



A Proposed Framework of an Intelligent Arabic Chatbot for Teaching Islamic History

O.M.EI-Saadawi¹, A.M.EI-Saadawi², M.M.EI-Saadawi³ ¹Mansoura University, Faculty of Engineering, Egypt

iansoura University, Faculty of Engineering, Egyp

²ZAD Group Company, Damam, KSA

³ Mansoura University, Faculty of Engineering, Egypt ¹omima.sa3dawi@gmail.com, ²ahmed@saadawi.me, ³m_saadawi@mans.edu.eg

ABSTRACT

A Chatbot is a conversational software agent, which interacts with users using natural language such as English or Arabic. Many Chatbots exist, with different knowledge-bases, programmed by the Chatbot builders. Reviews of relevant research show that there is a shortage in Arabic Chatbots. The purpose of this study is to propose a framework of an intelligent Arabic-based Chatbot to teach Islamic history, which can be used to create an Arabic Chatbot like human. The study presents an application on the use of the proposed Chatbot for teaching Prophet Muhammad Life (Seerah).

Keywords: Arabic Chatbot, Islamic History, Speech interaction

1. INTRODUCTION

Speech is one of the most powerful forms of communication between humans. Hence, using speech is the ambition of researchers' in the domain of human computer interaction in order to to improve speech interaction between the humans and the computer for simulating human-human speech interaction. Speech interaction using modern networked computing devices has received increasing interest in the past few years (Abdul-Kader, 2015).

Prior to the dissemination of computers, we could distinguish humans from non-humans on the basis of an ability to participate in conversations. However, nowadays we have hybrid systems operating between humans and non-humans with whom we can talk in ordinary language (Colby, 1999a).

Human computer interfaces were created to facilitate communication between humans and computers. For instance, information retrieval systems such as Google, Yahoo, MSN Ask Jevees, and Lycos are used to remotely access and search a large information system based on keyword matching. However, with the massive amount of information available via web pages, what a user really needs is only an answer to his/her request or question instead of documents or links. This is what a Chatbot system is intended to do.

The idea of Chatbot systems originated in the Massachusetts Institute of Technology (Weizenbaum 1966, 1967), where Weizenbaum implemented the ELIZA Chatbot to emulate a psychotherapist. The idea was simple and based on keyword matching. The input is

inspected for the presence of a keyword. If such a word is found, the sentence is mapped according to a rule associated with the keyword; if not, a connected free remark, or under certain conditions an earlier transformation, is retrieved. The next major Chatbot was PARRY (Colby 1973, 1999b). In contrast to ELIZA, instead of simulating a psychotherapist, PARRY modeled a paranoid patient during an interview with his therapist. Saygin et al. (2000) noticed that, "Both ELIZA and PARRY use certain tricks to be able to successfully perform in conversations. ELIZA directs the conversation away from herself by asking questions. ELIZA uses parts of the user's input in the output questions and seems to be following the conversations. In addition to these techniques, PARRY has little stories to tell and tend to insert these in the conversation." Since then, a range of new Chatbot architectures have been developed, such as: MEGAHAL (Hutchens, 1996), CONVERSE (Batacharia et al. 1999), ELIZABETH (Abu Shawar & Atwell, 2002), HEXBOT (2004).

Vrajitoru (2003) and Ratkiewicz (2004) proposed a new innovative pattern matching approach for Chatbots. The authors used a Genetic Algorithm (GA) with natural language to generate a new sentence from existing ones in order to improve the diversity of the response. Bhargava et al. (2009) presented a design of a new AIML based Chatbot of natural language speech and limited word inputs and outputs so as to use it in E-learning systems to enable disabled people to learn via speech. Mikic et al. (2009) modified the Chatbot CHARLIE to incorporate it into the platform INtelligent Educational System (INES) in order to improve the conversation between students and educational systems. Lokman and Zain (2009) presented a design of a proposed Chatbot that has the ability to remember previous conversations in order to work as a virtual adviser for diabetic patients. Lokman and Zain (2010) presented a modification and extension of the Chatbot ViDi by adding the prerequisite matching techniques in order to attain a conversational manner rather than a QA form and make it available to users on the internet via a web browser. Rosmalen et al. (2012) extended an existing serious game by adding a simple Chatbot to give the opportunity for trainees to be aware of work and activities on the first day of their employment. Pereira and Coheur (2013) presented a new corpus (knowledge base) that avoids overlapping, identifies personal questions, and rejects unwanted words or topics by combining available QA and dialogue formats.

