Information Modelling and Knowledge Bases XXXVII
M. Menšík et al. (Eds.)

doi:10.3233/FAIA251723

Co-Learning: Cognitive Load-Based Multilingual Learning Content Generation Model

Virach SORNLERTLAMVANICH[1],
Thatsanee CHAROENPORN[2], Pannathorn SATHIRASATTAYANON[3], Parin JATESIKTAT[4], Anatta SUESUWAN[5], and Parkhan NGAMWANNAKORN[6]
[1,2]*Asia AI Institute (AAII), Faculty of Data Science, Musashino University, Japan.*
[1]*Faculty of Informatics, Burapha University, Thailand.*
[3,4]*Sirindhorn International Institute of Technology, Thammasat University, Thailand.*
[5]*Jamsai Publishing Co., Ltd., Thailand.*
[6]*VIZDATA Co., Ltd., Thailand.*
ORCID ID: Virach SORNLERTLAMVANICH https://orcid.org/0000-0002-6918-8713
ORCID ID: Thatsanee CHAROENPORN https://orcid.org/0000-0002-9577-9082
6422782316@g.siit.tu.ac.th, 6422771707@g.siit.tu.ac.th, anatta.suesuwan@gmail.com, parkhan@vizdata.work

Abstract. With the growth of internet usage, countless educational videos are now available online. However, it can be a significant challenge for learners to identify the videos they need, especially in their preferred language and within their available time. Additionally, not all videos are suitable for subject-specific learning due to variations in length and presentation components. According to Sweller's Cognitive Load Theory, working memory during the learning process is highly limited. Learners must be selective about which information from sensory memory they choose to focus on. In our proposed Co-Learning model (a model of connective learning where all necessary knowledge is refined and interconnected to support effective learning within cognitive limitations), we leverage NLP approaches to enhance the learning experience. These approaches include video speech refinement, subtitle generation, dubbed video translation, summarization, classification,

[1] Corresponding Author: Virach Sornlertlamvanich, Asia AI Institute (AAII), Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo, 135-8181, Japan; E-mail: virach@musashino-u.ac.jp
Faculty of Informatics, Burapha University, 169 Long Had Bangsaen Road, Saensuk, Muang, Chonburi 20131, Thailand.

[2] Thatsanee Charoenporn, Asia AI Institute (AAII), Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo, 135-8181, Japan; E-mail: thatsane@musashino-u.ac.jp

[3] Pannathorn Sathirasattayanon, Sirindhorn International Institute of Technology, Thammasat University, 99 Moo 18, Paholyothin Road, Klong Nueng, Klong Luang, Pathumthani 12120, Thailand; E-mail: 6422782316@g.siit.tu.ac.th

[4] Parin Jatesiktat, Sirindhorn International Institute of Technology, Thammasat University, 99 Moo 18, Paholyothin Road, Klong Nueng, Klong Luang, Pathumthani 12120, Thailand; E-mail: 6422771707@g.siit.tu.ac.th

[5] Anatta Suesuwan, Jamsai Publishing Co., Ltd., 31/101 Phutthamonthon sai 2 Road, Salathammasop, Taweewattana, Bangkok 10170, Thailand; E-mail: anatta.suesuwan@gmail.com

[6] Parkhan Ngamwannakorn, VIZDATA Co., Ltd., 27/134 The Swiss Garden View Village, Nongrahang Road, Sam Wa Tawan Tok, Khlong Sam Wa, Bangkok 10510, Thailand; E-mail: parkhan@vizdata.work

keyword extraction for word cloud indexing, and quiz generation, thereby creating a multilingual, learner-efficient environment. In our preliminary survey, the generated content was well-received and effectively utilized for class adjustments with an acceptance rate of 93%.

Keywords. Connective Learning, Multi-lingual, Learning Video, Cognitive Load Theory

1. Introduction

In recent years, the proliferation of learning videos available online has significantly transformed the learning landscape. Platforms like YouTube, Coursera, and Khan Academy host an immense amount of learning content spanning diverse disciplines, from basic skills to advanced technical knowledge. While this abundance provides learners with unprecedented access to information, it also presents several challenges. For learners, identifying the most suitable videos that align with their specific needs, preferred time constraints, and language preferences can be overwhelming. Not all videos are optimized for subject-specific learning; issues such as excessive length, requirement of external source of knowledge, lack of structure, or inappropriate sequencing of learning components can hinder effective knowledge acquisition. Consequently, learners often experience cognitive overload, which detracts from their ability to process and retain information.

From the lecturers' perspective, producing high-quality educational content requires considerable effort and preparation. To ensure professional-quality videos, educators often need to allocate additional time beyond their regular teaching schedules. Recording learning content in a controlled studio environment necessitates meticulous planning, including drafting pre-designed scripts and practicing delivery to minimize issues such as repetition and redundancy, word corrections, hesitation and stuttering, errors in information, or other interruptions. This additional workload can deter educators from creating supplementary resources, thereby limiting the availability of high-quality learning materials.

These challenges are further compounded by the limitations of human cognitive capacity, as highlighted by Sweller's Cognitive Load Theory [1, 2, 3]. This theory posits that working memory—the part of memory responsible for holding and manipulating information during learning—is inherently limited. Effective learning requires careful management of cognitive load, ensuring that learners focus on relevant information without being overwhelmed by extraneous details. In a video-based learning environment, unstructured or poorly designed content can exacerbate cognitive load, reducing learners' ability to process and retain information effectively.

