Show, don't Tell. Reflections on the Design of Multi-modal Conversational Interfaces

Pietro Crovari[0000−0002−6436−4431], Sara Pidó[0000−0003−1425−1719], Franca Garzotto[0000−0003−4905−7166], and Stefano Ceri[0000−0003−0671−2415]

Politecnico di Milano,
Department of Electronics, Information and Bioengineering,
via Ponzio 34/5, 20133 Milan, Italy
{name.surname}@polimi.it

Abstract. Conversational Agents are the future of Human-Computer Interaction. Technological advancements in Artificial Intelligence and Natural Language Processing allow the development of Conversational Agents that support increasingly complex tasks. When the complexity increases, the conversation alone is no more sufficient to support the interaction effectively, but other modalities must be integrated to relieve the cognitive burden for the final user. To this aim, we define and discuss a set of design principles to create effective multi-modal Conversational Agents. We start from the best practices in literature for multi-modal interaction and uni-modal Conversational Interfaces to see how they apply in our context. Then, we validate our results with an empirical evaluation. Our work sheds light on a largely unexplored field and inspires the future design of such interfaces.

Keywords: Conversational Agent · Chatbot · Multi-modal Interaction · Design Principles · Interaction Design.

1 Motivations and Context

A chatbot is a user interface that communicates with the human being through the mean of Natural Language [6]. From the user perspective, chatbots are perceived as intuitive and efficient, since they can remove the friction of the interaction with the Graphical User Interface (GUI), and let the users focus on the task, rather than on the way they have to translate their intention into actions on the interface [35].

For this reason, chatbots are becoming ubiquitous in society. According to Radziwill and Benton, in the last decade more than one-third of online conversations involved a chatbot [31]. This trend is continuously growing; the authors in [14] predicted that soon people will prefer to interact with a chatbot to accomplish their tasks instead of using a "traditional" web application.

In recent years, the power of this technology has been combined with the latest technological advancements in subjects such as machine learning and deep

learning to develop chatbots for tasks with increasing complexity. Education, data science, data retrieval, and visualization are examples of application domains in which chatbots were successfully implemented to support the user [19, 13, 30, 24]. When the task's complexity increases, both in terms of quantity of information treated and the number of operations to concatenate to accomplish the task, the conversation alone is no more sufficient in supporting the user. When the information that must be shown to the user starts to be consistent and heterogeneous, empirical evidence shows that the conversation is no longer sufficient for most users [18]. In the same way, when the design of the conversation is not linear but is constituted by several possible branches, the users must be given a hint of what they can do.

As a consequence, the urge for the introduction of new modalities arises. When dealing with written conversational agents (i.e., chatbots), the most natural integration is the visual modality through the addition of visual content aside from the natural language interface. In this way, the conversation is supported by a whole new channel that can be exploited to support the users and provide visual feedback. Even if multi-modal conversational interfaces are increasingly adopted, to the best of our knowledge, very little work has been carried out to understand how to design these interfaces optimally.

In this context, our research takes place. Starting from the design principles present in literature to create optimal conversational interfaces, we want to see how they adapt or must be modified in a multi-modal setting, in particular where the conversation co-exists with visual interaction. We ran a literature review to understand how the problem of integrating a conversation with other modalities was faced. The main contribution of this work is a set of design principles resulting from the performed literature review applicable to multi-modal conversational interfaces, particularly to the ones where the conversation is integrated with a GUI. Then, we provide a concrete example of how the principles can be used to design such an interface. Finally, the interface is preliminarily evaluated to assess the result's quality and gather precious insights into the design process.

These principles have an "heuristic" nature and have been elicited on the basis of both a literature review and by distilling the authors' experience on the design, development, and evaluation of several conversational applications [2–4, 9–11, 33, 36, 37]. Different authors have proposed or used different guidelines for chatbot design, but - to the best of our knowledge – a catalogue (and a validation) of the most relevant ones is still missing in the current state of the art. Our principles can be regarded as design guidelines that complement other, more generic heuristics proposed in HCI (e.g., Nielsen's 10 heuristics for inspection-based usability evaluation [26]) since they address chatbot-specific design principles and can be helpful for two main reasons: during the design stage, to use them as a checklist to enhance the usability of chatbot specific product features from early in development; during usability evaluation - at the prototyping or deployment stage, to support expert's inspection[1] of chatbot usability.

