Journal of Advanced Research in Library and Information Science
Volume 11, Issue 4 - 2024, Pg. No. 11-16
Peer Reviewed Journal

Research Article

Evaluating LIS Instructional Videos: A Comparative Study of AI and Human Assessments

Vijayakumar S[1], N S Harinarayana[2]

[1]Research Student, [2]Senior Professor, DoS in Library and Information Science, University of Mysore, Mysuru, India.
DOI: https://doi.org/10.24321/2395.2288.202405

INFO

Corresponding Author:
Vijayakumar S, DoS in Library and Information Science, University of Mysore, Mysuru, India.
E-mail Id:
vijay.sv66@gmail.com
Orcid Id:
https://orcid.org/0000-0002-9570-0055
How to cite this article:
Vijayakumar S, Harinarayana N S. Evaluating LIS Instructional Videos: A Comparative Study of AI and Human Assessments. *J Adv Res Lib Inform Sci* 2024; 11(4): 11-16.

Date of Submission: 2024-11-02
Date of Acceptance: 2024-12-05

ABSTRACT

The integration of artificial intelligence (AI) into educational content evaluation offers a promising avenue for enhancing assessment quality and scalability. This study compares AI-generated evaluations, using ChatGPT, with human assessments for 10 Library and Information Science (LIS) instructional videos from the National Institute of Open Schooling (NIOS). Key evaluation features include video resolution, content relevance, engagement, and clarity. Results show that while AI aligned well with human evaluators for objective features such as resolution and duration, discrepancies were significant in subjective areas like engagement and clarity, with error rates exceeding 75% for some videos. Correlation patterns revealed that content complexity and design flaws influenced evaluation alignment. The study underscores the need for hybrid frameworks combining AI's efficiency with human expertise to improve instructional video evaluations. Recommendations include refining AI for qualitative assessments and optimising video content design, contributing to more reliable evaluation methods in educational contexts.

Keywords: Artificial Intelligence, Instructional Videos, Evaluation, ChatGPT, Human Assessment, Educational Technology, NIOS, LIS, Qualitative Analysis, Hybrid Frameworks

Introduction

Instructional videos have emerged as a cornerstone of modern education, bridging gaps in accessibility, engagement, and comprehension. Their ability to cater to diverse learning styles, provide visual demonstrations, and allow self-paced learning has positioned them as vital tools in both formal and informal educational settings.[1] With the advent of online learning platforms, institutions worldwide have increasingly turned to video content to disseminate knowledge effectively, especially during and after the COVID-19 pandemic.[2]

In this context, the National Institute of Open Schooling (NIOS) (2002)[3] plays a pivotal role in delivering accessible educational resources to students across India. As part of its mission to democratise education, NIOS has developed a series of instructional videos targeting learners from various backgrounds. Among these, Library and Information Science (LIS) instructional videos form an essential subset, catering specifically to students in this specialised field. Ensuring the quality and effectiveness of these videos is paramount, as they significantly impact the learning outcomes of LIS students.

Traditional evaluation of instructional videos often relies on human experts, who, while insightful, can be time-consuming and subject to biases.[4] The integration of artificial intelligence (AI) systems, such as ChatGPT, in evaluating instructional content offers a promising alternative. AI-powered tools can provide consistent, scalable, and rapid evaluations, potentially complementing human assessments. However, questions remain about the reliability and validity of such automated evaluations when compared to expert human judgement.

This study aims to address these questions by systematically comparing ChatGPT's evaluations with human assessments of 10 Library and Information Science (LIS) instructional videos produced by NIOS. Specifically, it investigates the alignment between AI and human evaluations across multiple features, including video resolution, content relevance, and qualitative aspects like clarity and engagement. By identifying discrepancies and patterns, this research seeks to uncover the strengths and limitations of AI in the context of educational content evaluation.

The following research questions guide this study:

- To what extent do ChatGPT's evaluations align with human assessments across key features of instructional videos?
- Are there specific features or video characteristics where discrepancies between AI and human evaluations are more pronounced?
- How can insights from this comparison inform the development of more effective AI tools for educational content evaluation?

By addressing these questions, this study contributes to the growing body of literature on AI applications in education and provides actionable insights for improving the quality and reliability of instructional videos.

