

Towards Investigating Biases in Spoken Conversational Search

Sachin Pathiyan Cherumanal

RMIT University
Melbourne, Australia
s3874326@student.rmit.edu.au

Johanne R. Trippas

RMIT University
Melbourne, Australia
j.trippas@rmit.edu.au

Falk Scholer

RMIT University
Melbourne, Australia
falk.scholer@rmit.edu.au

Damiano Spina

RMIT University
Melbourne, Australia
damiano.spina@rmit.edu.au

Abstract

Voice-based systems like Amazon Alexa, Google Assistant, and Apple Siri, along with the growing popularity of OpenAI’s ChatGPT and Microsoft’s Copilot, serve diverse populations, including visually impaired and low-literacy communities. This reflects a shift in user expectations from traditional search to more interactive question-answering models. However, presenting information effectively in voice-only channels remains challenging due to their linear nature. This limitation can impact the presentation of complex queries involving controversial topics with multiple perspectives. Failing to present diverse viewpoints may perpetuate or introduce biases and affect user attitudes. Balancing information load and addressing biases is crucial in designing a fair and effective voice-based system. To address this, we (i) review how biases and user attitude changes have been studied in screen-based web search, (ii) address challenges in studying these changes in voice-based settings like SCS, (iii) outline research questions, and (iv) propose an experimental setup with variables, data, and instruments to explore biases in a voice-based setting like Spoken Conversational Search.

CCS Concepts

• **Information systems** → **Search interfaces; Presentation of retrieval results;** • **Human-centered computing** → **Auditory feedback.**

Keywords

audio output; bias; information retrieval; conversational search

ACM Reference Format:

Sachin Pathiyan Cherumanal, Falk Scholer, Johanne R. Trippas, and Damiano Spina. 2024. Towards Investigating Biases in Spoken Conversational Search. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI Companion '24)*, November 04–08, 2024, San Jose, Costa Rica. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3686215.3690156>



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

ICMI Companion '24, November 04–08, 2024, San Jose, Costa Rica
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0463-5/24/11
<https://doi.org/10.1145/3686215.3690156>

1 Introduction

The adoption of generative AI in search engines such as OpenAI’s ChatGPT and Microsoft Copilot signifies a shift in traditional information retrieval. This transition moves away from the conventional “ten blue links” approach to a more conversational paradigm. While current generative AI technologies are mostly text-based, the increasing adoption of intelligent assistants like Google Assistant, Amazon Alexa, and Apple Siri indicates a move towards fully voice-based information interaction. Such Spoken Conversational Search (SCS) – a type of voice-based information access system, can meet diverse information needs for many, including visually impaired, low-literacy, and sighted users in situations where reading is impractical, like driving, cooking, or exercising.

Information access systems, like search engines, have traditionally been users’ main source of information for decades. However, voice-based systems like SCS, and multimodal systems that integrate voice, face challenges due to the transient and linear nature of the voice channel, which can limit the amount of information that can be presented. Unlike screen-based web search that comes with the luxury of visually inspecting the “ten-blue links” anytime during the search session. For instance, in an SCS query about controversial topics like “should zoos exist?”, the system should provide both sides (i.e., PRO and CON) to avoid biasing the user with one-sided perspectives. On a larger scale, this can adversely affect society by creating or reinforcing biases affecting users’ attitudes towards certain topics. Such biased exposure of perspectives and its impact have been extensively studied in screen-based web search, but less in voice-based settings like SCS. In this work, we review attitude changes in screen-based web search (Section 2.1), address challenges in studying these changes in voice-based settings like SCS (Section 2.2), outline research questions (Section 3), and propose an experimental setup (Section 4) with variables, data, and instruments to explore biases in SCS.

2 Background

2.1 Biases in Information Access

Search systems have become a primary source of information, but the resulting information overload [36] due to the vast amount of information available online, can cause stress, confusion, and reduce productivity [13]. Effectively navigating information requires filtering, prioritizing, and organizing of the information. Due to limited cognitive capacity, users often rely on mental shortcuts, known as cognitive biases, which is a “pattern of deviation” from

