Data availability statement

The data for this study are not publicly available due to IRB and privacy restrictions. All data is stored securely on university systems. De-identified data may be shared upon request through a Data Usage Agreement (DUA) via the University of Delaware Office of Research.

References

Adjin-Tettey, T. D. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent Arts & Humanities*, *9*(1), 2037229. https://doi.org/10.1080/23311983.2022.2037229

Ahlers, D. (2006). News consumption and the new electronic media. *Harvard International Journal of Press/Politics*, *11*(1), 29–52. https://doi.org/10.1177/1081180X05284317

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Arapakis, I., Lalmas, M., Cambazoglu, B. B., Marcos, M.-C., & Jose, J. M. (2014). User engagement in online news: Under the scope of sentiment, interest, affect, and gaze. *Journal of the Association for Information Science and Technology*, *65*(10), 1988–2005. https://doi.org/10.1002/asi.23096

Arsenault, A., & Castells, M. (2006). Conquering the minds, conquering Iraq: The social production of misinformation in the United States–A case study. *Information, Communication & Society*, *9*(3), 284–307. https://doi.org/10.1080/13691180600751256

Asad, F., & Parket, O. (2025). Traditional media vs. new media: Analyzing news consumption patterns in a contemporary media landscape. https://doi.org/10.13140/RG.2.2.20126.24649

Avram, M., Micallef, N., Patil, S., & Menczer, F. (2020). Exposure to social engagement metrics increases vulnerability to misinformation. *Harvard Kennedy School Misinformation Review*, *1*(5). https://doi.org/10.37016/mr-2020-033

Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). *Predicting factuality of reporting and bias of news media sources* [Paper presentation]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 3528–3539). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1389

Barthel, M., Mitchell, A., & Holcomb, J. (2016). *Many Americans believe fake news is sowing confusion*. Pew Research Center. https://www.pewresearch.org/journalism/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/ Pew Research Center report.

Bentley, F., Quehl, K., Wirfs-Brock, J., & Bica, M. (2019). *Understanding online news behaviors* [Paper presentation]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Association for Computational Linguistics. https://doi.org/10.1145/3290605.3300820

Bernecker, S., Flowerree, A. K., & Grundmann, T. (2021). *The epistemology of fake news*. Oxford University Press. https://doi.org/10.1093/oso/9780198863977.001.0001

Bhuiyan, M. D. M., Whitley, H., Horning, M., Lee, S. W., & Mitra, T. (2021a). Designing transparency cues in online news platforms to promote trust: Journalists' & consumers' perspectives. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–31. https://doi.org/10.1145/3479539

Bhuiyan, M. M., Horning, M., Lee, S. W., & Mitra, T. (2021b). NudgeCred: Supporting news credibility assessment on social media through nudges. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–30. https://doi.org/10.1145/3479571

Boh Podgornik, B., Dolničar, D., Šorgo, A., & Bartol, T. (2016). Development, testing, and validation of an information literacy test (ILT) for higher education. *Journal of the Association for Information Science and Technology*, *67*(10), 2420–2436. https://doi.org/10.1002/asi.23586

Braun, V., & Clarke, V. (2012). Thematic analysis. In Harris C., Paul M. C., Debra L. L., Panter A. T., David R., & Kenneth J. S. (Eds.), *APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association. https://doi.org/10.1037/13620-004

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, *8*(1), 108–117. https://doi.org/10.1016/j.jarmac.2018.09.005

Burgers, C., & Brugman, B. C. (2022). How satirical news impacts affective responses, learning, and persuasion: A three-level random-effects meta-analysis. *Communication Research*, *49*(7), 966–993. https://doi.org/10.1177/009365022110321

C2PA. (2025). *Coalition for content provenance and authenticity (C2PA)*. https://c2pa.org/.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, *42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Chowdhury, F. A., Saha, D., Hasan, M. R., Saha, K., & Mueen, A. (2022). *Examining factors associated with Twitter account suspension following the 2020 U.S. presidential election* [Paper presentation]. Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (Virtual Event, Netherlands) (ASONAM '21). Association for Computing Machinery, New York, NY, USA (pp. 607–612). ACM. https://doi.org/10.1145/3487351.3492715

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587 Originally published in 1988.

