Usability evaluation of a Conversational Artificial Intelligence Interface

Ligia Maza-Jimenez[1] and Pablo Torres-Carrión[1][0000-0002-7606-0582]

[1] Universidad Técnica Particular de Loja, San Cayetano Alto-Loja, Ecuador
{lemaza, pvtorres}@utpl.edu.ec

Abstract. In this research, the usability of a Conversational Artificial Intelligence environment is evaluated from the metrics of effectiveness, efficiency and satisfaction, taking as a case of the virtual assistant "Max" of the Private Technical University of Loja (UTPL). The evaluation is carried out virtually, with the participation of 45 members of the university community (15 students, 15 parents, and 15 outsiders). An initial informative test prior to interaction is proposed; and the usability questionnaire PSSUQ (Adaptation to Spanish CSUQ) after the interaction. The research has a quantitative approach, with a concurrent and quasi-experimental design of a transectional type, and a descriptive scope. The results obtained show that the level of effectiveness, efficiency and satisfaction metrics are low, so it is concluded that the interaction strategy of the UTPL Virtual Assistant "Max" should be reviewed.

Keywords: Usability, Conversational Artificial Intelligence, Chatbot, University

1   Introduction

Human-Computer Interaction (HCI) is the emerging area of computer science closest to the user, which studies the communication that occurs between computers and the people who use them [1], [2]. In this context, User Experience (UX) refers to the feeling of the person when interacting with a computer system [3] and manages mainly three dimensions or experiences: aesthetic, significant and affective; One of the key factors for a good user experience is usability. The international standard ISO 9241-11 defines it as the "Degree to which a product can be used by certain users to achieve specific objectives with effectiveness, efficiency and satisfaction in a context of specific use" [4]. According to this definition, usability is made up of three main elements, which are related to the characteristics and objectives of the users and the context of use: a) Effectiveness: it is defined as the precision and integrity with which users achieve specific objectives; b) Efficiency: on the other hand, as the resources spent in relation to the precision and integrity with which the users achieve these objectives; and, c) Satisfaction: as "being free of discomfort and positive attitudes towards the use of the product (system, service or environment)".

Research in usability, as well as other sciences, has been enriched by Artificial Intelligence (AI), which has allowed the establishment of intelligent interaction environments [5]. According to García [6], AI tries to explain mental functioning based on the development of algorithms to control different things, combining several fields, such as robotics, expert systems and others, which have the same objective, which is to try to create machines that can think for themselves. AI can be divided into three wide scenarios [7]: a) narrow or weak that is limited to one functional area; b) general, containing the power of reasoning, problem solving and abstract thinking; and, c) the super intelligence that is the maximum level when the AI exceeds the human intelligence. Within the weak AI is Conversational, with a strong relationship with HCI, which guarantees that simulated conversations can be made with a computer [8], opening the space to develop empathetic behaviors between the machine and the user.

Conversational Artificial Intelligence (IAC), according to Nieves [9] is responsible for the logic behind the robots, that is, it is the brain and soul of the chatbot. Without IAC, a bot is just a bunch of questions and answers. Additionally, the IAC is powered by Natural Language Processing (NLP), which focuses on human language interpretation, while developers present the basic framework of how a conversation can develop. Simply put, IAC and humans work together to create a virtual conversational experience.

Chatbots are programs that use natural language processing (NLP) in a question and answer system (QA systems) [10]. Its purpose is to simulate an intelligent dialogue with a human interlocutor, through text messages through a console or through voice. They started from the Turing Test [11] that originated in 1950 by Alan Mathison Turing, From the passage of time until today there is an interesting evolution of chatbots, the most current ones are hosted on websites to personal virtual assistants on mobile devices which include features that make them one of the most attractive tools for a company or institution. As a result of the health emergency due to the expansion of the coronavirus COVID-19, chatbots have gained ground in terms of online assistance, in order to keep society informed on different topics. For this reason, this study evaluates the usability of the UTPL "Max" virtual assistant, based on the effectiveness, efficiency and satisfaction metrics.

To this study, it is proposed a quantitative approach as the methodology, with a concurrent and quasi-experimental design, of a transactional or transversal type, with a descriptive and correlational scope. The results obtained indicate that there is a low level of usability due to different factors such as: presence of errors, difficulty in using the system, lack of friendly interface, imprecise information and preference for interaction with human beings.

2 Methodology

It is proposed to evaluate the usability of an active IAC environment in a higher education institution. The UTPL "Max" virtual assistant has been selected as the evaluation interface. Following the indications of ISO 9241-11, the effectiveness, efficiency and satisfaction metrics have been selected.