norm or rationality in thinking and reasoning, that influences the users' decision making process. For instance, during a search session, a user's attitude or decision can be heavily influenced by the order of the documents in the Search Engine Results Page (SERP) i.e., *order effect*, primarily due to position bias [22] – a user's tendency to interact more with items at the top of a SERP. A user's decision making may also be influenced by the balanced or imbalanced exposure of perspectives i.e., *exposure effect*. For instance, Gao and Shah [8] demonstrates a Google search query “coffee health” where the SERP fails to provide a balanced perspective on the benefits and the negative impacts of coffee on health. The two effects (i.e., order and exposure) are two of many phenomenon that leads to change in user attitude is called Search Engine Manipulation Effect (SEME). Biases have also shown to be easily “weaponized” [1] to manipulate public opinion. For instance, Epstein and Robertson [6] shows that biased search rankings can shift the voting preferences of undecided voters by 20% consequently impacting election outcomes. Over the past years, SEME and biases in rankings have been extensively studied in the context of traditional screen-based web search i.e., “the ten blue links”. A recent crowdsourced study investigated factors affecting user attitude change while interacting with the entire SERP, including exposure and order effects, perceived diversity, pre-existing beliefs, and willingness to process counter-attitudinal information [4]. Bink et al. [2] examined similar factors using SERP “snippets”. However, investigating these factors in a voice-based setting like SCS remains an open challenge, as discussed in Section 2.2. To the best of our knowledge, the only biases studied in voice-setting are related to the recognition quality of phrases spoken by different groups [18].

2.2 Challenges in Spoken Conversational Search

Spoken Conversational Search (SCS) is an information access system where communication is entirely verbal [30]. The voice channel poses cognitive challenges, leading to different user behavior compared to a screen-based search processes. For instance, in traditional web search, users have the flexibility of reviewing multiple documents in the SERP with ten-blue links [34, 37]. In contrast, in voice-only settings, the linear and transient nature of the channel may challenge users to keep up with information due to limited cognitive capacity. Another challenge is that users can easily review and refer back to their query with screens, which is inherently more difficult with voice queries. Moreover, previous work shows that users pay more attention to the first and last responses of a voice-based system [31] for faceted queries (e.g., broad intent with subtopics, like controversial queries). We are still unsure if the order of response influence user attitudes in a voice-based setting. Furthermore, when multiple perspectives need to be presented in an unbiased, it is important to understand if response order (i.e., order effects) our source influences user attitude changes [29]. One of the solutions to mitigate biases in screen-based web search has been to expose users to diverse perspectives. However, this becomes challenging in a voice-based setting due to user's limited cognitive capacity and the need to keep responses short to avoid additional cognitive load. Additionally, previous research indicates that user attitudes can change based on the proportion of perspectives consumed (i.e., exposure effects, as noted by Draws et al. [4]). Following

the same, we also explore how users' perceived diversity of voice responses and their open-mindedness impact their attitudes. With this knowledge, we define our research questions in Section 3.

3 Research Questions

The following research questions shall guide the experimental setup and protocol that will be used for our user study.

RQ1: Does the order of passages (corresponding to different stances) affect user attitude in a voice based search system like SCS?

RQ2: Does the variation in exposure of passages (corresponding to different stances) affect user attitude in a voice based search system like SCS?

RQ3: Do users actually perceive diversity in a voice-only search system? If so, is the ordering or the exposure of the passages perceived as diverse by the users of a voice based search system like SCS?

RQ4: Do factors like actively open-minded thinking (AOT) and perceived diversity help predict user attitude change in a voice-based search setting like SCS?

4 Experimental Setup

In this section, we explore the potential experimental setup needed to address our research questions through a between-subjects 4x5 user study.¹

4.1 Data

We use the annotated data released by Draws et al. [4], which includes topics (e.g., Is obesity a disease?), viewpoints (i.e., PRO or CON), and binary relevance (i.e., relevant or not relevant). To control for unknown effects from non-relevant documents, we use only relevant ones. This dataset features five topics with neutral or mild attitudes: (i) *Are social networking sites good for our society?*, (ii) *Should zoos exist?*, (iii) *Is cell phone radiation safe?*, (iv) *Should bottled water be banned?*, (v) *Is obesity a disease?*. While this data contains viewpoint labels in the range [-3,3], (-3 represents *strongly opposing* and +3 represents *strongly supporting*), for our study we pick arguments that have only the following labels -2, -1, +1 and +2 (due to lack of documents for the other viewpoints). For simplicity and to reduce the dimensions in our study, we group all the supporting passages and opposing passages to two different categories. The aforementioned documents are webpages, and not feasible in a voice-only setting due to audio limitations on search result presentation. Furthermore, previous work has shown that the length of voice responses can affect voice-interaction experiences [31, 32]. Section 4.1.1 details the steps we follow to control for the length of the passage.

