Co-design and testing of an educational chatbot for teaching political representation through history

Eleonora Belligni^{2,†}, Sara Capecchi^{1,*,†}, Davide Cellie^{1,†}, Rossana Damiano^{1,*,†}, Carla Scilabra^{2,†} and Marino Zabbia^{2,†}

¹Dipartimento di Informatica, Università di Torino ²Dipartimento di Studi Storici, Università di Torino

Abstract

This paper presents the design and implementation of a chatbot developed to support history teaching in primary schools. The project was inspired by two main goals: on the one side, we wanted to involve teachers and educators in the creation of the chatbot, asking them to co-design both the character and the content to be delivered; on the other side, we wanted to test the use of dramatization techniques for the design of the chatbot. The validation of the chatbot was conducted through an experimental phase involving five primary school classes. The classes experience with the chatbot received positive ratings, thus confirming its capability to engage the users with its behavior and appearance. The study also demonstrates the importance of teacher collaboration in the development of educational technologies: the design principles behind chatbots illustrate the broader mechanics of how generative AI systems operate in tasks like content creation, problem-solving, and personalized interactions. This allowed us to introduce AI-based tools to the teachers in a conscious and participatory manner.

Keywords

Educational chatbots, AI and education, artificial characters, dramatization

1. Introduction

In the last few years, thanks to the advent of AI-powered speech and text-based chats, the conversational mode for acquiring information from and engaging with artificial agents has become an everyday experience for many children. The architectures of today's end user applications, then, make chatbots potentially available in all areas reached by mobile services, since computation tends to occur on centralized remote severs, with only a major obstacle still represented by linguistic minorities. In this sense, chatbots represent an opportunity for teaching, as witnessed by several research projects that have explored their use in schools [1, 2, 3]. Thanks to the familiarity with them developed by the students, and the availability of off-the-shelf tools for creating them, in the near future chatbots may be autonomously developed by teachers and seamlessly integrated in teaching practices as a way to facilitate learning.

In this paper, we report about a project in which we experimented the use of an educational chatbot in primary school inspired by two main goals: on the one side, we wanted to involve teachers and educators in the creation of the chatbot, asking them to co-design both the character and the content to be delivered; on the other side, we wanted to test the use of dramatization techniques for the design of the chatbot. To do so, we relied on artificial intelligence technologies that allow the chatbot to be entirely scripted by the developers, so as to avoid the pitfalls of generative models in factual knowledge and to control closely the character creation.

¹st Workshop on Education for Artificial Intelligence (edu4AI 2024, https://edu4ai.di.unito.it/), co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024). 26-28 November 2024, Bolzano, Italy *Corresponding author.

[†]These authors contributed equally.

The chatbot design and the experiments were carried out as part of the project "Chi decide per noi? Percorsi tra Storia ed Educazione Civica dall'università alla scuola primaria" ("Who decides for us? Pathways between History and Civic Education from university to primary school") at the University of Turin. This project aimed to integrate the Civic Education programme of Italian primary school with education about political representation and participation from the perspective of their historical development. The project involved the University of Turin and a group of primary schools (grades IV and V) and relied on the use of traditional, media and digital teaching tools to convey the educational contents.

The involvement of teachers in the design of the chatbot had a twofold objective. Firstly, the participatory design methodology chosen for this project required teachers as the main users of the final product. Secondly, the design process allowed us to give the teachers a role as active creators of the interaction script between the chatbot and the users. The design of a chatbot serves as an excellent way to understand how generative AI systems work. Chatbots, especially those powered by AI like GPT models, simulate human-like conversation by generating text based on input. The design principles behind chatbots illustrate the broader mechanics of how generative AI systems operate in tasks like content creation, problem-solving, and personalized interactions. This allowed us to introduce ai-based tools to the teachers in a conscious and participative manner. This aspect is very important especially in the context of the Italian school. Indeed despite many contributions made by various organisations, the level of digital preparation of the Italian teachers is still limited [4, 5]. The single-cycle study courses in primary education do not include a syllabus that is attentive to new literacies, so in the laboratories of educational and learning technologies or of teaching technologies, the most varied syllabuses are to be found, ranging from teaching technical and procedural skills to coding, media education and teaching with technologies. The result is that even the new generations of teachers risk remaining substantially unprepared especially concerning new tools based on generative AI. Involving teachers in the design of the tools they will use in the classroom is a highly effective strategy for several reasons. This approach not only empowers educators but also ensures that the tools developed are truly aligned with their needs and classroom realities.

