ACCESSIBILITY, BIASES AND ETHICS IN CHATBOTS AND INTELLIGENT AGENTS FOR EDUCATION

Emmanuelle Gutiérrez y Restrepo¹,², Martín Baldassarre², Jesús G. Boticario¹

¹aDeNu Research Group. Artificial Intelligence Department. Computer Science School, UNED (SPAIN)
²Fundación Sidar – Acceso Universal (SPAIN)

Abstract

The use of chatbots is increasingly common in universities around the world to support both educational and administrative tasks. There is also a growing awareness of the importance of inclusive education and, as a consequence, the need to comply with accessibility requirements regarding web contents and web interface (WCAG). These requirements are also supported by specific laws in most countries. But such awareness does not yet reach the interfaces of conversational agents or chatbots that are being developed.

Within the framework of the European project ACACIA, co-funded by the Erasmus+ program of the European Union, Artemisa, a chatbot has been created dedicated to fight against sexual harassment and search for volunteers to promote the acceptance of diversity and tolerance, has been created using a platform that facilitates the generation and management of this type of artificial intelligences.

But to what extent Artemisa is an accessible chatbot or not? Is it ethically acceptable to make use of a tool that is intended to support inclusiveness but presents accessibility barriers to some users? What is the current state in terms of accessibility compliance of chatbots that work on social networks?

This article seeks to answer this and other questions related to accessibility in Chatbots, conversational agents and virtual assistants. Based on the answer to these questions it follows that there is a need for training in interculturalism and web accessibility to combat the biases that present such entities, which are endowed with artificial intelligence and, unfortunately, in some cases causing serious damage to some people.

Keywords: Conversational agents, Chatbot, accessibility, biases, ethics, empathy, artificial intelligence.

1 INTRODUCTION

Joseph Weizenbaum [1] of MIT in 1966 developed a program that was intended to be able to deceive humans by making them think that they were talking to another person, who was pretended to be a psychotherapist from the humanist or Rogerian school. This "entity" was called ELIZA and is considered the first conversational agent or chatbot, even though it was created long before the term 'ChatterBot' was coined three decades later to define conversational programs.

Until 2001, when it was considered that the first chatbot was actually created, several entities were developed such as ALICE [2], Albert One [3] that won the Loebner Prize Contest both in 1998 and in 1999 or Mitsuki [4] that continues to operate today and that He has won that prize four times. These chatbots were limited to holding conversations with more or less success since they only had basic functions. With the arrival of SmarterChild [5] everything changed. It was no longer a mere conversational bot, but almost a first virtual assistant sketch that became compatible with AIM, MSN Messenger, AOL Instant Messenger and ICQ, and went from being a simple entertainment to becoming a method to offer different services. It was a bot ahead of its time, a kind of father for the current Siri, Cortana or Google Now.

Chatbots, conversational agents or virtual assistants, are being used for some time by numerous universities to facilitate students to perform various procedures, such as obtaining information about a course, supporting enrolment or virtually moving through the campus, but also in more academic tasks, such as obtaining information on the degree of students’ satisfaction in a given course and even carrying out evaluations of a subject or module.

Today there are already several companies dedicated to creating chatbots for universities, which deal with informative or management tasks, and these have shown positive results. But the use of these intelligent agents as teachers is scarcer though it is much appreciated by students. These have not yet
shown substantial improvements from the academic point of view [6]. In relation to this, and contrary to expectations, the text-based interfaces were rated significantly higher in terms of enjoyment and utility and led to better learning outcomes compared to the conversational agent [7]. In a recent study, positive results have been obtained in the use of chatbots to improve the scalability of practical tasks in MOOCs (Massive Open Online Courses) and thus helping to overcome the limitations of personnel [8]. Notice that this was due to the personalization achieved in the generation of evaluations, without which most likely such positive results would not have been obtained.

1.1 Artemisa, defender of freedom and diversity

Artemisa [9] is a chatbot dedicated to both fight against sexual harassment and support the university community of the Peruvian National University of San Marcos (UNMSM). She was created as a collaborator of the “Apoya” Module (support module) of the Acacia Center of that university. Its usage is planned to be extended to cope with some needs in other Acacia CADEPs (Centers of support and educational and professional development, by its initials in Spanish) that make up the Acacia Centers Network.

Artemisa facilitates those who are suffering sexual harassment to file a formal complaint. It explains the UNMSM’s protocol of action against sexual harassment and allows to report on any situation of harassment. This involves sending documentation with photos or videos.