4.1.1 Passages. We use ChatGPT for aspect-based abstractive summarization [19] of each document. Following Wu et al. [35], we apply a zero-shot prompt chaining approach with the structure as follows. PROMPT 1: “From the text below, provide the justification, for [VIEWPOINT] the topic [TOPIC]. Justification: Text: [DOCUMENT CONTENT]”, where [VIEWPOINT] can be either “attacking” (i.e., CON) or “supporting” (i.e., PRO) and the [TOPIC] can be one of five discussed earlier (e.g., *Should zoos exist*). The output of PROMPT

¹More details about the pre-registration are made available at <https://osf.io/35ewj/>

1 is then chained as an input to PROMPT 2: “Now summarize the below text in less than 30 words. Text: [PROMPT 1 RESPONSE]”.

4.1.2 Text-to-Speech. After generating the passages, we use Amazon Polly² Text-To-Speech for synthesizing the voice responses, maintaining the same gender voice across all passages and avoiding audio manipulations (like pitch and pauses) to prevent impacting user information consumption [3]. By using the same voice and tone features of the speech across all the passages, we ensure that there are no external factors (e.g., persona, persuasiveness of the voice, gender) that impact the user [5].

4.1.3 Voice-Only Search Results. We now use synthesized audio passages to create a ranked list containing top-4 results for each topic (as we focus on voice-only setting) in contrast to Draws et al. [4] where top-10 text-based results were used to simulate a web search experience. In Figure 1a, we show the 16 possible permutations in which the two stances (PRO and CON) can be arranged. We also ensure that the synthetic scenario is proportion-agnostic, i.e., all permutations have a 1/1 ratio of PRO/CON arguments judged as relevant in the ground truth [24]. We then categorise the 16 permutations into two (i) *Exposure varying (E)*, where the number of PRO and CON passages are not equal (ii) *Exposure balanced (O)*, where the number of PRO and CON passages in the top-4 position are balanced. Given the 16 ranked lists, we compute the fairness score using a combination of rND@4, rKL@4, rRD@4 adapted by Pathiyar Chermanal et al. [24] specifically for conversational search on controversial topics (see Figure 1b).

We pick one unfair and a fair system each of the two categories described above. In summary our study uses four systems (or conditions), (i) Exposure varying and Unfair (E_{unfair}) – Permutation 4 (ii) Exposure varying and fair (E_{fair}) – Permutation 2 (iii) Exposure balanced and unfair (O_{unfair}) – Permutation 11 (iv) Exposure balanced and fair (O_{fair}) – Permutation 7 (see Figure 2). In the case of O_{fair} and O_{unfair} the exposure is maintained in the top-4 positions. Whereas E_{fair} and E_{unfair} vary in terms of exposure.

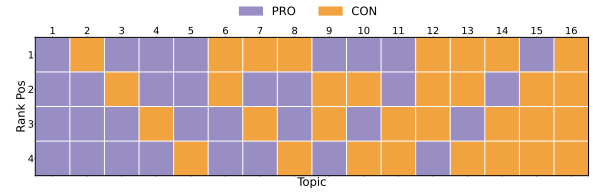
4.2 Participants

We propose to conduct a crowdsourced user study and identified the sample size required for a *Between-Subjects Analysis of Covariance ANCOVA analysis* using the G^* Power software. Consequently, we propose to recruit 400 participants from three English-speaking countries for our study (i.e., USA, UK, and Australia).

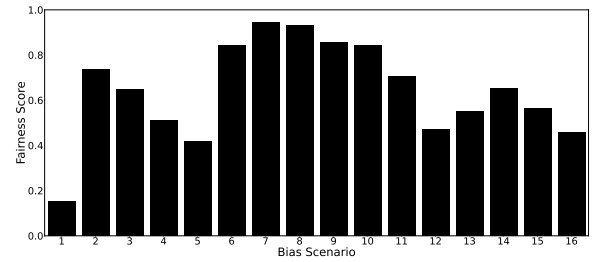
4.3 Experiment Protocol

The recruited participants will be given backstories to simulate a real-life information need. The backstory will provide context for the voice-only search task to promote a natural search behaviour [9]. The participants will then be provided with a sample task to familiarise with the experiment setup. The topic used for this sample task shall be different from the five topics that will be used in this study. To ensure reproducibility of the web search context, we use the same backstory used by Draws et al. [4]. The steps involved in the task are mentioned below:

²<https://aws.amazon.com/polly/>



(a) Permutations of argument stances across the top-4 positions using simulated scenarios.