The paper is organized as follows: after reviewing the related work in Section 2, we describe the design and creation of the chatbot in Section 3. Evaluation is illustrated and discussed in Section 4. Conclusion and future work end the paper.

2. Related work

[1] presents a systematic review of studies on the use of chatbots in education. The authors discussed the benefits obtained from the applications of chatbots in the educational domain: integration of contents, quick access, motivation, and engagement and immediate assistance. [2] analyzed 36 educational chatbots proposed in the literature by assessing each chatbot within seven dimensions: educational field, platform, educational role, interaction style, design principles, empirical principles, and challenges as well as limitations. The results show that there are almost no chatbot proposed for history educational topics. [6] presents the design of a "social bots of conviction" (BoCs) which shift the focus from offering information to provoking reflection. The BoC is designed as a digital experience to support history education for high school students (ages 14-18). The findings highlight the BoC's role in engaging the students in constructive dialogue with each other; and the ways in which it guided perspective taking and collective reflection about the past.

3. Chatbot design and implementation

The project "Chi decide per noi?" included different phases: the first was the lesson planning by the group of teachers, which included primary school teachers and university staff; the second consisted of participatory lessons in class on the historical evolution of the concepts and practices of participation and representation; the third consisted of the elaboration of laboratory experiences on the topics

covered in the lessons, which included the production of textual products (text and images), videos (stop-motion documentary), and of the chatbot. Funded by Fondazione CRT, the project was carried out from September 2022 to August 2024.

3.1. Technology and method

In order to avoid the well-know pitfalls of state-of-the-art chatbots (see [7] for a review and reflection on the motivations behind the so-called hallucinations), we decided to resort to the Artificial Intelligence Markup Language (AIML) for implementing the chatbot, a well-established technology for the creation of conversational agents.¹ AIML (Artificial Intelligence Markup Language) was developed by Richard Wallace since 1995 as a language for creating conversational agents or chatbots, notably used in the chatbot "A.L.I.C.E." (Artificial Linguistic Internet Computer Entity). It allows developers to define rules and patterns for chatbot responses through simple XML-based tags. AIML became widely adopted for building rule-based chatbots, leading to many open-source implementations and is still in use, with a community of developers working at the creation of AIML rule engines for most platforms and languages. Basically, an AIML program consists of a set of IF-THEN rules, called categories, which specify what the bot should answer in response to each type of user input. The antecedent of the rule, called pattern, describes the linguistic contribution of the user; the consequent, called templates. In practice, each category describes a pair of conversational turns composed of a user's contribution (a question, an assertion, etc.) followed by the chatbot's reply. For example, the consider the following listing, which illustrates the basic structure of the AIML <category tag:

```
1 <category>
    <pattern>Mi chiamo *</pattern>
2
     <template>
3
         Ciao, <set name = "username"><star /></set>, anzi ave <set name = "username">
4
         <star /></set>!, come diciamo noi Romani.
5
6
         Oggi vorrei parlarti di come i cittadini
         prendevano parte alla vita politica nell'Antica Roma.
7
8
         Cosa ti piacerebbe sapere?
    </template>
9
10 </category>
```

Lines 2 contains the pattern, which matches any sentences beginning with "Mi chiamo" ("My name is" in English) followed by any string. Lines 3 to 9 contain the template for the generation of the bot's turn, where a combination of the <set> and the <star> tags inserts in the bot's sentence the proper name extracted from the input.