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, *42*(2), 214–257. https://doi.org/10.1598/RRQ.42.2.2

Collier, D., & Mahoney, J. (1996). Insights and pitfalls: Selection bias in qualitative research. *World Politics*, *49*(1), 56–91. https://doi.org/10.1353/wp.1996.0023

Da San Martino, G., Shaar, S., Zhang, Y., Yu, S., Barrón-Cedeño, A., & Nakov, P. (2020). *Prta: A system to support the analysis of propaganda techniques in the news* [Paper presentation]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (pp. 287–293). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-demos.32

Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & De Alfaro, L. (2018). *Automatic online fake news detection combining content and social signals* [Paper presentation]. 2018 22nd Conference of Open Innovations Association (FRUCT) (pp. 272–279). IEEE. https://doi.org/10.23919/FRUCT.2018.8468301

Downey, B. (2015). The looking glass self and deliberation bias in qualitative interviews. *Sociological Spectrum*, *35*(6), 534–551. https://doi.org/10.1080/02732173.2015.1064803

Eisenberg, M. B. (2008). Information literacy: Essential skills for the information age. *DESIDOC Journal of Library & Information Technology*, *28*(2), 39–47. https://doi.org/10.14429/djlit.28.2.166

Epstein, D. A., Jacobson, B. H., Bales, E., McDonald, D. W., & Munson, S. A. (2015a). *From "nobody cares" to way to go!" A design framework for social sharing in personal informatics* [Paper presentation]. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 1622–1636). ACM. https://doi.org/10.1145/2675133.2675135

Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015b). *A lived informatics model of personal informatics* [Paper presentation]. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 731–742). ACM. https://doi.org/10.1145/2750858.2804250

Epstein, Z., Pennycook, G., & Rand, D. (2020). *Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources* [Paper presentation]. Proceedings of

the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20) (pp. 1–11). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376232

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feldman, L. (2011). The opinion factor: The effects of opinionated news on information processing and attitude change. *Political Communication*, *28*(2), 163–181. https://doi.org/10.1080/10584609.2011.565014

Feng, K. J. K., Ritchie, N., Blumenthal, P., Parsons, A., & Zhang, A. X. (2023). Examining the impact of provenance-enabled media on trust and accuracy perceptions. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW2), 1–42. https://doi.org/10.1145/3610061

Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., & Moran, S. (2018). *Falling for fake news: Investigating the consumption of news via social media* [Paper presentation]. CHI Conference on Human Factors in Computing Systems (pp. 1–10). ACM. https://doi.org/10.1145/3173574.3173950

Galan, L., Osserman, J., Parker, T., & Taylor, M. (2019). *How young people consume news and the implications for mainstream media.* A report by Flamingo commissioned by the Reuters Institute for the Study of Journalism. https://www.digitalnewsreport.org/publications/2019/young-people-consume-news/

Garimella, K., Smith, T., Weiss, R., & West, R. (2021). Political polarization in online news consumption. *Proceedings of the International AAAI Conference on Web and Social Media*, *15*(1), 152–162. https://doi.org/10.1609/icwsm.v15i1.18049

Garrett, R. K. (2009). Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, *59*(4), 676–699. https://doi.org/10.1111/j.1460-2466.2009.01452.x

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. (2009). *Statistical Science*, *24*(3), 343–360. https://doi.org/10.1214/09-STS301

Gaziano, C., & McGrath, K. (1986). Measuring the concept of credibility. *Journalism Quarterly*, *63*(3), 451–462. https://doi.org/10.1177/107769908606300301

Grieve, J., & Woodfield, H. (2023). *The language of fake news*. Cambridge University Press. https://doi.org/10.1017/9781009349161

Grinberg, N. (2018). *Identifying modes of user engagement with online news and their relationship to information gain in text* [Paper presentation]. Proceedings of the 2018 World Wide Web Conference (Lyon, France) (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva (pp. 1745–1754). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3178876.3186180

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, *10*(2022), 178–206. https://doi.org/10.1162/tacl_a_00454

Gupta, A., Kumar, N., Prabhat, P., Gupta, R., Tanwar, S., Sharma, G., Bokoro, P. N., & Sharma, R. (2022). Combating fake news: Stakeholder interventions and potential solutions. *IEEE Access*, *10*(2022), 78268–78289. https://doi.org/10.1109/ACCESS.2022.3193670

