I don’t understand you – Error handling as a key aspect for conversational design for chatbots for public services

Stephan Raimer [1], *Marleen Vanhauer* [2]

[1,2] *University of Applied Sciences for Administration and Service*

Rehmkamp 10, 24161 Altenholz, Germany

ABSTRACT

Chatbots have been increasingly adopted in the public sector domain to support or provide public services. As a new kind of interface they provide a natural language access that is available 24/7. However, the application of user-centred development models for chatbots in order to achieve a good usability and user experience is still little used.

This paper presents a case study for a chatbots project together with back-ground and requirements and uses this example to research recommendations on the process model, methods and evaluation techniques for user-centred design. Our research focus is how chatbots and conversational design support public services.

Keywords: Human Factors, Chatbots, AI, Service Design, Public Service

INTRODUCTION

In over 50 years (the very first text chatbots were Eliza at MIT, 1966 and Parry at Stanford, 1972) , the evolution of chatbots has progressed to the point where different variants of voice-assisted assistants shape our everyday lives. If we follow the lines of development, we can say that chatbots have existed for decades, but have not been widely adopted. However, a revolution only occurred with the launch of Apple's Siri on the iPhone 4s (Siri was introduced by Apple on 4 October 2011 in a product presentation of the iPhone 4s). Whereas previously the function of a communication interface was fulfilled by chatbots, a variety of user commands could now be realised directly through the virtual voice assistant Siri - including performing phone actions, scheduling events, checking basic information, reminders, handling device settings, navigating areas, searching the Internet as well as engaging with iOS-integrated applications.

The revolutionary maturation of artificial intelligence (driven i.e. by IBM Watson winning Jeopardy 2011, among others) in terms of natural language processing did the rest. Amazon introduced Alexa (Initial release 2013 as Echo smart speaker) where Google presented Dialog Flow (2017, became later part of Google Cloud Platform) as a natural language understanding platform for conversational user interfaces for web apps, devices, interactive voice response systems and the like. The essential components for conversation design were thus available and are now well established: natural language understanding as well as dialog management.

Today, a large number of leading companies use chatbots to increase their efficiency. Communication channels can be offered 24/7 and can be directly connected through frameworks - from Facebook to Twitter, Skype, Cortana, Alexa or Slack and many more (Chatbot statistics 2021 – landbot.io).

As an example of use in the public sector, in 2019 the first pilot applications were implemented in Germany's northernmost federal state, Schleswig-Holstein (chatbot Ina for the Integration Office of the Ministry of Social Affairs, Integrationsamt 2021). In addition, chatbots and digital voice assistants were defined as important fields of application in the federal state's digitalisation programme (The Prime Minster of Schleswig-Holstein 2021) as well as the AI strategy. For further expansion, it was decided to set up a state infrastructure for chatbots in Schleswig-Holstein (based on the popular RASA open source framework for chat and voice-based virtual assistants, see https://rasa.com/). So if we assume that chatbots will find wider use in businesses and the public sector, it becomes clear that good conversational design will play an essential role.

The aim of this paper is to contribute to the user centered design of chatbots with a special focus on conversational design and error handling. The concrete motivation for the authors is to analyse the preliminary steps of an ongoing project and to plan and, if necessary, optimise its further implementation. The methodological approach

is combining Design Science Research (Hevner et. al. 2004, Peffers et. al. 2007) and Information System Development (Nunamaker et. al. 1990). The structure is as follows: first we present our case study background and requirements. Next, we discuss process models and recommendations of practice in literature, especially from the context of Human Centred Design, Conversational Design and Error Handling. We close with a conclusion and an outlook on the next steps in our project and give the direction of our further research approach.

A CHATBOT AS CONTINUING EDUCATION ADVISOR FOR PUBLIC ADMINISTRATION EMPLOYEES

The background for this paper is provided by a current research and development project of the authors (Digital Competence Navigator). As part of the development of a continuing education portal on digital competencies for administrative employees, three access points to courses will be implemented: A detailed search function for courses, individual matchmaking after an anonymous competence test as well as a self-assessment through a dialogue option with a chatbot. Herein, the chatbot is intended to be the central interface for presenting the portal and its functions and explaining all questions about it. In addition, structured chatbot dialogues should lead users to individual course recommendations.

