Towards Detecting and Mitigating Cognitive Bias in Spoken Conversational Search

Kaixin Ji*
RMIT University
Melbourne, Australia
kaixin.ji@student.rmit.edu.au

Sachin Pathiyan Cherumanal*
RMIT University
Melbourne, Australia
s3874326@student.rmit.edu.au

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

Danula Hettiachchi
RMIT University
Melbourne, Australia
danula.hettiachchi@rmit.edu.au

Flora D. Salim
The University of New South Wales
Sydney, Australia
flora.salim@unsw.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

Spoken Conversational Search (SCS) poses unique challenges in understanding user-system interactions due to the absence of visual cues, and the complexity of less structured dialogue. Tackling the impacts of cognitive bias in today's information-rich online environment, especially when SCS becomes more prevalent, this paper integrates insights from information science, psychology, cognitive science, and wearable sensor technology to explore potential opportunities and challenges in studying cognitive biases in SCS. It then outlines a framework for experimental designs with various experiment setups to multimodal instruments. It also analyzes data from an existing dataset as a preliminary example to demonstrate the potential of this framework and discuss its implications for future research. In the end, it discusses the challenges and ethical considerations associated with implementing this approach. This work aims to provoke new directions and discussion in the community and enhance understanding of cognitive biases in Spoken Conversational Search.

CCS CONCEPTS

• Human-centered computing → Empirical studies in ubiquitous and mobile computing; **• Information systems → Users and interactive retrieval**.

KEYWORDS

Cognitive Bias, Spoken Conversational Search, Information Seeking, Physiological Signals, Wearable Sensors, Experimental Design

ACM Reference Format:
Kaixin Ji, Sachin Pathiyan 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 *26th International Conference on Mobile Human-Computer Interaction (MOBILEHCI Adjunct '24), September 30-October 3, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3640471.3680245

*Equal contributions.

1 INTRODUCTION

The rapid advancement of generative AI has been swiftly integrated into our everyday systems and acted as our personal assistants. For example, Bing Chat on search engines. This advancement marks a transition from traditional query-list-examine to conversational question-answering in information searches. Although such interaction is primarily text-based with limited access, the trend is evolving towards multimodal capabilities in personal devices, exemplified by the partnership between GPT-4o and Apple.[1] This offers broader accessibility through voice-based interaction, paving the way for Spoken Conversational Search (SCS). While this advancement can benefit various groups with limited access (e.g., visually impaired) [33] and those in situations where reading isn't feasible (e.g., driving or exercising) [80], delivering user-friendly yet relevant responses remains a challenge, especially due to limitations in cognition (in processing, analyzing, and interpreting information) and that of the voice channel itself [77, 115, 122]. Search engines act as intermediaries of knowledge making it crucial for such systems to curate relevant yet diverse content to foster balanced viewpoints (avoid "echo chambers" [8]) and overcome cognitive limitations and biases.[2]

However, screen-based web search benefits from well-defined tools and standard protocols to visualize and study bias behaviors, such as eye-tracking [17, 22, 42, 129] and click-through

[1]OpenAI and Apple announce partnership to integrate ChatGPT into Apple experiences. Retrieved 10 June, 2024 from https://openai.com/index/openai-and-apple-announce-partnership/

[2]For instance, researchers have already raised concerns about biases in personalized informatics [124], or Conversational Search [100].

logs [26, 59, 106]. Such methodologies are not established for SCS, which calls for instruments, methodology, and protocols that go beyond the visual paradigm [35]. Regarding this, our contributions in this position paper are three-fold: (i) discuss the applicability of behavior analysis tools used for web search to SCS and identify research opportunities for exploring cognitive biases in SCS, (ii) propose approaches to design experiments, from setup formats to measurements, with preliminary results demonstrating the potential of using multimodal physiological signals as a voice channel equivalent to eye-tracking in web search, (iii) outline challenges with suggestions and ethical considerations for adopting our approach to achieve accurate and representative results from multimodal signals.

2 BACKGROUND

2.1 Spoken Conversational Search

Conversational information seeking (CIS), the process of obtaining information through conversations (text, audio/voice, or multi-modal), is a fast-developing research area [91, 128]. CIS supports users to search for information through natural language. It enables users to ask questions, refine their questions, ask follow-up questions, or provide relevant feedback in a natural manner. The interaction of such systems could either be single-turn or multi-turn. In contrast to a single-turn, a multi-turn setting typically maintains the conversational context (e.g., co-reference resolution)[3] in a back-and-forth information exchange with the user [128]. Some advantages of multi-turn CIS include alleviating the cognitive burden on the user by breaking down the information, assisting with information need formulation, or providing highly personalized information for a given context [113]. While context management may be relatively trivial for a CIS system, users also have to perform context management subconsciously. This would require significant cognitive effort from the user, particularly when the conversation gets longer, and the task gets more complex. This paper focuses on SCS, a type of CIS, where communication between the user and system is entirely mediated verbally through audio [109]. Visual CIS interfaces often use screen-based cues like boldfacing important sections of text [20], or attributing sources within the responses of text-based CIS [61], large language model (LLM) based conversational agents [11, 62, 99]. These cues aid users in effortlessly finding information. However, in linear channels like SCS, users may struggle to keep up with presented information, due to limited cognitive capacity and audio features (e.g., prosody) that can affect their understanding [20].

2.2 Cognitive Biases in Information Seeking

Cognitive biases "are systematic errors in judgment and naturally occurring tendencies that skew information processes, due to limitations in cognitive, motivational, or environmental factors, which lead to sub-optimal or fundamentally wrong outcomes" [121]. It is based on the cognitive load theory [107] that humans have limited cognitive capacity, so they tend to favor mental shortcuts of other judgments (e.g., system ranking, or crowd opinions) [8, 104]. Information seekers often rely on perceived trustworthiness when accessing information, constructing mental models to link various pieces of information [59]. This process may influenced by cognitive biases [8]. In particular, the information cherry-picking will likely be affected by the order (rank) (*Order Effect*), imbalanced viewpoints (*Exposure Effect*), a prior judgment (*Confirmation Bias*), the first piece of information (*Anchoring Bias*) or *Misinformation* [18, 59, 82, 104]. This can lead to uncritical support for partisans, reinforce stereotypes, and spread misinformation [8, 21, 59]. Conversely, it can also help users effectively navigate overwhelming information. Therefore, the impacts of these biases must be studied to provide accurate in situ information in SCS. Common methods for measuring cognitive bias in web search include web-logging metrics like sentiment analysis, dwell time, clicks [26, 30, 59, 106], and eye-tracking [17, 22, 42, 129].

However, there is a lack of research on biases in voice-based systems. Eye-tracking is unavailable on these systems, and web logging has limitations in providing granular data [21, 106]. Additionally, recent work found inconsistent results using NASA-TLX for mental load [37], suggesting traditional self-reports may be unreliable. These emphasize the need for fine-grained data, such as physiological data from wearable technology.

