AN ARTIFICIAL INTELLIGENCE VIRTUAL TRAINER TO SERVE THE UNDERSERVED AND MAKE THEM EMPLOYABLE

D. Aditya[1], P.C.J. Otermans[1], M. Pereira[2]

[1]Otermans Institute (UNITED KINGDOM)
[2]London Metropolitan University (UNITED KINGDOM)

Abstract

The aim of Otermans Institute (OI) is to educate underserved populations with the main purpose of making them employable. One underserved population is refugees and internally displaced persons in United Nations (UN) camps in Iraq and with the introduction to and access to internet, the possibility of reaching these camps is possible. As a result of this introduction of internet access, OI has developed a conversational artificial intelligence (AI) virtual trainer that is primarily designed to teach employability skills. The vision is to deploy this conversational AI virtual trainer in camps to enhance access to learning. The aim of the current user study is to gather users' perceptions of a conversational AI virtual trainer. A total of four users were recruited in a UN refugee camp in Iraq. All users had engaged with the conversational AI virtual trainer and then filled in a survey consisting of four questionnaires to gage their perceptions of the conversational AI virtual trainer. The users rated the conversational AI virtual trainer as competent, friendly, socially present, socially attractive and trustworthy. However, users rated that the conversational AI virtual trainer was usable but could be improved. Future work will delve into methods of improving the usability of the conversational AI virtual trainer by conducting focus groups or interviews to gather information on how to improve. The implications of this research are global with an aim to provide access to learning for all populations regardless of background and availability of human trainers.

Keywords: Artificial intelligence, digital training, digital education, EdTech, virtual training.

1 INTRODUCTION

Artificial Intelligence (AI) can be defined as an area within computer science that is committed to solving cognitive problems associated with human intelligence such as problem solving, learning and pattern recognition. A more applied definition of AI would be that it is the development of machines capable of performing normal everyday tasks associated with human intelligence such as visual perception, speech recognition, decision-making and language translation. AI has made itself known in the educational domain and is a growing trend for the use of and improvement of access to education.

Otermans Institute (OI) is a global company which aims to upskill underserved populations with the objective of making them employable. OI is providing a possible and likely solution physically, digitally and using AI. From remote schools, foundation-run organisations to United Nations camps in Iraq, OI has supported over 30,000 underserved learners to date. OI is aiming to upskill a vast number of learners by 2025 to make them more employable. To do this, OI has created a conversational AI virtual trainer that can upskill learners in these populations at scale in their own time and based around individual learning and training needs. This solution called OI AI will offer a conversational AI virtual trainer providing continuous upskilling for such learners. With smart phone and internet access now increasing in such camps, the potential of it upskilling internally displaced people and refugee learners is massive especially when millions of people are displaced by violence or war globally.

Coupled with the need for a conversational AI virtual trainer in these populations, the literature also notes an increase in demand for educational lessons and access to learning [21]. Digital technologies have become an integral part of life and have changed how we look for information, how we communicate with each other and how we conduct ourselves. As a result of this introduction of digital technology, the education landscape is undergoing a fundamental change; one in which students will have classroom facilitators who are virtual facilitators, virtual trainers or virtual tutors. Digital technologies have already started to be incorporated into the curriculum of a number of courses which include Massive Open Online Courses (MOOCs) for online study such as coursera.com. Simultaneously, there are advancements in Augmented Reality (AR), Virtual Reality (VR) and AI each of which have been placed in the education domain.

A more specific form of AI that relates to the education domain is an Intelligent Tutoring System (ITS). An ITS can be described as a system that is designed based on models of learning in human-to-human

tutoring [1–3]. The development of ITS has created a space where learners have easier and faster access to learning. In other words, a teaching agent plays the role of human trainers and provides support in the administration of a curriculum. Typically, these systems have three components: the domain model, a student model, and a teaching model. However, to date these systems are typically developed requiring complex systems and knowledge of programming. This is not accessible to both educational institutions and students. There have been a number of ITS systems developed in a number of disciplines including physics [4], mathematics [5,6], language [7,8], computer science [9,10], medicine [11] and others. Examples of ITS systems include The AutoTutor [12], Why2-Atlas [4] and Beetle II system [4]. The authors acknowledge that a system has not been developed for employability skills.

