

# Exploring the Pedagogical Use of AI-Powered Chatbots Educational Perceptions and Practices

E. Rakovac Bekeš\*, V. Galzina \*\*

\* I. gymnasium Osijek, Osijek, Croatia

\*\* Faculty of Education Osijek and Faculty of Kinesiology Osijek  
ella.rakovac@skole.hr, vgalzina@foozos.hr

**Abstract** - By providing personalized and adaptable learning experiences, artificial intelligence (AI) has the potential to revolutionize education. One promising development in this field is the use of generative artificial intelligence technology, such as the ChatGPT conversational agent. Chatbot technology has the potential to revolutionize the way educational content is accessed, created, understood, and implemented. In this research paper, we examine the current state of AI-powered chatbots in education and discuss the advantages and disadvantages of using chatbots in this context. A case study of an AI unit planner (Copilot) built on top of the GPT-3 application in K-12 settings will also be presented, examining how chatbots are used to assist teachers in lesson design and the extent to which teachers are familiar with them. Finally, the implications of chatbot technology for the future of education are discussed, along with recommendations for future research directions.

**Keywords** - artificial intelligence; chatbots; education; ChatGPT; educational content

## I. INTRODUCTION

Artificial Intelligence (AI) has been hailed as a transformative technology that has the potential to revolutionize a variety of fields of human endeavor. AI algorithms and techniques have been developed and implemented in a variety of human endeavors, and the impact of AI on society is likely to increase as the field continues to advance. [1]

As stated in [1], there are numerous examples of using AI for a good purpose. AI has been utilized in the medical field to diagnose diseases, provide personalized treatment plans, and monitor patient health. Algorithms based on artificial intelligence have been trained to recognize patterns in medical imaging, allowing for earlier diagnosis and more precise test results. AI has the potential to provide physicians with real-time patient data, allowing them to make informed treatment decisions.

The financial industry is another area where AI has made substantial progress. Several financial procedures, including credit scoring and fraud detection, have been automated using AI algorithms. Financial institutions and investors can make more informed decisions with the assistance of AI systems that analyze market trends and offer investment advice.

The transportation industry also enjoys the advantages of AI. The development of autonomous vehicles that can transport people and goods safely and efficiently is currently underway. Additionally, AI systems can analyze

traffic patterns and optimize routes, thereby reducing congestion and enhancing transportation efficiency overall.

AI has had an impact on the field of education as well. AI systems have been used to personalize student learning experiences and provide real-time performance feedback. AI systems can also analyze student data to identify problem areas, allowing teachers to concentrate their efforts on these areas.

AI has the potential to revolutionize a variety of other fields, such as retail, manufacturing, and energy. The promise of artificial intelligence is that it can help us make better decisions, increase efficiency, and more effectively solve complex problems. Nevertheless, it is essential to be mindful of the limitations of AI and to ensure that its application is ethical, responsible, and sustainable.

It has the potential to be a transformative technology that can enhance a variety of facets of life. Even though it's important to be aware of AI's limits and possible risks, its benefits are clear, and as the technology gets better, its effects on society are likely to grow. As AI technology continues to advance, it's essential that we develop guidelines and protocols for its use.

Given the new trend of people using digital technologies in their vacant time and the fact that a big part of daily life takes place on social networks and the Internet in general, large companies saw this as an investment opportunity and began investing substantial financial resources to contribute to the development of these technologies while increasing their revenues. Because technology has made natural interaction possible, it is inevitable that it has also become a trend in education.