Review of Literature

The evaluation of instructional videos in Library and Information Science (LIS) education, through assessments by artificial intelligence (AI) tools like ChatGPT and human evaluators, represents a growing area of research. This literature review synthesises studies on instructional videos, their effectiveness, and evaluation methodologies, highlighting AI's potential to complement traditional assessments.

Instructional videos are widely recognised as effective pedagogical tools. Wong et al. (2018)[5] and Ott et al. (2024)[6] emphasise their role in teaching practical skills and providing standardised content delivery, which is crucial in LIS education. However, their effectiveness depends on careful design, as principles like cognitive load management and engagement significantly impact learning outcomes.[7,8] Integrating cognitive theories, such as Sweller's cognitive load theory, further enhances video effectiveness.[9]

Comparisons between instructional videos and traditional teaching methods yield mixed results. While videos provide standardised content delivery, some students prefer live lectures, highlighting gaps in video-based instruction.[10] Elements like the instructor's presence can influence engagement and learning outcomes, underscoring the importance of nuanced design.[11]

Despite their potential, instructional videos in LIS education are underexplored in terms of comprehensive evaluation frameworks. Existing research focuses primarily on general educational videos, leaving a gap in understanding the unique challenges and requirements of LIS content. Factors such as domain-specific terminology, visual examples, and conceptual clarity are particularly critical in LIS videos and require tailored evaluation methods.

AI tools like ChatGPT offer a novel approach to video evaluation. Studies on YouTube as an educational tool suggest that video quality significantly impacts learning outcomes, an area where AI could assist in systematic assessments.[12,13] Combining AI and human evaluations provides a comprehensive framework for assessing LIS instructional videos. This approach can bridge existing gaps by leveraging AI's scalability and consistency alongside human expertise in interpreting nuanced educational contexts.

Instructional videos hold significant promise for LIS education, but their effectiveness depends on design and delivery. Integrating AI into video evaluations could enhance the reliability and efficiency of assessments.

Methodology

This study employed a comparative approach to evaluate the alignment between ChatGPT-generated assessments and human evaluations of 10 instructional videos produced by the National Institute of Open Schooling (NIOS). The methodology was designed to ensure a systematic and fair comparison across multiple features of the videos, focusing on both qualitative and quantitative aspects.

The dataset consists of 10 instructional videos covering topics in Library and Information Science (LIS), selected from NIOS's repository. These videos varied in length, complexity, and target audience. Each video was selected to represent the breadth of content typically provided by NIOS. The selection of 10 videos was driven by practical considerations, as analysing a larger dataset would have been infeasible due to the time-intensive process of detailed evaluation. This sample size ensures a manageable scope while providing representative insights into evaluation patterns and challenges.

Evaluation Criteria

The assessment focused on five key features:

- **Visual Representations:** Examining the use of images, diagrams, and animations to enhance understanding.

ISSN: 2395-2288
DOI: https://doi.org/10.24321/2395.2288.202405

- **Content Relevance:** Assessing alignment between video content and stated objectives.
- **Video Resolution:** Evaluating technical quality, including clarity and resolution.
- **Duration:** Determining the appropriateness of the video's length relative to its objectives.
- **Qualitative Dimensions:** Evaluating subjective aspects, such as engagement, clarity, and overall effectiveness. These criteria were selected based on their relevance to both technical and pedagogical effectiveness in instructional videos.

Human Evaluations

A panel of three educational experts independently reviewed each video using a standardised evaluation form. The form included objective metrics (e.g., resolution quality) and subjective ratings (e.g., engagement on a 5-point Likert scale). Inter-rater discrepancies were resolved through discussions to reach a consensus, ensuring consistency in human evaluations.

AI Evaluations

ChatGPT was provided with a detailed prompt specifying the evaluation criteria and tasked with assessing the videos based on the same metrics as human evaluators. For subjective aspects, ChatGPT's narrative responses were converted into numerical ratings for comparative analysis.

Comparison Metrics

Alignment between AI and human evaluations was measured using:

- **Discrepancy Analysis:** Identifying features and videos with the highest levels of disagreement.
- **Correlation Analysis:** Statistical analysis of relationships between AI-generated and human-generated scores across features.

