ISBN: 978-989-8704-51-1 © 2023

"A PIECE OF CAKE" - SUSTAINABLE EDUCATION PRACTICES WITH TRANSFORMER MODELS

Francisco de Arriba-Pérez, Silvia García-Méndez and Francisco J. González-Castaño
Information Technologies Group, atlanTTic, University of Vigo
School of Telecommunications Engineering, Maxwell street, 36310 Vigo, Spain

ABSTRACT

New advances in Artificial Intelligence (AI), particularly transformer language models, have disruptively changed the current education and training practices. The latter advancements have forced a shift towards a more practical and sustainable learning process with the ultimate objective of creating active, high-quality, and continuous educational services. However, human intervention is still essential to keep control over the outcomes of AI-based solutions. Thus, this work contributes with a fully functional and well-designed Conversational Assistant for Knowledge Enhancement (CAKE), which integrates a transformer model in a chatbot platform with empathetic capabilities. The proposed environment is fully immersive and incorporates gamification. Natural Language Processing techniques are exploited through prompts. The most relevant features of CAKE are its flexibility regarding language, areas and topics covered, and the education level of the end users. Results obtained in two evaluation scenarios endorse the performance of the proposed solution and motivate us to continue this research line toward real-time monitoring of smart learning.

KEYWORDS

Artificial Intelligence, Intelligent Chatbot, Natural Language Processing, Prompt Engineering, Smart and Sustainable Education, Transformer Language Models

1. INTRODUCTION

Digitalization has spread in many fields. However, the impact of artificial intelligence (AI) has been even more notable in education and training (Rudolph et al., 2023). The latter process, when applied to the education sector, is transversal through all levels (elementary and high school, university studies, etc.) and seeks to shift from traditional and theoretical practices to more practical ones, promoting at the same time the sustainability of the learning process (Crompton & Burke, 2018). Accordingly, smart education has attracted the attention of the research community, coinciding with the progress on AI-based solutions and the need for a global framework for high-quality training services (Li et al., 2022). The fundamental change now lies in the real-time monitoring of the learning process thanks to these applications and technologies, including the analysis and detection of the emotional state of the students through sentiment and emotion analyses (Cope et al., 2021).

Since the disruptive transformer models were released, education has been subject to challenges that, when not addressed properly, may result in inadequate pedagogical practices. The educational community, researchers, and policymakers are responsible for providing well-designed applications (Baidoo-Anu & Owusu Ansah, 2023). In this line, transformer models like ChatGPT[1] have caused a huge buzz worldwide. These new AI-based solutions are changing business by creating new market opportunities and eliminating those whose Natural Language Processing (NLP) needs are already covered by them (Baidoo-Anu & Owusu Ansah, 2023). The latter disruptive change also applies to the education field. Among the benefits for students and teachers, the workload reduction (e.g., automatic assessment), the real-time monitoring of student knowledge acquisition, and classroom innovation (e.g., feedback mechanisms) deserve to be mentioned (Cope et al., 2021). However, these systems still require human intervention (Kitamura, 2023). Thus, fully functional and well-designed solutions are essential (Parsazadeh et al., 2018).

[1]Available at https://openai.com/blog/chatgpt, April 2023.

The most popular form of integration of the transformer models in commercial solutions is through chatbots (Brown et al., 2020). They are agents that have the capabilities to establish real-time and human-like conversations with the end users through natural language (text or voice) and are appropriate to our purpose due to their simplicity and quick response time (Tallyn et al., 2018). Since their creation, chatbots have been used in education, entertainment, etc. (Rodrigues et al., 2022) for question-answering and task-solving. In the particular case of the education sector, they contribute to creating an active, high-quality, and continuous training environment (Romero-Rodríguez et al., 2020). The future generation of learners, as digital natives, are willing to incorporate these technologies to get an immersive learning experience (Pedro et al., 2018).

Taking advantage of transformers and the chatbot paradigm, this work contributes with a multi-platform and language application that integrates sustainable learning practices, CAKE (Conversational Assistant for Knowledge Enhancement). The main advantage is its flexibility regarding language, areas and topics covered, and the education level of the end users, thanks to NLP techniques applied to prompt engineering.

