

## Article

# Evaluating an Artificial Intelligence (AI) Model Designed for Education to Identify Its Accuracy: Establishing the Need for Continuous AI Model Updates

Navdeep Verma \*, Seyum Getenet, Christopher Dann and Thanveer Shaik

School of Education, University of Southern Queensland, Queensland 4300, Australia; seyum.getenet@unisq.edu.au (S.G.); chris.dann@unisq.edu.au (C.D.); thanveer.shaik@unisq.edu.au (T.S.)

\* Correspondence: navdeepverma2032@gmail.com

**Abstract:** The growing popularity of online learning brings with it inherent challenges that must be addressed, particularly in enhancing teaching effectiveness. Artificial intelligence (AI) offers potential solutions by identifying learning gaps and providing targeted improvements. However, to ensure their reliability and effectiveness in educational contexts, AI models must be rigorously evaluated. This study aimed to evaluate the performance and reliability of an AI model designed to identify the characteristics and indicators of engaging teaching videos. The research employed a design-based approach, incorporating statistical analysis to evaluate the AI model's accuracy by comparing its assessments with expert evaluations of teaching videos. Multiple metrics were employed, including Cohen's Kappa, Bland–Altman analysis, the Intraclass Correlation Coefficient (ICC), and Pearson/Spearman correlation coefficients, to compare the AI model's results with those of the experts. The findings indicated low agreement between the AI model's assessments and those of the experts. Cohen's Kappa values were low, suggesting minimal categorical agreement. Bland–Altman analysis showed moderate variability with substantial differences in results, and both Pearson and Spearman correlations revealed weak relationships, with values close to zero. The ICC indicated moderate reliability in quantitative measurements. Overall, these results suggest that the AI model requires continuous updates to improve its accuracy and effectiveness. Future work should focus on expanding the dataset and utilise continual learning methods to enhance the model's ability to learn from new data and improve its performance over time.

Academic Editor: Will W. K. Ma

Received: 11 February 2025

Revised: 19 March 2025

Accepted: 20 March 2025

Published: 23 March 2025

**Citation:** Verma, N., Getenet, S., Dann, C., & Shaik, T. (2025). Evaluating an Artificial Intelligence (AI) Model Designed for Education to Identify Its Accuracy: Establishing the Need for Continuous AI Model Updates. *Education Sciences*, 15(4), 403. <https://doi.org/10.3390/educsci15040403>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** AI; video conferencing; online student engagement; teachers' behaviours; teachers' movements; design-based research

## 1. Introduction

Over the past decade, there has been substantial growth in online education within higher education institutions. This growth is due to its flexibility, accessibility, and cost efficiency (Castro & Tumibay, 2021; Dhawan, 2020). Further, COVID-19 has compelled higher education institutes worldwide to transition to online learning (Xie et al., 2021). Due to this sudden change, teachers encounter notable challenges in adapting to online learning, with student engagement emerging as the most prominent challenge (Alenezi et al., 2022). Studies have highlighted that fostering online student engagement is more complex than engaging students in traditional face-to-face learning (Gillett-Swan, 2017;

Hew, 2016). The potential of online learning and its trends brings forth new opportunities but also poses various challenges (Liang & Chen, 2012).

Incorporating AI can assist in addressing these challenges by identifying and evaluating discrepancies and offering suggestions for enhancing teaching effectiveness. AI opens up new avenues for learning and teaching (Limna et al., 2022). AI technologies' abilities to quickly analyse large datasets, recognise patterns, and make predictions support more personalised and effective learning experiences (Harry & Sayudin, 2023; Shaikh et al., 2022; Tahiru, 2021). For instance, AI-powered systems can recommend personalised learning paths, automate grading, and enhance educational resources (Nguyen, 2023). However, a critical challenge lies in evaluating the accuracy of AI models, especially when they are tasked with assessing complex human behaviours and movements, such as those of teachers, aimed at encouraging student engagement. Despite its potential, there is still much to learn about how accurately AI can interpret and predict the behaviours that enhance student engagement in online learning environments.

This study employed design-based research (DBR) to address these gaps by designing an AI model to identify engagement-enhancing teacher behaviours and movements during video conferences. During the initial phase of this DBR, the authors conducted a systematic literature review to determine the characteristics and indicators of engaging teaching videos Verma et al. (2023b). In the second phase, the authors, with the assistance of an AI expert, trained an AI model to replace the manual annotation of teaching videos based on teachers' behaviours and movements (Verma et al., 2023a), which expedites the process as manual annotation was identified as time-consuming (Beaver & Mueen, 2022). The identified characteristics and indicators were then applied to train the AI model using deep learning as an AI methodology. The current phase focuses on evaluating the AI model to ensure its accuracy and determine whether continuous AI model updates are necessary. Specifically, this study seeks to address the following research questions:

"How accurately can an AI model generate a report for characteristics and indicators of engaging teaching videos based on teachers' behaviours and movements?" (RQ1)

"Why is it important to continuously update the AI model designed to enhance online learning and teaching?" (RQ2)

By addressing these questions, this research aims to contribute to the ongoing effort to accurately and sustainably integrate AI into online learning.

## 2. Background

This section consists of three subsections. Section 2.1 presents the three distinct phases of the DBR, with a special focus on the current phase. Section 2.2 explores existing studies on evaluation methods in the field of education. Finally, Section 2.3 delves into studies that discuss evaluation methods within AI. Each section provides valuable insights and analysis into these important topics, highlighting their significance and implications in their respective domains.

### 2.1. Previous Phases

This study is the third phase of a DBR where the authors evaluate an AI model to ensure its accuracy and to determine whether continuous model updates are necessary. In the first phase, the authors conducted a systematic literature review to identify the characteristics and indicators of engaging teaching videos. The authors reviewed 34 studies and identified 11 characteristics crucial for enhancing student engagement in video conferencing based on teachers' behaviours and movements Verma et al. (2023b). Further, 47 indicators that can describe each characteristic were identified. The identification and

categorisation of these indicators into the 11 main characteristics are backed by the significant findings from the reviewed studies and research concerning online student engagement. These characteristics were organised into three overarching domains: Teachers' behaviours, movements, and use of technology Verma et al. (2023b). Appendix A.1 illustrates the main theme, characteristics, and indicators of engaging teaching videos.

Researchers have demonstrated significant interest in examining the influence of teachers' behaviours and movements on online student engagement (Cents-Boonstra et al., 2021; J. Ma et al., 2015). Verma et al. (2023b) strongly believe that the characteristics and indicators outlined in Appendix A.1 can be used as a benchmark for improving teachers' performance in online learning. Educational institutions can implement these indicators and characteristics of engaging teaching videos to enhance and regulate online teaching practices. Educational institutions worldwide can use this information to develop and offer training for teachers aimed at refining their skills in creating teaching videos that effectively boost online student engagement. However, identifying these engaging characteristics and indicators within recorded lecture videos requires human participation (Verma et al., 2023a). This manual identification and analysis process demands a significant amount of time and resources (Beaver & Mueen, 2022). Additionally, this approach may introduce human bias into the analysis. Therefore, in order to mitigate human bias and maintain efficiency in identifying engaging teaching videos, the authors collaborated with an AI expert to develop an AI model in phase 2. This tool generates a report on the characteristics and indicators of engaging teaching videos (Verma et al., 2023a).

In the second phase, the educational experts annotated 25 recorded lecture videos. The recorded lecture videos were presented to higher education students by lecturers from a university in Australia. The videos encompass a range of fields, including law, business, health, education, arts, and sciences, with an average length of 01:28:37 (Verma et al., 2023a). There were 13 female and 12 male speakers featured in the videos, and the authors secured ethical approval from the local university under the ethics approval number H20REA185. The manual annotation of these videos was performed individually using the Visual Geometry Group (VGG) Image Annotator (VIA) (Version 3) tool accessible from [https://www.robots.ox.ac.uk/~vgg/software/via/app/via\\_video\\_annotator.html](https://www.robots.ox.ac.uk/~vgg/software/via/app/via_video_annotator.html) (accessed on 11 January 2024). The manual annotation was carried out at the indicator level. Through the manual annotation of 25 recorded lecture videos, the authors identified 7 characteristics and 15 descriptive indicators, as detailed in Table 1. Based on the outcomes of this manual annotation, the AI expert assisted the authors during the development and training of an AI model designed to identify the characteristics and indicators of engaging teaching videos each time a video is processed.

