
WidgetExplorer: A Rapid Prototyping and Evaluation Platform to Design Chatbots for all User Groups

Patrick Stadler

German Research Center for
Artificial Intelligence (DFKI)
Berlin, Germany
patrick.stadler@dfki.de

Stefan Schaffer

German Research Center for
Artificial Intelligence (DFKI)
Berlin, Germany
stefan.schaffer@dfki.de

Abstract

The design and evaluation of Conversational User Interfaces (CUI) are a resource-intensive and complex process. To be as user-friendly and accessible to all users as possible, many different factors that can have an influence on such aspects must be considered when creating CUI applications (e.g., chatbots). Marginalized groups should be involved in the design process from the very beginning to make the application equally accessible to all people. We present WidgetExplorer, a tool for rapid prototyping and evaluating the design of CUIs. With WidgetExplorer, various aspects of conversation design, such as variants of visual elements of a CUI (i.e., widgets) or different wordings, can be sketched and tested in parallel and evaluated in any browser, lowering the technical barriers for both, stakeholders, and participants.

Author Keywords

CUI-Design; Prototyping; Evaluation; Co-Design

CCS Concepts

•**Human-centered computing** → **Usability testing**; Accessibility systems and tools;

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).

CUI@CHI'23, April 23, 2023, Hamburg, Germany

Introduction

The evaluation of design choices, user experience and usability of artificial intelligence applications is complex and has not been sufficiently addressed in scientific studies (Yang, Banovic, and Zimmerman 2018). Furthermore, marginalized groups and users are often not sufficiently included in the evaluation (e.g., elderly people, cf. Sin and Munteanu (2020)). Research shows that it is useful to involve laymen in the design process on Conversational User Interfaces (CUI) (Sun et al. 2022). Especially for projects with vulnerable groups, e.g., elderly, or sick people, care should be taken to integrate these people into the process as inclusively as possible, even if they are not tech experts. Another hurdle may exist due to the design of many user studies when potential subjects need to be on site to participate in the experiments or design sessions. This is where remote participation options might come in handy, especially when these sessions are independent of time and location (Walsh and Wronsky 2019). Therefore, a methodology is needed on how to involve as many different users as possible with little effort in the design process of a CUI application, for example a chatbot, so that the most diverse assessment of the later application can take place. This is also referred to as co-design.

In this context, design, and creation of inclusive chatbots are time-consuming and expensive in terms of training and adaptation of “real” dialog backends. At the same time, there are many adjustable parameters that influence the usability of chatbots (cf. Luger and Sellen 2016), e.g., font (cf. Candello, Pinhanez, and Figueiredo 2017), emojis (cf. Fadhil et al. 2018), anthropomorphic design (cf. Donkelaar 2018; Rietz, Benke, and Maedche 2019), and many other open questions and research approaches for design of VUIs / CUIs, including e.g., older people (cf. Sayago, Neves, and Cowan 2019). To be able to test all these as-

pects, chatbot prototypes must be developed over and over, be it low fidelity in the form of paper prototypes, real chatbots via Facebook Messenger or other systems (e.g., Rasa), just to be able to evaluate certain aspects of usability. This takes a lot of time and resources, plus there are often technical challenges especially with ML-based chatbots.

In this paper we present WidgetExplorer, a modular platform for rapid prototyping of chatbot interactions, where different scenarios can be compared against each other in parallel over the internet. Because WidgetExplorer is a web application, only a web browser is needed to sketch a user study and to participate. This can reduce the requirements and allow for simplified remote usability testing.

Rapid Prototyping of CUIs

There is a need for rapid prototyping (RP) in usability evaluation of applications with artificial intelligence elements, for example CUIs and VUIs. Generally, there are two common ways to achieve this: Wizard-of-Oz (WoZ) and prototypes (Yang et al. 2020). Free and commercial chatbot prototyping systems (often no-code) exist, e.g., SUEDE (Klemmer et al. 2000). However, even here, a chatbot usually must be assembled first in an elaborate way. WoZ approaches can also help to test a system independently of a real dialog system, but such experiments always require a wizard; the experiments cannot be performed without them (time, cost, effort). Moreover, certain aspects and consequences of artificial intelligence cannot be measured (Yang et al. 2020). Especially for small-scale studies, e.g., formulations in direct comparison, the effort is very high, because sometimes different chatbots must be developed and distributed. Direct parallel comparison is usually not possible, and comparisons are done sequentially. With an RP platform, such design decisions can be tested in advance with real users

Co-Design

"The term co-design refers to proactively involving non-designers including end-users, internal team members specialized in non-design disciplines, external stakeholders, or anyone affected by design in the design process" (Chen et al. 2020)

HCI evaluation and AI

"Challenges in iterative prototyping and testing human-AI interaction (in both convergent thinking stages): One core practice of HCI design and innovation is rapid prototyping, assessing the human consequences of a design and iteratively improving on it." (Yang et al. 2020)

from all groups and especially from disadvantaged groups. Simulations can combine the advantages of both worlds: a fast, low-cost prototype and controlled user interaction.