[1] Usability inspection is the generic name for a set of methods that are all based on having evaluators inspect a user interface [25], without involving user in the testing.

2 Design Principles

In this section, the design principles will be described carefully, to understand the underlying motivations and the consequences they imply.

To elicit these principles, we started from the best practices and the results in the literature for uni-modal conversational agents and multi-modal interfaces, to see if and how they apply for multi-modal conversational agents.

To accomplish our review, we proceeded as follows. We searched for relevant paper Google Scholar and Scopus engines, using the following query: *"(design principles OR guidelines) AND (conversational agent OR multi-modal interface)"*, filtering for paper published in the last 25 years (date¿1995). The resulting list was scanned to filter eligible papers according the following criterion: *from the title and/or the abstract it must be intended that the paper addresses the problem (also) from a design perspective.* 19 papers passed the selection process. To evict the principles, we read the documents integrally and we grouped them according to the design principles exploited in the described interfaces. This process originated seven recurrent themes that reflects the design principles for the design of multi-modal conversational interfaces reported in the paper.

Table 1. Design Principles for the design of multi-modal conversational agents

	Design Principles
P1	Show, don't tell.
P2	Separate feedback from support
P3	Show information only when necessary
P4	Design a light interface — emphasize content
P5	Show one modality at a time
P6	Don't overload multiple modalities beyond user preferences and capabilities
P7	Use multi-modality to resolve ambiguities

Show, Don't Tell. The availability of more communication channels is the most immediate consequence of introducing new modalities in the interaction. Thus, the information can be conveyed to the user in multiple ways.

When dealing with a uni-modal conversational interface, the agent must be designed to be self-explainable. The conversation must contain all the information necessary to continue the interaction, such as the results of the previous operations and some hints on the possible next actions the user can choose. When the choices are many, and the results complex, the conversation becomes verbose, going to increase the length of the messages, or even the number of the interaction required to select the desired operation, consequently reducing the usability of the chatbot [39].

To overcome this problem, we take inspiration from the well-known literature principle *Show, don't tell.* This idea has been formulated over a quote said by the Russian playwright Anton Chekhov, who said that in narration things should

not be described but shown through concrete examples [7]. In the same way, in a multi-modal Conversational Agent, information can be shown over multiple modalities, rather than being only textually described in the conversation itself. For instance, visual hints can orient the user through the conversation, giving a clear overview of the performed operations and removing the necessity of written summaries. Graphics can summarize the data retrieved through the conversations [41], and a table can summarize the choices with the previous utterances.

This technique brings a double advantage. In the design of the conversation, all the information reported through another modality can be omitted, creating shorter and more effective messages [18]. The number of messages can be reduced, reducing the cost of the conversations [39]. Second, the risk of loss of information in the conversation is minimized, since the meaningful one is conveyed exploiting the other modalities [41].

Separate Feedback from Support. A conversational agent typically provides two types of information to the user: feedback and support. The first comprises the results of the operations performed, whereas the latter illustrates what the user can or should do in the next interactions.

In a multi-modal interface, these kinds of information can be conveyed through different channels. For example, the results can be shown as graphs in a GUI, the completion of the operation can be represented through the change of the color of a button in the interface, the information on the operations the user can do can be embedded in the conversation or written in a dedicated pane.

According to the structure principle for GUIs introduced by Constantine and Lockwood [8], the users should have clear where to find the results they are seeking and where to look for support. Contrarily to the original principle, though, the division must be consistent not only between the modules of the interface but also between the different modalities.

Geranium [15] is an excellent example of a multi-modal conversational agent that exploits different channels for feedback and support. The application consists of an embodied multi-modal conversational agent for increasing the awareness of the urban ecosystem in children. The agent asks questions on the topic and comments on the answers. Children can choose the correct answer using a set of buttons that appears when the question is asked. When an answer is selected, the agent's avatar plays an animation that is happy or sad according to the given answer's correctness.

