JOURNAL OF CME
2024, VOL. 13, 2434322
https://doi.org/10.1080/28338073.2024.2434322

EUROPEAN CME Forum
Taylor & Francis
Taylor & Francis Group

BRIEF REPORT

Revolutionising Faculty Development and Continuing Medical Education Through AI-Generated Videos

Irene Contreras, Samia Hossfeld, Katharine de Boer, Jane Thorley Wiedler and Monica Ghidinelli

AO Education Institute, AO Foundation, Davos, Switzerland

ABSTRACT

Producing high-quality and engaging educational videos for continuing medical education (CME) is traditionally time-consuming and costly. Generative AI tools have shown promise in creating synthetic videos that mimic traditional lecture videos. We conducted a comparative analysis of four AI video generation platforms HeyGen, Synthesia, Colossyan, and HourOne using the Kano model. Our analysis revealed that HeyGen met most of our requirements. We created two videos and collected feedback from 25 learners. The feedback indicated that the videos were of good quality, engaging, and well-paced for learning. Only 32% recognised the videos as AI-generated, citing limited facial expressions, hand gestures and monotone vocal expression. Importantly, only 24% considered disclosure of AI-generated content necessary. This research indicates that AI-generated videos can be a viable alternative to traditionally produced educational videos. It offers an efficient, cost-effective solution for producing educational content. Ethical considerations regarding AI content disclosure should be addressed to maintain transparency.

ARTICLE HISTORY
Received 9 August 2024
Accepted 19 November 2024

KEYWORDS
AI-generated videos; synthetic videos; AI-generated lectures; AI video platforms; CME; faculty development; generative AI

Introduction

Artificial intelligence (AI) has emerged as a disruptive technology emerging across various. Within medical education, AI technologies are increasingly recognised for their potential to revolutionise traditional teaching methods and enhance educational content quality and accessibility.

In the context of continuing medical education (CME) and continuous professional development (CPD) a prominent use of generative AI tools is the creation of synthetic videos including virtual instructors, which resemble traditional lecture videos. These AI-generated videos have proven to be viable substitutes for traditionally produced lecture videos since learners can achieve comparable content acquisition and learning experiences [1], making educational content more accessible and scalable. The quality of the avatars has improved tremendously with the concomitant rise of deepfake technology enabling face-swapping and facial expression control for authentic appearance that is almost indistinguishable from real individuals [2]. The presence and gestures of the virtual instructor can positively impact learner behaviours, attitudes, and motivation [3] and can improve learning in online settings [4].

AI-generated synthetic videos offered significant advantages in terms of quality and accessibility of educational materials and facilitated the development of more engaging and interactive learning experiences [5]. The production costs are lower compared to traditional videos that demand extensive human labour, filming equipment, and specific software. Moreover, creating synthetic videos is more time-efficient for content experts and editors, as correcting errors or making updates only requires script editing and video regeneration, rather than another round of traditional filming and editing.

There are several AI tools on the market for generating synthetic videos, each offering slightly different features and capabilities. For our global organisation, the AO Foundation, the primary goal was to find an AI video generation tool that could create engaging and scalable content tailored to diverse audiences across different socio-cultural backgrounds and clinical specialties. Efficiency was a critical objective for us. By utilising AI tools, we hoped to significantly lower production costs and limit production, review, and iteration processes. Additionally, we sought to avoid cognitive fatigue, ensuring that learners remain focused and can absorb information effectively over longer

CONTACT Monica Ghidinelli monica.ghidinelli@aofoundation.org AO Foundation, AO Education Institute, Clavadelerstrasse 8, Davos 7270, Switzerland

Supplemental data for this article can be accessed online at https://doi.org/10.1080/28338073.2024.2434322

periods. Finally, we prioritised flexibility, allowing for the modification and updating of educational content in response to emerging research evidence and/or feedback from learners.

In this paper we explored and evaluated four AI video generation tools from the perspective of the content creator (user). We then generated two videos with the tool that matched most of our requirements and collected feedback from 25 learners.