## II. ARTIFICIAL INTELLIGENCE AND CHATBOTS IN EDUCATION

Artificial intelligence is a science and research aimed at creating intelligent machines, especially intelligent computer programs, according to [2]. When asked if artificial intelligence is a copy of human intelligence, McCarthy says that it can be, but not always. There are many definitions of artificial intelligence, but this is the most widely accepted one. That there is no consensus on the definition can be concluded from [3]. The issue is that intelligence itself is a problem. The same issue arises with the definition of artificial intelligence due to the complexity and varying understandings of intelligence. According to [2], artificial intelligence refers to the research and development of intelligent devices, particularly intelligent

computer systems. McCarthy responds that artificial intelligence can imitate human intelligence, but not always. There are numerous acknowledged definitions of artificial intelligence, but this one is the most prevalent. It can be concluded from [3] that there is no consistency over the definition. Problematically, intellect itself is the issue. The similar problem emerges with the definition of artificial intelligence due to the complexity and diversity of intelligence definitions.

AI is, at its most basic level, the use of computers and machines to do tasks by emulating human experience, decision-making, and other processes. In other words, AI happens when systems perform complex pattern matching and acquire knowledge in the process. There are numerous ways to understand the nature of artificial intelligence. There are two types of artificial intelligence: rule-based and machine learning-based. The second provides a recommendation or solution proposal based on decision-making guidelines. In this sense, it is the most fundamental shape. A system of intelligent tutoring is an example of this technology. It may provide students with detailed and individualized feedback [4]. AI fueled by machine learning is more effective since the computers can actually learn and improve over time, especially when presented with large, complicated datasets.

#### A. Chatbots

The term "chatbot" represents two defining factors: "chat" in lieu of conversational capabilities and "bot" shortly for robot. [5] Chat-bots are automated programs meant to perform instructions based on certain inputs and deliver feedback in a conversational manner. [6]

A chatbot is a computer program based on artificial intelligence that replicates human communication via textual and/or auditory means. However, different chatbot-related terms are used today that don't share the same interpretation: chatter bot, smart bot, educabot, quizbot, digital assistant, personal assistant, virtual tutor, conversational agent, etc. Typically, they are excellent applications of artificial intelligence (AI).

The evolution of chatbot technology has accelerated over the past few decades, due in part to recent developments in Natural Language Processing (NLP) and machine learning [7,8]. The origins of chatbots can be traced back to the 1950s, when Alan Turing wrote a paper in *Computing Machinery and Intelligence* posing the challenging issue "Can machines think?" (1950). Since then, numerous chatbots have been developed, including Eliza (1966), Parry (1972), Racter (1983), Jabberwacky (1988), and A.L.I.C.E. (1995), some of which are still in use. Originally, these chatbots were text-based and preprogrammed, based on Question - and - answer scripts, so their reactions were deemed predictable, and the interaction was not viewed as natural by human standards. Modern chatbots include new functions such as speech recognition and synthesis technology, customizable interaction, integration with third-party applications, omnichannel deployment, context-awareness, and multi-turn capacity [7,9,10]. As a result, a variety of chatbots are now integrated into many electronic devices, programs, and applications, such as messaging platforms (WhatsApp,

Telegram, Kik, Slack), interactive media and gaming platforms (Xbox, Roblox), and social media (Instagram, Facebook Messenger, Twitter). There are essentially no areas where chatbots do not currently exist or will not exist in the near future. As Fryer et al. [10] emphasized: 'During the first three decades of their existence, chatbots evolved from exploratory software to a wide variety of potential internet-based friends, advisors, and merchants. In the two decades that followed this initial rise, advancements in text-to-speech-to-text and the increasing usage of mobile devices and home assistants have made chatbots an integral part of the daily lives of many people. (p. 17) The word chatbot can be deceiving, as it can refer to a large number of programs with varying styles and functions.