(b) Fairness score (higher means more fair) computed using three fairness metrics [24].

Figure 1: From 1a and 1b, we identify the 4 topics/scenarios based on fairness score and narrow down our study to four scenarios (see Figure 2).

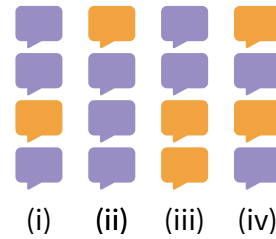


Figure 2: The four conditions used in the study, (i) Exposure varying and Unfair (E_{unfair}) (ii) Exposure varying and fair (E_{fair}) (iii) Exposure balanced and unfair (O_{unfair}) (iv) Exposure balanced and fair (O_{fair}). The four conditions represent top-4 positions in a ranking where each audio response has an associated stance/perspective i.e., PRO (supporting) or CON (opposing).

Step 1. Participants shall be given a brief overview of the task and asked to state their gender, age, familiarity with voice-only systems [14] and attitude towards one of the five debated topics on a 7-point Likert scale. The participant will be assigned the topic towards which they have a neutral attitude. For participants who does not have a neutral attitude, one of the 5 topics would be randomly assigned. Participants shall be assigned a topic toward which they have a mild pre-existing attitude (i.e., responding with “somewhat supporting”, “neutral”, or “somewhat opposing”).

Step 2. The participants shall then have the freedom to raise a spoken natural language query as this would help make the task

Table 1: Experiment variables considered in the study.

Variable Type	Variables	Description
Independent	Topic (x5)	Five topics the user will be exposed to in the experiment.
	Bias Scenario (x4)	One of the four search results conditions discussed earlier in Section 4.1.3 and shown in Figure 2.
Dependent	User Attitude	Measured twice—before the experiment and after exposure to bias— on a 7-point Likert Scale. The change will be calculated as in prior research in the web search context [4].
Covariates	Actively Open-minded Thinking	Measures the degree to which a person is willing to consider opposing perspectives— on a 7-point Likert Scale [11].
	Perceived Diversity	Measures the degree of diversity in perspectives perceived by the user, in the voice-only search results— on a 7-point Likert Scale [15].
Exploratory	Gender and Age	Martínez-Costa et al. [20]
	Education Level	Leroy and Kauchak [17] found that education level impacts comprehension of information presented via audio channels.
	Topic Familiarity	3-point Likert Scale.
	Declared Impairments	Taking into account individual variations for those with brain injuries or neurological conditions (e.g., neurodiverse population) [26].
	Language Proficiency	Muda et al. [21]

more realistic. However, the query would not affect the results provided for the assigned topic.

Step 3. Participants will listen to four voice responses corresponding to the top four results from the ranked list for the given topic. However, participants shall be randomly assigned one of the bias scenarios shown in Figure 2.

Step 4. Participants would then state their (updated) attitude and interest concerning their assigned topic.

Step 5. The participants would be asked to fill a post-questionnaire that consisted of the AOT scale [11] and the perceived diversity scale [15].

Step 6. Participants will receive a written transcription of the voice responses they heard and will then be asked to re-order the passages according to the original sequence.

5 Implications and Ethical Considerations

Understanding how the order (i.e., order effect) and proportion (i.e., exposure effect) of perspectives in a voice-only setting influence user attitudes can inform the design of fairness-aware presentation strategies, an ongoing challenge [27, 28]. Additionally, offline evaluation metrics for SCS and multimodal information access systems with voice interaction are still in their infancy, and often relies on web-search based user models (e.g., logarithmic decay) [7] that do not account for limitations in users’ working memory and recall in voice channels. Our setup and recall tests may reveal patterns in user attention, provoking the community to explore SCS specific user models. These patterns could also inform and guide practitioners in designing techniques to alert users to potential biases, similar to the auditory warnings by Pathiyen Cherumanal et al. [23] and the “nudges” by Gohsen et al. [10]. A limitation of our study is scalability due to the controlled knowledge base. However, advances

in Large Language Models (LLMs), like Retrieval-Augmented Generation (RAG) or in general, Generative Information Access [33], could enable future experiments to use a broader knowledge base, as shown by Pathiyen Cherumanal et al. [25].

