Paper accepted for presentation at CONVERSATIONS 2025 – the International Symposium on Chatbots and Human-Centred AI, hosted at TU Lübeck, Germany, November 12-13, 2025.

Across the Domains: Investigating User Interface and Language Features

Johanna Rockstroh[0009−0008−2320−7289], Laura Spillner[0000−0001−8490−8961], Nima Zargham[0000−0003−4116−0601], Rainer Malaka[0000−0001−6463−4828], and Nina Wenig[0009−0007−5660−5446]

Digital Media Lab, University of Bremen, Bremen
https://www.uni-bremen.de/dmlab/
{rockstro,l.spillner,zargham,malaka,nwenig}@uni-bremen.de

Abstract. Current chatbot approaches are often designed for an open domain, making them suitable for many different purposes, but also raising challenges when users have different needs depending on the task. Based on Large Language Models (LLMs), users increasingly expect human-like behavior when interacting with chatbots. In our project, we are tackling the differences in users' needs and challenges, depending on the topic they are talking about and the domain the chatbot is dedicated for. We are investigating how a chatbot's user interface and domain are intertwined, since both play a crucial role in helping users identify problematic information and evaluate the quality and reliability of data sources. A good user interface can help users feel supported, open themselves up more, and enable them to verify information. We focus on both visual components (e.g. buttons, avatars, indicators) and language features (e.g. tone of voice, register and conversational cues like repair). Our approach is prototype-driven, and we are collaborating with partners from education, journalism, and health to test various visual user interface elements and language features across these domains.

Keywords: Chatbots · User Interface · Communicative AI · Interfaces
.

1 Introduction

Since the introduction of ELIZA, a chatbot in the domain of psychotherapy [16], many chatbots have been introduced in various domains [7]. The emergence of ChatGPT introduced the open-domain usage of chatbots to the general public, as it allows users to discuss multiple topics without pre-defined tasks or limited domain knowledge. However, domain chatbots have many advantages; they can be tailored to a particular user group with specific *visual user interface components* (e.g., avatars or action buttons) and *language features* (linguistic and paralinguistic cues, such as register, tone, and emojis [14]). Against this background, the differences between users interacting with open-domain chatbots, which have general knowledge, and users communicating with "closed-domain"

chatbots, which focus on a specific topic, are of particular interest. It remains unclear how the user interface and domain context might influence the communication with and perception of these chatbots. While the use of both task-specific and open-domain chatbots is growing, a systematic comparison of features between different domains is still missing [6].

For the perception of its users, the user interface design of chatbots is fundamental, as it plays a vital role in attributing meaning and understanding [5]. Users always interact with chatbots through a user interface, which guides the conversation and frames the user experience. It influences the user's trust in the system, whether they want to interact with it and in what manner [4, 13]. Although there are many studies that evaluate the user interface in a specific domain, e.g., journalism [18], (mental) health [8, 10, 12] or education [11], a general comparison across disciplines is lacking [2, 6]. Features that enhance usability in one domain may hinder performance in others. It is therefore necessary to study the interaction of visual user interface features and language features across different domains, use cases, and user groups.

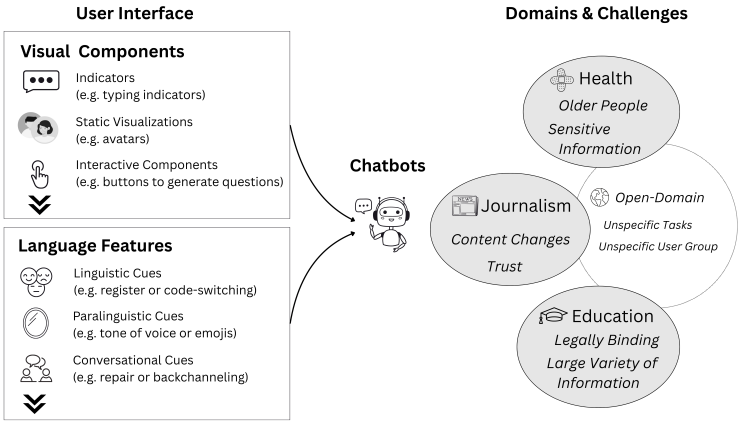

Fig. 1. The idea of our project is to systematically investigate visual user interface features and language features across different domains, focusing on health, education, and journalism in contrast to open-domain chatbots, where the user and task range are very broad.

