A Systematic Identification of Pedagogical Conversational Agents

Leo Natan Paschoal*, Aliane Loureiro Krassmann†, Felipe Becker Nunes‡, Myke Morais de Oliveira*, Magda Bercht†, Ellen Francine Barbosa* and Simone do Rocio Senger de Souza*
*University of São Paulo (ICMC-USP), São Carlos, SP – Brazil
†Federal University of Rio Grande do Sul (CINTED - UFRGS), Porto Alegre, RS - Brazil
‡Antonio Meneghetti Faculty (AMF), Santa Maria, RS - Brazil
E-mail: paschoalln@usp.br, alkrassmann@gmail.com, nunesfb@gmail.com, mykeoliveira@usp.br, bercht@inf.ufrgs.br, francine@icmc.usp.br, srocio@icmc.usp.br

***Abstract*—Conversational agents have been established to support educational practices, solving students' questions, recommending teaching materials, increasing the motivation of students who have little interest in certain content, among others. Although many studies are discussing the establishing of conversational agents, there is no global understanding of what has been investigated in this area. There are some previous secondary studies that have attempted to map the conversational agents used for educational purposes (i.e., pedagogical conversational agents). However, these studies are limited to include primary studies published in specific sets of conferences and journals. Therefore, an overview of the current state of the art associated with pedagogical conversational agents has not been produced yet. This research paper aims to contribute to the theme, through the presentation of planning, conduction, and results analysis of a systematic mapping on pedagogical conversational agents. We address the following questions: (i) in which areas have conversation agents been investigated?; (ii) at which education levels are conversational agents used?; (iii) which mechanisms do conversational agents use to interact with students?. We also present some discussions about research opportunities that can be explored in future works. This study aims to contribute to the topic of pedagogical conversational agents, helping researchers to have a deeper understanding of what has already been done and to guide their research to subjects not explored yet.**

***Index Terms*—Chatbot, Chatterbot, Pedagogical Agent.**

I. Introduction

Computer-based applications that make use of Artificial Intelligence (AI) are becoming increasingly common in today's society. Humans interact almost daily with some kind of AI-based application, from a simple text editor that checks for verbal concordance errors, to a vacuum cleaner that automatically recognizes the environment in which it is inserted [1]. A significant amount of software and hardware systems have been produced with some algorithm or approach that originated from the AI. In particular, one type of AI application that is gaining more and more space in the community and industry is conversational agents, also known as chatbots or dialogue systems.

Conversational agents are software systems designed to interact with humans or other systems in some natural language [2] by simulating human interactions [3]. It uses strategies to process natural language and generate outputs in the same format, such as text or voice messages [4]. Siri[1], Cortana[2], and Google Assistant[3] are some examples of conversational agents. In particular, they are capable of recognizing both voice and text input and they are available on iOS, Windows, and Android operating systems.

Applications of conversational agents have been explored in various types of configurations and environments, offering different types of services in different places of society. These applications are designed to perform business services [5], for entertainment [6], to support language learning [7], among others.

The use of conversational agents to support teaching and learning is another domain that gains more and more space. They are employed to meet different user profiles, different themes, and teaching modalities [8]. Conversational agents can be designed to contribute in different ways in learning environments [9]. Examples of the use of conversational agents to solve student questions [10] or to keep students attentive relevant information [11] are found. Other examples of use include improving the student motivation [11], offering behaviors that positively impact in the students' emotional states [12], guide students in problem-solving [13] and support the fresh students in a second language [7].

Conversational agents can be used in face-to-face education practices, in distance education to support learning environments, and in models that mix face-to-face and distance moments (i.e., blended learning) [8]. In the study by Paschoal et al. [9] for example, the authors address the possibility of using a conversational agent in the context of the Flipped Classroom pedagogical model [14], since, in this model, students tend to have greater difficulty in preparing themselves for classes. Thus, the conversational agent can solve dubious concepts and misinterpretations made by students at the time of study a priori to the class.

Given the different educational configurations that allow conversational agents to be incorporated as teaching support mechanisms, it becomes relevant to map and organize the

[1]More information available at: <https://www.apple.com/siri>.
[2]More information available at: <https://www.microsoft.com/en-us/cortana>.
[3]More information available at: <https://assistant.google.com/>.

knowledge that has already been produced on the theme. Previous studies were carried out to map the state of the art on pedagogical conversational agents [15]–[20]. However, these studies have some limitations.

Some of these studies have limitations associated with the sources that were used in the search for primary studies. In the study by Smutny and Schreiberova [21], for example, the authors sought to understand the conversational agents, considering those that were mapped by Botlist [4], which included only gray literature. In the study by Kuyven et al. [17], the authors defined some conferences and journals involving educational technology to conduct the mapping.

Another three previous studies were carried out to map the conversational agents. The study by Io and Lee [18] presents only data associated with the number of documents about conversational agents in the Web of Science and the number of citations of these studies. Similarly, Bernardini et al. [19] conducted a study that considers only documents indexed by Scopus. As studies are mentioning that pedagogical conversational agents are a theme studied by researchers from different areas (e.g., linguistic, mathematics, computing, health, among others) and specific topics in these areas (e.g., English), it is important to consider a wider variety of search sources. Dyba et al. [22], Kitchenham and Charters [23], Petersen et al. [24] recommend some search bases that index scientific studies from different areas of human knowledge and that have been referenced in studies that use systematic literature mapping as a method.

