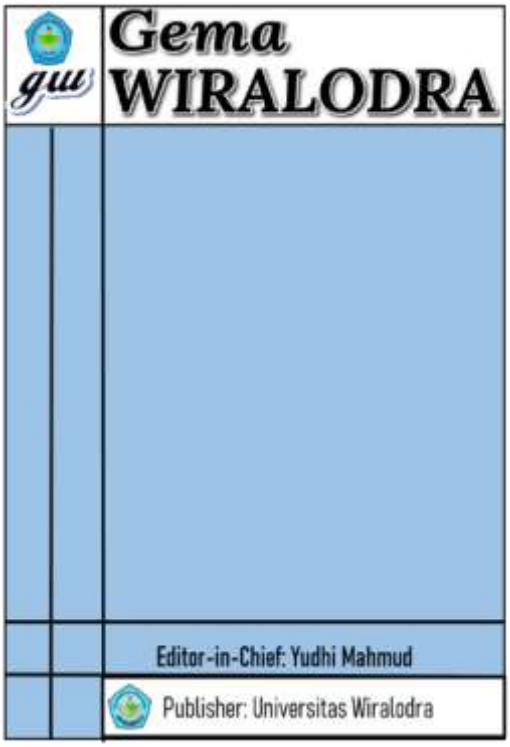
Gema Wiralodra

Publication details, including instructions for authors and subscription information:
https://gemawiralodra.unwir.ac.id



Optimizing AI-Based Video Summarization for Educational Media: A Comparative Evaluation of OpenAI, SumTube, and NoteGPT

Bachriah Fatwa Dhini[a*], Kusnindyah Puspito Hapsari[b]

[a*]Multimedia Learning Materials Production Center at Terbuka University, Indonesia, riri@ecampus.ut.ac.id

[b]Multimedia Learning Materials Production Center at Terbuka University, Indonesia, kusnindyah@ecampus.ut.ac.id

To cite this article:
Dhini, B. F., & Hapsari, K. P. (2026). Optimizing AI-Based Video Summarization for Educational Media: A Comparative Evaluation of OpenAI, SumTube, and NoteGPT. *Gema Wiralodra*, 17(1), 60 – 78.

To link to this article:
https://gemawiralodra.unwir.ac.id/index.php/gemawiralodra/issue/view/59

Published by:
Universitas Wiralodra
Jln. Ir. H. Juanda Km 3 Indramayu, West Java, Indonesia

Gema Wiralodra, 17(1), 60 – 78
https://gemawiralodra.unwir.ac.id/index.php/gemawiralodra
p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Orginal Article

Optimizing AI-Based Video Summarization for Educational Media: A Comparative Evaluation of OpenAI, SumTube, and NoteGPT

Bachriah Fatwa Dhini[a*], Kusnindyah Puspito Hapsari[b]
[a*]Multimedia Learning Materials Production Center at Terbuka University, Indonesia, riri@ecampus.ut.ac.id
[b]Multimedia Learning Materials Production Center at Terbuka University, Indonesia, kusnindyah@ecampus.ut.ac.id

***Correspondence: riri@ecampus.ut.ac.id**

ABSTRACT
Video content is increasingly central to distance and mobile learning, yet the volume of instructional media available to learners creates a pressing need for tools that can efficiently distill key content without compromising pedagogical integrity. This study comparatively evaluates three AI-based video summarization tools OpenAI, SumTube, and NoteGPT applied to authentic educational media from UT Radio Mobile at Universitas Terbuka. Eleven videos across two content categories were analyzed: promotional programs (Seputar UT) and module-based instructional content (Tutorial Radio). Using a comparative evaluative design, summary outputs were assessed against four integrated criteria accuracy, suitability, clarity, and processing time by expert validators, with inter-rater consistency reported across content categories. Results indicate that OpenAI achieved the highest accuracy for promotional content (93.3%) through narrative-preserving abstraction, while NoteGPT demonstrated stronger performance on instructional content (85.7%) but exhibited systematic semantic drift including terminological substitution, logical inversion, and topic conflation with meaningful consequences for novice learners. SumTube consistently delivered the fastest processing times (~4 minutes), making it suitable for time-constrained mobile learning contexts, though its extractive architecture rendered outputs susceptible to non-instructional content inclusion. Beyond tool benchmarking, the study demonstrates that summary quality has direct implications for cognitive load, knowledge retention, and learner engagement, and that prompt engineering functions as a form of pedagogical mediation that shapes the instructional alignment of AI-generated outputs. Findings support a hybrid, content-sensitive summarization model and offer theoretically grounded guidance for integrating AI summarization tools into distance and mobile education platforms.

Keywords: AI Video Summarization, OpenAI, NoteGPT, SumTube, Mobile Learning, Distance Education

1. Introduction

Video-based learning has become a dominant instructional modality in higher education as institutions increasingly adopt digital technologies to enhance teaching and learning processes. The integration of video into curricula enables instructors to combine audio, visual, and textual elements, thereby supporting diverse learning styles and improving learner engagement and comprehension (Vivekraj et al., 2019; Kadam et al., 2022). Through recorded lectures, tutorials, and discussion-based programs, students are afforded greater flexibility to access learning materials asynchronously and to revisit complex topics according to their individual pace, which has been shown to support retention and deeper understanding (Kadam et al., 2022). Consequently, video-based learning is widely recognized as a key driver of innovation in distance education and mobile learning environments.

The pedagogical utility of video-based instruction has been thoroughly examined within cognitive and multimedia learning paradigms. A prominent example is Mayer's Cognitive

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Theory of Multimedia Learning, which posits that individuals process information more efficiently when it is delivered via both verbal (narration or text) and visual (images or video) modalities. This dual-channel processing promotes more profound cognitive involvement and facilitates schema development, thereby enhancing knowledge retention and transfer (Vora et al., 2025). Furthermore, Paivio's Dual Coding Theory provides additional support, suggesting that the concurrent activation of visual and verbal systems improves understanding. Consequently, learners exposed to well-designed multimedia content experience cognitive reinforcement, particularly when visual elements are combined with verbal or textual explanations (Hussain et al., 2021).

Conversely, the escalating utilization of video in education has engendered novel pedagogical difficulties, especially when instructional material is presented in protracted formats. Learners often encounter cognitive overload when interacting with lengthy educational videos; the unceasing stream of information can surpass their processing capabilities, thereby hindering effective learning (Kwak et al., 2025). In asynchronous environments, the use of extended videos further intensifies issues of disengagement and procrastination, given that the lack of real-time interaction and feedback may diminish learners' motivation to maintain focus throughout the entire session (Marevac et al., 2025). These difficulties underscore a conflict between the inherent richness of video-based instruction and the practical constraints of learners' time, attention, and cognitive resources.

