Participatory Design for Creating Conversational Agents to Improve Web Accessibility

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

[1] *University of Bari Aldo Moro, 70126, Bari, Italy*
[2] *Politecnico di Milano, 20133, Milano, Italy*

Abstract

This contribution presents the PROTECT (imPROving ciTizEn inClusivity Through Conversational AI) project, which aims to improve web accessibility offering benefits to fragile populations through conversational agents, addressing primarily blind and visually impaired (BVI) users. The research adopts a participatory design approach, involving BVI users in the co-design and testing of the ConWeb prototypes, a middleware solution that facilitates conversational web browsing. This contribution reports about some performed user research and outlines the experimental design of a study that intends to investigate if and how CA technology can address some of the current limitations of screen readers, thereby significantly improving website accessibility and user experience.

Keywords

Conversational UIs, Participatory Design, PD, Conversational Web Browsing, Conversational Patterns, Conversational AI, BVI users

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 more than half of the population lacks basic digital skills [1]. This inequality is often linked to digital barriers. Thus, the development of accessible digital services 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, with the latest being the EU Directive (UE) 2019/882 [1] on the accessibility of products and services. Nevertheless, the Web remains a visual experience, inadequate for many users living with permanent or situational impairments, partly because assistive technologies, such as screen readers, can help, but they still have several problems.

Conversational AI has emerged as a technology for inclusive interaction with digital services, offering benefits to several users, from blind and visually impaired users to the elderly and other fragile populations. However, there is still a lack of concrete guidance for designers and developers about how to deliver effective conversational experiences on the Web [2];[3]. The PROTECT project (imPROving ciTizEn inClusivity Through Conversational AI) addresses this challenge by investigating how conversational interfaces can fill the gap in Web access inclusivity. As indicated on the website (https://protect.di.uniba.it/), the project has several goals, which include: i) identifying

Proceedings of the First International Workshop on Participatory Design & End-User Development - Building Bridges (PDEUD2024), October 2024, Uppsala, Sweden
* Corresponding author.
† These authors contributed equally.
maria.costabile@uniba.it (M. F. Costabile); rosa.lanzilotti@uniba.it (R. Lanzilotti); maristella.matera@polimi.it (M. Matera); antonio.piccinno@uniba.it (A. Piccinno); n.pinto15@studenti.uniba.it (N. Pinto); ludovica.piro@polimi.it (L. Piro); emanuele.pucci@polimi.it (E. Pucci); grazia.ragone@uniba.it (G. Ragone)
0000-0001-8554-0273 (M. F. Costabile); 0000-0002-2039-8162 (R. Lanzilotti); 0000-0003-0552-8624 (M. Matera); 0000-0003-1561-7073 (A. Antonio); 0009-0006-1849-7758 (N. Pinto); 0000-0002-5515-8396 (L. Piro); 0000-0003-2808-5619 (E. Pucci); 0000-0002-8774-1789 (G. Ragone)

the challenges posed by current technologies; ii) designing innovative technologies for the integration of Conversational AI in Web platforms; and iii) the design of techniques for AI model transparency and explainability to promote Conversational AI as a trustworthy technology for accessing the Web. The project goals are pursued with the support of stakeholders who are Public Administrations (PA), government agencies, and non-profit associations that are interested in the project.

We address here our primary goal: illustrating the human-centric process adopted to identify current challenges for more inclusive interactions with digital services on the web. In this first phase of the project, we target on blind or visually impaired (BVI) individuals, who generally use screen readers to access the web. They are involved in several sessions of our user research, aimed at understanding their needs and to identify conversational flows that can help BVI people to navigate the web without the barriers posed by current voice-based technologies. We adopt a participatory design approach (see for example [4, 5, 6]) by asking BVI people to co-design prototypes to explore design challenges and involving them in user testing to validate the developed prototypes.

After briefly discussing some works on Conversational AI for Web accessibility, we summarize the performed user research that also included a co-design session. We then describe the user study we are carrying out to investigate if and how CA technology can solve some of the current limitations of screen readers, with the aim of enhancing the website accessibility and user experience of BVI people.

As final remark, worth mentioning in this workshop, one of the goals of the PROTECT project is to define a component-based architectural model privileging flexibility for the integration and customization of conversational patterns to accommodate the variability of user needs. Specific attention will be posed to configurability aspects to provide the final users with proper End-User Development (EUD) mechanisms to give them the freedom to configure Web page access according to their specific needs and preferences [6]. Further EUD mechanisms will be considered to enable end users to take care of the security of their smart environments [7].