Method

Study Design

This study combined objective feature assessment and learner feedback to guide the adoption of effective AI video generation tools. The four generative AI platforms Synthesia (London, UK), Colossyan (Philadelphia, USA), HourOne (Tel Aviv, Israel), and HeyGen (Los Angeles, USA) were tested. The Kano model served as the basis for the objective feature assessment, which was carried out by one technical expert in learning management systems and three content developers for faculty development programmes. The learner survey was administered to 31 surgeons from Europe and South Africa that participated in the pilot of the *AO Coach-Refresher training* module, obtaining 25 responses (response rate of 81%).

Feature Survey Design

To conduct an objective analysis of the features provided by the AI tools, the Kano model [6] was selected as the primary evaluative framework. The Kano model is well-suited because it categorises features based on their impact on user satisfaction. The initial phase of applying the Kano model involved compiling an inventory of features relevant to Artificial Intelligence-Generated Content (AIGC) for video. The features were systematically classified into groups, including branding, export options, sharing options, templates, and avatars. For example, the avatars group of features included aspects such as avatar diversity, natural appearance, accurate lip synchronisation, and realistic emotions and movement. The full list is available in Appendix 1. The subsequent step in the Kano model process involved designing and administering a survey to capture user preferences regarding these features. Seventeen features were identified and evaluated from both functional and dysfunctional perspectives, requiring the users to respond to 34 questions. Both perspectives ask survey participants to consider the perspectives of presence and absence of each feature. The full questionnaire is available in Appendix 2.

To determine the essential features for comparison, the detailed list in Appendix 1 has been grouped into higher-level aspects leading to 20 items, including the 3 added for technical requirements which have been used for comparing the AI tools.

Videos

The two AI-generated videos with HeyGen were included in a 3-hour online training on coaching, combining them with different interactive activities and other non-AI-generated videos. One video of 70 seconds was used for the introduction to the training activities and one video of 1 minute and 51 seconds summarised the application of coaching principles in a feedback conversation, as well as an explanation of the Dunning-Kruger effect. The script was created by the content developer (script divided by scene) and then imported in to the HeyGen platform and optimised by using its in-system AI capabilities. The generated video content was edited using the available AI options and different avatar features and voices were tested to increase expressivity and fidelity. Punctuation was used to influence the diction, intonation, and expressivity of the avatars.

Data Analysis

Survey responses were analysed with Excel Version 16.0, Microsoft, 2024.

Ethical Considerations

All the participants were asked for their agreement to participate in the study. On the survey, we included the following statement of purpose, which disclosed our intended use of the data: “The information you provide is anonymised and will be used for research purposes”.

ChatGPT 4.0 (OpenAI, San Francisco, CA; accessed July 26, 2024) was utilized for manuscript editing.

Results

To find an optimal tool for generating AI videos based on our organisation’s needs we performed a comparative analysis of the four AI platforms Synthesia, Colossyan, HourOne and HeyGen.

The first step of the analysis involved classifying the features based on the survey results (Appendix 1) according to the Kano model process [7] leading to classifying each aspect into one of the following categories (Figure 1):