Conversational AI, further known as conversational agents, can be defined as intelligent software programs that learn through communication in a manner similar to humans. Originally, conversational agents provided standard responses to common questions, but they have since evolved into more complex software with the help of Natural Language Understanding (NLU), neural networks, and deep-learning technologies. They represent a sophisticated dialogue system that replicates written and/or spoken discussion with human users, generally through the Internet. They may be based on text or be read aloud, and they may respond with speech, visuals, virtual motions, or haptic-assisted physical gestures. In addition, some of them, such as IBM Watson, Apple Siri, Samsung Bixby, and Amazon Alexa, are intelligent personal assistants (IPAs) trained to perform specific duties [11]. Fryer and Carpenter [12] studied the interaction between two chatbots and 211 students and concluded that they might be utilized well for self-practice through regular conversation, although the authors considered that these programs are more beneficial for advanced language learners. Correspondingly, Hill et al. [13] compared 100 instant messaging interactions to 100 exchanges with the chatbot named Cleverbot across seven dimensions: words per message, words per conversation, messages per conversation, word uniqueness and use of profanity, abbreviations, and emoticons. The authors discovered notable differences, such as the fact that people communicated with the chatbot for longer periods of time (but with fairly short texts) than they did with another person, and that human-chatbot communication appears to lack much of the vocabulary found in conversations between people, which is consistent with later research on the use of chatbots by children [14]. Ayedoun et al. [15] employed a semantic approach to illustrate the positive influence of deploying a conversational agent on the willingness to communicate (WTC) in the setting of English as a Foreign Language (EFL) by exposing users to a variety of everyday conversation contexts. In their study, the conversational agent created an immersive environment in which students could imitate a variety of everyday English interactions in order to lessen their fear and boost their self-confidence. Until now, numerous articles on conversational agents in language acquisition have been published. Io and Lee [16] found various study gaps related to "the paucity of research from the human perspective" (p. 216) and stated that modern types of chatbot such as mobile chat apps, embedded website services, and wearable devices should receive special attention in future research.

Shawar [17] reviewed the usage of several chatbots in language learning and emphasized some of its benefits, including student delight, less language anxiety, endless repetition opportunities, and multimodal capabilities. In a separate assessment, Radziwill and Benton [18] analyzed a selection of qualities and quality concerns associated with the creation and deployment of chatbots. The authors considered criteria such as efficiency, efficacy, and customer happiness before concluding that chatbots can positively enhance learning but can potentially be used to engineer societal harm (misinformation, rumors). In a similar vein, Haristani [19] evaluated various types of chatbots and highlighted six benefits: lessened language anxiety, widespread availability, multimodal practice, novelty effect, rich range of contextual vocabulary, and useful feedback. Bibauw et al. [20,21] presented a comprehensive review of 343 dialogue-based systems-related papers. The authors described the interactional, instructional, and technological components as well as the favorable effects on student self-confidence and motivation. Recently, Huang et al. [22] examined 25 empirical studies and found five pedagogical uses of chatbots: as interactants, simulations, helplines, transmission, and recommendations. The authors highlighted various advantages, such as timeliness, usability, and personalization, as well as disadvantages, such as the perceived unnaturalness of the computer-generated voice and communication failures.

Dokukina and Gumanova [23] underlined the originality of employing chatbots in a new scenario based on microlearning, learning process automation, and an adaptive learning environment. However, some disadvantages of using chatbots in education have also been highlighted. Criticism is primarily focused on the planned character of their responses, which makes them somewhat predictable, as well as their restricted understanding (vocabulary range, purposeful meaning) and inefficient communication (off-topic, meaningless words) [24,25]. Students' (lack of) interest in such predictable technologies and the (in)effectiveness of chatbot-human interactions in language acquisition are the two most frequently stated obstacles. In the first scenario, [26] examined chatbot-human versus human-human talks across a variety of tasks and assessed the effect on student engagement. The researchers stated that human associate tasks can predict future course interest of students, whereas chatbot partner circumstances decrease this interest. In other words, after a certain period of time, students lose interest in chatbot conversations, a phenomenon known as the novelty effect.

### B. GPT-3

More than four years ago, OpenAI academics and engineers published their initial work on generative models—language models—artificial intelligence systems that could be pre-trained with a massive and diverse corpus of text via datasets, a method they termed generative pre-training (GP). [27] Language understanding performance in natural language processing (NLP) was enhanced in GPT-n through "generative pre-training of a language model on a diverse corpus of unlabeled text, followed by discriminative fine-tuning on each individual task," as reported by the authors. This reduced the requirement for human oversight and time-consuming hand-labeling. [27]

Microsoft introduced their Turing Natural Language Generation (T-NLG) system in February 2020 [28], claiming that it was the "biggest language model ever published with 17 billion parameters." It improved performance over any other language model on a range of tasks, including text summarization and question answering.