Compared to existing chatbot projects within the context of administration in Schleswig-Holstein, it should be noted that a large and diverse target group (all employees of the public administration at federal state and municipal level) is addressed whereas the thematic scope is also quite large.

Generalised, it can be said that at the start of each chatbot project an analysis of the following aspects should be considered: stakeholders and value proposition, current solutions and channels, context and conversational tasks, personality of the chatbot, background tasks, error handling and fallbacks (Deibel and Evanhoe 2021, Hathaway and Hathaway 2021).

In the field of public services, chatbot systems are becoming more and more common in public administration contexts. Compared to business use cases, where conversion rates count as the ultimate KPI, here the benefits are more towards 24/7 support services, availability and scalability (Chatbot statistics 2021 – landbot.io).

From conversational design perspective, users should be able to enjoy talking to the virtual assistant. Hence, the dialogue design must be adapted to a (defined) target group depending on the context and area of application. The chatbot is a (virtual) representative of a company or administrative body, and therefore represents them in conversations with users or customers. Consequently, the language of the chatbot

should be adapted to the corporate identity, values and language (Chatbot statistics 2021 – landbot.io, Deibel and Evanhoe 2021, Hathaway and Hathaway 2021).

Challenges and opportunities of chatbot design, especially looking at the user experience, give you on the one hand a great potential for engaging with users at scale. On the other hand, there are (potential) usability problems, i.e. users need to read a lot and can't just easily scan and skip to their relevant content.

With regard to chatbot types, the presented example relies on the aforementioned federal state's infrastructure and allows for AI or NLP based chatbots (as opposed to simpler, rule-based chatbots). The underlying technology of machine learning allows the chatbot to learn from user requests, in order to become more and more knowledgeable and intelligent. It builds upon the open-source conversational AI stack RASA and integrates the chatbot's content management-platform botario (see https://botario.com/en/), in order to provide automatic processing of requests.

USER-CENTRICITY, CONVERSATIONAL DESIGN AND ERROR HANDLING

After an initial literature review, the question can be asked, whether a discipline like Conversational UX Design is already established as a distinct discipline . However, it can be said that Conversational Design falls under the umbrella of UX Design and shares common processes, principles and techniques. Accordingly, a number of principles and guidelines for the development of chatbots can be considered (Hoehn 2019, Moore and Arar 2019, Shevat 2017, Sugisaki and Bleiker 2020).

The basic understanding of the general UX Design Process (for systems) also applies here. The fundamental steps (cp. ISO 9241-210:2019) within an iteration cycle are "understanding & specifying context of use", "specifying user requirements", "production of design solutions" and "evaluating design solutions against requirements". Well-established evaluation methods for the last step are thereby Usability Testing and Heuristic Evaluation.

Ren et. al. (2019) aimed to identify the state-of-the-art in chatbots usability and the applied Human–Computer Interaction (HCI) techniques. They analysed the state-of-the-art of usability in the development of chatbots, and collected usability methods for evaluating overall chatbot usability. According to this, the most frequently applied research methods were quantitative surveys (and qualitative interviews), with a specific focus on the following parameters: "[…] among efficiency, effectiveness and satisfaction, satisfaction is the one that has been evaluated more frequently. User satisfaction is directly impacted by chatbot usability and hence it is considered as the primary measure in most studies." (Ren et. al. 2019).

On the other hand, within an administrative project in Schleswig-Holstein at the beginning of the year for a municipal vehicle registration office, the "correctness" or

"expected conformity of answers" from a chatbot was used as a measure of success (which could be increased from 69% to almost 90% in about 4 weeks – Petersen 2021).

Höhn and Bongard-Blanchy (2021) proposed "12 Heuristics for Conversational UX Analysis" based on Nielsen's ten heuristics for User Interface Design (Nielsen 2005) and elaborated that "this evaluation framework is applicable to chatbots of different conversation styles". Table 1 gives an overview of Heuristics and assigned Sub-heuristic.