Generally, people perceive conversational agents as friendly, trustworthy and safe when they are able to communicate intellectually with a human. In addition, the way in which an ITS is perceived has been found to predict the likelihood of users continuing their use of the system [13,14]. A research study was conducted to explore and understand whether students accept ITS systems in terms of the Technology Acceptance Model. A total of 38 students in higher education were recruited to engage with the ITS system for four weeks in total. A total of 15 students were recruited for a focus group to further explore their experiences. The results showed that 64% of students found the system to be useful, 15% perceived ease of use and 21% reported intention to use the system [13].

The role of conversational agents in the classroom is essential to consider in education. While teaching agents have had an increase in use in the last year, the way in which they are designed should be specific to the user. Especially the content should be specific to the user. End-user requirements are limited for the space of UN refugee camps and internally displaced people which is a vast amount of people globally (see Norwegian Refugee Council, 2019). The aim of the current research is to evaluate a virtual trainer (ITS) with a sample of this population to explore means of improvement.

2 METHODOLOGY

This study was a pilot study. The study and the lesson were conducted in English; however, English was not the participants' first language. Therefore, this study is a pilot study to assess potential issues with understanding English and the instructions provided. The total participants recruited in this study are deemed sufficient in the literature for a pilot user study [17].

2.1 Participants

A total of four participants took part in the study who lived in a United Nations Refugee camp in Iraq. Ethical approval was given by the College of Health, Medicine and Life Science Research Ethics Committee, Brunel University London. English was not the participants' first language. OI already provides weekly digital lessons to this refugee camp via online meeting platform Zoom; as a result, access to participants was not an issue. Before the user study was conducted, the camp leader was approached with the idea to evaluate the conversational AI virtual trainer with a small sample of students and written permission was given.

2.2 Materials

2.2.1 Conversational AI Virtual Trainer

The conversational AI virtual trainer was developed using Deepfake which is a fake video created using digital software, machine learning and face swopping [15]. Deepfakes are computer-created artificial videos in which videos are combined to create new footage that depicts evens, statements or actions that did not happen. Deepfakes rely on artificial neural networks that recognise patterns in data. In total two iterations were conducted. The first iteration had a structured dialogue flow which designed the lesson revision structure from start to finish. Once this was tested and a smooth dialogue flow was established, a separate open domain dataset was added to build conversational ability to address questions and comments from students that were outside the dialogue flow. The second iteration used unsupervised techniques. The machine learning techniques utilized to assist in the development include Natural Language Understanding (NLU) and CORE. NLU assists computers in understanding and interpreting language by breaking down elements of speech [16]. The primary function of this was to handle intents and entities which are structured pieces of information inside a user message. Common examples of entities include locations, names of organisations and prices. The CORE component handled fulfillment or custom actions; an action that can run any code you want. This can be used to make an API call or to query a database. The second interaction is displayed in Figure 1.

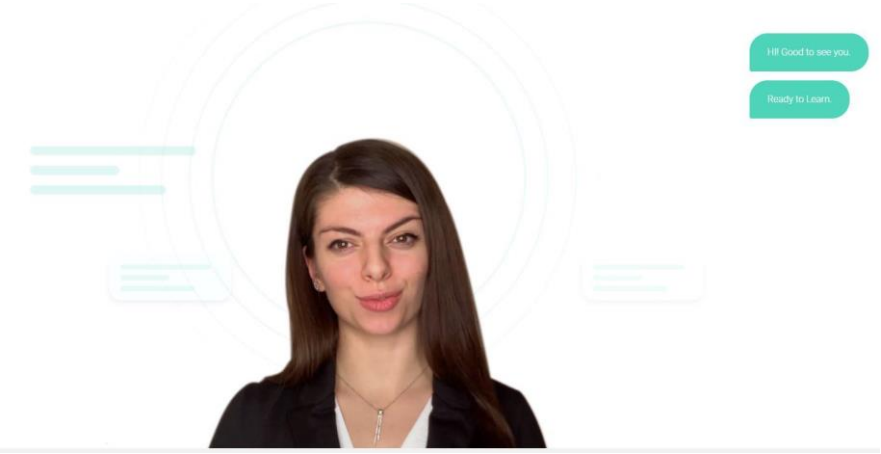

Figure 1. Virtual Trainer (2nd Iteration).