Ethical Considerations. It may be worth keeping in mind that individual differences may exist between users, such as neurological impairments affecting comprehension. This must be considered [26] in experiments that involves voice-based systems like SCS. A key challenge in systems like SCS is explainability [16] and the lack of explanations from SCS in such studies may impact users’ trust and perceived credibility. Furthermore, in voice-based settings exploring biases, participants are exposed to biased information and researchers may have to provide a textual summary afterward to prevent retention of biased content.

6 Conclusion

Given the impact of biases in traditional search and users’ growing preference for conversational search experience powered by Large Language Models (LLMs), such as Microsoft Copilot and OpenAI’s GPT-4o³ which prioritize natural language and voice, there is a critical need for research on biases in these settings due to a lack of exploration concerning biases in voice-based systems as discussed by Ji et al. [12, 13]. This study focuses specifically on voice responses, aiming to investigate biases within this modality.

Acknowledgments

This research has been carried out in the unceded lands of the Woi Wurrung and Boon Wurrung peoples of the eastern Kulin Nation. We pay our respects to their Ancestors and Elders, past, present, and emerging. This research is partially supported by the Australian Research Council (ARC, project nr. DE200100064 and CE200100005).

³<https://openai.com/index/hello-gpt-4o/>

References

- [1] Marwah Alaofi, Luke Gallagher, Dana McKay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. 2022. Where Do Queries Come From?. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) (SIGIR '22). Association for Computing Machinery, New York, NY, USA, 2850–2862. <https://doi.org/10.1145/3477495.3531711>
- [2] Markus Bink, Sebastian Schwarz, Tim Draws, and David Elswailer. 2023. Investigating the Influence of Featured Snippets on User Attitudes. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) (CHIIR '23). Association for Computing Machinery, New York, NY, USA, 211–220. <https://doi.org/10.1145/3576840.3578323>
- [3] Aleksandr Chuklin, Aliaksei Severyn, Johanne R. Trippas, Enrique Alfonseca, Hanna Silen, and Damiano Spina. 2019. Using Audio Transformations to Improve Comprehension in Voice Question Answering. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, September 9–12, 2019, Proceedings* (Lugano, Switzerland). Springer-Verlag, Berlin, Heidelberg, 164–170. https://doi.org/10.1007/978-3-030-28577-7_12
- [4] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. 2021. This Is Not What We Ordered: Exploring Why Biased Search Result Rankings Affect User Attitudes on Debated Topics. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) (SIGIR '21). Association for Computing Machinery, New York, NY, USA, 295–305. <https://doi.org/10.1145/3404835.3462851>
- [5] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive Synthetic Speech: Voice Perception and User Behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) (CUI '20). Association for Computing Machinery, New York, NY, USA, Article 6, 9 pages. <https://doi.org/10.1145/3405755.3406120>
- [6] Robert Epstein and Ronald E. Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences* 112, 33 (2015), E4512–E4521. <https://doi.org/10.1073/pnas.1419828112>
- [7] Alexander Frummet, Andrea Papenmeier, Maik Fröbe, Johannes Kiesel, Vaibhav Adlakha, Norbert Braunschweiler, Mateusz Dubiel, Satanu Ghosh, Marcel Gohsen, Christin Kreuz, Milad Momeni, Markus Nilles, Sachin Pathiyam Cherumanal, Abbas Pirmoradi, Paul Thomas, Johanne R. Trippas, Ines Zelch, and Oleg Zendel. 2024. Report on the 8th Workshop on Search-Oriented Conversational Artificial Intelligence (SCAI 2024) at CHIIR 2024. *SIGIR Forum* 58, 1 (aug 2024), 1–12. <https://doi.org/10.1145/3687273.3687282>
- [8] Ruoyuan Gao and Chirag Shah. 2020. Toward Creating a Fairer Ranking in Search Engine Results. *Information Processing & Management* 57, 1 (2020), 102138. <https://doi.org/10.1016/j.ipm.2019.102138>
- [9] Souvick Ghosh. 2019. Exploring Result Presentation in Conversational IR Using a Wizard-of-Oz Study. In *Advances in Information Retrieval: 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14–18, 2019, Proceedings, Part II* (Cologne, Germany). Springer-Verlag, Berlin, Heidelberg, 327–331. https://doi.org/10.1007/978-3-030-15719-7_43
- [10] Marcel Gohsen, Johannes Kiesel, Mariam Korashi, Jan Ehlers, and Benno Stein. 2023. Guiding Oral Conversations: How to Nudge Users Towards Asking Questions?. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) (CHIIR '23). Association for Computing Machinery, New York, NY, USA, 34–42. <https://doi.org/10.1145/3576840.3578291>
- [11] Uriel Haran, Ilana Ritov, and Barbara A. Mellers. 2013. The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making* 8, 3 (2013), 188–201. <https://doi.org/10.1017/S1930297500005921>
- [12] 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* (Washington DC, USA) (SIGIR '24). Association for Computing Machinery, New York, NY, USA, 1006–1017. <https://doi.org/10.1145/3626772.3657793>
- [13] Kaixin Ji, Sachin Pathiyam Cherumanal, Johanne R. Trippas, Danula Hettiachchi, Flora D. Salim, Falk Scholer, and Damiano Spina. 2024. Towards Detecting and Mitigating Cognitive Bias in Spoken Conversational Search. In *Adjunct Publication of the 26th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '24)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3640471.3680245>