To address these issues, we propose a novel Co-Learning (Connective Learning) model designed to meet the needs of both learners and lecturers. The Co-Learning model is a model of connective learning, where all necessary knowledge is refined and interconnected to support effective learning within cognitive limitations. We leverage Natural Language Processing (NLP) approaches to enhance the learning experience.

Specifically, it provides the following capabilities:

1. *Video Speech Refinement*: The system includes modules to smoothen recorded speech in videos, eliminating common issues such as pauses, repetitions, and stuttering, resulting in polished and professional-quality content.

2. *Multilingual Subtitle Support*: By converting source videos into multilingual transcriptions and adding subtitles, the system ensures that learners from diverse linguistic backgrounds can access and understand the material better.
3. *Content Summarization*: The system generates concise summaries of videos, enabling learners to review key points before delving into the full content. This feature helps reduce cognitive load by providing a clear overview of the material.
4. *Keyword Extraction for Word Cloud Indexing*: The system extracts key terms and phrases from video transcriptions to generate visually appealing word clouds. These word clouds provide learners with a quick visual summary of the content, highlighting the most important themes and concepts.
5. *Extraneous Knowledge Link*: The system integrates relevant external knowledge to supplement the learning content, providing learners with additional context and background information. This feature bridges gaps in understanding and enriches the learning experience by connecting core material with necessary ancillary knowledge.
6. *Quiz Generation*: To reinforce learning and assess comprehension, the system automatically generates quizzes based on the video content. These quizzes serve as an effective tool for learners to validate their understanding and identify areas that require further study.

The Co-Learning model embodies the principles of connective learning, fostering a collaborative and adaptive learning environment by connecting the constructive AI and NLP components to deal with the available learning contents. By applying NLP techniques, the system refines video subtitles for archiving, translating, summarizing, classifying, and labeling relevant keywords. These functionalities create a multilingual and learner-friendly ecosystem that caters to diverse educational needs. Furthermore, the system reduces the burden on lecturers by automating labor-intensive tasks, allowing them to focus on delivering quality education without the added stress of extensive content production.

In summary, the proposed Co-Learning model bridges the gap between the challenges faced by learners and lecturers in the current digital learning landscape. By addressing issues of cognitive load, accessibility, and content quality, the system empowers learners to engage with educational videos more effectively while enabling lecturers to produce high-quality materials efficiently.

In summary, our paper provides the following contributions:

- Model architecture: we propose the architecture of Co-Learning model that bridges the gap between the challenges faced by learners and lecturers in the current digital learning landscape.
- Learning content refinement: we connect the SOTA NLP techniques to refine and generate the content that leverage the learning efficiency according to the Cognitive Load Theory and human learning process.

The rest of this paper is organized as follows: Section 2 explores related works, examining Cognitive Load Theory and its implications for the limitations of human learning processes. Section 3 highlights the growth of educational videos available online. Section 4 presents the design of the Co-Learning model architecture, which leverages NLP techniques to streamline recorded videos and generate highly effective and engaging learning content. Section 5 details the implementation of the Co-Learning System for its evaluation in a classroom setting. Section 6 provides a subjective

evaluation and the results of improvements based on user feedback. Finally, Section 7 concludes the paper and discusses potential future extensions.

2. Related Works

Cognitive Load Theory, as proposed by Sweller et al. (2011), plays a crucial role in understanding how learners interact with educational content [3]. It focuses on the limitations of working memory and the strategies we can employ to optimize learning by managing cognitive load. Memory is composed of several distinct components, including sensory memory, working memory, and long-term memory. Each has a distinct role in the learning process.

2.1. The Role of Human Memory in Learning

Cognitive Load Theory is founded on the concept that learning occurs through the transfer of information from sensory memory to working memory, and ultimately to long-term memory [3]. Sensory memory is transient, capturing a vast amount of environmental stimuli, but most of it quickly fades unless it is selected for further processing. The selected information moves to working memory, which has a limited capacity [4]. This makes it crucial for learners to focus their attention selectively on relevant information; otherwise, working memory might become overloaded, reducing learning efficiency.

Working memory processes the selected information and, if managed correctly, can encode it into long-term memory [5], where it can be stored virtually indefinitely. Long-term memory is structured in schemas, which are mental representations that help make sense of new information by linking it to existing knowledge. This process of encoding new information into long-term memory is essential for learning, as it forms the basis of retention and retrieval [6].

2.2. Cognitive Load and Human Learning

Cognitive load, a central concept in Cognitive Load Theory, refers to the mental effort required to process information. There are three types of cognitive load:

1. *Intrinsic Cognitive Load*: This load is imposed by the inherent complexity of the learning task itself. For example, learning a complex mathematical concept requires more cognitive effort than learning a simple one. While intrinsic load cannot be fully controlled, it can be reduced through strategies like "chunking," which involves breaking down complex information into smaller, manageable units [3].
2. *Extraneous Cognitive Load*: This refers to the cognitive effort induced by factors unrelated to the task itself, such as distractions or poor presentation of the material. The goal is to minimize extraneous load by designing learning materials in ways that help learners focus on what is most relevant [7].
3. *Germane Cognitive Load*: This load is associated with the cognitive effort required to process and organize new information into long-term memory. It involves activities such as comparing new material with prior knowledge and forming connections to strengthen memory schemas. The more learners know

about a topic, the lower the germane load becomes, as they can more easily automate processes related to that topic [5].

For efficient learning, it is essential to ensure that the working memory capacity exceeds the combined demands of intrinsic, extraneous, and germane cognitive load [6]. In other words, to facilitate learning, we must reduce the unnecessary cognitive load (extraneous load) and manage the inherent complexity (intrinsic load) in a way that promotes germane load.