2.1 Research design

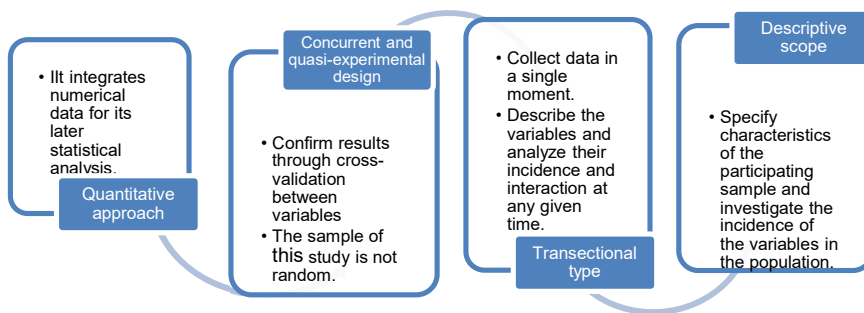Taking into account what this research mentions [12], it has the following design see Fig.1:

Fig. 1. Research design

2.2 Sample

Table 1. Sample of participants

TYPES OF USER	DESCRIPTION	n	%
Students	• Undergraduate degree at the Universidad Técnica Particular de Loja (English, Biochemistry, Accounting and Auditing, Medicine, Psychology, Economics and Architecture)	15	33,33
Parents	• Postgraduate at the Universidad Técnica Particular de Loja (Communication Sciences and Technologies)	15	33,33
Externs	• Presential and distance	15	33,33
TOTAL		45	100

The sample is of a non-probabilistic type, selected following the technique called accidental or snowball, which takes advantage of the people available at any given time for the purpose of the study [13]. It should be noted that the type of sample was conditioned by the national health emergency established on March 17, 2020 by Presidential Decree 1017 Agreement No. 00126-2020 [14], see Table 1.

2.3   Instruments

The instruments used for the present study are the following:

- Initial test: It is an ad-hoc survey that collects personal data and identification of the emotional state of the participants. It consists of 10 items, among which are: informed consent; personal data, classification of emotions, reason for emotions, effectiveness, efficiency and satisfaction in relation to the interaction of with the Virtual Assistant "MAX" of the UTPL.
- System usability questionnaire after the PSSUQ study (Adaptation to Spanish CSUQ): Whose objective is to evaluate the general satisfaction of users with an interface? Account of 16 items, which is answered by a Likert-type scale of 7 options classified from one to seven, is classified into three variables a) Utility of the corresponding system of item 1-6; b), Quality of the information to items 7-12; c) Quality of interface to items 13-16. The higher score, the higher level of satisfaction in general and a reliability (Cronbach's Alpha, $\alpha = .96$) [15].

2.4   Procedure

Stage 1. Interaction planning: This stage specifies the technology that will be used to capture the information related to the characteristics of user behavior while participating in the evaluation with IAC.

- In the first instance, the evaluation instruments were selected: Initial ad-hoc test which was developed using the Microsoft Forms tool, the system usability questionnaire after the PSSUQ study (adaptation to Spanish CSUQ) and as an environment IAC used the virtual assistant "MAX" from UTPL.
- The next step was to determine the sample that is divided into 3 groups: Students (Postgraduate in Information Sciences and Technologies and Undergraduate students corresponding to English, biochemistry, accounting and auditing, medicine, psychology, economy and architecture of the UTPL), Parents and External Persons, made up of 15 participants each; the same ones that have been contacted through social networks, WhatsApp and phone calls.
- To end the first stage, a preliminary test planning or piloting of the application of the questionnaires was carried out based on an application protocol guide, to clarify, specify and perfect the method of application of instruments.

Stage 2. Execution: The execution phase consists of three moments:

- In the first moment, the ad-hoc emotion identification questionnaire was applied online, which measures the usability of the virtual assistant "MAX" of the UTPL by the participants. This process lasted 5 min. approximately.
- In the second moment, the participants share the screen through the ZOOM web conferencing and video conferencing services platform and will interact with the UTPL virtual assistant "MAX"; This process will be recorded for a period of 2 min.
- Finally, in the third moment, the participant answers an online usability questionnaire after the PSSUQ study (adaptation to Spanish CSUQ) online, regarding their appreciations of the experience to verify the validity and reliability of the Chatbot instrument (Assistant Virtual "MAX" of the UTPL) used, the test lasted 3 min. approximately; at that time, the metrics of efficacy (average of complete tasks and usability problems) and efficiency (task time and errors) and satisfaction (satisfaction scales) were evaluated [16]

Stage 3: Analysis of results: In this section, the data was analyzed with the results obtained in the initial test and usability questionnaire of the system subsequent to the PSSUQ study, using descriptive tables in Excel.