Although the AIML programmer seeks to cover all possible user's intents in the given domain, using regular expressions to improve the patterns' capability to generalize, the coverage of an AIML script in most cases will still be limited, with the bot providing a generic reply when the user's input cannot be matched against any patterns. For this reason, for the chatbot to be able to interact in a fluent and natural way with the users, some techniques can be adopted to minimize unpredictability. In general, the more constraining are the chatbot's turns in terms of possible users answers, the more predictable and easier to handle will be the sentences produced by the user. If it is coherent with the context of the interaction, the chatbot's turns may also include multimedia elements such as images and animations to clarify better the chatbot's capability to deal with it. For example, the following template creates the menu displayed in Figure 1

```
1 <template>
2 Preferisci un nome femminile o maschile?
3 <button>
4 <text>Femminile</text>
5 <postback>femminile</postback>
6 </button>
7 <button>
8 <text>Maschile</text>
```

¹http://www.aiml.foundation/

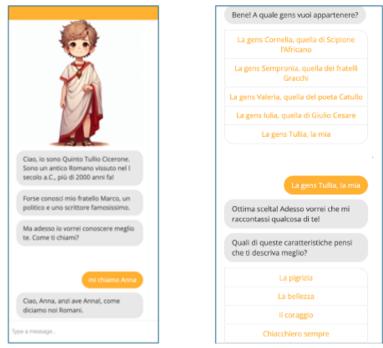


Figure 1: Screenshot of the chat showing the appearance of the character, a Roman child called Quintus Tullius Cicero (left) and the use of buttons to acquire input (right). Chatbot's turns are in grey, user's in yellow.

9 <postback>maschile</postback>
10 </button>
11 </template>

In the listing above, the <button> tags (lines 3 and 7) create two buttons with text "Femminile" ("Female") and "Maschile" ("Male") respectively. Each button, when selected, generates the text value set by the cpostback> button that follows it; this text, not visible by the user, is fed to the rule engine, and will eventually match the pattern contained in one or more categories, making the conversation advance.

Secondly, the conceptual framework of dramatization can offer a way to deal with the rigidity required in the creation of the chatbot by the aforementioned need to constrain the user input. A well-structured, consistent character, in fact, is not only more effective in bringing the user to behave in a certain way, but is at the same time engaging at the social and emotional level. Dramatization, tested among others by [8, 9, 10], relies on a set of techniques to create believable characters which can emotionally engage the user. To do so, the character should be able to communicate a clear identity and personality - created and sustained by providing information about its background, also in an anecdotal way –, display social capabilities – for example by greeting and questioning the user – and possibly declare explicit goals, better if not directly reachable without some kind of conflict.

3.2. Design and development

The design of the character was carried out by a multidisciplinary team formed by primary school teachers, Roman history experts, HCI and digital education experts and developers. The character of Roman child, younger brother of the famous Roman orator Cicero, was selected due to its capability to mediate between the students and the historical contents, being a fictional young person students can identify with. After introducing himself and greeting the user, the character expresses his desire to tell about himself and the life in the ancient Rome, teaching basic notions (e.g., Latin greetings) with a peer-to-peer attitude, "sending" images and short animations from time to time. For example, he greets the user with the Latin expression "Ave" and immediately after briefly explains the meaning of the word "ave" in very simple terms ("It is a word that comes from the imperative of the verb to have, in