2. Bringing Conversational AI to the Web

Conversational AI for Web Accessibility is increasingly adopted to provide access to data and services at various levels [8], from extending graphical user interfaces 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 [8], 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 website and conversational AI is achieved by multi-experience websites, which offer both visual and conversational interfaces on the same content and functionality [9]. However, developers still must define the conversational experiences by hand as a separate application that possibly clones the same functionalities.

Some approaches leverage conversational AI to offer alternative interaction paradigms for the web. Cambre et al. [10] explore open Internet technologies to build a plugin for the Firefox browser that allows users to express NL requests, which are then translated into Google Search queries to locate specific content items on the Web. Ripa et al. [11] focus on facilitating the end-user generation of bots from website content using an annotation tool that lets users structure the content feeding the bot and define the dialog flow. The end-user is responsible for conversation design. Other papers promote the idea of a Conversational Web [12], enabling users, especially those challenged by visual interaction, to fulfill their web browsing goals through conversations mediated by a Conversational Agent (CA). Progress toward this paradigm mainly refers to technical challenges and directions for tight integration between Web platforms and Conversational AI [13, 14]. More recent research focuses on the integration of voice and graphical UI development, making it easier for developers to build dynamic, intent-based multimodal applications [15]. Leveraging the capabilities of Large

Language Models, the resulting applications allow users to specify intents in natural language and dynamically shape their user interface needed to perform the tasks related to the expressed intents.

The big providers, e.g., Microsoft and Google, have introduced interpretative capabilities into their browsers (Edge and Chrome, respectively) that mainly provide content summaries for the web pages the users land on[1]. Despite these advances, there is still limited guidance on how to interpret the structure of an existing website and transpose content and functionality into conversational experiences. Indeed, most of these approaches outsource the design of conversational interactions. Also, a general assumption is that adopting natural language for the expression of the user intent is enough to make conversational interfaces usable and accessible. The PROTECT project aims to identify reusable conversational design patterns that can maximize usability and accessibility and be natively offered by websites.

3. The User Research

Our research capitalizes on the results from a previous work [16], in which some co-authors of the Politecnico di Milano were involved. It consisted of a series of formative studies aimed at redefining web browsing for BVI users. These studies, which are summarized in this section, involved a co-design session, as reported below. Authorized by the Politecnico di Milano's ethical committee, the studies involved BVI participants contacted through Italian BVI associations.

First, unstructured interviews were conducted with three assistive technology experts, two of whom were blind. These experts educate and assist BVI people in learning and using assistive technologies. They provided valuable insights into the challenges faced by BVI people: the complexity of understanding websites structures, the difficulties with screen readers, and the limitations of current CAs in aiding web navigation. Following these interviews, a questionnaire was launched to gather broader input from the BVI community. Distributed through BVI associations, the questionnaire collected responses from 23 people, highlighting widespread issues with website accessibility and indicating barriers posed by confusing web page layouts. To explore these issues further, semi-structured interviews were conducted with BVI participants who were actively engaged with digital services and assistive technologies. These interviews confirmed the complexity of learning the structures of the websites, also due to the variability in the content organization of the pages. To refine the challenges identified in the previous steps, a focus groups was organized with 5 participants. Two researchers moderated the focus group, observing the participants as they first browsed the web with screen readers and then interacted with a CA designed specifically to browse selected web pages. In using the prototype, participants were asked to elicit insights, also providing concrete examples, on how they would envision browsing web sites mediated by a CA. The participants provide several useful comments about the way the CA supported initial orientation and content browsing

After one week, a co-design session was organized with the same 5 participants to the focus group, moderated by the same two researchers. The objective was to identify further preferences of the participants on how to organize conversation for web browsing, which was the main activity taken into account. The participants were then asked to define the conversation for browsing a website of their choice. They provide some initial ideas, and the researchers helped them implement the conversations using DialogFlow and deployed the prototypes on Google Assistant. Participants played with the prototype and modified it in several iterations to refine their ideas. From the analysis of the proposed solutions, navigation and content reading emerged as the most significant design dimensions to consider. This participatory design [4]; [5] process was very much appreciated and allowed participants to refine their ideas and provide direct feedback on the proposed solutions; as in many other cases, participatory design was crucial in creating a valuable user-centered prototype.