The rest of the paper is organized as follows. Section 2 reviews relevant works on innovative educational applications. Section 3 describes the proposed solution, while Section 4 provides relevant details about implementation and experimental results. Finally, Section 5 concludes the article and proposes future research.

2. RELATED WORK

In the era of transformer models, education and training have derived towards smart education (Li et al., 2022). Consequently, there exists a wide variety of technology-based solutions applied in the classrooms. A relevant feature of these systems is that they are ubiquitous. Thus, users apply them both inside and outside of these environments (*e.g.*, in remote learning). Notwithstanding its easy, flexible, and ubiquitous nature, AI-based solutions remove significant barriers, most importantly, money and time, without neglecting the learner's needs (Kumar Basak et al., 2018). It is the case with popular apps for educational purposes (Crompton & Burke, 2018). In fact, educational apps rank second in popularity on Google Play (Singh & Suri, 2022). Since the learning process is composed of high-difficulty tasks like comprehension and problem-solving (Malik et al., 2020), these kinds of systems become essential to support both teachers and students. In this line, Intelligent tutoring systems (ITS) are educational applications based on AI which provide intelligent support for students in an adaptive and personalized manner, aka personalized adaptive learning (PAL) (Peng et al., 2019; Zawacki-Richter et al., 2019). Among the applications of ITS, one of the most popular consists of automatically providing feedback to the students, for example, about their writing skills (Garcia-Gorrostieta et al., 2018). Furthermore, when combined with traditional educational and training approaches, technology-enhanced learning ensures promising learning outcomes and user satisfaction (Mota et al., 2018). Note that the use of multi-modal sources of information is highly appropriate in smart education (Li et al., 2022).

AI-based solutions provide relevant benefits for the whole educational community (*e.g.*, workload reduction and real-time training for teachers and students, respectively) (Cope et al., 2021). Accordingly, ChatGPT has become a common NLP practice in education (Baidoo-Anu & Owusu Ansah, 2023). Apart from question generation, transformer models can generate distracting answers in a multiple-choice quiz use case (Rodriguez-Torrealba et al., 2022). It is the case of EduQuiz[2], a quiz generation system for the reading comprehension use case based on GPT-3 that creates correct and distracting answers. However, ChatGPT and transformer models, in general, still require human intervention (Kitamura, 2023). Moreover, when it comes to sustainability, these models need to be energy-efficient in terms of maintenance and training costs (algorithms, data processing, representation, storage, etc.). The data mining must be managed ethically and following regulatory compliance (Kasneci et al., 2023).

Regarding the use of transformer models, Bhat et al. (2022) integrated GPT-3 in an assessment application. The questions generated by the system were manually evaluated in terms of their usefulness with positive results. Moreover, Bulut et al. (2022) used GPT-2 for reading passage generation in a reading comprehension assessment solution. Note that the authors pointed out that further research is still needed to adapt the outcome of the AI-based model for students. Conversely, MacNeil et al. (2022) used GPT-3 to generate descriptions about code programming, while Moore et al. (2022) presented another assessment system, in this case, for chemistry. A more sophisticated approach was followed by Topsakal et al. (2022), who designed a language

[2]Available at https://github.com/RamonDijkstra/EduQuiz, April 2023.

learning chatbot with ChatGPT, augmented reality, and gamification. Finally, Zong et al. (2022) used GPT-3 for math problems about two linear equations (word problem classification, equation extraction, and word problem generation) with competitive results.

Based on the aforementioned works, chatbots seem appropriate for cooperative and self-regulated learning since they are easy to use and promote engagement (Smutny & Schreiberova, 2020). Accordingly, Ethnobot (Tallyn et al., 2018) has question generation capabilities but is limited to the ethnographic field. Aligned with our work, Ng et al. (2020) developed WeChat, which became the most popular assistant for collaborative education among Chinese learners. More recently, Liu (2022) presented a language learning application combining collaborative approaches with gamification in a competitive learning environment.