The study by Winkler and Söllner [15] considered some of the databases indicated by Dyba et al. [22], Kitchenham, and Charters [23], Petersen et al. [24]. There is also the definition of research questions. Despite this, Winkler and Söllner [15] do not disclose how many studies are used to answer each research question. Furthermore, it is not clear whether all the studies analyzed were used to answer the research questions, given that the authors choose to only describe qualitatively what they found with their studies.

Some studies explore conversational agents defined for a specific domain or specific characteristics of pedagogical conversational agents. In Montenegro et al. [25], the study mapped conversational agents specific to the health area, including health education. Hobert et al. [16] have focused on mapping the evaluation methods used in the context of educational conversational agents. Roos [20] analyzed just studies on pedagogical conversational agents based on the AIML language (Artificial Intelligence Markup Language).

Based on this perspective, this paper provides an overview of pedagogical conversational agents. Different from previous literature review work, this paper presents an embracing view about how conversational agents are employed to support teaching and learning. We believe that information about areas where these agents are applied, strategies used for agents communication, among others, can be useful to other researchers interested in implementing or in researching pedagogical conversational agents. Therefore, the mapping study conducted in this work addresses the following research questions:

(i) In which areas have conversation agents been investigated?;
(ii) At which education levels are conversational agents used?;
(iii) Which mechanisms do conversational agents use to interact with students?

This paper is organized as follows. Section II describes the method used to identify primary studies that describe pedagogical conversational agents. The results obtained through the methodological procedures are presented in Section III. In Section IV, we discuss the results and the threats to the study's validity and how they were mitigated. In Section V we present some limitations of the study. Finally, in Section VI we present our conclusions and perspectives for future research.

II. Systematic Mapping Process

The conduction of this Systematic Mapping Study (SMS) is based on a predefined protocol that considers the guidelines defined by Petersen et al. [26] and Petersen et al. [24]. Therefore, the SMS was divided into five stages: (i) definition of research questions, (ii) search for primary studies, (iii) screening of studies for inclusion and exclusion, (iv) keywording of abstracts, and (v) data extraction and the mapped studies. The mapping contemplated a delimited period, between 2007 and 2017.

A. Definition of research questions

For the definition of research questions, the Goal Question Metric (GQM) approach [27] was adopted, defining a set of metrics to validate the answers. In this study, the goal was to provide the state of the art on pedagogical conversational agents, focusing on: (i) which areas they were investigated; (ii) at what levels of education they have been used; (iii) what mechanisms they use to interact with students. Based on this goal, three research questions (RQ) and related metrics were established, as presented in Table I.

TABLE I
Research questions and metrics following the GQM approach

	Research questions	Metrics
RQ1	In which areas have conversation agents been investigated?	(1) Knowledge areas that have been identified in primary studies; (2) Number of occurrences of each knowledge area.
RQ2	At which education levels are conversational agents used?	(1) Educational levels of formal education that have been identified in the primary studies; (2) Number of occurrences of each educational level.
RQ3	Which mechanisms do conversational agents use to interact with students?	(1) Communication strategies have been used by conversational agents to communicate with the student (e.g., text, voice). (2) Number of occurrences of each communication strategy.

[4]More information available at: <botlist.co>.

Besides these RQs, we have also defined a set of additional issues that are associated with technical characteristics of the development of conversational agents. These questions were defined in order to provide an overview of the state of development of pedagogical conversational agents. This information becomes relevant as new knowledge-based models have emerged and more traditional models have been used less and less [28]. With this panorama, for example, it is possible to direct researchers to use certain models of knowledge representation and associated technologies. The questions are listed next:

- What design techniques are used in the development of pedagogical conversational agents?
- What programming languages are used in the development of pedagogical conversational agents?
- How have the conversational agents been graphically represented?

B. Search procedures

The six databases that were selected are presented in Table II. According to Dyba et al. [22] and Kitchenham and Charters [23], these databases are the most relevant sources in the Computer Science area. Additionally, the Engineering Village database was adopted based on the systematic mapping successful experience in the software engineering area reported by Petersen et al. [24].

TABLE II
DATABASES USED IN THE SEARCH

Database	Source
ACM Digital Library	http://portal.acm.org
Elsevier (Science Direct)	http://www.sciencedirect.com
Engineering Village	https://www.engineeringvillage.com/search
IEEE Xplore	http://ieeexplore.ieee.org
Scopus	http://www.scopus.com
Web of Science	https://webofknowledge.com

The search string was defined so that it could contemplate a big range of variations of the terms "conversational agent" and "education" existing in the literature. Therefore, the search string was built around the union of these keywords, with their spelling variants and associated terms. Figure 1 illustrates the variations of the terms that were used to define the search string.