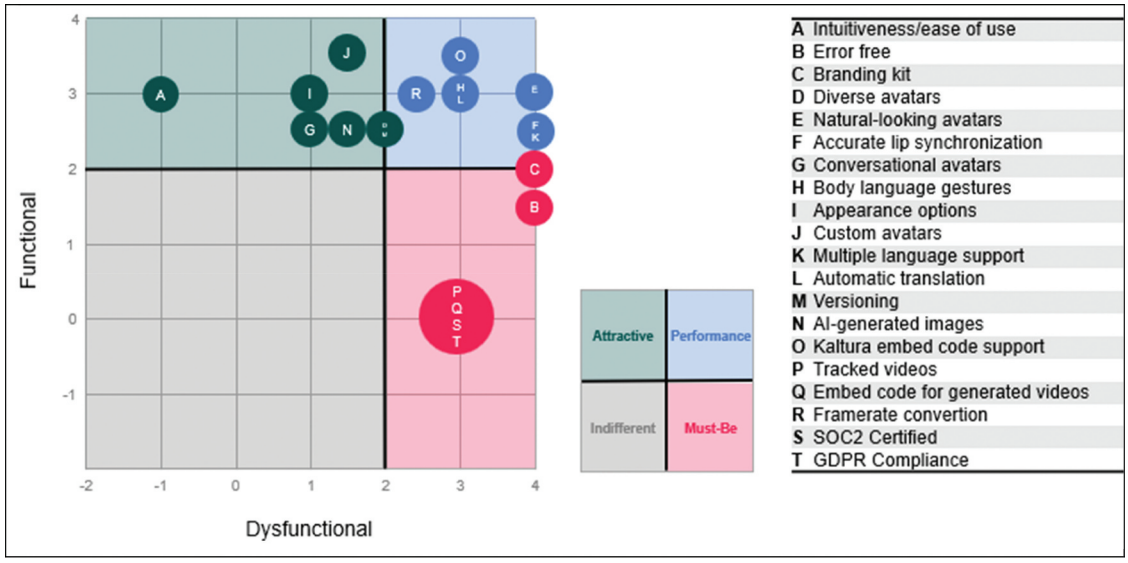

Figure 1. Set of AI-video platform key features determined using the Kano model.

Must-be: Features whose presence does not necessarily lead to satisfaction, but whose absence can cause dissatisfaction.

Performance: Features that have a direct correlation with satisfaction; the more of these features a product possesses, the higher the user satisfaction level.

Attractive: Features that increase user satisfaction when present, but whose absence has little impact on satisfaction.

Indifferent: Features that neither enhance nor diminish user satisfaction.

- Must-be: Features whose presence does not necessarily lead to satisfaction, but whose absence can cause dissatisfaction.
- Performance: Features that have a direct correlation with satisfaction; the more of these features a product possesses, the higher the user satisfaction level.
- Attractive: Features that increase user satisfaction when present, but whose absence has little impact on satisfaction.
- Indifferent: Features that neither enhance nor diminish user satisfaction.

Since the Kano model is designed for one tool, the second step of the analysis involved input from the content creators (users) who tested the four tools on how well each tool performed in each feature. Input was collected to rate the performance of each product for each feature on a scale from 1 to 4 reflected in the radar charts (Figure 2). Among all evaluated platforms, HeyGen achieved the highest overall rating, whereas HourOne received the lowest. Synthesia and Colossyan exhibited comparable ratings in the "must-be" category, except for tracked videos, which distinguished Colossyan, and SOC2 Certification which Synthesia offered but not Colossyan. In the "performance" category, Synthesia outperformed Colossyan in body language gestures. Conversely, Colossyan surpassed Synthesia in all "attractive" features, with the exception of versioning (Figure 2).

We then generated 2 videos for our *Coaching module* with HeyGen (Figure 3a) and collected feedback from 25 learners. Respondents rated the video quality between fair and excellent (Figure 3b). They found the content to be clear to extremely clear (Figure 3c), with 89% indicating that the videos were well-paced for learning (Figure 3d). Additionally, 76% of respondents considered the videos to be engaging to extremely engaging (Figure 3e), and 40% felt that the videos captured their attention very or extremely well (Figure 3f). Notably, only 32% of respondents clearly recognised that the videos were AI-generated, 36% realised that something was different and 32% did not recognise it (Figure 3g). We also asked what aspects of the video suggested to them that the video was generated using AI. Limited facial expression, monotony of voice tone, and limited body language were the ones most mentioned. Only 24% found disclosure that a video is generated using AI important (Figure 3h).

Discussion

In this work we evaluated four AI video generation platforms: Synthesia, Colossyan, HourOne, and HeyGen. HeyGen was the tool that met most of our requirements, followed closely by Colossyan. The feedback from participants indicates that AI-generated videos can be a viable alternative to traditionally produced educational videos.