The visual user interface does not only consist of menus and buttons, it also includes other visual components, such as the avatar of the chatbot [9] or the font [1]. Moreover, communication is not only about the content but also how it is conveyed through language features, such as domain-specific register, tone of voice, and conversational cues, such as repair [3] (see Figure 1). We argue that the interplay between domain, visual user interface and language features is often overlooked and insufficiently addressed in research on chatbots.

Følstad et al. [6] proposed an interdisciplinary research agenda for chatbot research in 2021. They note that although a large amount of studies on chatbots in different domains exists, and many new user groups from new application domains are likely to become of interest in the future, it is unclear to what extent research insights can be transferred across different domains. Since then, the number of chatbot users and application areas has increased immensely due to technological advancements. Therefore, we propose that it is prudent to focus specifically on the problem of transferability of research insights between domains, and the differences and interactions between them, through a systematic, replicable comparison between domains selected for their characteristics.

Building on this perspective, we outline a research agenda designed to systematically explore how insights about chatbot design and user experience can be transferred across domains. Specifically, we aim to address the following research questions:

> **RQ1** How do different domain topics and user interface features (visual components and language features) interact to influence the user experience in chatbot-based user studies?
> **RQ2** To what extent do variations in domain topic and user interface design (visual components and language features) affect user engagement, trust, and task outcomes with chatbots?

We plan to answer these questions through an agenda that encompasses three main steps:

1. State-of-the-Art in Research and Industry
   a) Analyze visual components in the user interfaces of existing chatbots across different domains. Annotate these components following an existing classification framework [15] to identify trends, and possibly extend classification for changes introduced by LLM-based bots.
   b) As investigating language features in proprietary chatbots is not easily doable, we survey the relevant literature to identify which language features have been implemented and evaluated in chatbots. We develop a framework to categorize these analogously to visual components.
2. Modular Implementation
   a) Implement first chatbot prototypes for different domains and test individual features of interest in those domains. Identify which features should be investigated systematically across domains.
   b) Based on the prototypes, build a modular chatbot system in which both visual components and language features can be swapped out easily. Domain knowledge is transferred through a RAG-based backend.
3. Cross-domain User Studies and Evaluation
   Conduct a series of systematic studies employing the modular chatbot system, to cross-compare the effects of different user interface features (visual components and language features) across domains.

Figure 1 shows an overview of our research agenda. We are already in the process of investigating (Step 1) existing bots and literature and are currently annotating them. Moreover, we are implementing and testing first chatbot prototypes (Step 2a) for different language features (e.g., hedging, figurative language, language alignment) and visual features (e.g., a 3D-avatar). To this end, we are collaborating with stakeholders from different domains to collect real-world data, tasks, and participants, starting with journalism, education, and health. Based on this we will start with the implementation of the modular chatbot system (Step 2b) in the next step.

2 Overview of User Interface Features

We investigate the posed research questions by focusing on three domains: journalism, health, and education. The chosen domains offer a wide variability of users (age and technical background), topics (e.g., website content in education, sensitive communication in health), and contexts (work-related journalism vs. private health information). This will allow us to create domain-specific chatbots and study how domain topics and user interface features influence user interaction and relate to one another. The tasks and domains are diverse, as are the users and their requirements. We expect, for example, that older users discussing sensitive topics have different expectations than young students focused on task completion. We are interested in how different users might perceive and interact with the same user interface and which features they consider important.