the sense of being well. When I say "Ave" to you, I am wishing you to be well"). To describe himself, the character discloses information about himself in a casual way, for example by mentioning his well known elder brother Cicero ("You may know my brother Marco, a very famous politician and writer"). While offering to talk about the expansion of Rome across the centuries and the evolution of its political system ("I would like to talk you about politicians in the ancient Rome"), the chatbot leaves the choice of the specific argument to the user by showing a list of buttons. The latter choice, apart from the need to constrain user input to manageable alternatives, is also functional to the interaction with groups of children, since it drastically simplifies the negotiation between the children about the choice by providing a limited, easily identified set of keywords. Finally, the character, like every child, wishes to play with his peers, so he engages in various games, making riddles that give him the opportunity to convey notions about ancient Rome and other small games. For example, after talking about Roman proper names, he proposed the user to invent their fantasy Roman name by choosing a personal name, a surname, and the name of a gens (a Roman clan) (see Figure 1, right).

After discussing the content to be conveyed as part of the educational goals of the project, the character creation methodology was shared with the team in a seminar delivered by the HCI and AI experts. Then, one of the teachers was commissioned to write the interaction between the chatbot and a sample user a script. The task explicitly included the request to break down the interaction into different phases, each characterized by a set of alternatives that would become user choice in the final program, and to insert graphic elements in the chatbot's turns. The script, revised by the historians, was handed to the developer for implementation. In the end, the chatbot consisted of 75 categories that covered four main topics: getting to know each other, Latin language, the expansion and form of governement of Rome, political life and career.

The chatbot, shared with the team on Github, was debugged and finally tested in an experimental session with a primary school class (5th grade) in April 2024 as part of a program run by the University of Turin ("A day at the university") in which primary and secondary school classes visit the university departments and attend educational labs organized by the staff.² The test was aimed at investigating two main aspects: on the one side, we wanted to observe how students reacted to the chatbot, and collect information about how to improve it from their point of view; on the other side, we wanted to collect the opinion of the teachers about the possible use of the chatbot in the classroom and its possible function as a teaching tool.

The presentation of the chatbot was done with the experimenter interacting with the chatbot (Quinto Tullio Cicerone) on the screen, reading the answers out loud. We used the simple test interface of the Pandorabots development environment that has a smartphone-like chat box on one side of the screen. After a while, it was evident that the simplest way to manage the interaction was to leave to the class the decision about what to write in the chat box or what option to select in case of multiple choices.

At the end of the presentation, we conducted a focus group, asking students and teachers for their opinion with some guiding questions: what do you think about the chatbot, how would you like it, what would you like it to talk about. Two main requests emerged from the discussion with the students and teachers. They took part in the activity with enthusiasm and generally liked the chatbot, but suggested that it should talk more about interesting topics for children, such as habits, daily life (clothing, transport, houses), and geography. Also, they suggested that the chatbot should be characterized and have an appearance of somebody they could relate to, namely a boy or a girl.

Concerning the use of the chatbot in the classroom, teachers were positive about this possibility, justifying their opinion on the basis of the fact that they already used multimedia material during lessons. In particular, we asked them if they thought it was difficult to manage interactivity, but they gave us a positive opinion on this aspect too. Furthermore, they indicated a series of elements that they believe are important for engaging children based on their experience with educational videos (which they often find on online platforms Italian accompanying textbooks). According to them, a major factor of engagement is the use of characters who speak in first person (versus a third person narrator); humorous elements in the narration and dialogues with other characters (e.g., when characters interact

 $^{^{2}} https://www.unito.it/ateneo/gli-speciali/bambine-e-bambini-un-giorno-alluniversita_{1} and 1 an$

in a video) also emerged as capable of engaging the students. The preference for a character who speaks in the first person was also confirmed by the children in reference to a video from the same project about Marco Polo ("a positive aspect is that he speaks"). Finally, we discussed the interaction mode with the teachers: in their opinion it would be appropriate for the chatbot to speak (rather than respond only in writing) but they believe that the current input mode (text and buttons) is preferable to spoken input in the classroom context, because it allowed for a more ordered interaction. Finally, they were not favorable to individual interactions with the chatbot on personal devices in the classroom, considered the group mode more manageable.