Summing up, this work contributes with an innovative application combining ChatGPT with NLP for the teachers and researchers to keep control over the outcome of the model, avoiding inadequate pedagogical practices, named CAKE. Regarding the pedagogical basis of the proposed solution, it seeks to contribute to the following educational perspectives: cognitive (enhance reasoning during the learning process), emotional (promote positive feelings during training), behavioral (improve the results of the learning process), and contextual (encourage long-term training inside and outside the classroom). Note that the cognitive and behavioral perspectives are related to the application of NLP techniques over the content of the dialogue, while the emotional one is handled with emotion analysis. Finally, since the solution can be accessed anywhere and anytime, the users can benefit from the contextual perspective.

3. PROPOSED METHOD

The CAKE solution is shown in Figure 1. It has multi-language support and was designed bearing in mind students and teachers. Moreover, it provides automatic adjustment of the difficulty level of the proposed tasks.

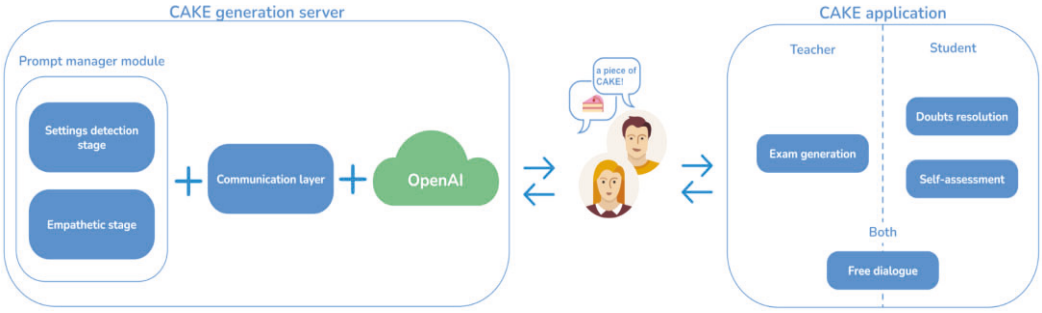

Figure 1. CAKE application for sustainable education and training

3.1 CAKE Generation Server

It is responsible for text generation. Note that NLP is exploited to both control the quality of the outcome and to ensure that an appropriate dialogue is established between students and teachers with the proposed solution. In addition, the generation server is in charge of past utterances management since the transformer models, and also ChatGPT, lack memory. Note that a cache of the six last interactions was established motivated by the cost-efficiency trade-off (for the free dialogue functionality, see Section 3.2).

3.1.1 Prompt manager module

This is the main module of the system. It defines the prompts of the model used for dialogue contextualization about a certain educational topic. The latter prompts are created using natural language and serve to direct the functioning of Open AI's ChatGPT model. Note that CAKE automatically adapts to the language of the users.

In the end, this module enhances the outcome of the system with empathetic utterances through feedback and personalization processes. For each of the following stages, a detailed prompt is defined in Table 1.

- **Empathetic stage**. There exist two prompts (7 and 8 in Table 1) related to the empathetic nature of the CAKE system. In order to promote engagement, the user utterances are analyzed to evaluate the student's

level of knowledge on the subject. The difficulty level is automatically handled based on the number of correct/incorrect answers provided by the students to prevent frustration. Moreover, for each question correctly answered and depending on the degree of difficulty, a score is assigned (low: 10 points, medium: 20 points, high: 30 points). A badge will be granted when a user reaches the 100 points threshold. The latter rewards will be stored in their Hall of Fame.

- **Settings detection stage.** It consists of detecting: (i) the language, (ii) the role of the user (student or teacher), (iii) the functionality (see Section 3.2), and (iv) the area and topic. For the specific case of the exam generation functionality, the number of questions and difficulty level are extracted from the teacher's utterance. Otherwise, a random number of medium-level questions is generated.

3.1.2 Communication layer

It acts as an abstraction layer between the CAKE generation server and Open AI's ChatGPT model. The communication layer defines a structure based on tags to identify the sender (CAKE or the end user) and the context. The tags used are (see Table 1):

- **"System" tag**. It sets the application context in terms of (i) language, (ii) the sender (CAKE or user) and role (student or user), (iii) functionality, (iv) the area and topic, and (v) the number of questions and difficulty level. The empathetic nature of CAKE (see prompts 7 and 8 in Table 1) is also configured with the "System" tag.
- **"CAKE" and "User" tags**. These tags identify the utterances of the CAKE system or the end user, respectively.