Show Information Only When Necessary. The presence of multiple channels to communicate with the users can cause a cognitive overload with a loss of usability, if not used properly [34].

To prevent this problem, the modalities should be complementary in their content, without being redundant in their information [28]. We need to think of the conversation as a part of the multi-modal interface, and not as a standalone channel. In this way, the information can be distributed over the various

channels, conveying the right information at the right moment and through the right channel. Otherwise, the repetitions created between the chatbot utterances and what is on the other channels create ambiguities in the interface, decreasing the usability of the system.

A good multi-modal chatbot design also deals with the removal of the information from the interface. When some data is no more necessary, it should be hidden to free space and lightening the cognitive burden.

This principle is widely adopted in Embodied Conversational Agents (ECA), where often the agents' utterances are transcribed in balloons that disappear when the interaction continues [5].

Design a Light Interface — Emphasize Content. Hearst and Tory [28] highlight how, when the user is engaged in the dialogue with a multi-modal conversational agent, the interface disappears to the user's eye since the only focus is on the provided data. A good design for such a chatbot is hence a design that minimizes the overall impact of the interface on the interaction. Only in this way users can fully concentrate on the focus of the conversation, which is the action they want to perform.

To satisfy this principle, the interfaces must be designed to have the main focus on the channel exploited to convey the information or the data. For example, if the chatbot is integrated into a dashboard for data visualization, considerable space has to be given to the graphs, instead of the conversation itself.

In the same study, the researchers noticed how the interface suddenly became the user's focus when it did not work correctly, or when the system gave unexpected (or undesirable) responses, as in the case of the interruption of the conversation flow. One example is when the dialogue reaches a dead end, leaving the task unaccomplished and the user unsatisfied [28]. This effect can be mitigated by a careful analysis of the dialogue tree, to ensure that each utterance can bring to a proper conclusion of the dialogue.

A good example is provided by Ava [20], a conversational agent that exploits this principle by presenting just two columns, one for the conversation and one for the generated Python notebook, where the interface almost disappears.

Show One Modality at a Time. Studies reveal that users, even if they like multi-modal interaction [28], in most occasions tend to interact with one mode at a time [27, 29].

Multi-modal interaction with a chatbot should follow the same principle. The user should be requested to use only a modality at a time. The final task can be multi-modal, but the multi-modality should originate in alternating different uni-modal actions, and not vice versa. For example, a conversational agent for education can be embedded in a visual interface where the tasks are described. After reading the assignment, students can dialogue with the chatbot to get to the solution, and then report the results in a separate dialogue box [38, 21].

Even if the channels are not exploited simultaneously, the information conveyed through the others will influence the conversation. In many cases, the

sentences will be simplified since complementary information will be exchanged through the other modalities. This consideration can support the design of the conversation; if a step is too critical or error-prone to be described with words, other modalities can be used instead.

Don't Overload Multiple Modalities Beyond User Preferences and Capabilities. If exploited properly, the multi-modality can facilitate the interaction for the user, but if the combination of channels does not result as natural and intuitive, it will only obstacle the accomplishment of the users' goals [12].

Thus, once the modalities have been established in the design phase of the conversational agent, it is fundamental to carefully decide the best channel over which the user can interact with the platform and the ones the system uses for providing the feedback. Additionally, similar interactions should involve similar modalities. For example, all the visualizations should be conveyed through the same pane, all the search results should be described in the conversation, and the possible action should be suggested through a dedicated list. This consistency will be appreciated by the user, that otherwise will remain unsatisfied from the interaction [16].

Every time the user or the chatbot sends a new message, this is added to the interface's conversation history. As a consequence, as the interaction continues, the amount of text in the dialogue grows, making the retrieval of information written in the messages always harder. For this reason, key information should be stored in other places than in the conversation to allow users to retrieve it at a glance.