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Hartwig, K., Doell, F., & Reuter, C. (2024). The landscape of user-centered misinformation interventions – A systematic literature review. *ACM Computing Surveys*, *56*(11), 1–36. https://doi.org/10.1145/3674724

Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). *Toward automated fact-checking: detecting check-worthy factual claims by ClaimBuster* [Paper presentation]. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Halifax, NS, Canada) (KDD '17) (pp. 1803–1812). Association for Computing Machinery. https://doi.org/10.1145/3097983.3098131

Hatem, G., Zeidan, J., Goossens, M., & Moreira, C. (2022). Normality testing methods and the importance of skewness and kurtosis in statistical analysis. *BAU Journal – Science and Technology*, *3*(2), 7. https://doi.org/10.54729/KTPE9512

Hawa, S., Lobo, L., Dogra, U., & Kamble, V. (2021). *Combating misinformation dissemination through verification and content driven recommendation* [Paper presentation]. 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE. https://doi.org/10.1109/ICICV50876.2021.9388406

Iizuka, R., Toriumi, F., Nishiguchi, M., Takano, M., & Yoshida, M. (2022). Impact of correcting misinformation on social disruption. *PloS One*, *17*(4), e0265734. https://doi.org/10.1371/journal.pone.0265734

Jaeger, P. T., & Taylor, N. G. (2021). Arsenals of lifelong information literacy: Educating users to navigate political and current events information in world of ever-evolving misinformation. *Library Quarterly*, *91*(1), 19–31. https://doi.org/10.1086/711632

Jahanbakhsh, F., & Karger, D. R. (2024). *A browser extension for in-place signaling and assessment of misinformation* [Paper presentation]. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642473

Jahanbakhsh, F., Zhang, A. X., & Karger, D. R. (2022). Leveraging structured trusted-peer assessments to combat misinformation. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–40. https://doi.org/10.1145/3555637

Jahn, L., Rendsvig, R. K., Flammini, A., Menczer, F., & Hendricks, V. F. (2025). A perspective on friction interventions to curb the spread of misinformation. *Npj Complexity*, *2*(1), 31. https://doi.org/10.1038/s44260-025-00051-1

Jones, A., & Kang, J. (2020). Media technology shifts: Exploring millennial consumers' fashion-information-seeking behaviors and motivations. *Canadian Journal of Administrative Sciences/Revue Canadienne Des Sciences de L'Administration*, *37*(1), 13–29. https://doi.org/10.1002/cjas.1546

Kang, H. (2021). Sample size determination and power analysis using the G* Power software. *Journal of Educational Evaluation for Health Professions*, *18*(2021), 17. https://doi.org/10.1017/S1930297500008640

Kate, S., Ahmer, A., & Tom, W. (2019). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–26. https://doi.org/10.1145/3359229

Khatiwada, P., Halko, L., Syed, N., Mahesh, A., Alvanpour, A., & Mauriello, M. L. (2025). Spotting online news: A mixed method study of online news engagement and perceptions on misinformation interventions. *Proceedings of the ACM on Human-Computer Interaction*, *9*(2), 1–30. https://doi.org/10.1145/3711071

Khatiwada, P., Mumma, I. A. N., Halko, L. U. K. E., Alvanpour, A., & Louis Mauriello, M. (2022). *Toward browser-based interventions to tackle misinformation online.*

Khatiwada, P., Pappu, V., E. Bagozzi, B., & Louis Mauriello, M. (2026). *When AI rewrites the news: How sentiment, framing, and LLM disclosure shape perceptions* [Paper presentation]. Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26). Association for Computing Machinery. https://doi.org/10.1145/3772318.3791527

Kirchner, J., & Reuter, C. (2020). Countering fake news: A comparison of possible solutions regarding user acceptance and effectiveness. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–27. https://doi.org/10.1145/3415211

Kocielnik, R., & Sidorova, N. (2015). Personalized stress management: Enabling stress monitoring with LifelogExplorer. *KI – Künstliche Intelligenz*, *29*(2), 115–122. https://doi.org/10.1007/s13218-015-0348-1