Table 1: 12 Heuristics for Conversational UX Analysis [17]

(Folds et al. 2008)

Heuristic	**Sub-heuristic**
1. Visibility of system status	(a) Presence of information about the chatbot's state in the entire process (b) Immediate feedback (did the last user action work?) (c) Compel user action (what does the chatbot think the user will do next?)
2. Match between system and the real world	(a) Chatbot uses the language familiar to the target users (b) Visual components (emojis, GIFs, icons) are linked to real-world objects (c) If metaphors are used, they are understandable for the user
3. User control and freedom	(a) Chatbot supports undo/redo of actions (b) Chatbot offers a permanent menu (c) Chatbot provides navigation options (d) Chatbot understands repair initiations
4. Consistency and standards	(a) Chatbot uses the domain model from the user perspective (b) Chatbot has a personality, consistency in language and style
5. Error prevention	(a) Chatbot prevents unconscious slips by meaningful constraints (b) Chatbot prevents unconscious slips by spelling error detection (c) Chatbot requests confirmation before actions with significant implications (d) Chatbot explains consequences of the user actions
6. Recognition rather than recall	(a) Chatbot makes the options clear through descriptive visual elements and explicit instructions (b) Chatbot shows summary of the collected information before transactions (c) Chatbot offers a permanent menu and help option
7. Flexibility and efficiency of use	(a) Chatbot understands not only special instructions but also synonyms (b) Chatbot can deal with different formulations (c) Chatbot offers multiple ways to achieve the same goal

8. Aesthetic and minimalist design	(a) Chatbot dialogues are concise, only contain relevant information (b) Chatbot uses visual information in a personality-consistent manner to support the user, not just random decoration
9. Help users recognise, diagnose, and recover from errors	(a) Chatbot clearly indicates that an error has occurred (b) Chatbot uses plain language to explain the error (c) Chatbot explains the actions needed for recovery (d) Chatbot offers shortcuts to fix errors quickly
10. Help and documentation	(a) Chatbot provides a clear description of its capabilities (b) Chatbot offers keyword search (c) Chatbot focuses its help on the user task (d) Chatbot explains concrete steps to be carried out for a task
11. Context understanding	(a) Chatbot understands the context within one turn (b) Chatbot understands the context within a small number of turns (usually 2-3 user-bot turn pairs) (c) Chatbot understands the context of a multi-turn conversation
12. Interaction management capabilities	(a) Chatbot understands conversation openings and closings (e.g., 'hello') (b) Chatbot understands sequence closings (e.g., 'ok' and 'thank you') (c) Chatbot understands repair initiations and replies with repairs (d) Chatbot initiates repair to handle potential user errors

We consider the aspect of error handling to be particularly relevant and want to give it special consideration in our project (Heuristics 5, 9 and 11 of Table 1 relate to this). Generally considered, misunderstandings can occur in any dialogue. This applies to communication between people just as it does to communication between humans and chatbots. Resolving these misunderstandings and continuing the dialogue effectively is called error handling (Nielsen 2001, ITVSH 2021). We assume that good error handling makes conversations with chatbots more helpful, convincing and successful and that it has great impact on the overall user experience (Overview of chatbot challenges, Chatbot statistics 2021 – landbot.io; Moore and Arar 2019, Shevat 2017, Sugisaki and Bleiker 2020).

CONCLUSIONS AND OUTLOOK

The literature research has yielded several essential results for us and our chatbots project. We can clearly confirm that prototyping conversational interfaces is different or “stranger” compared to GUI-based interactions and more iteration cycles are recommendable in development (Deibel and Evanhoe 2021).

We have also found that we need to optimise our project plan in terms of chatbot training and associated iteration cycles (we will extend this project phase by approx. 6 weeks, as found in Petersen 2021). We will try to involve as large a number of test users (at least 100+) as possible in order to provide for a good training base.

Finally, we will apply the Heuristics for Conversational UX Analysis to our prototypes and first development increments. We assume that we have found a good working basis here, which also represents comparative values for a different application context.

Since we have not yet found much standardisation for the UX design of chatbots, we also want to incorporate our approach and templates for this project and future research into the digital design system (ITVSH 2021) of the municipal IT Association of the federal state of Schleswig-Holstein.