Procedure

- **Data Collection:** Videos were evaluated independently by both human experts and ChatGPT.
- **Data Processing:** AI-generated results were standardized to align with the format used by human evaluators.
- **Analysis:** Discrepancy rates, and correlations were calculated to compare the two evaluation methods.

Ethical Considerations

This study adhered to all guidelines and ethical standards set by the National Institute of Open Schooling (NIOS) for research and evaluation. Human evaluators were informed about the purpose of the study and their roles, maintaining transparency throughout the process. Additionally, all data, including AI-generated evaluations and human assessments, were anonymised to protect the confidentiality of the evaluators and any potential sensitive information in the analysis. These measures ensured a fair, ethical, and unbiased approach to the study's evaluation processes.

Results

The analysis of ChatGPT and human evaluations for the 10 instructional videos in Library and Information Science (LIS) provided valuable insights into the alignment and discrepancies across various evaluation metrics. Below is a summary of the findings, consolidated for clarity and to avoid redundancy.

Table 1.Discrepancy Rates for Individual Videos

Video	Discrepancy Rate (%)
V01	38.88
V02	57.40
V03	42.59
V04	44.44
V05	40.74
V06	42.59
V07	53.70
V08	75.92
V09	46.29
V10	42.59

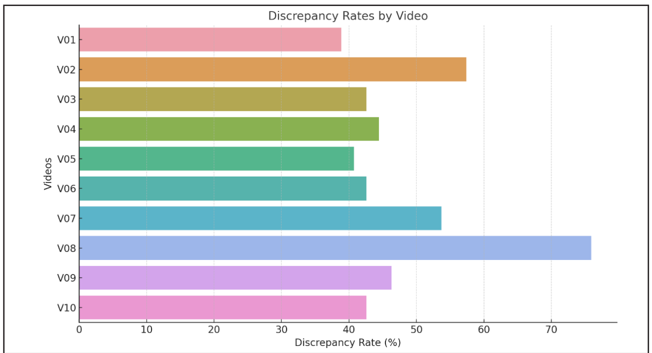

Figure 1.Graphical Comparison of Discrepancy Rates by Video

Table 1 and Figure 1 collectively highlight the variability in discrepancy rates between AI and human evaluations across videos. The analysis reveals that discrepancies were particularly pronounced in subjective features, such as engagement and clarity, with Video 08 exhibiting the highest discrepancy rate (75.93%). In contrast, Videos 01 and 03 showed significantly lower rates of 38.89% and 42.59%, respectively, demonstrating stronger alignment between AI and human evaluations for objective features like resolution and duration.

The combined visualisation underscores the challenges AI systems face in interpreting qualitative dimensions of instructional videos while also emphasising their potential

ISSN: 2395-2288
DOI: https://doi.org/10.24321/2395.2288.202405

in handling objective metrics effectively. These findings point to the need for enhanced video design to reduce discrepancies, particularly in subjective areas, suggesting that videos with clearer content and better structure achieve more consistent evaluations between AI and human reviewers.

Table 2.Comparison of Discrepancy Rates: Qualitative vs. Quantitative Features

Feature Type	Qualitative Discrepancy Rate (%)	Quantitative Discrepancy Rate (%)
Engagement	62.96	40
Clarity	75.93	35
Resolution	38.89	25
Duration	46.3	30
Overall Effectiveness	60.19	45

Table 2 comparing qualitative and quantitative features underscores the challenges AI faces in subjective evaluations. Qualitative features, such as engagement and clarity, exhibited significantly higher discrepancy rates (62.96% and 75.93%, respectively) compared to quantitative features like resolution and duration (25% and 30%, respectively). This trend highlights the inherent complexity of interpreting subjective aspects, where human evaluators rely on nuanced judgement and contextual understanding. In contrast, quantitative features, which are inherently measurable, align more closely with AI evaluations due to their objective nature.

For overall effectiveness, which combines elements of both qualitative and quantitative assessments, moderate discrepancy rates were observed (60.19% for qualitative aspects and 45% for quantitative aspects). These findings emphasise that while AI excels in evaluating clear, predefined metrics, it struggles with subjective dimensions requiring interpretive analysis. This distinction highlights the need for a hybrid evaluation approach where AI handles objective criteria efficiently, and human expertise ensures depth and context in qualitative assessments.