The 2020 version of Generative Pre-trained Transformer 3 (GPT-3) is an autoregressive language model that employs deep learning to generate text that resembles human language. Given an initial text as a prompt, it will generate a continuation of the prompt. GPT-3, which was in beta testing as of July 2020, is part of a movement in natural language processing (NLP) systems toward pre-trained language representations. [29]

The structure is a decoder-only transformation network with 2048-token-long contexts and an extraordinary 175 billion parameters, needing 800GB of storage space. The model was trained through generative pre-training; it is trained to anticipate the future token based on the previous tokens. On numerous tasks, the model showed impressive zero-shot and few-shot learning. [27]

It is the third-generation language prediction model in the GPT series, produced by OpenAI, a San Francisco artificial intelligence research center, as the successor to GPT-2. [29]

The exceptional quality of the writing produced by GPT-3 makes it impossible to identify whether it has been written by a person, which has both advantages and disadvantages.

A team of 31 researchers and engineers at OpenAI described the construction of GPT-3, a third generation "state-of-the-art" language model, in an arXiv preprint dated May 28, 2020. [29] The team enhanced GPT-3's capacity by more than two orders of magnitude relative to its predecessor, GPT-2, making it the largest non-sparse language model to date. (In a sparse model, many parameters are fixed, thus even though there are extra overall parameters, there is fewer meaningful data.) Because GPT-3 is similar in structure to its predecessors, its improved capacity and number of parameters account for its greater accuracy. GPT-3 is 10 times more capable than Microsoft's Turing NLG, the next largest NLP model at the time. [30]

In this article they cautioned of the possible risks of GPT-3 and advocated for further research to decrease risk. David Chalmers, an Australian philosopher, referred to GPT-3 as "one of the most fascinating and significant AI systems ever developed." A review published in The New York Times in April 2022 stated that GPT-3 is capable of writing unique language with human-like fluidity.

Since the training data for GPT-3 was exhaustive, it doesn't require additional training for specific language tasks. [30] The trained data contains occasional harmful language, and GPT-3 generates occasional harmful language while imitating its training data. According to research from the University of Washington, GPT-3 produced similarly harmful language as GPT-2 and CTRL, which are also natural language processing models. OpenAI has adopted a variety of ways to decrease the

amount of harmful language produced by GPT-3. As a result, GPT-3 generated less harmful language than its predecessor, GPT-1.

### C. ChatGPT

OpenAI introduced ChatGPT (Chat Generative Pre-Trained Transformer) [31] in November 2022. It is built upon OpenAI's GPT-3 family of big language models and is fine-tuned (a transfer learning technique) utilizing both controlled and reinforcement learning approaches.

It was released as a concept on November 30, 2022, and soon attracted notice for its thorough responses and clear answers spanning a wide range of academic fields. Its inconsistent factual accuracy was noted as a major flaw. After the launch of ChatGPT, OpenAI was valued at \$29 billion.

Although a chatbot's primary goal is to simulate human conversationalists, ChatGPT is multifunctional. For instance, it can write and debug computer programs, compose music, teleplays, fairy tales, and student essays, answer test questions (sometimes at a level above the average human test-taker), write poetry and song lyrics, emulate a Linux system, simulate an entire chat room, play games like tic-tac-toe, and simulate an ATM. [32] The training data for ChatGPT consists of person pages and data about Internet phenomenon and programming languages, such as message board systems and Python.