In 2016, social media platforms like Facebook allowed developers to create a Chatbot for their brand or service so that consumers could carry out some of their daily actions from within their messaging platform. This development of AI technology has excited everyone, as the possibilities for the way we communicate with brands have been exponentially expanded (Bayerque, 2016). Incorporating Chatbots into society has brought us to the beginning of a new era in technology: the era of the conversational interface. Although some commercial products have emerged recently in the market (e.g. Microsoft Cortana) as dialogue Chatbots, improvements need continuous research and lack a common solution. Each researcher needs to robustly document any successful improvements to allow the human computer speech interaction to agree to a common approach.

There are two main types of Chatbots, one functions based on a set of rules, and the other more advanced version, uses artificial intelligence. The first type is very limited. It can only respond to very specific commands. If the user says a wrong statement, it does not know what its meaning is. The second type has an artificial brain and it based on artificial intelligence technology. Users do not need to be specific when communicating with it. It understands language, not just commands. This bot becomes smarter as it learns from conversations it has with people.

The objectives of this paper is to introduce a proposed framework for an intelligent Arabic Chatbot that can help and guide designers for the efficient use of Chatbots and simulations used for teaching Islamic history.

2. CURRENT ARABIC APPLICATIONS OF CHATBOT

An Arabic Chatbot giving answers from the Qur'an was designed by Abu Shawar & Atwell (2004). They presented machine-learning techniques used to generate an Arabic Chatbot, which accepts user input in Arabic and generates replies extracted from the Qur'an. In principle, the Qur'an provides guidance and answers to religious and other questions; so they used the Qur'an as a training corpus for Chatbot. As the Qur'an is not a transcription of a conversation, they adapted the learning process to cope with the structure of the Qur'an in terms of "sooras" and "ayyas". The resulting system accepts user input in Arabic, and answers with appropriate "ayyas" from the Qur'an. Kanaan et al. (2009) described an approach for the construction of a question answering system that provides short answers to questions expressed in the Arabic language. The system utilized techniques from information retrieval and natural language processing to process a collection of Arabic text documents as its primary source of knowledge.

Recently, there are many simple Islamic applications of Chatbots. These applications include:

- ➢ Islamic Messages: Receive Daily Qur'an, Hadith and Islamic updates, and Educational Videos from Scholars. (Oct. 2015)
- IlmusSalaf: earth Africa ilmusSalaf introduces Islam in its most pure and clear form based on the Quran and Sunnah, taken upon the way of the Salaf. This bot gives you details of topics, website and social media links, audio/video clips, latest updates, and other information. (July 2016)
- Daily Islamic reminder: remind for verily, the reminding profits the believer. (Feb. 2016)
- Azkarbot: makes you remember morning Azkar and evening Azkar. It provides users with a "sujood" service, which gives you the right pray base on your feelings.
- Bismillah: Explore the Noble Quran in Arabic and English, and provides "tafseer" and audio. (Nov. 2015)

3. CHALLENGES FACING THE DEVELOPMENT OF ARABIC CHATBOTS

Arabic Chatbots can be helpful in many operations and help people who only know the Arabic language. However, the technology for Arabic Chatbot is still in its infancy due to some challenges surrounding the Arabic language. In contrast to English and other European languages, the Arabic language suffers from shortages in Neuro-linguistic programming (NLP) resources and tools.

Arabic language belongs to the family of Semitic languages that differs from Indo-European languages semantically, syntactically, and morphologically. Arabic is composed of 25 consonants and three long vowels. On other hand, Arabic is a derivative language where most Arabic words are derived from a root, generally composed of three consonants; occasionally the root can also be formed from two, four or rarely five consonants. Arabic words are classified into three categories (Abu Shawar, 2011):

- Original Arabic words: which include Arabic verbs and nouns that are formed according to Arabic derivation rules;
- Fixed Arabic words: which include words that do not belong to derivative rules, these words were modeled by Arabs in ancient times;
- Arabized words: which includes nouns that were taken from foreign languages and have become commonly used by Arab people.

Many challenges are facing the development of Arabic Chatbots (Xu et al. 2002).