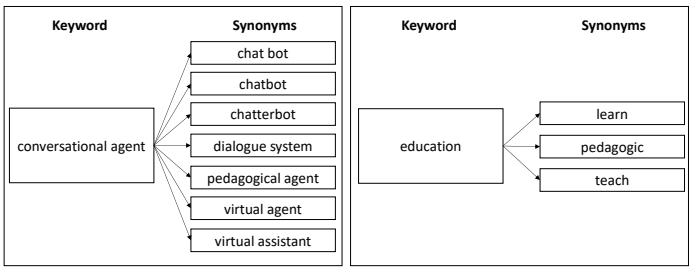

Fig. 1. Search string terms

The search string was created based on terms of Figure 1:

```
((''conversational agent'' OR chatbot OR
''chat bot'' OR chatterbot OR ''dialogue
system'' OR ''pedagogical agent'' OR
''virtual agent'' OR ''virtual
assistant'') AND (education OR learn OR
pedagogic OR teach)).
```

In order to select the studies, factors to identify the relevance of the primary studies in answering the research questions were defined as well as inclusion and exclusion criteria (Table III). These criteria were used as a parameter to assist the authors during the analysis of the primary studies, making it more objective, reliable and accurate.

TABLE III
INCLUSION AND EXCLUSION CRITERIA

Inclusion Criteria (IC)	Exclusion Criteria (EC)
IC–1: The primary study presents at least one conversational agent to be used in the educational context.	EC–1: The study is a technical report, a document, which is available in summary format, is a presentation, a call for papers or a summary of a conference. EC–2: The primary study is written in a language other than English. EC–3: The full text of the primary study is not available in the database. EC–4: The primary study does not meet the inclusion criteria.

C. Review execution

The process of selection of the primary studies was carried out following some stages:

- *1st stage:* Initially, the search was conducted in each database, resulting in a total of 5.075 primary studies. The metadata of these primary studies, containing title, summary, article URL, keywords, and DOI number (Digital Object Identifier) were exported from these bases in a BIB file format. Subsequently, it was imported into the Parsifal tool[5], which supports the automatization of activities while conducting systematic mappings and/or reviews of literature.
- *2nd stage:* In this phase, duplicated studies were removed, with the support of the Parsifal tool. After that, it has remained 2.619 primary studies.
- *3rd stage:* The remaining 2.619 primary studies were randomly distributed among three reviewers. In this initial analysis, they read the title, abstracts and keywords and applied the inclusion and exclusion criteria (first round). At the end of this phase, the number of primary studies was reduced to 548.
- *4th stage:* The reviewers carefully read in full each paper remaining after 3rd stage and kept only those relevant to the research questions. As a result of this phase, 101 studies were selected.
- *5th stage:* The 101 selected studies were analyzed and the data were extracted to answers the research questions.

[5]More information can be obtained at: <https://parsif.al/>.

III. Results and Synthesis

To present our findings, we have structured this section according to the research question. In this sense, initially, in Subsection III-A we present the distribution of selected studies over the years. Subsequently, in Subsection III-B we present the identification of the educational areas that are being supported by the pedagogical conversational agents. In Subsection III-C we reveal the relationship between conversational agents and the levels of education. The communication strategies that have been used by the pedagogical conversational agents are presented in Subsection III-D. Finally, in Subsection III-E we demonstrate how the conversational agents are being developed.

A. Overview of the selected studies

Figure 2 presents the temporal evolution of publications, pointing out the number of primary studies per year. We can observe that the amount of studies has increaded since 2007, despite some casualties at certain years (as in 2009 and in 2015). We believe this shows that the subject is becoming more and more relevant within the community, as there is an increase of interest for the scientific community in the area. We also managed to realize that there was a concentration of publications between 2010 to 2014. This concentration returned in 2016 and settled since then.

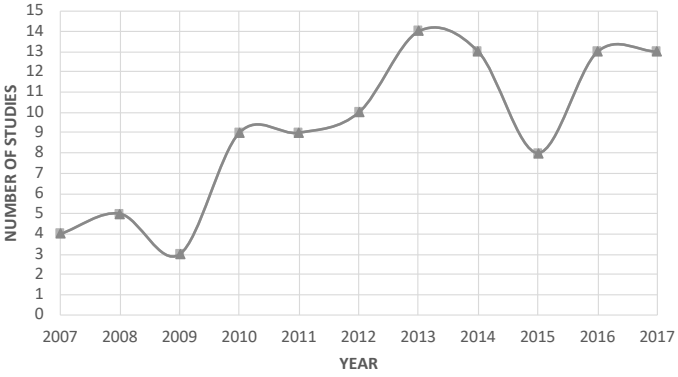

Fig. 2. Number of primary studies published per year

The primary studies were published in different scientific dissemination vehicles. We classify these publications into three categories, considering the type of vehicle in which they were published: scientific journals, conference proceedings and book chapters. Figure 3 illustrates this distribution, showing that most of the primary studies were published in Conferences (59); 36 primary studies were published in Journals; and a smaller number (six) in Book chapters.

B. Application areas

The first issue that we investigated was the area of knowledge in which conversational agents have been investigated and used as a teaching support mechanism. This information is useful for educators and researchers because they can locate similar studies more easily, observe scientific gaps and use the related studies as models or parameters to develop new scientific research.