**Table 1.** Characteristics and indicators identified in manual annotation (Verma et al., 2023a, p. 7).

Characteristics	Indicators
Encouraging Active Participation	<ul style="list-style-type: none"> <li>• Encouraging students' participation in discussion</li> <li>• Encouraging students to share their knowledge and ideas</li> <li>• Encouraging students to ask questions</li> <li>• Encouraging collaborative learning activities</li> <li>• Encouraging meaningful interaction</li> </ul>
Establishing Teacher Presence	<ul style="list-style-type: none"> <li>• Providing learning resources</li> <li>• Giving clear instructions</li> </ul>
Establishing Clear Expectations	<ul style="list-style-type: none"> <li>• Outlining the learning objectives</li> </ul>

Demonstrating Empathy	<ul style="list-style-type: none"> <li>• Using appropriate changes in tone of voice</li> </ul>
Using Nonverbal Cues	<ul style="list-style-type: none"> <li>• Facial expressions</li> <li>• Eye contact</li> <li>• Appropriate body language</li> </ul>
Using Technology Effectively	<ul style="list-style-type: none"> <li>• Enabling class recording for later review</li> <li>• Screen sharing and enabling chat, camera, and microphone</li> <li>• Varying the presentation media</li> </ul>

The engaging characteristics and indicators identified through manual video annotation were utilised to train prototype 1. Recognising challenges like misleading metrics and class imbalance, the model underwent refinement in prototype 2 by implementing the oversampling technique. By implementing the oversampling technique, the model was further improvised and demonstrated promising results, achieving an average precision, recall, F1-score, and balanced accuracy of 68%, 75%, 73%, and 79%, respectively, in categorising the annotated videos at the indicator level (Verma et al., 2023a).

The developed model has the potential to support higher education institutions in establishing moderation in lecture delivery. Moreover, it can significantly influence teaching and learning by providing teachers with reports on their technology utilisation effectiveness and identifying engagement-enhancing behaviours and movements present or lacking during their lecture delivery. To ensure the AI model's effectiveness and accuracy in generating reports, the current study evaluates its performance using a range of metrics.

## 2.2. Evaluation Methods in Education

Researchers have used various evaluation methods to evaluate the available instruments for measuring student engagement in education (Apicella et al., 2022; Giang et al., 2022; Shekhar et al., 2018).

Giang et al. (2022) conducted a validation of their proposed model to measure student engagement, which includes four sub-components, emotional engagement, cognitive engagement, participatory engagement, and agentic engagement, by employing a qualitative analysis approach, conducting interviews and focus group sessions as part of their data collection process. An interview in research is a data collection method where a researcher asks participants questions to gather information about their experiences, opinions, and perspectives (Kvale, 1996). Frequently, interviews are combined with other data collection methods to ensure a comprehensive and diverse range of information for analysis purposes (Turner, 2010).

In their recent study, Apicella et al. (2022) carried out an experimental case study to verify the effectiveness and validity of the tool they introduced to assess and monitor student engagement. A case study is commonly defined as a thorough and methodical examination of an individual, group, community, or another entity where the researcher carefully analyses detailed information about various factors or variables (Heale & Twycross, 2018).

Shekhar et al. (2018) employed a mixed-methods approach, combining quantitative and qualitative methods to assess the effectiveness and validity of the instruments they developed for observing active learning, instructor participation, student resistance, and student engagement. This combination of methods allowed for the validation of broader frameworks through qualitative analysis and the identification of specific elements to

incorporate into quantitative tools during the developmental stage, as Sandelowski (2000) suggested.

Chiu (2021) applied questionnaires in their study and adopted a quantitative analysis method to evaluate the model they provided, where they leveraged digital tools to fulfil the requirements of competence, relatedness, and autonomy, leading to active student engagement in online learning. A questionnaire serves as a methodical approach for gathering primary quantitative data in the literature. It typically consists of a sequence of written inquiries to which respondents are required to provide responses (Bell, 1999).

Lee et al. (2019) incorporated expert opinions and conducted reliability and validity analyses to ensure the accuracy and consistency of the model they proposed to enhance student engagement in e-learning environments. Expert opinion refers to a judgment by an individual with superior knowledge in a specific domain. It encompasses two key components: expertise and domain specificity (Pingenot & Shanteau, 2009).

### *2.3. Evaluation Methods in AI*

Several studies have explored using deep learning and computer vision techniques to evaluate AI-enabled tools that identify engagement-enhancing teacher behaviours and movements in video conferencing.

X. Ma et al. (2021) presented a deep learning-based approach to recognise online student engagement, employing both convolutional and recurrent neural networks. They analysed facial expressions, body movements, and gaze patterns to predict engagement levels.

Behera et al. (2020) focused on automatically analysing teachers' nonverbal behaviours in online learning settings. They employed computer vision techniques such as face detection, tracking, gesture recognition, and body pose estimation to extract meaningful features from video data. AI algorithms were applied to classify nonverbal behaviours and assess their impact on student engagement. In their research, Weng et al. (2023) conducted a systematic literature review on video-based learning analytics in online education. The review highlighted the importance of utilising computer vision techniques to analyse teachers' behaviours and their influence on online student engagement and learning outcomes. Ashwin and Guddeti (2019) explored the utilisation of deep learning techniques for automatic emotion recognition in educational videos. They used convolutional neural networks and recurrent neural networks to analyse teachers' and students' facial expressions and body movements, demonstrating the potential of deep learning models in capturing emotional cues and evaluating their impact on student engagement.

A handful of studies (Ashwin & Guddeti, 2019; Behera et al., 2020; X. Ma et al., 2021; Weng et al., 2023) highlight the use of deep learning and computer vision techniques in evaluating AI-enabled tools for identifying engagement-enhancing teacher behaviours and movements in video conferencing. They offer valuable perspectives on the capacity of these techniques to enhance student engagement and improve the quality of online learning experiences.

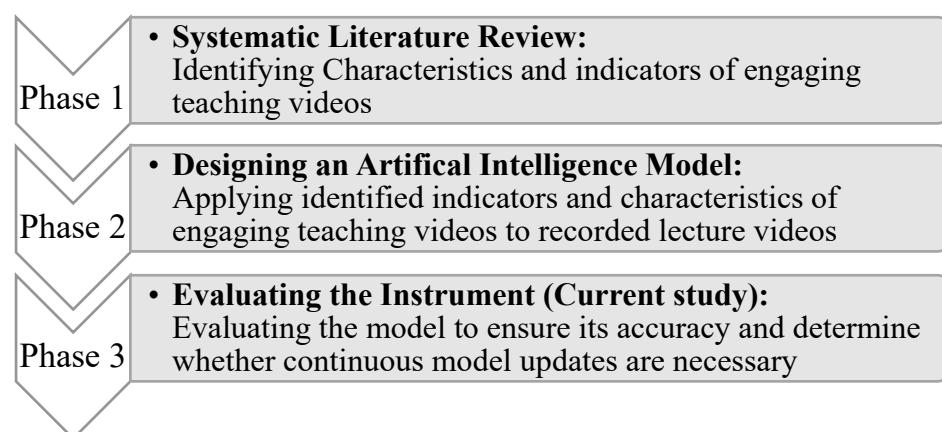
Existing research in education lacks evaluation methods specifically designed for measuring online student engagement using AI-enabled tools (Huang et al., 2023). Previous studies have focused on developing instruments and models for traditional face-to-face settings, utilising methods such as interviews, case studies, mixed-methods approaches, and questionnaires. The evaluation methods used to validate the instruments in education might not be suitable for the AI model created by the authors as these methods require human analysis, which can lead to bias (Heeg & Avraamidou, 2023).

This paper seeks to evaluate the AI model developed in the preceding phase through the use of various metrics such as Cohen's Kappa, Bland–Altman analysis, the Intraclass

Correlation Coefficient (ICC), and Pearson/Spearman correlation coefficients to assess its accuracy and identify whether it is necessary to perform continuous AI model updates.

### 3. Methods

The authors utilised a DBR approach to develop an AI model that generates reports on teachers' behaviours and movements whenever it processes a recorded lecture video. The DBR methodology has gained recognition in educational research, with many researchers highlighting its ability to support the development of practical research processes (Tinoca et al., 2022). Following the principles of the DBR methodology, this study has unfolded in three distinct phases. The phases of the DBR process are summarised in Figure 1.



**Figure 1.** Research phases.

**Phase 1, systematic literature review:** This phase involves a systematic review of the existing literature to identify the characteristics and indicators of engaging teaching videos. By analysing previous research, a foundational understanding of what constitutes effective teacher behaviours and movements in online teaching environments is established. In this study, the authors identified 47 indicators and 11 characteristics categorised into three main themes (see Appendix A). These identified indicators then guided the development of the AI model in subsequent phases.

**Phase 2, designing an AI model,** involves video annotation to create an AI model capable of analysing the characteristics and indicators identified in Phase I, to recognise and evaluate teachers' engagement-enhancing behaviours and movements in recorded lecture videos using Zoom. The model was designed through two prototypes.

#### *AI process*

The authors developed a deep learning model to learn a teacher's movements in a recording with the support of an AI expert. This is achieved by recording the temporal coordinates extracted from the tool's manual video annotation. Temporal coordinates are markers in the video timeline that help identify specific points in time. Selected lecture videos were split based on these coordinates, and we transformed them into a stack of image frames. The pre-processed frames were then labelled with corresponding teaching indicators, and we prepared the data model for training. Next, the data were split into two sets—training and testing—for model training and evaluation. An AI expert fed the training set to the convolutional neural network (CNN) model to learn the actions in the image frames and their corresponding labels. Finally, the test set was used to evaluate the performance of the CNN model.

