Improving Citizen Inclusivity through Conversational AI. The PROTECT Approach

Maristella Matera[1,*], Ludovica Piro[1], Emanuele Pucci[1], Maria Francesca Costabile[2], Rosa Lanzilotti[2] and Antonio Piccinno[2]

[1]Politecnico di Milano, DEIB, Piazza Leonardo da Vinci, 32, Milano, 20133, Italy
[2]Università degli Studi di Bari Aldo Moro, Dipartimento di Informatica, Via Orabona, 4, Bari, 70125, Italy

Abstract

This paper illustrates the research work of the PROTECT project, carried out as a collaboration between the HINT Human-centric INteractive Technology) lab of Politecnico di Milano and the IVU (Interaction, Visualization, Usability and UX) lab of the University of Bari. The goal of PROTECT is to exploit Conversational AI to define new methodologies and technologies that will increase the inclusivity of Web resources, offering benefits primarily to those users that can take advantage of conversational user interfaces, from blind and visually-impaired users to the elderly and other fragile populations. Addressing this fragile population is challenging, in particular for Public Administrations.

Keywords

Conversational AI, Digital Inclusion, Web Accessibility, AI for PAs

1. Introduction

Digital inclusion is a primary right for all citizens and represents a "must-have" for granting access to knowledge, education, and work. However, four out of ten people in Italy still do not use the Internet regularly, and over half of the population lacks basic digital skills [1]. This inequality is often linked to digital barriers. The development of accessible digital services thus becomes essential to guarantee everyone's right to access information and be included in every aspect of society.

Inclusion, active participation, dignity, and accessibility are the main objectives of regulatory interventions, such as the latest being the EU Directive (UE) 2019/882 [2] on the accessibility of products and services: from 2025, Member States will be required to grant all citizens, and specifically people with disabilities, full and effective participation in society, providing information access through multiple channels and perceivable and comprehensible modalities. Nevertheless, the Web remains a visual experience, inadequate for many users living with permanent or situational impairments. In these conditions, assistive technologies, e.g., screen readers, can help. Still, their reading paradigm is not without problems [3, 4]: many websites are hard to interpret as they are still designed regardless of accessibility and usability guidelines, and the information published on the Web is conceived strictly for visual consumption. Conversational AI is emerging as a technology for inclusive interaction with digital services [5], offering benefits to several users, from blind and visually-impaired users to the elderly and other fragile populations [6]. However, there is still a lack of concrete guidance for designers and developers about how to deliver effective conversational experiences on the Web. It is paramount to address this challenge by investigating how conversational interfaces can fill the gap in Web access inclusivity, especially in the context of services provided by a Public Administration (PA).

The main goal of the research project described in this article is to define a new paradigm sustaining the notion of *Conversational Web Browsing*, to enable users to browse the Web through Natural-Language interaction mediated by a Conversational Agent (CA). One limitation emerging from the literature is that current CAs help search and locate something interesting on the Web, but then they stop after opening a website [7]. To overcome this limitation, there is a need to i) deeply understand the users' needs by means of a *human-centered process* aimed at identifying the new paradigm for conversational Web browsing; ii) propose technological solutions for the *integration of AI models and Web architectures*; iii) investigate factors that can increase *users' trust* in the new technologies, through techniques for the transparency and explainability of Conversational AI models. Some authors of this article are already addressing the first two points by providing the notion of *Conversational Web* (ConWeb) and defining its challenges and patterns thanks to a human-centered approach [8]. The PROTECT

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
maristella.matera@polimi.it (M. Matera); ludovica.piro@polimi.it (L. Piro); emanuele.pucci@polimi.it (E. Pucci); maria.costabile@uniba.it (M. F. Costabile); rosa.lanzilotti@uniba.it (R. Lanzilotti); antonio.piccinno@uniba.it (A. Piccinno)

project aims at extending the analysis to PAs online resourcs; this requires the support of stakeholders who are PAs, government agencies, and non-profit associations. Some of them have already expressed their interest in the project outcomes and will be included in the design process we are defining for the extension of ConWeb.

After summarizing relevant related works in Section 2, Section 3 illustrates the ConWeb paradigm and how it can bring Conversational AI to enhance the inclusivity of online resources published by PA. Section 4 then draws our conclusions.