AdApt [17] is an agent designed to support the retail sector, specifically the search for available apartments in Stockholm. Users can exploit two channels for the interaction: they can communicate vocally with the agent or interact with a map shown on the screen. Their Wizard-of-Oz study showed how users used different channels for different purposes, coherently with the system design.

Use Multi-modality to Resolve Ambiguities. Natural Language is ambiguous for its nature [32]. When the operations to perform become complex, these ambiguities can compromise the overall result of the interaction. New modalities can be introduced in the interface to eliminate this problem: when an ambiguity generates, the new modality can solve the ambiguity.

For example, in a music chatbot, the agent can make users listen to a short preview of the song to ensure the one the user is referring to [3]; in an e-commerce website the virtual assistant can show pictures of the product to understand the user's tastes and recommends items accordingly [23]; in end-user development, the conversation can ask to point out items on the screen to understand precisely what the user is talking about [22].

3 Exemplifying Our Principles

3.1 Case Study in the Bioinformatics Domain

We designed a multi-modal interface for an bioinformatics application in which the chatbot supports data retrieval and exploration. These are intrinsically complex tasks, for the complexity of the domain and because they require competence in search and analysis techniques, a common skill among computer scientists which often biologists and clinicians lack. This is a complex task since it requires a good understanding of Computer Science, skill that often biologists and clinicians lack [10]. For this reason, Bolchini et al. [1] highlighted the importance of a new family of tools that these users can use in autonomy. The proposed conversational agent is shown in Figure 1.

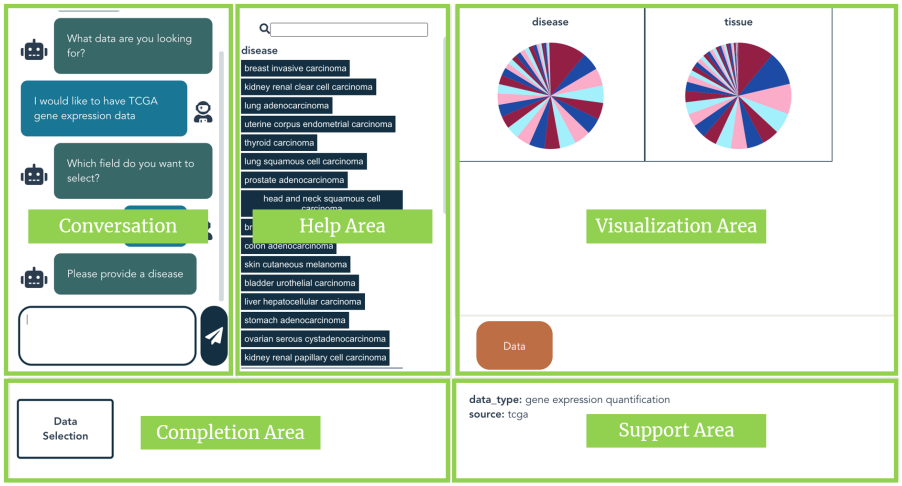

Fig. 1. The interface of the multi-modal chatbot we designed for the study. In the upper row the conversation takes place, next to the support and the visualization area. In the lower section, the completion and the support area helped the user during the interaction. Table 2 illustrates how the principles have been followed in the interface design.

The multi-modality is given by the conversational channel used as the mean tool for the communication and the visual channel used for user support, orientation, and feedback. The interfaces were designed carefully, following the described principles, as described in Table 2. The GUI was divided into five sections, divided into two rows, each one dedicated to a specific function. In the upper part, the conversation occupied the leftmost part, followed by the help and visualization area. In the second row, the completion and the support area took place.

The user could communicate with the system through the chatbot. At every step of the conversation, the help pane is populated with a set of information useful to support the user. For example, at the beginning of the conversation, the users are shown the possible operations, whereas when data are filtered all the possible values to filter are displayed. The user can click on the terms in this area to automatically copy them inside the chat area. When data are selected, descriptive summarizations are shown in the visualization area in the form of pie-charts. When the mouse pointer passes over the graphs, the name of the category and the count of its samples appear. The support area is populated with the query parameters selected by the user, such that the status of the query is comprehensible at a glance. Finally, when the operation has completed, the box in the completion area changes color to represent the end of the task. At the same time, a download button appears in the same box to export the retrieved data.