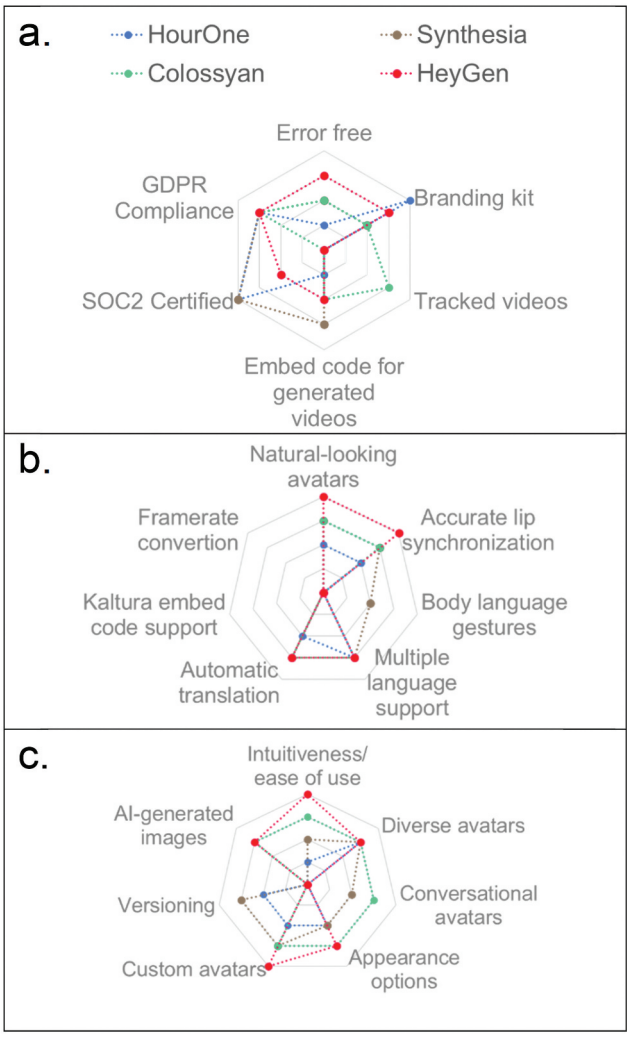

Figure 2. Radar chart of AI-tools features. a. Features whose presence does not necessarily lead to satisfaction, but whose absence can cause dissatisfaction; b. features that have a direct correlation with satisfaction; the more of these features a product possesses, the higher the user satisfaction level. c. Features that increase user satisfaction when present, but whose absence has little impact on satisfaction.

User experience

All the compared AI tools generally adhere to best practice to design effective instructional videos [8] making it easy to follow. We used it for non-clinical content however it might be suited for clinical content. The key benefits of AI-driven video production tools were the time and cost-effectiveness, ability for rapid editing, and the possibility to adapt the video, or parts of the video, to target a particular audience (e.g. regional-wise) without much trouble. The downside is that a full preview mode is not available in any of the tested tools so every time a video must be generated with associated waiting time.

Learner perception and effectiveness

Learners found the videos generated using HeyGen to be of good quality, engaging, and well-paced for learning. Notably, only 32% of respondents clearly recognised that the videos were AI-generated, while 36% sensed something different. Commonly cited indicators of AI generation included limited facial expression, monotony of voice tone, and limited body language. Despite technological advancements, avatars still lack a completely natural appearance, which may distract or interfere with the learner's experience. Studies have shown that voice tone and faculty's nonverbal expressiveness (for example gestures when they speak) can affect learners' motivation, as well as learning outcomes [2,9,10]. Educators should be aware of these aspects when utilising Artificial Intelligence-Generated Content (AIGC) tools.

Challenges and Considerations in Using AI for CME Video Production

The deployment of AI in this context comes with several challenges and considerations that must be carefully managed to ensure successful and ethical implementation.

Quality and accuracy of input content

The quality and accuracy of input content directly impacts the effectiveness and quality of AI-generated video content. AI algorithms rely on accurate, complete, and high-quality data to function correctly. Inaccurate, incomplete, or erroneous input content can undermine the educational goals, even if the video excels in presentation and visuals.

Data privacy, security and legal considerations

One primary concern when using AI in video production is data privacy and security, particularly when utilising AI-generative tools that allow for the customisation of avatars with real faces, creating deepfake videos. While this feature may seem attractive, enabling the showcasing of content experts without the need for travel or recording time, it raises significant concerns about data privacy and security. These platforms can become prime targets for cyberattacks, potentially leading to data breaches and unauthorised access to confidential information. Consequently, we refrained from testing this option as we currently lack a legal framework.