Consequently to the observations made during the first test, a few redesign elements were put into practice. Texts were shortened, especially given the limited space allowed for by the phone (or phone-like) interface. Although it was possible to keep the original texts by using a full screen interface, we observed that the chat provided by the environment employed for the test (visible in Figure 1), which is similar to that of a standard smartphone, introduced a very powerful element of credibility to the interaction with the chatbot, instilling the feeling of chatting with a person from the past, distant in space and time but co-present in the chat space. Also, to balance the weight of the text, we decided to insert more images and animations. Finally, in order to make the activity more compelling, we decided to introduce in the script the "name game" mentioned in the previous section.

4. Testing the chatbot

We carried out a series of educational workshops in primary schools with the aim of validating the chatbot. In particular, we were interested in i) understanding whether the users (i.e., primary school students) had a positive experience while playing our serious game in the context of an in-class educational workshop, since this is a prerequisite for its adoption and ii) assessing the learning outcomes, iii) to obtain a general assessment of our approach and web platform, together with some suggestions from the teachers involved.

4.1. Recruitment

The classes involved in the experiment were selected as part of the 'Chi decide per me' project through the University of Turin's educational activities.

We established a formal agreement with all those schools; we distributed the informed consent to the parents of all children possibly involved in our experiment and asked them for the signed privacy consent form. Only those children whose parents had signed the consent form were allowed to participate in our experiment. Overall, 5 V grade classes were involved in our study.

4.2. Activities and materials

All classes involved had some classroom lectures on history and politics in ancient Rome in April 2024 and worked for the same number of hours and on the same topics.

Afterwards, in May 2024 the classes were divided into 2 groups: control classes and experimental classes. The activities were organized as follows. *Experimental classes*: 3 classes did a revision activity on the different habits, daily life and geography in the ancient Rome using the chatbot. The process begun with the researcher introducing the chatbot and explaining its role as a source of information. Each time the chatbot prompted a response or question, the students paused to decide collaboratively on the next question they would like to ask. With the assistance of a class teacher, the researcher encouraged active participation by ensuring that every child has the chance to contribute to the conversation, promoting teamwork and consensus-building. After discussing the possible questions, the students voted on a final choice. The selected question was then submitted to the chatbot, and the class waited for its response. Once the chatbot provided an answer, the students analyzed the information together, and the cycle repeated, with the class deciding the next step based on the chatbot's answers.

Statement 1	If I had this chatbot on my iPad, I think that I would like to play it a lot	
Statement 2	I was confused many times when I was playing	
Statement 3	I thought the chatbot was easy to use	
Statement 4	I would need help from an adult to continue to use the chatbot	
Statement 5	I always felt like I knew what to do next when I played with the chatbot	
Statement 6	Some of the things I had to do when using the chatbot did not make sense	
Statement 7	I think most of my friends could learn to use the chatbot very quickly	
Statement 8	8 Some of the things I had to do to use the chatbot were kind of weird	
Statement 9	atement 9 I was confident when I was using the chatbot	
Statement 10	10 I had to learn a lot of things before using the chatbot well	
Statement 11	I really enjoyed using the chatbot	
Statement 12	I liked the images I saw	
Statement 13	If we had more time, I would keep using the chatbot	
Statement 14	I plan on telling my friends about the chatbot	

Table 1

Statement of the SUS questionnaire adapted for 9-11 children.

Control classes: children reviewed history topics by reading fictional interviews with characters from Ancient Rome. The students were divided into small groups, and each group was assigned two or three interviews featuring a different Roman figures. The interviews covered various historical themes and matched the chatbot interaction scripts.