2.3 Neural Activities for Cognitive Bias

The human brain is divided into several regions in charge of different functionalities. For example, the frontal lobe handles decision-making, motivation, and focus, while the temporal lobe is responsible for auditory and language processing [66]. Investigating how neural activities traveled across regions provides a window look into the flow and processing of information within the brain [64, 71]. For example, researchers have measured the workload change in web browsing [47] and understood search intentions [75, 76] or keyword relevance [123]. Listening effort refers to the cognitive resources people spend on listening [34, 89]. The audio information is first stored as a "buffer" in working memory, then processed for comprehension, and then potentially stored in long-term memory [89, 93]. During this process, information that is discrepant with the current mental model or perceived as irrelevant will not be [93], or requires more effort [94] to interpret further. Compared to visually impaired individuals, sighted users generally have a diminished ability to understand and interpret audio information [16] as evidenced by increased cognitive effort in audio-only scenarios [97]. This heightened effort may hinder their capacity for reasoning and critical thinking, that is essential for mitigating cognitive bias [8]. By understanding such cognitive activities involved, we can understand if users encounter bias - e.g., if information only reaches language regions or proceeds to memory retrieval.For instance, users expend more cognitive effort and attention when assessing information aligned with their beliefs [74]. Additionally, initial judgments during utterances can shape final decisions on

[3]By "context", we mean the information exchanged during the conversation necessary to interpret the users' response, e.g., the history, preferences, and so on.

a voice's believability [46].These results suggest a hypothesized process where language regions activate first, followed by comprehension and working memory assessment. If the information is deemed irrelevant or dissident, it will be discarded without further processing across brain regions, leading to biased decisions [71]. Advances in wearable devices have enabled physiological sensing to detect cognitive bias in web searches, relying on grounded theories (e.g.,cognitive load theory [107], orienting responses [103], cognitive dissonance [28, 90], and dual-thinking system theory [24]). Multi-modal data are discussed later in Section 5.1.

3 RESEARCH OPPORTUNITIES

Exploring cognitive biases in SCS offers research avenues, such as characterizing search stages, understanding user behavior, and developing bias detection or mitigation approaches.

How to Characterize Cognitive Bias at the Different Stages of the SCS Process? Cognitive bias may occur at each stage of a visual-based search process [8], i.e., querying, consuming the search results, and judging relevance and satisfaction. Previous work suggests variations in search stages or actions [67] (e.g., query formulation/reformulation, results scanning, selection, and assessment) and user behaviors between screen and audio-only channels [113]. Similarly, cognitive biases manifest differently in these search stages for screen and audio-only channels [50]. For instance, users can review and refer back to their query more easily on screens than with voice queries [96]. In SCS, queries are often in natural language [23, 40], and the arrangement of words may reveal user intent [102] and perhaps even reveal any underlying biases. For instance, a user's choice of query formulation between, "Why is renewable energy inefficient?" and "What are the efficiencies of renewable energy?" may indicate preconceived beliefs, potentially leading to biased search results. Furthermore, detecting cognitive biases in the query stage can be complicated by users' false memories(misremembered attributes of searched items), as they may not easily accept misremembering [52]. To this end, we highlight the significance of investigating cognitive processes at various stages of SCS interaction (e.g., detecting false memories at the query stage).

What Is the Role of Clarifying Questions in SCS? How Is It Related to Cognitive Bias? In CIS, the dialogic nature makes query reformulation and clarifying questions more critical and frequent, supporting conversational actions [3, 112, 128]. Users often iteratively refine queries by referring to previous responses to narrow down or expand their initial query [128]. Cognitive biases may influence this iterative process. For example, if the information aligns with users' beliefs they may accept it without further questioning. Conversely, if it opposes their beliefs, they may reformulate the query to find results that align with their expectations. This means that considering a user's reformulation/clarifying questions can help to detect potential bias. Consequently, presenting strategies for clarifying options becomes as important as providing relevant responses in SCS. Different presentation strategies may affect user satisfaction and their arrangement and format may reinforce certain types of biases (e.g., confirmation bias). this is a research challenge that has not been explored in SCS.

Can Voice Modulation Be Used to Characterize Cognitive Bias? While eye-tracking is not feasible in voice interactions, audio attributes (e.g., pitch and speed) from both the system and user reveal information about motivations, emotions, and personal traits [58]. For instance, Jiang et al. [46] indicated that perceived *information believability* is affected by the confidence in the voice of the system. Additionally, a recent work, found higher trust in female-voice agents that higher pitch reduces participants' decision-making reliance on the provided information [38]. These examples illustrate how voice modulation in systems affects information perception. Currently, we lack understanding of how system voice modulations might influence user beliefs or reinforce biases like *confirmation bias*, presenting an open research challenge. One potential solution is to slow down the system when discussing controversial opinions, allowing users ample time to absorb and consider. Besides, an important direction is the relationship between biases and user voice modulation. For instance, a skeptical tone and higher pitch when querying, "Is climate change REALLY [accentuate] happening?" may indicate *confirmation bias* towards the belief that climate change is not a real issue.

How to Leverage Content Manipulation to Mitigate Harms of Cognitive Bias? Cognitive bias does not always have a negative effect [69]. While it can skew perceptions and decisions, it also helps balance perspectives [53]. For instance, *Availability Bias* refers to placing greater importance on readily available or easily recalled information. A way to counteract it is by presenting less readily available information first. However, this solution may raise concerns about group fairness and misinformation spread. Recognizing and understanding the impact of cognitive bias helps address potential pitfalls and leverage its potential to create effective and user-friendly search experiences. Furthermore, audio interventions in voice-based conversations (e.g., nudging for clarifying questions [37], or warning users of presence of misinformation in a voice-based setting [19]) offer a potential solution to inform users of potential biases in SCS.

4 CASE STUDY: ARGUMENT SEARCH

Expanding on our identified research opportunities, we introduce a SCS specific use case called Spoken Conversational Argumentative Search (SCAS) and discuss its implications, data, topics, and methodology for experimentation (see Section 5). SCAS systems respond to a user's spoken query on controversial topics with multiple argument stances or viewpoints (i.e., PRO and CON). Users can rely on SCAS to provide them with balanced arguments on topics of interest. Let us consider an example in which, a user asks "is universal basic income good for society?". If the system only provides one side (i.e., PRO) of the issue, the user tends to be blind-sided by not having any information about other perspectives [36]. Such a biased exposure of perspectives is an important open challenge [85] if left unaddressed, may negatively impact society [12, 26, 114, 126]. Biases can arise from data itself as much as they can from algorithms [86] and presentation strategy in voice-only settings. Hence, choosing appropriate data is crucial when studying cognitive biases in SCS to control for unknown effects (from the data).

Data. For our specific case study, designing experiments requires argumentative topics (e.g., "should zoos exist?") and documents/passages supporting (PRO) and opposing (CON) the topics. A crowdsourced study by Draws et al. [26] collected opinions from 100 participants on 18 topics from the ProCon.org debate portal[4]; only a few topics identified with mild pre-existing viewpoints. Incorporating these topics into future experiments on cognitive bias is crucial to avoid heavily polarized subjects and better detect the effects of cognitive bias. The current dataset, with only 280 search results, may be insufficient for longer conversations. Therefore, we propose expanding the collection with the args.me corpus [2], which not only includes arguments with stances (PRO or CON) but also offers additional granularity by providing sub-topical perspectives (e.g., Capitalism, Healthcare, and Poverty) for each document. This increased granularity will also aid in mitigating unknown effects in future experiments.