#### *Data pre-processing*

During the data pre-processing step, the AI expert captured the temporal coordinates provided by the video annotation tool. For example, suppose a lecture recording displayed the teaching indicator “Clear and concise explanation of information” at the temporal coordinates (3051.315, 3053.256). In that case, the recorded lecture was divided into video segments highlighting and extracting the teaching indicator. Then, each video was split into segments of image frames and annotated each frame with the “Clear and concise explanation of information” teaching indicator. These annotated image frames are represented as 2D matrices and serve as inputs for the convolution layer of the deep learning model, as described in the subsequent subsection.

#### *Deep learning model*

The AI expert developed the CNN model as a deep learning approach for classifying two-dimensional (2D) data images. The CNN model offers the advantage of reducing the high dimensionality of images while preserving their information. Figure A2 illustrates the learning process of the CNN model. First, the input image frames, pre-processed in the previous step, are passed to a two-dimensional (2D) convolution layer, which uses a set of filters to divide the image frame into smaller sub-images and analyse them individually. The convolution layer’s output is then passed to the pooling layer, which estimates the maximum value for a feature set and creates a down-sampled group feature. The pooled features can be flattened into a 2D array and then processed in the output layer of the CNN model. The output layer provides a probability for each label classification, which can be optimised using a threshold value to classify the features into a label.

As shown in Figure 1, the present study, Phase 3, focuses on the third phase of this DBR, where authors have evaluated the AI model to ensure its accuracy and determine whether continuous updates are required. The authors have used multiple statistical methods to ensure the model’s accuracy. As part of the evaluation process, the model processed two recorded lecture videos and then generated results, identifying indicators of engaging teaching videos. Meanwhile, human experts who are well-versed in the domain independently analysed the same set of videos and provided their findings. The AI model was evaluated using multiple statistical methods to identify the statistical agreement and consistency between the findings of an AI model and two human experts in evaluating specific segments of video data.

#### *3.1. Data Collection*

The evaluation of the AI model’s ability to identify engagement-enhancing teacher behaviours and movements in video conferencing involved the participation of two human experts who manually annotated two videos and the AI-generated reports. The results obtained from the AI model and the two human experts were carefully analysed using various metrics.

Two videos of varying durations were utilised, one lasting 49 min and 3 s with 11 segments and the other lasting 58 min and 40 s with 23 segments, featuring presenters with different camera settings. The research was carried out with ethical clearance obtained from a regional university in Australia (ethics approval number H20REA185). However, demographic information about the lecturers, such as age, location, and academic background, was not collected.

#### *3.2. Video Analysis*

This section explores two distinct approaches for processing and analysing a set of videos to identify teachers’ engagement-enhancing behaviours and movements. It highlights the annotation process carried out by human experts and the use of an AI model designed by the authors in the previous phase to achieve a similar objective.

### 3.2.1. Expert Involvement

The two human experts conducted an annotation process guided by the 7 characteristics and 15 descriptive indicators of engaging teaching videos identified in the previous phase (refer to Table 1). Having two experts for comparison brings in diverse perspectives and broader insights and potentially leads to more comprehensive solutions or decisions. Additionally, it reduces the chances of individual bias influencing the outcomes, leading to a more balanced and reliable evaluation. To complete the manual annotation process, the Visual Geometry Group Image Annotator (VIA) tool was used (refer to Appendix A.2).

### 3.2.2. AI Reports

The AI model employed a deep learning model known as a convolutional neural network (CNN) to process the same set of recorded lecture videos. Its main goal was to identify the teachers' engagement-enhancing behaviours and movements based on the characteristics and indicators it had been trained with, similar to what the human experts utilised for manual annotation. By examining visual cues and patterns, the model generated detailed reports highlighting the teachers' behaviours and movements that enhance student engagement.

### 3.3. Data Analysis

The analysis involved multiple statistical methods to assess the agreement and consistency between the findings of an AI tool and two human experts in evaluating specific segments of video data. Cohen's Kappa was used to measure the inter-rater agreement for categorical items, considering the possibility of agreement occurring by chance. To analyse the differences between their assessments, Bland–Altman analysis was employed to explore the agreement between the AI tool and the experts. The Intraclass Correlation Coefficient (ICC) was calculated to assess the reliability and agreement of the quantitative measurements between the AI tool and both experts. Lastly, the Pearson and Spearman correlation coefficients were computed to measure the linear and rank-order relationships between the AI tool's assessments and those of the experts.

## 4. Results

Tables 2 and 3 serve as invaluable resources, offering a clear outline of the analyses conducted on each video and facilitating a deeper understanding of the comparative evaluations undertaken by both human experts and the AI model. Table 4 presents the statistical agreement and consistency analysis between the AI model and experts evaluating video 1 and video 2 data. The combined analysis results are discussed in detail, pointing out the findings for each statistical method used.

**Table 2.** AI and experts' findings from video 1.

Video 1	AI Model	Expert 1	Expert 2
Segment 0	1	1	14
Segment 1	6	8	6
Segment 2	6	8	6
Segment 3	14	8	14
Segment 4	1	14	8
Segment 5	15	7	15
Segment 6	7	7	No identified indicator
Segment 7	5	9	No identified indicator
Segment 8	2	8	No identified indicator
Segment 9	5	9	No identified indicator



Segment 10	9	9	No identified indicator
Segment 11	5	9	No identified indicator

**Table 3.** AI and experts' findings from video 2.

Video 2	AI Model	Expert 1	Expert 2
Segment 0	1	14	15
Segment 1	10	8	15
Segment 2	5	7	5
Segment 3	5	4	5
Segment 4	1	7	2
Segment 5	12	12	4
Segment 6	5	7	2
Segment 7	10	12	10
Segment 8	5	7	7
Segment 9	7	12	7
Segment 10	1	12	1
Segment 11	1	12	No identified indicator
Segment 12	5	9	No identified indicator
Segment 13	1	12	No identified indicator
Segment 14	1	12	No identified indicator
Segment 15	9	7	No identified indicator
Segment 16	5	7	No identified indicator
Segment 17	14	15	No identified indicator
Segment 18	5	12	No identified indicator
Segment 19	14	12	No identified indicator
Segment 20	1	9	No identified indicator
Segment 21	14	15	No identified indicator
Segment 22	1	1	No identified indicator
Segment 23	5	12	No identified indicator

**Table 4.** Statistical agreement and consistency analysis between the AI tool and experts.

Statistical Measure	AI Tool vs. Expert 1	AI Tool vs. Expert 2	Interpretation
Cohen's Kappa	0.09	0.07	Slight agreement
Bland–Altman Analysis			
-Mean Difference	4.92	2.24	Moderate variability in differences
-Standard Deviation of Differences	4.55	6.18	
-95% Limits of Agreement	(−4.00, 13.84)	(−9.87, 14.35)	
Intraclass Correlation Coefficient (ICC2k)	0.45	0.45	Moderate reliability
Pearson Correlation Coefficient	0.09	−0.02	Weak linear relationship
Spearman Correlation Coefficient	0.09	−0.10	Weak rank-order relationship

#### 4.1. Explanation of Findings

This section analyses the findings for the two distinct videos at each level. Tables (List the tables) showcase the outcomes of AI processing and expert analysis, forming the foundation for further exploration and discussion.

#### 4.1.1. Video 1 Results

Table 2 highlights video 1 segments (0 to 11) and the results obtained from the AI model and Experts 1 and 2.

The findings from video 1, as analysed by both the AI model and experts, are organised into four columns. The first column displays the video segments. The second column lists the indicators identified by the AI model. The third column presents the indicators identified by Expert 1, while the fourth column outlines the indicators identified by Expert 2. (Refer to Figure A4 in Appendix A.2 for the complete list of indicators.)

#### 4.1.2. Video 2 Results

Further, Table 3 presents video 2 segments (0 to 23) and the results from the AI model and Experts 1 and 2.

The findings from video 2 follow the same format, with four columns. The first column displays the video segments, the second contains the indicators identified by the AI model, the third presents the indicators identified by Expert 1, and the fourth outlines those identified by Expert 2. (Refer to Figure A4 in Appendix A.2 for the complete list of indicators.)

Table 4 summarises the result of the statistical agreement and consistency analysis between the AI model and expert findings, followed by a detailed explanation of the results.

In this study, multiple statistical methods were employed to assess the agreement and consistency between the findings of an AI model and two human experts in evaluating specific segments of video data. The analysis involved the calculation of Cohen's Kappa, Bland–Altman analysis, the Intraclass Correlation Coefficient (ICC), and Pearson/Spearman correlation coefficients to comprehensively explore the degree of similarity between the AI-generated results and the expert assessments.