2. Related Works

2.1. Bringing Conversational AI to the Web

Conversational AI is adopted to grant access to data and services at different levels [9], from extending GUIs of apps and websites to adding natural language (NL) front ends to Web services, processes, and data repositories. On the Web, Conversational AI is exploited to build pop-up bots [9], i.e., assistants embedded within websites to offer conversational help and escalation to human support. However, these solutions do not focus on website navigation and content fruition. A tighter integration between websites and Conversational AI is achieved by multi-experience websites, which offer both visual and conversational interfaces on the same content and functionality [10]. However, the developer must still define the conversational experience by hand as a detached application. Other approaches leverage Conversational AI to offer alternative interaction paradigms for the Web. Cambre et al. [7] explore open Internet technologies to build a plugin for the Firefox browser that allows the users to express NL requests then translated into Google Search queries for locating specific content items on the Web. Ripa et al. [11] focus on facilitating the end-user generation of bots out of website content, relying on an annotation tool that lets the users structure the content feeding the bot and define the dialog flow. The end-user is in charge of conversation design. Some papers then promote the idea of a Conversational Web [12] to enable users, especially those challenged by visual interaction, to fulfill their Web browsing goals by engaging in conversations mediated by a CA. One key principle for this paradigm is enabling access to the Web even when websites are not equipped with ad-hoc conversational extensions. Progress toward this paradigm mainly refers to technical challenges and directions for tight integration between Web platforms and Conversational AI [13]. In general, there is still limited guidance on how to interpret the structure of an existing website and transpose content and functionality into conversational experiences.

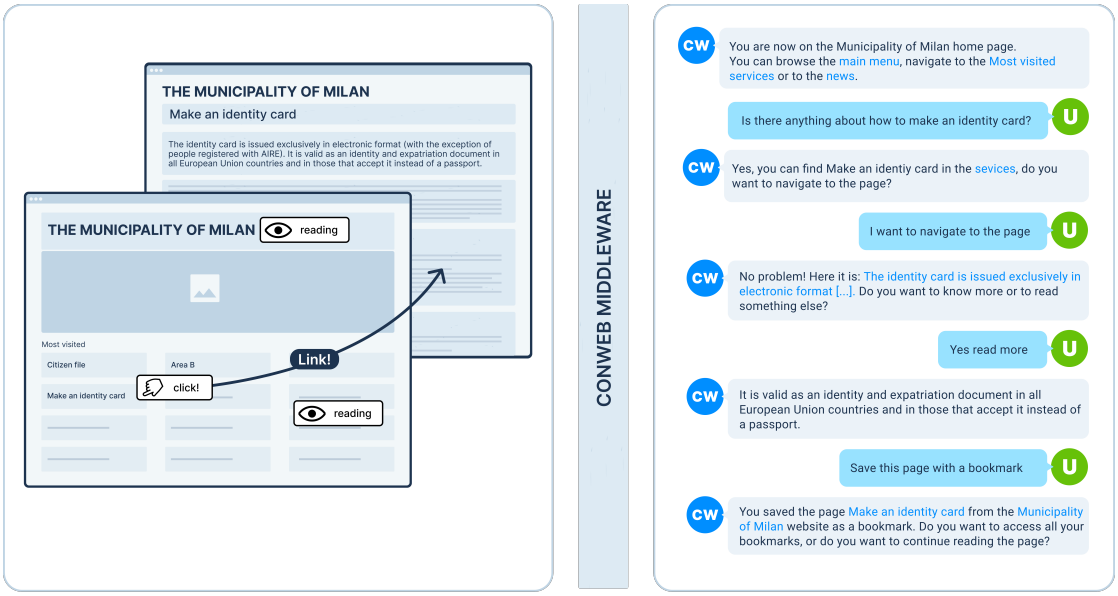
The ConWeb approach [8] tries to overcome this lack and proposes a paradigm that enables users to browse the Web through conversation. As an alternative to operating on graphical user interfaces using keyboards, mice, or screen readers, the users can express their browsing goals and access the websites through dialog-based interactions with a CA. The studies discussed in [8] show the effectiveness of this paradigm for people challenged by visual interaction. PROTECT will leverage these results and extend them to identify and validate an exhaustive library of conversational design patterns that can be natively offered by websites. One important aspect will be to identify how the benefits these patterns offer to blind and visually-impaired users can be extended to a larger population of users.