ChatGPT has received extensive and strong criticisms from educators, reporters, artists, researchers, and public advocates in the months following its introduction. James Vincent of the website *The Verge* viewed ChatGPT's viral success as sign that artificial intelligence has entered the mainstream. But users have noticed on ChatGPT's "hallucinate" habit, since it fabricates facts, even references. [33]

### III. AN IMPACT ON TEACHING AND LEARNING

When considering the broader issue and what it means to teach and learn in the age of AI, embracing teaching and learning with the help of AI is a difficult and inadequately elaborated topic that teachers need to address. In fact, researchers such as Seldon and Abidoye [34] refer to education as the "Cinderella of the AI tale," pointing to the underdeveloped and generally overlooked issue of employing AI in learning and teaching settings. In contrast, Holmes et al. [35] believed that it would be presumptuous to assume that AI will have no effect on teaching and learning, not just from a technological standpoint but also from a pedagogical and ethical one, as well as from the perspective of the competencies of educators. The primary distinction between AIED and other educational technology solutions is that AIED aims to facilitate the creation of adaptable and personalized learning environments so each pupil has their own educational opportunities. AIED systems would ideally be capable of making computational conclusions that aid teachers in gaining a comprehensive picture of how students learn best and how optimal learning is influenced by past knowledge, teaching methods, and learning and physical settings.

ChatGPT writing scientific article introductions and abstracts raises ethical concerns. It has started to be stated as having co-authored several publications. [36] In *The Atlantic*, Stephen Marche wrote that its implications on academic circles, especially application essays, is unknown.

Since January 4, 2023, the NYC Department of Education has blocked ChatGPT from public school internet and devices.

One of the lead data scientists and creators of a Croatian recruitment app made for successful hiring tested ChatGPT on the Croatian language graduation exam. The output test was corrected by a professor of the Croatian language who corrects real graduation exams. [37] ChatGPT performed better on literature tasks containing text (80%), while it performed slightly worse on tasks without text (73%), confirming once again that this AI tool requires context for more accurate conclusions.

The author of this test is convinced that the language results would be improved if ChatGPT received more information about them, and he demonstrated this on a question which required selecting the answer with a vowel from the four provided options. ChatGPT generated only 150-200 words, some of which were in English and Serbian, so after multiple attempts it switched to "classic GPT" i.e., GPT-3, which can generate longer texts due to its fewer character limitations.

GPT-3 tore the essay apart from the beginning, and the aforementioned portal also published a thorough correction of it, in which the professor who made the correction reached the following conclusion: The essay focuses primarily on the work's content. The interpretation of the characters' relationships and the description of the events are extremely precise. Literary-historical and literary-theoretical concepts, textual quotations proving claims, and comparisons to other literary texts (reader experience) are absent. Spelling and grammar rules are strictly adhered to, while syntactic rules are slightly disregarded.

While technology cannot impact the educational system on its own, synergy between technology and educators is required for optimal results. Technology is not a hazard to the educational process or the methods of instruction; it serves to provide possible solutions to specific issues and concerns, and the best results will arise from technology and teacher cooperation.

### IV. EDUCATIONAL COPILLOT – AI UNIT PLANNER

Educational Copilot is an AI tool that is helpful for educators. It allows teachers to create unit plans, presentations, quizzes, handouts, projects, writing prompts, student reports and even generate a parent email within a few minutes. The main objective of its creators [38] is to provide educators with high-quality, AI-generated lesson planning and preparation templates and resources. They target to save educators time and energy so that they can concentrate on what truly matters: their students.

This tool saves teachers a lot of their time mining Google for valuable information, and it is significantly

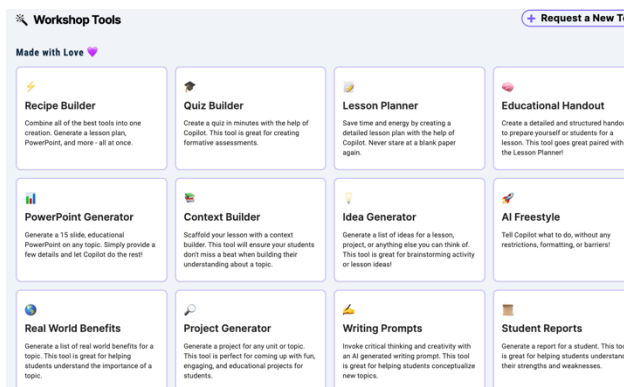


Figure 1. Screenshot of educational CoPilots' available tools at the moment

better used. Copilot, like most other tools, is developed on GPT-3 large language models. This means it is created and trained on massive amounts of data. It interprets the prompts you give it (instructions) to generate outputs. Therefore, the more detailed the prompts, the better the outputs.