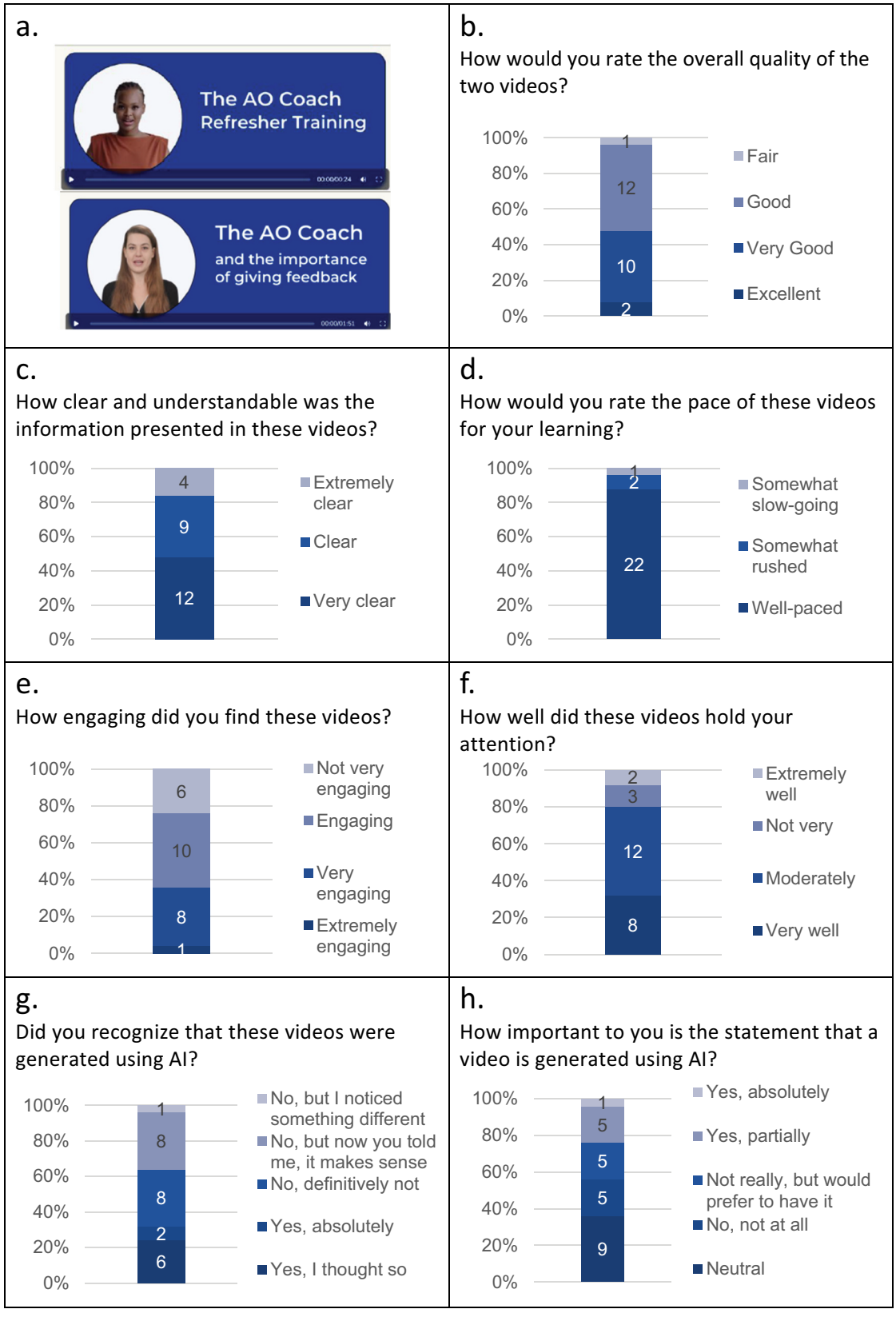

Figure 3. Learner feedback. a. AI-generated videos b,c. Analysis of survey answers $N = 25$.

Ethical considerations

Regarding the ethical considerations in the deployment of AI in CME/CPD video production, educators and content creators must proactively address potential biases in AI algorithms to ensure fairness, accuracy, and inclusivity. As our work in the AO Foundation is mainly in the sphere of orthopaedic surgery with a clearly identified homogeneity of the workforce [11]; [12], it was especially important for us to offer diverse avatars to our audience.

In our survey, only 24% of respondents deemed it necessary to disclose that a video was AI-generated, raising ethical considerations regarding transparency in educational content. We believe that maintaining high standards of authenticity and trust is crucial for educational content. Establishing ethical guidelines and standards for AI use in CME/CPD can further ensure responsible development and deployment.