This is where educational content design becomes critical. According to Cognitive Load Theory, videos, as learning resources, should be designed with attention to video length, structure, and content. For example, research by Guo et al. (2014) found that video length significantly impacts learner engagement, with videos under six minutes resulting in near-total engagement [8]. Thus, keeping videos short and manageable can help optimize working memory capacity, leading to more efficient learning.

2.3. NLP Techniques for Effective Content Design

To further align with Cognitive Load Theory and enhance the learning experience, we propose the use of Natural Language Processing (NLP) techniques to process educational content from videos. These techniques are connected to create learning materials that are optimized for cognitive load and human memory processing:

A. Managing Intrinsic Cognitive load (Essential: managing the core of the learning content)
   1. *Video Length Shortening*: By ensuring that videos are no longer than six minutes, we align with findings by Guo et al. (2014) that shorter videos lead to higher engagement and better retention.
   2. *Providing Video Synopses*: By offering a clear, structured summary of each video, we can enhance its clarity and reduce cognitive overload, helping learners focus on the most important information [6].
   3. *Content Summarization*: Summarizing video content provides learners with a quick overview, which supports efficient memory encoding by highlighting the most important points [6].

B. Reducing Extraneous Cognitive Load (Environment: reducing the absence of circumstancing knowledge)
   4. *Indexing for Searchability*: Keyword search and scene indexing make it easier for learners to locate specific information within a video, reducing the need to process irrelevant content and minimizing extraneous cognitive load [9].
   5. *Related Knowledge Source Link*: Connecting related knowledge sources facilitates a more holistic understanding of the material, promoting deeper learning by integrating new knowledge into existing schemas [5].

C. Increasing Germane Cognitive Load (Relevant: making it easy to find the relevant knowledge)
   6. *Video Categorization*: Proper categorization improves the accessibility of educational content, helping learners to easily find and select relevant materials [10].

7. *Keyword Extraction*: Identifying key concepts and terms within the videos allows for a more effective representation of the content and can help learners quickly grasp the core ideas [11].

By leveraging Cognitive Load Theory, we can design more effective learning experiences that minimize cognitive overload and maximize the potential for encoding information into long-term memory. Incorporating NLP techniques into learning content design—such as shortening video lengths, categorizing content, and providing summaries—supports the human memory process, leading to more efficient learning outcomes. Through these strategies, learners can engage with material in ways that align with the limitations and strengths of their cognitive architecture, enhancing both retention and understanding.

3. Availability of Educational Videos

The growth of educational videos is leaping forward, both in the number of videos produced and presented on the internet and the number of user engagements. The statistics from a market survey in 2018 by Marketing Charts illustrate that video viewership on Facebook and YouTube increased by up to 51.6% for businesses on Facebook and 23.4% on YouTube[7]. In the educational field, there was also a significant increase of 10.2% on Facebook and 11.2% on YouTube, marking the third-highest growth in comparison to other sectors.

This surge in demand for educational videos has spurred a corresponding increase in the supply of video content tailored for learning. The need for online learning has grown exponentially, particularly due to the shift in education practices driven by the COVID-19 pandemic. This global crisis accelerated the adoption of online lecture formats and remote learning platforms, as traditional in-person education became largely inaccessible. As a result, educators and institutions turned to digital platforms to bridge the gap, marking a significant shift in the landscape of educational video usage.

There are two explicit types of online learning videos available on the internet. The first type involves direct learning systems integrated into formal educational frameworks, such as E-learning platforms developed by educational institutions. Examples include TUXSA, the online Master's degree program offered by Thammasat University in collaboration with SkillLane (https://www.skilllane.com/academic/tuxsa?), and Chula MOOC provided by Chulalongkorn University (https://mooc.chula.ac.th/courses). These platforms are designed to deliver structured, curriculum-based content.

The second type encompasses additional learning opportunities through open online courses and publicly available resources. Renowned platforms such as Khan Academy (https://www.khanacademy.org), Udemy (https://www.udemy.com), Fast AI (https://course.fast.ai), and Coursera (https://www.coursera.org) provide learners with access to a vast array of video content across diverse subjects. Additionally, videos commonly distributed on YouTube offer informal yet valuable learning opportunities. Platforms like Open Culture (http://www.openculture.com) categorize videos by academic field and accessibility, enabling learners to easily locate relevant resources.

[7] https://www.marketingcharts.com/charts/video-view-growth-facebook-youtube-q2-2018-genre

These services demonstrate the immense potential of video-based learning in catering to a variety of educational needs.

The COVID-19 pandemic also underscored the importance of leveraging virtual meeting tools, such as ZOOM, Microsoft Teams, and Google Meet, for educational purposes. These tools not only facilitate real-time interaction between educators and learners but also serve as a platform for recording and disseminating lectures. The integration of such environments into the learning process allows lecturers to create high-quality content with minimal disruption. They can record live lectures or pre-record sessions, which can then be refined and repurposed for broader audiences. This practice has led to a significant increase in the volume and quality of educational videos available online.

According to reports, prior to the pandemic in December 2019, ZOOM reported approximately 10 million daily meeting participants. By December 2020, this number surged to 350 million, reflecting the platform's rapid integration into educational and professional settings. At the height of the pandemic, over 90,000 schools worldwide were utilizing ZOOM to facilitate online learning, underscoring its critical role in maintaining educational continuity. In Australia, ZOOM usage during the pandemic peaked at 70 times higher than in 2019, exceeding 1.2 billion meeting minutes in a single month during the height of lockdowns.