3 Results

This section is described according to the objective set based on the initial test and usability questionnaire of the system after the PSSUQ study, measuring the usability of the virtual assistant "MAX" of the UTPL, see Fig 2.
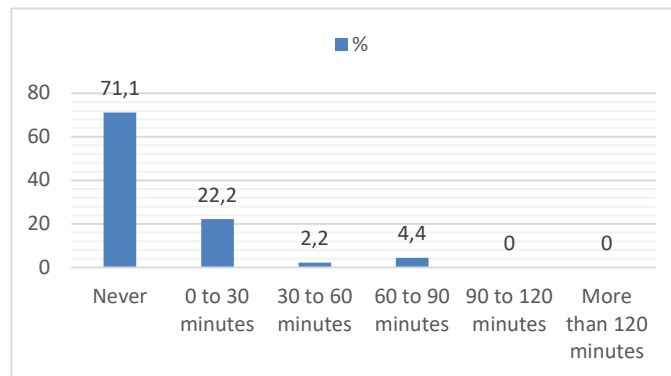

Fig. 2. On average, how long a week do you use the UTPL Virtual Assistant "MAX"?

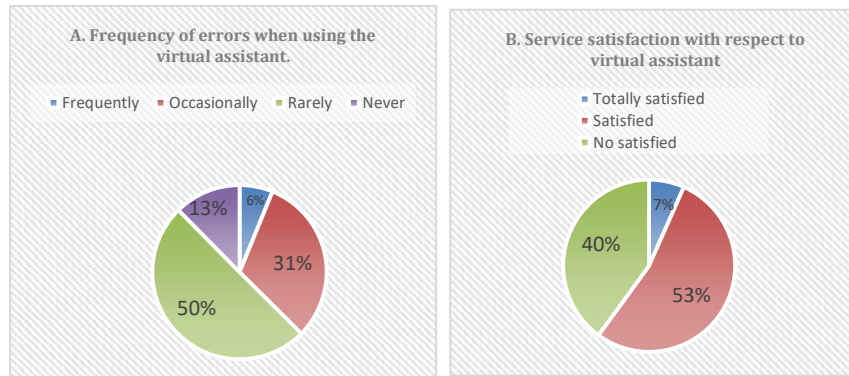71.1% of users mention never using the UTPL "MAX" virtual assistant a week.

Fig. 3. Usability parameters according to the initial test

As part of the research, the system usability questionnaire after the PSSUQ study is applied after the interaction see Fig. 3, 4:
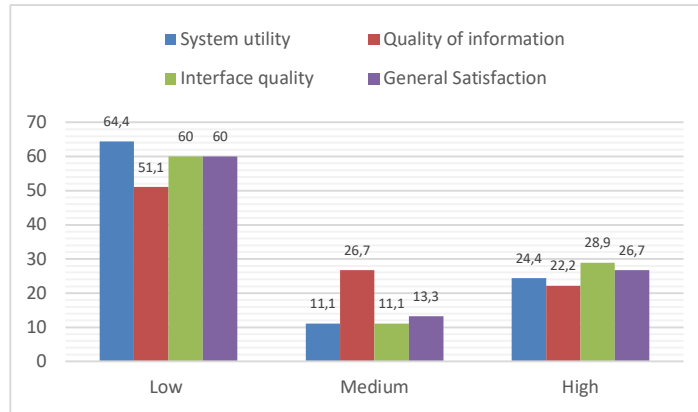

Fig. 4. Variables of the PSSUQ usability questionnaire.

4   Discussion of results

In this study 71.1% of the users did not use the virtual assistant "MAX" of the UTPL, while the participants who did it 50% rarely found errors. Likewise, 53% are satisfied with the service at the time of interaction. For this reason, the effectiveness, efficiency and satisfaction metrics have a medium level of usability. Regarding the system usability questionnaire after the PSSUQ study, the variables of quality systems, information quality, interface quality and perceived satisfaction obtained a low level of more than 50%, which means that review the interaction strategy of the Virtual Assistant "Max"

of the UTPL. The same occurs with the study carried out by Valle [17] in Mexico, with the aim of using a conversational assistant to help resolve doubts of university students in their degree process, obtained as a result that the people who used the virtual assistant they rated their interaction low, compared to people who interacted with a human to obtain academic information, as well as indicating that the design and responses that do not give enough information could be improved. Along these same lines, Peralta [18], in his study carried out in Peru, with the aim of measuring the level of improvement in personalized assistance in the process of obtaining a degree in a university through the use of chatbot, found that the effectiveness of virtual assistant depends on the use of an agile methodology, simplicity of usability and operation of the program. Taking into account the aforementioned studies, the similarity that exists with this study, regarding the level of usability, can be evidenced.