The issue is exacerbated within educational media that utilize segmented or broadcast-style formats. Programs that integrate instructional material with advertisements or non-instructional segments can impede learners' concentration and the coherence of their understanding. Research indicates that frequent interruptions elevate cognitive load by compelling learners to repeatedly redirect their attention, thus impeding comprehension and retention of essential material (Marevac et al., 2025; Kwak et al., 2025). Likewise, fragmented program structures can disrupt narrative flow, hindering learners' ability to formulate coherent mental representations of the presented content (Peronikolis & Panagiotakis, 2024). These challenges are especially pertinent in mobile learning environments, where learners frequently interact with content during brief, discontinuous sessions.

Considering these difficulties, video summarization has become a crucial approach for enhancing the accessibility of educational video materials. By distilling lengthy videos into more concise forms that highlight key concepts, summarization allows learners to obtain vital information without the need to watch complete recordings, thus facilitating more effective time allocation and concentrated learning (Peronikolis & Panagiotakis, 2024). Summaries can serve as preliminary organizers, study aids, or quick-reference tools, assisting learners in navigating extensive video content more efficiently. Given the expanding volume of educational video, summarization is increasingly recognized as a fundamental element of distance and mobile learning platforms.

Video summarization techniques have evolved significantly and are broadly categorized into extractive, abstractive, and narrative-based approaches. Extractive summarization involves selecting keyframes or audio-visual snippets directly from the source material, preserving their original form. This method is computationally efficient and retains semantic fidelity to the original content (Marevac et al., 2025).

By contrast, abstractive summarization generates new content, usually textual, that reinterprets the meaning of the video. This method demands complex modeling, such as natural language processing (NLP) and visual understanding through deep learning. It offers flexibility in presenting the material and can bridge linguistic or stylistic gaps but may also introduce semantic drift if models lack fine-tuning (Pang et al., 2024).

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

The third approach, narrative-based summarization, seeks to maintain the storytelling flow by selecting segments that not only represent core ideas but also preserve a coherent structure. This method is particularly useful in educational videos where conceptual continuity is vital. Maintaining narrative logic aids learners in building relationships between subtopics and understanding thematic progressions (Kumar et al., 2023). Each approach has pedagogical implications. Extractive methods may be more appropriate for technical or procedural content where exact phrasing or imagery matters, while abstract and narrative methods suit concept-heavy topics that benefit from reformulation and coherence-building (Vora et al., 2025; Hussain et al., 2021).

Advances in artificial intelligence have significantly expanded the potential of video summarization in educational settings. AI-based approaches employ machine learning and natural language processing techniques to analyze video transcripts, identify salient segments, and generate concise representations of content with minimal human intervention (Hussain et al., 2021; Vora et al., 2025). With increasing demands for scalable, real-time summarization, artificial intelligence has transformed how educational videos are processed. Deep learning models, particularly Convolutional Neural Networks (CNNs) for image detection and Recurrent Neural Networks (RNNs) for sequence modeling, have enabled systems to extract temporal and spatial features simultaneously (Seo et al., 2021; Vora et al., 2025). Moreover, AI-driven summarization enables personalization, allowing summaries to be adapted to individual learning preferences, goals, and prior knowledge, which has been shown to enhance learner engagement and perceived usefulness (Hussain et al., 2021).

AI-based summarizers now leverage massive datasets to improve context recognition, speech-to-text accuracy, and scene segmentation. They can detect important visual cues, key speech patterns, and even transitions between instructional sections. These capabilities make automated summarization more adaptive and context-aware, especially with reinforcement learning or transformer-based architectures that allow continual learning (Apostolidis et al., 2021). (Gonzalez et al., 2023) found that students preferred summaries under time constraints, finding them helpful for overview and reinforcing details. The literature provides strong evidence for the instructional effectiveness of AI-generated video summaries, with controlled studies demonstrating quiz score improvements of 17.7% to 24.4% and significant gains in engagement and task efficiency (Stavrinou et al., 2025).

Furthermore, by integrating NLP with computer vision, AI-driven summarizers can generate abstractive summaries that mirror human-like understanding of lecture content. They are not merely replicating visual tokens but synthesizing them into semantically meaningful narratives (Hussain et al., 2021). As such, AI facilitates more refined and learner-sensitive summarization outcomes, reducing manual effort while improving content relevance.

Video summarization, in addition to its efficiency benefits, significantly enhances accessibility and inclusivity within educational contexts. Condensed summaries are particularly beneficial for learners facing time limitations, short attention spans, or challenges in engaging with extensive multimedia materials, thereby facilitating more comprehensive participation in learning activities (Hussain et al., 2021). Furthermore, personalized summaries cater to diverse learning styles by presenting information in formats that align with individual requirements, consequently mitigating obstacles to comprehension and engagement (Peronikolis & Panagiotakis, 2024; Kadam et al., 2022). A growing body of research underscores personalization as a critical element of effective educational video summarization. Conventional, uniform summaries may fail to account for individual variations in learning styles, prior knowledge, or content preferences. Personalized summaries seek to optimize relevance and engagement by customizing extracts to each learner's cognitive profile and

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

performance history (Peronikolis & Panagiotakis, 2024). This personalization can be achieved through learner modeling, wherein systems monitor behaviors such as playback patterns, pausing frequency, and quiz results to infer areas of interest or difficulty. Subsequently, summaries are generated to emphasize content most pertinent to learner needs, thereby enhancing motivation and reducing unnecessary cognitive load (Kadam et al., 2022; Wasim et al., 2022). As higher education increasingly prioritizes equitable access to learning resources, summarization technologies present a promising avenue for supporting inclusive educational practices.

However, despite growing interest in AI-based video summarization, important gaps remain in understanding how different summarization tools perform across varying types of educational content. Comparative evaluations of commercial tools remain exceptionally rare (Olowoniyi et al., 2025; Dongre et al., 2026) and processing time is severely underreported despite its practical importance for deployment in distance education and mobile learning contexts (Yuan et al., 2026). Prior studies have largely focused on technical accuracy or generic benchmarks, with limited attention to real-world educational media that combine instructional material, promotional information, and non-instructional interruptions such as advertisements. There is a need for empirical evaluation that examines not only the accuracy of summaries, but also their suitability to content intent, narrative coherence, and practical usefulness for learners in authentic learning environments.

The reviewed literature confirms that video summarization, when grounded in cognitive learning theory and enhanced through AI and personalization, has strong potential to improve learning efficiency, accessibility, and engagement. However, few studies systematically evaluate how different AI summarization tools perform across distinct types of educational content—for instance, differentiating between promotional/informational videos and structured learning modules. There is also limited work that benchmarks tool performance against human-generated summaries, particularly with respect to accuracy, narrative fidelity, and processing time.