Table 2. Application of the principles in the define of the interface shown in Fig. 1

	Application of each principle in the design of Fig.1's interface
P1	Different visualizations are used as feedback and orientation in the process workflow.
P2	Help and visualizations are in different section of the GUI.
P3	Information in the GUI is dynamically changed according the state of the conversation
P4	Interface is designed around the essential elements – no superfluous information
P5	Relevant information is showed only through one modality at a time
P6	Actions and functionalities are defined for every modality
P7	Hints in Help Area support users in the setting of parameters

3.2 Exploratory Study

We performed an exploratory study to verify that the close adherence to our principles results into multi-modal conversational applications that users perceive as usable and effective. To this aim, we developed a multi-modal interface that integrates a conversational agent in a GUI and was designed according to our design principles; then we ran a small (n=16) empirical study, devoted to investigate the perceived aspects related both to usability (such as errors performed or task difficulty) and to conversation specific quality issues (such as understandability of the dialogue).

Subjects. We recruited 16 participants on a volunteer base (6 Female, 10 Male, avg. age: 28.61) through a mailing list of our research group's collaborators. These people have a heterogeneous background – mainly Computer Science and Engineering, Biomedical Engineering, and Biology – but with a shared research interest for computational genomics.

Procedure. Due to the current pandemic emergency, we ran an online survey. Participants had a link to the online platform, and one to the research survey. The survey was divided into sections. In the introductory one, participants were introduced to the procedure. They were then invited to use the platform to answer five tasks, described in Appendix A, and report the result in the questionnaire. Since no tutorial was given, users had to understand how to use the platform in autonomy. After completing the tasks, we invited participants to continue exploring the platform in autonomy as long as they believed to have discovered all the functionalities. In the last part, ten questions investigated the perception of the platform from the volunteers' perspective (Appendix B). Some of them asked the users to express their opinion in a grade from 1 (very low) to 5 (very high). In others, they could freely express their thoughts as text. All the conversations were anonymously logged in the backend.

Results. 13 participants were able to complete the study procedure. The other three had issues in the connection that could not support such a data-intensive process. Accounting all the tasks for all the participants, 85% of the tasks were accomplished. In general, users found the system easy to use (3.63/5) and intuitive (4.09/5), despite the difficulty of the tasks proposed. The modalities resulted well integrated into the system (4/5). All participants that completed the evaluation were able to find out all the ways to interact with the interface. From the analysis of the conversations, we see that the users preferred to communicate with the chatbot through keywords rather than with full sentences. Two participants did not even use a single sentence in the whole interaction, limiting themselves to a few nouns or adjectives per utterance. The most liked features of the system were the multi-modal interface (6/13), the ease-of-use (4/13), and the freedom of expression left to the user (2/13). The least liked ones were some bugs found in the conversation and in the Natural Language Understanding (4/13), and the fact that people felt a little constrained by the system's actual capabilities (3/13).

Discussion. The analysis of the above results allow us to comment on how adherence to design principles can enhance the power of chatbots. We are aware that the sample population does not fully reflect the target population of the final platform, as it includes several people with computer science training; prior work shows that people with higher levels of computer science training (or more advanced technical knowledge) results in them being more forgiving of failure [40]. However, modern biologists using bioinformatics analysis tools such as the ones accessed through our platform must also have a computational background, so we expect our results to be confirmed when we will be able to recruit a wider and more balanced set of evaluators.

Multi-modal channels of communication can provide new information that is hardly conveyed through the conversation alone [P1]. We provided the user with support information and visualizations, making the interaction with the conversation easier. Even if the subjects had background expertise in computer

science, we don't believe this fact affects our findings, as the interaction with the platform does not require or expect any strong computer science skill, but only basic ability in file retrieval, which is well known to most biologists.