4.3. Measures

At the end of the activity we administered questionnaires to each participant. In this process, the informed consent form was presented to the children in a clear and age-appropriate manner. The researcher began by explaining what informed consent means, emphasizing that it is a way for them and their parents to understand the study and decide whether they want to participate. The explanation included a discussion of the children's rights, such as the option to withdraw from the research at any time without any consequences. The implications of their participation were outlined, including what the study would involve, how their information would be used, and the confidentiality of their responses. The children were encouraged to ask questions to ensure they fully understood everything. After the explanation, the informed consent forms were distributed. Finally, two questionnaires were handed out to those who had provided consent, and they were asked to fill it out:

- To measure the chatbot usability, as well as users' experience and engagement with the chatbot we used that adaptation of System Usability Scales (SUS) for children between ages of 7-11 [11]. We added an item (see item 12 in Table1). This test was administered only to the classes belonging to experimental group. Users are asked to assess their level of agreement with each item using a five-point Likert scale, ranging from "strongly disagree" to "strongly agree" (see Table1).
- We assessed knowledge and skills using a a multiple choice test created by the teachers on the most important concepts covered in the course³.

The study was approved by the University of Turin Ethics Committee.

4.4. Results

In this section, we illustrate the data collected about the use of digital media by the students involved in the workshops, the results of the adapted SUS administered to the test group to evaluate the usability of the chatbot, and the results of the multiple choice test that all students completed at the end of each activity (both test and control groups).

³Linktothetesthttps://acesse.one/historyquestionnaire

computer	46	55,42%
tablet	46	55,42%
smartphone	53	63,86%
console	31	37,35%
other	0	0,00%
>1 hour	34	40,96%
<1 hours	22	26,51%
few times/week	24	28,92%
few times/month	2	2,41%
never	1	1,20%
	tablet smartphone console other >1 hour <1 hours few times/week few times/month	tablet46smartphone53console31other0>1 hour34<1 hours22few times/week24few times/month2

social		76	91,57%
	facebook	7	8,43%
	tiktok	24	28,92%
	twitch	4	4,82%
	roblox	37	44,58%
tumo	discord	3	3,61%
type	instagram	14	16,87%
	twitter	1	1,20%
	youtube	76	91,57%
	pinterest	21	25,30%
	altro	0	0,00%
account		48	57,83%
	whatsapp	59	71,08%
	telegram	6	7,23%
chat tuno	messenger	7	8,43%
chat type	other chats	5	6,02%
	no chat	18	21,69%
	other	0	0,00%
chatbot		25	30,12%

Table 2	
Results of the questionnaire about the use of devices, internet, social networks, and chats (all participants, n=83).	

Table 2 reports the results of the questionnaire about the use of digital media. Concerning the use of devices, the results confirm the familiarity of the sample students with the most widespread devices: 55.42% of the participants use a computer, and 55.42% of the participants use a tablet; the smartphone appears to be the most used (63.86%), while only 37,35% of the participants use a game console. Less than half of the participants (40.96%) use the internet more than one hour a day; 26.51% of the participants declare to use the internet less than one hour a day; 28.92% of the participants use the internet few times a week; many fewer participants (2,41%) use the internet only a few times a month; finally, only 1 participant declare to never use the internet. Concerning the use of social networks, more than half of the participants (57.83%) have their own personal account. The large majority (91.57%), however, use social networks. The most used social network is YouTube (91.57%), followed by Roblox (44.58%), TikTok (28.92%), Pininterest (25,30%), Instagram (16.18%), Facebook (8.43%), Twitch (4.82%), Discord (3.61%), and Twitter (1.20%). Concerning the use of chats, that we wanted to investigate in order to assess the familiarity of the participants with this interaction mode, the large majority of the participants (71.08%) use Whatsapp, followed by Messenger (8.43%) and Telegram (7.23%), and other chat types (6.02%). A a small but significant minority (21,69%) don't use any chat. Finally, only 30 participants (30.12%) declare they have interacted with chatbots before. These results show that, even if some individual differences, the students in both groups have a good familiarity with digital devices and social media, and that they regularly use chats.