Accordingly, this study evaluates and compares three AI-based video summarization tools OpenAI, SumTube, and NoteGPT within the context of UT Radio Mobile, an educational platform delivering instructional and general informational content. Using authentic educational video datasets, the research examines tool performance across content types and analyzes trade-offs among accuracy, processing time, and level of summary detail. By incorporating human validation and content-specific assessment criteria, this study proposes an applied, learner-centered evaluation framework that offers evidence-based guidance for selecting effective AI-driven video summarization solutions in mobile higher education environments.

2. Method

This study adopts a comparative evaluative design to investigate the effectiveness of three AI-based video summarization tools OpenAI, SumTube, and NoteGPT in the context of UT Radio Mobile. The goal is to determine which tool provides the most accurate, relevant, and efficient summaries for two types of educational video content: general informational programs (Seputar UT) and instructional tutorial content (Tutorial Radio). The research aligns comparative frameworks frequently used in AI summarization studies, which benchmark multiple tools using standardized datasets and mixed evaluation metrics (Sun et al., 2021; Kumar et al., 2023). Experimental logic underpins the approach, where the impact of each tool on summary quality is evaluated based on predefined criteria and human validation.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Orginal Article

Based on Figure 1, the research design consists of four phases: (1) literature review and framework development, (2) data selection and categorization, (3) tool implementation and summary generation, and (4) evaluation using quantitative and qualitative measures. Phase 1, a comprehensive desk study was conducted to identify pedagogical, computational, and evaluation frameworks for AI-based video summarization. Emphasis was placed on learning theories, summarization strategies (extractive vs. abstractive), personalization techniques, and benchmark metrics. Phases 2, researchers identified relevant UT Radio programs in both content categories. The primary dataset includes 11 educational videos sourced from UT Radio's official repository and YouTube platform. These videos are grouped into two categories (1) 5 Seputar UT videos: Featuring promotional, institutional, or general university information and (2) 6 Tutorial Radio videos: Featuring structured academic tutorials on subjects such as physics, mathematics, statistics, archival studies, and economics. Content selection was guided by expert interviews with UT Radio data managers to ensure thematic relevance and content diversity. Videos were chosen based on their popularity, educational value, and representativeness of UT Radio's broadcast formats.

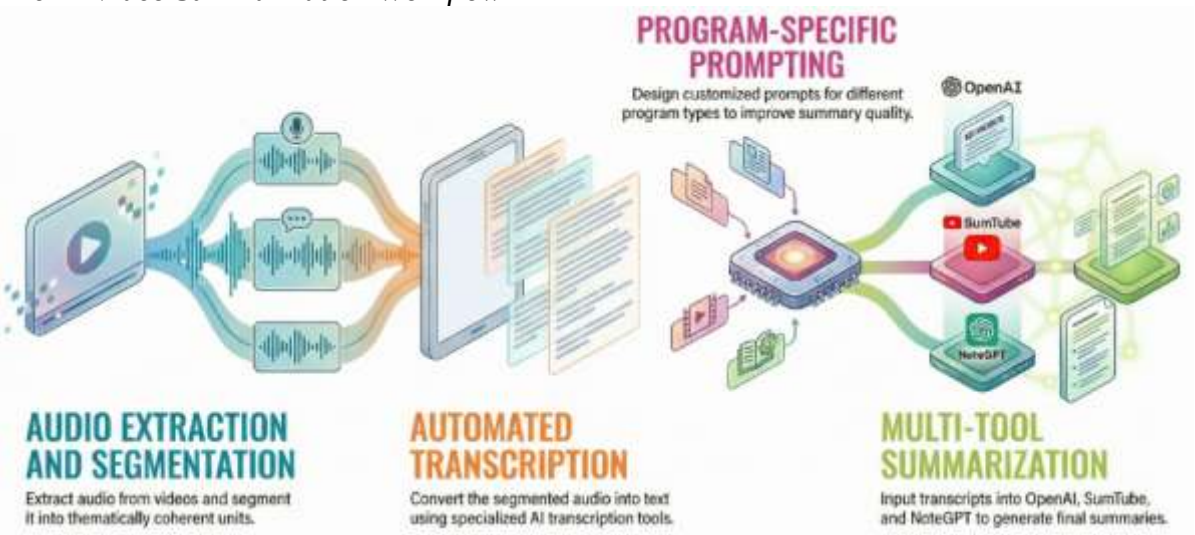
Figure 1

The 4-Phase Research Design Workflow

Figure 2 describes a structured and consistent video summarization workflow was implemented to ensure methodological rigor and comparability across tools. The process began with audio extraction, in which downloaded video files were converted into audio format to facilitate downstream processing. Subsequently, the audio files underwent segmentation, dividing the content into thematically coherent units to improve transcription accuracy and contextual relevance. Each segment was then processed through AI-based transcription tools to generate textual representations of the spoken content. Following transcription, a prompt engineering phase was conducted, where program-type-specific prompts were carefully designed to enhance summary quality and relevance. Finally, the generated transcripts were submitted to three AI summarization tools OpenAI, SumTube, and NoteGPT to produce concise summaries.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Figure 2
The AI Video Summarization Workflow



Based on Table 1, to optimize output quality, prompt characteristics were tailored according to program type. For Seputar UT, prompts emphasized the extraction of five key talking points while explicitly excluding advertisements, greetings, and introductory remarks to maintain informational focus. In contrast, for Tutorial Radio, prompts were designed to prioritize the identification and synthesis of core instructional content aligned with defined learning objectives. This differentiated prompt strategy ensured that summaries were contextually appropriate and pedagogically meaningful across varying program formats.

Table 1
Prompt Design

Program Type	Prompt Characteristics
Seputar UT	Emphasizes five key talking points; excludes advertisements, greetings, and introductions
Tutorial Radio	Focuses on extracting core instructional content aligned with learning objectives

Final Phase Evaluation Criteria and Scoring Procedure. Each AI-generated summary was decomposed into discrete summary points individual statements or claims produced by the tool. Expert validators, all of whom are practitioners in educational media production at Universitas Terbuka, assessed each summary point against the original video transcript using three integrated evaluative criteria. These criteria served as qualitative anchors guiding the validator's binary judgment on each point, as described below.

Accuracy refers to factual correctness: a summary point is judged correct only if its content can be verified directly against the source transcript without distortion, addition of information not present in the original, or omission of critical qualifiers.