Artemisa also encourages and supports the user in becoming a volunteer to collaborate with the “Apoya” Module (Fig. 1 View of the three options Artemis offers to become a volunteerFig. 1), covering the following tasks:

- Becoming a Custodian Angel to support people with disabilities, people who are suffering bullying or any type of harassment.
- Becoming a Professor for Diversity to help other teachers in implementing in their classes the Apoya Guides regarding various social problems.
- Becoming a Community Driver to help the Apoya Module in creating communicative channels dealing with tolerance and acceptance of diversity.

When the user chooses the option to inform himself, he is presented with four options from where to obtain information: The UNMSM’s Protocol of Action, the Peruvian national legislation and the norms of the UNMSM. These cover how to detect the harassment and what to do about it, along with curricular orientations; i.e., how to deal with the subject of sexual harassment in the class, no matter the specific given subject.
These curricular guidelines come from the Guide on sexual harassment published for the Apoya Modules [10], which is one of the 10 guides that deal with different social problems that can lead to university exclusion and whose use is promoted from the Acacia Centres Network.

The aim of these series of guides is learning tolerance and acceptance of diversity and dealing with the following topics: a) Armed conflict, b) Ethnic and cultural diversity, c) Drug addiction, d) Religious intolerance, e) Violence against gender, f) Domestic violence, g) Sexual diversity, h) Sexual harassment, i) Disability, and j) Psychosocial problems.

All of them share a content structure, as follows:

- Glossary
- Legal and educational framework
- Characterization of the actors
- Curricular attention and orientation guidelines
- Survey for detection

They are, therefore, valuable tools for teachers, as they guide them in detecting and dealing with people, fellow teachers, administrators, technicians and students who may be going through situations that may lead them to drop out of college. They also explain how to include such topics in the subject taught, even when this seems to be far away from these issues. These guidelines are intended to contribute to the university community being more tolerant and inclusive, which entails accepting diversity, respecting one another and their respective differences.

Artemisa is inserted in a program to make the university truly inclusive. It has to contribute to improve the university climate, change attitudes and prevents the university from generating any kind of exclusion. From these requirements follows that it must be accessible, able to be used by any person regardless of whether or not she has any kind of disability. It must also looks and be inclusive avoid any kind of bias that can make anyone feel excluded. It must not be either sexist or racist. In other words, she must have an ethical behaviour with anyone.

2 METHODOLOGY

To facilitate the understanding of the issue, we will consider “inaccessibility” as if it were just another bias and thus we evaluate Artemisa with this view in mind. This is justified in as much as presenting accessibility barriers to those who are supposed to be supported generates exclusion, which is the essence of any other bias.

It should not be necessary to explain why it is important that a chatbot, or any other software with an interface that communicates with humans, should be accessible to all of them. Especially when that application works in the educational field. Bear in mind that one in seven people in the world has a disability, i.e., a billion people. Needless to say that all of us will end up having some disabilities because of our age. The world is committed to an education throughout life. Universities must be inclusive if we want a fairer and better world. Today the accessibility of digital content is an obligation required by law in most countries.

In order to evaluate Artemisa, the participation of people with different conditions of knowledge and use of computer applications was sought, as well as different access modes. The aim was to quickly check the perceived degree of accessibility of the interface provided to Artemisa by Messenger, not to perform a complete accessibility audit on the (WCAG 2.1) [11] guidelines. This is because the author of the chatbot, even if she wanted to, she could not make any changes to that interface. Later we will discuss this question and its importance.

Among possible users, we selected a paradigmatic case, a blind person who has used different combinations of browsers and screen readers, from a notebook with Windows 7.

The browsers used were Internet Explorer 11, Google Chrome 73, and Mozilla Firefox 66.

The Screen Readers used were JAWS 2018 and NVDA 2019.1.1.

These users interacted freely with the chatbot and subsequently reported.
3 RESULTS

Since this article is intended for the 11th annual International Conference on Education and New Learning Technologies, we will not present the technical details of the resulting evaluation, but we will discuss some of the problems detected and the general result of the evaluation.

For a screen reader user, who access Artemisa from the Apoya Module webpage of the UNMSM's Acacia Centre, it appears at the bottom of the page. In fact, at the bottom of the page there are two untitled frames, which makes identification difficult.

The second frame contains an unlabelled button, which serves to hide or display the Chatbot, but only for people with vision, since for the screen reader it is always available, but it does not report the change of state of the same. The first frame contains the Chatbot interface.