She was programmed to greet users and ask them if they were happy to begin learning. If users interacted with her without starting the learning process, she interacted with them like answering their questions and prompted them to start the lesson. This can be seen below in Figure 2.

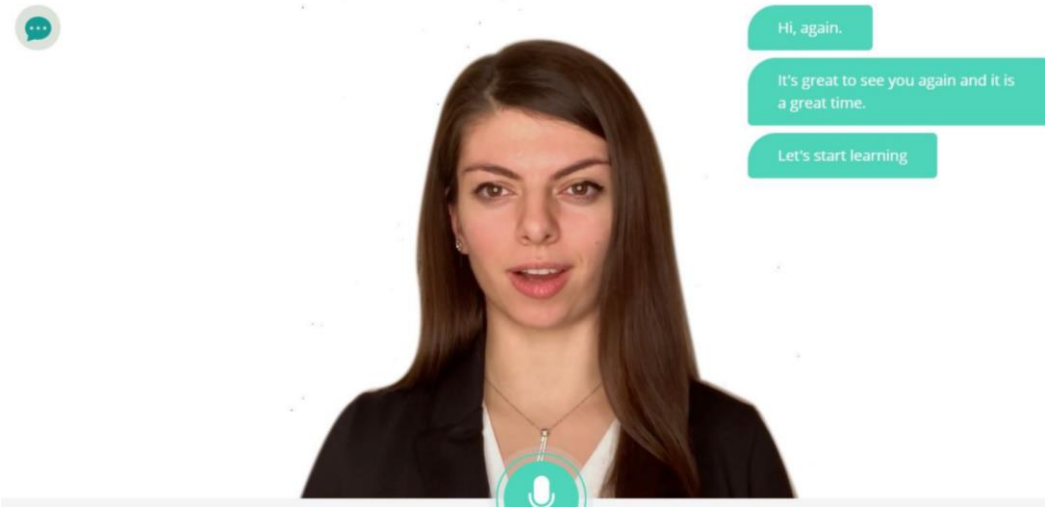

Figure 2. Virtual Trainer Start Questions.

Once learning mode was enabled by learners by stating or typing in ‘let’s learn’ she started the lesson. The lesson was on goal setting and the core concept of the SMART methodology to set goals. She started by asking learners about what they thought goals were and moved onto asking them about each element of SMART: Specific, Measurable, Achievable, Realistic and Time-bound. Once the lesson ended, she gave an overview of the correct answers.

Due to low academic and English spoken abilities of the users, as they have been displaced from their homes and are non-native English speakers, the probability of them giving incorrect answers to the questions asked by her was high. Hence, she has been programmed to encourage users when they made mistakes. She is also programmed to encourage users who may get distracted to keep focus on the lesson until it is completed. These encouragements can be seen in Figure 3. Such distractions were evident as users showed a lot of excitement to interact with her as soon as they got access to her.