---

[1] See for example the co-pilot functions integrated in Microsoft Edge: https://www.microsoft.com/en-us/edge/learning-center/how-to-get-an-article-summary-using-ai?form=MA13I2

Participants were interviewed and observed interacting with the CA and with screen readers; this helped identify the challenges faced by BVI users and the characteristics needed to enhance their experience. Insights from this process guided the researchers in crafting a CA that addresses these challenges by implementing features based on participants' feedback.

Leveraging insights from the participatory design process, the research team developed ConWeb, a prototype designed to support a conversational web browsing paradigm, which incorporates conversational mechanisms for navigating, browsing, and reading web content. ConWeb functions as a sophisticated middleware that operates between the web client and web servers. Each time a webpage is accessed, ConWeb parses its HTML code and constructs a Content Navigation Tree (CNT) by identifying and organizing the page's elements, facilitating easier navigation for users.

The team conducted a preliminary remote evaluation of ConWeb with a small sample of 4 participants. This initial evaluation involved configuring a CA to navigate three Wikipedia pages, including the Home Page and two pages about the Solar System. The simplicity of the Wikipedia structure made it an ideal testbed for initial trials. After a brief introduction and general instructions, participants were assigned tasks to test orientation, content browsing, and content reading on various Wikipedia pages. The CA included scaffolding commands to aid navigation, and participants explored independently while researchers observed. Three of the four participants completed the tasks successfully. One participant encountered difficulties highlighting the need for more specific and context-aware responses from the CA. Despite this, users found the interaction natural. Participants reacted positively to the prototype, appreciating the natural and consistent voice-controlled navigation and ease of learning. In navigation tasks, users noted reduced mental effort due to the CA's simplification of navigation and helpful reminders. The direct questioning approach was praised for its efficiency in finding information. During content reading tasks, users favored the chunking and segmentation of content for intuitive navigation and appreciated the ability to pause and return to previous segments. The CA was considered reliable, though users desired more customization options. They saw potential for its use in augmenting screen readers.

4. The User Study

This section describes one of the current activities of the PROTECT project, i.e., the organization and execution of a user test involving a more significant user sample, with a wider age range and a balanced gender distribution; the aim is to validate the findings of the previous study, which only involved 4 participants, and to possibly provide further insights. Moreover, the study in [16] primarily focused on informational websites with regular structures, such as Wikipedia. We are now extending our research by considering the Bari municipality website (https://www.comune.bari.it/). The Bari municipality is one of the public administrations interested in the PROTECT project; its website presents a more complex and dynamic structure compared to Wikipedia.

4.1. Objectives, Research questions, Design and Participants

The main objective of our study is to investigate if and how the CA technology can solve some of the current limitations of screen readers, thereby significantly enhancing the website accessibility and user experience. The study consists in a test conducted with the "Thinking aloud" protocol, which also aims to acquire information on mental processes during the interaction. At the end of the test, two questionnaires are administered to the participant. The "thinking aloud" protocol involves instructing participants to verbalize their thoughts, feelings, and reactions as they interact with assistive technologies and conversational agents while web browsing. This provides researchers with valuable insights into real-time decision-making and user experience preferences. Observation is fundamental in our study. User behavior and reactions during the test sessions and afterwards can be closely monitored, thanks to the analysis of the audio and videos recording the interactions. This allows us to capture nuances and details that may not emerge through other forms of data collection, offering a more complete and in-depth view of users' experiences and needs.

For the test, we adopt a within-subject design, i.e., each participant is exposed to two conditions: A) interaction with the screen reader; B) interaction with ConWeb as conversational agent. Under both conditions, participants are encouraged to use the systems autonomously, without external intervention. This design allows us to directly compare the performance and experiences of the same participants using both modes of interaction. The study is conducted in a controlled environment to ensure consistency between conditions.

The participants involved in our study are adults, either blinds or with a severe visual impairment that affects their ability to use digital devices conventionally. They are contacted through an email sent by the presidents of local sections of the Italian Union of the Blind and Visually Impaired. In the e-mail they get a brief introduction to the study. At the time of writing this contribution, 11 participants already performed their test. Several other people already agreed to perform the test, so by the beginning of September we are confident to get at least 20 BVI people involved in the test. Thus, we will be able to discuss at the workshop the test findings.