3.2 CAKE Application

The client application was developed following the popular instant messaging approach of applications like WhatsApp. Note that the chat can be ceased with the command "terminate". The following functionalities are available in the CAKE application:

- **Free dialogue (F)**. This functioning mode is intended for both students and teachers for them to search for information or even generate or solve more sophisticated context-dependent questions.
- **Exam generation (E)**. It consists in generating exam questions for the teacher role. There are a total of 3 difficulty levels: low, medium, and high.
- **Doubts resolution (D)**. This functionality is intended for students to solve their doubts. Questions about exercises and problems can be formulated to obtain the correct answer and a detailed explanation.
- **Self-assessment (S)**. It was designed for students to practice for an exam, with empathetic feedback.

Table 1. Prompts used

ID	Detection	Prompt	Temp.	Result
1	Language	SYSTEM: If I told you this phrase/word: "X", in what language is it? Return {"language":""}. If it is not a language, return "not_applicable".	0	LANG
2	Role	SYSTEM: If I told you "X", what would you say I am? A teacher or a student? Answer with one word.	0	ROLE = T/S
3	Functionality	SYSTEM: I am a [ROLE]. If I told you "X", what would you say I have chosen from the available options: [AVAILABLE OPTIONS]? Answer with only one of the options from the list.	0	MODE = F/E/D/S
4	Number of questions, area, topic, level	SYSTEM: If I told you "X", what number, area, topic, difficulty level am I referring to? Answer in English using only the next format: {"number":"", "area":"", "topic":"", "difficulty_level":""}.	0	NUM, AREA, TOPIC, LEVEL
5	Question generation	SYSTEM: I am a [ROLE] with advanced knowledge, let's talk in [LANG]. I want to generate [NUM] exam questions in [LANG] about [AREA] and [TOPIC] with [LEVEL] difficulty.	1	QUESTION
6	Answer generation (AG)	SYSTEM: I am a [ROLE] with advanced knowledge, let's talk in [LANG]. I want you to answer this question [QUESTION].	1	ANSWER

7	AG + empathy	SYSTEM: I am a [ROLE] with advanced knowledge, let's talk in [LANG]. For the question [QUESTION], I have answered [ANSWER]. Tell me in [LANG] if it is correct, partially correct or incorrect. Try to detect in my answer frustration [0.0 to 1.0]. Always explain to me why it is incorrect and the correct answer. Adapt the answer to a student with [LEVEL] knowledge (maximum 100 words). Use this format: {"eval":"", "explanation":"", "frustration":"", is_correct:"0.0/0.5/1.0"}.	1	EVAL, EXPLANATION FRUSTRATION IS_CORRECT
8	Free dialogue + empathy	SYSTEM: The assistant speaks only in [LANG]. The assistant is dynamic (it never repeats the same thing twice), creative, intelligent and very kind. The assistant always asks a question when it finishes speaking. USER: I am a [ROLE] with advanced knowledge. Let's talk in [LANG]. I wanted to talk about "X".	1	RESPONSE

4. EXPERIMENTAL RESULTS

In this section, the implementations used are detailed to ensure reproducibility. Moreover, the results obtained are presented and discussed.

4.1 CAKE Generation Server

It was programmed in Python (version 3.8) using a Flask server (version 2.2.2) with a Gunicorn traffic balancer (version 20.1.0). The ChatGPT model (version 3.5) provided by OpenAI was used.

4.1.1 Prompt manager module

For the ChatGPT model, the *temperature* configuration is shown in Table 1 for each stage in the CAKE scheme. Note that this parameter allows adjusting the randomness of the generated texts (1 stands for random, and 0 for deterministic). Thus, it contributes to creating human-like dialogues.

4.1.2 Communication layer

It makes requests from our Flask server to the ChatGPT model using the OpenAI Python Library[3]. This library allows us to interact with the model using a REST API.

4.2 CAKE Application

The Flutter programming language was used to make it multi-platform, *i.e.*, to be integrated with Android and iOS smartphones and tablets, and also as a web application with macOS and Windows operating systems. For the instant messaging design, we used the Flyer Chat library[4].

