IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 5 November 2024, accepted 16 November 2024, date of publication 20 November 2024,
date of current version 29 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3503059

RESEARCH ARTICLE

Gaze-Driven Adaptive Learning System With ChatGPT-Generated Summaries

JAYASANKAR SANTHOSH[1,2], (Member, IEEE), ANDREAS DENGEL[1,2], AND SHOYA ISHIMARU[3], (Member, IEEE)

[1]SDS Department, German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany
[2]Department of Computer Science, RPTU Kaiserslautern-Landau, 67663 Kaiserslautern, Germany
[3]Department of Computer Science, Osaka Metropolitan University, Naka-ku, Sakai, Osaka 599-8531, Japan

Corresponding author: Jayasankar Santhosh (Jayasankar.Santhosh@dfki.de)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by DFKI Ethics Committee under Application No. ALS - 31/24.

ABSTRACT Enhancing student engagement and comprehension is crucial for effective learning. However, tracking and improving these dynamic states in real-time remains a significant challenge. This study addresses this gap by integrating real-time engagement prediction from gaze data with an adaptive learning system that utilizes ChatGPT-generated summaries to enhance student engagement and learning outcomes. Our experiment with twenty two (N=22) university students demonstrates the effectiveness of gaze data in predicting real-time engagement levels and the impact of adaptive interventions on student engagement, objective and subjective comprehension, and cognitive load. To predict the self-reported engagement and comprehension levels, two deep neural network models, InceptionTime and Transformers were employed. The Transformers model achieved better outcomes, with an average accuracy of 68.15% in predicting engagement levels across a 5-fold StratifiedGroupKFold cross-validation. The results revealed that the experimental group, which received the AI-driven interventions, exhibited significantly better learning outcomes, higher engagement, and better objective comprehension results compared to the control group. Additionally, we observed strong correlations between gaze metrics, engagement levels, and learning outcomes, suggesting that real-time adaptive interventions can dynamically enhance the educational experience. This study advances the field of educational technology by demonstrating the benefits of integrating gaze tracking and AI in learning environments, laying the foundation for dynamic learning interfaces that adapt to individual engagement levels, potentially improving both comprehension and involvement.

INDEX TERMS Adaptive learning, E-learning, ChatGPT API, eye-tracking, real-time engagement, personalized interventions, comprehension enhancement.

I. INTRODUCTION

In the rapidly evolving domain of educational technology, the objective of enhancing student engagement and learning outcomes remains a fundamental priority [1], [2], [3]. Engagement and comprehension are critical elements in education, directly influencing a student's ability to absorb, retain, and apply knowledge effectively [4], [5]. The widespread adoption of online learning platforms has transformed the way students interact with educational content, but it also presents several challenges, including the lack of personalization, limited engagement, and inadequate feedback [6], [7], [8]. To address these challenges, there is a growing need for adaptive learning systems that can leverage real-time data to provide personalized feedback and support.

Recent advancements in real-time data analytics and artificial intelligence offer promising solutions to these challenges, with the potential to create more responsive and adaptive learning environments [9], [10], [11]. For instance, frameworks such as INSIGHT have demonstrated the effectiveness

The associate editor coordinating the review of this manuscript and approving it for publication was Olarik Surinta.

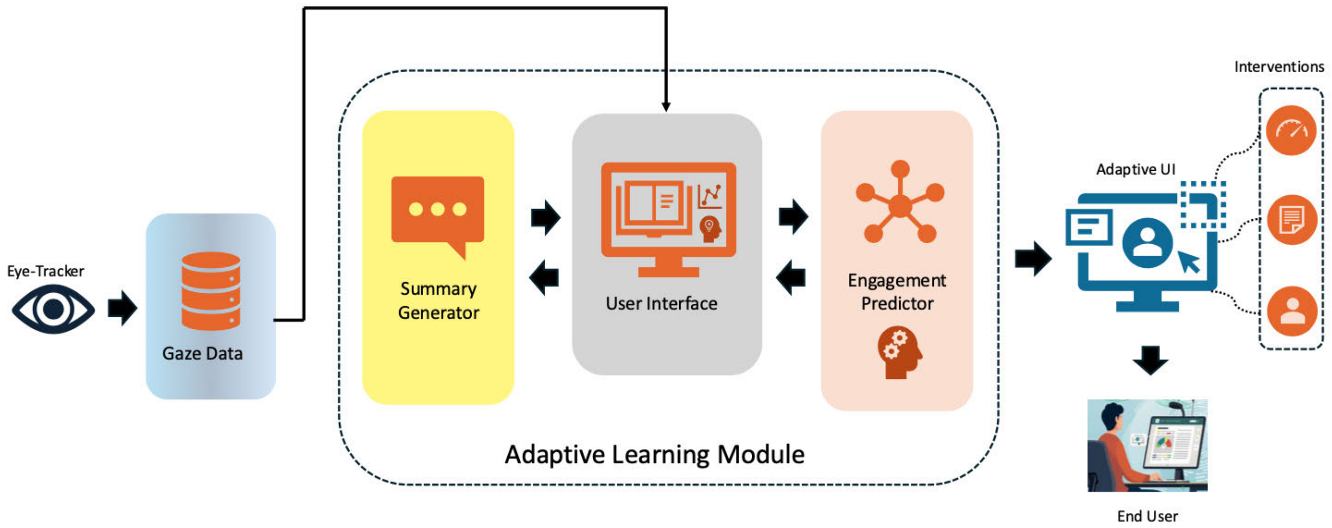


FIGURE 1. Architecture of the gaze-driven adaptive learning system, illustrating its key components.

of online reinforcement learning in training data-driven pedagogical policies that optimize student learning in narrative-centered learning environments [12]. Adaptive e-learning refers to a personalized learning approach where educational content is tailored to align with students' individual learning styles and preferences [13], [14], [15], [16]. Similarly, Wu et al. [17] proposed an intelligent tutorial-generating system, Self-GT, which incorporates cognitive computing and generative learning to generate personalized tutorials tailored to individual learners' learning preferences. This approach has shown effectiveness in generating personalized tutorials and has been successfully applied to an online self-aid learning system. However, current adaptive learning systems often rely on static models of learner behavior, which can fail to capture the dynamic and complex nature of human learning [18]. Additionally, these systems typically require extensive manual annotation and labeling of learning materials, which can be time-consuming and labor-intensive. Furthermore, current systems often lack the ability to provide real-time feedback and support, which can hinder their effectiveness in promoting learner engagement and motivation [19], [20]. Engagement, however, is a complex construct influenced by various factors, including the learner's emotional and cognitive states. The ability to accurately and promptly monitor these states provides invaluable insights into the learning process, allowing for timely interventions that can re-engage and refocus learners.

Despite these advancements, implementing effective real-time engagement monitoring and intervention systems presents several critical challenges [21], [22]. The integration of real-time analytics into existing educational frameworks requires careful consideration of privacy implications, ethical concerns, and the need for robust technological infrastructure. Traditional methods often struggle with capturing and responding to the immediate emotional and cognitive states of learners, particularly during critical moments of disengagement [23], [24]. While studies like Alshammari et al. [25] have demonstrated that adapting e-learning materials based on learning styles and knowledge levels can enhance outcomes, the challenge lies in developing systems that can provide truly personalized, real-time interventions. Current adaptive systems, despite their sophistication, still face limitations in accurately measuring and responding to the dynamic nature of student engagement. As such, while the potential for AI and data analytics in adaptive e-learning is significant, addressing these fundamental challenges in real-time engagement monitoring and intervention is essential for realizing their full impact on enhancing learner engagement and success.

To address these fundamental challenges in real-time engagement monitoring and intervention, our research introduces a novel gaze-driven adaptive learning system. While previous attempts at engagement detection have relied on delayed or indirect measures, our approach utilizes real-time gaze tracking, a non-invasive and continuous method of monitoring student engagement levels. The importance of this research lies in its potential to overcome several critical limitations of current adaptive learning systems: the lack of real-time intervention capabilities, the inability to detect subtle changes in engagement, and the challenge of providing personalized, contextually relevant support.

The system we propose is designed to provide tailored interventions through ChatGPT-generated summaries based on real-time learner engagement levels, ultimately aiming to optimize learning outcomes for students by delivering relevant and timely information. The motivation for this work stems from our previous research, where we evaluated and compared various methods for detecting engagement/interest [26], [27] and developed user applications to visually represent changes in users' affective responses [28],

[29]. In that earlier study, we identified real-time prediction as an area for future exploration, and this current research seeks to address that gap.

The significance of our gaze-driven adaptive learning system lies in its ability to integrate real-time data, allowing it to respond dynamically to the unique needs of each learner. By continuously analyzing gaze patterns, the system can detect moments when a learner may be disengaging from the material. In such instances, it can intervene promptly with concise, contextually relevant information designed to re-engage the learner and enhance their understanding of the subject matter. This research addresses two fundamental questions:

- RQ1: How effectively can real-time gaze data be used to predict and monitor student engagement levels in adaptive learning environments?
- RQ2: To what extent do AI-generated adaptive interventions, triggered by real-time engagement predictions, improve student learning outcomes and comprehension?

Figure 1 depicts the system's architecture, highlighting its key components. The system includes a user interface that visualizes eye movements and sends the fixation coordinates and pupil diameter to the engagement predictor for detecting the engagement levels on timely intervals. The predicted engagement level is then displayed on a dashboard within the user interface. The summary generator utilizes a Large Language Model (LLM), specifically a ChatGPT-based API, which processes the paragraph send from the user-interface as a prompt and displays the summary on the user interface only when the engagement level is low.