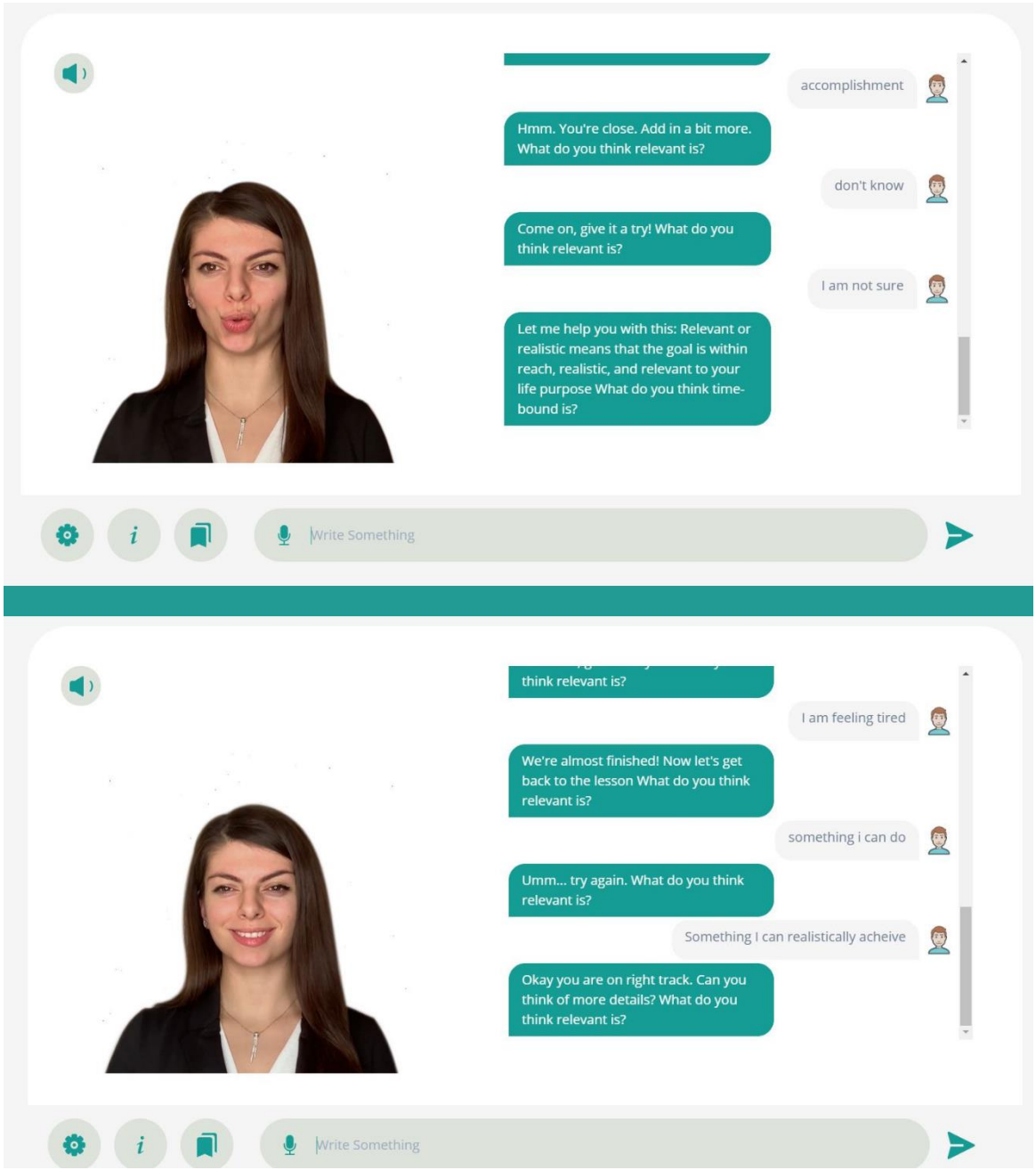

Figure 3. Conversational encouragement example.

2.2.2 Questionnaires

Participants were asked to complete an online survey using Qualtrics consisting of four questionnaires to measure how they perceived the conversational AI virtual trainer. The questionnaires included: 1) System Usability Scale, 2) Competence and Warmth Scales, 3) Social Attractiveness and Trustworthiness Scale, and 4) Social Presence Measure.

System Usability Scale. The system usability scale (SUS) was used to assess how 'usable' the conversational AI virtual trainer was [17]. Usability of the conversational AI virtual trainer can be described as how easy to use and understand the receiver perceives the conversational AI virtual trainer. The SUS is a 10-item questionnaire using a 5-point Likert scale for each item that range from (1) 'strongly disagree' to (5) 'strongly agree'. This scale is commonly used to assess the usability of a range of systems [18]. A high numeric value produced by ratings suggests that the system is usable. The benchmark for an unusable system is below 68 and a rating below 50 is cause for concern [18,19].

Additional questions were added to evaluate the usability of the conversational AI virtual trainer which included:

a) To what extent do you think the virtual trainer is understandable from a trainee perspective?
b) How easy do you think it might be to use this virtual trainer in the future?

Question a) was rated on a 5-point Likert scale ranging from (1) 'very difficult to understand' to (5) 'very easy to understand'. Question b) was rated on a 5-point Likert scale ranging from (1) 'very difficult' to (5) 'very easy to use' in the future.