5 METHODOLOGY

This section outlines an experimental framework for studying cognitive biases in SCS, covering potential experimental setups and data collection, including behavioral and physiological data. Additionally, we showcase preliminary results from an information-seeking experiment as an example of this approach and the potential of physiological data. The less structured nature of conversational interactions and the lack of clear indicators of comprehension or focus, i.e., listening effort (see Section 2.3), make it challenging to identify and measure specific biases in SCS. This section outlines an experimental framework for studying cognitive biases in SCS, including possible setups and measurements. Table 2 categorizes applicable measurements into Behavioral and Physiological Responses. It also showcases preliminary results as an example of this approach and the potential of physiological data. To accommodate the various needs of research questions and their associated experiments, including feasibility, scalability, research method (qualitative, quantitative, mixed), Table 1 covers potential experiment set-ups, including their advantages and disadvantages.

5.1 Measurements

Behavioral Responses. In SCS, *natural language utterances* function as queries [96] and the **Voice modulation** (see Section 3) of these raised queries distinguishes it from traditional screen-based search. In a case of rectifying system errors[5], users typically adjust volume, rephrase commands, or change pronunciation [120]. When the system's response contradicts their beliefs, users, especially those less tech-savvy, might confuse cognitive biases with system errors. They may then try familiar methods used for system errors to get preferred outcomes, potentially introducing bias. Speaking of querying, users' listening habits can also be used to investigate biases in SCS. Listening effort or speech intelligibility is assessed by the recalled accuracy, as indicators of attention and language-related cognitive processing [93], in **recall/recognition tasks** like word/sentence recognition and sentence comprehension [16]. This may also reveal biases, as users often comprehend biased information more easily due to lower cognitive load [10], but it still lacks granularity. It is worth noting that confounding variables like language proficiency [51] and working memory capacity [32, 89] can also impact listening performance [16] potentially introducing biases in information comprehension. A potential solution to address these pitfalls is adapting Brief-IAT [105] – a version of the Implicit Association Test (IAT) [39] designed to assess bias [25] . However, implementing a reliable bias assessment in the SCS remains an open challenge and requires more attention from the community.

Physiological Responses. Cognitive bias can be measured by examining differences in cognitive processes, emotions, and engagement. For instance, a user may be more engaged and emotionally aroused at the end of an audio segment. Multi-modal sensing with wearables can capture these responses, offering a scalable and comprehensive way to 'visualize' cognitive bias in SCS, analogous to using eye-tracking for screen-based IR systems. **Electroencephalography (EEG)** gathers brain electrical activity, aiding in studying cognitive and emotional processes such as memory, attention, and responses to stimuli [15, 55, 64]. EEG has shown promising results in web search, detecting relevance judgment at both article level [4, 41] and word level [123], and identifying information needs in Q&A scenarios [70]. Two common ways EEG signals are analysed [73] are Event-Related Potentials (ERP) and Frequency Band Analysis. **ERP** is a time-locked analysis describing cognitive activity after an event's onset [64]. Typically analyzing signals within a short time window (e.g., 1 second) [31, 70, 123], which may potentially help with detecting biases in each turn of a conversation in SCS. On the other hand, **Frequency Band Analysis** is typically used with longer stimuli durations (e.g., 1 minute) and can potentially help explore biases at the whole session level, rather than just per turn of the conversation. The latter explores various wave frequencies linked to cognitive states (e.g., alpha for attention [71, 74], theta for memory [72], beta for active thinking engagement [125]) [55]. The works above focus on brain waves in the frontal cortex related to human attention, memory, decoding, and retrieval. While they were explored in a screen-based IR context, we emphasize their potential in SCS as well, to explore cognitive biases. With current wearable EEG devices (e.g., headbands [78] and earbuds [29]) being integrated ubiquitously into earphones [1, 7], we foresee opportunities to expand research on biases in SCS through crowdsourced studies, thus lowering barriers for many researchers. Additionally, **peripheral signals** from commercial wearables, such as Electrodermal Activity (EDA), Photoplethysmography (PPG), and

[4]https://www.procon.org/debate-topics/ [Accessed: 9 Feb 2024]

[5]A *system error* occurs when the system fails to provide users with the desired results, regardless of whether it is caused by an incorrect response or lack of a response.

Table 1: A breakdown of different experiment set-ups (i.e., Lab, Field, and Crowdsourced) in *SCS*. LLM: large language model

Features	Lab Study	Field Study	Crowdsourced Study
Control	High	Low; unobserved factors in real-world	Moderate; depends on the design of platform or task
Data Quality	High and detailed; due to highly controlled and optimal environment	Low; real-world noise and factors may affect data	Moderate; less controlled than lab studies.
Scalability	Low; requires physical attendance on both participants and researchers	Moderate; enables more participants than lab studies but still limited	High; enables larger participant pool from diverse locations. LLM applications like Retrieval Augmented Generation (RAG) [61] show potential for controlled studies [83, 87]
Ecological Validity	Low; the artificial setting may influence behavior	High; since participants are in natural environments	Moderate; the absence of a physical entity (e.g., smart speaker) may influence user information perception [57]
Setup	Wizard of Oz (WOZ) [27, 111, 116]	Participants are provided with pre-configured voice agents and wearable devices to take home [120]. Comfortable and portable devices may facilitate longitudinal studies.	Crowdsourcing platforms like Prolific enable simulating always-on voice assistants for hypothetical scenarios. Consumer products like Apple AirPods with EEG [7] will make crowdsourced studies more feasible.
Related Works	[13, 45, 79, 109, 111, 113]	[118–120]	[43, 108]

Table 2: A Breakdown of studied measures by data type (Behavioral vs. Physiological) and user interaction mode (screen-based vs. voice). Bold text highlights studies on cognitive biases, emphasizing the limited research on cognitive biases in voice search (i.e., SCS).

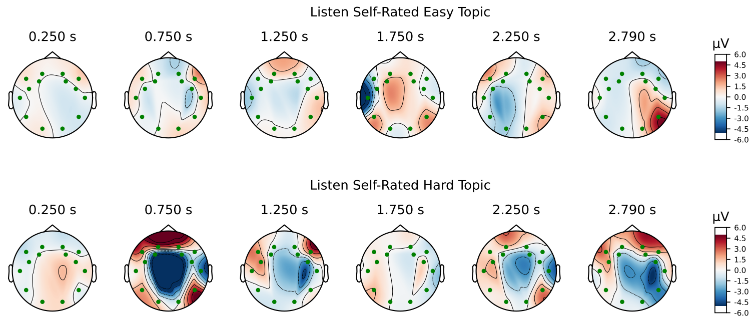
	Data Type	Screen-based		Voice	
		Construct	Related Work	Construct	Related Work
Behavioral	Web-logging (e.g., dwell time, clicks)	**Cognitive Bias**	[26, 59, 106]	–	
Behavioral	Transcripts & Voice Modulation (e.g., pitch, speed)	–		Perceived Trust	[38, 63]
Behavioral	Task Performance (e.g., sentiments of query/utterance, recall rate)	**Cognitive Bias** Search Experience	[30] [68, 98]	Listening Effort Search Experience	[16, 49, 51, 89, 97] [49, 98]
Behavioral	Motion, Facial Expression, Gaze	–		Engagement	[81, 84, 84]
Physiological	Brain Signals (e.g., EEG)	Cognitive Workload Search Experience **Cognitive Bias**	[47, 72] [4, 41, 70, 75, 123] [10, 71, 74, 125]	Perceived Trust	[46]
Physiological	Peripheral Sensing (e.g., EDA, PPG)	**Cognitive Bias**	[14, 71, 90]	–	
Physiological	Pupillary Responses	Selective Attention	[41, 93]	Selective Attention Distraction Listening Effort	[93] [65] [89]