Building on these research questions, our study makes significant contributions to the field of adaptive learning through systematic evaluation of our proposed system. Our controlled experimental study shows that participants in the experimental group who received these adaptive interventions achieved better learning outcomes compared to those in the control group. Our findings reveal that the system not only enhances learners' objective comprehension through tailored interventions but also maintains sustained engagement with reduced visual strain. These results underscore the potential of real-time adaptive interventions to dynamically enhance the educational experience, setting the stage for future improvements in adaptive e-learning technologies.

The contributions of this research are as follows:

- Development of a novel and innovative adaptive learning system that integrates real-time gaze based engagement prediction and AI-generated text summaries.
- A thorough and comparative validation of the system's effectiveness in improving student engagement and comprehension from gaze and survey metrics.
- Insights into the relationship between cognitive processing, visual effort, and adaptive learning interventions.

The structure of this paper is as follows: Section II reviews the related work in the fields of gaze tracking, adaptive learning systems, and AI-generated content. Section III details the methodology, including the design of the adaptive learning system, the experimental setup, and the data collection process. Section IV presents the results of the experiment, analyzing the effectiveness of real-time gaze data in predicting engagement, the impact of adaptive interventions on student outcomes, and the system's effect on cognitive load and visual effort. Section V discusses the implications of the findings, potential applications, and limitations of the study. Finally, Section VI concludes the paper and suggests directions for future research.

II. BACKGROUND AND RELATED WORK

Recent years witnessed a significant advancement in the field of e-learning mainly aimed at enhancing student engagement and improving the learning outcomes. As the learning environments become increasingly complex, understanding the factors that influence student engagement and comprehension has become important. Traditional learning systems often struggle to adapt to the dynamic needs of learners, leading to gaps in engagement and comprehension. This section provides a comprehensive overview of the existing literature related to engagement and comprehension in learning environments, gaze tracking for engagement detection, adaptive learning systems, and the role of artificial intelligence in education. By examining these areas, we aim to highlight the current gaps in research and practice, ultimately establishing the foundation for our proposed system. This system seeks to bridge these gaps by integrating real-time gaze based engagement prediction and AI-generated summaries to create a more responsive and personalized learning experience.

A. ENGAGEMENT AND COMPREHENSION IN DIGITAL LEARNING

Engagement detection in digital learning focuses on establishing a learning environment that is adaptive, responsive, and centered around the student [30], [31]. Understanding this concept can be beneficial across multiple fields, particularly in education, where customized strategies can be designed to improve reading comprehension and enjoyment [32], [33]. Recent advancements in digital learning technologies have revolutionized how engagement and comprehension are monitored and enhanced. Abedi and Khan [34] advanced the state-of-the-art in engagement detection by developing a hybrid ResNet-TCN architecture that effectively captures both spatial and temporal aspects of student engagement in online learning environments, demonstrating superior performance compared to existing methods. While significant progress has been made in online learning environments, similar advances are being pursued in physical classrooms. For instance, EduSense [35] represents a breakthrough in classroom analytics, offering a comprehensive sensing system that captures visual and audio features correlated with effective instruction, demonstrating the potential for automated, continuous feedback mechanisms in traditional learning environments. Understanding these developments across both online and physical learning spaces is crucial for developing effective learning interventions and improving educational outcomes [36].

Recent research has revealed several key aspects of engagement detection and measurement in digital learning environments. Benabbes et al. [37] found a novel approach that can effectively predict learner engagement in online courses, revealing that most learners are observers and highlighting a nonlinear correlation between learner engagement and success. Ishimaru et al. [38] found that the variations in pupil diameter and nose temperature have high correlation with the cognitive states of students, like interest in learning materials in Physics. Wang [39] identified critical patterns of attention decline during the learning process and emphasized the importance of implementing effective feedback mechanisms. The effectiveness of different digital learning tools has been extensively studied. Bikowski and Casal [40] investigated how non-native English speaking students engage with interactive digital textbooks, while Hashim and Vongkulluksn [41] examined the impact of e-readers on reading comprehension, revealing that while these tools aid self-regulation, they may inadvertently affect reading enjoyment. Recent technological advances have introduced new possibilities, with Hew et al. [42] demonstrating the potential of chatbots in enhancing student goal setting and social presence in online learning environments. The relationship between engagement and performance has also been established, with Krasodomska and Godawska [43] finding significant correlations between student engagement in blended learning environments and academic performance, while noting demographic variations in these relationships.

Despite these advancements, several challenges remain in the field. Current research emphasizes the need for more effective methods of analyzing real-time data, balancing technology integration with human interaction, and addressing privacy and ethical considerations in data collection [44], [45]. Additionally, the requirement for robust technological infrastructure and the need to ensure inclusive access to digital learning resources continue to be significant concerns in the field.

B. GAZE BASED ENGAGEMENT DETECTION

Gaze-based engagement detection has become a powerful tool for assessing attention and involvement during learning interactions, providing valuable insights into user engagement levels [46], [47], [48]. The analysis of gaze patterns enables the identification of moments of high engagement or distraction, facilitating more responsive and adaptive user experiences [49], [50], [51]. The evolution of gaze-based engagement detection has seen significant advancement in recent years. Early work by Bidwell and Fuchs [52] established the foundation by creating an automated gaze system to classify student engagement through video data analysis and expert validation. Building on this, Carolis et al. [53] demonstrated how student engagement could be automatically measured through behavioral cues, particularly gaze behavior, emphasizing these measures' importance in the learning process.

Recent technological advances have led to more sophisticated approaches in gaze-based engagement detection. Chen et al. [54] developed a state-of-the-art multi-modal deep neural network that predicts student engagement by analyzing both gaze direction and facial expressions in collaborative learning settings, revealing that students with higher gaze ratios and positive expressions demonstrated better test performance. Similarly, Sharma et al. [55] introduced an innovative concentration index based on eye gaze and emotion weights, providing a more comprehensive approach to detecting student engagement during learning activities. The latest developments in the field have focused on addressing real-world implementation challenges. Mathew et al. [56] made significant contributions by introducing GESCAM, a novel dataset and network capable of identifying gaze fixations within complex classroom scenes, offering insights into human attention across diverse educational contexts. These advances build upon earlier work, such as Jacob et al. [26], who demonstrated the effectiveness of eye-tracking metrics in detecting user interest during reading tasks and proposed methods for real-time interest prediction. Current research emphasizes the need for more robust methods that can handle varying environmental conditions, different learning contexts, and diverse student populations. Additionally, the integration of real-time gaze tracking with adaptive learning systems, while promising, requires further investigation to establish practical and scalable solutions that can enhance the learning experience while maintaining accuracy and reliability.

C. ADAPTIVE LEARNING SYSTEMS

Adaptive learning has evolved significantly, moving from simple personalized content delivery to sophisticated AI-driven systems that dynamically adjust to learner needs [57], [58]. Modern adaptive learning systems use advanced algorithms and artificial intelligence to modify content, delivery, and pace of instruction based on real-time learner performance and engagement [59], [60], [61], [62].

Recent developments have significantly enhanced adaptive learning capabilities. Hussain et al. [63] demonstrated state-of-the-art performance with their novel multi-layer topic modeling approach, which integrates the Felder-Silverman learning style model with fuzzy logic and sentiment analysis to accurately detect and adapt to different learning styles. Sayed et al. [64] introduced APPEAL, an advanced platform that combines cognitive, behavioral, and affective learner modeling with AI to provide comprehensive personalized learning experiences, showing significant improvements in learning effectiveness and student satisfaction. The integration of emotional and physiological data has emerged as a crucial advancement in adaptive learning systems. Wei et al. [65] initiated the use of eye-tracking technology to sense learners' interest and emotional states, while Tyng et al. [66]

provided neuroimaging evidence for the critical role of emotions in learning and memory. Sargazi et al. [67] developed an AI-based decision framework that identifies and implements micro-brake activities based on learners' emotional states, demonstrating improved learning performance through emotional regulation.

The latest developments in adaptive learning systems focus on deep learning and neural network integration. Omar et al. [68] shows how neural networks can enhance learning management systems by creating personalized learning paths and enabling real-time content adaptation based on learner behavior. A comprehensive review by Essa et al. [69] highlights the growing trend toward using machine learning for intelligent, adaptive e-learning environments, while also identifying the need for more comparative studies of deep learning methods.

D. LARGE LANGUAGE MODELS (LLM) FOR DIGITAL LEARNING

Digital learning is being revolutionized by Large Language Models (LLMs), which are able to generate human-like text and provide personalized educational experiences [70]. These models show advanced capabilities in creating customized learning materials, answering student queries, and offering real-time explanations [71], [72], [73].

Recent advances in LLM applications for education have shown promising results across various domains. In programming education, Gabbay et al. [74] demonstrated GPT-4's superior performance in generating code assignment feedback compared to traditional Automated Test-based Feedback tools. Azaiz et al. [75] found that GPT-4 Turbo provides more structured and consistent feedback in programming courses, though noting some limitations in consistency. The impact of LLMs on writing and language learning has been particularly significant. Meyer et al. [76] conducted a large-scale study with 459 upper secondary students, revealing that GPT-3.5-turbo-generated feedback significantly enhanced essay revision performance and student motivation. In the domain of language learning, Xu et al. [77] advanced the field by developing mHyER, a novel method utilizing LLMs for zero-shot exercise retrieval, demonstrating superior performance in personalizing language learning exercises.

However, the integration of LLMs in education presents both opportunities and challenges. Arora et al. [78] identified concerning trends of student over-reliance on LLMs in advanced coursework, emphasizing the need for updated curricula that incorporate effective prompting strategies. To address accuracy concerns, Abu et al. [79] developed an innovative approach using knowledge graphs to enhance LLM prompts, significantly improving the precision and reliability of generated educational content.