Suitability refers to topical alignment with the instructional purpose of the video. For Seputar UT programs, a point is suitable if it reflects the informational intent of the content while excluding non-instructional elements such as greetings, advertisements, or broadcast closings. For Tutorial Radio programs, a point is suitable if it is aligned with the defined learning objectives of the academic subject being discussed. Clarity refers to linguistic and semantic coherence: a point is clear if it is expressed without ambiguity, avoids terminology inconsistent with the source, and can be understood without reference to the original video.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

A summary point was scored as correct (1) only if it satisfied all three criteria simultaneously. A point failing any one criterion was scored as incorrect (0). This integrated binary approach was adopted to reflect the inseparability of these dimensions in practical educational use. A summary point that is factually accurate but topically irrelevant, or relevant but semantically ambiguous, provides no meaningful learning value. The composite score for each video-tool pair was then computed using the following formula:

$$score = \frac{\sum correct\ points}{\sum total\ summary\ points}\ x\ 100 \qquad (1)$$

To illustrate with a concrete example from the data: for Video 1 (Seputar UT Registration Requirements), NoteGPT generated 7 summary points. The validator judged only 1 point as satisfying all three criteria, yielding a score of 1/7 × 100 = 14.2%. SumTube generated 8 summary points for the same video, of which 7 were accepted, yielding 7/8 × 100 = 87.5%. OpenAI generated 5 points, all accepted, yielding 100%. This example illustrates both how the formula operates and how substantially tool performance can vary on the same content.

Scores therefore range from 0% to 100%, where higher values indicate greater proportion of summary points meeting the combined standard of accuracy, suitability, and clarity. To aid interpretation, the following bands were applied (Table 2). Processing time was recorded separately as the total elapsed time in minutes and seconds from transcript submission to complete summary generation, measured using a digital stopwatch for each video-tool combination.

To establish the credibility and consistency of the human evaluation process, an inter-rater consistency analysis was conducted based on the evaluative decisions made by expert validators across all video-tool combinations. Each designated validator all of whom hold expertise in educational media production independently assessed all three tools' summary outputs for their assigned video against the same source transcript, applying the four evaluation criteria operationally defined above.

Table 2
Scoring Band Interpretation for Summary Quality Assessment

Score Range	Interpretation
85–100%	High fidelity — summary closely mirrors key content
70–84%	Moderate fidelity — minor omissions or imprecisions
55–69%	Low fidelity — notable gaps or partial accuracy
Below 55%	Inadequate — summary substantially misrepresents content

Given that each video was assigned to one primary validator who evaluated all three tool outputs for that video, inter-rater consistency was computed as pairwise percentage agreement across tool-quality decisions. For each video, an evaluative decision was classified as either acceptable (score ≥ 75%) or unacceptable (score < 75%) for each tool. The proportion of concordant pairwise decisions among the three tool evaluations representing the degree to which quality thresholds were applied consistently across tools within each video was then averaged across all videos in each content category. This approach provides a transparent and conservative estimate of evaluative consistency given the study's single-validator-per-video design.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**
Orginal Article

The 75% threshold for acceptable summary quality was established during a pre-evaluation calibration session, in which all validators reviewed two sample summaries jointly and aligned their understanding of each scoring criterion before conducting independent assessments. In cases of borderline scores, validators were required to document their rationale with reference to the source transcript.

3. Results

Prior to presenting the comparative tool performance data, the consistency of the human evaluation procedure is reported. Table 4 dan 5 presents pairwise percentage agreement across tool-quality decisions for each video, disaggregated by content category. Agreement was computed as the proportion of concordant accept/reject decisions (threshold ≥ 75%) across all three pairwise tool comparisons per video, then averaged across videos within each category.

The results reveal a comparative evaluation of OpenAI, SumTube, and NoteGPT summarization tools applied to six Seputar UT and five Tutorial Radio program videos. Seputar UT programs are promotional and informational in nature, while Tutorial Radio programs consist of module-based learning videos representing subjects from each faculty. The quantitative overview of tool performance across both content categories is presented in Table 3.

Based on Table 3, the comparative results reveal distinct performance patterns across content types. For Seputar UT programs, OpenAI demonstrated the highest average accuracy (93.3%), indicating strong narrative fidelity and topic alignment, although it required the longest processing time (approximately 19 minutes). SumTube achieved moderate accuracy (83.8%) but was significantly faster, completing summaries in approximately 4 minutes, making it the most time-efficient tool. NoteGPT produced comparatively lower accuracy (75.2%) with a moderate processing time of approximately 10 minutes, suggesting variability in handling promotional or informational content.

Table 3
Quantitative Results: Average Accuracy and Processing Time by Tool and Content Category

Program Type	Prompt Characteristics	Avg. Accuracy	Avg. Processing Time
Seputar UT	OpenAI	93.3%	~19 minutes
	SumTube	83.8%	~4 minutes
	NoteGPT	75.2%	~10 minutes
Tutorial Radio	OpenAI	80.0%	~15 minutes
	SumTube	78.5%	~4 minutes
	NoteGPT	85.7%	~11 minutes

In contrast, for Tutorial Radio programs, which involve more complex instructional material, NoteGPT achieved the highest average accuracy (85.7%), indicating stronger retention of instructional details. OpenAI followed with 80.0% accuracy and an average processing time of approximately 15 minutes, maintaining solid conceptual clarity but requiring longer computation. SumTube again emerged as the fastest tool (approximately 4 minutes) with an accuracy of 78.5%, showing consistent speed performance but slightly lower precision for concept-heavy academic content. Overall, the data indicate that tool effectiveness varies depending on content type, with trade-offs between accuracy and processing efficiency.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

The inter-rater consistency analysis is presented in Tables 4 and 5. For Seputar UT programs, Table 4 reports the pairwise agreement across tool-quality decisions per video, with key findings showing that OpenAI achieved the highest accuracy (93.3%) across all videos, SumTube demonstrated the fastest average processing time (~4 minutes), and NoteGPT showed greater variability including off-topic or redundant information in several outputs. These results reflect OpenAI's strength in narrative fidelity and relevance for broad, promotional content, consistent with the literature on context-sensitive summarization (Kadam et al., 2022). Expert validators assessed summary outputs based on topic alignment, factual accuracy, clarity, and processing efficiency.

Table 4
Inter-Rater Consistency: Pairwise Percentage Agreement by Content Category and Video: Seputar UT Programs

No.	Category	Topic	NoteGPT	SumTube	OpenAI	Agreement
1	Registration	Upload Complete Requirements for Prospective Master's & Doctoral Students	Reject (14.2%)	Accept (87.5%)	Accept (100%)	33.3%
2	Study Program	Strategies to Make Learning Chemistry Interesting	Accept (100%)	Accept (87.5%)	Accept (80%)	100.0%
3	Professor	One Hour with UT Professor #4: Learner–Content Interaction	Accept (80%)	Accept (87.5%)	Accept (100%)	100%
4	Outstanding Students	Inspirational Student Talk	Accept (85.7%)	Reject (60%)	Accept (100%)	33.3%
5	Learning Services	Benefits of Assignment Workshops & Exam Clinics	Accept (100%)	Accept (90%)	Accept (100%)	100%
6	Promotional Outreach	UT's Efforts to Provide Inclusive Teaching Materials	Reject (71.4%)	Accept (90%)	Accept (80%)	33.3%
Mean % Agreement						**66.6%**

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Based on Table 4, Tutorial Radio programs present deeper educational complexity and demand greater semantic understanding from summarization tools. Evaluation focused on summary precision, avoidance of irrelevant content, and preservation of instructional detail. Key findings from Table 5 show that NoteGPT performed best in instructional detail retention, scoring 85.7% on average; OpenAI excelled on conceptual clarity in content-heavy topics; and SumTube remained the fastest tool but occasionally included advertisements or off-topic content. These variations support literature findings that instructional videos require sophisticated semantic parsing (Kumar et al., 2023).