2.2. Design Guidelines and Patterns for Conversational AI

The literature highlights the benefits of conversational paradigms for a better user experience for people with disabilities, especially blind and visually-impaired people [14]. At the same time, several studies have highlighted the unmet needs of this population when it comes to designing CAs [5, 14, 15]. Prominent challenges relate to the input mechanisms, the control over the presented information, the interaction modalities, and even privacy when the users interact through voice [5, 14, 15]. To address these challenges, contributions go from general Human-AI interactions guidelines [16] to industry and platform-specific guidelines [17], and recommendations for accessible Conversational AI [18, 19]. Branham and Roy argue that these proposals do not properly meet the needs of people living with disability and in particular the blind population [3]. Among the efforts toward accessible Conversational AI, Leister et al. assessed the applicability of Web content accessibility guidelines (WCAG 2.1) for CA design and derived 23 design considerations for accessible conversational interfaces [18]. Stanley et al. synthesized their findings in 157 recommendations [19]. While extremely valuable, these accessibility design considerations are general and not targeted to the needs and capabilities of users living with a disability. More specific guidance comes from a few empirical studies that do leverage the skills and capabilities of blind users [20, 21]. However, these studies do not address Web browsing tasks. In general, design patterns for Conversational AI are unexplored [22]. One of the goals of the PROTECT project is to study and develop conversational patterns for Web browsing.

2.3. Trustworthy Conversational AI

Recent works highlighted how problems occurring in the interaction with CAs refer to the lack of trust in this tech-



(a) Automated Web browsing actions

(b) Natural language dialogue

Figure 1: Example of interaction with ConWeb on the Municipality of Milan website. (a) shows the simplified structure and possible visual actions on a Home Page of the Milan Municipality website, while (b) presents the interaction mediated by ConWeb to browse the same content through conversation. When ConWeb recognizes browsing intents of the user, it automatically builds the dialog in (b) for accessing the requested content and functionality and triggers related browsing actions in (a).

nology, especially in relation to the lack of transparency that NLP models exhibit when the content is transformed during the shift from text to the conversation [23, 18]. Some studies have highlighted the need for users to customize this transformation [24]. The users should for example be enabled to understand and customize mechanisms for content skimming, summarization, and indexing, to strengthen their control and avoid unwanted side effects of content filtering or nudging. Improving the transparency of Conversational AI models, designing adequate conversational patterns to provide explanations, and allowing the users to control and customize the models acquire even more priority when speaking about services provided by the PA.

3. Conversational AI for Public Administrations

The PROTECT goal is to extend the notion of ConWeb and its patterns [8] to public online services. Despite the aspects already addressed by the literature, several steps still need to be taken to provide completely accessible services. The starting point in PROTECT will be the knowledge about Conversational Web Browsing already acquired through the design of the ConWeb platform[8]. After describing the main characteristics of ConWeb, this section shortly introduces the challenges the PROTECT project will focus on.

3.1. ConWeb

The idea behind ConWeb has been developed thanks to an extensive human-centered study with blind and visually-impaired participants which comprehended interviews, focus groups, and co-design sessions to understand the main challenges behind Web navigation, and come up with strategies to overcome such challenges (i.e., *patterns*) [8].

To explain the main idea behind the framework, we illustrate a scenario of a user browsing the Municipality of Milan website Home Page by conversing with a conversational agent (e.g., a smart speaker or a voice-based browser plugin). As represented at the top of Fig. 1b, when landing on the Home Page, the user can be introduced by ConWeb to a short description along with the main organisation of the website. The user could also, at any point, ask about the content available in a given context, e.g., by uttering “Is there anything about how to make an identity card?”. The user can then navigate the website by following up on one of the available options (e.g., “I want to navigate to [...]”). As represented in