Cohen's Kappa was used to measure the inter-rater agreement for categorical items, taking into account the possibility of agreement occurring by chance. The results indicated slight agreement between the AI model and the experts, with Cohen's Kappa values of 0.09 for Expert 1 and 0.07 for Expert 2. These low Kappa values suggest that the AI model's categorical assessments are only marginally aligned with those of the human experts, with a considerable amount of disagreement present.

When analysing the differences between their assessments, Bland–Altman analysis was employed to explore the agreement between the AI model and the experts. For the comparison between the AI model and Expert 1, the mean difference was 4.92, with a standard deviation of 4.55. The 95% limits of agreement ranged from −4.00 to 13.84. Similarly, the comparison with Expert 2 yielded a mean difference of 2.24, with a standard deviation of 6.18 and 95% limits of agreement from −9.87 to 14.35. These results reveal a moderate degree of variability in the differences between the AI model and the experts, indicating that while there is some level of agreement, the variability is substantial enough to warrant further refinement of the AI model.

The Intraclass Correlation Coefficient (ICC) was calculated to assess the reliability and agreement of the quantitative measurements between the AI model and both experts. The ICC value (ICC2k) for the comparison was 0.45, indicating moderate reliability. This suggests that while there is some consistency in the measurements between the AI model and the experts, the level of agreement is not strong enough to be considered highly reliable.

Finally, the Pearson and Spearman correlation coefficients were computed to measure the linear and rank-order relationships between the AI model's assessments and those of the experts. The Pearson correlation coefficient for the AI model and Expert 1 was 0.09, indicating a weak positive linear relationship, while the correlation with Expert 2 was

−0.02, reflecting a weak negative linear relationship. Similarly, the Spearman correlation coefficients showed a weak positive rank-order correlation of 0.09 with Expert 1 and a weak negative rank-order correlation of −0.10 with Expert 2. These results suggest that the AI model's findings have a minimal linear or monotonic relationship with the expert assessments.

The statistical analyses reveal that the AI model's assessments exhibit slight to moderate agreement and consistency with those of the human experts. While there is some level of alignment, the relatively low agreement metrics indicate that there is significant room for improvement in the AI model's performance. Enhancing the AI model, perhaps through additional training with a more diverse dataset or by refining its algorithms, could potentially increase its reliability and consistency with expert evaluations. This would be crucial for ensuring the AI tool's effectiveness and accuracy in real-world applications.

## 5. Discussion

Researchers (e.g., Apicella et al., 2022; Giang et al., 2022; Shekhar et al., 2018) have developed various evaluation methods such as interviews, case studies, mixed-methods approaches, and questionnaires to validate instruments and ensure their effectiveness in education. However, existing research in education lacks evaluation methods specifically designed for measuring online student engagement using AI models (Heeg & Avraamidou, 2023; Huang et al., 2023). Therefore, the authors employed multiple statistical methods to measure the developed AI model's accuracy and identify whether it requires continuous model updates.

### 5.1. Exploration of Research Findings

Upon evaluating the model trained in 2022 by annotating 25 recorded lecture videos by education experts, the results revealed that the model requires updating. This is mainly due to the significant increase in expert knowledge concerning human characteristics over the past two years, while the model's knowledge has not changed. Further, research in this field indicates that AI models require regular updates to maintain their effectiveness (Li et al., 2023; Murtaza et al., 2022; Ocaña & Opdahl, 2023; Roshanaei et al., 2024).

In relation to the RQ1: How accurately can an AI model generate a report on the characteristics and indicators of engaging teaching videos based on teachers' behaviours and movements? The findings revealed that the AI model's ability to identify the characteristics and indicators of engaging teaching videos was only marginally aligned with expert analyses. The main reason for these results is the evolving nature of the human mind. From the development of the model to its evaluation, the experts' understanding has evolved significantly, enabling them to recognise more characteristics and indicators from the recorded lecture sessions, while the knowledge embedded in the AI model remains static. If the AI model was trained on more data, such as more videos manually annotated by experts, these results would likely reflect a stronger alignment between the experts' assessments and the AI model, indicating a significant improvement in the model's performance and accuracy. This overall result was drawn from multiple statistical methods, including Cohen's Kappa, Bland–Altman analysis, the ICC, and Pearson and Spearman correlation coefficients, which indicated limited agreement between the AI model and the human experts. Specifically, Cohen's Kappa values were low at 0.09 for Expert 1 and 0.07 for Expert 2, suggesting minimal alignment with expert findings. Bland–Altman analysis showed a mean difference of 4.92 (SD = 4.55) for Expert 1 and 2.24 (SD = 6.18) for Expert 2, with 95% limits of agreement ranging from −4.00 to 13.84 and −9.87 to 14.35, respectively, demonstrating moderate variability in differences. The ICC value (ICC2k) of 0.45 indicated moderate reliability, while Pearson and Spearman correlation coefficients

revealed weak relationships: 0.09 with Expert 1 and  $-0.02$  with Expert 2 for Pearson, and 0.09 and  $-0.10$  for Spearman, respectively. These findings highlight significant room for improvement in the AI model's performance, suggesting that a further update is needed to enhance its accuracy and consistency with expert evaluations.

In relation to the RQ2: Why is it important to continuously update the AI model designed to enhance online learning and teaching? The evaluation findings indicate only a slight to moderate alignment of the AI model's performance outcome with the experts' analysis results, emphasising the need for further improvement through continuous model updates. Apart from the findings of this study, various factors support the importance of continuously updating AI models. AI models are trained and rely on historical data, which may become outdated as the data environment evolves. Such changes can significantly impact the AI model's performance, making regular updates necessary to keep the model's performance from declining (Li et al., 2023). Roshanaei et al. (2024) describe regular updates and patches for AI models as the process of refreshing them to address any weaknesses in their design or data handling processes. AI models need to be regularly updated to keep up with new information (Ocaña & Opdahl, 2023). Pinykh et al. (2020) recommend the incorporation of feedback from match results and adjusting algorithms as part of the continuous training and updating of AI models to improve their predictive accuracy over time. Further, model updates can be influenced by other factors such as the availability of new or higher-quality training data, user feedback, learning algorithm advancements, and the need to ensure fairness in the model (X. Wang & Yin, 2023). Murtaza et al. (2022) highlight that continuously updating AI learning models with new training data can enhance the learning experience. Therefore, keeping models up to date ensures that AI models can continuously offer relevant, effective, and fair support in online learning environments.

### 5.2. Implications

This study holds significant implications for the use of AI models in education. Firstly, this three-phase research project provides the characteristics and indicators of engaging teaching videos that can improve online student engagement. These characteristics and indicators can help teachers and educational institutions enhance their pedagogical approaches.

Secondly, this study provides a procedure to train AI models for education. Further, by creating an AI model in phase 2, this research proves that AI can be used to create models and tools to replace the manual identification process. This can avoid challenges such as time consumption, cost, and potential human bias. According to De Silva et al. (2024), one of the multifaceted benefits of AI is its ability to automate processes, leading to increased efficiency in terms of both time and cost.

Thirdly, this study highlights the importance of model monitoring and validation. Monitoring and validating AI systems to ensure accuracy and fairness are crucial. Aldoseri et al. (2023) highlighted that inaccurate, biased, or irrelevant outcomes derived from low-quality data can have adverse effects on decision-making processes grounded in AI outputs, emphasising the importance of validation to enable AI systems to generate dependable and valuable outcomes. Thus, this study employed various metrics to guarantee the reliability of the evaluation results for the developed AI model, assessing its accuracy and identifying the importance of continuous AI model updates. This establishes the need for a policy that requires educational institutions to regularly enhance and update AI models to maintain accuracy and reliability and ensure the models remain relevant.

Moreover, if the AI model accurately identifies these characteristics and indicators of engaging teaching videos effectively, it can provide teachers with significant support in

various aspects, such as saving time, enhancing learning, and reinforcing professional development. Regarding professional growth and continual improvement, AI-generated reports are instrumental in aiding teachers in recognising both the strong points and areas needing improvement in their lecture delivery concerning engagement. Similarly, processing engaging recorded lecture videos using the AI model provides teachers with valuable insights into what resonates most effectively with their students. This empowers them to make well-informed decisions for future learning experiences, ultimately resulting in improved teaching and learning outcomes. Further, this research also provides a manual annotation procedure that can assist AI engineers in developing similar AI models.

## 6. Limitations and Future Directions

While the authors have developed an AI model to understand student engagement based on teachers' behaviours and movements in video conferencing, certain limitations must be recognised. Firstly, significant differences in outcomes have been identified, attributed to factors such as human bias, evolution, and the limited training of the AI model due to a small dataset containing few indicators and variations. These factors underscore the need to enhance the AI model's performance to better align with the analyses conducted by human experts. Additionally, the reliance on a small dataset for evaluation emphasises the need for assessments on larger datasets by processing and analysing more lecture videos to comprehensively evaluate the model's performance.