WidgetExplorer

WidgetExplorer is a rapid prototyping platform for design evaluations of CUIs and chatbots. In addition to different UI components (cf. Valério et al. 2017), so-called widgets, variants of dialogs can also be evaluated, e.g., which wording suits better for a certain use case or how different avatars affect the user acceptance. The option to display multiple variants next to each other opens new possibilities for usability evaluation. In addition, the entire application can be hosted as a service on the Internet, eliminating the need for costly in-place experiments. With WidgetExplorer, up to three variants can be displayed side by side, with the first variant colored in red, the second in green, and the third in blue. To rate, the subjects are to select the variant they liked the most based on the associated colored card. In addition, there are the two additional options “All variants are equally good” and “Wouldn’t use any variant” per sub-experiment for evaluation. This makes the evaluation process in general easy to carry out even by laypersons and people with less technical experience, which facilitates an inclusive design process. Separate “Play” buttons allow subjects to determine the order in which the scripted sub-experiments are played. In addition, each sub-experiment can be replayed as often as needed.

The dialogs and steps are pre-scripted, i.e., stories are created based on a pre-defined flow path and actions (see box “Available action types”). Such actions include waiting for a user interaction or running a timer. Internally, the stories are defined via a JSON structure. However, programming knowledge is not necessary as the entire process of scripting, previewing and deploying of a WidgetExplorer study

Available action types

For the description of the simulated dialogues by means of stories the following action types can be used in WidgetExplorer:

- Sending a bot message
- Show dialog
- Wait for user input
- Simulate user input in the text field
- Show widget
- Wait for a timer
- Simulate a click on a button a widget

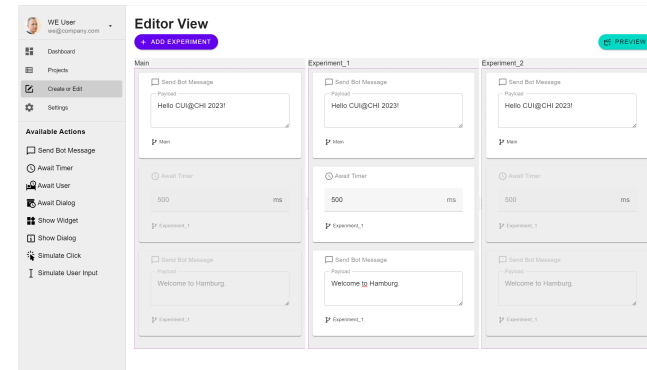


Figure 1: The provided frontend for maintaining and prototyping WidgetExplorer studies in a visual way

can be done in a visual way by using the provided frontend (see figure 1). This ensures that anybody can prototype and sketch dialogs and experiments, regardless of their technical knowledge. This can help to open the process of prototyping chatbot applications to a broader audience making it more inclusive and inviting.

A WidgetExplorer study consists of one or more experiments, in one or more variants (conditions). The dialogs can be branched at any point, i.e., depending on the condition. As WidgetExplorer is a low-code or no-code application, this is also done by using the provided frontend. Only the process of setting up the application by deploying the web apps and the database service requires some technical expertise (i.e., deploying a docker container on a web server).

Possible applications and benefits

Here are examples of what WidgetExplorer can be used for and why in what ways it can be helpful for inclusive design:

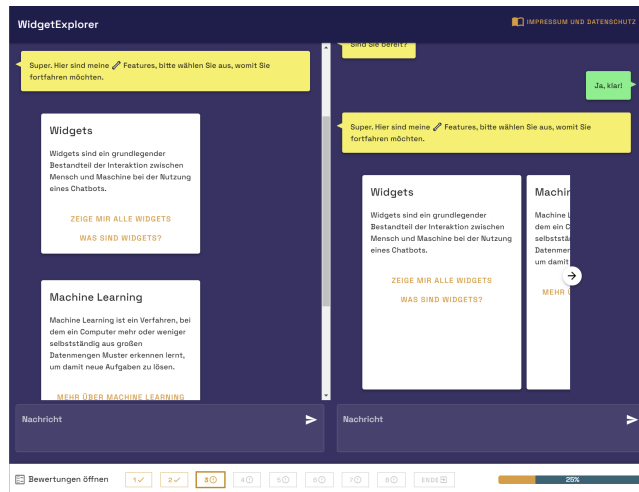


Figure 2: Evaluation of two presentation options for a carousel widget

Conversation design for inclusive language (various formulations); participatory design process during life cycle of the CUI; investigate which widgets and which widget variations are useful in different situations and contexts.