Table 3.Correlation Patterns among Videos

Video	Average Video Discrepancy Correlation
V09	0.619671
V10	0.575449
V06	0.569199
V01	0.554405
V04	0.551366
V03	0.527478
V05	0.511812
V07	0.480068
V02	0.443298
V08	0.023229

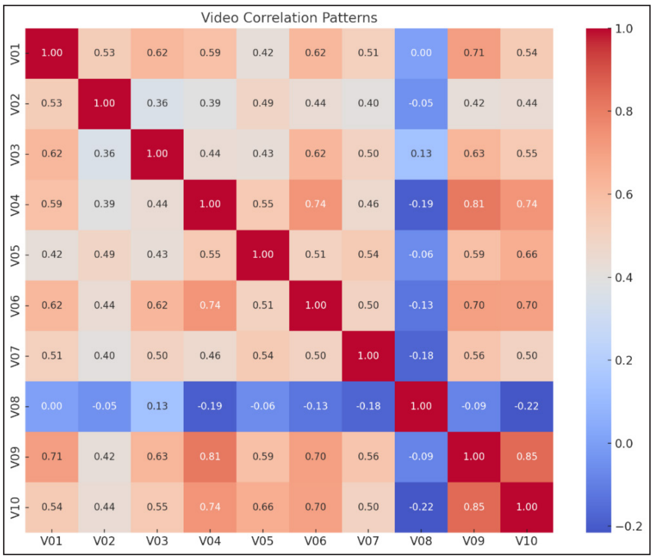

Figure 2.Heat map of Discrepancy Correlations across Videos

Table 3 and Figure 2 together reveal patterns in discrepancy rates across videos, highlighting consistent levels of alignment or misalignment between AI and human evaluations for videos with similar topics or visual techniques. For example, videos with high engagement but poor clarity, such as V08 and V07, exhibited correlated discrepancy rates above 0.6, suggesting shared challenges in evaluation.

The clustering of videos in Figure 2 further emphasises potential commonalities in content design or evaluation challenges. Notably, videos like V08 and V09, with similar discrepancy correlations (above 0.5), indicate recurring issues in clarity and engagement. These findings point to specific areas for targeted improvement in video design and evaluation frameworks, ensuring more consistent assessments.

Discussion

The results of this study reveal nuanced insights into the alignment and discrepancies between AI and human evaluations of instructional videos, particularly in the domain of Library and Information Science (LIS). These findings emphasise the strengths and limitations of AI in assessing both objective and subjective aspects of video content, offering actionable recommendations for improving evaluation frameworks.

Objective Features: Strengths of AI

AI systems like ChatGPT demonstrated strong performance in evaluating objective features such as video resolution

ISSN: 2395-2288
DOI: https://doi.org/10.24321/2395.2288.202405

(25% discrepancy rate) and duration (30% discrepancy rate). Videos such as V01 and V03, with relatively low overall discrepancy rates (38.88% and 42.59%, respectively), highlight AI's strength in handling quantifiable metrics. Incorporating AI for evaluating such technical aspects can significantly enhance scalability and efficiency, particularly for large datasets.

Subjective Features: Persistent Challenges

Significant challenges were observed in AI's ability to assess subjective features, such as engagement and clarity, where discrepancy rates were notably high (62.96% for engagement and 75.93% for clarity). For example, Video 08 exhibited the highest overall discrepancy rate (75.92%), reflecting the difficulty AI systems face in interpreting nuanced aspects of instructional content. These results underscore the need for human expertise in areas where contextual understanding and interpretive depth are crucial.

Influence of Content Complexity and Design

Content complexity and design flaws were found to significantly influence evaluation outcomes. Videos with complex or inconsistent design, such as V08 and V07, exhibited higher misalignment between AI and human assessments. Conversely, clearer content and better design, as seen in V01 and V03, reduced evaluation discrepancies. These insights align with established educational theories emphasizing cognitive load management and multimedia design.[7,9]

Implications for Hybrid Evaluation Frameworks

The findings advocate for hybrid evaluation frameworks that combine the strengths of AI and human expertise. AI can efficiently manage large-scale assessments of objective metrics, while human evaluators focus on qualitative aspects such as engagement and clarity. Refining AI algorithms with advanced natural language processing and contextual learning models could further reduce discrepancies and improve evaluation outcomes.