In future research, the findings from this final phase may be incorporated for improvement. The results reveal that the AI model developed in this study to identify engagement-enhancing behaviours and movements needs continuous updates to address the challenges posed by evolving data. This study also establishes the importance of continuous model updates. As noted by Žliobaite et al. (2015) and Roshanaei et al. (2024), the performance of predictive models can degrade if they lack mechanisms for regular updates and adaptation to new data, highlighting the importance of continuous updates in preventing such vulnerability in AI models. In their study, C. Wang et al. (2024) suggested various triggers to perform model updates. Firstly, they introduced periodic updates, in which model updates are performed at intervals such as quarterly, monthly, or weekly. Secondly, they suggested performance-driven updates, where models are refreshed when their accuracy metrics fall below a predefined threshold. Lastly, they suggested a data-driven approach, where models are updated upon accumulating significant data. Another recommended approach is continual learning (CL), which enables AI models to be updated with new data without the need to retrain them from the beginning (Nikoloutsopoulos et al., 2024). Continual learning refers to an AI model's ability to continuously learn from new data streams while retaining its previous knowledge. In this process, the model improves its performance by adapting to new data and updating its knowledge base as new information becomes available.

## 7. Conclusions

As detailed in the explanation of findings, the AI model evaluation involved various statistical methods used to perform a statistical agreement and consistency analysis, comparing the AI model's findings with those of human experts. The results showed relatively low agreement between the AI model's ability to identify the characteristics and indicators of engaging teaching videos and the experts' analysis. While the AI model shows potential, the results highlight significant room for improvement, suggesting further updates are needed to improve the model's accuracy and achieve strong to excellent alignment with expert evaluations.

**Author Contributions:** N.V.: Conceptualization, Methodology, Formal Analysis, Writing—Original Draft and Review and Editing. S.G.: Conceptualization, Writing—Original Draft and Review and Editing. C.D.: Conceptualization, Writing—Original Draft and Review and Editing. T.S.: AI Methodology, Formal Analysis, Review and Editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

**Institutional Review Board Statement:** This research obtained ethics approval from the local university under the ethics approval number H20REA185, approval date 19 February 2021.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Please contact the authors for a data request.

**Conflicts of Interest:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
CNN	Convolutud Neural Network
COVID-19	Coronavirus Disease 2019
DBR	Design-based Research
VIA	VGG Image Annotator

## Appendix A

### Appendix A.1

Main theme, characteristics, and indicators of engaging teaching videos (Verma et al., 2023a, p. 11).

Main Theme	Characteristics	Indicators
Teachers' Behaviours	Encouraging Active Participation	<ul style="list-style-type: none"> <li>• Encouraging students' participation in discussion</li> <li>• Encouraging students to share their knowledge and ideas</li> <li>• Encouraging students to ask questions</li> <li>• Encouraging collaborative learning activities</li> <li>• Encouraging meaningful interaction</li> <li>• Encouraging students to turn on their webcams</li> </ul>
	Establishing Teacher Presence	<ul style="list-style-type: none"> <li>• Clear and concise explanations of information</li> <li>• Recognising and considering learners' individual differences</li> <li>• Using an appropriate style of presentation</li> <li>• Allowing sufficient time for students' information processing</li> <li>• Providing learning resources</li> <li>• Giving clear instructions</li> <li>• Using a range of teaching strategies</li> <li>• Appropriate speed of lecture delivery</li> </ul>
	Establishing Social Presence	<ul style="list-style-type: none"> <li>• Maintaining constant teacher–student interaction</li> </ul>

		<ul style="list-style-type: none"> <li>• Encouraging student–student interaction (peer collaboration)</li> <li>• Active and constructive communication</li> <li>• Taking on multiple roles</li> </ul>
	Establishing Cognitive Presence	<ul style="list-style-type: none"> <li>• Giving students a sense of puzzlement (trigger)</li> <li>• Providing opportunities for students to reflect (exploration)</li> <li>• Leading students to think and learn through discussion with others (integration)</li> <li>• Helping students apply knowledge to solve issues (resolution)</li> </ul>
	Questions and Feedback	<ul style="list-style-type: none"> <li>• Addressing students’ questions and providing prompt feedback</li> <li>• Asking for questions and feedback</li> <li>• Clarifying misunderstanding</li> </ul>
	Displaying Enthusiasm	<ul style="list-style-type: none"> <li>• Motivating students</li> <li>• Displaying positive emotion</li> </ul>
	Establishing Clear Expectations	<ul style="list-style-type: none"> <li>• Outlining the learning objectives</li> <li>• Outlining teachers’ expectations of students’ behaviours and responsibilities</li> </ul>
	Demonstrating Empathy	<ul style="list-style-type: none"> <li>• Using appropriate changes in tone of voice</li> <li>• Ensuring the learning environment is a respectful, safe, and supportive one</li> <li>• Showing concern</li> </ul>
	Demonstrating Professionalism	<ul style="list-style-type: none"> <li>• Demonstrating in-depth and up-to-date knowledge</li> <li>• Displaying appropriate behaviours</li> </ul>
Teachers’ Movements	Using Nonverbal Cues	<ul style="list-style-type: none"> <li>• Facial expressions</li> <li>• Gestures</li> <li>• Eye gazes</li> <li>• Silence</li> <li>• Eye contact</li> <li>• Physical proximity</li> <li>• Appropriate body language</li> </ul>
Use of Technology	Using Technology Effectively	<ul style="list-style-type: none"> <li>• Screen sharing and enabling chat, camera, and microphone</li> <li>• Varying the presentation media</li> <li>• Providing technical support to students</li> <li>• Providing multiple communication channels</li> <li>• Providing interactive software tools</li> <li>• Enabling class recording for later review</li> </ul>

#### Appendix A.2. Manual Video Annotation Procedure

VGG Image Annotator (VIA) software (Version 3) was used in this manual video annotation process to annotate Zoom-based lecture recordings. VIA is an open-source project-based annotation software for annotating images, audio, and videos, available at [https://www.robots.ox.ac.uk/~vgg/software/via/app/via\\_video\\_annotator.html](https://www.robots.ox.ac.uk/~vgg/software/via/app/via_video_annotator.html) (accessed on 11 January 2024).

In this project, the researchers used the following steps to annotate the videos:

**Step 1: Creating a new project:** Open the VIA annotation tool by clicking the link above. Add the project name on the top left-hand side (refer to Figure 1). The project name should be the same as the recorded lecture name.

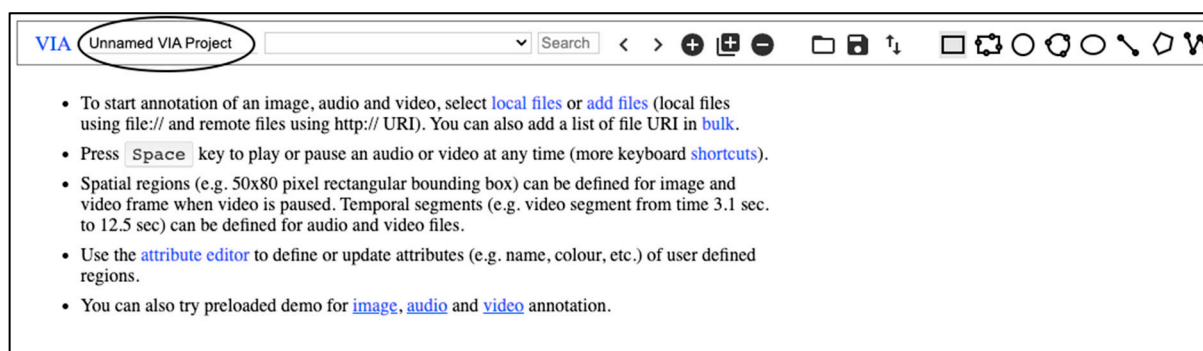


Figure A1. Create a new project.

**Step 2: Adding a video file:** The second step is to add a video by clicking the plus icon (refer to Figure A1). Select the video to be annotated from the desktop or cloud storage.

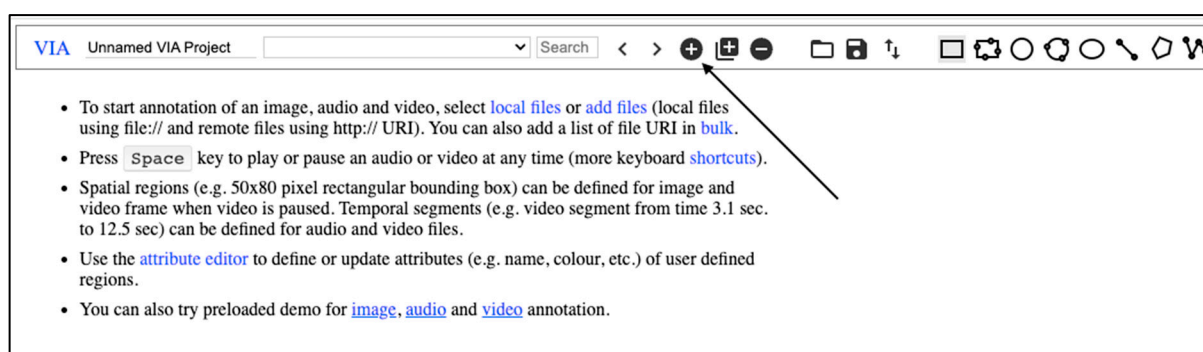


Figure A2. Add a video.