Table 5
Inter-Rater Consistency: Pairwise Percentage Agreement by Content Category and Video: Tutorial Radio Programs

No.	Faculty	Topic	NoteGPT	SumTube	OpenAI	Agreement
1	Physics Education	Energy and Momentum	Accept (85.7%)	Accept (88%)	Accept (100%)	100%
2	Statistics	Basic Concepts of Data Collection & Presentation	Reject (71.4%)	Accept (75%)	Reject (60%)	33.3%
3	Archival Studies	History of Archives	Accept (85.7%)	Reject (60%)	Reject (60%)	33.3%
4	Mathematics Education	Groups and Subgroups	Accept (100%)	Accept (81.8%)	Accept (80%)	100.0%
5	Development Economics	Employment, Poverty & Decentralization	Accept (85.7%)	Accept (87.5%)	Accept (100%)	100.0%
	Mean % Agreement					73.3%

Note: Accept = score ≥ 75%; Reject = score < 75%. Agreement computed as proportion of concordant pairwise decisions across three tool combinations per video.

The mean pairwise agreement of 66.6% for Seputar UT and 73.3% for Tutorial Radio indicate moderate to acceptable evaluative consistency across the three tool assessments. The lower agreement observed in Seputar UT is attributable to three videos (V1, V4, V6) in which one tool produced a markedly divergent result most notably Video 1, where NoteGPT scored only 14.2% against SumTube and OpenAI scores of 87.5% and 100% respectively. This divergence does not reflect inconsistency in the validators' application of criteria but rather reflects genuine performance discontinuities between tools on specific content types, which is itself a substantive finding of the study. For Tutorial Radio, the lower agreement in Videos 2 and 3 similarly reflects real performance gaps across tools rather than evaluator inconsistency, a pattern consistent with the greater semantic complexity of instructional content discussed in Section 4.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

These findings confirm that the scoring procedure was applied in a principled and threshold-consistent manner across validators, supporting confidence in the validity of the comparative results reported below. Future replications of this study are recommended to adopt a formal double-rating protocol in which at least two validators independently score each video to enable the computation of Cohen's Kappa as a more statistically precise reliability index.

A cross-comparison of the three AI-based summarization tools OpenAI, SumTube, and NoteGPT reveals distinct performance profiles shaped by their underlying processing approaches, as summarized in Table 6. OpenAI consistently demonstrates superior accuracy and narrative fidelity, particularly in content requiring coherent storytelling and contextual sensitivity. Its summaries are typically concise, often structured into approximately five key points, reflecting a high level of abstraction and semantic filtering. However, this strength in deep parsing and conceptual synthesis comes with longer processing times, indicating a trade-off between precision and efficiency.

In contrast, SumTube stands out for its speed and broad content coverage. It generates summaries significantly faster than the other tools, making it well-suited for time-sensitive applications or mobile learning contexts where immediacy is essential. While its accuracy remains moderate and generally reliable for promotional or informational material, it occasionally captures unrelated segments such as advertisements or transitional content, suggesting a more extractive and surface-level summarization strategy. Meanwhile, NoteGPT excels in producing detailed and elaborative paraphrased summaries, making it particularly useful for drill-down review of instructional materials where learners may benefit from expanded explanations. Nevertheless, its tendency toward semantic drift and mixed-topic inclusion indicates challenges in maintaining strict topical boundaries and conceptual precision. Overall, the comparative patterns confirm that each tool offers unique advantages, and their effectiveness depends on the instructional context, content complexity, and intended learning outcomes.

Table 6
Comparison Performance:Narrative Style vs. Speed

	Accuracy	Processing Time
OpenAI	★★★★★	★★☆☆☆
SumTube	★★★★☆	★★★★★
NoteGPT	★★★☆☆	★★★☆☆

Table 6 highlights a clear comparative performance gradient among the three tools in terms of accuracy and processing time. OpenAI achieves the highest accuracy rating (★★★★★), confirming its strength in producing precise and coherent summaries, yet it scores lower in processing speed (★★☆☆☆), reflecting longer generation times. SumTube, by contrast, demonstrates the fastest processing performance (★★★★★) while maintaining relatively strong—but slightly lower—accuracy (★★★★☆), indicating a favorable balance for speed-oriented applications. Meanwhile, NoteGPT occupies a middle position, with moderate accuracy (★★★☆☆) and average processing time (★★★☆☆), suggesting a balanced but less specialized performance profile. Overall, the comparison underscores a fundamental trade-off between accuracy and speed, with each tool offering distinct advantages depending on user priorities and instructional needs. These outcomes affirm that no single tool is universally optimal; rather, suitability depends on content type and usage goals.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**
Orginal Article

4. Discussion

Why Content Type Determines Tool Effectiveness: A Cognitive Load Interpretation

The performance differential observed between tools across content categories is not incidental it reflects a theoretically predictable interaction between the structural complexity of the source material and the summarization architecture of each tool. Promotional content, such as Seputar UT, is characterized by relatively linear narrative structures, explicit topic transitions, and institutionally formulaic language. These features reduce the semantic disambiguation burden on the summarization model, which explains why OpenAI a model optimized for abstractive, context-sensitive synthesis, achieved its highest accuracy (93.3%) precisely in this category. The structural regularity of promotional content aligns well with OpenAI's capacity for narrative abstraction, enabling it to identify and preserve key informational units without being destabilized by conceptual ambiguity.

By contrast, Tutorial Radio content involves domain-specific terminology, multi-step instructional sequencing, and implicit logical dependencies between concepts features that impose substantially higher demands on semantic parsing. The fact that NoteGPT outperformed OpenAI in this category (85.7% vs. 80.0%) suggests that its more elaborative, paraphrase-intensive approach may inadvertently preserve instructional granularity that a more concise abstractive model tends to compress. This is consistent with Mayer's Cognitive Theory of Multimedia Learning, which holds that instructional effectiveness depends not merely on information reduction but on the preservation of the relational structure between concepts (Mayer, 2009). A summary that compresses too aggressively risks severing the conceptual scaffolding that learners rely upon to construct coherent mental models. This concern is particularly acute in AI-assisted learning environments, where the alignment between AI output and cognitive architecture determines whether the technology genuinely supports or inadvertently undermines learning (AlShaikh et al., 2024).