Moreover, the use of ZOOM-like environments has proven effective in fostering interactive learning experiences. Lecturers can utilize features such as screen sharing, breakout rooms, and polls to enhance engagement and participation during live sessions. These tools also simplify the content production process by enabling educators to record sessions directly and upload them to learning management systems or public platforms.

The potential for continued growth in the use of online video content is immense. With the increasing acceptance of hybrid and fully online educational models, video-based learning is poised to become a cornerstone of modern education. Institutions and educators are likely to continue investing in tools and platforms that enable the creation, refinement, and distribution of high-quality learning materials, ensuring accessibility and inclusivity for a global audience. This ongoing evolution underscores the transformative power of educational videos in bridging knowledge gaps and democratizing education for learners worldwide.

4. Co-Learning Model Architecture

The Co-Learning model architecture is designed based on the principles of Cognitive Load Theory, aiming to maximize the effectiveness of learning experiences within cognitive limitations. Following the guidelines for NLP techniques in effective content design, as discussed in Section 3.3, the model utilizes appropriate NLP tools to process input lecture materials. These materials include textual components such as syllabi, presentations, and lecture notes, as well as multimedia components like lecture videos.

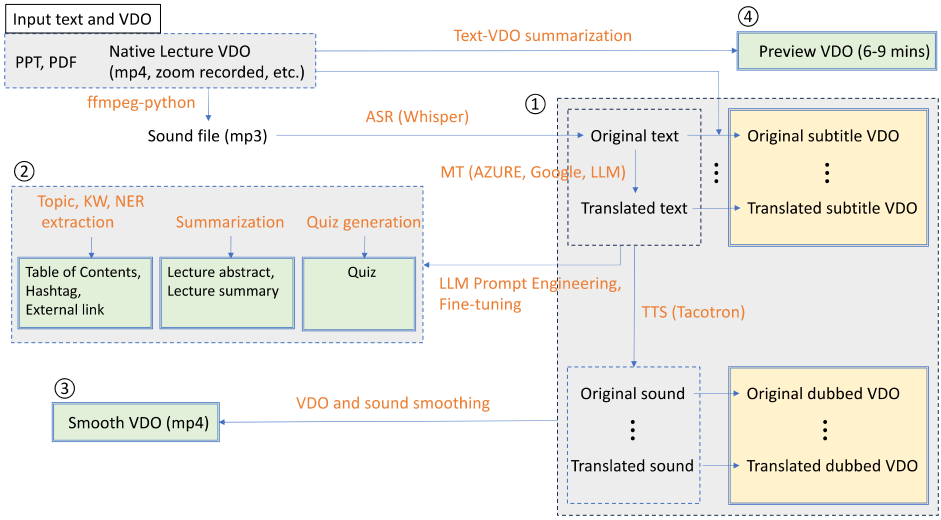

Figure 1. Co-Learning model architecture

Figure 1 illustrates the model architecture. The model primarily utilizes a lecture video recorded via platforms like ZOOM, supplemented with related materials (e.g., syllabus, presentation file, or lecture note). Initially, the model generates videos with subtitles and translated dubbed versions in Module ①) by employing ffmpeg-python[8] to extract the audio component and output it in MP3 format.

The ffmpeg-python is a Python wrapper for FFmpeg, a powerful open-source multimedia framework used for processing video and audio files. The library allows users to access FFmpeg's features programmatically, enabling efficient manipulation of media files without the need for direct command-line input. With ffmpeg-python, users can perform operations such as format conversion, video/audio extraction, resizing, editing, and even adding effects through an intuitive Pythonic interface.

Next, Whisper[9] is used to generate subtitle text for the lecture, which is output in SRT file format. Whisper is an advanced Automatic Speech Recognition (ASR) system developed by OpenAI. It is designed to transcribe spoken language from audio files into text with high accuracy. Whisper leverages a large-scale transformer-based architecture trained on a diverse dataset of multilingual and multitask audio data. This allows it to handle a wide range of languages, accents, and audio conditions effectively. Additionally, Whisper is capable of performing tasks such as translation, speaker identification, and audio classification, making it versatile for various applications.

At this stage, a subtitle editing tool is provided to correct recognition errors or fine-tune the lecture content as needed. Subsequently, the subtitle file is translated to produce a dubbed video in the target language by employing Tacotron. As a result of Module ① , a translated dubbed video with subtitle is generated. Tacotron is a sequence-to-sequence model developed for text-to-speech (TTS) synthesis, designed to generate highly intelligible and natural-sounding speech directly from text. Tacotron replaced traditional TTS pipelines, which often relied on a combination of multiple modules (e.g., text

[8] https://github.com/kkroening/ffmpeg-python
[9] https://github.com/openai/whisper

analysis, phoneme-to-spectrogram conversion, and vocoding). Instead, Tacotron unified this process into a single neural network architecture capable of directly synthesizing mel-spectrograms from input text [12]. The model has been widely adopted and further refined for various applications, including multilingual TTS, voice cloning, and personalized speech synthesis.

Next, text processing techniques using a large language model (LLM) are employed to automatically generate a table of contents, summary, and quizzes, providing learners with clear and accessible information in Module ②.