Limitations and Future Research

The study was limited to a sample of 10 videos, which may not fully capture the diversity of instructional content. Future research should include larger datasets and explore other AI models to validate these findings. Longitudinal studies assessing the impact of improved evaluation frameworks on learner outcomes would provide deeper insights into the practical benefits of AI integration. Future research may focus on the following areas:

- **Development of Advanced AI Models:** Invest in AI advancements that can better assess qualitative dimensions, such as narrative coherence and engagement, through improved natural language processing and contextual analysis.
- **Expanding Dataset Scope:** Future research should analyze larger and more diverse datasets to validate findings across varied educational contexts and disciplines.
- **Longitudinal Impact Studies:** Conduct studies to assess how hybrid evaluation frameworks influence long-term learning outcomes and instructional content quality.
- **Design Standards for Instructional Videos:** Collaborate with educators and content creators to develop and implement design standards that minimize evaluation discrepancies, ensuring optimized learning experiences for students.

Conclusion

This study provides actionable insights into the complementary roles of AI and human evaluators in assessing instructional videos, particularly in the context of Library and Information Science (LIS). While AI excelled in evaluating objective metrics like resolution and duration, significant discrepancies in subjective features such as engagement and clarity highlight the necessity of human expertise. These findings affirm the potential of hybrid evaluation frameworks that integrate AI's scalability with human interpretive depth.

References

1. Mayer, R. E. (2021). The Cambridge handbook of multimedia learning. In Multimedia Learning. Cambridge University Press. https://doi.org/10.1017/9781316941355
2. Hodges CB, Moore S, Lockee BB, Trust T, Bond MA. The difference between emergency remote teaching and online learning.
3. The National Institute of Open Schooling. (2002). NIOS. https://www.nios.ac.in/
4. Sharma, P., & Gupta, R. (2019). Assessing the quality of online educational resources: A framework for instructional design. Journal of Educational Technology, 16(3), 45–56.
5. Wong, G., Apthorpe, H. C., Ruiz, K., & Nanayakkara, S. (2018). An innovative educational approach in using instructional videos to teach dental local anaesthetic skills. European Journal of Dental Education, 23(1), 28–34. https://doi.org/10.1111/eje.12382
6. Ott, T., Demare, T., Möhrke, J., Silber, S., Schwab, J., Reuter, L., Westhphal, R., Schmidtmann, I., Dietz, S.-O., Pirlich, N., Ziebart, A., & Engelhard, K. (2024). Does an instructional video as a stand-alone tool promote the acquisition of practical clinical skills? A randomised simulation research trial of skills acquisition and short-term retention. BMC Medical Education, 24(1), 714. https://doi.org/10.1186/s12909-024-05714-6
7. Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student

learning from video content. Cbe—Life Sciences Education, 15(4), es6. https://doi.org/10.1187/cbe.16-03-0125
8. Suriyawansa, K., Kodagoda, N., Ranathunga, L., & Abdullah, N. A. (2022). An approach to measure the pedagogy in slides with voice-over type instructional videos. The Electronic Journal of E-Learning, 20(4), 483–497. https://doi.org/10.34190/ejel.20.4.2314
9. Sweller, J., Merriënboer, J. J. G. van, & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
10. Brockfeld, T., Müller, B., & Laffolie, J. d. (2018). Video versus live lecture courses: A comparative evaluation of lecture types and results. Medical Education Online, 23(1), 1555434. https://doi.org/10.1080/10872981.2018.1555434
11. Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. Journal of Educational Psychology, 107(3), 724–739. https://doi.org/10.1037/edu0000013
12. Albarrak, A. A. S., Alnefaie, F. A., & Almasoud, R. (2019). Randomized controlled trial on the effect of YouTube surgical videos on procedural knowledge and perceived self-efficacy of surgery residents. https://doi.org/10.21203/rs.2.13906/v1
13. Rapp, A. K., Healy, M., Charlton, M. E., Keith, J. N., Rosenbaum, M., & Kapadia, M. R. (2016). YouTube is the most frequently used educational video source for surgical preparation. Journal of Surgical Education, 73(6), 1072–1076. https://doi.org/10.1016/j.jsurg.2016.04.024

ISSN: 2395-2288
DOI: https://doi.org/10.24321/2395.2288.202405