Skin Temperature (SKT), can complement EEG [6, 15]. **EDA** measures the variations in skin's electrical conductance driven by sweat gland activity. **PPG** uses light to measure blood volume changes and to derive heart rate, blood oxygen levels, and other related metrics. **SKT** reflects the balance between the body's heat production and heat loss. These data indicate emotional responses from different aspects. For example, high arousal triggered by stressful events, often increase perspiration (sweating), leading to elevated EDA levels [9, 15, 54], or a rapid increase in heart rate (manifests as shorter intervals between PPG peaks [15, 54, 88]). Besides, EDA decreases when individuals are highly engaged (and thus less aroused) [54] and SKT generally decreases in low valence [54]. Furthermore, **pupillary responses** have been used to investigate selective attention [93], auditory distraction [65], and listening efforts [89]. For voice interaction, wearable eye-tracking glasses, e.g., Pupil Labs Neon glasses [56], can provide such a channel. However, pupil data is most suitable for lab studies with consistent lighting.

5.2 Preliminary Results

We used the EEG and EDA data collected by Ji et al. [44] for illustration purposes.[6] They collected various physiological signals from wearable devices in a lab study with simulated information search settings. Each participant completed a search task and rated the perceived difficulty in understanding the provided information on 12 topics. We analyzed data from 7 participants who received search results in audio formats on both most *easy* ($\overline{\mu}$ 1.3/5.0) and *hard* ($\overline{\mu}$ 3.0/5.0) topics (according to self-ratings). Although bias was not the target manipulation, the difficulty reveals changes in cognitive efforts

[6]EEG are cleaned following Eugster et al. [31], divided into 3sec segments. EDA are cleaned, baseline-corrected following Bota et al. [15], aggregated with a 1sec window.



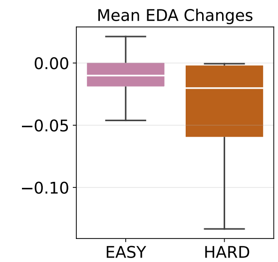


Figure 1: Preliminary EEG (left) and EDA (right) results ($N = 7$) of grand average on listening to search results (about 1 minute) on self-rated *easy* (Antarctica exploration – R03.353) and *hard* topics (Freighter ship registration – T04.743). In the left figure, deeper colors indicate greater neural activity. Cool colors (negative voltage) represent inhibitory, i.e., suppressing or restricting neural responses, while warm colors (positive) represent excitatory, i.e., promoting or enhancing responses [64]. The dots represent the placement of 14 electrodes.

required to receive the information. Figure 1 demonstrates clear differences between the *easy* and *hard* topics in both results. Overall, there was less neural activation on *easy*. Increased positive voltages around 1.75s in most regions suggest focused attention and engagement. Meanwhile, the left temporal negative may indicate reallocating cognitive resources from auditory processing to other areas needing more processing power. On *hard*, heightened activation was observed early at 0.75s. Pronounced prefrontal/frontal peaks suggest deeper processing and working memory load related to understanding the information. Enhanced activation at temporal regions, which handles the auditory and language processing, indicates increased comprehension effort and knowledge recall. EDA exhibits more consistency on *easy*, while much greater variability and fluctuation on *hard*. This suggests increased arousal or stress when absorbing difficult information in the audio. **In summary**, these preliminary results suggest that observing users' auditory information consumption is viable and warrants further exploration. Multi-modal signals may offer insights into fast and slow thinking systems [24, 48], and combining behavioral data with wearable signals could accurately identify user behavior, preferences and biases in SCS.

6 LESSONS LEARNED & ETHICAL CONSIDERATIONS

SCS interactions are less structured than screen-based interactions, which complicates analysis. Physiological data show distinct changes in receiving audio search results but interpreting cognitive biases is still complex. To ensure collect reliable data, these factors should be considered when designing the experiment: (i) data with more channels (e.g., 14+ channel EEG) offers direct insights but involves noise and requires specialized designs and expertise, while with fewer channels (e.g., peripherals) is easier to analyze, (ii) longer activities provide more reliable data, but SCS often involves short tasks, (iii) confounding variables like fatigue, interest, health, and specific activities (e.g., speech) may significantly impact. It is important to ensure optimal contact between sensors with specific body areas (e.g., see [9]). Furthermore, given biases are abstract concepts, the related hypotheses should be deconstructed into specific constructs, like engagement or cognitive load, and further into direct indicators that are measurable, reliable, and objective [95, 117], such as skin conductance or reaction time. During analysis, the requirements of signal processing on frequency can make certain features unavailable or distorted, especially those associated with high frequency in PPG [88]. Besides, analyzing SCS transcripts requires extensive effort and qualitative approaches as demonstrated in earlier works (see [110, 113]). For **ethical considerations**, it is crucial that informed consent and participant awareness of the exposure levels as physiological data could compromise privacy by revealing thoughts and emotions [127]. For example, the protocol used by Arnau-González et al. [5] could be adopted in this case. To protect cognitive liberty [92], caution is essential when developing strategies to mitigate biases using multi-modal signals for real-time content manipulation. It is also crucial to account for individual variations (e.g., minority groups, neurological conditions) for accurate and representative results [101].

Authors' Positionality. This paper reflects the perspectives shaped by the interdisciplinary backgrounds and views of our author team, which includes computer science researchers in information retrieval, conversational search, human-computer interaction, and pervasive computing. Some of the authors have significantly influenced these perspectives from their work on exploring cognitive bias in screen- or voice-based search, and personal experience as members of the neurodiverse community. The authors acknowledge the complexities surrounding cognitive biases. This paper aims to support a comprehensive discussion on understanding and utilizing biases in SCS. We acknowledge the gap in including perspectives from minority groups, First Nations peoples [60, 126], or people with disabilities.

7 CONCLUSIONS

Drawing insights from information-seeking, psychology, cognitive science, and wearable sensors, this paper highlights the under-explored area of cognitive biases in sophisticated voice-only systems like SCS, and advocates further research. We argue that traditional web search instruments are insufficient for studying cognitive biases and envision further research opportunities. Furthermore, we propose a general experimental approach for studying cognitive biases in SCS and report preliminary results demonstrating the feasibility and significance of using physiological responses. Additionally, we discuss the challenges and ethical considerations in adopting this approach.