The Accuracy-Speed Trade-off as a Design Tension, not a Technical Limitation

The inverse relationship between accuracy and processing speed with OpenAI being the most accurate but slowest, and SumTube the fastest but least precise should not be interpreted purely as a technical constraint. Rather, it reflects a fundamental design philosophy embedded in each tool's architecture. SumTube's extractive approach prioritizes computational efficiency by selecting surface-level textual segments from transcripts without deep semantic modeling. This makes it highly responsive but structurally shallow, which explains its tendency to include non-instructional content such as advertisements and transitional segments. This behavior is not a malfunction it is the predictable output of a system that has not been designed to distinguish instructional intent from incidental content. Recent work on multimodal video summarization confirms that distinguishing informative from non-informative segments requires fused audio-visual-textual feature modeling that extractive-only systems inherently lack (Marevac et al., 2025).

This distinction has important implications for deployment decisions. In mobile learning contexts where learners access content in fragmented, time-constrained sessions, the value of a summary is partly temporal a four-minute turnaround serves a fundamentally different learning behavior than a nineteen-minute wait. However, if the resulting summary lacks instructional coherence, the efficiency gain is pedagogically hollow. This tension suggests that the appropriate performance benchmark for educational summarization tools should not be accuracy or speed in isolation, but rather a context-weighted composite that reflects the relative priority of each dimension for a given use case. Future tool development should consider

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

making this trade-off explicit and user-configurable, rather than embedding it invisibly in system defaults.

Semantic Drift in NoteGPT: Mechanisms and Pedagogical Consequences

The semantic drift observed in NoteGPT's outputs warrants deeper examination beyond the surface-level observation that it "paraphrases excessively." Semantic drift in abstractive summarization arises when a model prioritizes linguistic fluency over semantic fidelity that is, when the generation process optimizes for coherent-sounding output rather than strict correspondence to the source. In educational contexts, this is particularly consequential because learners who have not yet mastered the subject matter lack the prior knowledge needed to detect inaccuracies or fill conceptual gaps introduced by drift. Unlike expert readers who can cross-reference a distorted summary against existing schema, novice learners may internalize the distorted version as accurate, leading to misconceptions that are subsequently difficult to correct (Darabi & Ghinea, 2014). This challenge is compounded by evidence that LLM-based systems are substantially less effective at identifying incorrect or drifted reasoning than they are at producing fluent text, making automated quality control unreliable without human validation (Choudhury et al., 2025).

Three distinct drift mechanisms were identified across NoteGPT's outputs in this study. The first is terminological substitution, in which domain-specific vocabulary is replaced with colloquially adjacent but disciplinarily imprecise language. In the Statistics tutorial on data collection and presentation, NoteGPT replaced the technical terms validitas (validity) and reliabilitas (reliability) which carry precise psychometric meanings central to the learning objective with the general phrase "the right and trustworthy tools." For a learner encountering these concepts for the first time, this substitution constitutes a form of terminological erasure: the paraphrase preserves surface meaning while stripping away the disciplinary scaffolding that makes the content educationally functional. Critically, the omitted terms are ones the learner will be expected to recall and apply in subsequent coursework and assessments their absence from the summary creates a gap that passive review cannot compensate for.

The second mechanism is logical inversion, in which conditional or causal relationships are paraphrased without preserving their directionality. In the Physics tutorial on energy and momentum, the source transcript stated explicitly that an object's momentum remains constant unless an external force acts upon the system establishing conservation of momentum as the default condition. NoteGPT's summary rendered this as the claim that "external forces affect momentum, and under certain conditions momentum can be maintained," implying that conservation is a special case rather than the baseline principle. A learner relying on this summary would internalize a structurally incorrect understanding of a fundamental physics law one that is likely to generate systematic errors in subsequent problem-solving tasks and will conflict with future instruction in ways that are effortful to unlearn.

The third mechanism is topic conflation, in which sequentially or conceptually distinct content segments are merged into a single undifferentiated summary point, erasing organizational distinctions that are pedagogically significant. This was observed in the Archival Studies tutorial on the history of archives, where the source video sequentially addressed two historically and epistemologically distinct developments: the emergence of archival practice in ancient civilizations, followed by the formalization of archival science as a recognized academic discipline in the nineteenth century. NoteGPT collapsed these into a single statement that archives have existed for a long time and have developed into an internationally recognized field eliminating the conceptual distinction that formed the tutorial's central learning objective.

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Learners who internalize this conflated version lose the historical framework needed to organize subsequent knowledge about the discipline's development.

Across these three mechanisms, a shared underlying pattern is apparent: NoteGPT's generation process consistently privileges sentence-level fluency over cross-sentence semantic structure. This behavior is characteristic of abstractive models that have not been fine-tuned on domain-specific educational corpora, where disciplinary vocabulary, logical structure, and conceptual sequencing carry instructional weight that general-purpose language models are not trained to preserve. Addressing this limitation would require either fine-tuning on curated educational datasets or implementing post-generation factual consistency checks against the source transcript before summaries are surfaced to learners (Choudhury et al., 2025).

Pedagogical Implications: How Summarization Quality Shapes Learning Outcomes, Engagement, and Cognitive Processing

The performance differences observed across tools carry direct and distinct consequences for three dimensions of learner experience central to educational effectiveness: cognitive load management, knowledge retention, and sustained engagement.

With respect to cognitive load, Sweller's Cognitive Load Theory distinguishes between intrinsic load (the inherent complexity of the material), extraneous load (unnecessary processing demands imposed by poor presentation), and germane load (effortful processing that contributes to schema formation) (Sweller et al., 1998). AI-generated summaries intervene primarily at the level of extraneous load by reducing the volume of content a learner must process, they free cognitive resources for deeper engagement with the material's conceptual structure. However, this benefit is conditional on summary quality. A high-accuracy summary from OpenAI reduces extraneous load without distorting the intrinsic structure of the content, thereby supporting germane processing. By contrast, a semantically drifted summary from NoteGPT may paradoxically increase extraneous load: the learner must expend additional cognitive effort reconciling unfamiliar paraphrases with prior knowledge, or more problematically may process the distorted version as accurate, constructing flawed schema that requires effortful correction later. Recent empirical work confirms that poorly aligned AI outputs can trigger split attention effects and shallow processing that negate the cognitive benefits of summarization altogether (Meng et al., 2025). This suggests that the cognitive benefit of AI-driven summarization is not guaranteed by compression alone; it is contingent on the semantic fidelity of the output.

With respect to knowledge retention, research in educational psychology consistently demonstrates that retention is enhanced when learners encounter content that is structurally coherent, terminologically consistent, and aligned with the organizational logic of the original material (Kintsch, 1988). OpenAI's strength in narrative fidelity, particularly for promotional content, suggests that its summaries are more likely to support durable retention because they preserve the logical sequencing and key informational anchors of the source. SumTube's occasional inclusion of non-instructional content, such as advertisements and transitional remarks, introduces irrelevant material that competes for memory encoding and dilutes the salience of genuinely instructional content. For learners, revisiting summaries as study aids a common use case in distance education this dilution effect is compounded across multiple review sessions, progressively weakening the signal-to-noise ratio of the learning resource and reducing its utility as a revision tool.