To refine the output from the LLM, prompt engineering plays a crucial role in producing the desired text. Prompt engineering is the process of designing effective inputs to optimize the responses generated by the LLM [13, 14]. The key strategies include:

A. *Include Specific Details*: Clearly outline the context, requirements, and desired information to obtain more relevant and accurate answers.
B. *Adopt a Persona*: Ask the model to assume a specific role or perspective (e.g., "Act as a teacher" or "Respond as a financial advisor").
C. *Use Delimiters*: Clearly separate distinct parts of the input (e.g., use quotation marks, triple backticks, or other markers to indicate sections like context, questions, or examples).
D. *Specify Steps*: Break down tasks into clear, sequential steps to guide the model in completing complex processes.
E. *Provide Examples*: Include example inputs and desired outputs to help the model understand the format and style you are expecting.
F. *Control Output Length*: Specify whether the response should be brief, detailed, or within a particular word or sentence limit.
G. *Reference Texts*: Instruct the model to use a provided reference text when generating answers to ensure accuracy and consistency.

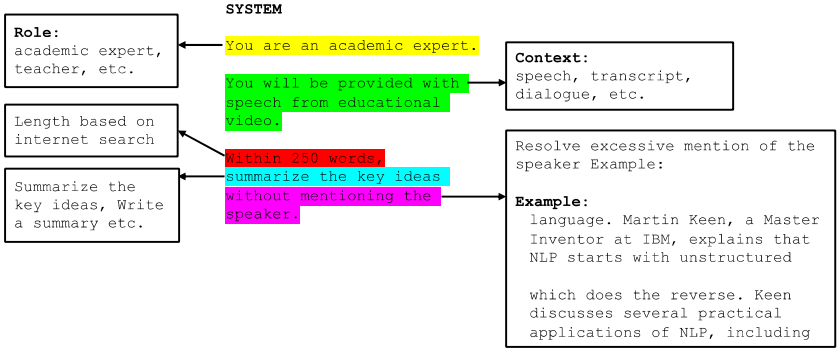

Figure 2. Prompt for video summary generation

Figure 2 illustrates an example of a prompt designed to generate a summary within a length of 250 words. The content is tailored to highlight the key ideas from the transcribed text. To ensure the text reads naturally and focuses on the key ideas, the prompt is crafted to exclude mentions of speaker names, which are commonly referenced during lectures. The version of summary can be ranged from 250 to 500 words according to the visibility essence of the contents.

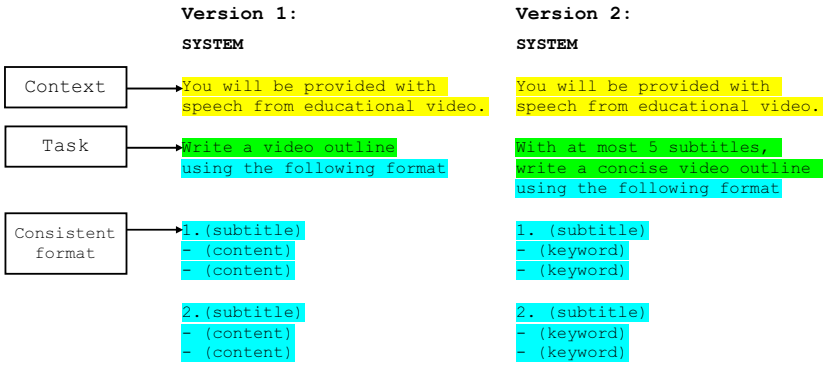

Figure 3. Prompt for video outline generation

Figure 3 illustrates an example of a prompt designed to generate a video outline, providing structural information about the lecture. The prompt in Version 2, shown on the right-hand side of the figure, produces more concise details compared to Version 1, shown on the left-hand side. In this case, the prompt constrains the generated outline to include up to 5 subtitles with concise details. The appropriate version can be selected to minimize information overload for learners. The output of this process is a table of contents for the learning video.

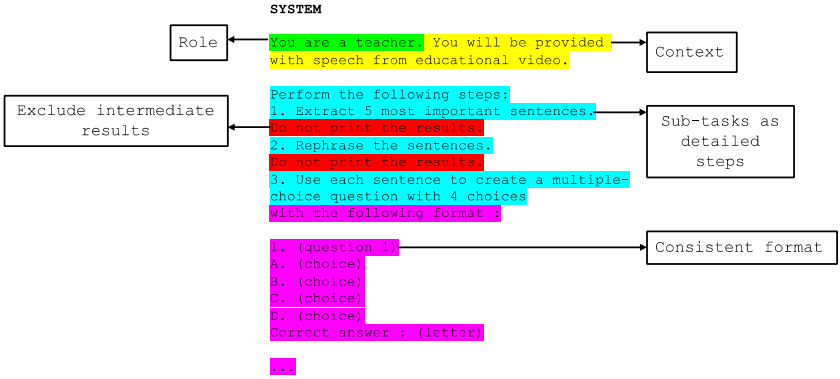

Figure 4. Prompt for quiz generation

Figure 4 illustrates an example of a prompt designed to generate a short quiz from the essence of the lecture. Therefore, a teacher role is assigned to generate a quiz based on the provided video content. A quiz is useful for reminding the learner about the content. It is well accepted by the learners when the quiz spontaneously provides the result of the answer. It is an important step to confirm the learner's understanding.

To address issues such as surrounding noises, stuttering, redundancy or conversational speech that may hinder on-demand usability, text processing adjusts the content to produce smoother video as process in Module ③.