REFERENCES

Chatbot Statistics 2021 (March 2021): State of the Market & Opportunities, Website: https://landbot.io/blog/chatbot-statistics-compilation

Deibel, D., Evanhoe, R.: Conversations with things – UX Design for Chat and Voice, Rosenfeld, New York 2021

Google Assistant (November 18, 2021) Conversation Design – Errors - Google Developer, Website: https://developers.google.com/assistant/conversation-design/errors

Hathaway, T., Hathaway, A.: Chatting with Humans: User Experience Design (UX) for Chatbots, BA-Experts, 2021

Hevner, A.R.; March, S.T.; Park, J.; Ram, S. (2004): Design Science in Information Sys-tems Research. In: MIS Quarterly (Vol. 28 No. 1), S. 75–105.

Hoehn, S.: Artificial Companion for Second Language Conversation. Springer (2019)

Hoehn, S., Bongard-Blanchy, K. (2021): Heuristic Evaluation of COVID-19 Chatbots, In: Følstad A. et al. (eds) Chatbot Research and Design. CONVERSATIONS 2020. Lecture Notes in Computer Science, vol 12604. Springer, Cham. https://doi.org/10.1007/978-3-030-68288-0_9

IBM Conversational UX Design CHI 2017 Workshop, Website: https://researcher.watson.ibm.com/researcher/view_group.php?id=7539

Integrationsamt (November 18, 2021) Chatbot Ina - new digital assistance of the Integration Office (Chatbot Ina - neue digitale Assistenz des Integrationsamtes), Website: https://www.schleswig-holstein.de/DE/Landesregierung/IAMT/Projekte/teilhabe_Chatbot_Fachinhalt.html

ISO 9241-210:2019, Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems, Website: https://www.iso.org/standard/77520.html

ITVSH (November 2021) Design System | ITVSH - Design System, Website: https://www.design-system.sh/

Moore, R.J., Arar, R.: Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework. ACM Books (2019)

Nielsen, J. (2001): Error Message Guidelines, Website: https://www.nngroup.com/articles/error-message-guidelines/

Nielsen, J. (2005): Ten usability heuristics, Website: http://www.nngroup.com/articles/ten-usability-heuristics/

Nunamaker, J.F.; Chen, M.; Purdin, T.D.M. (1990): Systems Development in Information Systems Research. In: Journal of Management Information Systems 7 (3), S. 89–106.

Peffers , K., Tuunanen, T. , Rothenberger, M.A. & Chatterjee, S. (2007): A Design Science Research Methodology for Information Systems Research, Journal of Management Information Systems, 24:3, 45-77, DOI: 10.2753/MIS0742-1222240302

Petersen, L., Höhn, T., A. Höhn (2021): Chatbots: Digital assistants for citizen-centred administration (Chatbots: Digitale Assistenten der bürgernahen Verwaltung) In: Die Gemeinde – Zeitschrift für die kommunale Selbstverwaltung Schleswig-Holstein, 04/2021, pp 96-100, ISSN 0340-3653

Ren, R., Castro, J.W., Acuna, S.T. and de Lara, J. (2019): Evaluation Techniques for Chatbot Usability: A Systematic Mapping Study, International Journal of Software Engineering and Knowledge Engineering, Vol. 29, Nos. 11&12 (2019) 1673–1702, DOI: 10.1142/S0218194019400163

Shevat, A. (2017): Designing Bots: Creating Conversational Experiences. O'Reilly

Sugisaki, K., Bleiker, A. (2020): Usability guidelines and evaluation criteria for conversational user interfaces: a heuristic and linguistic approach. In: Proceedings of the Conference on Mensch und Computer. pp. 309–319

The Prime Minister of the State of Schleswig-Holstein (Der Ministerpräsident des Landes Schleswig-Holstein), Digitalisierungsprogramm 2021/2022 (March 2021) Website: https://www.schleswig-holstein.de/DE/Fachinhalte/D/Digitalisierung/ExterneLinks/digitalisierungsprogramm2122.pdf?__blob=publicationFile&v=1