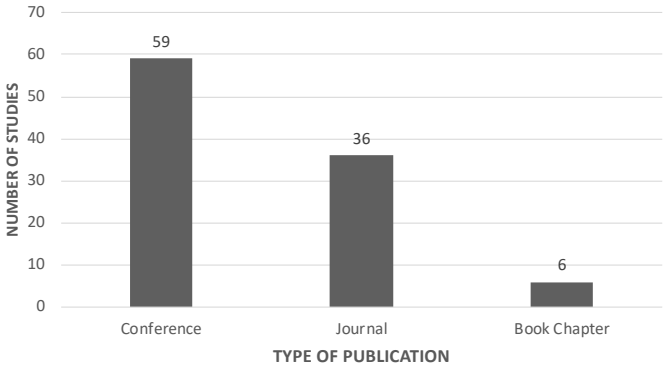

Fig. 3. Primary studies organized by publication type

Table IV shows that 30 studies did not mention the area of application of the conversational agents, or did not relate the conversational agent to a specific domain. In these studies, the authors present generic conversational agents that can be used for education, without being a conversational agent intended to support the teaching of specific content (e.g., a study seeks to investigate the effects of an agent's presence and different types of feedback on learning, motivation and cognitive load in a multimedia-learning environment [12]).

Fourteen studies focused on language teaching and learning, which we classified in the Linguistics area. Such studies establish contributions to the study of languages, such as English and Arabic, mentioning that conversational agents encourage student training, easing the learning of a new language. They enable the student to exercise the formulation of phrases and questions, to understand the meaning of words and to exercise the spelling of terms.

We also found 13 primary studies with a focus on Computer Science education, in which conversational agents are used to support teaching and learning practices in different subareas of computing. These studies mention the importance of using a mechanism that supports natural language guidance in solving students' doubts and difficulties while they learn themes as programming, database and computer networks.

In a smaller proportion, we found 12 studies about the teaching of medicine, psychology and nursing, which were classified in the area of Health, and 8 studies focusing on Mathematics, with particular emphasis on linear algebra and analytical geometry. Other areas of application of the extracted studies are summarized in Table IV.

With the results presented in Table IV, we intend to contribute to researchers and teachers who are interested in the theme can consider the mapped studies and the conversational agents described in such studies. For example, teachers who have been teaching languages can reuse one of these conversational agents in their classes, according to the language that is been taught. Still, these results contribute to those seeking for learning, since they will be able to find and use the more appropriate conversational agents to solve doubts,

find teaching materials, among others.

TABLE IV
APPLICATION DOMAIN OF EACH PRIMARY STUDY

Knowledge area	Number of studies
Linguistics	14
Computer Science	13
Health	12
Mathematics	8
History	5
Physchics	5
Biology	4
Special Education	4
Engineering	3
Chimistry	2
Social Science	1

C. Education levels

We also identify the educational level that each conversational agent was developed. This information is important because to show which pedagogical conversational agents have been developed to support the teaching of the content to a specific target audience. We carefully analyzed the audience mentioned by each primary study and classified the studies based on the International Standard Classification of Education levels (ISCED) [29], which is the standard adopted by UNESCO (United Nations Educational, Scientific and Cultural Organization) - See Table V.

TABLE V
CLASSIFICATION OF EDUCATION LEVELS

ISCED levels	Description
Level 0	Early childhood education
Level 1	Primary education
Level 2	Lower secondary education
Level 3	Upper secondary education
Level 4	Post-secondary non-tertiary education
Level 5	Short-cycle tertiary education
Level 6	Bachelor's or equivalent level
Level 7	Master's or equivalent level
Level 8	Doctoral or equivalent level

When reading the primary studies, we observed that 30 papers address conversational agents to be used in undergraduate courses, which were classified in the sixth level of ISCED. It is worth noting that in some studies, conversational agents were applied to students from different levels of education. Therefore, we had to classify them in more than one category/educational level. For example, 22 studies mentioned the establishment and application of conversational agents to masters and doctoral students. Table VI presents the distribution of primary studies by educational levels.

We were unable to observe conversational agents supporting students in the category 0. This may be because students at this level are under the age of five and are beginning to learn how to communicate, in the literacy stage. As the interaction with the conversational agents usually occurs through text and voice, we assume that researchers must believe that this type of software system is not suitable for students of this educational level.

In addition to not being able to identify studies at ISCED level 0, we were also unable to recognize conversational agents used at levels 4 and 5. Perhaps the authors of the primary studies did not have the opportunity to apply conversational agents at these levels of education. However, we believe that some of the conversational agents that have been defined for level 6 can somehow be extended or simply used at these levels. It is worth noting that, for this to occur, it is necessary to carefully understand the context in which the agents were used. A conversational agent dedicated to teaching programming, for example, can be used by students of an educational level other than undergraduate. On the other hand, a conversational agent that was made for a specific theme of a medical course will have less applicability at educational levels such as ISCED level 4 and ISCED level 5.

Finally, among the 101 studies selected, three of them did not clearly indicate the educational target audience associated with their conversational agents, considering a more generic scope. For this reason, they were not considered in the Table VI.

