ARC Centre of Excellence for **Automated Decision-Making and Society**

Quantifying and Measuring Bias and Engagement in Automated Decision-Making

May 2024

Authors: Damiano Spina, Danula Hettiachchi, & Anthony McCosker

Contributors: Kaixin Ji, Sachin Pathiyan Cherumanal, Weronika Ławejska, Marwah Alaofi, Hmdh Alknjr, Nuha Abu Onq, Flora D Salim, Falk Scholer, Mark Sanderson & Jenny Kennedy

Partner: RMIT ABC Fact Check

Copy editing: Dr Nelson Mok AE (Nevermore Editing)

> ‘This research provided RMIT ABC Fact Check with valuable insights into the relative effectiveness of existing presentation strategies such as the use of veracity indicators and summaries, whilst alerting us to important considerations for content delivery on emergent platforms and the change in audience expectations they bring.’
>
> —Devi Mallal, Media and Research Lead at RMIT ABC Fact Check

Suggested citation

Spina, D., Hettiachchi, D., & McCosker, A. (2024). Quantifying and Measuring Bias and Engagement in Automated Decision-Making. Melbourne: ARC Centre of Excellence for Automated Decision-Making and Society RMIT University. DOI: 10.60836/k8xh-en92

Acknowledgement of Country

In the spirit of reconciliation, we acknowledge the Traditional Custodians of Country throughout Australia and their connections to land, sea and community. We pay our respect to their Elders past, present and future, and extend that respect to all Aboriginal and Torres Strait Islander peoples today.

ARC Acknowledgement

This research was partially funded by the Australian Government through the Australian Research Council’s Centre of Excellence for Automated Decision Making and Society (ADM+S) [CE200100005].

Australian Government
Australian Research Council

CONTENTS

EXECUTIVE SUMMARY

There is no doubt that misinformation has become endemic and threatens the fabric of our internet platform, news services and the social and political institutions that rely on them. Automated moderation and news feed curation systems can help to sift and restrict misinformation, but they cannot correct misinformation once it circulates and takes hold. Consequently, fact-checking efforts manually performed by professionals and independent fact-checking agencies have become essential for maintaining trust in online information flows. However, there is limited guidance as to the most effective means for presenting fact-checking information and how to best inform the widest spectrum of people engaging with online content. Thus, fact-checkers need more research showing what works when explaining truth from falsehood.

Working with experienced fact-checking professionals, we examined the way people engaged with different types of online fact-checking and established benchmarks for measuring the effectiveness of different techniques and designs. An interdisciplinary team of researchers from the ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S) designed a three-phase study to understand the biases and preferences that people have when they engage with fact-checking material.

The first phase used creative co-design methods with fact-checking professionals and media experts to examine different approaches to presenting fact-checking content. The second phase brought these findings into a user study with 76 crowdworkers to test the efficacy of different design strategies. The third phase used sensing technologies to characterise user behaviours with physiological signals, particularly in relation to cognitive load, affective arousal and valence.

The findings painted a detailed picture of the importance of fact-checking presentation design and the biases and contexts that people bring to their engagement with those designs. This entails careful consideration of the content, structure and presentation of information.

We found that crucial presentation elements include a clear and consistent veracity indicator and a dedicated summary at the start of the report. Credibility is established through the clear citation of sources, and this helps to build trust. Personalisation is also vital for reaching target audiences, and it is one of the most difficult elements to achieve if restricted to text and image reports. Finally, conducting user studies in controlled laboratory settings can help establish reliable physiological signals that characterise cognitive load and affective responses during different stages of engagement with both screen-based and audio-only communication channels.

Ultimately, a better understanding of engagement with misinformation and fact-checking will help improve search engines, intelligent assistants and large language model–based conversational agents in providing not only relevant and useful information but also access to trustworthy and reliable information. The design, evaluation and optimisation of these automated decision-making systems hinges on defining frameworks that model user interactions. Rather than simplifying user behaviours using metrics like the relevance of delivered items, the way forward lies in augmenting these frameworks to encompass new dimensions such as fairness, trustworthiness and inclusion alongside traditional quality definitions.