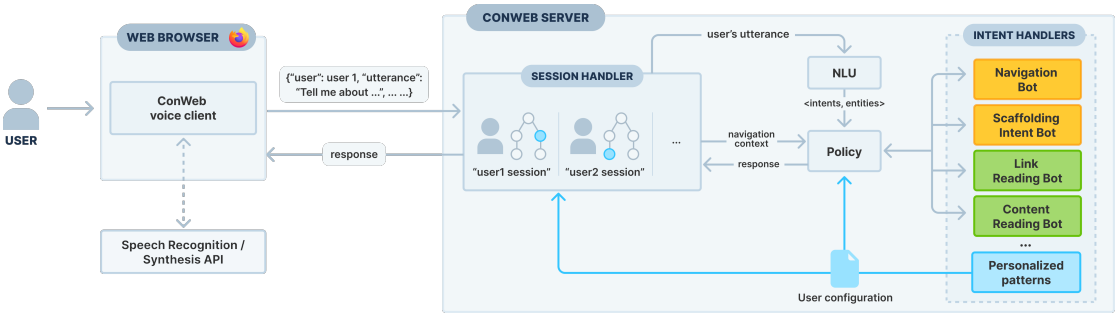


Figure 2: Conceptual architecture of the ConWeb platform. The *INTENT HANDLERS* support the flexible introduction of conversational patterns and personalization mechanisms [24].

Fig. 1a, these conversational requests also trigger navigation within or across pages in the website (e.g., from the Home to the identity card page). Ultimately, the user can browse the structure of the content or read the available content.

This interaction is organized around a library of conversational patterns identified through a human-centered process that involved 26 blind and visually-impaired users [8]. It covers basic navigation intents in the main phases of a Web browsing journey, offering mechanisms for the initial orientation, for navigation through intelligible and quick commands, and for digesting content through segmentation and summarization of page content. Scaffolding intents to control both the navigation and the conversation are also provided.

To enable such interaction, a middleware sitting between the user and the website identifies the offerings and content of the website that can be accessed through the conversational medium, interprets user intents and associated entities from user utterances, and automatically perform related actions on the website (e.g., click, extract information). Translating these interactions into architectural choices for the design of the ConWeb platform has required focusing on the following aspects [24]:

- A conversational-browsing model must be built when the website is first accessed to index and present to the users the available conversation nodes and the navigation structures that can sustain conversational browsing.
- A conversation node does not necessarily correspond to an entire Web page; it can be a content paragraph, a navigation menu, a link, or any other element in the Web page that can be presented independently from the others and has a role in the progressive exploration of the website content. A context representation characterising the navigation status must be handled to let the users move easily backward, i.e., along previous conversation nodes, and forward, i.e., to identify and explore new reachable nodes.
- To extract browsing-relevant intents and entities from the user utterances, an NLP engine must be adequately trained to start from the website content.
- Recognized intents and entities must be matched with navigation – and content-reading actions as deriving from the conversational-browsing model.
- The resulting CA must recognise website-specific intents as well as scaffolding intents related to auxiliary commands for the user to control the conversation.

Figure 2 describes the conceptual architecture of the current prototype of *ConWeb*. The *ConWeb voice client* manages the interaction with the conversational agent, also handling the transformation of the users' voice requests into text, and of the server responses into voice. In the current implementation, it is a plugin for the Firefox and Chrome Web browsers; we have already planned additional deployments through virtual assistants (e.g., Alexa), or on dedicated smart objects.

On the server side, the user's utterances are interpreted by an *NLU* engine (RASA[1] in the current implementation). The *Policy* module further elaborates the extracted intents and entities by contextualizing them with respect to the user's navigation tracked by a *SESSION HANDLER*. At the first access to a Web page, the SESSION HANDLER builds a "domain knowledge" by automatically extracting from the HTML code some features of the website content and functionality. Also, based on this domain knowledge, the *Policy* module can trigger the *INTENT HANDLERS* serving the user's request in a given navigation state. The Intent Handlers manage the conversation and perform the appropriate Web browsing actions on behalf of the user. Based on their output, the Policy module builds the response to be sent to the client.

[1]https://rasa.com/

3.2. Main Goals of the PROTECT Project

Based on the results already achieved with the ConWeb platform, PROTECT will pursue the following goals:

- Identifying challenges posed by current technologies for a more inclusive interaction with PA digital services on the Web, by means of a human-centred process including extended (large-scale) user studies.
- Defining design methodologies that embed design patterns for conversational user interfaces for Web browsing of PA applications.
- Designing technologies for integrating Conversational AI in Web platforms that will be available as open-source software and contribute to making proposals towards new standards for the Web.
- Designing techniques for AI model transparency and explainability to promote Conversational AI as a trustworthy technology for accessing the PA Web applications.