The introduction of the new interactions paradigm did not burden the users, who find our platform intuitive and easy-to-use. In fact, it enhanced the estimation of the "intelligence" of the platform making users disappointed when the conversational agent did not wholly match their expectations in terms of computational capabilities [P4].

The division of the interface in functional areas has been particularly appreciated, since it gave the possibility of understanding at a glance what had already been done and which were the possible next steps [P2]. The help area revealed particularly useful at the beginning when users were not confident with the possibilities offered by the system and played a pivotal role in making participants discover all the functionalities of the system. The users appreciated a lot the dynamically changing content of this area, capable of providing the right information at the right time [P3]. Visualization area acted at the decision-making level, informing the users on the selected data and therefore letting them make the best choices on how to continue the data exploration process. Participants liked the interplay of the conversation with two Help and Visualization Areas since, at every step of the conversation, they were able to find most relevant information on the visual interface, while they relied on the conversation only as a guide throughout the process [P5]. In addition, the support area was appreciated in the short-term strategy, since it allowed users to understand whether the users' utterances were interpreted correctly and the desired operations were executed successfully. As expected, the side effect of introducing a new modality was to make the users' sentences shorter, thereby easing the task of the Natural Language Understanding unit in the backend, which had simpler utterances to parse [P7].

On the other hand, new interaction modalities imply greater attention in the design of the interface. To guarantee the consistency of the information on the various channels implies a careful analysis of each moment of the interaction. At design time, it is necessary to have the complete description of every state of the system, what is shown to the user, and probably even more importantly, what is removed from the interaction. In fact, in an initial prototype of the system, we noticed how much the content not removed at the right time from the interface could induce confusion in the user, even if experienced ones like the chatbot's designer themselves [P6].

4 Conclusions

Chatbots are more and more exploited to accomplish tasks that are increasingly complex, e.g., in terms of process and amount of data involved . Still, conversation alone might not be sufficient and would benefit from the integration of other interaction modes. The introduction of additional modalities facilitates users who need to be supported continuously through the interaction, and can

enrich structured assistance and feedback. Even if multi-modal conversational interfaces are increasingly adopted, in literature very little has been done to tackle the optimization of their design.

With this paper, we provided a set of guidelines on the design of effective multi-modal chatbots, which are summarized in Table 1. We started from multi-modal and conversational literature to elicit our principles and then verify them with a preliminary empirical evaluation.

We are aware that our work presents some limitations. First, our principles should not be seen as a guide, but be considered as a starting point on which the interface designer can reflect to produce the interface. Even if result of a comprehensive analysis, we tackled only the surface of this problem. Our work should be considered the starting point for a broader discussion that includes experts from different domains that can contribute to their point of view. Finally, even if promising, the exploratory study should be considered a preliminary step in evaluating the principles, given the small number of participants we were able to get involved in due to the pandemic emergency. For this reason, we will proceed with a complete usability evaluation with a wider sample including more biologists with limited technical skills.

Our contribution is a first attempt to shed light on a largely unexplored field.Within the bioinformatics domain, we will apply our principles to more complex tasks. We will then challenge our design principles by putting them at work in other domains, going beyond bioinformatics data retrieval. Finally we will continue our investigation on design principles, by broadening our approach and adding to the problem a multidisciplinary perspective, including in the discussion experts in related subjects such as cognitive sciences, linguistics, and psychology.

References

1. Bolchini, D., Finkelstein, A., Perrone, V., Nagl, S.: Better bioinformatics through usability analysis. Bioinformatics **25**(3), 406–412 (2009)
2. Catania, F., Di Nardo, N., Garzotto, F., Occhiuto, D.: Emoty: an emotionally sensitive conversational agent for people with neurodevelopmental disorders. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (2019)
3. Catania, F., Luca, G.D., Bombaci, N., Colombo, E., Crovari, P., Beccaluva, E., Garzotto, F.: Musical and conversational artificial intelligence. In: Proceedings of the 25th International Conference on Intelligent User Interfaces Companion. pp. 51–52 (2020)
4. Catania, F., Spitale, M., Fisicaro, D., Garzotto, F.: Cork: A conversational agent framework exploiting both rational and emotional intelligence. In: IUI Workshops (2019)
5. Cauell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H.: Designing embodied conversational agents. Embodied conversational agents **29** (2000)
6. Chandel, S., Yuying, Y., Yujie, G., Razaque, A., Yang, G.: Chatbot: efficient and utility-based platform. In: Science and Information Conference. pp. 109–122. Springer (2018)