INTRODUCTION

Navigating online news, media and information has never been more difficult despite the curatorial work of algorithmic news feeds, search engines and other automated decision systems. In this project, we focused on strategies for effective presentation of fact-checking content (e.g., reports including more structured content or more information about the the author to enhance credibility) as part of a broader study into quantifying and measuring bias and engagement in automated decision-making systems.

We aimed to explore the following research questions:

- How do users perceive fairness, bias and trust, and how can these perceptions be measured effectively?
- Can bias be measured by observing users' interactions with search engines or intelligent assistants?
- To what extent can sensors in wearable devices and interactions inform the measurement of bias and engagement?

To address these questions in the context of fact-checking, we developed a three-phase mixed-methods approach: participatory research (Phase 1), online user studies via crowdsourcing (Phase 2) and controlled laboratory user studies (Phase 3).

This report summarises the methodologies, findings and implications of the three phases. We hope that this report will provide readers with a clear understanding of the key contributions made within the Quantifying and Measuring Bias and Engagement project. For a more comprehensive insight into the project, we invite you to read the research outputs listed in the References section.

PHASE 1: CO-DESIGNING PRESENTATION STRATEGIES

With the rapid growth of online misinformation, it is crucial to have reliable fact-checking methods. Recent research on finding checkworthy claims and automated fact-checking has made significant advancements. However, there is limited guidance regarding the presentation of fact-checking content to effectively convey this verified information to users. Thus, in Phase 1, we address this research gap by exploring the critical design elements in fact-checking reports.

METHODOLOGY

We conducted a hands-on design workshop to develop a set of fact-checked content presentation strategies for screen interfaces [1]. Through the workshop, we also aimed to understand the process of delivering fact-checked content through voice-only interaction and to develop relevant basic presentation strategies. The 4-hour workshop involved 10 participants representing fact-checking professionals, communication experts and researchers, who were divided in three groups.

The workshop included five hands-on activities and discussions. As shown in Figure 1, in the first activity, participants were asked to identify and list different fact-checking elements in eight example fact-checking reports extracted from eight different global fact-checking websites. In the second activity, participants discussed, within their group, the importance of the elements they noted in the first activity. In the third activity, participants developed fact-checking presentation strategies for screens. A presentation strategy may consist of elements identified in the previous activity or adapted versions of those elements. Participants used Lego blocks, using colours to represent the element type, vertical height to represent the expected attention/intensity level, and the occupied surface area of the block to represent the space requirement on the website.