In a first step, we are currently working to systematize the possible influence of language features and visual user interface features on user interaction, based on existing research and implementations. As a starting point, we follow the classification of visual user interface features proposed by Traubinger et al. [15] to annotate and categorize visual components found in chatbots on the market across the following domains: open-domain, education, health and social companions. At the moment, we are further consolidating insights from prior studies on the effects of different language features, including linguistic and paralinguistic features, on user interactions. Based on these insights, we are building a theoretical framework around user interface designs and language features. Such a domain-specific embedding can inform future research and user interface design, such as for the domain prototypes we are developing.

During the annotation of visual features and the chatbot's linguistic capabilities, the distinction between open-domain and domain-specific systems emerged as a central yet ambiguous aspect of the analysis. Firstly, we have to consider how domain-specific systems can be related to open-domain ones and where boundaries or overlaps in feature sets may occur. Secondly, it is important to ask what happens when users engage with open-domain bots to address specific, use-case-oriented problems. Finally, these reflections point toward practical implications for system design and adaptation, paving the way for the following discussion on implementation strategies.

3 Implementation and Evaluation

Developing conversational bots requires integrating Large Language Models with interface and interaction design. In Step 2, we are implementing a modular chatbot system that uses Retrieval-Augmented Generation (RAG) [17] in the backend to enable us to utilize it across application domains. In a RAG-based chatbot, domain knowledge is chunked and vectorized, allowing relevant chunks to be selected in response to user queries. Our research focuses on features layered over this foundation, which strongly shape the user's perception of the interaction: appearance, interaction parameters, and language features.

We are constructing the visual user interface on a modular basis, allowing features to be tested individually or in combination. Parallel to this, we are building a modular framework in the backend, which allows us to integrate various language features, such as language alignment, backchanneling, or a specific conversational style, into the RAG-based foundation. We are collaborating with researchers from the three target domains to collect real-world data, documents from which to retrieve chatbot answers, existing user preferences and usage strategies, and possible user queries for testing prototypes, and to gain insight into specific user groups and stakeholders to better understand their needs.

To systematically compare the design and the language features, we will conduct field studies and experimental online studies, as well as lab studies and participatory workshops. We are planning to evaluate three kinds of data: a) quantitative measures of users' perception of the bots and the effectiveness of their collaboration, including usability and engagement metrics and measures of trust, performance, and task success; b) conversation analysis e.g. in terms of how the language style of the users has changed in response to that of the chatbot; and c) qualitative interviews to explore how users perceive and interact with the user interfaces in more detail: Do they perceive that the user interface adapts to their needs? Does the user interface support understanding of the task or domain? Recruitment for domain-specific testing will be done in collaboration with the domain experts.

Some questions concerning implementation and systematic evaluation are still under investigation: Firstly, we have to understand which technical approaches are feasible to vary language features in a controlled manner, e.g., through prompting, modification of output texts through a second LLM, or rule-based modification. We are currently running studies with first prototypes to test these. Secondly, we have to carefully consider how we can compare different tasks across domains, without oversimplifying domain-specific nuances. To this end, the ongoing discussions with domain experts are especially important. Because systematically comparing features on what is otherwise the same interface across different tasks and users is quite novel, we expect it to provide a variety of insights into the chosen domains. Most importantly, we aim to gain a better understanding of how the effects of the user interface design, language features, and the application domain interact with each other.

Acknowledgments. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article is based on research conducted in the DFG funded Research Unit 5656 'Communicative AI', project 'Interfaces: Implementing user-centered ComAI' (funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