Contribution to the CUI@CHI Workshop

WidgetExplorer is intended to be a tool for designing accessible CUIs in early stages of development. A first study provides empirical evidence that using WidgetExplorer users can recognize even subtle differences between different widget variants. A real dialog system is not necessarily required to evaluate chatbot components. Simulations can also provide realistic data and impressions that can be crucial, especially in the design phase. We want to share our insights on CUI design and discuss how our approach could improve the design of inclusive CUIs.

- Candello, Heloisa, Claudio Pinhanez, and Flavio Figueiredo. 2017. "Typefaces and the Perception of Humanness in Natural Language Chatbots." In *Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems*, 3476–87. <https://doi.org/10.1145/3025453.3025919>.
- Donkelaar, Laury. 2018. "How Human Should a Chatbot Be?: The Influence of Avatar Appearance and Anthropomorphic Characteristics in the Conversational Tone Regarding Chatbots in Customer Service Field." PhD thesis, University of Twente.
- Fadhil, Ahmed, Gianluca Schiavo, Yunlong Wang, and Bereket A. Yilma. 2018. "The Effect of Emojis When Interacting with Conversational Interface Assisted Health Coaching System." In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 378–83. New York NY USA: ACM. <https://doi.org/10.1145/3240925.3240965>.
- Klemmer, Scott R., Anoop K. Sinha, Jack Chen, James A. Landay, Nadeem Aboobaker, and Annie Wang. 2000. "Suede: A Wizard of Oz Prototyping Tool for Speech User Interfaces." In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology - UIST '00*, 1–10. San Diego, California, United States: ACM Press. <https://doi.org/10.1145/354401.354406>.
- Luger, Ewa, and Abigail Sellen. 2016. "“Like Having a Really Bad PA”: The Gulf Between User Expectation and Experience of Conversational Agents." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–97. San Jose California USA: ACM. <https://doi.org/10.1145/2858036.2858288>.
- Rietz, Tim, Ivo Benke, and Alexander Maedche. 2019. "The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance."
- Sayago, Sergio, Barbara Barbosa Neves, and Benjamin

- R Cowan. 2019. "Voice Assistants and Older People: Some Open Issues." In *Proceedings of the 1st International Conference on Conversational User Interfaces - CUI '19*, 1–3. Dublin, Ireland: ACM Press. <https://doi.org/10.1145/3342775.3342803>.
- Sin, Jaisie, and Cosmin Munteanu. 2020. "An Empirically Grounded Sociotechnical Perspective on Designing Virtual Agents for Older Adults." *Human-Computer Interaction* 35 (5-6): 481–510. <https://doi.org/10.1080/07370024.2020.1731690>.
- Sun, Lu, Yuhan Liu, Grace Joseph, Zhou Yu, Haiyi Zhu, and Steven P Dow. 2022. "Comparing Experts and Novices for AI Data Work: Insights on Allocating Human Intelligence to Design a Conversational Agent." In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 10:195–206. <https://doi.org/10.1609/hcomp.v10i1.21999>.
- Valério, Francisco A. M., Tatiane G. Guimarães, Raquel O. Prates, and Heloisa Candello. 2017. "Here's What I Can Do: Chatbots' Strategies to Convey Their Features to Users." In *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, 1–10. Joinville Brazil: ACM. <https://doi.org/10.1145/3160504.3160544>.
- Walsh, Greg, and Eric Wronsky. 2019. "AI + Co-Design: Developing a Novel Computer-supported Approach to Inclusive Design." In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, 408–12. Austin TX USA: ACM. <https://doi.org/10.1145/3311957.3359456>.
- Yang, Qian, Nikola Banovic, and John Zimmerman. 2018. "Mapping Machine Learning Advances from HCI Research to Reveal Starting Places for Design Innovation." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–11. Montreal QC Canada: ACM. <https://doi.org/10.1145/3173574.3173704>.
- Yang, Qian, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. "Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. Honolulu HI USA: ACM. <https://doi.org/10.1145/3313831.3376301>.