The tools in Copilot cannot search the internet, they generate the outputs from training data. This means there might sometimes be incorrect information or misconceptions. The tools in Copilot basically act as automated prompting and formatting interfaces, to save teachers work.

Periodically, Copilot will 'fine-tune' the model based on the input and output data in the system. Over time, this should make it more tailored to educational use.

### A. Evaluation study

We conducted a study to evaluate the quality of educational Copilot and better comprehend teachers' willingness to use it. The primary objective was to assess the instructional quality of created lesson plans and effects of AI on teachers' preparation time and overall satisfaction. This study investigated the hypothesis that a teacher's use of Copilot would increase the teacher's willingness to use AI-supported tool and to evaluate effects on planning and preparation. A modified lesson plan analysis protocol (LPAP) was used as a tool for evaluation [39].

Participants were 34 middle school teachers from four schools and three subjects (math, history, and biology). All participating teachers had at least five years of experience teaching at the middle school level and no prior experience using AI tools in or for the classroom.

### B. Usual lesson plans vs Copilot lesson plans

Using Copilot, we created three lesson plans for each of the listed subjects (Mathematics, Biology, History). The initial data entry consisted of the titles of the currently used textbooks in the classrooms and the curriculum outcomes stated for a specific topic. Presentations, key-questions, handouts, and additional worksheets for homework were included in the lesson plans created for two hours of class time. We provided teachers with a choice between lesson plans developed with Copilot and those developed by other co-experts which were offered in the methodological

manual (non-AI). Teachers were not informed that AI created some of these offered lesson plans. Offered plans were identical in general metadata (title, summary, keywords, etc.) in terms of definition of educational objectives (subject domain, topic domain, general learning outcomes, curriculum standards) and needed follow-up content. Teachers had to evaluate both lesson plans for their subject and decide which one they would prefer to use in their teaching. They filled out a modified LPAP Evaluation Questionnaire which includes 5-point Likert items for evaluating the instructional quality of each lesson in the following categories: educational objectives, educational resources, lesson stages, lesson approaches, lesson evaluation. At the end of the questionnaire there was a selection criterion for proposed lesson plans.

All three lesson plans from the teacher's manual (non-AI) were rated "good" or "very good" by 76% of teachers. Only 4% rated these lesson plans as excellent in regards to their pedagogical value. Lesson plans created using Copilot achieved a slightly better average rating, with 81% of teachers rating them "good" or "very good" and 7% rating these lesson plans as "excellent". Furthermore, when they had to choose which of the two proposed lesson plans they would use in their teaching, 78% stated they would choose the one created using Copilot, 10% those from the manual, and the rest were not sure about their selection.

### C. Effects

The teachers were informed, following the initial selections, that some of the selected lesson plans were created using an AI tool. The Copilot tool was identified and a link to it was provided. Teachers were required to create working materials for classes as required, evaluate the quality of the material and the tool using the same modified LPAD protocol, and list the advantages and disadvantages of the aforementioned. Nearly 90% of teachers stated that the tool is extremely useful and simple to use, 78% that they will use it frequently, and 95% that it is a significant time saver that will give them more time to spend with students and on personal growth. As for disadvantages, all History teachers agreed that the facts were slightly inaccurate and required corrections. Most teachers (63%) find the language barrier to be the greatest disadvantage; inputs can be in any language, but outputs are only in English. It takes time to translate them if they are going to be implemented.

Concerning the perception of using AI in education, alarming percentages have been observed. 74% believe that artificial intelligence is a threat to the educational sector and their professions, and 86% have never used AI in and for teaching.