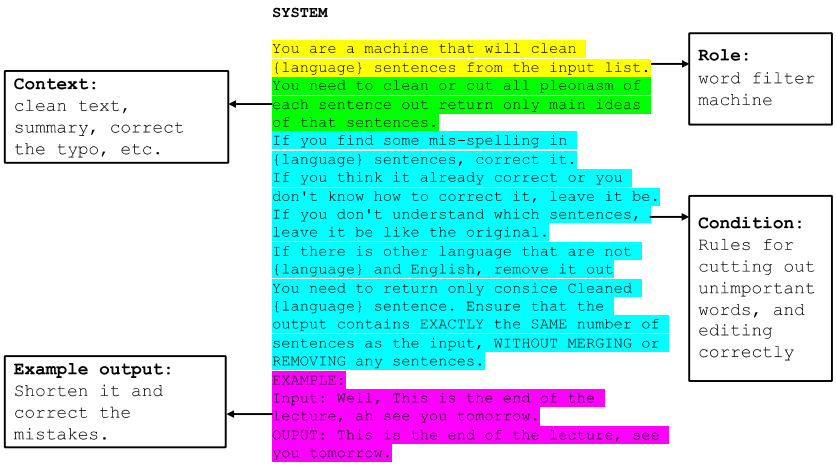

Figure 5. Prompt for video cleanup and smoothing

To clean up and smooth speech in a video, a carefully designed prompt is provided to a LLM to generate concise and polished text from transcriptions. The process involves identifying and correcting errors, such as typos or misused words, while removing filler words or non-essential phrases that may detract from clarity. The prompt instructs the LLM to preserve the original meaning of each sentence and avoid merging or removing any sentences from the input, ensuring that the structure of the transcription remains intact. It also requires the model to handle multilingual content by excluding irrelevant languages and focusing solely on the target language. For instance, extraneous interjections like “well”, “um” or “ah” are cleaned up, and grammatical inaccuracies are corrected to produce refined output. The example output demonstrates how a verbose and error-prone sentence is transformed into a concise and grammatically correct version, thereby improving readability and maintaining the speech's coherence and flow. By following these detailed rules, the LLM effectively generates high-quality, concise transcripts that can be used for subtitles or other educational purposes. As a result, the prompt in Figure 5 shows an example of the output which the terms “well” and “ah” are cancelled from the speech.

To reduce the cost of using external available LLM, prompt compression is a choice that can reduce the size of the prompt. Figure 6 exhibits a result of prompt compression. The output is kept the same while the condition part of the prompt is significantly shorten but still keep the original purpose of intention. The model of LLMLingua-2[10] is used for this purpose.

[10] https://github.com/microsoft/LLMLingua?tab=readme-ov-file

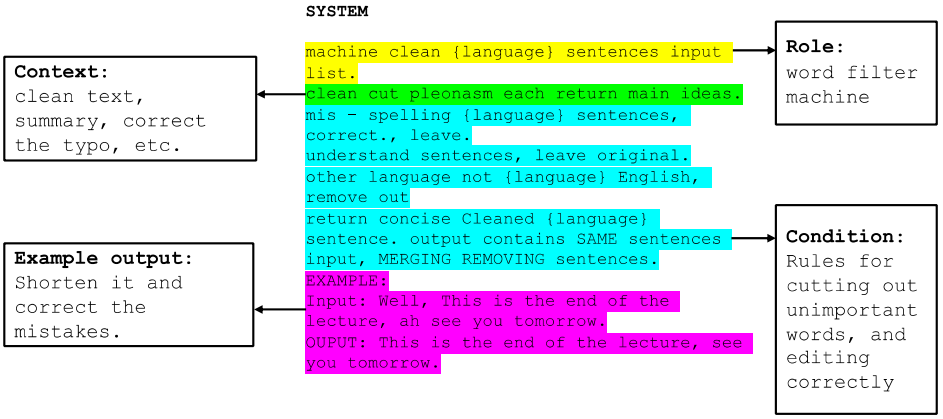

Figure 6. Result of prompt compression

Finally, to aid learners in deciding whether to watch a video, a summary video of about 6-9 minutes is provided as produced in Module ④, giving an overview of the content. This improves learning efficiency by allowing learners to grasp the overall structure before diving into the full material.

5. Co-Learning System Implementation

The system is experimentally implemented in a public cloud system. The target videos are collected with the corresponding subtitle files and archived in the cloud database. The subtitle text files are translated into any target languages (such as Japanese, Chinese, and Thai) by Google Translate API. The resulting translation file is manipulated as a source file for each language processing. From the source file, keyword extraction, summarization and video synchronization are conducted in parallel with a relating unique ID to realize the video multilingual services.

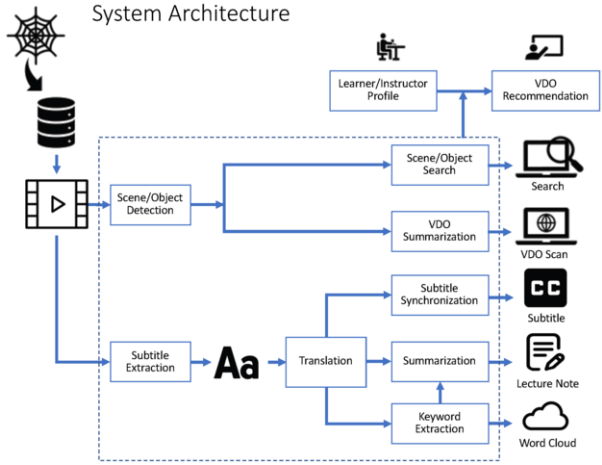

Figure 7. Co-Learning system implementation

Figure 7 shows the system architecture of the proposed connective learning [15]. The flow is started from video file collection. Video subtitles are extracted and processed in accordingly to their video contents. Text processing techniques are applied to extract the keywords, summarized and indexed.