The first time it is accessed the user needs to log into Facebook Messenger. This could be done without accessibility problems.

After logging into Messenger, a user interface with the following elements is presented:

- An unlabelled button, which activates the following options:
  - Deactivate messages
  - Silence conversation
  - See in Messenger
  - Go to the Facebook page
  - Add Messenger to your website

- A second untagged button, which activates the following options:
  - Menu
  - Main Menu
  - Start Over
  - Close pop-ups and return

- A text box where you can enter the message.

- An unlabelled button that sends an emoticon or emoji, the thumbs up sign that, therefore, can only be perceived by people with vision.

The navigation with the keyboard does not present any accessibility issues, both with the keyboard shortcuts of the screen reader and with the tabulator.

When sending a message, the bot responds immediately, but returning to focus on the text box that allows the user to write, which forces him to leave it to scroll up and read the answer. The answer can be a simple text, an enriched text, or link texts for the user to select an action.

An error message in English is always included in the response as a link: "This message did not send. Click to try again. "This message is not displayed, it is only read by the screen reader. Also sometimes other messages appear that are only read with the screen reader as "Sticker Like, thumbs up of" or "customerchat app_id = 1678638095 ". The latter only with the NVDA.

If the answer is short, the navigation is normal, but if the answer is long or contains several lines it is very difficult to find where the answer begins. The response texts have a button role, and the actions are identified as links.

As a navigation strategy we chose to use the quick navigation shortcut of the screen reader "Shift + B", until the beginning of the answer and then the tabulator.

The answers may contain an emoji or emoticon, and these are not always read as such by the screen reader. Even an emoji image had as an alternative text the emoji, when it should be a text with printable characters. If the answer is an image, an alternative text is not provided.

The identification of the author of each message is not clear, and sometimes even confuses.

It was verified that all message history can be found in the Messenger App on the iPhone, and the reading of the history is better than on the web, but it was not possible to click on the links of the answers. This would be an accessibility problem in the Messenger App.
The blind reviewer conclude that: “The Chatbot accessibility is quite good, but it would be desirable to be able to improve it. If the platform does not allow to customize or improve the detected problems, it would be better to look for another platform or build your own”.

As positive points, he highlighted:

- The message history always available in Messenger, and the last messages are displayed if the page is reopened in any browser.
- Navigation in message history with the keyboard is good.
- Interaction is fast
- The answers in general are clear and friendly.

We have, then, a Chatbot quite accessible but that can be improved. The reviewer has detected some faults but has positively assessed some aspects of both the interaction and the functioning of the Chatbot, including its clarity and an aspect of its personality, friendliness, which also has a positive influence.

### 3.1 Who is responsible for accessibility?

Some authors have already studied the issue of the chatbots accessibility [12], [13], [14], [15], [16], [17], [18] reaching practically the same conclusions, that is, that most of the existing chatbots present accessibility barriers to users, which would be easy to eliminate if the accessibility guidelines were applied when designing and creating those intelligent entities.

But who is responsible for that? The creation of Chatbot nowadays and surely more and more in the future, is based on the existence of platforms that facilitate the definition of its components and the algorithms that serve for the understanding between the bot and the human.

Many of these platforms are created so that people who do not have programming skills can easily define the behaviour of the Chatbot, its responses, its 'understanding' of what the human writes or says and the inclusion of the content to be shared. But regretfully, most do not provide access to the source code of the interface, which would improve accessibility.

The logical recommendation, then, would be to create the Chatbot from scratch, without relying on any platform or choosing one that does provide access to the source code of the interface. But there is another actor here, if we want our Chatbot to work with instant messaging services like Facebook Messenger for example. In this case we will not have the possibility of modifying the elements of the interface of these applications, so we will have to conform or, in any case, resort to the community of developers to note the accessibility barriers that the application is generating. Fortunately, the large commercial houses, nowadays, have teams dedicated to improving the accessibility of their products, and it is possible that our complaints are heard.

From the above follows that it is necessary that the developers are trained in accessibility. They need to acquire the knowledge, skills and abilities to be able to detect accessibility barriers and eliminate them.

This training in accessibility, in addition, is especially important to not get carried away by the recommendations of companies that offer frameworks that may be wrong. For example, although it is convenient to use Microsoft Azure to create our Chatbot, since it is quite configurable, we must avoid using the embedded webchat control provided by Microsoft, as the bot's user interface will not be accessible. An accessible user interface (UI) will be achieved by creating your own UI in HTML5, CSS and JavaScript.