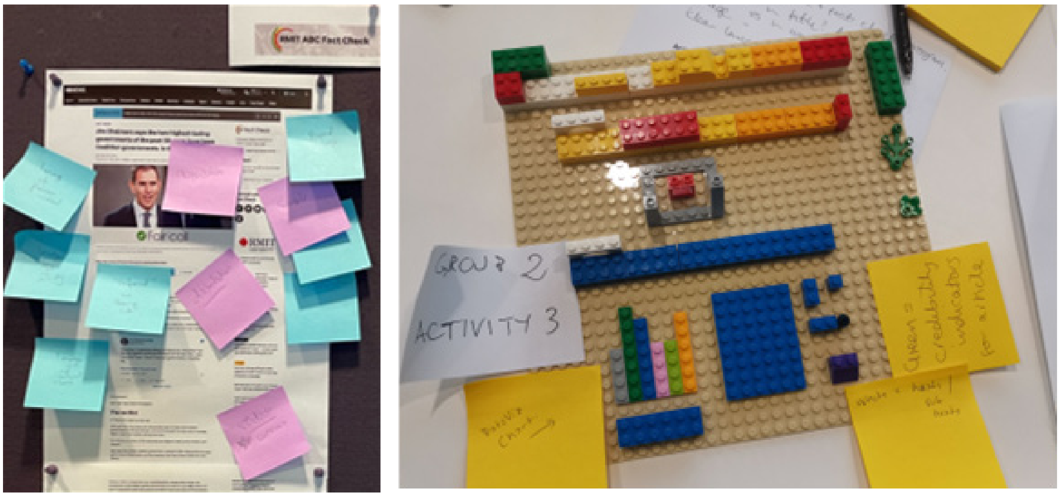


Figure 1: Hands on activities as part of design workshop.

To foster in-depth discussions about how fact-checking presentations could affect different individuals, we developed six personas that were provided to groups in the next activities. Informed by the Australian Digital Inclusion Index, each persona included four characteristics in addition to basic demographic information: technology attitude, basic technology skills, advanced technology skills and English proficiency.

In the fourth activity, each group picked two personas with characteristics contrasting with respect to the personas' interest (or willingness) to consume fact-checking content. Participants then created a scenario using the persona and a topic from the fact-checking reports previously shown to them . The scenario could include details such as news consumption channels, motivation to use the fact-checking website and so on. Afterwards, participants modified the fact-checking presentation they developed in the third activity for each persona and scenario.

In the final activity, participants developed a fact-checking presentation strategy for voice interaction . Similar to the third activity, participants used Lego blocks to represent the element type, level of expected attention/intensity, and the turns in conversation.

Throughout the workshop, participants shared and discussed their thoughts and findings with other groups and facilitators in debriefing sessions between activities. For our qualitative analysis, we focused mainly on the debriefing sessions.

MAIN FINDINGS

As summarised in Figure 2, and synthesised from workshop discussions, we report key considerations for designing fact-checking reports for screens and voice interfaces. For instance, fostering perceived trust towards the fact-checking organisation and the author was perceived as an important design requirement in both screen and audio interactions.

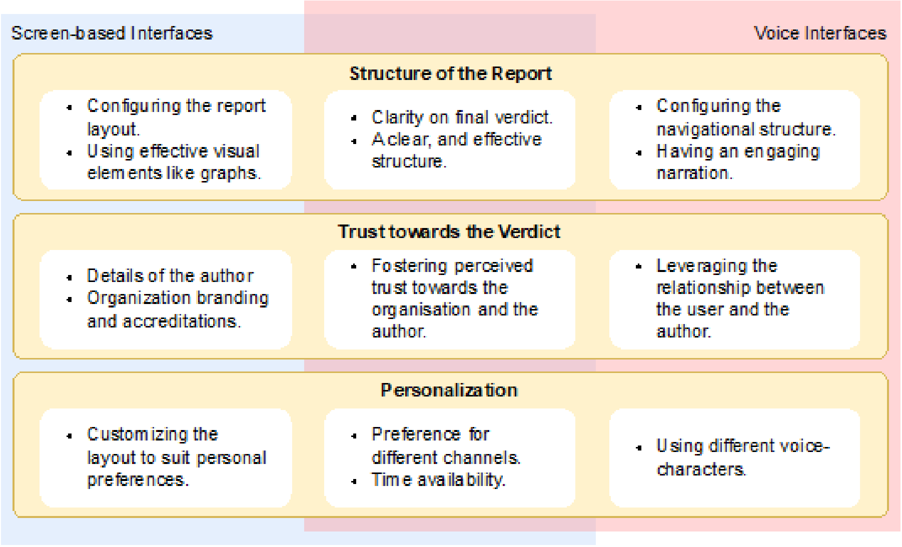

Screen-based Interfaces
Voice Interfaces
Structure of the Report
Configuring the report layout.
Using effective visual elements like graphs.
Clarity on final verdict.
A clear, and effective structure.
Configuring the navigational structure.
Having an engaging narration.
Trust towards the Verdict
Details of the author
Organization branding and accreditations.
Fostering perceived trust towards the organisation and the author.
Leveraging the relationship between the user and the author.
Personalization
Customizing the layout to suit personal preferences.
Preference for different channels.
Time availability.
Using different voice-characters.