4.2. Procedure

The procedure consists of the following three steps:

- Introduction. After signing a consent form for data collection during the study, participants complete a demographic questionnaire to collect data such as age, gender, computer skills, previous experience with assistive technologies, familiarity with specific websites, and other relevant information. The researchers introduce ConWeb through two demo presentations that provide some examples of interaction with the conversational agent.
- Task Execution. Each participant interacts with a screen reader, e.g., Jaws, NVDA (condition A) and with the ConWeb conversational agent prototype (condition B) to complete four specific tasks, one at a time. The order of conditions A or B follows a randomized sequence to ensure impartiality and reduce possible biases in data analysis. Video recording of each interaction is performed to capture and revisit every session aspect, ensuring no crucial details are missed during data analysis.
- Questionnaire Administration. For each condition, after executing the tasks, participants complete the SASSI [17] and UEQ [18] questionnaires.

4.3. Tasks

In [15], patterns for conversational web navigation were defined. The 4 tasks to be executed in the current study to test various aspects of the user experience have been defined according to some of these patterns.

Task 1. Find how to book an online appointment in the "Sportello Virtuale" section.
Pattern: View in the Large / Navigational Context
Description: The user interacts with the Homepage and the "Sportello Virtuale" section to identify options for booking an online appointment.

Task 2. Find the most visited services.
Pattern: View in the Large / Link Predictability
Description: The user explores the site to locate and preview the most visited services, facilitating navigation.

Task 3. Find the procedure for requesting the Electronic Identity Card.
Pattern: Exploration of Thematic Areas / Link Predictability
Description: The user explores thematic areas to find the procedure for requesting the Electronic Identity Card, with link previews to avoid unnecessary steps.

Task 4. Find how to pay the TARI waste tax.
Pattern: Exploration of Thematic Areas / Content Segmentation

Description: The user navigates through the website pages to find information on how to pay the TARI waste tax, with content segmented to facilitate reading.

The four tasks are executed by participants in a randomized order.

5. Ethical Considerations

This study follows ethical principles such as respecting human rights, ensuring privacy, and protecting participant well-being. Personal data will be confidential and anonymous, used only for research. Participants are informed about risks and benefits and gave informed consent. Data is securely stored and accessible only to the research team. Participants face no harm, discomfort, or consequences if they withdraw. All procedures comply with Ethics Review Committee guidelines [19].

6. Conclusion

This contribution has presented the PROTECT project, sponsored by the Italian Ministry of University and Research, that aims to investigate the potential of CAs to significantly improve website accessibility. The project primarily targets BVI people. By adopting a participatory design approach, the project ensures that the solutions developed are closely aligned with the real needs and preferences of the target users. The initial co-design and testing of the ConWeb prototype has shown that conversational agents have the potential to overcome the limitations of screen readers, providing a more intuitive and efficient way to navigate and interact with web content.

The aim of the described user study is testing the ConWeb prototype with a larger group of participants and a more complex and dynamic web structure to evaluate the robustness and versatility of the solutions. The study is ongoing, by involving more participants. Once we reach a suitable number of participants, we will analyze the results. Additional studies are planned in the project to establish comprehensive guidelines and best practices for designing conversational web interfaces that are universally accessible, thereby contributing to greater digital inclusion and equity. Another goal of the project, which is of interest to this workshop, is to address configurability aspects. We will capitalize on our experience on EUD to study and experiment proper EUD mechanisms capable to give end users the freedom to configure Web page access according to their needs and preferences.

Acknowledgements

This work is partially supported by the Italian Ministry of University and Research (MUR) under grant PRIN 2022 “PROTECT project (imPROving ciTizEn inClusivity Through Conversational AI)”—CUP: H53D23008150001.

References

[1] 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), Technical Report, https://eur-lex.europa.eu/eli/dir/2019/882/oj, 2019. URL: https://eur-lex.europa.eu/eli/dir/2019/882/oj.