PROTECT will adopt a "research-through design" approach, in which prototypes will be created to explore foundational aspects. Since the beginning of the project, end users will be involved to elicit not just requirements but also the users' actual needs, practices, and values.

Rapid prototyping will support discussions on the opportunity (and the potential risks as well) of adopting new technologies for making websites inclusive. Rapid prototyping will also enable continuous co-design and user-based evaluation, with a unique opportunity to really involve users in discussions for assessing innovation, avoiding both dreamlike illusions and fearful attachment to customary practices.

The main technological outcome of the project will be a demonstrator of a platform for Conversational Web browsing, offering services that can support the conversational patterns identified through a human-centered process. This outcome will be complemented by the proposition of Web technologies extensions enabling a "by-design" integration of Conversational AI into Web platforms. The technical feasibility of this integration has already been investigated in previous studies conducted at the Politecnico di Milano [8]. Starting from these preliminary findings, PROTECT will develop and integrate new components handling the conversational patterns that support the interaction design choices deriving from the conducted user studies.

The validation of the approach will be conducted in collaboration with PAs (Milan and Bari Municipalities) and governmental working groups focusing on usability and accessibility (AgID and DGTCSI-ISCTI of the Ministero delle Imprese e del Made in Italy), that have expressed their interest in the project.

4. Conclusions

The PROTECT project aims to inquire about foundational aspects of technologies and interaction paradigms enhancing Web access inclusivity. Its main goal is to deliver methodological and technological tools for inclusive design and in particular: i) a toolkit for the design of CAs for Web access (i.e., design guidelines and a library of design patterns), and ii) open Web technologies (i.e., Web services and their APIs) that can enable the creation of conversation-augmented Web platforms and that in the end can also lead to the proposition of new Web technology standards. Other relevant outcomes will refer to the provision of educational materials that can help PAs adopt new development practices and can also teach Masters' and Ph.D. students how to approach inclusive design.

The applied methodology, guiding the project activities, is based on a "research-through design" approach, which emphasizes the production of prototypes as vehicles for inquiring about foundational aspects of a research challenge. This approach is commonly used in the field of Human-Computer Interaction and Interaction Design, while it is rarely used in other research fields, such as Web Engineering and AI.

The project's positioning is in the field of Human-Computer Interaction at the crossroads with the fields of Web Engineering and Artificial Intelligence. Specifically, it focuses on this new trend in investigating how to design AI systems that can empower human beings. The rising interest in this topic shows that there is strong demand for conceptual and methodological tools to design human-centered AI systems better.

Acknowledgements

We are grateful to the Milan Municipality (Direzione Innovazione Tecnologica e Digitale), the Bari Municipality, the Agenzia per l'Italia Digitale (AgID), the Direzione Generale per le Tecnologie delle Comunicazioni e la Sicurezza Informatica e la Sicurezza Informatica – Istituto Superiore delle Comunicazioni e delle Tecnologie dell'Informazione (DGTCSI-ISCTI) of the Ministero delle Imprese e del Made in Italy, for expressing their interest in the PROTECT project.

References

[1] RI22, Digital economy and society index (desi) 2022, la performance dei paesi europei nel campo della digitalizzazione, 2022. URL: www.researchitaly.mur.gov.i.

[2] EU19, European union. directive (eu) 2019/882 of the european parliament and of the council of 17

april 2019 on the accessibility requirements for products and services (text with eea relevance)., 2019. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0882.

[3] S. M. Branham, A. R. Mukkath Roy, Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users, in: Proc. of ASSETS 2019, ACM, New York, NY, USA, 2019, p. 446–458.

[4] S. Yan, P. G. Ramachandran, The current status of accessibility in mobile apps, ACM Trans. Access. Comput. 12 (2019).

[5] A. Abdolrahmani, R. Kuber, S. M. Branham, "siri talks at you": An empirical investigation of voice-activated personal assistant (vapa) usage by individuals who are blind, in: Proc of ASSETS 2018, ACM, 2018, p. 249–258.

[6] M. Heron, V. L. Hanson, I. W. Ricketts, Accessibility support for older adults with the access framework, Int. J. of Human-Computer Interaction 29 (2013) 702–716.