Konstantinou, L., & Karapanos, E. (2025a). *Behavior change interventions combating online misinformation: A scoping review* [Paper presentation]. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713127

Konstantinou, L., & Karapanos, E. (2025b). How do users perceive nudges against online misinformation? A repertory-grid study. *International Journal of Human–Computer Interaction*, *41*(24), 15551–15569. https://doi.org/10.1080/10447318.2025.2499163

Konstantinou, L., Panos, D., & Karapanos, E. (2025). Exploring the design of technology-mediated nudges for online misinformation. *International Journal of Human–Computer Interaction*, *41*(1), 515–542. https://doi.org/10.1080/10447318.2023.2301265

Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., Ali, A., Bak-Coleman, J., Barzilai, S., Basol, M., Berinsky, A. J., Betsch, C., Cook, J., Fazio, L. K., Geers, M., Guess, A. M., Huang, H., Larreguy, H., Maertens, R., … Wineburg, S. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behaviour*, *8*(6), 1044–1052. https://doi.org/10.1038/s41562-024-01881-0

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science (New York, N.Y.)*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lee, H., Kang, S., & Lee, U. (2022). Understanding privacy risks and perceived benefits in open dataset collection for mobile affective computing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*(2), 1–26. https://doi.org/10.1145/3534623

Lee, S., & Suh, K.-S. (2022). Deliberate news consumption through the quantified self and the self-regulatory process. *Information & Management*, *59*(2), 103585. https://doi.org/10.1016/j.im.2021.103585

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest: Journal of the American Psychological Society*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Li, I., Dey, A., & Forlizzi, J. (2010a). A stage-based model of personal informatics systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 557–566). ACM.

Li, I., Dey, A., & Forlizzi, J. (2010b). *A stage-based model of personal informatics systems* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Atlanta, Georgia, USA) (CHI '10) (pp. 557–566). Association for Computing Machinery. https://doi.org/10.1145/1753326.1753409

Li, I., Medynskiy, Y., Froehlich, J., & Larsen, J. (2012). *Personal informatics in practice: Improving quality of life through data* [Paper presentation]. CHI'12 Extended Abstracts on Human Factors in Computing Systems (pp. 2799–2802). Association for Computing Machinery. https://doi.org/10.1145/2212776.2212724

Li, J. (2025). Not all skepticism is "healthy" skepticism: Theorizing accuracy-and identity-motivated skepticism toward social media misinformation. *New Media & Society*, *27*(1), 522–544. https://doi.org/10.1177/14614448231179941

Loerakker, M. B., Niess, J., Bentvelzen, M., & Woźniak, P. W. (2024). Designing data visualisations for self-compassion in personal informatics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *7*(4), 1–22. https://doi.org/10.1145/3631448

Lu, Z., Li, P., Wang, W., & Yin, M. (2022). The effects of AI-based credibility indicators on the detection and spread of misinformation under social influence. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–27. https://doi.org/10.1145/3555562

Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., & Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences*, *118*(23), e2019527118. https://doi.org/10.1073/pnas.2019527118

Ma, Y., Sarsenbayeva, Z., Knibbe, J., & Goncalves, J. (2025). Exploring the effects of location information on perceptions of news credibility and sharing intention. *International Journal of Human-Computer Studies*, *193*(C), 103378. https://doi.org/10.1016/j.ijhcs.2024.103378

Martel, C., & Rand, D. G. (2023). Misinformation warning labels are widely effective: A review of warning effects and their moderating features. *Current Opinion in Psychology*, *54*(2023), 101710. https://doi.org/10.1016/j.copsyc.2023.101710

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

Media Literacy Now. (2025). *What is media literacy? What is digital literacy? What is information literacy? What is news literacy? What is digital citizenship?* https://medialiteracynow.org/what-is-media-literacy-what-is-digital-literacy-what-is-information-literacy-what-is-news-literacy-what-is-digital-citizenship/

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. https://doi.org/10.1002/asi.20672

Micallef, N., Avram, M., Menczer, F., & Patil, S. (2021). Fakey: A game intervention to improve news literacy on social media. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–27. https://doi.org/10.1145/3449080

Mitchell, A., Gottfried, J., Barthel, M., Sumida, N., & Mitchell, A. (2018). *Distinguishing between factual and opinion statements in the news*. Pew Research Center's Journalism Project. https://www.pewresearch.org/journalism/2018/06/18/distinguishing-between-factual-and-opinion-statements-in-the-news/