Competence and Warmth Scales. The competence and warmth scale (CSW) [20] was used to investigate whether the participants thought that the conversational AI virtual trainer was both competent and warm. The competence and warmth scale contains 9 items where 5 items measure competence (competent, confident, independent, competitive and intelligent) and 4 items which measure warmth (tolerant, warm, good natured and sincere). The reliability for the competence scale has previously been recorded between .85 - .90 [20]. Warmth was detected with .82. The higher the score the higher the rated competence or warmth.

Social Attractiveness and Trustworthiness. The Social Attractiveness and Trustworthiness scale (SATS) [21] was used to measure the participants perception of the conversational AI virtual trainer's social attractiveness and trustworthiness. The scale is comprised of 6 items 4 of which relate to social attractiveness (likeable, sociable, pleasant and friendly) and 2 of which relate to trustworthiness (trustworthy and reliable). Social attractiveness was reported as reliable in the literature with a Cronbach's α value of .88 and trustworthiness. The Cronbach's α has not yet been reported for trustworthiness. The higher the score the higher the rated social attractiveness or trustworthiness.

Social Presence Measure. The Social Presence Measure (SPM) was used to investigate how users perceived the conversational AI virtual trainers' social presence following an interaction [14]. The scale is comprised of 5 items each are rated on a 5-point Likert scale ranging from (1) 'not at all' to (5) 'very strong'. Sample questions include *"I felt that the trainer was present*" and *"I felt that the trainer was very detached in her interaction with me"* (negatively loaded). The test has been proven to be reliable in previous literature with a Cronbach α value of .72 [14]. The higher the score the higher the rated social presence.

2.3 Procedure

On the day, users (students) had been given a simple lesson on 'What are SMART goals?' which lasted 30 minutes. An extra 15 minutes were given to account for set-up, connection issues. OI begins their training with a lesson on goal setting. This helps learners set expectations for the upcoming lessons and to set their own goals of what they wish to achieve from the programme. The learning outcomes of the session was to get students to understand what goals are, understand the importance of setting goals, understand what SMART goals are, and be able to set their own SMART goals.

Following this session, participants were then given the opportunity to engage with the conversational AI virtual trainer whereby she would ask students questions which pertained to the lesson (see section 2.2.1). Participants had engaged (see section 2.2.1) with the conversational AI virtual trainer for 20 minutes. Following this, participants were sent a link to Qualtrics which contained a participant information sheet and were given the opportunity to ask questions before the consent form was signed. All participants had given consent to take part in the study. Then, participants filled in the questionnaires in the following order: SUS, CSW, SATS and SPM. Then participants were given a debrief on Qualtrics and were asked if they had any further questions. Participants were thanked for their participation in the study.

3 RESULTS

The results from the questionnaires revealed some concerns over the usability of the conversational AI virtual trainer. The users of the study found that the conversational AI virtual trainer was below benchmark, set at 68, for the usability of a program. However, the users found the conversational AI virtual trainer to competent, tolerant, have a sense of social presence when interacting with the conversational AI virtual trainer and found the trainer to be trustworthy. The results for all users can be seen in Table 1 below.

Table 1. User Evaluation of Conversational AI Virtual Trainer.

	Usability	Competence	Tolerance	Social Presence	Trustworthiness	Social Attractiveness
User 1	62.50	25	20	18	11	17
User 2	35	25	20	21	10	15
User 3	62.50	25	19	20	20	26
User 4	62.50	25	20	21	10	15

Participants also expressed that they found the conversational AI virtual trainer difficult to understand (2 users), better to understand (1 user) and easy to understand (1 user). In addition, users expressed that they found the conversational AI virtual trainer difficult to use (2 users), okay to use (1 user) and easy to use (1 user). When they were asked how likely they were to use the conversational AI virtual trainer again participants expressed that they were unlikely to use this again (1 user), neutral (1 user) and were likely to use this again (2 users).