Table 4.4 (a) illustrates the results of the adapted System Usability Scale (SUS) administered to the participants in the test group (n=53). As it can be observed, the overall rating (77.5) reflects the students' enjoyment of the experience observed by the experimenter, and the standard deviation (8.71) does not suggest meaningful differences between the participants. Positive statements consistently received a rating about 2.5%, and negative statements received a rating below 2.5%. Concerning the single ratings, the highest agreement is about statement 3 ("I found the chatbot simple to use"). This result, together with the rating of statement 9 ("I felt confident while we used the chatbot"), seem to confirm the perceived ease of use by the participants. However, since the use of the chatbot occurred in a group context, with the experimenter executing the actions (or writing the sentences) suggested by the group, we think that these results are more suggestive than certain.

Concerning the four statements tailored to the reception of the experience with the chatbot (integrated from the statements 11 to 13 by [11]: 11 "I really enjoyed the chatbot", 12 "I liked the images I saw", 13 "I would have continued using the chatbot if we still had time", 14 "I will tell my friends to use

[statement	avg. rating	st. dev.
-	1	3,63	0,87
	2	2,09	1,01
	3	4,2	0,67
	4	1,94	0,77
(a)	5	4,04	0,84
(a)	6	2,35	0,93
-	7	3,85	0,97
	8	2,2	1,01
	9	3,87	0,99
	10	1,87	0,69
	overall rating	77,5	8,71

	statement	avg. rating	st. dev.
(b)	11	4,04	1,29
	12	4,19	1,04
	13	3,81	1,02
	14	3,57	0,09

Table 3

Usability score (test group, n=53). Table (a) reports the scores of the standard SUS scale adapted for children. Table (b) reports the scores of the statements 11 to 14 added to the standard SUS by [11].

the chatbot) the ratings are reported in Table 4.4 (b). All statements received positive ratings, thus confirming the capability of the chatbot to engage the users with its behavior and appearance. Ratings, in fact, ranged from 3.57 (statement 14) and 4.19 (statement 12), showing the students' willingness to repeat the experience.

Finally, in order to test the role of the chatbot in teaching, namely its capability to help the students to learn the notions provided by the chatbot during the interaction with the students, we asked the participants from both groups to take a quiz about Roman history (structured as a multiple choice test), as illustrated in the previous section. We collected 50 valid copies of the quiz (3 participants dropped off) for the test groups and 30 copied for the control group. The average score for the test group was 55.72% of correct answers; the average score for the control group was 52.65% of correct answers. The two groups were compared with a non-parametric test (Mann Whitney U test) but no statistically significant difference was found.

We computed the overall score obtained by each participant from the questionnaire on the use of digital media by summing the declared number of devices, accounts and chat systems, and we computed the correlation between the obtained score and the SUS score using Spearman's Rho. However, no statistically significant correlation was found. This result is not surprising given two factors: on the one side, the SUS scores do not have a large standard deviation, showing small differences in the individual reception of the chatbot; on the other side, the large majority of the participants appeared to be very familiar with digital media, so we didn't expect this aspect to affect the reception of the chatbot.

5. Conclusion and Future Work

The impact on the chatbot activity with children was really effective: the collaborative approach in deciding the question to prompt fosters critical thinking, active listening, and decision-making skills. It also helps students understanding how to gather and evaluate information in a group setting through an engaging experience that was positively evaluated by the participants. Throughout the activity, the teacher serves as a facilitator, guiding discussions and ensuring that all students are engaged in the process. The impact on teachers was really positive:

- When teachers are involved in the design process, they feel a sense of ownership over the technology. This engagement helps them overcome potential fears or resistance to adopting new digital tools. By contributing their insights and ideas, they build confidence in their ability to use these technologies effectively, even if they start with limited digital skills.
- Teachers know their classroom environments, students' needs, and curricular goals better than anyone else. By involving them in the design, the tools can be tailored to fit real-life teaching scenarios. This reduces the risk of creating technology that is impractical or difficult for teachers

to integrate into their daily lessons, leading to higher adoption rates and more effective use of the tools.