ACKNOWLEDGMENTS

This research is supported by the Australian Research Council (CE200100005, DE200100064). We wish to extend our deep gratitude to Manasa Kalanadhabhatta, whose expertise in neuroscience was invaluable in ensuring the accuracy and clarity of the neuroscience-related sections.

REFERENCES

[1] Evan Ackerman and Eliza Strickland. 2022. Are You Ready for Workplace Brain Scanning?: Leveraging brain data will make workers happier and more productive, backers say. *IEEE Spectrum* 59, 12 (2022), 46–52. https://doi.org/10.1109/MSPEC.2022.9976479

[2] Yamen Ajjour, Henning Wachsmuth, Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. 2019. Data Acquisition for Argument Search: The args.me corpus. In *42nd German Conference on Artificial Intelligence (KI 2019)*, Christoph Benzmüller and Heiner Stuckenschmidt (Eds.). Springer, Berlin Heidelberg New York, 48–59. https://doi.org/10.1007/978-3-030-30179-8_4

[3] Mohammad Aliannejadi and Johanne R. Trippas. 2022. Conversational Information Seeking: Theory and Evaluation: CHIIR 2022 Half Day Tutorial. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 365–366. https://doi.org/10.1145/3498366.3505843

[4] Marco Allegretti, Yashar Moshfeghi, Maria Hadjigeorgieva, Frank E. Pollick, Joemon M. Jose, and Gabriella Pasi. 2015. When Relevance Judgement is Happening? An EEG-Based Study. 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, 719–722. https://doi.org/10.1145/2766462.2767811

[5] Pablo Arnau-González, Stamos Katsigiannis, Miguel Arevalillo-Herráez, and Naeem Ramzan. 2021. BED: A New Data Set for EEG-based Biometrics. *IEEE Internet of Things Journal* 8, 15 (2021), 12219–12230. https://doi.org/10.1109/JIOT.2021.3061727

[6] Resham Arya, Jaiteg Singh, and Ashok Kumar. 2021. A survey of multidisciplinary domains contributing to affective computing. *Computer Science Review* 40 (2021), 100399. https://doi.org/10.1016/j.cosrev.2021.100399

[7] Erdrin Azemi, Ali Moin, Anuranjini Pragada, Jean Hsiang-Chun Lu, Victoria M Powell, Juri Minxha, and Steven P Hotelling. 2023. Biosignal sensing device using dynamic selection of electrodes. US Patent App. 18/094,841.

[8] Leif Azzopardi. 2021. Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 27–37. https://doi.org/10.1145/3406522.3446023

[9] Ebrahim Babaei, Benjamin Tag, Tilman Dingler, and Eduardo Velloso. 2021. A Critique of Electrodermal Activity Practices at CHI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 177, 14 pages. https://doi.org/10.1145/3411764.3445370

[10] Bence Bago, Darren Frey, Julie Vidal, Olivier Houdé, Gregoire Borst, and Wim De Neys. 2018. Fast and slow thinking: Electrophysiological evidence for early conflict sensitivity. *Neuropsychologia* 117 (2018), 483–490. https://doi.org/10.1016/j.neuropsychologia.2018.07.017

[11] Garbiel Bénédict, Ruqing Zhang, and Donald Metzler. 2023. Gen-IR@SIGIR 2023: The First Workshop on Generative Information Retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 3460–3463. https://doi.org/10.1145/3539618.3591923

[12] Amin Bigdeli, Negar Arabzadeh, Shirin Seyedsalehi, Morteza Zihayat, and Ebrahim Bagheri. 2023. Understanding and Mitigating Gender Bias in Information Retrieval Systems. In *European Conference on Information Retrieval*. Springer, 315–323. https://doi.org/10.1007/978-3-031-28241-6_32

[13] Nattapat Boonprakong, Xiuge Chen, Catherine Davey, Benjamin Tag, and Tilman Dingler. 2023. Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases When Facing Different Opinions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 27, 19 pages. https://doi.org/10.1145/3544548.3580917

[14] Nattapat Boonprakong, Xiuge Chen, Catherine Davey, Benjamin Tag, and Tilman Dingler. 2023. Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases When Facing Different Opinions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 27, 19 pages. https://doi.org/10.1145/3544548.3580917

[15] Patricia J. Bota, Chen Wang, Ana L.N. Fred, and Hugo Placido Da Silva. 2019. A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE Access* 7 (2019), 140990–141020. https://doi.org/10.1109/ACCESS.2019.

2944001

[16] Danielle Bragg, Cynthia Bennett, Katharina Reinecke, and Richard Ladner. 2018. A Large Inclusive Study of Human Listening Rates. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174018

[17] Georg Buscher, Andreas Dengel, Ralf Biedert, and Ludger V. Elst. 2012. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Trans. Interact. Intell. Syst.* 1, 2, Article 9 (jan 2012), 30 pages. https://doi.org/10.1145/2070719.2070722

[18] Donald O Case, James E Andrews, J David Johnson, and Suzanne L Allard. 2005. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. *Journal of the Medical Library Association* 93, 3 (2005), 353.

[19] Sachin Pathiyan Cherumanal, Ujwal Gadiraju, and Damiano Spina. 2024. Everything We Hear: Towards Tackling Misinformation in Podcasts. *arXiv preprint arXiv:2408.00292* (2024). https://arxiv.org/pdf/2408.00292

[20] 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

[21] Russ Clay, Jessica M. Barber, and Natalie J. Shook. 2013. Techniques for Measuring Selective Exposure: A Critical Review. *Communication Methods and Measures* 7, 3-4 (2013), 147–171. https://doi.org/10.1080/19312458.2013.813925

[22] Michael J. Cole, Chathra Hendahewa, Nicholas J. Belkin, and Chirag Shah. 2014. Discrimination between Tasks with User Activity Patterns during Information Search. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Gold Coast, Queensland, Australia) *(SIGIR '14)*. Association for Computing Machinery, New York, NY, USA, 567–576. https://doi.org/10.1145/2600428.2609591

[23] Fabio Crestani and Heather Du. 2006. Written versus spoken queries: A qualitative and quantitative comparative analysis. *Journal of the American Society for Information Science and Technology* 57, 7 (2006), 881–890. https://doi.org/10.1002/asi.20350

[24] Kahneman Daniel. 2017. *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

[25] Tilman Dingler, Benjamin Tag, David A. Eccles, Niels van Berkel, and Vassilis Kostakos. 2022. Method for Appropriating the Brief Implicit Association Test to Elicit Biases in Users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 243, 16 pages. https://doi.org/10.1145/3491102.3517570

[26] 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

[27] Mateusz Dubiel, Martin Halvey, Leif Azzopardi, and Sylvain Daronnat. 2018. Investigating how conversational search agents affect user's behaviour, performance and search experience. In *The second international workshop on conversational approaches to information retrieval*. https://doi.org/10.1037/0022-3514.51.1.55

[28] Roger A Elkin and Michael R Leippe. 1986. Physiological arousal, dissonance, and attitude change: evidence for a dissonance-arousal link and a" don't remind me" effect. *Journal of personality and social psychology* 51, 1 (1986), 55.

[29] Emotiv. 2023. *MN8 - 2 Channel EEG Earbuds*. https://www.emotiv.com/mn8/ Accessed: October 21, 2023.