At the same time, the video files are analyzed to detect the objects and scene representations. The preliminary experiment on video analysis is conducted to support video summarization and scene search. Finally, video recommendation based on learner view history and profile can be considered, and the instructor curriculum fulfillment function can be extended.

The system efficiently provides video playback, summary, word cloud annotated with a hyper link, scene search under the multilingual service environment. As a result, a learner can browse the summary and word cloud to understand the structure of the content before starting the video playback. A hyper link to external webpages supports the additional explanation. Scene search can direct the learner to the desired scene. The available learning videos are finally connected to realize the efficient learning environment.

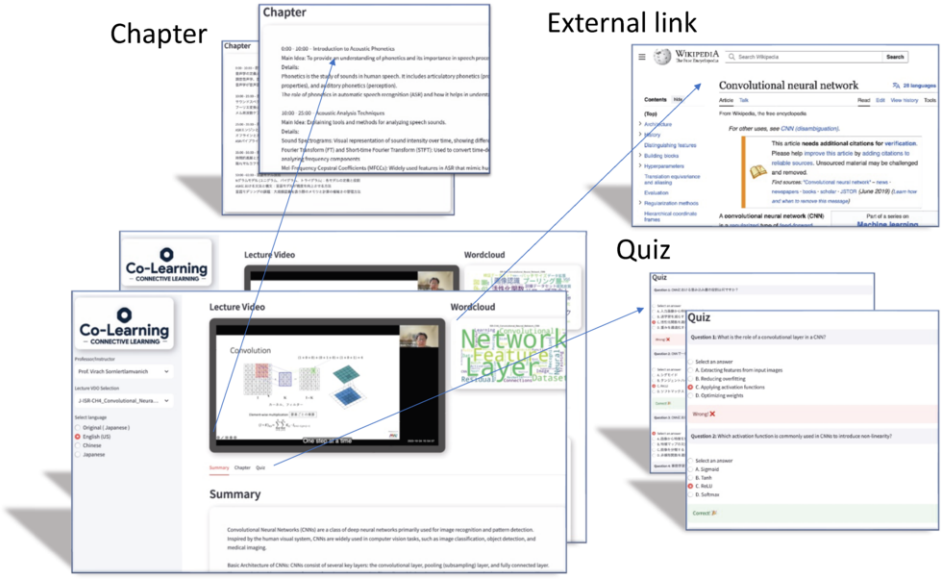

Figure 8. Co-Learning system for multi-lingual service

Figure 8 illustrates a learning-effective integrated service of the proposed Co-Learning system for a multilingual learning environment. The platform is composed of interconnected components designed to enhance the learner's experience. At its core is the lecture video interface, which delivers educational content supported by subtitles and concise summaries. The well-structured intrinsic content expression guides learners to engage with the video in a logical and comprehensible manner, facilitating effective understanding.

Surrounding the lecture video, complementary resources, such as word clouds, emphasize key terms, enabling learners to quickly grasp critical concepts while reducing extraneous cognitive load. External references, such as Wikipedia and additional study

materials, are integrated to provide broader context and deeper understanding. These references are dynamically linked to the lecture content, allowing learners to seamlessly explore related topics.

The content is organized chapter-wise, with time-indexed video frames to ensure accessible navigation. Additionally, instant quizzes are provided to assess learners' understanding, reinforcing comprehension and increasing germane cognitive load.

This strategic design manages intrinsic cognitive load by presenting structured content, minimizes extraneous cognitive load by reducing interruptions and irrelevant information, and increases germane cognitive load to foster meaningful learning. The approach aligns with Cognitive Load Theory as outlined by Sweller et al. (2011).

Furthermore, multilingual content is readily switchable upon request, ensuring accessibility for diverse audiences. By integrating these elements, the Co-Learning platform offers a holistic, effective, and inclusive learning experience, tailored to meet varied linguistic and educational needs.

6. Subjective Evaluation and Improvement based on Suggestions

The evaluation was conducted with students enrolled in an Image and Speech Recognition System class. The system provides video lectures in English, Japanese, and Chinese, with a duration of approximately 100 minutes each. Japanese serves as the original language. The evaluation period spanned four weeks, following weekly classes. A total of 89 students were registered for the class, with responses collected from 79 students for Video title 1, 77 for Video title 2, 82 for Video title 3, and 68 for Video title 4. Table 1 summarizes the number of students enrolled, the number of responses received, and the number of positive and negative feedback instances for each video, along with representative comments for each category.

The system was very well received, achieving a high positive-to-negative acceptance ratio of 71:8 (90%) due to the introduction of an innovative concept in cognitive load-aware design for learning content. However, the acceptance rate dropped during the second and third rounds of evaluation after feedback was incorporated, with ratios of 47:30 (61%) and 30:52 (37%), respectively. This decline was primarily attributed to dissatisfaction with subtitle display and video quality, rather than the interactive functions provided. Finally, the user acceptance rate increased significantly to 63:5 (93%) after addressing the main issues that were causing dissatisfaction.

Key observations from the table include: a high prevalence of positive feedback across all videos, with students consistently finding the system helpful for class revision. The use of subtitles was widely appreciated for aiding in content comprehension. The instant quiz feature was deemed effective in confirming student understanding, and the video summaries were considered helpful for reviewing and summarizing the content. Negative feedback primarily focused on issues related to subtitle display, such as obscuration of content due to font size and timing discrepancies. Other concerns included the presence of stuttering speech, unclear audio, and difficulty accessing the quiz tab.

Further analysis can involve qualitative analysis of comments to gain a deeper understanding of student concerns and suggestions. A comparative analysis with the existing system can identify specific areas where the new system outperforms or falls short. Additionally, analyzing feedback by language can reveal potential differences in user experience across the three languages.