- Participation in the design process naturally encourages teachers to learn more about the digital tools they will use. Through hands-on involvement, teachers can build their technical skills in a supportive, low-pressure environment. This gradual learning helps them feel more comfortable and competent with AI based technology.
- When tools are co-designed with input from teachers, they are more likely to be sustainable and effective in the long term. Teachers are more inclined to continue using tools that they helped shape, ensuring that the investment in technology leads to meaningful, ongoing improvements in the classroom.

As future work, we intend to replicate the methodological framework and technology employed to create the chatbot, extending it to the other historical and political contexts explored in the overarching project. To do so, we intend to co-create with the teachers similar characters that are able to engage positively with the students.

References

- C. W. Okonkwo, A. Ade-Ibijola, Chatbots applications in education: A systematic review, Computers and Education: Artificial Intelligence 2 (2021) 100033. URL: https://www.sciencedirect.com/ science/article/pii/S2666920X21000278. doi:https://doi.org/10.1016/j.caeai.2021.100033.
- [2] M. A. Kuhail, N. Alturki, S. Alramlawi, K. Alhejori, Interacting with educational chatbots: A systematic review, Education and Information Technologies 28 (2022) 973–1018. URL: https: //doi.org/10.1007/s10639-022-11177-3. doi:10.1007/s10639-022-11177-3.
- [3] D. Z. Ruofei Zhang, G. Cheng, A review of chatbot-assisted learning: pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions, Interactive Learning Environments 0 (2023) 1–29. URL: https://doi.org/10.1080/10494820.2023.2202704. doi:10. 1080/10494820.2023.2202704. arXiv:https://doi.org/10.1080/10494820.2023.2202704.
- [4] C. G. Stephanie, N. Joanna, B. Adonis, M. Eve, P. Agnieszka, R. Maria, T. Karen, L. Koen, R. B. Nicolas, M. Marco, G. V. Ignacio, What did we learn from schooling practices during the COVID-19 lockdown?, Scientific analysis or review KJ-NA-30559-EN-N (online), Luxembourg (Luxembourg), 2021. doi:10.2760/135208(online).
- [5] M. Ranieri, Le competenze digitali degli insegnanti, 2022, pp. 49–60. doi:10.36253/ 978-88-5518-587-5.6.
- [6] D. Petousi, A. Katifori, S. McKinney, S. Perry, M. Roussou, Y. Ioannidis, Social bots of conviction as dialogue facilitators for history education: Promoting historical empathy in teens through dialogue, in: Proceedings of the 20th Annual ACM Interaction Design and Children Conference, IDC '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 326–337. URL: https://doi.org/10.1145/3459990.3460710. doi:10.1145/3459990.3460710.
- [7] M. T. Hicks, J. Humphries, J. Slater, Chatgpt is bullshit, Ethics and Information Technology 26 (2024) 38.
- [8] R. Damiano, C. Gena, V. Lombardo, F. Nunnari, A. Pizzo, A stroll with carletto: adaptation in drama-based tours with virtual characters, User Modeling and User-Adapted Interaction 18 (2008) 417–453.
- [9] E. Borini, R. Damiano, V. Lombardo, A. Pizzo, Dramasearch. character-mediated search in cultural heritage, in: 2009 2nd Conference on Human System Interactions, IEEE, 2009, pp. 554–561.
- [10] A. Pizzo, V. Lombardo, R. Damiano, Interactive storytelling: a cross-media approach to writing, producing and editing with AI, Taylor & Francis, 2023.
- [11] C. Putnam, M. Puthenmadom, M. A. Cuerdo, W. Wang, N. Paul, Adaptation of the system usability scale for user testing with children, in: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–7. URL: https://doi.org/10.1145/3334480.3382840. doi:10.1145/3334480.3382840.