[30] 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

[31] Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2016. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific reports* 6, 1 (2016), 38580. https://doi.org/10.1038/srep38580

[32] Kara D Federmeier, Suzanne R Jongman, and Jakub M Szewczyk. 2020. Examining the role of general cognitive skills in language processing: a window into complex cognition. *Current directions in psychological science* 29, 6 (2020), 575–582. https://doi.org/10.1177/0963721420964095

[33] Christos Fidas, Stella Sylaiou, Tsvi Kuflik, Fotis Liarokapis, Panayiotis Koutsabasis, Manolis Wallace, Dimitrios Koukopoulos, Angeliki Antoniou, Katerina Mania, and Ioanna Lykourentzou. 2023. Mobile and Multimodal HCI Design Approaches in Museums for People with Impairments. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction* (Athens, Greece) *(MobileHCI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 33, 4 pages. https://doi.org/10.1145/3565066.3609509

[34] Alexander L Francis and Jordan Love. 2020. Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdisciplinary Reviews: Cognitive Science* 11, 1 (2020), e1514.

[35] Alexander Frummet, Andrea Papenmeier, Maik Fröbe, Johannes Kiesel, Vaibhav Adlakha, Norbert Braunschweiler, Mateusz Dubiel, Satanu Ghosh, Marcel Gohsen, Christin Kreutz, Milad Momeni, Markus Nilles, Sachin Pathiyan 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

[36] 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

[37] Marcel Gohsen, Johannes Kiesel, Mariam Korashi, Jan Ehlers, and Benno Stein. 2023. Guiding Oral Conversations: How to Nudge Users Towards Asking Questions?. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2023)*. ACM, New York, 34–42. https://doi.org/10.1145/3576840.3578291

[38] Kylie L. Goodman and Christopher B. Mayhorn. 2023. It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance. *Applied Ergonomics* 106 (2023), 103864. https://doi.org/10.1016/j.apergo.2022.103864

[39] Anthony G Greenwald, Debbie E McGhee, and Jordan LK Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of personality and social psychology* 74, 6 (1998), 1464.

[40] Ido Guy. 2018. The Characteristics of Voice Search: Comparing Spoken with Typed-in Mobile Web Search Queries. *ACM Trans. Inf. Syst.* 36, 3, Article 30 (mar 2018), 28 pages. https://doi.org/10.1145/3182163

[41] Jacek Gwizdka, Rahilsadat Hosseini, Michael Cole, and Shouyi Wang. 2017. Temporal Dynamics of Eye-Tracking and EEG During Reading and Relevance Decisions. *Journal of the Association for Information Science and Technology* 68, 10 (2017), 2299–2312. https://doi.org/10.1002/asi.23904

[42] Christopher G. Harris. 2019. Detecting Cognitive Bias in a Relevance Assessment Task Using an Eye Tracker. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 36, 5 pages. https://doi.org/10.1145/3314111.3319824

[43] Danula Hettiachchi, Zhanna Sarsenbayeva, Fraser Allison, Niels van Berkel, Tilman Dingler, Gabriele Marini, Vassilis Kostakos, and Jorge Goncalves. 2020. "Hi! I am the Crowd Tasker" Crowdsourcing through Digital Voice Assistants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376320

[44] 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, 12 pages. https://doi.org/10.1145/3626772.3657793

[45] Kaixin Ji, Damiano Spina, Danula Hettiachchi, Flora Dylis Salim, and Falk Scholer. 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* (Cancun, Quintana Roo, Mexico) *(UbiComp/ISWC '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 5 pages. https://doi.org/10.1145/3594739.3610679

[46] Xiaoming Jiang, Kira Gossack-Keenan, and Marc D Pell. 2020. To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology* 73, 1 (2020), 55–79.

[47] Angel Jimenez-Molina, Cristian Retamal, and Hernan Lira. 2018. Using psychophysiological sensors to assess mental workload during web browsing. *Sensors* 18, 2 (2018), 458.

[48] Daniel Kahneman. 2003. A Perspective on Judgment and Choice: Mapping Bounded Rationality. *American Psychologist* 58, 9 (2003), 697. https:

//doi.org/10.1037/0003-066X.58.9.697

[49] Clare-Marie Karat, Christine Halverson, Daniel Horn, and John Karat. 1999. Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 568–575. https://doi.org/10.1145/302979.303160

[50] Anna Khofman. 2023. Exploring cognitive biases in voice-based virtual assistants.

[51] Johannes Kiesel, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. 2018. Toward Voice Query Clarification. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (Ann Arbor, MI, USA) *(SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 1257–1260. https://doi.org/10.1145/3209978.3210160

[52] Johannes Kiesel, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. 2019. Clarifying False Memories in Voice-Based Search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK) *(CHIIR '19)*. Association for Computing Machinery, New York, NY, USA, 331–335. https://doi.org/10.1145/3295750.3298961

[53] Johannes Kiesel, Damiano Spina, Henning Wachsmuth, and Benno Stein. 2021. The Meant, the Said, and the Understood: Conversational Argument Search and Cognitive Biases. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 20, 5 pages. https://doi.org/10.1145/3469595.3469615

[54] Sylvia D. Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological Psychology* 84, 3 (2010), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010 The biopsychology of emotion: Current theoretical and empirical perspectives.

[55] J. Satheesh Kumar and P. Bhuvaneswari. 2012. Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study. *Procedia Engineering* 38 (2012), 2525–2536. https://doi.org/10.1016/j.proeng.2012.06.298 International Conference on Modelling Optimization and Computing.

[56] Pupil Labs. 2023. *Neon: Eye tracking for research and beyond*. https://pupil-labs.com/products/ Accessed: October 21, 2023.

[57] Sunok Lee, Minji Cho, and Sangsu Lee. 2020. What If Conversational Agents Became Invisible? Comparing Users' Mental Models According to Physical Entity of AI Speaker. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 88 (sep 2020), 24 pages. https://doi.org/10.1145/3411840

[58] Juan David Leongómez, Katarzyna Pisanski, David Reby, Disa Sauter, Nadine Lavan, Marcus Perlman, and Jaroslava Varella Valentova. 2021. Voice modulation: from origin and mechanism to social impact. , 20200386 pages. https://doi.org/10.1098/rstb.2020.0386

[59] Stephan Lewandowsky, KH Ecker Ullrich, and M Seifert Colleen. 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13, 3 (2012), 106–131. https://doi.org/10.1177/1529100612451018

[60] Jason Edward Lewis, Angie Abdilla, Noelani Arista, Kaipulaumakaniolono Baker, Scott Benesiinaabandan, Michelle Brown, Melanie Cheung, Meredith Coleman, Ashley Cordes, Joel Davison, Kūpono Duncan, Sergio Garzon, D. Fox Harrell, Peter-Lucas Jones, Kekuhi Kealiikanakaoleohaililani, Megan Kelleher, Suzanne Kite, Olin Lagon, Jason Leigh, Maroussia Levesque, Keoni Mahelona, Caleb Moses, Isaac ('Ika'aka) Nahuewai, Kari Noe, Danielle Olson, 'Ōiwi Parker Jones, Caroline Running Wolf, Michael Running Wolf, Marlee Silva, Skawennati Fragnito, and Hēmi Whaanga. 2020. Indigenous Protocol and Artificial Intelligence Position Paper. (2020). https://doi.org/10.11573/SPECTRUM.LIBRARY.CONCORDIA.CA.00986506

[61] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS'20)*. Curran Associates Inc., Red Hook, NY, USA, Article 793, 16 pages.