TABLE VI
NUMBER OF STUDIES BY EDUCATIONAL LEVEL

Education levels	Number of studies
Level 6	30
Level 7 and level 8	22
Level 1	15
Level 2	10
Level 3	8
Level 6 and level 7	5
Level 2 and level 3	3
Level 3 and level 6	2
Level 1, level 3, and level 6	1
Level 2 and level 6	1
Level 3, level 7, and level 8	1

D. Communication strategies

The third research question of our systematic mapping focused on understanding which means of communication have been used by the pedagogical conversational agents. This information becomes relevant as we can better understand how the interaction between the system and the students occurred. Despite this importance, in four studies we were unable to identify the communication strategies used by the conversational agents. Among the 97 remaining studies, we found that communication by text messages was the medium most used, with some of these (28) using voice to complement text.In one study, the authors mentioned that the conversational agent was able to talk through voice, which was generated by voice synthesizers. Table VII presents the media that was used by the educational conversational agents to interact in with students.

TABLE VII
NUMBER OF STUDIES BY COMMUNICATION STRATEGIES

Communication strategies	Number of studies
Text	69
Text and voice	28
Voice	2

To complement our analysis, we related the communication strategy with the distribution of studies over the years. As the period considered in the mapping included the movement before and after the emergence of technologies that allow the communication of humans by voice commands with software systems (e.g., Siri [6], Google Assistant [7]), we assumed that we would be able to have an understanding of the pedagogical conversational agents that were influenced by the emergence of such technologies. Therefore, Figure 4 shows this relationship.

As a result, we can notice that since 2010 our community has intensified efforts to produce pedagogical conversational agents that can recognize and reproduce voice commands. In particular, there was the emergence of pedagogical conversational agents that can interact both by voice commands and by text messages (i.e., merging the two communication strategies). Before 2010, few solutions considered communication through voice. In the years after 2010, we noticed that the number of conversational agents that can interact with students through text and voice has increased. In 2014, for example, the number of pedagogical conversational agents that mix the two communication strategies surpassed the number of agents that exclusively use text-based communication. Although this has occurred, it is important to consider that the number of text-based conversational agents has also expanded. Therefore, when looking at Figure 4, we assume that the increase in the number of conversational agents that can interact by voice and text has grown in the same proportion as conversational agents that use exclusively text.

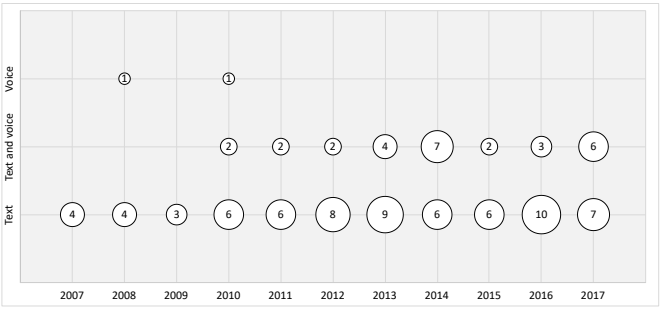

Fig. 4. Relationship between the communication strategy and the period the primary study was published

E. Technical characteristics of pedagogical conversational agents

Different types of techniques, methods, and frameworks have been used during the development of conversational agents. Previous studies have attempted to map these design techniques considering conversational agents that have participated in international competitions such as the Loebner Prize [30]. This type of study aims to help the community at understanding what trends exist in the development of this type of software. From this, we can, for instance, direct new research, produce warnings about the use of inappropriate or outdated technologies, among others.

The first aspect observed was the design technique employed into conversational agents. We have identified 27 different types of design techniques that have been adopted. Some of these design techniques were used in more than one primary study (e.g., pattern matching), while others were used only in one study (e.g., lexical network analysis). Also, there are studies addressing the use of more than one technique (e.g., neural networks and bayesian models) and studies that did not mention the techniques used during the development of the pedagogical conversational agents. We summarized the techniques that were used in more than one primary study in Table VIII. As more than 50% of the studies did not mention the techniques used, this data is also presented in Table VIII.

TABLE VIII
DESIGN TECHNIQUES MORE USED IN THE DEVELOPMENT OF PEDAGOGICAL CONVERSATIONAL AGENTS

Design technique	Number of primary studies
Uninformed	53
Pattern Matching	18
Ontologies	4
Bag of Words	3
Markov Chain Models	2

The 21 primary studies that were not counted in Table VIII used test techniques that are not repeated. In Table IX we have organized the design techniques that have not been previously presented. In the first column we list the techniques found in studies that mentioned only one design technique, and in the second column, the techniques that were identified in studies that use more than one technique.