**Step 3: Define the attributes:** Once the video is added, define the attributes by clicking on 1 (refer to Figure A2). In this step, two attributes have been created by typing the attribute name in 2 (refer to Figure A2) and clicking Create. In this project, the first attribute was created to identify the engaging teaching video indicators and the second to highlight the presenter's location in the video.

While defining the attributes, the following information was inserted (refer to Figure A3):

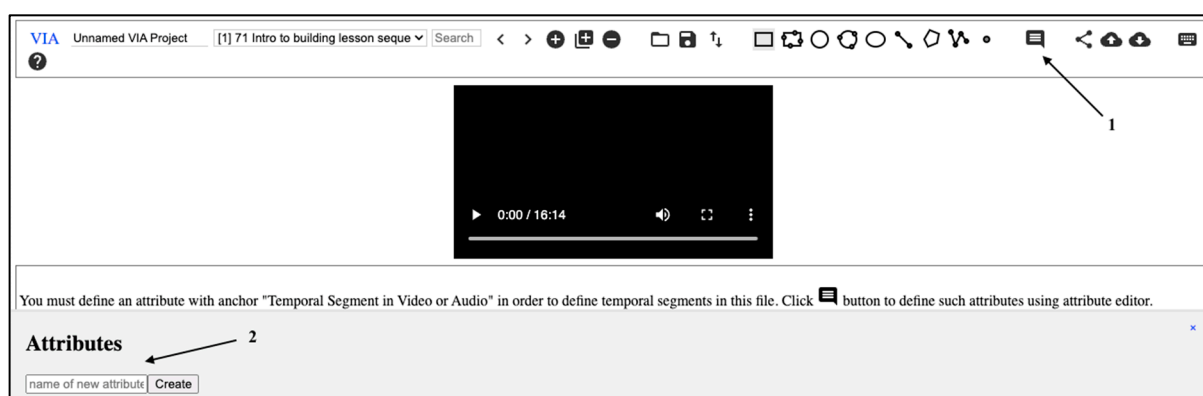


Figure A3. Define the attributes.



**Attribute 1:** The name of the first attribute is “Engaging teaching video indicators”. The anchor is set to “Temporal Segment in Video or Audio” as researchers identified the indicators in small video segments. The text function is selected for the input type (refer to Figure A4).


**Attribute 2:** The name of the second attribute is “Presenter location”. The attribute is created to signal the presenter’s location in the video. The anchor is set to “Spatial region in a video frame” as an area is highlighted to indicate the presenter’s location. The input type is set as Select. In the options section, the researchers have typed “presenter” to

**Name** = Presenter location

**Anchor** = Spatial region in a video frame



**Input Type** = Select

**Options** = \*Presenter (Note: if there are multiple presenters in a video, we can add \*presenter 1, presenter 2)

You must define an attribute with anchor "Temporal Segment in Video or Audio" in order to define temporal segments in this file. Click  button to define such attributes using attribute editor.

### Attributes

name of new attribute

	Id	Name	Anchor	Input Type	Description	Options	Default Value	Preview
	1	Engaging teaching vi	Temporal Segment in V	TEXT		-	-	
	2	Presenter location	Spatial Region in a Vid	SELECT		Presenter	Not Defined	Presenter

**Figure A4.** Attribute 1 and 2.

**Step 4: Adding indicators to Attribute 1 (engaging teaching video indicators):** After defining the attributes, the next step is adding the indicators. The researchers added the indicators at the bottom left-hand side by writing the indicator name and then clicking Add (refer to Figure A5). The following indicators have been added.

Indicators	Description
1. Encouraging students' participation in discussion	Teachers to engage students in discussions or debates to attract their interest and motivate a deeper understanding
2. Encouraging students to share their knowledge and ideas	Teachers to ask for students' participation in active learning methods by sharing their perceptions, knowledge, and ideas
3. Encouraging students to ask questions	Teachers to create a safe and open environment that allows students to ask their questions, to enhance the student interaction experience
4. Encouraging collaborative learning activities	Teachers to create opportunities for students to interact with each other through group activities or collaborative work
5. Encouraging meaningful interaction	Teachers to construct a welcoming and efficient online learning environment by fostering regular and meaningful communication with students and providing meaningful answers to students' enquiries
6. Providing learning resources	Teachers to provide students with various learning resources, videos, etc., to increase students' active participation
7. Giving clear instructions	Teachers to be clear and detailed in communicating the instructions, expectations, roles, and responsibilities, to show commitment to meeting the course goals
8. Outlining the learning objectives	Teachers to clearly outline and communicate the topics and instructions to increase student engagement in online learning
9. Using appropriate changes in tone of voice	Teachers to read and respond to perceived restlessness by using appropriate changes in tone of voice or changes in direction

10. Facial expressions	Teachers to maintain appropriate facial expressions such as smiling and nodding
11. Eye contact	Teachers to maintain eye contact with students in online learning
12. Appropriate body language	Teachers to maintain appropriate body language in the online classroom
13. Enabling class recording for later review	Teachers to increase the value of the online learning experience by enabling class recording, which allows students access to classroom sessions from the comfort of their home and if they want to review afterwards
14. Screen sharing and enabling chat, camera, and microphone	Teachers to assure students of their presence and positively impact student engagement and satisfaction by communicating in real-time through a chat, camera, microphone, and screen sharing
15. Varying the presentation media	Teachers to vary the presentation media (e.g., videos, slides, note sharing, etc.) to capture students' attention and foster engagement



**Figure A5.** Adding indicators.

Step 5: Drawing a boundary box by clicking on 1 to signal the presenter's location by clicking on 2 (Attribute 2: presenter location): The researchers drew a boundary box to indicate the presenter's location (refer to Figure A6).

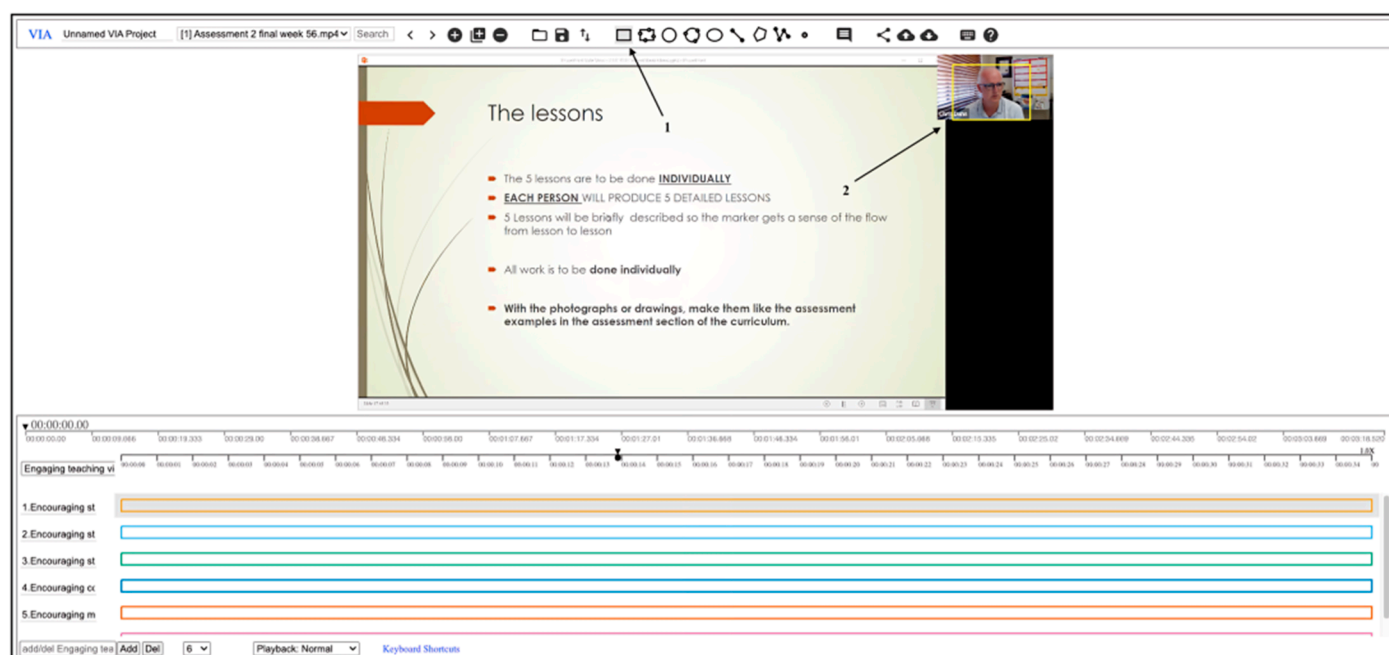


Figure A6. Drawing boundary box.