Figure 2: Themes identified as the result of the workshop with practitioners.

IMPLICATIONS

Our workshop findings highlight that **there is no one-fits-all presentation strategy for fact-checking content**. We believe fact-checking organisations need to put more emphasis on specific factors, such as a consistent layout or engaging narration when delivering reports across screen-based and voice interfaces. However, **there is also an expectation and need for personalisation**, and this means that, for fact-checking content to be successful, it must be accessible and enticing to different audience segments. Similarly, our persona and scenario-based exercise discussions showed that user attributes could be used to inform the design requirements of fact-checking reports.

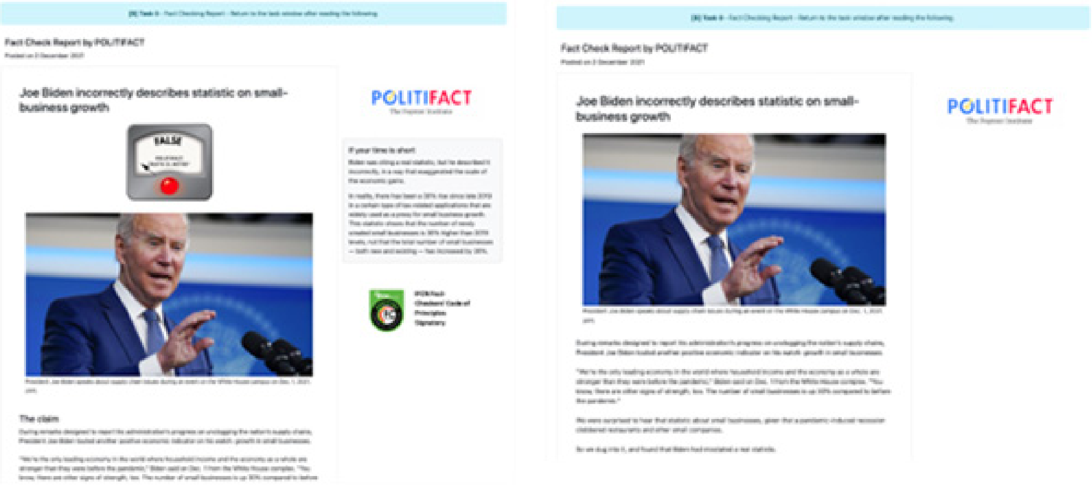
PHASE 2: EVALUATING PRESENTATION STRATEGIES VIA ONLINE USER STUDIES

Drawing on the outcomes of the Phase 1 workshop, we then conducted an online study with 76 crowdworkers to better understand the efficacy of different design strategies for improving the credibility and overall presentation of fact-checking articles.

METHODOLOGY

We identified six key design elements of fact-checking reports based on workshop outcomes and prior work. By adding credibility elements such as accreditation, additional author details and sources, we aimed to improve user trust in fact-checking reports. With presentation elements, we aimed to make it easy for users to navigate reports and find the relevant information they seek.

While our collaborative workshops focused on both screen and voice interfaces, our first attempt at experimental implementation and evaluation in Phase 2 focused on screen-based interfaces. Currently, screen-based implementations (e.g., fact-checking websites) are predominantly used to disseminate fact-checking reports. In our study, we used eight fact-checking articles: four from PolitiFact and four from RMIT ABC Fact Check. All fact-checking reports were of statements related to the economy, environment or demographics—with no subjective/ideological statements or statements directed at others—made by politicians at the national level from two major parties in the US or Australia. We balanced the dataset in terms of veracity and the speaker’s political party. Based on the credibility and presentation features, we developed four study conditions: baseline, improved presentation, improved credibility, and improved presentation and credibility. As shown in Figure 3, our experimental setup allowed us to dynamically adjust the presentation of articles by adding necessary components according to the study condition.