TABLE IX
DESIGN TECNIQUES USED IN THE DEVELOPMENT OF PEDAGOGICAL CONVERSATIONAL AGENTS

	Only one technique	More than one technique
Design techniques	Bazzar agent framework	Human-centered Distributed Conversational Modeling and Case-based reasoning
	Cross-collection Latent Dirichlet Allocation	Neural Networks and Bayesian models
	Cleverbot	Parsing and Ontologies
	Decision Tree	Pattern Matching and Concept Maps
	Finite State Machine	Pattern Matching and FreeLing
	Fuzzy Logic	Pattern Matching and Parsing
	Lexical Network Analysis	Pattern Matching and Rules created by the authors
	MASCARET meta-model	Policy Iteration
	Naive Bayes	RUANLP algorithm
	Neural Networks	
	OpenNLP	
	Parsing	

After organizing this systematic mapping, we were able to make three further observations. First, most studies have not described the design techniques adopted, making it difficult to have an overview of the use of such techniques. In this sense, we highlight for the next works the need to clearly address which design techniques are being used. Second, there are

[6]Siri was launched by Apple in 2011.
[7]Google Assistant was launched in 2016.

authors who believe that they need to develop pedagogical conversational agents with more than one technique, because every technique has limitations and adopting only one cannot accurately perform all types of necessary linguistic treatments. Actually, one of the biggest discussions in the context of conversational systems is the definition of the best techniques to achieve natural language processing with greater precision, bringing the interaction between humans and machines closer to the "real world" [31]. Therefore, some authors have mixed some techniques for this to happen. Finally, the Pattern Matching technique [32] has been addressed in a significant set of studies and may indicate a trend that authors follow when implementing their pedagogical conversational agents.

To complement the analysis of design techniques, we mapped the programming languages that have been used in the development of pedagogical conversational agents. The design technique is related to the linguistic treatments that will be done, whereas the languages are used to implement these techniques. For example, Pattern Matching is a technique that has a certain behavior capable of contemplating four types of linguistic treatment (morphosyntactic, semantic, discursive and pragmatic) [9], and the AIML is one of the main language defined to implement this technique [32].

Similar to the previous analysis, we observed that 53 studies did not mention the programming language used. When studies mention a language, most of them describe the use of XML-based markup languages, such as AIML. We agree that this type of language is often used to implement conversational agents based on Pattern Matching. Therefore, our previous observation made on the use of Pattern Matching can be supported by the results presented in Table X.

TABLE X
PROGRAMMING LANGUAGES USED IN THE DEVELOPMENT OF PEDAGOGICAL CONVERSATIONAL AGENTS

Programming languages	Number of studies
Uninformed	53
XML	29
Java	8
Node.js	3
PHP	3
C#	1
C# and .NET	1
C++	1
Java, PHP and JavaScript	1
PHP and Python	1

Finally, the last analysis carried out in the primary studies concerns the representation of pedagogical conversational agents. Some studies have discussed the pedagogical benefits of using representations of agents (e.g., characters, 3D avatars) [33]. Also, there is a front of studies that have been concerned with strategies to support the definition of embodied conversational agents, in which the interaction stops considering only verbal communication (e.g., text or voice) and includes non-verbal communication (i.e., the bodily movements performed by the conversational agent) [34]. Thereby, we intend to understand how pedagogical conversational agents have been represented.

As a result, we observed that there is no common sense in the way the conversational agents are visually represented. Table XI presents the complete data on the conversational agents' representation/appearance. We have noticed that 29 primary studies describe that their conversational agent does not have an appearance, 17 studies do not mention whether the conversational agent has any avatar or image to represent it, and 16 studies describe the appearance of the conversational agent like a full humanoid 3D body. In a smaller proportion, 2D heads (8 studies), humanoid 2D heads (8 studies), fictitious character full 2D body (7 Studies), humanoid 3D head (7 studies), among others, were the choice of representation.

It is worth mentioning that in three studies we found more than one pedagogical conversational agent and we were able to observe that these conversational agents have different appearances. For this reason, in the 'Appearance' column in Table XI, there are three studies that cite more than one appearance.

TABLE XI
PEDAGOGICAL CONVERSATIONAL AGENTS APPEARANCE

Appearance	Number of studies
No representation	29
Uninformed	17
Full humanoid body 3D	16
Character head 2D	8
Humanoid head 2D	8
Full character body 2D	7
Humanoid head 3D	7
Full humanoid body 2D	4
Character head 3D	1
Character head 3D and character head 2D	1
Full character body 3D	1
Full character body 3D and humanoid head 3D	1
Humanoid head 3D, humanoid head 2D, character head 3D and character head 2D	1

After the analysis carried out, we believe that there is no trend towards a specific type of appearance during the establishment of conversational agents. Some pedagogical conversational agents have a 2D representation, others 3D. Some are represented in full bodies, others only by the head. Also, some of them do not have any representation. We believe that agents usually have representations when authors are interested in understanding the impact of a given appearance on issues associated with the teaching and learning process (e.g., expressing emotions to stimulate the student).

IV. THREATS TO VALIDITY

In this study, we attempted to map the state of the art on pedagogical conversational agents. Although we have followed a systematic methodological procedure, some risks could have threatened the validity of the results obtained [35]. Therefore, during the planning and conduct of systematic mapping, we identified and carried out some actions to mitigate these risks, which we discuss next.

- *Absence of relevant primary studies:* To mitigate this threat, we defined a protocol before performing the mapping. Initially, we defined the research questions.