Even with these advances in LLM applications for education, several significant challenges exist in their effective implementation. A primary concern is ensuring the accuracy and reliability of AI-generated educational content, particularly as studies have shown limitations in content validity and consistency [80]. The integration of LLMs in education requires careful consideration of ethical implications, including privacy concerns, fairness in access, and potential impacts on critical thinking skills [81]. There is a growing need to strike a balance between utilizing LLM capabilities and preventing student over-reliance on these tools, while simultaneously developing effective strategies for integrating these technologies into existing educational frameworks. Educational institutions also face the challenge of maintaining high-quality learning experiences while utilizing automated systems, with particular emphasis on ensuring transparency and oversight in AI-generated content [82]. The widespread adoption of LLMs in education raises important considerations about equity and accessibility, as research indicates that not all students have equal access to or benefit equally from these technologies [83].

III. METHODOLOGY

This study employed a mixed-methods approach to design, develop, and evaluate a gaze-driven adaptive learning system with LLM generated summaries. The system was designed to provide personalized learning experiences for learners by adapting to their individual affective state. To achieve this, we integrated eye tracking technology with a ChatGPT-based summary generation system, which enabled the system to provide real-time interventions and support to learners. We utilized a Tobii 4C remote eye-tracker with pro license, to capture the gaze coordinates and pupil diameters at a robust sampling rate of 90 Hz. In this section, we describe the methodology used to design and develop the system, including the system architecture, engagement prediction model, and ChatGPT API integration.

A. SYSTEM ARCHITECTURE

The adaptive learning system comprises of four main components: a gaze tracking module, an engagement prediction model, a ChatGPT API integration, and a user interface. The system flow is as follows:

1) GAZE TRACKING MODULE

The module collects gaze tracking data from learners as they interact with the system. This data includes the x and y coordinates of the learner's gaze point, as well as other metadata such as timestamp and pupil diameter. The raw gaze data is processed to extract fixation data, which consists of the points where the learner's gaze has stabilized for a certain period, along with the corresponding pupil diameter values at each fixation point. This processing involves filtering out noise and artifacts from the raw data, and identifying the fixation points based on the learner's gaze behavior. The fixation data, which includes both gaze coordinates and pupil diameter values, is stored in an array, which is updated in real-time as the learner interacts with the system. Our system uses both gaze data and fixation data, but for different purposes. The gaze data represents the raw, unprocessed data

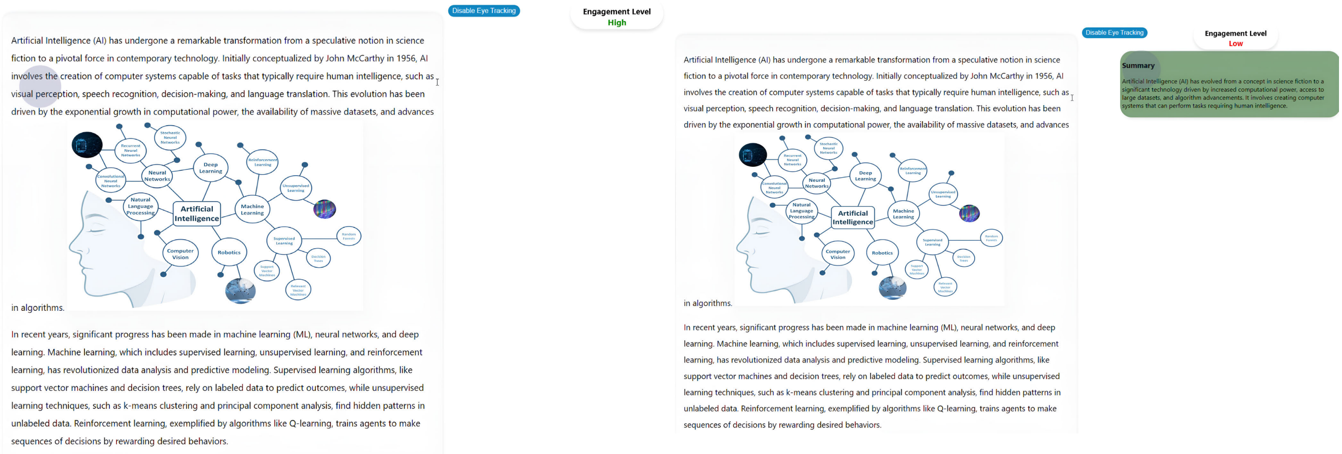


FIGURE 2. **The system interface with real-time engagement prediction dashboard and ChatGPT based summary generation feature. The figure on the left depicts the system with high detected engagement and the figure on the right depicts system with low detected engagement for a learning material.**

collected from the eye tracker, which includes the x and y coordinates of the learner's gaze point and the fixation data on the other hand, represents the processed data that indicates where the learner's gaze has stabilized for a certain period, and includes both gaze coordinates and pupil diameter values. The raw gaze data is used to create a real-time visual representation of the user's gaze, while the fixation data is used to predict engagement levels, as it provides more insight into learner behavior and attention. The fixation data is transmitted at 5-second intervals, allowing the system to make timely predictions about the user's engagement level and respond accordingly.

2) ENGAGEMENT PREDICTION MODEL

The engagement prediction model is a deep learning algorithm that has been pre-trained on data from our previous research [27], [29], which focused on interest and engagement detection using various evaluation metrics. This model takes a combination of fixation data and pupil diameter information from the gaze tracking module and outputs a predicted engagement score. The model is trained on a dataset of labeled examples, where each example consists of a set of fixation points, pupil diameter values, and a corresponding engagement level. When a new set of fixation data and pupil diameter values is received, the model processes the data by normalizing the data, and making a prediction using the trained deep learning model. The engagement score is predicted as binary values with '0' representing low engagement and '1' representing high engagement. The final engagement score is then sent back to the user interface as a response, where the engagement level is displayed as an interactive dashboard to the learner.

3) CHATGPT BASED SUMMARY GENERATION

To generate the LLM-based summary of the learning material, we utilized a ChatGPT-based API, which is integrated into our system. In our summarization process, we employed the gpt-3.5-turbo model from OpenAI, known for its ability to generate high-quality text outputs efficiently. This model is particularly well-suited for tasks that require natural language understanding and generation, making it an ideal choice for summarizing educational content. Based on the learner reading behavior, when a request is made to summarize a paragraph, the API is called with a prompt that includes the paragraph text. The prompt used for summarization is carefully structured to concentrate solely on the provided content, minimizing the risk of irrelevant information and enhancing the relevance of the generated summary. We limited responses to 80 tokens to ensure our summaries are concise and informative, capturing key points effectively and not being too long which could distract the user from original content. The temperature was set to 0.5, striking a balance between creativity and coherence, which enables the model to generate engaging and accurate outputs that remain true to the original content. Upon receiving a request, the ChatGPT API processes the input paragraph using the specified prompt, generating a set of summaries $\{S_1, S_2, \ldots, S_n\}$. Each generated summary S_i undergoes a refinement process, which includes removing unwanted elements, such as URLs, and ensuring that the summary concludes with a complete sentence. Finally, the most suitable summary S_{final} is selected through the function $\text{Process}(S_i)$ and returned as a response to the main user interface.

$$S_{\text{final}} = \text{Process}(S_i) \text{ where } S_i \in \{S_1, S_2, \ldots, S_n\} \quad (1)$$

where:

- S_{final} is the final selected summary.
- S_i is the i-th generated summary.
- $\{S_1, S_2, \ldots, S_n\}$ is the set of all generated summaries.
- $\text{Process}(S_i)$ is the function that includes the following steps:
  - Input Processing: The ChatGPT API processes the input paragraph with the prompt.

TABLE 1. Post-survey for experiment participants: questions 1-3 were administered to all participants, while Questions 4-6 were specific to the experimental group.

No	Question	Type	Scale
1.	Rate the overall intensity of your engagement while reading the document.	Engagement	1-7 (1: lowest, 7: highest)
2.	Rate the overall level of comprehension or understanding while reading the document.	Comprehension	1-7 (1: lowest, 7: highest)
3.	Rate your self-confidence while answering the comprehension-based questionnaire.	Self-Confidence	1-7 (1: lowest, 7: highest)
4.	How well do you think your gaze behavior predicted by the system reflects your level of interest or engagement with the document?	Gaze-Reflect	1-7 (1: lowest, 7: highest)
5.	Rate the helpfulness of the summary in capturing the main points of the document.	Summary-Help	1-7 (1: lowest, 7: highest)
6.	How often did you refer back to the summary while reading?	Summary-Refer	1-7 (1: lowest, 7: highest)

- Summary Generation: The model generates multiple summaries based on the input.
- Refinement: The generated summaries are processed to remove unwanted parts (e.g., URLs) and ensure they end with complete sentences.
- Selection: The final summary is selected based on criteria such as coherence and relevance.

4) USER INTERFACE

The user interface, developed using React.js, provides a seamless and user-friendly experience. The participants begin by reading instructions, followed by a list of learning materials to read. When a participant selects a material, the system navigates to the corresponding content, where a toggle button allows them to enable or disable eye-tracking. Once eye-tracking is enabled, the system sends the participant's fixation data and pupil diameter to the engagement predictor at five second intervals. The predictor then returns a value indicating the participant's level of engagement, which is displayed in real-time on the dashboard integrated into the interface. The dashboard displays this value as either 'high' or 'low' and if the engagement level is 'low', the system responds by displaying a summary of the material adjacent to the paragraph being read. This adaptive intervention is designed to re-engage learners, helping to improve their comprehension and overall engagement if they become distracted or disinterested.