Table 1. User experience evaluation results of Co-Learning

	Video title 1	Video title 2	Video title 3	Video title 4
No. of students	79	80	83	72
No. of responses	79	77	82	68
Positive	71	47	30	63
Comments	Ability to prepare and review videos with a summary and a quiz	Improved quiz result display, making it easier to use than before	Enhanced word cloud functionality compared to the previous version	Organizing non-wordcloud information into three tabs creates a neater, more user-friendly interface
	Clear and easy-to-understand subtitles that facilitate learning	Improved word cloud displaying more relevant content	Better Chinese subtitles and audio quality than before	Wordcloud and other features have become much more user-friendly, making the tool feel highly effective
	Useful functions such as summaries, English subtitles, and voice narration	Conveniently locating the exact time to play corresponding videos while reviewing summary content	Effective learning through the quiz format, with summaries assigned to each topic for better comprehension	Implemented improvements enhance usability, with wordcloud showing significant improvements
	Instant quizzes effectively reinforce understanding			
Negative	8	30	52	5
Comments	Difficulty figuring out how to navigate to another page in the quiz	Subtitles disappearing from the screen after a certain period	White subtitles making them slightly difficult to read	Improved ease of viewing when items were separated, though overlapping subtitles sometimes reduced readability
	Wordcloud displaying different keywords for different languages	Difficulty reading slide content due to subtitle color and font	Picture quality issues and overlapping subtitles being slightly distracting	Video being easier to watch if the subtitle color were corrected
	Quiz selection resetting when changing a page	Irrelevant words being extracted in the wordcloud	Concerns about subtitle timing and readability, though reviewing them in the video remains convenient	
		Chinese pronunciation sounds unnatural	Poor image quality and difficulty reading text in the painting	

By addressing the identified issues and incorporating user feedback, the e-learning system can be further improved to enhance the learning experience for students.

7. Conclusion

The proposed Co-Learning model offers a transformative approach to digital education by integrating learning videos with summarized, well-structured keywords linked to external resources. This design not only improves accessibility but also reduces educational inequality through multilingual content and optimized learning strategies. Built on the principles of Cognitive Load Theory, the model enhances learning outcomes by addressing three core aspects: presenting structured content to manage intrinsic cognitive load, minimizes extraneous cognitive load by reducing interruptions and irrelevant information, and increases germane cognitive load to foster meaningful learning. Through advanced NLP techniques, the system refines learning materials by providing polished multilingual subtitles, concise summaries, keyword visualizations, external knowledge links, and auto-generated quizzes. Initial evaluations demonstrated a high acceptance rate of 71:8 (90%), underscoring the effectiveness of its cognitive load-aware design. However, issues with subtitle display and video quality led to decreased

acceptance rates of 61% and 37% in subsequent rounds. After addressing these concerns, the system achieved a final acceptance rate of 63:5 (93%), showcasing its adaptability and commitment to continuous improvement. The Co-Learning model presents a scalable solution for creating an engaging, efficient, and inclusive learning environment while reducing the burden on educators. Future developments will focus on integrating real-time feedback, adaptive learning features, and further automation to extend its applicability. By addressing modern educational challenges, the Co-Learning model sets a new benchmark for accessible and effective learning ecosystems.

References

[1] Sweller J. Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4; 1994; pp. 295–312.
[2] Sweller J. Element interactivity and intrinsic, extraneous, and germane cognitive load. Educational Psychology Review, 22; 2020; pp. 123- 138.
[3] Sweller J, Ayres P, Kalyuga S. Cognitive load theory. Springer. New York; 2011.
[4] Baddeley A. The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, 4(11); 2000; pp. 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2
[5] Mayer R E, Moreno R. Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist, 42(1); 2007; pp. 27–42.
[6] Sweller J. Cognitive load theory and educational psychology. Educational Psychology Review, 22(2); 2010; pp. 123–138.
[7] Sweller J. Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2); 1988; pp.257–285.
[8] Guo P J, Kim J, Robin R. How video production affects student engagement: An empirical study of MOOC videos. ACM Conference on Learning at Scale (L@S 2014); 2014.
[9] Chung C S, et al. Video indexing and searching: Past, present, and future. IEEE Transactions on Circuits and Systems for Video Technology, 25(2); 2015; pp. 215–227.
[10] Mayer R E. The Cambridge Handbook of Multimedia Learning. Cambridge University Press; 2005.
[11] Liu S, et al. Efficient content-based video retrieval with keyword extraction. Multimedia Tools and Applications, 75(6); 2016; pp. 3261–3278.
[12] Shen J, Pang R, Weiss R J, Schuster M. Jaitly N, Yang Z. Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada; 2018; pp. 4779-4783.
[13] Shao M, Basit A, Karri R, Shafique M. Survey of Different Large Language Model Architectures: Trends, Benchmarks, and Challenges. IEEE Access, 12; 2024; pp. 188664-188706. doi: 10.1109/ACCESS.2024.3482107.
[14] Sun L, Shi Z. Prompt Learning Under the Large Language Model. International Seminar on Computer Science and Engineering Technology (SCSET), New York, NY, USA; 2023; pp. 288-291. doi: 10.1109/SCSET58950.2023.00070.
[15] Sornlertlamvanich V, Aksorn N, Charoenporn T. Multi-lingual Support in Connective Learning Scheme for Refining and Connecting the Open Educational Videos. Proceedings of the Language Technologies for All (LT4All), UNESCO, France; December 5-6, 2019; pp. 20-22.