**Step 6: Identifying the indicators from the video:** Manual annotation is performed after defining the attributes and indicating the presenter's location. In this process, the video is played, and indicators are identified in small segments (refer to arrows in Figure A7). To start the temporal segment, click "a", and to stop it, click "Shift" + "a".

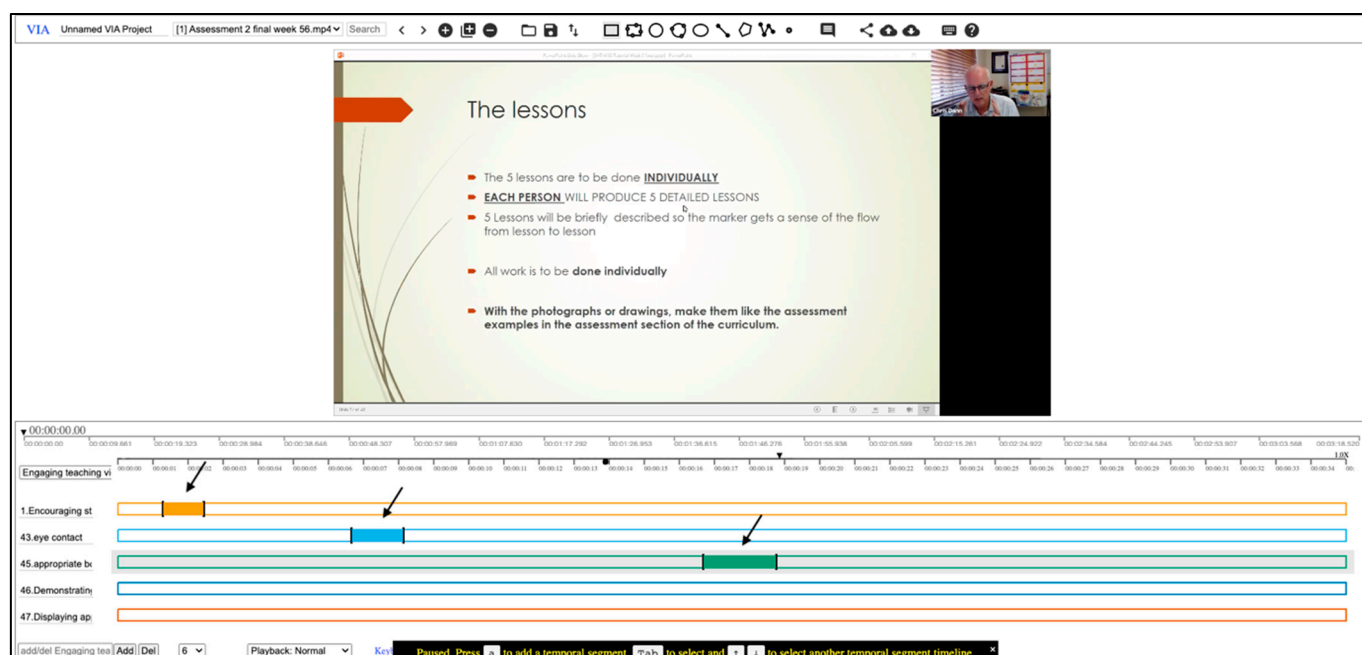


Figure A7. Identifying the indicators.

**Step 7: Saving and Exporting the Project for Machine Learning:** Once the annotation is complete, save the project by clicking on 1 and selecting the project's location. Similarly, click on 2 to export the project (refer to Figure A8).

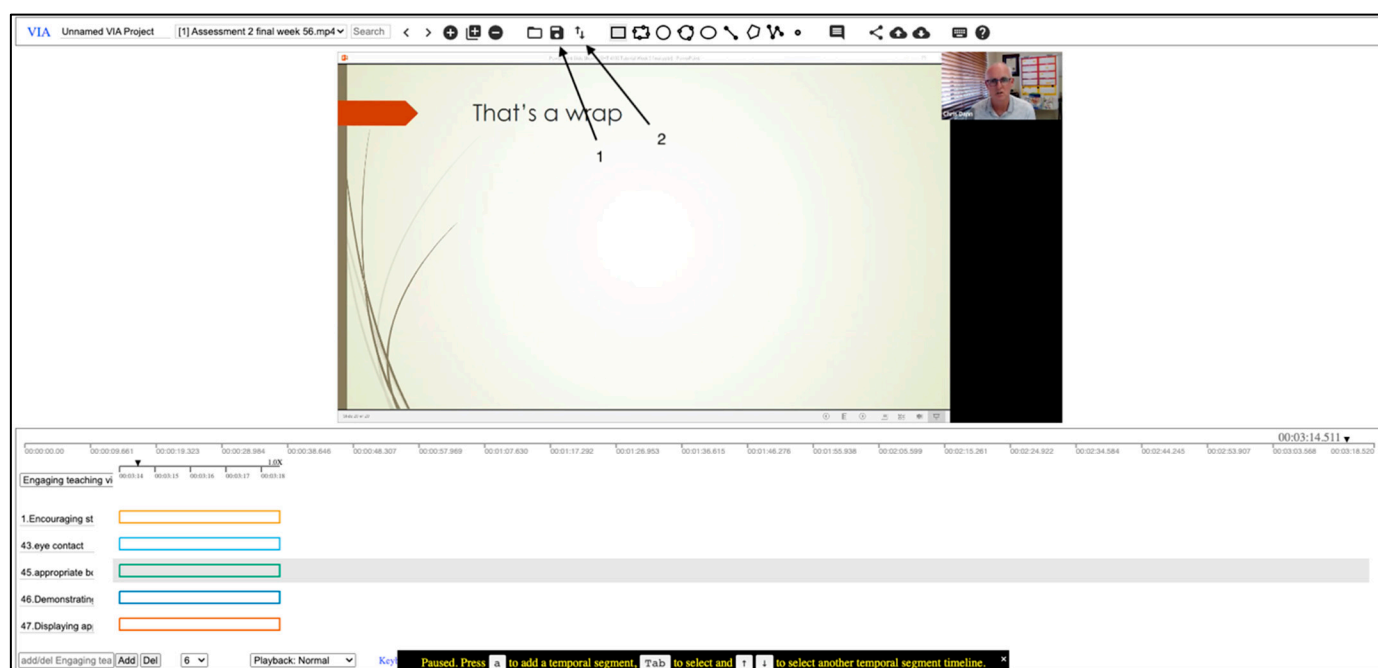


Figure A8. Save and export.

## References

- Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for Artificial Intelligence: Concepts, opportunities, and challenges. *Applied Sciences*, 13(12), 7082. <https://doi.org/10.3390/app13127082>.
- Alenezi, E., Alfadley, A. A., Alenezi, D. F., & Alenezi, Y. H. (2022). The sudden shift to distance learning: Challenges facing teachers. *Journal of Education and Learning*, 11(3), 14. <https://doi.org/10.5539/jel.v11n3p14>.
- Apicella, A., Arpaia, P., Frosolone, M., Improta, G., Moccaldi, N., & Pollastro, A. (2022). EEG-based measurement system for monitoring student engagement in learning 4.0. *Scientific Reports*, 12(1), 5857. <https://doi.org/10.1038/s41598-022-09578-y>.
- Ashwin, T. S., & Guddeti, R. M. R. (2019). Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks. *Education and Information Technologies*, 25(2), 1387–1415. <https://doi.org/10.1007/s10639-019-10004-6>.
- Beaver, I., & Mueen, A. (2022). On the care and feeding of virtual assistants: Automating conversation review with AI. *AI Magazine*, 42(4), 29–42. <https://doi.org/10.1609/aaai.12024>.
- Behera, A., Matthew, P., Keidel, A., Vangorp, P., Fang, H., & Canning, S. (2020). Associating facial expressions and upper-body gestures with learning tasks for enhancing intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 30(2), 236–270. <https://doi.org/10.1007/s40593-020-00195-2>.
- Bell, J. (1999). *Doing your research project: A guide for first-time researchers in education and social science* (3rd ed.). Open University Press.
- Castro, M. D. B., & Tumibay, G. M. (2021). A literature review: Efficacy of online learning courses for higher education institution using meta-analysis. *Education and Information Technologies*, 26, 1367–1385. <https://doi.org/10.1007/s10639-019-10027-z>.
- Cents-Boonstra, M., Lichtwarck-Aschoff, A., Lara, M. M., & Denessen, E. (2021). Patterns of motivating teaching behaviour and student engagement: A microanalytic approach. *European Journal of Psychology of Education*, 37, 227–255. <https://doi.org/10.1007/s10212-021-00543-3>.
- Chiu, T. K. F. (2021). Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic. *Journal of Research on Technology in Education*, 54(Suppl. 1), S14–S30. <https://doi.org/10.1080/15391523.2021.1891998>.
- De Silva, D., Kaynak, O., El-Ayoubi, M., Mills, N., Alahakoon, D., & Manic, M. (2024). Opportunities and challenges of Generative artificial intelligence: research, education, industry engagement, and social impact. *IEEE Industrial Electronics Magazine*, 2–17. <https://doi.org/10.1109/mie.2024.3382962>.
- Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. <https://doi.org/10.1177/0047239520934018>.
- Giang, T. T. T., Andre, J., & Lan, H. H. (2022). Student engagement: Validating a model to unify in-class and out-of-class Contexts. *Journal of Education and Learning*, 8(4), 1–14. <https://doi.org/10.1177/21582440221140334>.