[2] Følstad, A., Araujo, T.B., Law, E.L., Brandtzaeg, P.B., Papadopoulos, S., Reis, L., Báez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. Computing, 103, 2915 - 2942. https://doi.org/10.1007/s00607-021-01016-7

[3] Akram, S., Buono, P. & Lanzilotti, R. Recruitment chatbot acceptance in a company: a mixed method study on human-centered technology acceptance model. Pers Ubiquit Comput (2024). https://doi.org/10.1007/s00779-024-01826-4

[4] Schuler, D., & Namioka, A. (Eds.). (1993). Participatory Design: Principles and Practices (1st ed.). CRC Press.
[5] Fogli, D., Parasiliti Provenza, L. (2009) Information System Customization-Toward Participatory Design and Development of the Interaction Process, in: Proceedings of: International Conference on Enterprise Information Systems (ICEIS), Milan, Italy, 2009, pp. 72-77. https://doi.org/10.5220/0001992400720077
[6] Ardito, C., Buono, P., Costabile, M. F., Lanzilotti, R., & Piccinno, A. (2012). End users as co-designers of their own tools and products. Journal of Visual Languages & Computing, 23(2), 78-90. https://doi.org/10.1016/j.jvlc.2011.11.005
[7] Breve B., Desolda G., Deufemia V., Greco F., Matera M (2021). An End-User Development Approach to Secure Smart Environments. In Proc. of International Symposium on End User Development (IS-EUD '21). Springer, LNCS, vol. 12724, pp. 36-52. https://doi.org/10.1007/978-3-030-79840-6_3
[8] Baez, M., Daniel, F., Casati, F., and Benatallah. B. (2020). Chatbot integration in few patterns. IEEE Internet Computing 25, 3 (2020), 52–59. https://doi.org/10.1109/MIC.2020.302460
[9] Planas, E., Daniel, G., Brambilla, M., Cabot, J. (2021) Towards a model-driven approach for multiexperience ai-based user interfaces., Softw Syst Model 997–1009. https://doi.org/10.1007/s10270-021-00904-y
[10] Cambre, J., Williams, A., C., Razi, A., Bicking, I., Wallin, A., Tsai, J., Kulkarni, C., Kaye, J. (2021) Firefox voice: An open and extensible voice assistant built upon the web. In: Proceedings of the 2021 CHI conference on human factors in computing systems, pp. 1–18. https://doi.org/10.1145/3411764.3445409
[11] Ripa, G., Torre, Firmenich, M., S., Rossi, G. (2019), End-user development of voice user interfaces based on web content. In: End-User Development, Springer International Publishing, Cham, 2019, pp. 34–50. https://doi.org/10.1007/978-3-030-24781-2_3
[12] Baez, M., Daniel, F., Casati, F. (2019) 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, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers, Springer-Verlag, Berlin, Heidelberg, 2019, p. 94–110. https://doi.org/10.1007/978-3-030-39540-7
[13] Chittò, P., Baez, M., Daniel, F., Benatallah, B. (2020) Automatic generation of chatbots for conversational web browsing. In: G. Dobbie, U. Frank, G. Kappel, S. W. Liddle, H. C. Mayr (Eds.), Conceptual Modeling, Springer International Publishing, Cham, pp. 239–249. https://doi.org/10.1007/978-3-030-62522-1_17
[14] Pina, A., Baez, M., Daniel, F. (2020) Bringing cognitive augmentation to web browsing accessibility. In: H. Hacid, F. Outay, H.-y. Paik, A. Alloum, M. Petrocchi, M. R. Bouadjenek, A. Beheshti, X. Liu, A. Maaradji (Eds.), Service-Oriented Computing – ICSOC 2020 Workshops, Springer International Publishing, Cham, 2021, pp. 395–407. https://doi.org/10.1007/978-3-030-76352-7_37
[15] Yang, J., Shi, Y., Zhang, Y., Li, K., Wan Rosli, D., Jain, A., Zhang, S., Li, T., Landay, J.A., and Lam, M.L. (2024). ReactGenie: A Development Framework for Complex Multimodal Interactions Using Large Language Models. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 483, 1–23. https://doi.org/10.1145/3613904.3642517
[16] Pucci, E., Possaghi, I., Cutrupi, C.M., Baez, M., Cappiello, C., Matera, M., (2023) Defining patterns for a conversational web. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. https://doi.org/10.1145/3544548.3581145
[17] Hone, K., Graham, R. (2000), Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). Natural Language Engineering. 2000;6(3-4):287-303. https://doi.org/10.1017/S135132490000249

[18] Laugwitz, B., Schrepp, M. & Held, T. (2008). Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (Ed.): USAB 2008, LNCS 5298, pp. 63-76. https://doi.org/10.1007/978-3-540-89350-9_6

[19] Research Ethics Review Committee. (2024). Guidelines on submitting research proposals for ethics review. World Health Organization. https://www.who.int/groups/research-ethics-review-committee/guidelines-on-submitting-research-proposals-for-ethics-review.