- Orthographic variations are prevalent in Arabic.
- Arabic has a very complex morphology.
- Arabic words are often ambiguous due to the tri-literal root system.
- Synonyms are widespread, perhaps due to variety in expression which is appreciated as part of good writing style.
- Broken plurals are common.
- The absence of diacritics in the written text creates ambiguity, and therefore, complex morphological rules are required to identify the tokens and to parse the text.
- The writing direction is from right-to-left, with some of the characters change their shapes based on their location within the sentence.
- Capitalization is not used in Arabic, which makes it hard to identify proper names, acronyms, and abbreviations.
- Lack of adequate resources (corpora, morphological analyzers, lexicons, part-of-speech taggers, etc.).

4. METHODS USED IN CHATBOT DESIGN

Building a Chatbot needs highly professional programming skills. It also needs experienced developers to achieve basic levels of realism. There is a complicated development platform behind any Chatbot which will only be as good as its knowledge base which maps a user's words into the most appropriate response. The bot developer usually builds the knowledge base as well. However, there are some platforms which provide a learning environment. Writing a perfect Chatbot is very difficult because it needs a very large database and must provide reasonable answers to all interactions.

A Chatbot can be separated into three sections: Responder, Classifier and Graph master (Anik, 2016):

- 1) **Responder**: this part assumes the interfacing part between the bot's fundamental schedules and the client. The undertakings of the responder are: exchanging the information from the client to the Classifier and controlling the info and yield.
- 2) **Classifier**: it is the part between the Responder and the Graph master. This current layer's capacities are: separating and normalizing the info, sectioning the information entered by the client into intelligent segments, exchanging the standardized sentence into the Graph master, preparing the yield from the Graph master, and taking care of the instructions of the database syntax (e.g. AIML).
- 3) **Graph master**: is the part used for pattern matching that does the accompanying assignments: sorting out the mind's substance, stockpiling and holding the example coordinating calculations.

On other hand, there are a number of approaches to create a knowledge base for a Chatbot and includes writing by hand and learning from a corpus. Learning here means saving new phrases and then using them later to give appropriate answers for similar phrases (Pereira & Coheur, 2013).

To design any Chatbot, the designer must be familiar with a number of techniques such as:

- 1) **<u>Parsing</u>**: this technique includes analyzing the input text and manipulating it by using a number of Natural Language Processing (NLP) functions.
- 2) <u>Pattern matching</u>: it is the technique that is used in most Chatbots, and it is quite common in question-answer systems depending on matching types, such as natural language enquiries, simple statements, or semantic meaning of enquiries (Meffert, 2006).
- 3) <u>AIML</u>: it is one of the core techniques that are used in common Chatbot design.
- 4) <u>Chat Script</u>: is the technique that helps when no matches occur in AIML. It concentrates on the best syntax to build a sensible default answer. It gives a set of functionalities such as variable concepts, facts, and logical and/or.
- 5) <u>**Relational database**</u>: is a technique used recently in Chatbot design in order to make the Chatbot remember previous conversations.
- 6) <u>Markov Chain</u>: is used in Chatbots to build responses that are more applicable probabilistically and, consequently, are more correct. The idea of Markov Chains is that there is a fixed probability of occurrences for each letter or word in the same textual data (Mladenic & Bradesko, 2012).
- 7) <u>Language tricks</u>: these are sentences, phrases, or even paragraphs available in Chatbots in order to add variety to the knowledge base and make it more convincing. The types of language tricks are:
 - Canned responses.
 - Typing errors and simulating key strokes.
 - Model of personal history.
 - Non Sequitur (not a logical conclusion).

Each of these language tricks is used to satisfy a specific purpose and to provide alternative answers to questions (Mladenic & Bradesko, 2012).

8) Ontology's: they are also named semantic networks, and refer to a set of concepts that are interconnected relationally and hierarchically. The aim of using ontologies in a Chatbot is to compute the relation between these concepts, such as synonyms, hyponyms and other relations which are natural language concept names. The interconnection between these concepts can be represented in a graph enabling the computer to search by using particular rules for reasoning (Mladenic & Bradesko, 2012).

5. A PROPOSED FRAMEWORK FOR AN INTELLIGENT ARABIC CHATBOT FOR TEACHING ISLAMIC HISTORY

The objective of this paper is to introduce a proposed framework for an intelligent Arabicbased Chatbot that can help and guide designers for the efficient use of Chatbots and simulations for teaching Islamic history. The program is composed of three phases: interaction, processing and learning phases as explained in Figure 1.