Fact Check Report by POLITIFACT

Joe Biden incorrectly describes statistic on small-business growth

FALSE
POLITIFACT
If your time is short
The claim
Fact Check Report by POLITIFACT
Joe Biden incorrectly describes statistic on small-business growth
POLITIFACT

Figure 3: Fact checked presentation strategies used in the online study.

We used a within-subject experimental design, whereby all participants read and rated all eight fact-checking reports generated using the four experimental conditions (i.e., two reports per condition) . Each article was randomly assigned to one of the conditions. Each participant who launched the task from the Amazon Mechanical Turk platform completed three stages. Stage 1 comprised two questionnaires: the Conspiracy Mentality Questionnaire, which measures susceptibility to conspiracy, and the Credibility of Science Scale, which measures how individuals perceive scientific information and research as trustworthy, reliable and valid. Stage 2 included eight tasks, where users were asked to read a fact-checking report and answer a set of questions. All user engagements (e.g., scrolls and clicks) within the report page were recorded and tracked. Finally, in Stage 3, participants completed a demographics survey that asked their age, gender, highest level of education, political inclination and general questions about their experience with fact-checking.

MAIN FINDINGS

Our results showed that both presentation and credibility improvements helped users to be more accurate when interpreting fact-checking reports. In our within-subject experiment, each participant rated articles related to all four study conditions. A non-parametric paired Wilcoxon signed rank test indicated statistically significant higher task accuracy (Z=148.5, p<0.05) in conditions with presentation improvements (M=89.1) compared to conditions without presentation improvements (M=82.9). Considering credibility improvements, a Wilcoxon signed rank test indicated statistically significant higher task accuracy (Z=104.0, p<0.05) in conditions with credibility improvements (M=89.1) compared to conditions without credibility improvements (M=82.9).

Using a generalised linear model, we further analysed the influence of user attributes on their ability to accurately interpret the verdict of the fact-checking articles. The model indicated three significant predictors. The results indicate that the ability to accurately interpret the fact-checking report increases with age. When considering previous encounters with sharing misinformation, participants who had previously shared information that they later discovered through a social media user to be misinformation had a significantly lower ability to accurately interpret fact-checking reports compared to others who did not previously share misinformation. In addition, an increased belief in science (i.e., a higher score on the Credibility of Science Scale) resulted in more accurate veracity judgements.

IMPLICATIONS

Designing fact-checking reports requires careful consideration of the content, structure and presentation of information. Based on our research, we propose a set of design guidelines for effectively presenting fact-checking reports on screens.

- **Presentation**: Communicating the final verdict in a straightforward manner will help a broader range of users, particularly those not interested in reading the entire article. Two basic steps are having a clear and consistent veracity indicator and phrasing the verdict through the report heading. Including a dedicated summary for each report is also helpful. Considering the main report's structure and text, it should be easy to read and understand. Use clear and concise language and avoid technical jargon or complex sentence structures. Use headings, bullet points and visual aids to break up the text and make it easier to digest.
- **Credibility**: Ensure the report is credible by citing credible sources and providing links to additional information. It is essential to build trust with the user by providing details on who did the fact-checking and how. Information about accreditation and the author are simple additions that can potentially further enhance the perceived credibility of the report. However, as evident from our results, credibility additions can also be detrimental to overall usability, increasing the user's burden and the number of interactions required to find essential information in the report.
- **Personalisation**: Furthermore, understanding the audience is critical for effective presentation. For instance, our results highlight the need for specific interventions and support for individuals with a lower belief in science. Research shows that, while fact-checking reports are effective in curbing misinformation, they can often fail to reach the target audience [8]. As our workshop discussions on fact-checking consumption for various personas suggest, personalisation could be an effective strategy to reach broader audiences. The need for personalisation is further supported by our quantitative observations on how user attributes affect the user's ability to accurately interpret such reports.