4.3 Case Study

In order to evaluate the effectiveness of the proposed solution, two evaluation scenarios were designed.

Scenario 1. It evaluates the exam generation and self-assessment functionalities with the research questions in Table 2. Particularly, CAKE is used to generate 150 questions: 5 questions about 5 topics with 3 levels of difficulty (low, medium, high) in 2 languages (English and Spanish). The selected areas and topics are natural sciences - the cell, social sciences - history, mathematics - derivatives, language - grammar, and arts - painting.

Scenario 2. Explanatory answers are generated for each of the questions in the first scenario. The research questions used for evaluation are also displayed in Table 2.

[3]Available at https://github.com/openai/openai-python, April 2023.
[4]Available at https://github.com/flyerhq/flutter_chat_ui, April 2023.

Table 2. Research questions used for evaluation

Number	Scenario	Research question
1		Is the language correct?
2	1	Are the questions related to the selected area and topic?
3		Are the questions appropriate to the expected level of difficulty?
4		Is the language correct?
5	2	Is the answer related to the selected area and topic?
6		Is the answer appropriate to the expected level of difficulty?

4.4 Evaluation Results

The evaluation scenarios were composed of 300 utterances that were manually annotated by a teacher and an NLP researcher. Table 3 details the evaluation results for the two aforementioned scenarios. The best results are obtained for English, as expected since this is the original language of the ChatGPT model. Regarding area and topic, and difficulty level, the results are similar between the two languages in the two evaluation scenarios.

Table 3. Percentage of affirmative answers of the two annotators for the English and Spanish languages

Language	Annotator	RQ1	RQ2	RQ3	RQ4	RQ5	RQ6
English	1	100.00	100.00	88.00	100.00	100.00	88.00
	2	100.00	100.00	93.33	100.00	100.00	90.67
	Avg.	**100.00**	**100.00**	**90.67**	**100.00**	**100.00**	**89.34**
Spanish	1	89.93	98.67	86.67	98.65	98.67	90.67
	2	88.00	98.67	88.00	98.65	98.67	88.00
	Avg.	**88.96**	**98.67**	**87.34**	**98.65**	**98.67**	**89.34**

5. CONCLUSION

Smart education has become a relevant field of research with its own entity driven by the latest AI advances. Among them, the most disruptive change is associated with the transformer language models like OpenAI's ChatGPT. These AI-based solutions exploit the practical perspective of the learning process to promote the sustainability of real-time and continuous, high-quality, and active learning. Even though smart educational systems provide adaptive and personalized support for students and are also able to detect emotions in their speech, human intervention is still required to take the most of them for the global education community.

Consequently, we take advantage of the willingness of the new generation of educators and learners towards technology adoption to propose CAKE, a functional and well-designed solution that integrates transformers in the chatbot paradigm, combined with NLP techniques through prompt engineering. Its sustainability lies in its multi-platform and language-independent features. Moreover, it is flexible regarding areas and topics covered and the education level of the users. CAKE has empathetic capabilities and promotes engagement due to the easy, self-regulated learning process and thanks to the gamification techniques used. Note that it has been designed for both students and teachers with several functionalities (free dialogue, exam generation, doubts resolution, self-assessment). The generation server is composed of the prompt manager with empathy handling, the communication layer, and the OpenAI's model. Figure 2 shows the proposed solution in the self-assessment use case. The results obtained were evaluated in two scenarios in two different languages (English and Spanish). The first analyzes the exam generation and self-assessment functionalities for the teacher and student roles, respectively; while the second focuses on the doubts resolution and self-assessment functionalities for the students. Both the questions (scenario 1) and answers (scenario 2) generated by CAKE were manually annotated, obtaining promising results regarding language, area and topic, and level of difficulty.

In future work, we plan to apply data mining techniques for the system to learn from the end users to create more personalized learning experiences. Moreover, prompt engineering will be exploited to enhance the management of CAKE regarding non-English and mathematical languages. Ultimately, sustainability issues of transformer models, like the carbon footprint impact and governance issues, will be discussed.