5   Conclusions

In this way, it is concluded that the low levels of usability obtained in this study are due to different factors such as: presence of errors, difficulty in using the system, lack of friendly interface, imprecise information and preference for interaction with human beings. In this sense it is important to take into account what was mentioned by Sánchez [19], the usability of a product is the quality attribute that is given to an interface based on its easy interaction and the satisfaction it produces when used. Therefore, it is recommended that the virtual assistant be widely disseminated, and that an acceptable service provision be offered that guarantees its usability.

References

1. Marcos M.-C., "HCI (human computer interaction): concepto y desarrollo," *El Prof. la Inf.*, vol. 10, no. 6, pp. 4–16, 2001.
2. Sears A. and Jacko J. A., *Human-Computer Interaccion*, Second. 2007.
3. Boada N., "¿Por qué es tan importante el User Experience o Experiencia del Usuario?," 2017. [Online]. Available: https://www.cyberclick.es/numerical-blog/por-que-user-experience-o-experiencia-del-usuario. [Accessed: 13-Mar-2019].
4. International Organization for Standardization and ISO, "INTERNATIONAL STANDARD ISO 9241-11," 1998.
5. Mira J., "Inteligencia artificial, emoción y neurociencia," *Arbor*, vol. 162, no. 640, pp. 473–506, Apr. 1999.
6. García A., *Inteligencia artificial : fundamentos, práctica y aplicaciones*. RC Libros, 2012.
7. Gomes C. C. and Preto S., "Artificial intelligence and interaction design for a positive emotional user experience," vol. 722. Springer Verlag, The Research Centre for Architecture, Urbanism and Design, Universidade de Lisboa, Faculdade de Arquitetura, Rua Sá Nogueira, Polo Universitário, Alto da Ajuda, Lisbon, 1349-055, Portugal, pp. 62–68, 2018.
8. Brinquis C., "IA conversacional: conversaciones reales con un ordenador.," 2019. [Online]. Available: https://www.incentro.com/es-es/blog/stories/ai-conversacional-conversaciones-reales-con-un-ordenador/. [Accessed: 13-Apr-2019].

9. Nieves B., "IA Conversacional: definición y conceptos básicos.," 2018. [Online]. Available: https://planetachatbot.com/ia-conversacional-conceptos-basicos-y-la-definicion-107529e213c1. [Accessed: 04-Dec-2019].

10. Fitrianie S., "My_Eliza, A Multimodal Communication System," in *Proceedings of Euromedia*, 2002, pp. 1–187.

11. Turing A. M., "Computing Machinery and Intelligence," 1950.

12. Hernández Fernández R., C., and Baptista P., *Metodología de la investigación.*, Sexta. México, 2014.

13. Espinosa P., Hernández H., López R., and Lozano S., "Muestreo de Bola de Nieve," *Departamento de Probabilidad y Estadística UNAM*, 2018. [Online]. Available: https://es.scribd.com/document/379661920/Proyectofinal-Bola-de-Nieve. [Accessed: 15-Jul-2020].

14. C. E. Constitucional, "Registro Oficial - Órgano de la República del Ecuador," Ecuador, 2020.

15. Hedlefs M. I., González A. de la G., Sánchez M. P., and Garza A. A., "Adaptación al español del Cuestionario de Usabilidad de Sistemas Informáticos CSUQ / Spanish language adaptation of the Computer Systems Usability Questionnaire CSUQ.," *Revista Iberoamericana de las ciencias compitacionales e informática*, 2016. [Online]. Available: https://www.reci.org.mx/index.php/reci/article/view/35. [Accessed: 15-Jul-2020].

16. Omar Mohamed H., Yusoff R., and Jaafar A., "Quantitive analysis in a heuristic evaluation for usability of educational computer game (UsaECG)," in *Proceedings - 2012 International Conference on Information Retrieval and Knowledge Management, CAMP'12*, 2012, pp. 187–192.

17. Valle L., López J., and Garcia M., "Desarrollo e implementación de un bot conversacional como apoyo a los estudiantes en su proceso de titulación," *Int. Conf. Robot. Comput. 2013*, no. May, pp. 361–365, 2013.

18. Peralta A. G., "Chatbot para la asistencia personalizada en el proceso de obtención de título en la modalidad de tesis para los bachilleres de la escuela profesional de ingeniería de computación y sistemas de la UPAO," Universidad Privada Antenor Orrego, 2018.

19. Sánchez W., "La usabilidad en Ingeniería de Software : definición y características," *Ingnovación. Rep. Investig.*, no. 2, pp. 7–21, 2011.