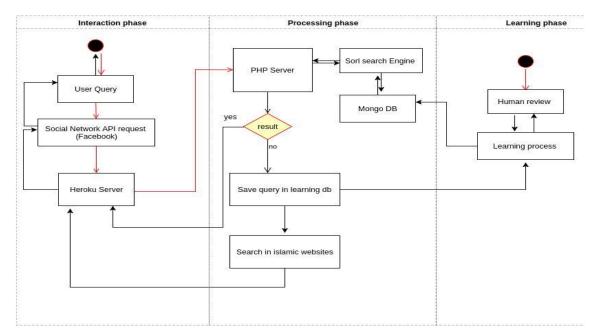


Figure 1. Structured programming of the proposed Islamic history Chatbot

The first phase in the Chatbot programming, is the interaction phase. This phase includes the user query which may be incorporated with any social network such as Facebook, Skype, Line, Telegram, etc. The data is stored in the Heroku cloud server. Cloud servers do not suffer from the usual server hardware problems and they have all cloud computing benefits; i.e. they are stable, fast and secure. On the other hand, cloud servers are economically and more efficient than the standard dedicated servers.

The second phase is the processing phase. This phase includes the use and implementation of many programming languages for handling, storing, processing, searching and recalling the data. Firstly, the Prophet Muhammad's Biography (life) is handled with a suitable data base so that it can be processed fast and easily. Many database programs can be used for handling the stored data, however, in case of Islamic history or the Prophet Muhammad's Seerah, there will be big-data involved. Hence, the content has to be managed and delivered, and it must be available on mobile and social platforms. The most suitable database which satisfies all these requirements is the MongoDB. In addition to the fact that it is an open source free program, MongoDB has many other advantages such as:

- A document database in which one collection holds different documents. Number of fields, content and size of the document can differ from one document to another.
- > Document Oriented Storage: Data is stored in the form of JSON style documents.
- Conversion/mapping of application objects to database objects not needed.
- Deep query-ability: MongoDB supports dynamic queries on documents using a document-based query language that is nearly as powerful as SQL.
- Uses internal memory for storing the (windowed) working set, enabling faster access of data.
- Replication and high availability
- ➢ Fast in-place updates and ease of scale-out

Figure 2 shows a sample of the data stored in the MongoDB.

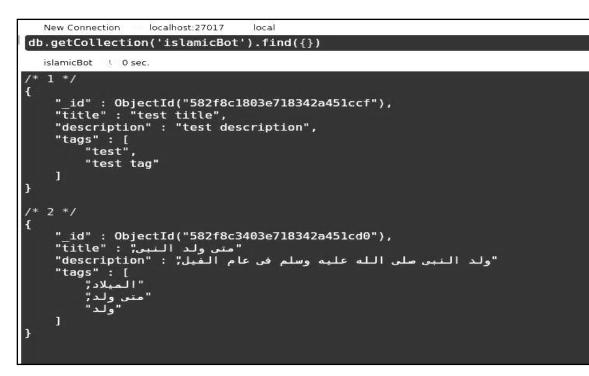


Figure 2. Sample of the data stored in the MongoDB

In this study, the PHP programming language is used to program, design, interpret, and compile website code. PHP has the following features (Hudson, 2013):

- > It provides more secured websites due to its high quality encryption.
- > It provides fast browsing capability due to its applications.
- ▶ It is compatible with almost every web-server and operating system.
- > Its applications are cheaper than other technologies.
- ▶ It has more functions to deal with MongoDB.

The PHP is accompanied with Apache Solr search engine, which is the most popular and powerful search engine. Solr runs as a standalone full-text search server. It supports Arabic language and it also has so many advantages over other available search engine (Nayrolles, 2014).

The Solr search engine starts with the words recorded in the MongoDB and if it cannot find the answer in the stored database, the query will be recorded and then a developed PHP application starts searching in Islamic web pages with an "application programming interface" (API). This interface supports the answer of any query from other web sites or applications. In case that the programmed PHP application cannot find the answer in those web pages, it gives the user a link to other Islamic web pages which do not apply API. This means that there will be different alternatives for answering each query.

The Third phase is the learning phase. This phase is responsible for learning and updating the bot according to the queries of the users. The bot will be recordable and learnable. The program will record all the queries of the users and then the program manager will update the bot memory by adding new words or queries with the appropriate answers. The last process can be implemented manually or automatically using some advanced programs such as Neuro Linguistic Programming (NLU). Unfortunately, the current NLU programs cannot support Arabic language and requires significant work to modify such code so that they can

understand and analyze Arabic words. In this study, the data and answers to queries' will be updated manually. Nevertheless, the framework shown in figure 1 can be easily modified for using NLU programs just by replacing the "Human review" block by "NLU modifications" block.