PHASE 3: MEASURING BIAS AND ENGAGEMENT IN CONTROLLED LABORATORY USER STUDIES

Information access systems are becoming more complex, and our understanding of user behaviour during information-seeking processes is mainly drawn from qualitative methods, such as observational studies or surveys. Leveraging the advances in sensing technologies, the laboratory user studies carried out during Phase 3 of this project aimed to characterise user behaviours with physiological signals, particularly in relation to cognitive load, affective arousal and valence [3].

METHODOLOGY

We have performed two experiments in controlled laboratory user studies.

The first user study (*N*=7), reported by Ji et al. [4, 5], assessed the robustness and sensitivity of capturing physiological signals across four information-processing activities (IPAs; read, listen, speak and write) using multiple sensors simultaneously to collect electrodermal activity, blood volume pulse, gaze and head motion signals.

The second study (*N*=24), reported by Ji et al. [3], aimed to characterise user behaviours in information-seeking tasks (e.g., interacting with a search engine) with physiological signals, particularly in relation to cognitive load, affective arousal and valence. The study asked participants to perform search tasks and included both text and audio-only modalities.

MAIN FINDINGS

The results of the first study [4, 5] validated our proposed methodology, as we were able to control confounding variables in the complex experimental design. In particular, we observed consistent trends across participants, and there were 10 statistically significant features across the four IPAs [5]. Our results provide preliminary quantitative evidence of differences in physiological responses when users encounter IPAs, revealing the need to inspect signals separately by IPA.

The second study [3] allowed us to corroborate the finding that physiological signals captured with multiple sensors can characterise information-seeking processes in terms of cognitive and affective arousal. In particular, the results showed that participants experienced significantly higher cognitive load at the first stage (when familiarising themselves with the information need), with a subtle increase in alertness, while the formulation of the query required higher attention . Affective responses were more pronounced when judging the relevance of information, suggesting greater interest and engagement as knowledge gaps were resolved. We did not observe statistically significant differences in terms of affective valence.

To the best of our knowledge, this is the first study to explore user behaviours in a search process using a more nuanced quantitative analysis of physiological signals, offering valuable insights into cognitive load and emotional responses during information searches.

IMPLICATIONS

Our studies demonstrate that multiple sensors can be used to effectively characterise complex IPAs and tasks. By conducting user studies in controlled settings in the laboratory, we can collect **reliable physiological signals that characterise cognitive load and affective responses** occurring at different stages of interactive information processes in both screen-based and audio-only communication channels.

We believe our methodology (validated by effectively characterising informational search stages) can be used to **better understand how people interact with more novel (and less explored) interfaces**, such as voice-enabled intelligent assistants (e.g., Amazon Alexa), large language model (LLM)-based conversational search systems (e.g., Microsoft Copilot) and news consumption with richer interfaces (e.g., fact-checked content or trustworthy explanations).

CONCLUSION AND NEXT STEPS

The three phases in this project advanced knowledge on how to effectively quantify and measure bias and engagement when users interact with fact-checked content and IPAs. An interdisciplinary approach—including artificial intelligence (AI) and data science, human-computer interaction, media and communication, wearable computing, and cognitive science—allowed us to combine both qualitative and quantitative methods to address the problem using a mixed-methods research framework.

The broader availability of LLMs provides many opportunities for human–AI fact-checking [9]. For instance, retrieval-augmented generation can be explored as a means to combine multiple fact-checking reports and dynamically synthesise a digestible summary of the relevant fact-checking information a user seeks.

LLMs can also be powerful aids in designing and deploying conversational systems capable of delivering accurate fact-checking information. With our chatbot, called Walert, we undertook an initial attempt at using LLMs to build and evaluate customised fact-based conversational agents [7].