With respect to learner engagement, engagement in asynchronous learning environments is fragile and heavily dependent on perceived relevance and cognitive manageability (Fredricks et al., 2004). A summary that is too long, too vague, or topically inconsistent risks triggering

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

disengagement before the learner reaches the core instructional content. SumTube's processing speed advantage is pedagogically meaningful in mobile learning contexts characterized by brief, interrupted sessions: a four-minute summary turnaround is more likely to fit within a learner's available attention window than a nineteen-minute wait, supporting initial access and reducing the friction associated with asynchronous content consumption. However, engagement sustained by speed is qualitatively different from engagement sustained by content relevance. The former supports entry; the latter supports deeper processing, consolidation, and return visits. An ideal summarization tool for mobile education would optimize both dimensions providing rapid delivery without sacrificing the topical coherence that motivates continued engagement. Taking together, these three dimensions confirm that tool selection in educational settings must be guided not only by technical performance metrics but by a theory-informed understanding of how summary characteristics interact with learner cognition, motivation, and memory.

Prompt Engineering as a Pedagogical Intervention

One underappreciated finding of this study is the role of prompt design in modulating summary quality. The differentiated prompt strategy employed excluding advertisements and greetings for Seputar UT and emphasizing instructional alignment for Tutorial Radio effectively functioned as a form of pedagogical mediation between the raw capabilities of the model and the specific learning requirements of the content. This suggests that prompt engineering is not merely a technical optimization strategy but a pedagogical act that encodes assumptions about what constitutes educationally valuable content. This view is supported by recent systematic reviews demonstrating that prompt engineering functions as an instructional design skill one that requires educators to translate pedagogical intent into structured AI inputs to produce outputs that are both contextually appropriate and learning-aligned (Qian, 2025; Cain, 2024).

The implication is that effective integration of AI summarization into educational platforms cannot be achieved through tool deployment alone. It requires the deliberate involvement of instructional designers or content specialists who can translate pedagogical intent into prompt specifications. Institutions that deploy summarization tools without this expertise risk generating technically competent but pedagogically misaligned output summaries that are fluent and coherent, but that emphasize the wrong content for the intended learning outcome.

Toward a Hybrid Summarization Model for Mobile Learning

Taken together, the findings suggest that no single summarization paradigm is sufficient for the full range of educational video content encountered in mobile learning environments. The evidence points toward the need for a hybrid model in which tool selection or tool weighting within a single pipeline is dynamically determined by content characteristics. Promotional and informational content benefits from narrative-preserving abstractive models such as OpenAI; complex instructional content may benefit from the elaborative detail of NoteGPT, provided that semantic drift is mitigated through post-processing; time-sensitive or preview-oriented use cases are well-served by the speed of SumTube, with the caveat that outputs require filtering for non-instructional content. The viability of such content-aware hybrid architectures is supported by recent technical work demonstrating that fusing audio, visual, and textual features within a single summarization pipeline substantially improves segment classification accuracy over single-modality approaches (Marevac et al., 2025).

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Such a hybrid system would need to be supported by an automatic content classification layer that identifies the structural type of the input video before routing it to the appropriate summarization model. The prompt engineering framework developed in this study provides a starting point for this classification logic. Future research should investigate whether content-type classifiers trained on educational video metadata can reliably automate this routing decision, reducing the burden on instructional designers while maintaining pedagogical alignment.

5. Conclusion

This study comparatively evaluated three AI-based video summarization tools OpenAI, SumTube, and NoteGPT using authentic educational media from UT Radio Mobile at Universitas Terbuka. Applying an integrated scoring framework grounded in accuracy, suitability, clarity, and processing time across two content categories promotional (Seputar UT) and instructional (Tutorial Radio) the study produced findings that extend beyond tool benchmarking to illuminate the pedagogical conditions under which AI summarization either supports or undermines educational outcomes.

OpenAI demonstrated the highest accuracy overall, particularly for promotional content (93.3%), owing to its capacity for narrative-preserving abstraction. SumTube consistently delivered the fastest processing times (approximately four minutes), making it well-suited for time-constrained mobile learning contexts, though its extractive architecture rendered it susceptible to including non-instructional segments. NoteGPT achieved the strongest performance on instructional content (85.7%), yet its outputs exhibited systematic semantic drift manifested as terminological substitution, logical inversion, and topic conflation that poses meaningful risks for novice learners who lack the prior knowledge needed to detect and correct distorted information.

Three broader implications emerge from these findings. First, content type is a primary determinant of tool effectiveness: the structural characteristics of the source material interact with each tool's summarization architecture in ways that are theoretically predictable and practically consequential. Matching tools to content type rather than applying a single tool universally is essential for preserving the pedagogical integrity of AI-generated summaries. Second, the accuracy-speed trade-off identified in this study is not a technical limitation, but a design tension embedded in each tool's architecture, one that should be made explicit and user-configurable in educational deployment contexts. Third, prompt engineering functions as a pedagogical intervention: the differentiated prompt strategies employed in this study demonstrated that the quality and instructional alignment of AI-generated summaries depend not only on the selected tool but on the specificity, and pedagogical intentionality of the input provided. These finding positions prompt design as a core competency for instructional designers working with AI summarization tools.

From a cognitive perspective, the findings carry direct implications for learner experience. High-fidelity summaries reduce extraneous cognitive load and support knowledge retention by preserving the structural and terminological coherence of the source. Low-fidelity or semantically drifted summaries, by contrast, may increase extraneous load, introduce misconceptions, and undermine engagement particularly in the asynchronous, mobile-first learning contexts that characterize distance education at scale.

These findings collectively support a hybrid summarization model in which tool selection is dynamically guided by content type, instructional purpose, and temporal constraints. Future implementations should explore automated content classification layers capable of routing video content to the most appropriate summarization approach, reducing reliance on manual

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

prompt engineering while maintaining pedagogical alignment. Future research should also address the limitations of this study, including the single-validator-per-video design, the absence of fine-tuned domain-specific models as comparison baselines, and the restriction to Bahasa Indonesia content from a single institutional context. Extending this work to multilingual corpora, larger video datasets, and learner-facing outcome measures including comprehension tests and retention assessments will be essential for establishing the generalizability of these findings and for advancing the development of truly learner-adaptive summarization systems.

Acknowledgments
We would like to extend our deepest gratitude to the Research and Community Service Center, known as LPPM Universitas Terbuka, for the allocation of research funds. We also express our deepest thanks to the operational team and listeners of UT Radio.