Mourão, R. R., & Robertson, C. T. (2019). Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information. *Journalism Studies*, *20*(14), 2077–2095. https://doi.org/10.1080/1461670X.2019.1566871

Munzner, T. (2025). *Visualization analysis and design* [Paper presentation]. Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Courses (SIGGRAPH Courses '25). Association for Computing Machinery. https://doi.org/10.1145/3721241.3733989

Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, *20*(11), 4366–4383. https://doi.org/10.1177/1461444818773059

Nagel, M.-T., Schäfer, S., Zlatkin-Troitschanskaia, O., Schemer, C., Maurer, M., Molerov, D., Schmidt, S., & Brückner, S. (2020). How do university students' web search behavior, website characteristics, and the interaction of both influence students' critical online reasoning? *Frontiers in Education*, *5*, 565062. https://doi.org/10.3389/feduc.2020.565062

Net Promoter Score. (2018). Net promoter score. *Work*, *10*(10), 10. https://cdn.gnapartners.com/production/pdfs/ga_nps_final_digital.pdf

Noh, Y.-G., Han, M., Jeon, J., & Hong, J.-H. (2025). *BIASsist: Empowering news readers via bias identification, explanation, and neutralization* [Paper presentation]. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713531

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books. https://dl.icdst.org/pdfs/files4/4bb8d08a9b309df7d86e62ec4056ceef.pdf Accessed: 2026-04-09.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

O'Connor, C., & Weatherall, J. O. (2019). *The misinformation age: How false beliefs spread*. Yale University Press.

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review (2020)*, *1*(4). https://doi.org/10.37016/mr-2020-024

Omar, B., Ng, S. K., & Ismail, N. (2023). Effects of message, medium, and motivational factors on news engagement and mobile news consumption: Evidence from Malaysia. *Online Journal of Communication and Media Technologies*, *13*(3), e202325. https://doi.org/10.30935/ojcmt/13116

Oppenlaender, J., Abbas, T., & Gadiraju, U. (2024). The state of pilot study reporting in crowdsourcing: A reflection on best practices and guidelines. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–45. https://doi.org/10.1145/3641023

Oviatt, S. (2006). *Human-centered design meets cognitive load theory: Designing interfaces that help people think* [Paper presentation]. Proceedings of the 14th ACM International Conference on Multimedia (Santa Barbara, CA, USA) (MM '06) (pp. 871–880). Association for Computing Machinery. https://doi.org/10.1145/1180639.1180831

Palen, L., & Liu, S. B. (2007). *Citizen communications in crisis: Anticipating a future of ICT-supported public participation* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 727–736). Association for Computing Machinery. https://doi.org/10.1145/1240624.1240736

Patel, S. S., Moncayo, O. E., Conroy, K. M., Jordan, D., & Erickson, T. B. (2020). The landscape of disinformation on health crisis communication during the COVID-19 pandemic in Ukraine: Hybrid warfare tactics, fake media news and review of evidence. *JCOM, Journal of Science Communication*, *19*(5), AO2. https://doi.org/10.22323/2.19050202

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*(2019), 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200. https://doi.org/10.1111/jopy.12476

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), 2333. https://doi.org/10.1038/s41467-022-30073-5

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Pointon, M., Walton, G., Turner, M., Lackenby, M., Barker, J., & Wilkinson, A. (2023). Information discernment and online reading behaviour: An experiment. *Online Information Review*, *47*(3), 522–549. https://doi.org/10.1108/OIR-02-2021-0101

Radinsky, J., & Tabak, I. (2022). Data practices during COVID: Everyday sensemaking in a high-stakes information ecology. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, *53*(5), 1221–1243. https://doi.org/10.1111/bjet.13252

Rapp, D. N., & Salovich, N. A. (2018). Can't we just disregard fake news? The consequences of exposure to inaccurate information. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 232–239. https://doi.org/10.1177/2372732218785193

Reuter, C., Lee Hughes, A., & Buntain, C. (2025). Combating information warfare: State and trends in user-centred countermeasures against fake news and misinformation. *Behaviour & Information Technology*, *44*(13), 3348–3361. https://doi.org/10.1080/0144929X.2024.2442486