B. GAZE FEATURE METRICS

To gain a better understanding of the gaze behavior and attention patterns of the participants, a range of features were extracted from the raw gaze data. These features were carefully selected to capture the differences of gaze behavior and provide insights between the experimental and control groups.

The extracted features included:

- **Fixation metrics**: mean fixation duration, standard deviation of fixation duration, fixation count, and mean pupil diameter. These metrics provide information about the participant's ability to focus and maintain attention on specific areas of the learning material.
- **Saccade metrics**: mean saccade length, mean saccade angle, mean saccade velocity, standard deviation of saccade length, standard deviation of saccade angle, and standard deviation of saccade velocity. These metrics capture the participant's eye movement patterns, including the speed and direction of their gaze.
- **Blink metrics**: blink count and blink rate. These metrics provide information about the participant's level of fatigue, distraction, or disengagement.

The extraction of these gaze metrics was motivated by the need to understand the differences in gaze behavior between the experimental and control groups. By analyzing these features, we aimed to identify patterns and trends that could inform the development of more effective adaptive learning systems.

C. MODEL TRAINING AND EVALUATION

To predict the engagement and comprehension levels reported by the participants, we employed two deep neural networks: an InceptionTime network, and Transformer network, following the architecture specifications detailed in our previous work [29]. The selection of these models was based on their demonstrated effectiveness in processing sequential data and their ability to capture both local and global patterns in time series data. These models were trained on raw gaze data, including x and y coordinates and pupil diameter values, to learn the complex patterns and relationships between gaze behavior and the predicted variables.

The InceptionTime network architecture is designed to capture diverse temporal patterns in time series data using a multi-branch convolutional approach. It consists of multiple inception modules, each with four parallel branches: a 1×1 convolutional layer with 64 filters, a 3×3 convolutional layer with 32 filters, and a 5×5 convolutional layer with 64 filters, all followed by batch normalization and an activation function. The fourth branch includes a max pooling layer with a pool size of 3, followed by a 1×1 convolutional layer with 32 filters, batch normalization, and an activation function. These branches are concatenated to integrate features at different scales, enhancing the model's ability to learn complex patterns. The model is constructed by stacking these inception modules, followed by a global average pooling layer to reduce spatial dimensions and a dropout layer with a rate of 0.3 to prevent overfitting. The final output layer is a dense layer with a sigmoid activation function, making it suitable for binary classification tasks.

The Transformer network utilizes a multi-head attention mechanism to capture complex temporal dependencies. The model begins with an input layer that processes data sequences through a multi-head attention layer with two heads and a key dimension of eight, incorporating dropout for regularization. The output is normalized and combined with the original input using a residual connection. A feed-forward network is followed, consisting of a 1D convolutional layer with 64 filters and ReLU activation, another convolutional layer to restore input dimensions, and additional layer normalization with a residual connection. The data is then passed through dense layers with dropout to prevent overfitting, ending in a final dense layer for predictions. This architecture, using the Adam optimizer and binary cross-entropy loss, is well-suited for binary classification tasks and identifying complex patterns in time series data.

To evaluate the performance of our models, we used Stratified Group K-Fold Cross-Validation with K=5, combining the benefits of stratified cross-validation and GroupKFold. This approach ensures that each fold maintains the proportion of classes (low vs. high comprehension/engagement) while also respecting the group structure (experimental vs. control). By using Stratified Group K-Fold Cross-Validation, we can ensure that our model is evaluated on a representative sample of the data, with both groups and classes balanced in each fold and want to ensure that our model is generalizable to both groups.

IV. USER STUDY AND DATA COLLECTION

The user study and data collection was designed and conducted to investigate the effectiveness of the adaptive learning system in enhancing learners' comprehension, engagement, and self-confidence in a learning environment. The primary objective of the study was to determine whether the adaptive learning system, which utilizes gaze tracking and ChatGPT-generated summaries, can improve learners' understanding and retention of learning material, increase their engagement and motivation, and boost their confidence in their ability to learn.

A. PARTICIPANTS

The study involved recruiting 22 university students, comprising 11 male and 11 female participants, with ages ranging from 22 to 29 years (M = 25, SD = 2.5). The study was approved by the DFKI Ethics Committee and was conducted in accordance with the requirements mentioned by the committee. To ensure that language proficiency did not influence participants' comprehension and overall reading experience, we specifically recruited individuals who exhibited advanced proficiency in English, as determined by standardized language assessments and self-reported language use. All participants joined the experiment after providing informed consent, and they had the freedom to withdraw from the study at any point if they chose to do so.

B. EXPERIMENTAL DESIGN

The experiment involved a carefully designed reading task followed by a quiz and survey, aimed at investigating the impact of adaptive learning on comprehension, engagement, and self-confidence. To achieve this, we curated a collection of ten documents, each selected to elicit distinct levels of engagement and comprehension.

To capture the participants' gaze patterns and eye movements as they read the documents, we employed a Tobii 4C eye tracker with a pro license with a sampling frequency of 90 Hz, mounted to a display monitor. The study employed a between-subjects design, where 22 participants were randomly assigned to either an experimental group (n=11) or a control group (n=11). The control group did not receive any real-time engagement prediction or summary generation, whereas the experimental group received both.

After reading each document, participants were presented with a quiz comprising four comprehension-based questions (objective comprehension). The post-survey included self-reported ratings of their comprehension, engagement, and self-confidence (in answering the quiz questions) and some other questions using a 7-point Likert scale which can be referred from Table 1. The scale ranged from 1 (indicating the lowest level) to 7 (indicating the highest level), allowing participants to reflect on their perceived understanding, interest, and confidence in their responses. By combining the objective comprehension scores from the quizzes with the subjective self-reported ratings, we were able to obtain a more accurate and comprehensive picture of the participants' comprehension levels. This integrated approach helped to reduce the potential for biases in self-reported ratings and provided a more robust understanding of the differences in comprehension levels between the experimental and control groups.

V. RESULTS

The results of the study are presented in this section, which provides an overview of the findings from the gaze metrics analysis, survey response analysis, and predictive modeling analysis. The study aimed to investigate the effectiveness of the adaptive learning system in enhancing learner engagement, comprehension, and self-confidence. The results are presented in three subsections, each focusing on a different aspect of the study.

A. GAZE METRICS ANALYSIS

The gaze metrics analysis is an important factor as it provides a quantitative measure of the gaze behavior of the participants belonging to the experimental and control group. The gaze metrics could provide insights into the cognitive processes underlying the gaze behavior and identify potential differences between the experimental and control groups. A range of features including mean fixation duration,

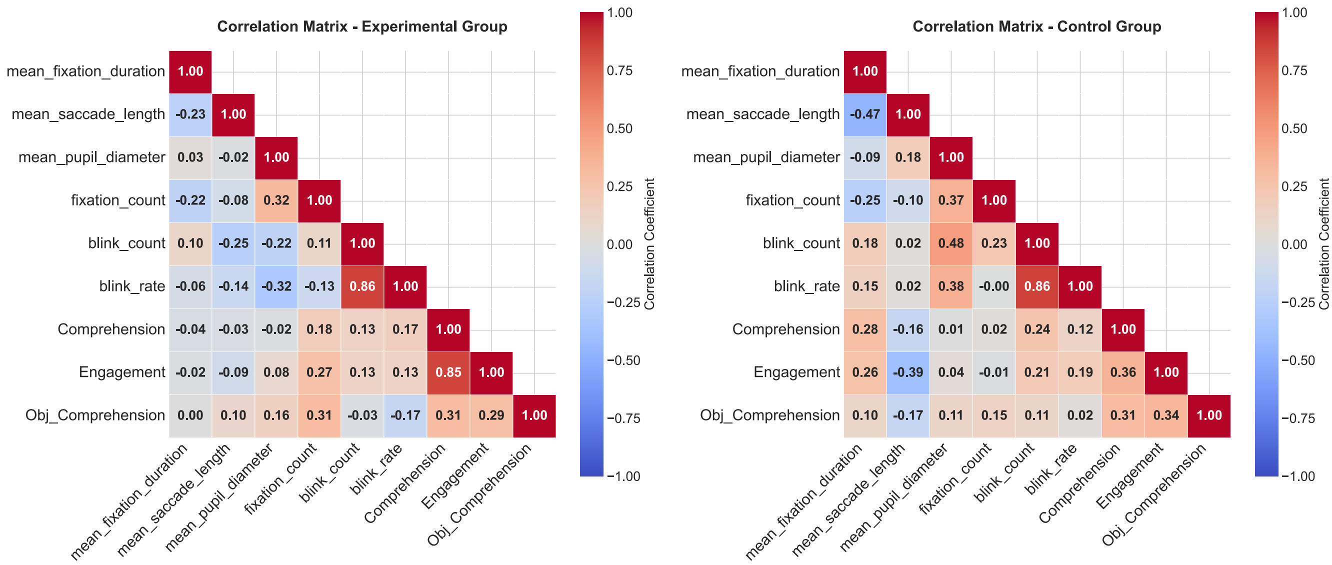


FIGURE 3. **Correlation matrices of eye-tracking metrics, comprehension, engagement, and confidence for experimental and control groups.**

mean saccade length, mean saccade angle, mean saccade velocity, standard deviation of fixation duration, standard deviation of saccade length, standard deviation of saccade angle, standard deviation of saccade velocity, mean pupil diameter, fixation count, saccade count, blink count, and blink rate were extracted from the raw gaze data to identify the variations in the gaze pattern across both groups.