From the set of questions, we use previous systematic mappings on conversational agents to define the search string (e.g., studies [15]–[20], [25]). This string was validated by four researchers who have been working for at least five years on the topic of pedagogical conversational agents. In the sequence, we defined the databases that would be used in the mapping, aiming at selecting relevant sources in the Computer Science area, used before in systematic mapping studies of the state of the art. Therefore, we believe that we have covered a significant number of studies for selection.

- *Lack of quality in primary studies:* To mitigate the risk of selecting studies with low quality, we defined a set of inclusion and exclusion criteria in the protocol. These criteria were defined by the four researchers involved in the study. For example, we specified that only primary studies that described experiences of conversational agents that had undergone some type of evaluation would be included.
- *Data extraction and incorrect classification:* To support data extraction from primary studies and to carry out an appropriate classification, we draw on previous work to define a data extraction form. We met to discuss when some information was not obvious to answer the research questions. Thus, no data extracted was selected at random.

V. Limitations of this study

Although we took steps to minimize threats to the validity of the study, our research still has some limitations. However, we believe that these limitations do not compromise the validity of the study. Particularly, we assume that the main limitation of the study is related to the period considered for systematic mapping. This mapping was carried out in the context of a larger study in early 2018, and we did not consider publications before 2007 in order to address pedagogical conversational agents that have been developed with more up to date technologies, maintaining a 10 year period. Although we have not updated the systematic mapping, we believe that the panorama presented herein has not been severely modified in the last recent years. New conversational agents might have been established, but we believe that the trends observed have not changed in a two-year period. Also, the protocol defined and used in systematic mapping allowed us to obtain a representative sample of primary studies that address already established pedagogical conversational agents. Even so, we may have lost relevant studies published in the last two years, and as a result, we have prepared a lab package with the protocol for a systematic mapping, with the complete list of selected primary studies, the extracted data and the analyzes performed, to allow it to be updated or replicated by researchers who are interested in the topic. The laboratory package has been registered in the Zenodo repository[8], an open access mechanism that can be accessed through the DOI register available at: <10.5281/zenodo.3942106>.

[8]More information available at: < https://zenodo.org/>.

VI. Conclusions and Future Work

This study presents a systematic mapping of pedagogical conversational agents, aiming at understanding how it is possible to take advantage of this supporting mechanism and to disseminate its educational use. Through a search in six international databases, 2.619 studies were first identified, and 101 were selected after the application of the inclusion and exclusion criteria. The detailed analysis of these 101 studies brought important results, with the identification of areas and levels of education for which the conversational agents were developed, the communication strategies used by agents to interact with students, design techniques, programming languages and how conversational agents have been represented.

With this systematic mapping, we hope that our colleagues in the field can use the results to direct new research and to support decision-making processes. From our results, it is possible, for instance: (i) to define a conversational agent for an area of knowledge that has not yet been investigated; (ii) to make an effort to understand why there are conversational agents produced for the same area; (iii) to define studies to verify which educational levels are most suitable for the use of conversational agents since at some levels the agents have not been explored; (iv) to help developers to select conversational agent development techniques that are more widely used; among others.

Further research could help us to understand in greater depth the roles of pedagogical conversational agents in which education can most benefit from, and what kind of pedagogical practice conversational agents can help the most. We believe that conversational agents have a potential that outweighs the ability to help students train a new language, to solve questions about a computing topic, or recommend material on an engineering topic, and it is necessary to understand this potential. We also hope that future studies could address the environments in which conversational agents have been made available (e.g., Intelling Tutoring Systems, Learning Management Systems, Vitual Worlds, among others). Since 2015, there have been studies addressing conversational agents produced to support MOOCs (Massive Open Online Courses), but in what other educational environments can they be used?

Acknowledgment

This study was financed by the University of São Paulo and the Brazilian funding agencies: the State of São Paulo Research Foundation - FAPESP (under processes no. 2013/07375-0, 2017/10941-8, and 2018/26636-2), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and CNPq (under process no. #312922/2018-3).

REFERENCES

[1] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Pearson, 3 ed., 2009.

[2] O. V. Deryugina, "Chatterbots," *Scientific and Technical Information Processing*, vol. 37, no. 2, p. 143–147, 2010.

[3] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge, "Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study," *Digital health*, vol. 5, pp. 1–12, 2019.

[4] A. Fadhil and S. Gabrielli, "Addressing challenges in promoting healthy lifestyles: The al-chatbot approach," in *International Conference on Pervasive Computing Technologies for Healthcare*, p. 261–265, 2017.

[5] J. T. S. Quah and Y. W. Chua, "Chatbot assisted marketing in financial service industry," in *International Conference on Services Computing*, pp. 107–114, 2019.

[6] A. Sciuto, A. Saini, J. Forlizzi, and J. I. Hong, ""hey alexa, what's up?": A mixed-methods studies of in-home conversational agent usage," in *Designing Interactive Systems Conference*, p. 857–868, 2018.

[7] Z. Guo and T. Inoue, "Using a conversational agent to facilitate non-native speaker's active participation in conversation," in *Conference on Human Factors in Computing Systems*, p. 1–6, 2019.

[8] A. Kerry, R. Ellis, and S. Bull, "Conversational agents in e-learning," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 169–182, 2009.