6. CASE STUDY (PROPHET MUHAMMAD BIOGRAPHY)

In this section, samples of random queries to the designed Chatbot are presented which show the different answer possibilities: direct answer, or a set of related links. Figure 3 shows a sample of a conversation with the designed Chatbot.

The proposed Chatbot can generate either a direct answer or related links for rewordings of the same query as the matching process is based on matching the words found in the user input. In Figure 2, the significant words are "الميلاد" and "ولد النبي" and "ولد النبي" in addition to the first word "متي". According to the Chatbot structure, the same answer can be generated for different queries with the same meaning. Figure 4 shows a part of the programming code used for building the proposed Chatbot.



Figure 3. A sample of the conversation with the designed Chatbot.



Figure 4. Part of the programming code used for building the proposed Chatbot

7. CONCLUSION

A Chatbot is a conversational agent that interacts with users using natural language. The main disadvantage of Chatbots is found in the need for manual development of knowledge. In a Chatbot, the bot memory can be updated by adding new words or queries with the appropriate answers. It can be updated manually or automatically. This paper has presented a proposed framework for an intelligent Arabic Chatbot that can help and guide designers for the efficient use of Chatbots and simulations for teaching Islamic history. The framework reflects the distinctive characteristics of intelligent Arabic Chatbots for the creation of a Chatbot similar to a human. The designed Chatbot can be incorporated with any social network such as Facebook, Skype, Line, Telegram ... etc. The framework program is composed of three main phases: interaction phase, processing phase and learning phase. The proposed framework is applied to design an application used for learning the Seerah of Prophet Muhammad. The proposed application was programmed using Mongo database, and PHP language accompanied with an Apache Solr search engine. The data was stored in a Heroku cloud server. The paper presented a sample of random queries to the designed Chatbot, together with the Chatbot's answer. According to the Chatbot structure, the same answer can be generated for different queries with the same meaning. Although some commercial products have emerged recently in the market as dialogue Chatbots, improvements need continuous research and currently lack a common solution. From the Islamic point of view, researchers or designers needs to robustly document any successful improvements to allow the human computer speech interaction to agree on a common approach for both Arabic and Islamic applications.

8. REFERENCES

- Abdul-Kader, S. and Woods, J. (2015). Survey on Chatbot Design Techniques in Speech Conversation Systems", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 6, No. 7, pp. 72-80.
- Abu Shawar, B. (2011). A Chatbot as a Natural Web Interface to Arabic Web QA. International Journal of Emerging Technologies in Learning (IJET), Vol. 6, Issue 1.
- Abu Shawar, B. and Atwell, E. (2002). A comparison between Alice and Elizabeth chatbot systems. School of Computing research report 2002.19. Leeds: University of Leeds.
- Abu Shawar, B. and Atwell, E. (2004). An Arabic chatbot giving answers from the Qur'an. In Bel, B. and Marlien, I. (Eds) Proc TALN04: XI Conference sur le Traitement Automatique des Langues Naturelles, Fez, Morocco.
- Anik, C., Jacob, C., Mohanan, A. (2016). A Survey on Web Based Conversational Bot Design. Journal of Emerging Technologies and Innovative Research, Vol. 3, Issue 10.
- Batacharia, B., Levy, D., Catizone, R., Krotov, A., Wilks, Y. (1999). CONVERSE: a conversational companion. In Y. Wilks (Eds), Machine conversations, Boston/Dordrecht/London: Kluwer, pp. 205-215

Bayerque, N. (2016). A short history of chatbots and artificial intelligence. Available at:

http://venturebeat.com/2016/08/15/a-short-history-of-chatbots-and-artificial-intelligence/