[62] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive Conversational Agents in the Post-ChatGPT World. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 3452–3455. https://doi.org/10.1145/3539618.3594250

[63] Mia Lindgren. 2023. Intimacy and emotions in podcast journalism: A study of award-winning Australian and British podcasts. *Journalism Practice* 17, 4 (2023), 704–719.

[64] Steven J. Luck and Emily S. Kappenman. 2016. *Electroencephalography and Event-Related Brain Potentials* (4 ed.). Cambridge University Press, 74–100. https://doi.org/10.1017/9781107415782.005

[65] Alexandre Marois, John E Marsh, and François Vachon. 2019. Is auditory distraction by changing-state and deviant sounds underpinned by the same mechanism? Evidence from pupillometry. *Biological Psychology* 141 (2019), 64–74.

[66] Andrés Martínez-Maldonado, Rosa Jurado-Barba, Ana Sion, Isabel Domínguez-Centeno, Gabriela Castillo-Parra, Julio Prieto-Montalvo, and Gabriel Rubio. 2020. Brain functional connectivity after cognitive-bias modification and behavioral changes in abstinent alcohol-use disorder patients. *International Journal of Psychophysiology* 154 (2020), 46–58. https://doi.org/10.1016/j.ijpsycho.2019.10.004 Neurophysiologic impact of non-pharmacologic interventions for cognition.

[67] David Maxwell and Leif Azzopardi. 2016. Agents, simulated users and humans: An analysis of performance and behaviour. In *Proceedings of the 25th ACM international on conference on information and knowledge management*. 731–740.

[68] Shiri Melumad. 2023. Vocalizing search: How voice technologies alter consumer search processes and satisfaction. *Journal of Consumer Research* (2023), ucad009.

[69] Miriam J Metzger and Andrew J Flanagin. 2013. Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of pragmatics* 59 (2013), 210–220.

[70] Dominika Michalkova, Mario Parra-Rodriguez, and Yashar Moshfeghi. 2022. Information Need Awareness: An EEG Study. 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, 610–621. https://doi.org/10.1145/3477495.3531999

[71] Randall K. Minas, Robert F. Potter, Alan R. Dennis, Valerie Bartelt, and Soyoung Bae. 2014. Putting on the Thinking Cap: Using NeuroIS to Understand Information Processing Biases in Virtual Teams. *Journal of Management Information Systems* 30, 4 (2014), 49–82. https://doi.org/10.2753/MIS0742-1222300403 arXiv:https://doi.org/10.2753/MIS0742-1222300403

[72] Zainab Mohamed, Mohamed El Halaby, Tamer Said, Doaa Shawky, and Ashraf Badawi. 2018. Characterizing focused attention and working memory using EEG. *Sensors* 18, 11 (2018), 3743.

[73] Santiago Morales and Maureen E. Bowers. 2022. Time-frequency analysis methods and their application in developmental EEG data. *Developmental Cognitive Neuroscience* 54 (2022), 101067. https://doi.org/10.1016/j.dcn.2022.101067

[74] Patricia Moravec, Randall Minas, and Alan R Dennis. 2018. Fake news on social media: People believe what they want to believe when it makes no sense at all. *Kelley School of Business research paper* 18-87 (2018).

[75] Yashar Moshfeghi and Joemon M. Jose. 2013. An effective implicit relevance feedback technique using affective, physiological and behavioural features. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Dublin, Ireland) *(SIGIR '13)*. Association for Computing Machinery, New York, NY, USA, 133–142. https://doi.org/10.1145/2484028.2484074

[76] Yashar Moshfeghi, Peter Triantafillou, and Frank Pollick. 2019. Towards predicting a realisation of an information need based on brain signals. In *The world wide web conference*. 1300–1309.

[77] Cosmin Munteanu and Gerald Penn. 2014. Speech-based interaction: myths, challenges, and opportunities. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 567–568. https://doi.org/10.1145/2628363.2645671

[78] Muse. 2023. *Muse EEG-Powered Meditation & Sleep Headband*. https://choosemuse.com/ Accessed: October 21, 2023.

[79] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3173574.3173580

[80] Adit Nair, Roshan Mathew, Zihao Mei, Rezylle Milallos, Pratheep Kumar Chelladurai, and Tae Oh. 2023. Investigating the User Experience and Challenges of Food Delivery Applications for the Blind and Visually Impaired. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction* (Athens, Greece) *(MobileHCI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 13, 6 pages. https://doi.org/10.1145/3565066.3608688