Moreover, we must be careful with the recommendations that they provide us to design our chatbot. For example, the Microsoft Azure Bot Service documentation in the "Design and control conversation flow" section says [19]:

> While your bot may be procedural centric, the user may decide to do something entirely different or ask a question that may be unrelated to the current topic. In the example above, the user asks a question rather than providing the yes/no response that the bot expects. How should your bot respond?

- Insist that the user answer the question first.
- Disregard everything that the user had done previously, reset the whole dialog stack, and start from the beginning by attempting to answer the user's question.
There is no right answer to this question, as the best solution will depend upon the specifics of your scenario and how the user would reasonably expect the bot to respond. For more information see Handle user interruption.

That design recommendation is wrong. Clearly, from the point of view of accessibility, there is a correct answer. The first two options are incorrect, since the system must remember the data that the user has entered into a system. The system must be friendly and allow the user to correct errors, change his mind or recover the session according to criteria 2.2.1, 2.2.5 and 2.2.6 of the WCAG 2.1 [11].

Precisely the document Understanding of the criterion 2.2.6 gives an example similar to the one used in the Microsoft Azure suggestions that can help us to understand why those are wrong [20]:

During the completion of an online process, such as to reserve a hotel room or purchase a plane ticket, a user with a cognitive impairment may become overwhelmed with lengthy instructions and data input required to complete the process. The user may not be able to complete the process in one sitting and may need to take a break. Users should be able to leave a process without losing their current place within the process, and without losing information that has already been entered. If users cannot take a break and check their work, many will often be unable to complete a task correctly.

This Success Criterion works in tandem with Success Criterion 2.2.1 Timing Adjustable, but is specifically focused on notification of timeouts related to user inactivity.

Therefore, training in accessibility is essential to prevent designers and developers from falling into recommendations that they may consider important to follow, since they come from large companies that are supposed to know what is to be done, but that will actually take them to create barriers for their users so that their chatbots will become agents of exclusion.

These facts lead us to reflect on the ethical issues involved from several perspectives. It is worth making the following questions: Is it ethical to create an inaccessible Chatbot when it will work in an educational environment? Is the use of an instrument for good that presents barriers to some users ethically acceptable? Is not training in accessibility essential to prevent intelligent agents from being agents of evil? To what extent can training on accessibility affect the ethical behaviour of a Chatbot?

3.2 Ethical considerations in the creation of Chatbots for education

A chatbot is an artificial intelligent entity and, therefore, must be subject to the ethical standards applicable to artificial intelligence. Very recently, research has begun [21] but guidelines have not yet been defined or policies developed or laws enacted that address the specific ethical issues raised by the use of Artificial Intelligence in Education [22]. In Europe we are in the process of defining our Ethics guidelines for trustworthy AI. A piloting process will be set up as a means of gathering practical feedback on how the assessment list, that operationalises the key requirements, can be improved. The published guidelines already understand that accessibility is an essential principle to be fulfilled by any entity with artificial intelligence. Thus, guideline number 5 Diversity, non-discrimination and fairness; arises from the idea that “in order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle” [23].

That fifth requirement includes: Avoidance of unfair bias, Accessibility and universal design, Stakeholder Participation. The relationship between these three actions is probably not very evident, but they do and are very close, as we proposed in [15] where we explained the unexpected but fundamental relationship between the strategy defined for the Educational and Professional Development and Support Centres, results from the ACACIA European project, and the future of artificial intelligence.

Fixing discrimination in algorithmic systems is not something that can be solved easily. It’s a process ongoing, just like discrimination in any other aspect of society [24].

3.2.1 Impact of training on accessibility and acceptance of diversity

There are already several tools to avoid biases in artificial intelligences, such as HoloClean [25], HoloDetct [26], Ethika [27], and AIFairness360 [28], with different types of approximations to the problem. But it is possible that the data you use to test the performance of your model has the same biases as the data you used to train it. Because, after all, the biases come from the humans who have created the applications, even from implicit misconceptions, because the machines learn from what
people know or think they know implicitly. We see this, for example, in the space of mathematical language, the words for flowers are close to those that mean "pleasant", while those referring to insects are close to those that express unpleasant things. This reflects the common vision of the relative merits of flowers and insects, although it is absolutely unjust because, for example, a ladybug is pretty and pleasant even if it is an insect, while the flower of the Rafflesia arnoldii stinks and is unpleasant.