References

1. Candello, H., Pinhanez, C., Figueiredo, F.: Typefaces and the perception of humanness in natural language chatbots. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3476–3487. Association for Computing Machinery, New York, NY, USA (2017), https://doi.org/10.1145/3025453.3025919
2. Chaves, A.P., Gerosa, M.A.: How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. International Journal of Human–Computer Interaction **37**(8), 729–758 (May 2021). https://doi.org/10.1080/10447318.2020.1841438
3. Dingemanse, M., Enfield, N.J.: Interactive repair and the foundations of language. Trends in Cognitive Sciences **28**(1), 30–42 (2024)
4. Dubiel, M., Daronnat, S., Leiva, L.A.: Conversational agents trust calibration: A user-centred perspective to design. In: Proceedings of the 4th Conference on Conversational User Interfaces. CUI '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3543829.3544518, https://doi.org/10.1145/3543829.3544518
5. Feine, J., Gnewuch, U., Morana, S., Maedche, A.: A Taxonomy of Social Cues for Conversational Agents. International Journal of Human-Computer Studies **132**, 138–161 (2019). https://doi.org/10.1016/j.ijhcs.2019.07.009
6. Følstad, A., Araujo, T., Law, E.L.C., Brandtzaeg, P.B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., Luger, E.: Future directions for chatbot research : An interdisciplinary research agenda. Computing : archives for informatics and numerical computation **103**(12) (2021). https://doi.org/10.1007/s00607-021-01016-7
7. Følstad, A., Skjuve, M., Brandtzaeg, P.B.: Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design, p. 145–156. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-17705-8_13, http://dx.doi.org/10.1007/978-3-030-17705-8_13
8. Harrington, C.N., Egede, L.: Trust, Comfort and Relatability: Understanding Black Older Adults' Perceptions of Chatbot Design for Health Information Seeking. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–18. CHI '23, Association for Computing Machinery, New York, NY, USA (Apr 2023). https://doi.org/10.1145/3544548.3580719
9. Harrington, C.N., Egede, L.: Trust, Comfort and Relatability: Understanding Black Older Adults' Perceptions of Chatbot Design for Health Information Seeking. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–18. CHI '23, Association for Computing Machinery, New York, NY, USA (Apr 2023). https://doi.org/10.1145/3544548.3580719, https://dl.acm.org/doi/10.1145/3544548.3580719
10. Lee, Y.C., Cui, Y., Jamieson, J., Fu, W., Yamashita, N.: Exploring Effects of Chatbot-based Social Contact on Reducing Mental Illness Stigma. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–16.

CHI '23, Association for Computing Machinery, New York, NY, USA (Apr 2023). https://doi.org/10.1145/3544548.3581384
11. Ramandanis, D., Xinogalos, S.: Designing a Chatbot for Contemporary Education: A Systematic Literature Review. Information **14**(9), 503 (Sep 2023). https://doi.org/10.3390/info14090503
12. Sah, Y.J., Peng, W.: Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. Computers in Human Behavior **45**, 392–401 (2015). https://doi.org/10.1016/j.chb.2014.12.055
13. Schmitt, A., Wambsganss, T., Leimeister, J.M.: Conversational agents for information retrieval in the education domain: A user-centered design investigation. Proc. ACM Hum.-Comput. Interact. **6**(CSCW2) (nov 2022). https://doi.org/10.1145/3555587
14. Spillner, L., Rockstroh, J., Peña, P.R., Wenig, N., Zargham, N., Cowan, B.R.: Beyond the Illusion: LLMs and the Case for Pragmatic Cues in Conversation. In: Proceedings of the 7th ACM Conference on Conversational User Interfaces. pp. 1–6. ACM, Waterloo ON Canada (Jul 2025). https://doi.org/10.1145/3719160.3737614, https://dl.acm.org/doi/10.1145/3719160.3737614
15. Traubinger, V., Gaedke, M.: Interaction Design Patterns of Web Chatbots, p. 309–317. Springer Nature Switzerland, Cham (Switzerland) (2024). https://doi.org/10.1007/978-3-031-62362-2_22
16. Weizenbaum, J.: Eliza—a computer program for the study of natural language communication between man and machine. Commun. ACM **9**(1), 36–45 (Jan 1966). https://doi.org/10.1145/365153.365168, https://doi.org/10.1145/365153.365168
17. Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J., Cui, B.: Retrieval-augmented generation for ai-generated content: A survey (2024), https://arxiv.org/abs/2402.19473
18. Zhao, X.: Towards Personalized User Interface Design For News Chatbots: A Pilot Study. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–6. CHI EA '23, Association for Computing Machinery, New York, NY, USA (Apr 2023). https://doi.org/10.1145/3544549.3583951