Figure 3 depicts the correlation matrices for the gaze-metrics to the subjective comprehension, engagement, and objective comprehension belonging to the control and experimental groups. In the experimental group, the gaze metrics show weaker correlations to engagement and subjective comprehension levels which could suggest that the intervention is modulating the typical relationship between eye-movements and cognitive responses. In the control group, longer fixation duration is associated with better comprehension and engagement while it is reversed in the experimental group, possibly due to the intervention causing more strategic reading patterns. The saccade length has a positive correlation with objective comprehension in experimental group while a negative correlation in the control group. This suggests that participants in experimental group viewing the generated summary leading to higher saccade length has a higher objective comprehension compared to the control group. The experimental group shows weak negative correlations between saccade length and cognitive outcomes, while the control group shows high negative correlations with engagement and subjective comprehension. For the pupil diameter, the experimental group shows weak correlations with the cognitive measures while the control group shows stronger positive correlations with comprehension and engagement. From the observations, the correlation matrices reveal that the intervention in the experimental group appears to be modifying the typical relationships between the gaze-metrics and cognitive outcomes (comprehension and engagement)suggesting that the real-time engagement prediction and summary generation are indeed influencing reading behavior and cognitive processing.

To gain a deeper understanding of the gaze-metrics for the different levels of engagement and comprehension, figure 4 provides a comprehensive analysis of gaze-metrics across varying levels of comprehension and engagement for both control and experimental groups. In the control group, there is a noticeable trend where mean fixation duration increases with higher engagement and comprehension levels while the saccade length decreases with higher engagement and comprehension levels suggesting that participants in the control group maintain consistent focus on the material, reflecting a traditional reading pattern. On the other hand, the experimental group, which received real-time engagement predictions and summary interventions, shows more controlled and stable metrics, such as a flatter fixation duration and a less steep saccade lengths with higher engagement and comprehension levels suggesting that the intervention might be stabilizing these metrics possibly by keeping participants more focused on relevant parts of the material. The experimental group shows a notable increase in pupil diameter for lower comprehension and engagement levels compared to the control group. The increased pupil diameter in the experimental group could indicate increased arousal or a higher cognitive effort in the experimental group, possibly due to the additional processing of summaries and the more engaging nature of the adaptive reading experience. The higher fixation count in the experimental group especially at higher levels of comprehension and engagement could suggest that participants are making more fixations, likely because they are focusing more intensely on the material, facilitated by the real-time predictions and

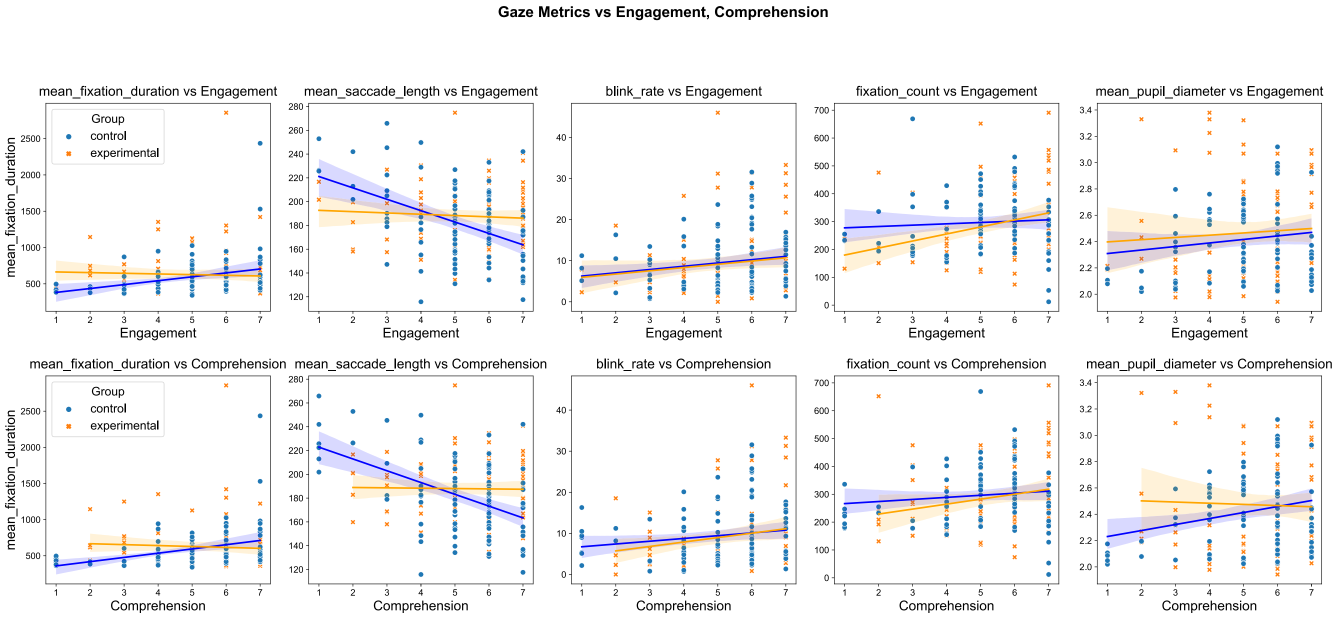


FIGURE 4. **Comparison of eye-tracking metrics between control and experimental groups across engagement and comprehension levels (1-7).**

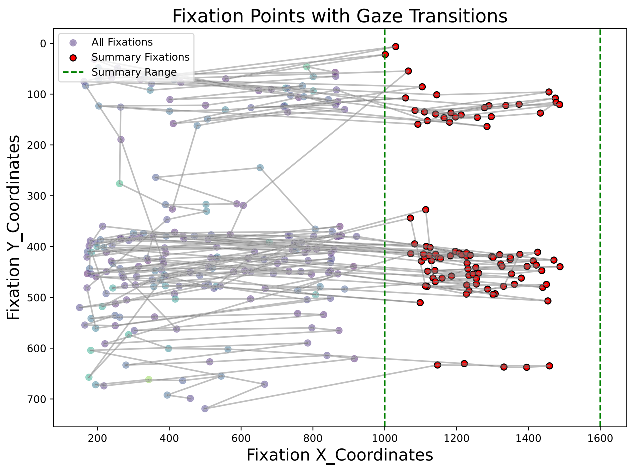


FIGURE 5. **The transition of fixations between the actual text and generated summary for a participant in experimental group detected with low engagement.**

summaries. These patterns suggest increased engagement and focused reading for experimental group, potentially due to the intervention and the similar trends in both comprehension and engagement metrics imply the intervention may enhance both simultaneously. The intervention appears to help stabilize some gaze metrics, reduce unnecessary eye movements, and potentially increase cognitive engagement, which is reflected in the metrics' relationship with both engagement and comprehension suggesting that real-time interventions might be aiding participants in focusing better and processing information more effectively.

Figure 5 illustrates the dynamic interaction between the original text and the generated summaries. The visualization shows a shift from the main text to the summary, triggered by the low engagement levels and indicated by a cluster of fixation points in the summary section. This concentration of fixations suggests higher attention to the concise, relevant information provided. The figure also demonstrates a subsequent transition back to the original content, with renewed fixation patterns in the main text area. This pattern implies that the summaries not only captured attention but also effectively re-engaged participants with the original material. The intervention appears to have served as a bridge, helping participants refocus and return to the main content with potentially enhanced understanding and engagement. Our findings suggest that the summaries served two purposes: they offered quick, relevant information and also encouraged readers to dive back into the detailed content with renewed interest. This approach seems to create a more dynamic and engaging reading experience.

B. SURVEY RESPONSE ANALYSIS

The post-survey mainly aimed at collecting the responses from the learners about their comprehension, engagement and self-confidence levels for both the experimental and control group. The experimental group had additional three questions to get their impressions on the real-time engagement tracking system and the summary response helpfulness.

Figure 6 illustrates the distribution of comprehension, engagement and confidence responses by the two groups (control vs. experimental). It can be inferred from the plot that experimental group with the adaptive learning system had a higher engagement and confidence levels compared to the control group with some improvement in subjective comprehension. To determine the significance of these differences, we conducted t-tests for each of the three

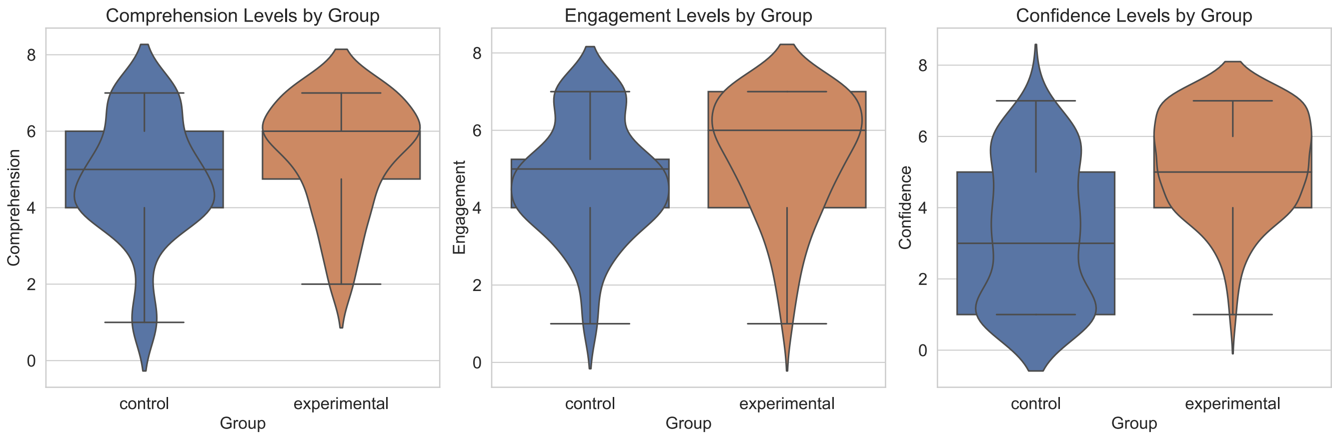


FIGURE 6. The distribution of subjective comprehension, engagement, and confidence levels by group (control vs. experimental).