4 DISCUSSION

The aim of this user study was to evaluate how users perceive the conversational AI virtual trainer developed by OI. This understanding will facilitate further development of the conversational AI virtual trainer to equip it to provide unsupervised teaching to underserved learners in refugee and internally displaced people in camps in Iraq. Overall, users reported positive feelings after being introduced to the conversational AI virtual trainer. The results showed that users found the conversational AI virtual trainer to be trustworthy, socially attractive, demonstrates a sense of social presence, tolerable and competent. However, they found the usability of the conversational AI virtual trainer to be the only limitation. However, there are some language barriers which could be what produced the low usability scores. The use of Deep Fake as a means of developing the conversational AI virtual trainer proved useful as participants reported the conversational AI virtual trainer as socially present, trustworthy and socially attractive. In addition to that, the conversation design using NLU and CORE also proved useful as the users reported the conversational AI virtual trainer as competent and tolerant.

There were some limitations of the current study. The first was that the language of the users was not English and required a lot of explaining by the researchers. Perhaps to ensure a better interaction with learners next time, an interpreter could be briefed on the study and included in the user evaluation process. The second was that this user evaluation study was conducted using an online survey only which means there is an absence of (non)verbal communication. An asynchronous online interaction lacks richness and spontaneity of face-face interaction [22].

Future work could conduct a focus group to gather information about the perceptions of the conversational AI virtual trainer for a deeper understanding. Focus groups have been found to be a powerful tool in system development. However, the power of a focus group should lie in exploring the usability of a system by discovering what the users want from the system [23]. This user study was the first of many in the development of the conversational AI virtual trainer that meets individual user's needs. The long-term goal of this project is to develop an intelligent embodied conversational AI virtual trainer.

5 CONCLUSIONS

OI has a vision to promote access to learning in underserved countries and communities. In the path to achieving that vision, a conversational AI virtual trainer has been developed to equip such populations with employability skills. The aim of this study was to evaluate this conversational AI virtual trainer from four users in a UN refugee camp in Iraq. The users expressed that the conversational AI virtual trainer was socially attractive, warm, competent, trustworthy, tolerant and socially present. However, the usability of the conversational AI virtual trainer will need to be improved. Future work could look at using focus groups or interviews to explore the usability of the conversational AI virtual trainer on a deeper level. The implications of this work are important to equip a global population with employability skills without potential barriers of access to human trainers.

ACKNOWLEDGEMENTS

We thank the management of Barzani Community Foundation (BCF) United Nations (UNHCR) camp for supporting us to conduct this research for the benefit of their camp members. We also thank Mr Hisham Sabir Khan who helped facilitate this relationship. We thank Dipesh Sharma, Jason Francis and Mark Adriaanse who advised, under agreement, on the project. We also thank Innovate UK (UKRI) for seeing credibility in our work and for funding us separately to build a similar educational conversational AI for South Africa.