[9] L. N. Paschoal, L. F. Turci, T. U. Conte, and S. R. S. Souza, "Towards a conversational agent to support the software testing education," in *Brazilian Symposium on Software Engineering*, p. 57–66, 2019.

[10] A. C. Graesser, Z. Cai, B. Morgan, and L. Wang, "Assessment with computer agents that engage in conversational dialogues and trialogues with learners," *Computers in Human Behavior*, vol. 76, pp. 607 – 616, 2017.

[11] S. D. Craig and N. L. Schroeder, "Reconsidering the voice effect when learning from a virtual human," *Computers & Education*, vol. 114, pp. 193 – 205, 2017.

[12] L. Lin, R. K. Atkinson, R. M. Christopherson, S. S. Joseph, and C. J. Harrison, "Animated agents and learning: Does the type of verbal feedback they provide matter?," *Computers & Education*, vol. 67, pp. 239 – 249, 2013.

[13] E. V. B. Aguiar, L. M. R. Tarouco, and E. Reategui, "Supporting problem-solving in mathematics with a conversational agent capable of representing gifted students' knowledge," in *Hawaii International Conference on System Sciences*, pp. 130–137, 2014.

[14] J. Bergmann and A. Sams, *Flip your classroom: Reach every student in every class every day*. Washington, D.C., EUA: International Society for Technology in Education, 2012.

[15] R. Winkler and M. Söllner, "Unleashing the potential of chatbots in education: A state-of-the-art analysis," in *Academy of Management Annual Meeting*, pp. 1–40, 2018.

[16] S. Hobert, "How are you, chatbot? evaluating chatbots in educational settings: Results of a literature review," *Lecture Notes in Informatics*, vol. 17, pp. 259–270, 2019.

[17] N. L. Kuyven, C. A. Antunes, V. J. B. Vanzin, J. L. T. Silva, A. L. Krassmann, and L. M. R. Tarouco, "Chatbots in education: a systematic literature review (in portuguese)," *RENOTE - Revista Novas Tecnologias na Educação*, vol. 16, no. 1, pp. 123–132, 2018.

[18] H. N. Io and C. B. Lee, "Chatbots and conversational agents: A bibliometric analysis," in *International Conference on Industrial Engineering and Engineering Management*, pp. 215–219, 2017.

[19] A. A. Bernardini, A. A. Sônego, and E. Pozzebon, "Chatbots: An analysis of the state of art of literature," in *Workshop on Advanced Virtual Environments and Education*, pp. 1–6, 2018.

[20] S. Roos, "Chatbots in education: A passing trend or a valuable pedagogical tool?," Master's thesis, Uppsala University, Disciplinary Domain of Humanities and Social Sciences, Faculty of Social Sciences, Department of Informatics and Media., Uppsala, Sweden, 2018.

[21] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the facebook messenger," *Computers & Education*, vol. 151, p. 103862, 2020.

[22] T. Dyba, T. Dingsoyr, and G. K. Hanssen, "Applying systematic reviews to diverse study types: An experience report," in *International Symposium on Empirical Software Engineering and Measurement*, p. 225–234, 2007.

[23] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Tech. Rep. 2.3, Keele University and University of Durham, Keele, UK, 2007.

[24] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, pp. 1 – 18, 2015.

[25] J. L. Z. Montenegro, C. A. Costa, and R. R. Righi, "Survey of conversational agents in health," *Expert Systems with Applications*, vol. 129, pp. 56–67, 2019.

[26] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *International Conference on Evaluation and Assessment in Software Engineering*, pp. 68–77, 2008.

[27] V. R. Basili, "Quantitative evaluation of software engineering methodology, 1985," in *Pan Pacific Computer Conference*.

[28] S. Hussain, O. Ameri Sianaki, and N. Ababneh, "A survey on conversational agents/chatbots classification and design techniques," in *Workshops of the International Conference on Advanced Information Networking and Applications*, pp. 946–956, 2019.

[29] ISCED, "International standard classification of education," Tech. Rep. 1, United Nations Educational, Scientific and Cultural Organization, Quebec - Canada, 2012.

[30] A. A.-K. S and J. Woods, "Survey on chatbot design techniques in speech conversation systems," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 7, pp. 72–80, 2015.

[31] S. Langer and M. Hickey, "Augmentative and alternative communication and natural language processing: Current research activities and prospects," *Augmentative and Alternative Communication*, vol. 15, no. 4, pp. 260–268, 1999.

[32] R. S. Wallace, *The Anatomy of A.L.I.C.E.*, pp. 181–210. Dordrecht: Springer Netherlands, 2009.

[33] J. T. Doswell, "Pedagogical embodied conversational agent," in *IEEE International Conference on Advanced Learning Technologies*, pp. 774–776, 2004.

[34] J. Cassell, "Embodied conversational interface agents," *Communications of the ACM*, vol. 43, no. 4, pp. 70–78, 2000.

[35] C. Wohlin, P. Runeson, M. Hst, M. C. Ohlsson, B. Regnell, and A. Wessln, *Experimentation in Software Engineering*. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg, 1 ed., 2012.