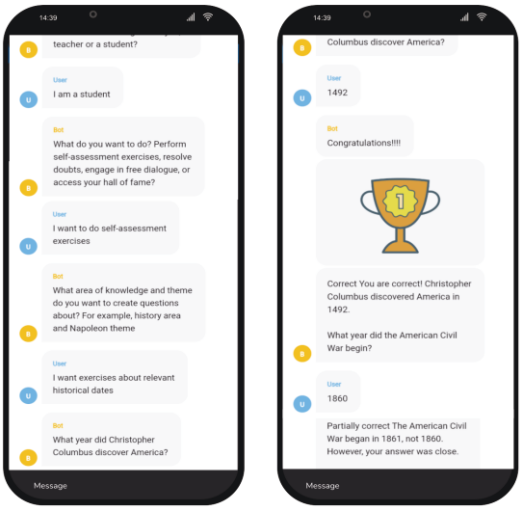

Figure 2. System application example

ACKNOWLEDGEMENT

This work was partially supported by Xunta de Galicia grants ED481B-2021-118, ED481B-2022-093, and ED431C 2022/04, Spain; and Ministerio de Ciencia e Innovación grant TED2021-130824B-C21, Spain.

REFERENCES

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *SSRR*, 1-20. 10.2139/ssrn.4337484

Bhat, S., Nguyen, H. A., Moore, S., Stamper, J., Sakr, M., & Nyberg, E. (2022). Towards Automated Generation and Evaluation of Questions in Educational Domains. *Proceedings of the International Conference on Educational Data Mining*, 701-704. 10.5281/zenodo.6853085

Brata, K. C., Brata, A. H., & Lukman, E. P. (2019). Hanasu: Interactive Japanese Language M-Learning Application to Support Listening and Speaking Exercise. *Proceedings of the International Conference on Education and Multimedia Technology*, 311-315. 10.1145/3345120.3345155

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., … & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.

Bulut, O., & Yildirim-Erbasli, S. N. (2022) Automatic story and item generation for reading comprehension assessments with transformers. *International Journal of Assessment Tools in Education*, 9, 72-87. 10.21449/ijate.1124382

Cope, B., Kalantzis, M., & Searsmith, D. (2021). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 53(12), 1229-1245. 10.1080/00131857.2020.1728732

Crompton, H., & Burke, D. (2018). The use of mobile learning in higher education: A systematic review. *Computers & Education*, 123, 53-64. 10.1016/j.compedu.2018.04.007

García-Gorrostieta, J. M., López-López, A., & González-López, S. (2018). Automatic argument assessment of final project reports of computer engineering students. *Computer Applications in Engineering Education*, 26(5), 1217-1226. 10.1002/cae.21996

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 1-9. 10.1016/j.lindif.2023.102274

Kitamura, F. C. (2023). ChatGPT is shaping the future of medical writing but still requires human judgment. *Radiology*, 230171. 10.1148/radiol.230171

Kumar Basak, S., Wotto, M., & Bélanger, P. (2018). E-learning, M-learning and D-learning: Conceptual definition and comparative analysis. *E-Learning and Digital Media*, 15(4), 191–216. 10.1177/2042753018785180

Li, N., Shen, Q., Song, R., Chi, Y., & Xu, H. (2022). MEduKG: A Deep-Learning-Based Approach for Multi-Modal Educational Knowledge Graph Construction. *Information*, 13(2), 1-18. 10.3390/info13020091

Liu, I.-F. (2022). Gamified mobile learning: effects on English learning in technical college students. *Computer Assisted Language Learning*, 1-24. 10.1080/09588221.2022.2080717

MacNeil, S., Tran, A., Mogil, D., Bernstein, S., Ross, E., & Huang, Z. (2022). Generating diverse code explanations using the GPT-3 large language model. *Proceedings of the ACM Conference on International Computing Education Research*, 2, 37-39. 10.1145/3501709.3544280

Malik, S. I., Al-Emran, M., Mathew, R., Tawafak, R. M., & Alfarsi, G. (2020). Comparison of E-Learning, M-Learning and Game-based Learning in Programming Education - A Gendered Analysis. *International Journal of Emerging Technologies in Learning*, 15(15), 133-146. 10.3991/ijet.v15i15.14503