Therefore, it is essential to provide training in accessibility and acceptance of diversity to designers and developers. The Sidar Foundation [29] reference entity in Spanish on digital accessibility, as a cooperating entity of the ACACIA project [30], provides its experience and system of courses for this purpose, applying the lessons learned in such implementation in other projects. One of the educational resources that the Foundation has designed for its courses are a series of "personas" who represent different subjects, different users, who have different needs and preferences when navigating and interacting with web content or mobile applications. The appearance of these "personas" has been updated in the ACACIA project and now they are also used by the Acacia Centers in the training program that is used to teach on accessibility.

From them, a new series of 10 "people" has been created, corresponding with 10 documents or Guides that deal with different social problems that can lead to university exclusion. The purpose of this guides is to improve the level of tolerance and acceptance of diversity in the university environment. The themes that deal are: Armed conflict [31], Ethnic and cultural diversity [32], Drug addiction [33], Religious intolerance [34], Gender violence [35], Domestic violence [36], Sexual diversity [37], Sexual harassment [10], Disability [38], Psychosocial problems [39].

These personas are especially useful for designers and developers to take into account the diversity of users and in this way they can create inclusive content and applications that meet their needs and preferences. Also to avoid the so-called trap of portability [24]. These personas have to become the "best friends" of the designers and developers, since it is difficult that in real life they have the opportunity to personally meet real people who are in these very diverse situations. Thus, they will have them in mind when designing and programming.

The biases of artificial intelligences can come from the author's prejudices, from the data with which it is fed before publishing it or from the data with which it is fed daily after publishing it. That is why different approaches to the solution for the elimination of biases are being made, from algorithms that help detect and mitigate hidden biases within the training data to those that mitigate the biases learned by the model regardless of the quality of the data.

It will be important to tackle the problem by fighting from different fronts. But it seems that the community of artificial intelligences can come from the author's prejudices, from the data with which it is fed before publishing it or from the data with which it is fed daily after publishing it. That is why different approaches to the solution for the elimination of biases are being made, from algorithms that help detect and mitigate hidden biases within the training data to those that mitigate the biases learned by the model regardless of the quality of the data.

But interestingly, it is not the same with the story of Xiaoice [41], which was created by the same company but for the Chinese public and that has been working and growing since 2014 without incident.
and delighting its users, because today it is a poet, singer, designer, journalist, among many other skills acquired through interaction with its users. Clearly, cultural and educational differences seem to play an important role here [15].

This fact leads us to think how important is inclusive education and how important it can be training in accessibility, tolerance and acceptance of diversity as we propose from the ACACIA project.

4 CONCLUSION

Bearing in mind the increasing use of chatbots in education, in this article we have reported a preliminary analysis on the accessibility of the prototype Artemisa chatbot, which reflects that although the perceived accessibility is good, there are pending problems. These issues relate to those derived from the usage of current chatbot platforms.

In as much as these platforms and their usage present accessibility barriers they do not comply with the ethical issues involved. In this view we considered who is responsible for the degree of accessibility of the chatbot, understanding that although ultimately the responsibility is the owner of the website in which it is inserted, in the case of institutions of higher education, platforms and existing frameworks should provide developers with the possibility to intervene in the elements of the user interface or guarantee compliance with the accessibility guidelines.

Another important issue regarding the usage of a particular chatbot platform is the indications or suggestions of good practices that they give in the web sites or in the documentation of the platforms or frameworks. In some cases these mislead the developers and designers, as it can be seen in the example here discussed. Therefore, training in accessibility is essential to avoid that designers and developers fall into recommendations that come from sources that they consider solvent, but that can lead them to create barriers of accessibility if they do not have a deep knowledge of the applicable techniques and the navigation strategies that users have.

Avoidance of unfair bias, Accessibility and universal design, Stakeholder Participation; are requirements of the European ethical guidelines for artificial intelligence that can be achieved as long as we adequately train designers and developers in accessibility, tolerance and acceptance of diversity, as we proposed in [15] where we explained the unexpected but fundamental relationship between the strategy defined for the Acacia Centres, results from the ACACIA European project, and the future of artificial intelligence.

This training is not only important for designers and developers, it is for the whole society, because as it is discussed in this paper, artificial intelligences learn from their interaction with humans and acquire the biases they transmit to them.

All of the above leads us to conclude that training in accessibility, tolerance and acceptance of diversity is an effective means to fight against the biases currently present by artificial intelligences while creating a truly inclusive education as we propose from the ACACIA project.

ACKNOWLEDGEMENTS


REFERENCES