The values indicate that teachers valued the Copilot tool for generating lesson plans as preferable and that content made and given utilizing AI capabilities is of good to very good instructional and pedagogical quality but that perception of it is not so positive amongst the teacher population.

## V. CONCLUSION

The integration of AI and chatbots, such as ChatGPT and GPT-3 based tools, in education has the potential to

revolutionize the way teachers teach and students learn. However, despite this potential, many teachers are still ignorant of the possibilities of AI and the role it can play in education. This ignorance is a barrier to the effective use of AI in education and is something that must be addressed. By providing training and support to teachers, they can gain the knowledge and skills they need to use AI ethically and effectively in their classrooms. This will not only benefit students by providing them with more engaging and interactive learning experiences but will also help teachers to better understand and make use of the latest technology in education. The use of AI in education is still in its early stages, but as it becomes more widespread, it is important to ensure that teachers have the knowledge and skills they need to use it responsibly and effectively. With the right support and training, teachers can harness the power of AI to transform education for the better.

#### REFERENCES

- [1] B. Marr, "10 Wonderful Examples Of Using Artificial Intelligence (AI) For Good", Forbes, 2020. Accessed on: December 13, 2022. [Online].
- [2] J. McCarthy, "What is artificial intelligence", 2007. Accessed on: January 6, 2023. [Online].
- [3] P. Wang, "On Defining Artificial Intelligence", Journal of Artificial General Intelligence, vol. 10, 2019, pp.1-37.
- [4] J. A. Kulik, J.D. Fletcher, "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review," Review of Educational Research, vol. 86 (1), pp. 42-78, 2015.
- [5] Chocarro, R.; Cortin  as, M.; Marcos-Mata  as, G. "Teachers' attitudes towards chatbots in education: A technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics". Educ. Stud. 1-19. 2021.
- [6] Ischen C., Araujo T., Voorveld H., van Noort G., Smit E. "Privacy concerns in chatbot interactions. Chatbot research and design". CONVERSATIONS 2019. Lecture Notes in Computer Science, vol 11970 (pp. 34-48). Springer, 2020.
- [7] Luo, B.; Lau, R.Y.; Li, C.; Si, Y.-W. "A critical review of state-of-the-art chatbot designs and applications". Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2022.
- [8] Adamopoulou, E.; Moussiades, L. "Chatbots: History, technology, and applications". Mach. Learn. Appl. 2020.
- [9] Okonkwo, C.W.; Ade-Ibijola, A. "Chatbots applications in education: A systematic review". Comput. Educ. Artif. Intell. 2021.
- [10] Fryer, L.; Coniam, D.; Carpenter, R.; L  pus, neanu, D. "Bots for language learning now: Current and future directions". Lang. Learn. Technol. 24, 8-22.. 2020.
- [11] de Barcelos Silva, A.; Gomes, M.M.; da Costa, C.A.; da Rosa Righi, R.; Barbosa, J.L.V.; Pessin, G.; De Doncker, G.; Federizzi, G. "Intelligent personal assistants: A systematic literature review". Expert Syst. Appl. 147. 2020.
- [12] Fryer, L.; Carpenter, R. "Bots as language learning tools". Lang. Learn. Technol. Vol.10, 8-14. 2006.
- [13] Hill, J.; Ford, W.R.; Farreras, I.G. "Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations". Comput. Hum. Behav. Vol. 49, 245-250. 2015.
- [14] Xu, Y.; Wang, D.; Collins, P.; Lee, H.; Warschauer, M. "Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner". Comput. Educ. 161. 2021.