REFERENCES

1. Devedzic V. Education and the Semantic Web [Internet]. Vol. 14, International Journal of Artificial Intelligence in Education. IOS Press; 2004 [cited 2021 May 9]. Available from: http://galeb.etf.bg.ac.yu/~devedzic/
2. Nwana HS. Intelligent tutoring systems: an overview. Artif Intell Rev [Internet]. 1990 Dec [cited 2021 May 9];4(4):251–77. Available from: https://link.springer.com/article/10.1007/BF00168958
3. Wenger E. Artificial Intelligence and Tutoring Systems: Computational and Cognitive ... - Etienne Wenger - Google Books [Internet]. 2014 [cited 2021 May 9].
4. Vanlehn K, Jordan PW, Rosé CP, Bhembe D, Böttner M, Gaydos A, et al. The architecture of why2-atlas: A coach for qualitative physics essay writing. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) [Internet]. 2002 [cited 2021 May 9];2363:158–67. Available from: https://link.springer.com/chapter/10.1007/3-540-47987-2_20
5. Sabo KE, Atkinson RK, Barrus AL, Joseph SS, Perez RS. Searching for the two sigma advantage: Evaluating algebra intelligent tutors. Comput Human Behav. 2013 Jul 1;29(4):1833–40.
6. Zhang B, Jia J. Evaluating an intelligent tutoring system for personalized math teaching. In: Proceedings - 2017 International Symposium on Educational Technology, ISET 2017. Institute of Electrical and Electronics Engineers Inc.; 2017. p. 126–30.
7. Mahmoud MH, Abo El-Hamayed SH. An intelligent tutoring system for teaching the grammar of the Arabic language. J Electr Syst Inf Technol. 2016 Sep 1;3(2):282–94.
8. Sung YT, Liao CN, Chang TH, Chen CL, Chang KE. The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. Comput Educ. 2016 Apr 1;95:1–18.
9. Wang D, Han H, Zhan Z, Xu J, Liu Q, Ren G. A problem solving oriented intelligent tutoring system to improve students' acquisition of basic computer skills. Comput Educ. 2015 Feb 1;81:102–12.
10. Koedinger KR, Brunskill E, Baker RSJD, McLaughlin EA, Stamper J. New potentials for data-driven intelligent tutoring system development and optimization. AI Mag [Internet]. 2013 Sep 15 [cited 2021 May 9];34(3):37–41. Available from: https://ojs.aaai.org/index.php/aimagazine/article/view/2484
11. Frize M, Frasson C. Decision-support and intelligent tutoring systems in medical education. In: Clinical and Investigative Medicine [Internet]. 2000 [cited 2021 May 9]. p. 266–9. Available from: https://www.researchgate.net/publication/12339815
12. Al Emran M, Shaalan K. A Survey of Intelligent Language Tutoring Systems. In: Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014. Institute of Electrical and Electronics Engineers Inc.; 2014. p. 393–9.
13. Karacı A, Üniversitesi K, Piri Z, İbrahim Akyüz H, Bilgici G. Student Perceptions of an Intelligent Tutoring System: A Technology Acceptance Model Perspective Siber güvenlik View project Social media use View project Student Perceptions of an Intelligent Tutoring System: A Technology Acceptance Model Perspective. Artic Int J Comput Appl [Internet]. 2018 [cited 2021 May 9];182(22):975–8887. Available from: http://79.123.169.199:8080/index.aspx
14. Li J, Kizilcec R, Bailenson J, Ju W. Social robots and virtual agents as lecturers for video instruction. Comput Human Behav [Internet]. 2015 [cited 2021 May 9]; Available from: http://dx.doi.org/10.1016/j.chb.2015.04.005
15. Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC. Deepfakes: Trick or treat? Vol. 63, Business

Horizons. Elsevier Ltd; 2020. p. 135–46.

16. McShane M. Natural language understanding (NLU, not NLP) in cognitive systems. AI Mag [Internet]. 2017 Dec 1 [cited 2021 May 11];38(4):43–56. Available from: https://ojs.aaai.org/index.php/aimagazine/article/view/2745

17. Brooke J. SUS - A quick and dirty usability scale. Usability Eval Ind [Internet]. 1996 [cited 2018 Jan 30];189(194):4–7. Available from: http://hell.meiert.org/core/pdf/sus.pdf

18. Kortum PT, Bangor A. Usability Ratings for Everyday Products Measured with the System Usability Scale. Int J Hum Comput Interact. 2013 Jan;29(2):67–76.

19. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. J Human–Computer Interact Intl J Human–Computer Interact [Internet]. 2008 [cited 2018 Mar 19];24(6):574–94. Available from: https://www-tandfonline-com.ezproxy.brunel.ac.uk/doi/pdf/10.1080/10447310802205776?needAccess=true

20. Fiske ST. Warmth and competence: Stereotype content issues for clinicians and researchers. Can Psychol [Internet]. 2012 [cited 2021 May 9];53(1):14–20. Available from: /pmc/articles/PMC3801417/

21. Nass C, Isbister K, Lee E-J. Truth is beauty: Researching embodied conversational agents. Embodied conversational agents [Internet]. 2000 [cited 2018 Aug 8];374–402. Available from: https://pdfs.semanticscholar.org/0517/22a89ea75583cfe58f6d218db373776b4c6e.pdf

22. Salmons J. Qualitative Online Interviews: Strategies, Design, and Skills [Internet]. 2015 [cited 2021 May 9].

23. McQuarrie EF, Krueger RA. Focus Groups: A Practical Guide for Applied Research. J Mark Res. 1989;26(3):371.