- [14] Johannes Kiesel, Kevin Lang, Henning Wachsmuth, Eva Hornecker, and Benno Stein. 2020. Investigating Expectations for Voice-Based and Conversational Argument Search on the Web. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (Vancouver BC, Canada) (CHIIR '20). Association for Computing Machinery, New York, NY, USA, 53–62. <https://doi.org/10.1145/3343413.3377978>
- [15] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22 (2012), 441–504. <https://doi.org/10.1007/s1257-011-9118-4>
- [16] Weronika Lajewska, Damiano Spina, Johanne Trippas, and Krzysztof Balog. 2024. Explainability for Transparent Conversational Information-Seeking. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Washington DC, USA) (SIGIR '24). Association for Computing Machinery, New York, NY, USA, 1040–1050. <https://doi.org/10.1145/3626772.3657768>
- [17] GONDY Leroy and David Kauchak. 2019. A comparison of text versus audio for information comprehension with future uses for smart speakers. *JAMIA Open* 2, 2 (05 2019), 254–260. <https://doi.org/10.1093/jamiaopen/ooz011>
- [18] Lanna Lima, Vasco Furtado, Elizabeth Furtado, and Virgilio Almeida. 2019. Empirical Analysis of Bias in Voice-based Personal Assistants. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) (WWW '19). Association for Computing Machinery, New York, NY, USA, 533–538. <https://doi.org/10.1145/3308560.3317597>
- [19] Hui Lin and Vincent Ng. 2019. Abstractive Summarization: A Survey of the State of the Art. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (Jul. 2019), 9815–9822. <https://doi.org/10.1609/aaai.v33i01.33019815>
- [20] María-Pilar Martínez-Costa, Fernando López-Pan, Nataly Buslón, and Ramón Salaverria. 2023. Nobody-fools-me perception: Influence of Age and Education on Overconfidence About Spotting Disinformation. *Journalism Practice* 17, 10 (2023), 2084–2102. <https://doi.org/10.1080/17512786.2022.2135128>
- [21] Rafal Muda, Gordon Pennycook, Damian Hamerski, and Michał Bialek. 2023. People are worse at detecting fake news in their foreign language. *Journal of Experimental Psychology: Applied* (2023). <https://doi.org/10.1037/xap0000475>
- [22] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Gery Gay, and Laura Granka. 2007. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication* 12, 3 (04 2007), 801–823. <https://doi.org/10.1111/j.1083-6101.2007.00351.x>
- [23] Sachin Pathiyam Cherumanal, Ujwal Gadiraju, and Damiano Spina. 2024. Everything We Hear: Towards Tackling Misinformation in Podcasts. In *Proceedings of the 25th International Conference on Multimodal Interaction (ICMI '24)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3678957.3678959>
- [24] Sachin Pathiyam Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft. 2021. Evaluating Fairness in Argument Retrieval. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (Virtual Event, Queensland, Australia) (CIKM '21). Association for Computing Machinery, New York, NY, USA, 3363–3367. <https://doi.org/10.1145/3459637.3482099>
- [25] Sachin Pathiyam Cherumanal, Lin Tian, Futoon M. Abushaqra, Angel Felipe Magnossão de Paula, Kaixin Ji, Halil Ali, Danula Hettiachchi, Johanne R. Trippas, Falk Scholer, and Damiano Spina. 2024. Walert: Putting Conversational Information Seeking 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* (Sheffield, United Kingdom) (CHIIR '24). Association for Computing Machinery, New York, NY, USA, 401–405. <https://doi.org/10.1145/3627508.3638309>
- [26] Laurianne Sitbon, Gerd Berger, and Margot Brereton. 2023. Perspectives of Neurodiverse Participants in Interactive Information Retrieval. *Foundations and Trends® in Information Retrieval* 17, 2 (2023), 124–243. <https://doi.org/10.1561/15000000086>
- [27] Damiano Spina, Danula Hettiachchi, and Anthony McCosker. 2024. *Quantifying and Measuring Bias and Engagement in Automated Decision-Making*. Technical Report. ARC Centre of Excellence for Automated Decision-Making and Society, RMIT University, Melbourne, Australia. <https://doi.org/10.60836/k8xh-en92>
- [28] Damiano Spina, Johanne R. Trippas, Paul Thomas, Hideo Joho, Katrina Byström, Leigh Clark, Nick Craswell, Mary Czerwinski, David Elswailer, Alexander Frummet, Souvick Ghosh, Johannes Kiesel, Irene Lopatovska, Daniel McDuff, Selina Meyer, Ahmed Mourad, Paul Owoicho, Sachin Pathiyam Cherumanal, Daniel Russell, and Laurianne Sitbon. 2021. Report on the future conversations workshop at CHIIR 2021. *SIGIR Forum* 55, 1, Article 6 (jul 2021), 22 pages. <https://doi.org/10.1145/3476415.3476421>
- [29] Johanne R Trippas, Luke Gallagher, and Joel Mackenzie. 2024. Re-evaluating the Command-and-Control Paradigm in Conversational Search Interactions. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management* (Boise, ID, USA) (CIKM '24). ACM, New York, NY, USA. <https://doi.org/doi.org/10.1145/3627673.3679588>
- [30] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, Hideo Joho, and Mark Sanderson. 2018. Informing the Design of Spoken Conversational Search: Perspective Paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) (CHIIR '18). Association for Computing Machinery, New York, NY, USA, 32–41. <https://doi.org/10.1145/3176349.3176387>
- [31] Johanne R. Trippas, Damiano Spina, Mark Sanderson, and Lawrence Cavedon. 2015. Results Presentation Methods for a Spoken Conversational Search System. In *Proceedings of the First International Workshop on Novel Web Search Interfaces and Systems* (Melbourne, Australia) (NWSearch '15). Association for Computing