Ross, R. M., Rand, D. G., & Pennycook, G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision Making*, *16*(2), 484–504. https://doi.org/10.3352/jeehp.2021.18.17

Russo, A., Jankowski, A., Beene, S., & Townsend, L. (2019). Strategic source evaluation: Addressing the container conundrum. *Reference Services Review*, *47*(3), 294–313. https://doi.org/10.1108/RSR-04-2019-0024

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*(2023), 110273. https://doi.org/10.1016/j.knosys.2023.110273

Sallami, D., Ben Salem, R., & Aïmeur, E. (2023). *Trust-based recommender system for fake news mitigation* [Paper presentation]. Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and

Personalization (Limassol, Cyprus) (UMAP '23 Adjunct) (pp. 104–109). Association for Computing Machinery. https://doi.org/10.1145/3563359.3597395

Schwaderer, C. (2025). When crisis conversations converge: How government messaging and online sense-making align during crises. *Risk, Hazards & Crisis in Public Policy*, *16*(4), e70042. https://doi.org/10.1002/rhc3.70042

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology*, *10*(3), 1–42. https://doi.org/10.1145/3305260

Sidnam-Mauch, E., Ivancsics, B., Monroe, A., Washington, E., Francis Ii, E., Caine, K., Bonneau, J., & McGregor, S. E. (2022). Usable cryptographic provenance: A proactive complement to fact-checking for mitigating misinformation. In *Proceedings of the International AAAI Conference on Weblogs and Social Media* (Vol. 16). AAAI Press. https://doi.org/10.36190/2022.55

Singh, P., & Singh, A. K. (2025). *The digital revolution: Exploring new media in the 21st century*. Crown Publishing.

Skillman, M., Cross-Barnet, C., Friedman Singer, R., Rotondo, C., Ruiz, S., & Moiduddin, A. (2019). A framework for rigorous qualitative research as a component of mixed method rapid-cycle evaluation. *Qualitative Health Research*, *29*(2), 279–289. https://doi.org/10.1177/1049732318795675

Spezzano, F., Shrestha, A., Fails, J. A., & Stone, B. W. (2021). That's fake news! Reliability of news when provided title, image, source bias & full article. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–19. https://doi.org/10.1145/3449183

Starbird, K. (2017). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 230–239. https://doi.org/10.1609/icwsm.v11i1.14878

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, *60*(3), 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x

Sundar, S. S., Jia, H., Bellur, S., Oh, J., & Kim, H.-S. (2022). *News informatics: Engaging individuals with data-rich news content through interactivity in source, medium, and message* [Paper presentation]. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery. https://doi.org/10.1145/3491102.3502207

Truszkowski, W., Rouff, C. A., Bailin, S. C., & Rilee, M. (2005). Progressive autonomy – An incremental agent-based approach. In *Proceedings of the International Conference on Software Engineering Research and Practice (SERP'05)* (Vol. 1). Association for Computing Machinery.

Ulu, H. (2019). Examining the relationships between the attitudes towards reading and reading habits, meta-cognitive awarenesses of reading strategies and critical thinking tendencies of pre-service teachers. *International Journal of Contemporary Educational Research, 6*(1), 169–182. https://api.semanticscholar.org/CorpusID:197687237

Van Der Meer, T. G., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. *Journalism Studies*, *24*(6), 803–823. https://doi.org/10.1080/1461670X.2023.2187652

VanTol, J., Nikitin, M., Helzerman, N., Allison, M., & Spradling, M. (2022). *OpenLabel*. An Open-Source Media Labeling Web Browser Extension.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science (New York, N.Y.)*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vraga, E. K., Bode, L., & Tully, M. (2022). Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research*, *49*(2), 245–267. https://doi.org/10.1177/0093650219898094

Vraga, E., Tully, M., Kotcher, J. E., Smithson, A.-B., & Broeckelman-Post, M. (2015). A multi-dimensional approach to measuring news media literacy. *Journal of Media Literacy Education*, *7*(3), 41–53. https://doi.org/10.23860/jmle-7-3-4

Walker, M., & Matsa, K. E. (2021). *News consumption across social media*. Pew Centre Research. https://www.jstor.org/stable/resrep63516