- Bhargava, V. and Maheshwari, N. (2009). An Intelligent Speech Recognition System for Education System. In International Conference on Smart Technologies for Materials, Communication, Controls, Computing & Energy in Chennai, India (ICST2011).
- Colby, K. (1973). Simulation of belief systems. In R. Schank and K. Colby (Eds), Computer models of thought and language, San Francisco: Freeman, pp. 251-286.
- Colby, K. (1999a). Comments on human-computer conversation. In Y. Wilks (Eds), Machine Conversations. Kluwer, Boston/Drdrecht/London, pp. 5-8.
- Colby, K. (1999b). Human-computer conversation in a cognitive therapy program. In Y. Wilks (Eds), Machine conversations. Kluwer, Boston/Drdrecht/London. pp. 9-19.
- Hudson, P. (2013). Practical PHP Programming. Online book, published by TuxRadar, Available at: <u>http://www.tuxradar.com/practicalphp</u>
- Hutchens, J. (1996). How to pass the Turing test by cheating. School of Electrical, Electronic and Computer Engineering research report TR97-05. Perth: University of Western Australia.
- HEXBOT (2004). HEXBOT chatbot website. http://www.hexbot.com/
- Kanaan, G., Hammouri, A., Al-Shalabi, R., Swalha, M. (2009). A new question answering system for the Arabic language. American Journal of Applied sciences. Vol. 6, No. 4, pp. 797-805
- Lokman A.S., and Zain, J.M. (2009). An architectural design of Virtual Dietitian (ViDi) for diabetic patients. 2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT 2009), pp. 408-411.
- Lokman A.S., and Zain, J.M. (2010). Extension and prerequisite: An algorithm to enable relations between responses in chatbot technology. Journal of Computer Science, Vol. 6, No. 10, pp. 1212.
- Meffert, K. (2006). Supporting design patterns with annotations. 13th Annual IEEE International Symposium and Workshop on Engineering of Computer Based Systems (ECBS'06), Technical University, Ilmenau. 8, pp. 445.
- Mikic, F.A., Burguillo, J.C., Llamas, M., Rodríguez, D.A., Rodríguez, E. (2009). CHARLIE: An AIML-based Chatterbot which Works as an Interface among (INES) and Humans. pp. 1-6.
- Mladenic, D. and Bradesko, L. (2012). A survey of chabot system through a Loebner prize competition.

Nayrolles, M. (2014). Mastering Apache Solr: practical guide to get to grips with Apache Solr. Create Space Independent Publishing Platform, USA.

- Pereira. M.J. and Coheur, L. (2013). Just Chat-a platform for processing information to be used in chatbots.
- Rosmalen, P., Eikelboom, J., Bloemers, E., Winzum, K., Spronck, P. (2012). Towards a Game-Chatbot: Extending the Interaction in Serious Games. 6th European Conference on Games Based Learning, 4-5 October 2012, Cork, Ireland.
- Saygin, A., Cicekli, I., Akman, V. (2000). Turing test: 50 years later. Minds and Machines, Vol. 10, No. 4, pp. 463-518.
- Vrajitoru, D. (2003). Evolutionary Sentence Building for Chatterbots. Genetic and Evolutionary Computation Conference (GECCO), Chicago, IL, pp. 315-321.
- Vrajitoru. D. and Ratkiewicz, J. (2004). Evolutionary sentence combination for Chatterbots. In The IASTED International Conference on Artificial Intelligence and Applications (AIA 2004), Innsbruck, Austria: ACTA Press, pp. 287–292.
- Weizenbaum, J. (1966). ELIZA-A computer program for the study of natural language communication between man and machine. Communications of the ACM, Vol. 10, No. 8, pp. 36-45.
- Weizenbaum J. (1967). Contextual understanding by computers. Communications of the ACM, Vol. 10, No. 8, pp. 474-480.
- Xu, J., Fraser, A., Weischedel, R. (2002). Empirical studies in strategies for Arabic Retrieval. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 86-91.

Biodata

| Omima M. El-Saadawi was born in Mansoura, Egypt in 1990. She received her B.Sc. (2011), Special Diploma in Specific Education (2013) and M.Sc. (2016) from Mansoura University, Egypt. She is now pursuing her Ph.D. at Mansoura University. From 2013, she was a computer developer at Mansoura University, Faculty of Engineering. Her research interests include computer uses in education, applications of AI to education, educational games design, and modern teaching methods. |
|---|
| Ahmed M. El-Saadawi was born in Mansoura, Egypt in 1988. He received his B.Sc. (2010), from MET Academy, Egypt. From 2013 he was a Software Engineer at islamway.net and islamqa.info one of the most popular Islamic websites in Middle East. From 2016 his interest is to use the new chat bots and AI engine technology to build an Islamic bots. |
| Magdi M. El-Saadawi was born in Mansoura, Egypt in 1959. He received his B.Sc. and M.Sc. from Mansoura University, Egypt in 1982 and 1988, respectively, and his Ph.D. from Warsaw University of Technology in 1997. He was a teaching assistant at El-Mansoura University from 1983–1992. From 1997, he was a staff member of the Electrical Engineering Department, Mansoura University, and has been a professor since May 2011. His fields of interest include AI applications, power system analysis, and learning technologies. |