[7] J. Cambre, A. C. Williams, A. Razi, I. Bicking, A. Wallin, J. Tsai, C. Kulkarni, J. Kaye, Firefox voice: An open and extensible voice assistant built upon the web, in: Proc. of CHI 2021, ACM, New York, NY, USA, 2021.

[8] E. Pucci, I. Possaghi, C. M. Cutrupi, M. Baez, C. Cappiello, M. Matera, Defining Patterns for a Conversational Web, in: Proc. of CHI 2023, 2023, pp. 471–475.

[9] M. Báez, F. Daniel, F. Casati, B. Benatallah, Chatbot integration in few patterns, IEEE Internet Comput. 25 (2021) 52–59.

[10] E. Planas, G. Daniel, M. Brambilla, J. Cabot, Towards a model-driven approach for multiexperience ai-based user interfaces, Softw. Syst. Model. 20 (2021) 997–1009.

[11] G. Ripa, M. Torre, S. Firmenich, G. Rossi, End-user development of voice user interfaces based on web content, in: A. Malizia, S. Valtolina, A. Morch, A. Serrano, A. Stratton (Eds.), End-User Development, Springer International Publishing, Cham, 2019, pp. 34–50.

[12] M. Baez, F. Daniel, F. Casati, Conversational web interaction: Proposal of a dialog-based natural language interaction paradigm for the web, in: Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Springer-Verlag, Berlin, Heidelberg, 2019, p. 94–110.

[13] P. Chittò, M. Báez, F. Daniel, B. Benatallah, Automatic generation of chatbots for conversational web browsing, in: G. Dobbie, U. Frank, G. Kappel, S. W. Liddle, H. C. Mayr (Eds.), Proc. of ER 2020, volume 12400 of *LNCS*, Springer, 2020, pp. 239–249.

[14] A. Abdolrahmani, K. M. Storer, A. R. M. Roy, R. Kuber, S. M. Branham, Blind leading the sighted: Drawing design insights from blind users towards more productivity-oriented voice interfaces, ACM Trans. Access. Comput. 12 (2020).

[15] A. Pradhan, K. Mehta, L. Findlater, "accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities, in: Proceedings of CHI 2018, ACM, 2018, p. 1–13.

[16] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, E. Horvitz, Guidelines for human-ai interaction, in: Proceedings of CHI 2019, ACM, 2019, p. 1–13.

[17] WebAIM. Alexa brand guidelines for amazon developers., 2022. URL: https://developer.amazon.com/en-US/alexa/branding/alexa-guidelines.

[18] K. Lister, T. Coughlan, F. Iniesto, N. Freear, P. Devine, Accessible conversational user interfaces: Considerations for design, in: Proc. of the 17th International Web for All Conference, W4A '20, ACM, 2020.

[19] J. Stanley, R. t. Brink, A. Valiton, T. Bostic, B. Scollan, Chatbot accessibility guidance: A review and way forward, in: Proceedings of Sixth International Congress on Information and Communication Technology, Springer, 2022, pp. 919–942.

[20] D. Choi, D. Kwak, M. Cho, S. Lee, "Nobody Speaks That Fast!" An Empirical Study of Speech Rate in Conversational Agents for People with Vision Impairments, in: Proc. of CHI 2020, ACM, 2020, p. 1–13.

[21] J. Hong, C. Vaing, H. Kacorri, L. Findlater, Reviewing speech input with audio: Differences between blind and sighted users, ACM Trans. Access. Comput. 13 (2020).

[22] S. Bouguelia, H. Brabra, S. Zamanirad, B. Benatallah, M. Báez, H. Kheddouci, Reusable abstractions and patterns for recognising compositional conversational flows, in: M. L. Rosa, S. W. Sadiq, E. Teniente (Eds.), Proc. of CAISE 2021, volume 12751 of *LNCS*, Springer, 2021, pp. 161–176.

[23] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, J. D. Weisz, Expanding explainability: Towards social transparency in ai systems, in: Proc. of CHI 2021, ACM, New York, NY, USA, 2021.

[24] M. Baez, C. M. Cutrupi, M. Matera, I. Possaghi, E. Pucci, G. Spadone, C. Cappiello, A. Pasquale, Exploring challenges for conversational web browsing with blind and visually impaired users, in: CHI 2022 - Extended Abstracts, ACM, 2022.