6. References

Apostolidis, E., Adamantidou, E., Metsai, A. I., Mezaris, V., & Patras, I. (2021). AC-SUM-GAN: Connecting Actor-Critic and Generative Adversarial Networks for Unsupervised Video Summarization. Ieee Transactions on Circuits and Systems for Video Technology, 31(8), 3278–3292. https://doi.org/10.1109/tcsvt.2020.3037883

AlShaikh, R., & Al-Malki, N. (2024). The implementation of the cognitive theory of multimedia learning in the design and evaluation of an AI educational video assistant utilizing large language models. Heliyon, 10(3), e25361. https://doi.org/10.1016/j.heliyon.2024.e25361

Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. TechTrends, 68, 47–57. https://doi.org/10.1007/s11528-023-00896-0

Choudhury, S., Sharma, M. N., Sharma, R., Gandal, G. R., & Garg, A. (2025). AI-based educational video summarization. ShodhKosh: Journal of Visual and Performing Arts, 6(2s), 272–280.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Darabi, K., & Ghinea, G. (2014). Personalized video summarization by highest quality frames. In *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* (pp. 1-6). IEEE.

Dongre, S., Jain, K., Kale, I., Kumbhare, V., Lohote, S., & Lonare, S. (2026). EduVision: An AI-Driven Framework for Automated Video Summarization and Intelligent Note Generation. In A. K. Tripathi, A. K. Saha, & V. Shrivastava (Eds.), *Proceedings of World Conference on Artificial Intelligence: Advances and Applications* (Vol. 1766, pp. 392–402). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-13806-4_32

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95(2), 163–182.

Gonzalez, H., Li, J., Jin, H., Ren, J., Zhang, H., Akinyele, A., Wang, A., Miltsakaki, E., Baker, R., & Callison-Burch, C. (2023). Automatically Generated Summaries of Video Lectures May Enhance Students' Learning Experience. *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, 382–393. https://doi.org/10.18653/v1/2023.bea-1.31

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S. W., & Victor Hugo C. de Albuquerque. (2021). A Comprehensive Survey of Multi-View Video Summarization. Pattern Recognition, 109, 107567. https://doi.org/10.1016/j.patcog.2020.107567

Hussain, T., Muhammad, K., Ullah, A., Cao, Z., Baik, S. W., & Victor Hugo C. de Albuquerque. (2020). Cloud-Assisted Multiview Video Summarization Using CNN and Bidirectional LSTM. Ieee Transactions on Industrial Informatics, 16(1), 77–86. https://doi.org/10.1109/tii.2019.2929228

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Kadam, P., Vora, D., Mishra, S., Patil, S., Kotecha, K., Abraham, A., & Gabralla, L. A. (2022). Recent Challenges and Opportunities in Video Summarization With Machine Learning Algorithms. Ieee Access, 10, 122762–122785. https://doi.org/10.1109/access.2022.3223379

Kumar, A., Pandey, S. K., Swarup, C., Singh, K. U., Singh, T., Ojha, M. K., & Mishra, P. (2023). Statistical Evaluation of Video Summarization Models From an Empirical Perspective. Traitement Du Signal, 40(2), 555–565. https://doi.org/10.18280/ts.400214

Kwak, S.-S., Lee, J.-D., & Park, S. K. (2025). The Effective Highlight-Detection Model for Video Clips Using Spatial—Perceptual. Electronics, 14(18), 3640. https://doi.org/10.3390/electronics14183640

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Marevac, E., Kadušić, E., Živić, N., Buzađija, N., Tabak, E., & Velić, S. (2025). Multimodal Video Summarization Using Machine Learning: A Comprehensive Benchmark of Feature Selection and Classifier Performance. Algorithms, 18(9), 572. https://doi.org/10.3390/a18090572

Meng, N., Mat Deli, M., & Abdul Rauf, U. A. (2025). Educational technology and AI: Bridging cognitive load and learner engagement for effective learning. SAGE Open, 15(2). https://doi.org/10.1177/21582440251395930

Olowoniyi, R., Willie, A., & Klenk, M. (2025). *Comparative Architectures and Pedagogical Efficacy of Generative Artificial Intelligence Tools for the Automated Summarization of Online Course Lectures*.

Pang, Z., Nakashima, Y., Otani, M., & Nagahara, H. (2024). Unleashing the Power of Contrastive Learning for Zero-Shot Video Summarization. Journal of Imaging, 10(9), 229. https://doi.org/10.3390/jimaging10090229

Peronikolis, M., & Panagiotakis, C. (2024). Personalized Video Summarization: A Comprehensive Survey of Methods and Datasets. Applied Sciences, 14(11), 4400. https://doi.org/10.3390/app14114400

Qian, Y. (2025). Prompt engineering in education: A systematic review of approaches and educational applications. Journal of Educational Computing Research. https://doi.org/10.1177/07356331251365189

Seo, M., Suh, Y., & Jeong, K. (2021). Deep Learning-Based Video Summarization. International Journal on Advanced Science Engineering and Information Technology, 11(6), 2488. https://doi.org/10.18517/ijaseit.11.6.12888

Stavrinou, L., Constantinides, A., Belk, M., Vassiliou, V., Liarokapis, F., & Constantinides, M. (2025). The Reel Deal: Designing and Evaluating LLM-Generated Short-Form Educational Videos. *Proceedings of the 3rd International Conference of the ACM Greek SIGCHI Chapter*, 188–196. https://doi.org/10.1145/3749012.3749048

p – ISSN: **1693 – 7945**
e – ISSN: **2622 – 1969**

Sun, G., Wu, H., Zhu, L., Xu, C., Liang, H., Xu, B., & Liang, R. (2021). VSumVis: Interactive Visual Understanding and Diagnosis of Video Summarization Model. Acm Transactions on Intelligent Systems and Technology, 12(4), 1–28. https://doi.org/10.1145/3458928

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296.

Vivekraj, V. K., Sen, D., & Raman, B. (2019). Video Skimming. Acm Computing Surveys, 52(5), 1–38. https://doi.org/10.1145/3347712

Vora, D., Kadam, P., Mohite, D. D., Kumar, Nilesh, Kumar, Nimit, Radhakrishnan, P., & Bhagwat, S. S. (2025). AI-driven Video Summarization for Optimizing Content Retrieval and Management Through Deep Learning Techniques. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-87824-9

Wasim, M., Ahmed, I., Ahmad, J., Nawaz, M., Alabdulkreem, E., & Ghadi, Y. Y. (2022). A Video Summarization Framework Based on Activity Attention Modeling Using Deep Features for Smart Campus Surveillance System. Peerj Computer Science, 8, e911. https://doi.org/10.7717/peerj-cs.911

Yuan, L., Xu, J., & Zhan, Z. (2025). Effects of Pedagogical Agent-Generated Summaries on Video-Based Learning: Evidence from Eye-Tracking and EEG. *Education Sciences*, *16*(1), 39. https://doi.org/10.3390/educsci16010039