Our laboratory user studies showed how various physiological indicators of cognitive load could provide rich insights into IPAs. We plan to study more complex constructs such as cognitive biases (e.g., confirmation bias [2]) or trust [6]. We anticipate that the findings of this further research will help us create more effective, personalised and context-aware information access and news consumption systems.

While all the above novel technologies can be highly useful, LLMs, in particular, tend to generate inaccurate information, potentially harming the integrity of the fact-checking presentation, as well as the broader system. Therefore, it is crucial to evaluate the effectiveness and feasibility of novel LLM-based methods for fact-checking. The research methods we proposed in this project provide initial guidance for such evaluations and for creating LLM-supported human-in-the-loop applications for better accountability. We plan to develop comprehensive evaluation frameworks in the next steps.

REFERENCES

[1] Danula Hettiachchi, Kaixin Ji, Jenny Kennedy, Anthony McCosker, Flora D Salim, Mark Sanderson, Falk Scholer and Damiano Spina. 2023. Designing and evaluating presentation strategies for fact-checked content. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), 21–25 October 2023. Association for Computing Machinery, New York, NY, USA, 751–761. https://doi.org/10.1145/3583780.3614841

[2] Kaixin Ji. 2023. Quantifying and measuring confirmation bias in information retrieval using sensors. In Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), 8–12 October 2023. Association for Computing Machinery, New York, NY, USA, 236–240. https://doi.org/10.1145/3594739.3610765

[3] Kaixin Ji, Danula Hettiachchi, Flora D Salim, Falk Scholer and Damiano Spina. 2024. Characterizing information seeking processes with multiple physiological signals. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), 14–18 July 2024. Association for Computing Machinery, New York, NY, USA, 1006–1017. https://doi.org/10.1145/3626772.3657793

[4] Kaixin Ji, Damiano Spina, Danula Hettiachchi, Flora Dilys Salim and Falk Scholer. 2023. Examining the impact of uncontrolled variables on physiological signals in user studies for information processing activities. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23), 23–27 July 2023. Association for Computing Machinery, New York, NY, USA, 1971–1975. https://doi.org/10.1145/3539618.3591981

[5] Kaixin Ji, Damiano Spina, Danula Hettiachchi, Falk Scholer and Flora D Salim. 2023. Towards detecting tonic information processing activities with physiological data. In Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), 8–12 October 2023. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3594739.3610679

[6] Weronika Ławejska, Damiano Spina, Johanne R Trippas and Krisztian Balog. 2024. Explainability for transparent conversational information-seeking. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), 14–18 July 2024. Association for Computing Machinery, New York, NY, USA, 1040–1050. https://doi.org/10.1145/3626772.3657768

[7] Sachin Pathiyan Cherumanal, Kaixin Ji, Danula Hettiachchi, Falk Scholer, Futoon Abushaqra and Damiano Spina. 2024. Walert: Putting conversational search knowledge into action by building and evaluating a large language model-powered chatbot. In Proceedings of the 2024 Conference on Human Information Interaction and Retrieval (CHIIR '24), 10–14 March 2024. Association for Computing Machinery, New York, NY, USA, 401–405. https://doi.org/10.1145/3627508.3638309

[8] Lauren L. Saling, Devi Mallal, Falk Scholer, Russell Skelton, and Damiano Spina. 2021. No one is immune to misinformation: An investigation of misinformation sharing by subscribers to a fact-checking newsletter. PLOS ONE 16, 8 (August 2021), e0255702. https://doi.org/10.1371/journal.pone.0255702

[9] Damiano Spina, Mark Sanderson, Daniel Angus, Gianluca Demartini, Dana Mckay, Lauren L Saling and Ryen W White. 2023. Human-AI cooperation to tackle misinformation and polarization. Commun ACM 66, 7 (July 2023), 40–45. https://doi.org/10.1145/3588431

admscentre.org.au

Australian Government

Australian Research Council

This Centre is funded by the
Australian Government through
the Australian Research Council

ADM + S