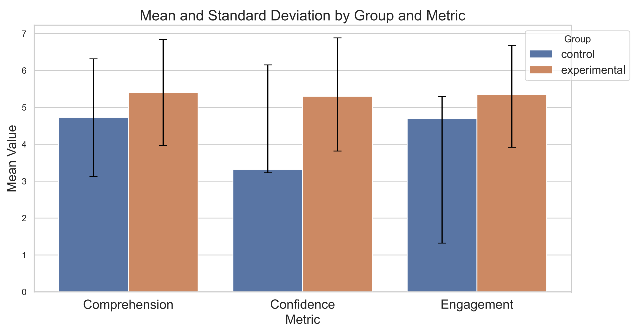


FIGURE 7. Mean and standard deviation of comprehension, confidence, and engagement by group.

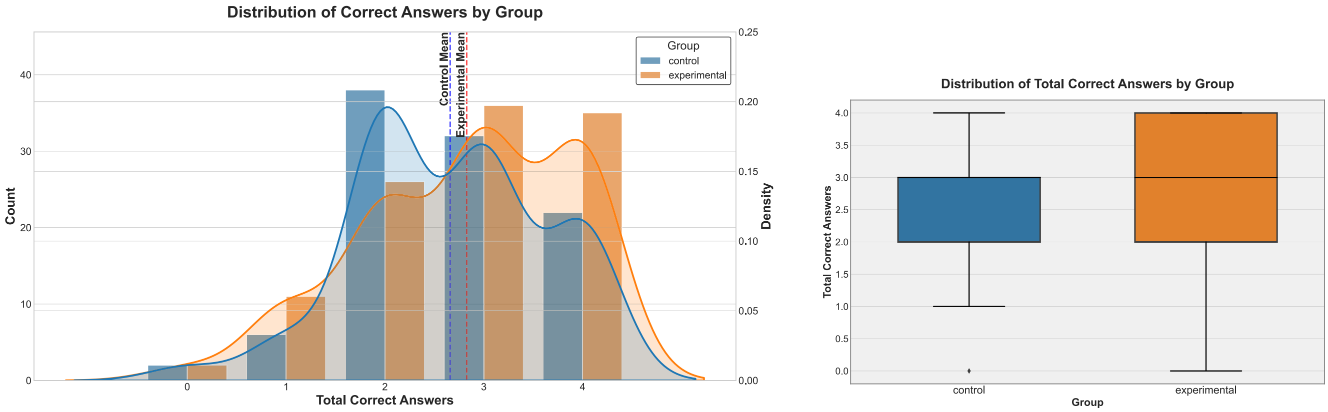
subjective variables. The results showed that the experimental group had significantly higher engagement (t-statistic = 3.12, p-value = 0.0021) and confidence (t-statistic = 8.22, p-value < 0.0001) levels compared to the control group. Additionally, the experimental group showed a significant improvement in comprehension (t-statistic = 3.17, p-value = 0.0018) compared to the control group. Figure 7 provide insights into the mean differences across the subjective ratings of the three metrics between the control and experimental groups. The experimental group's access to the adaptive system likely contributed to higher engagement and confidence levels, with some improvement in comprehension. The variability in responses suggests that while the tools were beneficial, their effectiveness varied among individuals.

To test our main hypothesis that the adaptive learning system improves objective comprehension and to minimize potential bias from the subjective comprehension ratings, an objective comprehension quiz was administered for each document, consisting of four topic-specific questions. The results of the objective questionnaire revealed promising differences in performance between the control and experimental groups, largely supporting our hypothesis. Figure 8a depicts the count of total correct answers for the two groups and figure 8b shows the overall distribution of correct answers between the groups. The experimental group had a higher proportion of participants achieving perfect scores (4/4), indicating that the adaptive interventions were effective in enhancing comprehension for a significant portion of the group. While the experimental group showed overall improvement, we observed that some participants in this group still scored lower (1/4 or 2/4). This variation suggests that the effectiveness of the adaptive interventions, particularly the generated summaries, may not have been uniform across all participants. Several factors could have contributed to this varied effectiveness like individual learning styles, prior knowledge of the topic, or varying levels of engagement with the adaptive features which could influence how participants benefited from the summaries. Despite these variations, the higher proportion of perfect scores in the experimental group supports our hypothesis that the adaptive learning system generally improves objective comprehension. These findings not only validate the potential of our approach but also highlight areas for future refinement.

C. PREDICTIVE MODELING ANALYSIS

The engagement and comprehension ratings reported by the participants were predicted using two deep neural networks: an InceptionTime model and a Transformer network. The raw gaze data were used to train the models for predicting the engagement and comprehension responses from the participants. It was formulated as a binary classification task, where ratings (1)-4) were treated as 'low' and ratings (5-7) as 'high'. To evaluate model performance, we employed a 5-fold StratifiedGroupKFold cross-validation strategy, ensuring that each fold contained a balanced representation of both experimental and control groups, as well as high and low engagement/comprehension ratings.

Based on the evaluation metrics presented in Table 2, a comparative analysis reveals that the Transformers model consistently outperformed the InceptionTime model. For engagement prediction, the Transformers model achieved an



(a) Correct-Answer count (b) Objective Comprehension Distribution

FIGURE 8. Distribution of total correct answers (out of 4 questions) for objective comprehension across control and experimental groups.

accuracy of 68.15% and an F1-score of 65.82, compared to InceptionTime's accuracy of 62.71% and F1-score of 60.45. Similarly, in comprehension prediction, the Transformers model demonstrated superior performance with an accuracy of 69.60% and an F1-score of 64.32, while the InceptionTime model recorded an accuracy of 60.32% and an F1-score of 59.18.

TABLE 2. Summary of evaluation metrics for engagement and comprehension.

Models	Engagement		Comprehension	
	Accuracy	F1-score	Accuracy	F1-score
InceptionTime	62.71	60.45	60.32	59.18
Transformers	**68.15**	**65.82**	**69.60**	**64.32**

To provide a more detailed view of model performance across the 5-fold cross-validation, Figure 9 illustrates the engagement accuracy of both models for each fold. The Transformers network consistently outperformed the InceptionTime model across all folds. The InceptionTime model's accuracy ranged from 61.5% to 64.1%, with an average of 62.71%. In contrast, the Transformers model demonstrated higher and more stable performance, with accuracies ranging from 66.9% to 69.3%, averaging 68.15%. The superior performance of the Transformers model can be attributed to its ability to capture long-range dependencies in sequential data, which is particularly beneficial for analyzing gaze patterns over time.

VI. DISCUSSION

In this section, we delve into the key contributions, outcomes,and insights that have been achieved through this study. This includes both the theoretical implications of the research and the practical applications that might result from it. This section considers the the novel real-time adaptive system and it's implications to the learning outcomes and the relevance of the research question to ongoing discussions in the field. This section also provides insights into the study by critically examining its challenges and limitations.

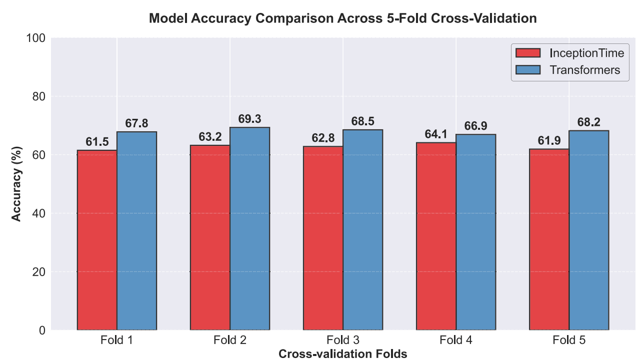


FIGURE 9. Model accuracy comparison across 5-fold cross-validation.

A. GAZE BASED REAL-TIME ENGAGEMENT PREDICTION

The user interface of the system is designed to capture the real-time gaze data recorded from the participants and to display the predicted engagement levels in real-time. An interactive dashboard was designed and displayed on the user interface to visually depict the engagement levels to the user based on the predicted engagement values. The raw gaze data was pre-processed and extracted fixation data along with the pupil diameter was send to the pre-trained engagement prediction model for detecting the engagement levels of the user with an interval of 5 seconds. To assess the frequency with which the summary was displayed on the interface, the count of summaries was recorded after participants in the experimental group read each document. This approach aimed to establish a direct correlation between the summary counts and the engagement ratings provided by the users, as well as to evaluate how the post-survey responses corresponded with the generated summary counts. In addition, participants in the experimental group were asked to rate the gaze behavior predicted by the system in the

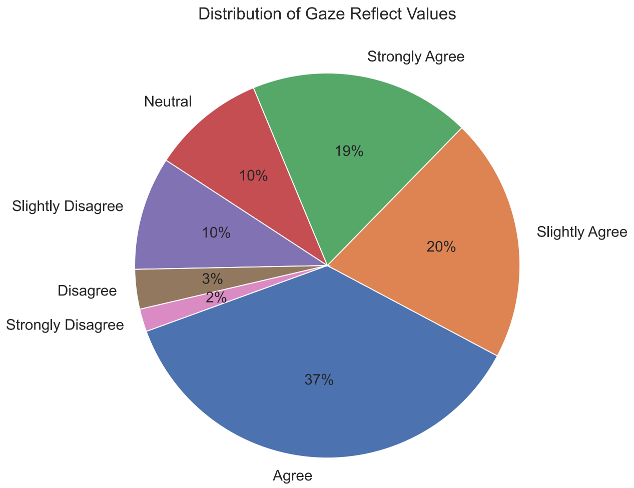


FIGURE 10. **Distribution of ratings provided by the participant reflecting the effectiveness of the system in predicting the gaze behavior for engagement prediction.**

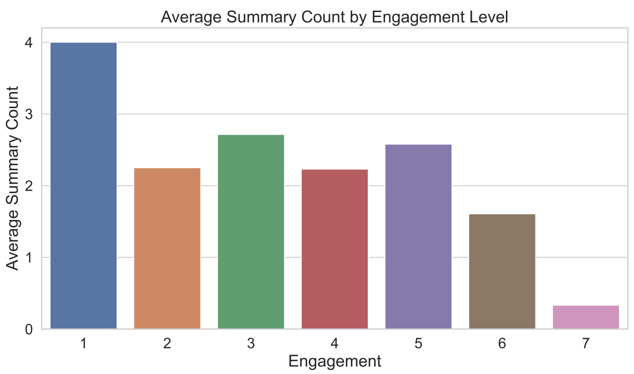


FIGURE 11. **The average summary count corresponding to different levels of engagement as rated by the participants.**

post-survey, which served as an indicator of their interest or engagement with the document.