Limitations

Considering the rapid advancements in the AIGC domain, it is important to note that the analysis of AI-platform features was based on information obtained up to mid-April 2024.

The selection of the AI generative platform is based on the needs of our organisation and might be different for other CME/CPD providers.

Conclusion

The integration of AI-generated videos represents a significant advancement in the delivery of CME and CPD. As technology continues to evolve, it is essential to leverage these innovations to enhance the quality and efficacy of medical education, ultimately improving patient care and healthcare outcomes. With strong security measures, clear ethical guidelines, and a commitment to accurate content, AI can revolutionise the way we deliver educational content.

Acknowledgments

We thank Hour One, Colossyan and Synthesia for providing the enterprise version of their platforms for testing purposes.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Authors Contribution

Irene Contreras, work conception and design, data acquisition, article drafting.

Samia Hossfeld work conception and design, data acquisition, data analysis, data interpretation, article drafting, manuscript revision.

Katharine de Boer, manuscript revision.

Jane Thorley Wiedler, manuscript revision.

Monica Ghidinelli, work conception and design, data acquisition, data analysis, data interpretation, article drafting.

ORCID

Monica Ghidinelli http://orcid.org/0000-0002-7378-6273

References

[1] Leiker D, Gyllen A, Eldesouky I, et al. Generative AI for learning: investigating the potential of synthetic learning videos. ArXiv, abs/2304.03784. 2023. doi: 10.48550/arXiv.2304.03784

[2] Renier LA, Shubham K, Vijay RS, et al. A deepfake-based study on facial expressiveness and social outcomes. Sci Rep. 2024;14(1):3642. doi: 10.1038/s41598-024-53475-5 PMID: 38351036; PMCID: PMC10864353.

[3] Hudson I, Hurter J. Avatar types matter: review of avatar literature for performance purposes. In: Lackey S Shumaker R, editors. Virtual, augmented and mixed reality. VAMR 2016. Lecture notes in computer science(). Vol. 9740. Cham: Springer; 2016. doi: 10.1007/978-3-319-39907-2_2

[4] Wang F, Li W, Mayer RE, et al. Animated pedagogical agents as aids in multimedia learning: effects on eye-fixations during learning and learning outcomes. J Educ Phychol. 2018;110(2):250–268. doi: 10.1037/edu0000221

[5] Kaplan A, Haenlein MS. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz. 2019;62(1):15–25, ISSN 0007-6813, doi: 10.1016/j.bushor.2018.08.004

[6] Kano N, Seraku N, Takahashi F, et al. "Attractive quality and must-be quality", Hinshitsu. The J Jpn Soc For Qual Control. 1984 Apr; 41, 39–48:3.

[7] Sauerwein E, Bailom F, Matzler K, et al. The Kano model: how to delight your customers. Int Work Seminar Production Econ. 1996;1(4):313–327.

[8] Mayer RE. Evidence-based principles for how to design effective instructional videos. J Appl Res Memory Cognition. 2021;10(2):229–240. doi: 10.1016/j.jarmac.2021.03.007

[9] Mayer RE, DaPra CS. An embodiment effect in computer-based learning with animated pedagogical agents. J Exp Phychol: Appl. 2012;18(3):239–252. doi: 10.1037/a0028616

[10] Imhof M, Välikoski T, Laukkanen A, et al. Cognition and interpersonal communication: the effect of voice quality on information processing and

person perception. Stud In Commun Sci. 2014;14 (1):37–44, ISSN 1424-4896, doi: 10.1016/j.scoms.2014.03.011

[11] Onuoha A, Meadows A, Faraj M, et al. Comparative analysis of racial and gender diversity in orthopedic surgery applicants and residents from 2007 and 2019. J Orthopaedic Exper Innov. 2022;3(1). doi: 10.60118/001c.31412

[12] Shah KN, Ruddell JH, Scott B, et al. Orthopaedic surgery faculty: an evaluation of gender and racial diversity compared with other specialties. JB JS Open Access. 2020;5 (3):e20.00009–e20.00009. doi: 10.2106/jbjs.Oa.20.00009