7. Chekhov, A.: The Unknown Chekhov: Stories & Other Writings Hitherto Untranslated. Macmillan (1999)
8. Constantine, L.L., Lockwood, L.A.: Software for use: a practical guide to the models and methods of usage-centered design. Pearson Education (1999)
9. Crovari, P., Catania, F., Garzotto, F.: Crime story as a tool for scientific and technological outreach. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–10 (2020)
10. Crovari, P., Catania, F., Pinoli, P., Roytburg, P., Salzar, A., Garzotto, F., Ceri, S.: Ok, dna! a conversational interface to explore genomic data. In: Proceedings of the 2nd Conference on Conversational User Interfaces. pp. 1–3 (2020)
11. Cutrupi, C.M., Fadda, S., Valcarenghi, G., Cosentino, G., Catania, F., Spitale, M., Garzotto, F.: Smemo: a multi-modal interface promoting children's creation of personal conversational agents. In: Proceedings of the 2nd Conference on Conversational User Interfaces. pp. 1–3 (2020)
12. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: A survey of principles, models and frameworks. In: Human machine interaction, pp. 3–26. Springer (2009)
13. Fast, E., Chen, B., Mendelsohn, J., Bassen, J., Bernstein, M.S.: Iris: A conversational agent for complex tasks. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–12 (2018)
14. Følstad, A., Brandtzæg, P.B.: Chatbots and the new world of hci. interactions **24**(4), 38–42 (2017)
15. Griol, D., Callejas, Z.: An architecture to develop multimodal educative applications with chatbots. International Journal of Advanced Robotic Systems **10**(3), 175 (2013)
16. Grudin, J.: The case against user interface consistency. Communications of the ACM **32**(10), 1164–1173 (1989)
17. Gustafson, J., Bell, L., Beskow, J., Boye, J., Carlson, R., Edlund, J., Granström, B., House, D., Wirén, M.: Adapt—a multimodal conversational dialogue system in an apartment domain. In: The Sixth International Conference on Spoken Language Processing (ICSLP), Beijing, China. pp. 134–137 (2000)
18. Hearst, M., Tory, M.: Would you like a chart with that? incorporating visualizations into conversational interfaces. In: 2019 IEEE Visualization Conference (VIS). pp. 1–5. IEEE (2019)
19. Hobert, S., Berens, F.: Small talk conversations and the long-term use of chatbots in educational settings–experiences from a field study. In: International Workshop on Chatbot Research and Design. pp. 260–272. Springer (2019)
20. John, R.J.L., Potti, N., Patel, J.M.: Ava: From data to insights through conversations. In: CIDR (2017)
21. Kerlyl, A., Hall, P., Bull, S.: Bringing chatbots into education: Towards natural language negotiation of open learner models. In: International Conference on Innovative Techniques and Applications of Artificial Intelligence. pp. 179–192. Springer (2006)
22. Li, T.J.J., Radensky, M., Jia, J., Singarajah, K., Mitchell, T.M., Myers, B.A.: Pumice: A multi-modal agent that learns concepts and conditionals from natural language and demonstrations. In: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. pp. 577–589 (2019)
23. Liao, L., Zhou, Y., Ma, Y., Hong, R., Chua, T.S.: Knowledge-aware multimodal fashion chatbot. In: Proceedings of the 26th ACM international conference on Multimedia. pp. 1265–1266 (2018)