Moore, S., Nguyen, H. A., Bier, N., Domadia, T., & Stamper, J. (2022). Assessing the Quality of Student-Generated Short Answer Questions Using GPT-3. *Proceedings of European Conference on Technology Enhanced Learning*, 243-257. 10.1007/978-3-031-16290-9_18

Mota, J. M., Ruiz-Rube, I., Dodero, J. M., & Arnedillo-Sánchez, I. (2018). Augmented reality mobile app development for all. *Computers & Electrical Engineering*, 65, 250-260. 10.1016/j.compeleceng.2017.08.025

Ng, S. F., Azlan, M. A. K., Kamal, A. N. A., & Manion, A. (2020). A quasi-experiment on using guided mobile learning interventions in ESL classrooms: Time use and academic performance. *Education and Information Technologies*, 25, 4699-4719. 10.1007/s10639-020-10191-7

Oyelere, S. S., Suhonen, J., Wajiga, G. M., & Sutinen, E. (2018). Design, development, and evaluation of a mobile learning application for computing education. Education and Information *Technologies*, 23, 467-495. 10.1007/s10639-017-9613-2

Parsazadeh, N., Ali, R., & Rezaei, M. (2018). A framework for cooperative and interactive mobile learning to improve online information evaluation skills. Computers & Education, 120, 75-89. 10.1016/j.compedu.2018.01.010

Pedro, L. F. M. G., Barbosa, C. M. M. de O., & Santos, C. M. das N. (2018). A critical review of mobile learning integration in formal educational contexts. *International Journal of Educational Technology in Higher Education*, 15, 1-15. 10. 10.1186/s41239-018-0091-4

Peng, H., Ma, S., & Spector, J. M. (2019). Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment. *Smart Learning Environments*, 6(1), 1-14. 10.1186/s40561-019-0089-y

Rodrigues, C., Reis, A., Pereira, R., Martins, P., Sousa, J., & Pinto, T. (2022). A Review of Conversational Agents in Education. *Proceedings of the International Conference on Technology and Innovation in Learning, Teaching and Education*, 461-467. 10.1007/978-3-031-22918-3_37/COVER

Rodriguez-Torrealba, R., Garcia-Lopez, E., & Garcia-Cabot, A. (2022). End-to-End generation of Multiple-Choice questions using Text-to-Text transfer Transformer models. *Expert Systems with Applications*, 208, 1-12. 10.1016/j.eswa.2022.118258

Romero-Rodríguez, J.-M., Aznar-Díaz, I., Hinojo-Lucena, F.-J., & Cáceres-Reche, M.-P. (2020). Models of good teaching practices for mobile learning in higher education. *Palgrave Communications*, 6(1), 1-7. 10.1057/s41599-020-0468-6

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6(1), 1-22, 10.37074/jalt.2023.6.1.9

Singh, Y., & Suri, P. K. (2022). An empirical analysis of mobile learning app usage experience. *Technology in Society*, 68, 1-14. 10.1016/j.techsoc.2022.101929

Tallyn, E., Fried, H., Gianni, R., Isard, A., & Speed, C. (2018). The Ethnobot: Gathering ethnographies in the age of IoT. *Proceedings of the Conference on Human Factors in Computing Systems*, 1-13. 10.1145/3173574.3174178

Topsakal, O., & Topsakal, E. (2022). Framework for A Foreign Language Teaching Software for Children Utilizing AR, Voicebots and ChatGPT (Large Language Models). *The Journal of Cognitive Systems*, 7(2), 33-38. 10.52876/jcs.1227392

Troussas, C., Krouska, A., & Sgouropoulou, C. (2020). Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education. *Computers & Education*, 144, 1-18. 10.1016/j.compedu.2019.103698

Wang, Z., Lan, A., & Baraniuk, R. (2021). Math Word Problem Generation with Mathematical Consistency and Problem Context Constraints. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 5986-5999. 10.18653/v1/2021.emnlp-main.484

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education - where are the educators?. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. 10.1186/s41239-019-0171-0

Zong, M., & Krishnamachari, B. (2022). Solving math word problems concerning systems of equations with GPT-3. *Proceedings of the AAAI Symposium on Educational Advances in Artificial Intelligence*, 1-8.