Wang, J. S., Haider, S., Tohidi, A., Gupta, A., Zhang, Y., Callison-Burch, C., Rothschild, D., & Watts, D. J. (2025). *Media bias detector: Designing and implementing a tool for real-time selection and framing bias analysis in news coverage* [Paper presentation]. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713716

Wang, Q., Khatiwada, P., Chouhan, A., Mahesh, A., Mwaria, J., Tran, D. D., Barner, K. E., & Mauriello, M. L. (2026). *"The explanation makes sense": An empirical study on LLM performance in news classification and its influence on judgment in human-AI collaborative annotation*. arXiv preprint arXiv:2602.19690 (2026).

Warner, M., Strohmayer, A., Higgs, M., & Coventry, L. (2025). A critical reflection on the use of toxicity detection algorithms in proactive content moderation systems. *International Journal of Human-Computer Studies*, *198*, 103468. https://doi.org/10.1016/j.ijhcs.2025.103468

Wineburg, S., & McGrew, S. (2019). Lateral reading: Reading less and learning more when evaluating digital information. *Teachers College Record*, *121*(11), 1–40. https://doi.org/10.2139/ssrn.3048994

Wirfs-Brock, J., & Quehl, K. (2019). News from the background to the foreground: How people use technology to manage media transitions: A study of technology-mediated news behaviors in a hyper-connected world. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(3), 1–19. https://doi.org/10.1145/3351268

Yaqub, W., Kakhidze, O., Brockman, M. L., Memon, N., & Patil, S. (2020). *Effects of credibility indicators on social media news sharing intent* [Paper presentation]. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–14). ACM. https://doi.org/10.1145/3313831.3376213

Zeng, L., Starbird, K., & Spiro, E. S. (2016). *Rumors at the speed of light? Modeling the rate of rumor transmission during crisis* [Paper presentation]. 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 1969–1978). IEEE. https://doi.org/10.1109/HICSS.2016.248

Zhang, A. X., Ranganathan, A., Metz, S. E., Appling, S., Sehat, C. M., Gilmore, N., Adams, N. B., Vincent, E., Lee, J., & Robbins, M. (2018). *A structured response to misinformation: Defining and annotating credibility indicators in news articles* [Paper presentation]. Companion Proceedings of the Web Conference 2018 (pp. 603–612). ACM. https://doi.org/10.1145/3184558.3188731

Zhou, X., Pan, W., & Jiang, S. (2025). Combating health misinformation through fact-checking: The role of active versus passive social media involvement. *International Journal of Human–Computer Interaction*, *41*(21), 13723–13736. https://doi.org/10.1080/10447318.2025.2476705

About the authors

Prerana Khatiwada is a doctoral candidate in Computer and Information Sciences at the University of Delaware. Her research focuses on human–computer interaction, media literacy, and human-centered AI, with an emphasis on designing and evaluating behavioral interventions and tools that support critical engagement with digital information.

Nabiha Syed is an alumnus of the University of Delaware and currently works as an Analyst at Deloitte. She is leading the development of internal generative and agentic AI solutions focused on digital web accessibility. Her interests lie at the intersection of human–computer interaction, accessibility, and AI-driven product development.

Luke Halko is a Software Engineer and University of Delaware alumnus who has experience with web development and machine learning. His research interests include human–computer interaction and public policy.

Ashrey Mahesh is a senior undergraduate computer engineering student in the Department of Electrical and Computer Engineering at the University of Delaware. His research interests lie at the intersection of human–computer interaction and generative AI/machine learning.

Ricky Kiamilev is a graduating senior at University of Delaware studying Computer Science and Political Science with a minor in Economics. His research interests include human–computer interaction and media literacy. He has interned at National Education Association and serves in the New York State Assembly, focusing on AI's workforce impact.

Antonia Vázquez is an undergraduate student majoring in finance at the University of Delaware, with research interests in finance and social policy. She is interested in exploring how social structures influence economic and educational outcomes.

Matthew Louis Mauriello is an Assistant Professor in the Department of Computer and Information Sciences at the University of Delaware. His research focuses on human–computer interaction, human-centered AI, and interactive systems for health, education, and sustainability, with an emphasis on designing intelligent systems that amplify human expertise.