Figure 10 illustrates the distribution of ratings provided by the participants, reflecting the effectiveness of the system in predicting the gaze behavior for real-time engagement prediction, on a scale from 1 to 7, where '1' indicates 'strongly disagree' and '7' indicates 'strongly agree.' A significant portion of participants, approximately 37% (rating: '6 - agree'), felt that the system effectively captured gaze behavior and provided accurate real-time engagement predictions. In addition, 19% rated the system as 'strongly agree,' while 20% indicated 'slightly agree.' Conversely, only 2% and 3% of participants rated the system as 'strongly disagree' and 'disagree,' respectively. These results suggest that the majority of participants recognized the system's effectiveness in predicting real-time engagement levels.

The summary count from each document, from the participants belonging to experimental group were tracked to correlate with the subjective responses of the engagement levels and it could be observed from the Figure 11 that the highest average summary count is observed at the lowest engagement level (1), indicating that participants who reported low engagement accessed summaries more frequently which aligns with the system's design to display summaries as a support mechanism during disengagement. This suggests that the system effectively provides summaries as a support mechanism when engagement is low. As engagement levels increase from 1 to 7, the average summary count decreases implying that participants who are more engaged (higher ratings) tend to rely less on summaries, possibly because they are better able to comprehend the material without additional support. At the highest engagement levels (6 and 7), the average summary count is significantly lower, suggesting that participants are confident in their understanding and do not require summaries. The plot illustrates a clear inverse relationship between engagement levels and the frequency of summary access, showing the adaptive nature of the system in providing support based on real-time engagement metrics. This finding highlights the importance of monitoring engagement to tailor educational interventions effectively to enhance learning outcomes.

Our system achieved 68.15% accuracy and a 65.82% F1-score for engagement prediction using gaze data, which is competitive with recent approaches in the field. Chen et al. [54] integrated gaze directions and facial expressions in a multi-modal deep neural network (MDNN) for predicting student engagement, achieving high accuracy in collaborative learning settings. Similarly, Sharma et al. [55] devised a method combining eye and head movements with facial emotional cues, creating an engagement index that effectively categorized students into different engagement levels. While Gupta et al. [84] achieved higher accuracies using deep learning architectures like ResNet-50 for facial emotion analysis, these methods often require complex setups and are limited by strict front-facing requirements. Our system addresses these limitations by relying solely on gaze data, achieving 68.15% accuracy while maintaining effectiveness regardless of head position or orientation, which is particularly valuable in natural learning environments.

Our approach follows a comprehensive validation along with real-time feedback mechanism. Unlike previous studies that primarily focused on model accuracy metrics, our system implements a dual validation approach combining technical validation through model performance with user-centric validation, where 76% of participants agreed with the system's predictions. Our system uses real-time visualization via an interactive dashboard, providing continuous engagement monitoring at 5-second intervals, compared to longer intervals or post-hoc analysis in previous studies. This real-time feedback mechanism, combined with the dual validation approach and automated interventions, creates a feedback loop where learners can verify and benefit from the system's predictions immediately. The high user agreement rate with our system's predictions, despite slightly lower technical accuracy compared to multimodal approaches, suggests that real-time feedback and immediate interventions may be more valuable in practical learning scenarios than marginally higher prediction accuracies achieved through more complex, multi-modal systems.

B. IMPACT OF THE SYSTEM IN LEARNING OUTCOMES

The system is designed to display summaries only when real-time engagement levels are detected as low, effectively providing targeted support to learners. The aim of the system is to reengage the participants in case of disengagement detected and to provide tailored summaries as interventions to improve the engagement or attention and also the comprehension levels. While previous research on adaptive e-learning environments [9] showed significant improvements in student engagement through learning style-based personalization, our system takes a different approach by implementing real-time, engagement-based interventions through automated summary generation. The summaries were generated using a ChatGPT based API which takes the text as prompt and provide the summary of that particular text.

The survey responses aimed to collect feedback from learners regarding their comprehension, engagement, and self-confidence levels for both the experimental and control groups. The experimental group, which utilized the adaptive learning system, reported higher levels of engagement and confidence compared to the control group, along with some improvement in subjective comprehension. Statistical analysis through t-tests confirmed these findings, indicating significantly higher engagement (t-statistic = 3.12, p-value = 0.0021) and confidence (t-statistic = 8.22, p-value < 0.0001) levels in the experimental group, as well as a notable improvement in comprehension (t-statistic = 3.17, p-value = 0.0018). These findings align with Liu et al. [85], who found significant positive correlations between learning achievement and engagement, as well as between engagement and learning attitude. Their study demonstrated that adaptive feedback systems not only improved engagement but also showed that students with greater positive engagement demonstrated enhanced self-directed learning capabilities in distance learning activities. However, the variability in responses suggests that while the adaptive system was beneficial for many participants, its effectiveness was not uniform across all learner. Some learners found the summaries particularly helpful, while others did not perceive the same level of benefit. The objective comprehension quiz results revealed that although the experimental group had a higher number of participants achieving perfect scores (4/4), there were also participants who scored lower (1/4). This variability aligns with previous findings [85] suggesting that effective learning outcomes are achieved when students maintain positive engagement and attitudes throughout the learning process, supporting our observation that the summary generation system's effectiveness varied based on individual engagement levels. This indicates that the summary generation may not have been equally effective for all individuals, suggesting a need for further refinement of the system to enhance its overall impact on learning outcomes. The mixed results highlight the importance of tailoring educational interventions to meet diverse learner needs, ensuring that all students can benefit from adaptive learning technologies.

The gaze metrics analysis was performed to get the quantitative assessment of learners gaze behavior in both the experimental and control groups. The participants in the experimental group often relied on summaries provided as interventions when their engagement levels were low, which indicates that the summaries served to redirect their attention. The pupil diameter and fixation durations in the experimental group were notably higher at lower engagement and comprehension levels compared to the control group. This increase suggests that participants in the experimental group experienced enhanced cognitive processing when interacting with the generated summaries. The larger pupil diameter is often associated with heightened cognitive load and arousal, indicating that participants were actively engaging with the material to improve their understanding. When engagement levels are low, the introduction of summaries appears to allow participants to focus more on the content, allowing them to extract relevant information effectively. The increased fixation duration reflects a higher engagement with the summaries, as participants invest more time processing the information to improve their comprehension. This cognitive effort, driven by the summaries, serves as a valuable intervention that helps re-engage learners and supports their understanding of the material.

Figure 12 depicts the variation in mean pupil diameter across a learning material for participant belonging to control (*P02*) and experimental group (*P18*). In comparing the data for participants *P18* and *P02*, there are notable differences that highlights the potential advantages of the interventions provided to participant *P18*. Participant *P18* consistently exhibits a higher mean pupil diameter compared to *P02*, suggesting greater cognitive load or sustained attention during tasks which could be due to the interventions provided. This physiological indicator aligns with *P18*'s better performance in both objective comprehension and engagement ratings. The participant *P18* achieved higher objective comprehension scores, mostly 4s, and maintains high engagement levels, with scores of 6 or 7 across all documents. In contrast, participant *P02* shows a lower mean pupil diameter across all the documents compared to *P18*. Participant *P02*'s objective comprehension scores vary significantly, with several scores as low as '1' or '2', and engagement levels fluctuate, indicating inconsistent involvement. These differences highlight the potential benefits of the interventions received by *P18*, which may contribute to enhanced focus, understanding, and engagement.