- [15] Ayedoun, E.; Hayashi, Y.; Seta, K. "A conversational agent to encourage willingness to communicate in the context of English as a foreign language". Procedia Comput. Sci. 2015.
- [16] Io, H.N.; Lee, C.B. "Chatbots and conversational agents: A bibliometric analysis". Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10-13 December 2017; IEEE: Piscataway, NJ, USA, pp. 215-219. 2017.
- [17] Shawar, B.A. "Integrating CALL systems with chatbots as conversational partners". Comput. Syst. 21, 615-626. 2017.
- [18] Radziwill, N.M.; Benton, M.C. "Evaluating quality of chatbots and intelligent conversational agents". arXiv 2017, arXiv:1704.04579.
- [19] Haristiani, N. "Artificial Intelligence (AI) chatbot as language learning medium: An inquiry". J. Phys. Conf. Ser. 2019.
- [20] Bibauw, S.; Franc  ois, T.; Desmet, P. "Computer Assisted Language Learning". Routledge: London, UK, pp. 827-877. 2019.
- [21] Bibauw, S.; Franc  ois, T.; Desmet, P. "Dialogue Systems for Language Learning: Chatbots and Beyond". The Routledge Handbook of Second Language Acquisition and Technology; Routledge: Abingdon, UK, pp. 121-135, ISBN 1-351-11758-0.2022.
- [22] Hwang, G.-J.; Chang, C.-Y. "A review of opportunities and challenges of chatbots in education". Interact. Learn. Environ. 2021.
- [23] Dokukina, Irina & Gumanova, Julia. "The rise of chatbots – new personal assistants in foreign language learning". Procedia Computer Science. 169. 542-546. 2020.
- [24] Allouch, M.; Azaria, A.; Azoulay, R. "Conversational Agents: Goals, Technologies, Vision and Challenges". Sensors 2021, 21. 2021.
- [25] Pe  rez-Mari  n, D. "A Review of the Practical Applications of Pedagogic Conversational Agents to Be Used in School and University Classrooms". Digital. 18-33. 2021.
- [26] Huang, W.; Hew, K.F.; Fryer, L.K. "Chatbots for language learning-Are they really useful? A systematic review of chatbot-supported language learning". J. Comput. Assist. Learn. 38, 237-257. 2022.
- [27] Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya. "Improving Language Understanding by Generative Pre-Training". p. 12. 2018.
- [28] Sterling, B. "Web Semantics: Microsoft Project Turing introduces Turing Natural Language Generation (T-NLG)". Wired. ISSN 1059-1028. 2020.
- [29] Brown, T. B.; Mann, B. et al. "Language Models are Few-Shot Learners". Advances in Neural Information Processing Systems. Curran Associates, Inc. 33: 1877-1901.2022.
- [30] Sagar, R. "OpenAI Releases GPT-3, The Largest Model So Far". Analytics India Magazine. 2020.
- [31] Roose, K. "The Brilliance and Weirdness of ChatGPT". The New York Times. 2022.
- [32] Edwards, B. "No Linux? No problem. Just get AI to hallucinate it for you". Ars Technica. 2022.
- [33] Mintz, S. "ChatGPT: Threat or Menace? Are fears about generative AI warranted?". Inside Higher Ed. 2023.
- [34] Seldon, A.; Abidoye, O. "The Fourth Education Revolution: Will Artificial Intelligence Liberate or Infantilise Humanity". The University of Buckingham Press: Buckingham, UK, 2018.
- [35] Holmes, W.; Bialik, M.; Fadel, C. "Artificial Intelligence in Education: Promises and Implications for Teaching and Learning". Center for Curriculum Redesign: Boston, MA, USA, 2019.
- [36] Bushard, B. "Fake Scientific Abstracts Written By ChatGPT Fooled Scientists", Study Finds. Forbes. 2023.
- [37]   idov I. Nlinfo. " Znanstvenik testirao ChatGPT na dr  avnoj maturi iz hrvatskog" Accessed on: February 05, 2023. [Online].
- [38] Educational Copilot, Decatur, Texas, <https://educationcopilot.com/>
- [39] Ndiokubwayo, K. et al. "Lesson plan analysis protocol (LPAP): A useful tool for researchers and educational evaluators". Heliyon 8. 2022. <https://doi.org/10.1016/j.heliyon.2022.e08730>