- Machinery, New York, NY, USA, 13–15. <https://doi.org/10.1145/2810355.2810356>
- [32] Johanne R. Trippas, Damiano Spina, Mark Sanderson, and Lawrence Cavedon. 2015. Towards Understanding the Impact of Length in Web Search Result Summaries over a Speech-only Communication Channel. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (Santiago, Chile) (SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 991–994. <https://doi.org/10.1145/2766462.2767826>
- [33] Johanne R. Trippas, Damiano Spina, and Falk Scholer. 2024. Adapting Generative Information Retrieval Systems to Users, Tasks, and Scenarios. In *Information Access in the Era of Generative AI*, Ryan W. White and Chirag Shah (Eds.). Springer Nature Switzerland AG, Cham, Switzerland.
- [34] Alexandra Vtyurina, Charles LA Clarke, Edith Law, Johanne R Trippas, and Horatiu Bota. 2020. A mixed-method analysis of text and audio search interfaces with varying task complexity. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 61–68.
- [35] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. 2022. PromptChainer: Chaining Large Language Model Prompts through Visual Programming. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 359, 10 pages. <https://doi.org/10.1145/3491101.3519729>
- [36] Tyson Yunkaporta. 2023. *Right story, wrong story: Adventures in Indigenous thinking*. Text Publishing. <https://hdl.handle.net/10779/DRO/DU:24502765.v1>
- [37] Hamed Zamani, Johanne R Trippas, Jeff Dalton, Filip Radlinski, et al. 2023. Conversational information seeking. *Foundations and Trends® in Information Retrieval* 17, 3-4 (2023), 244–456.