24. Messina, A., Augello, A., Pilato, G., Rizzo, R.: Biographbot: a conversational assistant for bioinformatics graph databases. In: International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. pp. 135–146. Springer (2017)
25. Nielsen, J.: Usability inspection methods. In: Conference companion on Human factors in computing systems. pp. 413–414 (1994)
26. Nielsen, J.: Ten usability heuristics (2005)
27. Oviatt, S.: Mulitmodal interactive maps: Designing for human performance. Human–Computer Interaction **12**(1-2), 93–129 (1997)
28. Oviatt, S.: Ten myths of multimodal interaction. Communications of the ACM **42**(11), 74–81 (1999)
29. Oviatt, S., DeAngeli, A., Kuhn, K.: Integration and synchronization of input modes during multimodal human-computer interaction. In: Proceedings of the ACM SIGCHI Conference on Human factors in computing systems. pp. 415–422 (1997)
30. Paixão-Côrtes, W.R., Paixão-Côrtes, V.S.M., Ellwanger, C., de Souza, O.N.: Development and usability evaluation of a prototype conversational interface for biological information retrieval via bioinformatics. In: International Conference on Human-Computer Interaction. pp. 575–593. Springer (2019)
31. Radziwill, N.M., Benton, M.C.: Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv:1704.04579 (2017)
32. Ratnaparkhi, A.: Maximum entropy models for natural language ambiguity resolution (1998)
33. Rouhi, A., Spitale, M., Catania, F., Cosentino, G., Gelsomini, M., Garzotto, F.: Emotify: emotional game for children with autism spectrum disorder based-on machine learning. In: Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion. pp. 31–32 (2019)
34. Sarter, N.B.: Multimodal information presentation: Design guidance and research challenges. International journal of industrial ergonomics **36**(5), 439–445 (2006)
35. Shawar, B.A., Atwell, E.: Chatbots: are they really useful? In: Ldv forum. vol. 22, pp. 29–49 (2007)
36. Spitale, M., Catania, F., Crovari, P., Garzotto, F.: Multicriteria decision analysis and conversational agents for children with autism. In: Proceedings of the 53rd Hawaii International Conference on System Sciences (2020)
37. Spitale, M., Silleresi, S., Cosentino, G., Panzeri, F., Garzotto, F.: " whom would you like to talk with?" exploring conversational agents for children's linguistic assessment. In: Proceedings of the Interaction Design and Children Conference. pp. 262–272 (2020)
38. Tegos, S., Demetriadis, S., Psathas, G., Tsiatsos, T.: A configurable agent to advance peers' productive dialogue in moocs. In: International Workshop on Chatbot Research and Design. pp. 245–259. Springer (2019)
39. Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A.: Paradise: A framework for evaluating spoken dialogue agents (1997)
40. Webster, J., Martocchio, J.J.: Microcomputer playfulness: Development of a measure with workplace implications. MIS quarterly pp. 201–226 (1992)
41. Zhi, Q., Metoyer, R.: Gamebot: A visualization-augmented chatbot for sports game. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–7 (2020)

Appendix A: Tasks of the user study

1. Can you extract the samples from TCGA with assembly GRCh38?
2. Try to download the URLs list regarding transcription factors for cervical adenocarcinoma.
3. Download the URLs list regarding h3k4me3 target extracted with chip seq.
4. Find and download the URLs for extracting the data regarding the simple nucleotide variation.
5. Explore the functionalities of the system in autonomy. List the functionalities you discovered.

Appendix B: Evaluation Questions

Open Questions:

1. What have you liked of the system?
2. What have you NOT liked about GeCo Agent?
3. Which were the main error you and/or the platform did?

Likert-scale Questions (scale 1 Totally Disagree – 5 Totally Agree)

1. Assigned tasks were difficult to accomplish
2. I found the platform easy to use
3. I found that the various functions in the platform were well integrated.

Yes/No Questions

1. Did you understand that you could click on the suggestion in the upper-central column to paste the text in the chat box?
2. Did you understand that you could answer just with the keywords (e.g. "which data do you want?" - "Annotations")?
3. Did you understand that you could use sentences instead of keywords (e.g. "which data do you want?" - "I would like Annotations")?
4. Did you understand that going with the mouse pointer over the pie charts you could see their details?
5. Did you understand that clicking on the download button that appears inside the box in the lower left panel you could download the URLs list file?