- [81] Yukiko I. Nakano and Ryo Ishii. 2010. Estimating User's Engagement from Eye-Gaze Behaviors in Human-Agent Conversations. In *Proceedings of the 15th International Conference on Intelligent User Interfaces* (Hong Kong, China) *(IUI '10)*. Association for Computing Machinery, New York, NY, USA, 139–148. https://doi.org/10.1145/1719970.1719990
- [82] Alamir Novin and Eric Meyers. 2017. Making Sense of Conflicting Science Information: Exploring Bias in the Search Engine Result Page. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR '17)*. Association for Computing Machinery, New York, NY, USA, 175–184. https://doi.org/10.1145/3020165.3020185
- [83] Julius Odede and Ingo Frommholz. 2024. JayBot – Aiding University Students and Admission with an LLM-based 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, 391–395. https://doi.org/10.1145/3627508.3638293
- [84] Ryota Ooko, Ryo Ishii, and Yukiko I Nakano. 2011. Estimating a user's conversational engagement based on head pose information. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*. Springer, 262–268. https://doi.org/10.1007/978-3-642-23974-8_29
- [85] Sachin Pathiyan Cherumanal. 2022. Fairness-Aware Question Answering for Intelligent Assistants *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 3492. https://doi.org/10.1145/3477495.3531682
- [86] Sachin Pathiyan 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
- [87] Sachin Pathiyan 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
- [88] Tam Pham, Zen Juen Lau, SH Annabel Chen, and Dominique Makowski. 2021. Heart rate variability in psychology: A review of HRV indices and an analysis tutorial. *Sensors* 21, 12 (2021), 3998. https://doi.org/10.3390/s21123998
- [89] Erin M. Picou, Todd A. Ricketts, and Benjamin WY. Hornsby. 2011. Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research* 54, 5 (2011). https://doi.org/10.1044/1092-4388(2011/10-0154)
- [90] Gavin W Ploger, Johnanna Dunaway, Patrick Fournier, and Stuart Soroka. 2021. The psychophysiological correlates of cognitive dissonance. *Politics and the Life Sciences* 40, 2 (2021), 202–212. https://doi.org/10.1017/pls.2021.15
- [91] Filip Radlinski and Nick Craswell. 2017. A theoretical framework for conversational search. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*. 117–126.
- [92] Stephen Rainey, Stéphanie Martin, Andy Christen, Pierre Mégevand, and Eric Fourneret. 2020. Brain recording, mind-reading, and neurotechnology: ethical issues from consumer devices to brain-based speech decoding. *Science and engineering ethics* 26 (2020), 2295–2311.
- [93] Mor Regev, Erez Simony, Katherine Lee, Kean Ming Tan, Janice Chen, and Uri Hasson. 2019. Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cerebral Cortex* 29, 10 (2019), 4017–4034.
- [94] Tobias Richter and Johanna Maier. 2017. Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist* 52, 3 (2017), 148–166.
- [95] René Riedl, Fred D Davis, and Alan R Hevner. 2014. Towards a NeuroIS research methodology: intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems* 15, 10 (2014), 4.
- [96] Ning Sa and Xiaojun Yuan. 2020. Examining users' partial query modification patterns in voice search. *Journal of the Association for Information Science and Technology* 71, 3 (2020), 251–263.
- [97] Pascal Schneiders. 2020. What remains in mind? Effectiveness and efficiency of explainers at conveying information. *Media and Communication* 8, 1 (2020), 218–231.
- [98] Vidya Setlur and Melanie Tory. 2022. How do you Converse with an Analytical Chatbot? Revisiting Gricean Maxims for Designing Analytical Conversational Behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 29, 17 pages. https://doi.org/10.1145/3491102.3501972
- [99] Chirag Shah and Emily M. Bender. 2022. Situating Search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 221–232. https://doi.org/10.1145/3498366.3505816
- [100] Nikhil Sharma, Q. Vera Liao, and Ziang Xiao. 2024. Generative Echo Chamber? Effects of LLM-Powered Search Systems on Diverse Information Seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, 17 pages. https://doi.org/10.1145/3613904.3642459
- [101] Laurianne Sitbon, Gerd Berget, and Margot Brereton. 2023. Perspectives of Neurodiverse Participants in Interactive Information Retrieval. *Found. Trends Inf. Retr.* 17, 2 (jul 2023), 124–243. https://doi.org/10.1561/1500000086
- [102] Anastasia Smirnova. 2020. Word Order Communicates User Intent in Search Queries. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3375207
- [103] Evgeny N Sokolov, John A Spinks, Risto Näätänen, and Heikki Lyytinen. 2002. *The orienting response in information processing*. Lawrence Erlbaum Associates Publishers.
- [104] Michael Soprano, Kevin Roitero, David La Barbera, Davide Ceolin, Damiano Spina, Gianluca Demartini, and Stefano Mizzaro. 2024. Cognitive Biases in Fact-Checking and Their Countermeasures: A Review. *Information Processing & Management* 61, 3 (2024), 103672. https://doi.org/10.1016/j.ipm.2024.103672
- [105] Natarajan Sriram and Anthony G Greenwald. 2009. The Brief Implicit Association Test. *Experimental psychology* 56, 4 (2009), 283–294.
- [106] Masaki Suzuki and Yusuke Yamamoto. 2021. Characterizing the Influence of Confirmation Bias on Web Search Behavior. *Frontiers in Psychology* 12 (2021), 771948.
- [107] John Sweller. 2011. CHAPTER TWO - Cognitive Load Theory. Psychology of Learning and Motivation, Vol. 55. Academic Press, 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8
- [108] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2020. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 153 (sep 2020), 23 pages. https://doi.org/10.1145/3369807
- [109] 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
- [110] Johanne R Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. 2017. A conversational search transcription protocol and analysis. In *Proc of sigir 1st international workshop on conversational approaches to information retrieval (cair'17), cair*, Vol. 17.
- [111] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. 2017. How Do People Interact in Conversational Speech-Only Search Tasks: A Preliminary Analysis. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR '17)*. Association for Computing Machinery, New York, NY, USA, 325–328. https://doi.org/10.1145/3020165.3022144
- [112] 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
- [113] Johanne R. Trippas, Damiano Spina, Paul Thomas, Mark Sanderson, Hideo Joho, and Lawrence Cavedon. 2020. Towards a Model for Spoken Conversational Search. *Information Processing & Management* 57, 2 (2020), 102162. https://doi.org/10.1016/j.ipm.2019.102162
- [114] Kees Van den Bos. 2018. *Why people radicalize: How unfairness judgments are used to fuel radical beliefs, extremist behaviors, and terrorism*. Oxford University Press.
- [115] Yolanda Vazquez Alvarez and Stephen A. Brewster. 2010. Designing spatial audio interfaces to support multiple audio streams. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 253–256. https://doi.org/10.1145/1851600.1851642

[116] Alexandra Vtyurina, Denis Savenkov, Eugene Agichtein, and Charles LA Clarke. 2017. Exploring conversational search with humans, assistants, and wizards. In *Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems*. 2187–2193.

[117] Emily Wall, Leslie M Blaha, Lyndsey Franklin, and Alex Endert. 2017. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *2017 ieee conference on visual analytics science and technology (vast)*. IEEE, 104–115.

[118] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2021. Developing the Proactive Speaker Prototype Based on Google Home. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 292, 6 pages. https://doi.org/10.1145/3411763.3451642

[119] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2022. Understanding User Perceptions of Proactive Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 185 (dec 2022), 28 pages. https://doi.org/10.1145/3494965

[120] Jing Wei, Benjamin Tag, Johanne R Trippas, Tilman Dingler, and Vassilis Kostakos. 2022. What Could Possibly Go Wrong When Interacting with Proactive Smart Speakers? A Case Study Using an ESM Application. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 276, 15 pages. https://doi.org/10.1145/3491102.3517432

[121] Andreas Wilke and Rui Mata. 2012. Cognitive bias. In *Encyclopedia of human behavior*. Academic Press, 531–535.

[122] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. 2020. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 34, 9 pages. https://doi.org/10.1145/3379503.3403563

[123] Ziyi Ye, Xiaohui Xie, Yiqun Liu, Zhihong Wang, Xuesong Chen, Min Zhang, and Shaoping Ma. 2022. Towards a Better Understanding of Human Reading Comprehension with Brain Signals. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 380–391. https://doi.org/10.1145/3485447.3511966

[124] Sofia Yfantidou, Pavlos Sermpezis, Athena Vakali, and Ricardo Baeza-Yates. 2023. Uncovering Bias in Personal Informatics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 139 (sep 2023), 30 pages. https://doi.org/10.1145/3610914

[125] Heeseung Yu and Eunkyoung Han. 2023. People see what they want to see: an EEG study. *Cognitive Neurodynamics* (2023), 1–15.

[126] Tyson Yunkaporta. 2023. *Right Story, Wrong Story: Adventures in Indigenous Thinking*. Text Publishing Company.

[127] Rafael Yuste. 2023. Advocating for neurodata privacy and neurotechnology regulation. *Nature Protocols* 18, 10 (2023), 2869–2875.

[128] 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.

[129] Arne Freya Zillich and Lars Guenther. 2021. Selective Exposure to Information on the Internet: Measuring Cognitive Dissonance and Selective Exposure with Eye-Tracking. *International Journal of Communication* 15 (2021), 20.