- Gillett-Swan, J. (2017). The challenges of online learning: Supporting and engaging the isolated learner. *Journal of Learning Design*, 10(1), 20–30. <https://doi.org/10.5204/jld.v9i3.293>.
- Harry, A., & Sayudin, S. (2023). Role of AI in education. *Interdisciplinary Journal and Humanity (Injurity)*, 2(3), 260–268. <https://doi.org/10.58631/injurity.v2i3.52>.
- Heale, R., & Twycross, A. (2018). What is a case study? *Evidence-Based Nursing*, 21(1), 7–8. <https://doi.org/10.1136/eb-2017-102845>.
- Heeg, D. M., & Avraamidou, L. (2023). The use of Artificial intelligence in school science: A systematic literature review. *Educational Media International*, 60(2), 125–150. <https://doi.org/10.1080/09523987.2023.2264990>.
- Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, 47(2), 320–341. <https://doi.org/10.1111/bjet.12235>.
- Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial Intelligence-Enabled personalised recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, 194, 104684. <https://doi.org/10.1016/j.compedu.2022.104684>.
- Kvale, S. (1996). *Interview views: An Introduction to qualitative research interviewing*. Sage Publications.
- Lee, J., Song, H., & Hong, A. J. (2019). Exploring factors, and indicators for measuring students' sustainable engagement in e-Learning. *Sustainability*, 11(4), 985. <https://doi.org/10.3390/su11040985>.
- Li, J., Lin, F., Yang, L., & Huang, D. (2023). AI service placement for Multi-Access Edge Intelligence systems in 6G. *IEEE Transactions on Network Science and Engineering*, 10(3), 1405–1416. <https://doi.org/10.1109/tNSE.2022.3228815>.
- Liang, R., & Chen, D. T. V. (2012). Online learning: Trends, potential and challenges. *Creative Education*, 3(8), 1332. <https://doi.org/10.4236/ce.2012.38195>.
- Limna, P., Jakwatanatham, S., Siripattanakul, S., Kaewpuang, P., & Sriboonruang, P. (2022). A review of artificial intelligence (AI) in education during the digital era. *Advance Knowledge for Executives*, 1(1), 1–9. Available online: <https://ssrn.com/abstract=4160798> (accessed on 5 January 2024).
- Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, 24, 26–34. <https://doi.org/10.1016/j.iheeduc.2014.09.005>.
- Ma, X., Xu, M., Dong, Y., & Sun, Z. (2021). Automatic student engagement in online learning environment based on Neural Turing Machine. *International Journal of Information and Education Technology*, 11(3), 107–111. <https://doi.org/10.18178/ijiet.2021.11.3.1497>.
- Murtaza, M., Ahmed, Y., Shamsi, J. A., Sherwani, F., & Usman, M. (2022). AI-Based personalised E-Learning systems: Issues, challenges, and solutions. *IEEE Access*, 10, 81323–81342. <https://doi.org/10.1109/access.2022.3193938>.
- Nguyen, N. D. (2023). Exploring the role of AI in education. *London Journal of Social Sciences*, 6, 84–95. <https://doi.org/10.31039/ljss.2023.6.108>.
- Nikoloutsopoulos, S., Koutsopoulos, I., & Titsias, M. K. (2024, May 5–8). Kullback-Leibler reservoir sampling for fairness in continual learning. 2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN) (pp. 460–466), Stockholm, Sweden. <https://doi.org/10.1109/icmlcn59089.2024.10624806>.
- Ocaña, M. G., & Opdahl, A. L. (2023). A software reference architecture for journalistic knowledge platforms. *Knowledge-Based Systems*, 276, 110750. <https://doi.org/10.1016/j.knsys.2023.110750>.
- Pianykh, O. S., Langs, G., Dewey, M., Enzmann, D. R., Herold, C. J., Schoenberg, S. O., & Brink, J. A. (2020). Continuous Learning AI in radiology: Implementation principles and early applications. *Radiology*, 297(1), 6–14. <https://doi.org/10.1148/radiol.2020200038>.
- Pingenot, A., & Shanteau, J. (2009). Expert opinion. In M. W. Kattan (Ed.), *Encyclopedia of medical decision making*. Sage Publications, Inc. Available online: [https://www.researchgate.net/publication/263471207\\_Expert\\_Opinion](https://www.researchgate.net/publication/263471207_Expert_Opinion) (accessed on 2 January 2024).
- Roshanaei, M., Khan, M. R., & Sylvester, N. N. (2024). Enhancing Cybersecurity through AI and ML: Strategies, Challenges, and Future Directions. *Journal of Information Security*, 15(3), 320–339. <https://doi.org/10.4236/jis.2024.153019>.
- Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection, and analysis techniques. *Research in Nursing & Health*, 23(3), 246–255. [https://doi.org/10.1002/1098-240X\(200006\)23:3<246::AID-NUR9>3.0.CO;2-H](https://doi.org/10.1002/1098-240X(200006)23:3<246::AID-NUR9>3.0.CO;2-H).
- Shaikh, A. A., Kumar, A., Jani, K., Mitra, S., García-Tadeo, D. A., & Devarajan, A. (2022). The role of Machine Learning and Artificial Intelligence for making a digital classroom and its sustainable impact on education during COVID-19. *Materials Today Proceedings*, 56, 3211–3215. <https://doi.org/10.1016/j.matpr.2021.09.368>.
- Shekhar, P., Prince, M. J., Finelli, C. J., DeMonbrun, M., & Waters, C. (2018). Integrating quantitative and qualitative research methods to examine student resistance to active learning. *European Journal of Engineering Education*, 44(1–2), 6–18. <https://doi.org/10.1080/03043797.2018.1438988>.

- Tahiru, F. (2021). AI in education. *Journal of Cases on Information Technology*, 23(1), 1–20. <https://doi.org/10.4018/jcit.2021010101>.
- Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-Based research in the educational field: A Systematic literature review. *Education Sciences*, 12(6), 410. <https://doi.org/10.3390/educsci12060410>.
- Turner, D. J. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report*, 15(3), 754–760. <https://doi.org/10.46743/2160-3715/2010.1178>.
- Verma, N., Getenet, S., Dann, C., & Shaik, T. (2023a). Characteristics of engaging teaching videos in higher education: a systematic literature review of teachers' behaviours and movements in video conferencing. *Research and Practice in Technology Enhanced Learning*, 18, 040. <https://doi.org/10.58459/rptel.2023.18040>
- Verma, N., Getenet, S., Dann, C., & Shaik, T. (2023b). Designing an artificial intelligence tool to understand student engagement based on teacher's behaviours and movements in video conferencing. *Computers & Education: Artificial Intelligence*, 5, 100187. <https://doi.org/10.1016/j.caeai.2023.100187>
- Wang, C., Yang, Z., Li, Z. S., Damian, D., & Lo, D. (2024). Quality assurance for Artificial intelligence: A study of industrial concerns, challenges and best practices. *arXiv*, arxiv:2402.16391. <https://doi.org/10.48550/arxiv.2402.16391>.
- Wang, X., & Yin, M. (2023, April 23–28). *Watch out for updates: Understanding the effects of model explanation updates in ai-assisted decision making*. 2023 CHI Conference on Human Factors in Computing Systems (pp. 1–19), Hamburg, Germany. <https://doi.org/10.1145/3544548.3581366>.
- Weng, X., Ng, O.-L., & Chiu, T. K. F. (2023). Competency development of pre-service teachers during video-based learning: A systematic literature review and meta-analysis. *Computers & Education*, 199, 104790. <https://doi.org/10.1016/j.compedu.2023.104790>.
- Xie, J., A, G., Rice, M. F., & Griswold, D. E. (2021). Instructional designers' shifting thinking about supporting teaching during and post-COVID-19. *Distance Education*, 42, 1–21. <https://doi.org/10.1080/01587919.2021.1956305>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.