Although the differences in gaze metrics between the experimental and control groups were not statistically significant, the observed trends still provide valuable insights into participant behavior. The gaze metrics, including fixation duration and pupil diameter, indicated that participants in the experimental group engaged with the generated summaries,

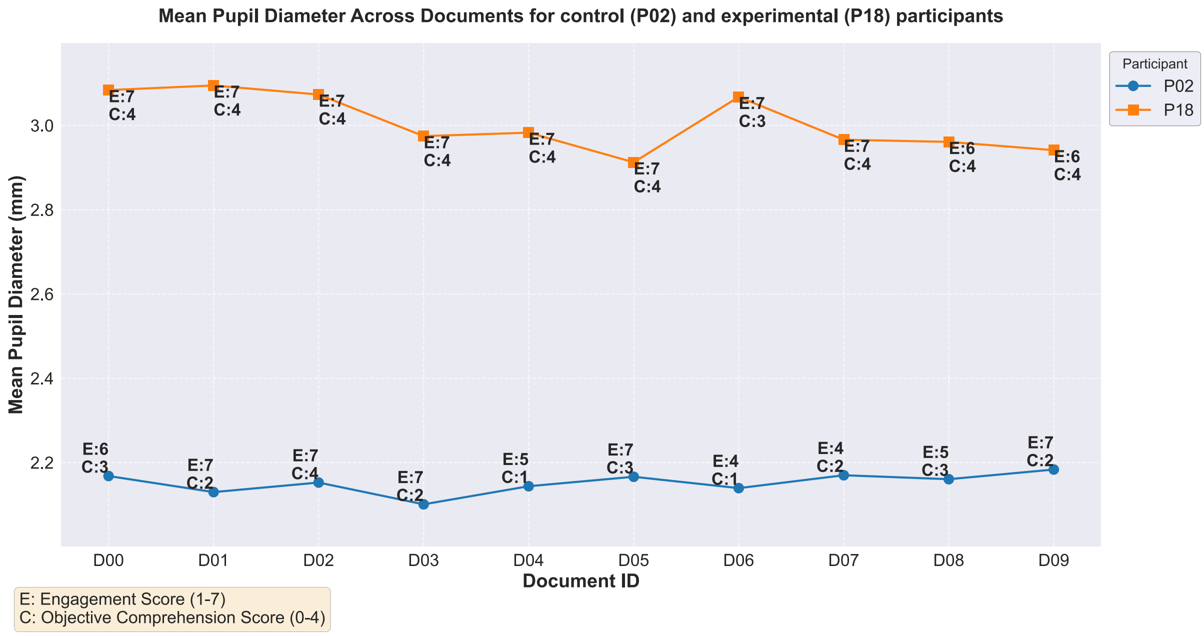
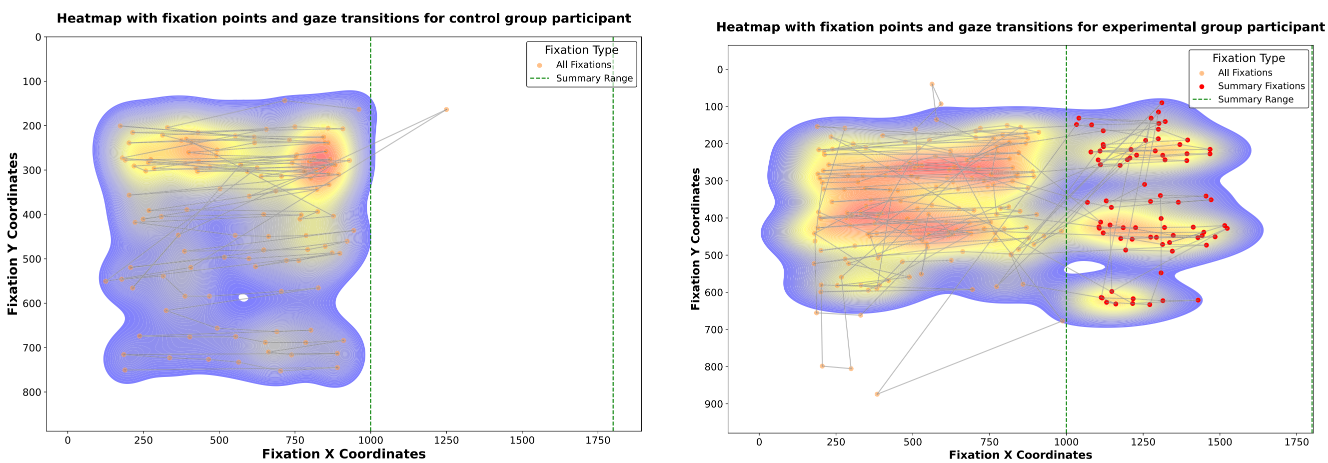


FIGURE 12. The variation in pupil diameter across control(P02) and experimental(P18) participants over the learning material with the reported engagement and computed objective comprehension score.



(a) Control Group Participant

(b) Experimental Group Participant

FIGURE 13. The heatmap with fixation points and gaze transitions for control group participant and experimental group participant reading the same learning material and with low reported engagement levels.

suggesting a potential for increased cognitive processing even if the differences did not reach conventional levels of significance. This lack of statistical significance may reflect the variability in individual responses to the summaries, highlighting that while the adaptive learning system may not have uniformly impacted gaze behavior across all participants, it still holds promise as a tool for enhancing engagement and comprehension. Overall, these findings highlight the beneficial role of adaptive learning interventions in fostering deeper cognitive engagement and improving learning outcomes, even in challenging contexts.

C. VISUAL EFFORT AND COGNITIVE OUTCOMES

To explore the the relationship between cognitive processing, visual effort, and adaptive learning interventions, it is vital to identify the key metrics that capture the nuances of cognitive and visual engagement, such as fixations, saccade patterns, and gaze transitions each offering unique insights into the reader's cognitive processes and visual strategies. The research by Skaramagkas et.al [86] has established that eye-tracking metrics such as fixations, saccades, and pupil diameter serve as reliable biosignals for understanding visual attention and cognitive workload. These metrics, along with gaze transitions, each offer unique insights into the reader's cognitive processes and visual strategies.

Our study design incorporates both control and experimental groups, allowing for comparative analysis of reading behaviors and task-solving approaches. While similar control-experimental designs have been employed in recent studies, our methodology differs in several key aspects. Heyd-Metzuyanim et al. [87] combined eye-tracking with discourse analysis to identify problem-solving ineffectiveness in geometric tasks, and da Silva Soares et al. [88] investigated physiological aspects of cognitive effort during mental rotation tests in naturalistic educational environments. Our approach extends beyond these studies by implementing real-time engagement prediction and providing immediate adaptive interventions based on gaze patterns, rather than conducting post-hoc analysis of problem-solving strategies. In our experiment, participants in the control group engage with the content using their natural reading and problem-solving strategies, providing a baseline for typical cognitive and visual behaviors, while the experimental group receives tailored interventions designed to optimize their learning experience.

The adaptive interventions, exclusively implemented for the experimental group, are expected to alter the participants' interaction with the learning material. As a result, we anticipate observing distinct differences in gaze transitions, fixation durations, and overall visual exploration patterns between the two groups. The control group's eye movements are likely to reflect more traditional reading patterns, while the experimental group may exhibit more targeted and efficient visual strategies influenced by the adaptive interventions.

Figure 13 illustrates the heatmap with fixation points and gaze transitions for control group participant and experimental group participant reading the same learning material and with low reported engagement levels in the post-survey. The presence of summary fixations (red dots) within the green dashed lines for the experimental group, and their absence in the control group, directly reflects the adaptive intervention strategy. For the experimental participant, this adaptive feature offers targeted support, providing a concise overview of key information when engagement drops, personalized learning pace, allowing participants to quickly catch up for better content understanding and engagement recovery. The summary fixations suggest participants are utilizing this additional resource for potentially more efficient comprehension and when their engagement level drops. In contrast, the control group's gaze patterns likely reflect a more consistent, but potentially less responsive, interaction with the content. This difference underscores the potential of adaptive interventions to provide timely, targeted support based on real-time engagement levels.

The visual effort and cognitive processing patterns differ notably between the control and experimental group. The control group participant (Figure13) exhibits more concentrated fixations and shorter gaze transitions, suggesting a potentially lower visual effort but also a more linear, traditional reading approach. This pattern may indicate a consistent cognitive load throughout the task. In contrast, experimental group participant (Figure13a) demonstrate a wider distribution of fixations and longer gaze transitions due to the interventions, pointing to potentially higher visual effort as they explore more of the content. This increased visual exploration likely corresponds to more diverse cognitive strategies and varying levels of cognitive load. The adaptive interventions provided to the experimental group appear to encourage a more dynamic interaction with the content, potentially leading to deeper cognitive engagement. While the experimental group may exert more visual effort, this increased activity could be indicative of more thorough information processing and potentially more effective learning. The relationship between visual effort and cognitive processing in this context suggests that the adaptive interventions may be promoting a more active and comprehensive approach to learning, although with increased visual demands.

D. POTENTIAL IMPACT ON EDUCATIONAL PRACTICE

The findings from this study have significant potential implications for educational practice across various learning environments. Our gaze-based adaptive learning system aligns with recent developments in learning analytics, where tools are developed to provide practical insights for learners [50]. Similar to how VizChat addresses cognitive overload concerns through AI-generated explanations [89], our system provides contextually relevant interventions based on real-time gaze data. While traditional adaptive learning systems focus on learning styles and content delivery [9], our system uniquely implements engagement detection and immediate interventions. For instructors, this technology offers a powerful tool to identify when students are struggling with content, enabling timely adjustments in teaching pace and methods.

The combination of eye-tracking technology and AI-generated summaries enables scalable, personalized learning support, particularly valuable in large classes where individual teacher attention is limited. The system's ability to deliver customized support addresses diverse learning needs, from providing additional help to offering more challenging content. This approach shows particular promise in online courses and self-paced programs, where maintaining student engagement is crucial for successful learning outcomes. The collected gaze patterns and engagement data enable curriculum refinement by helping educators identify challenging concepts that consistently cause disengagement. Teachers can utilize the system's analytics to develop more effective teaching strategies and identify areas for professional growth, while the real-time detection of engagement patterns provides an early warning system for students at risk of falling behind. The system's ability to track engagement and comprehension complements traditional assessment methods, providing insights into learning processes rather than just outcomes. The system adapts to different learning needs, helping students who may have attention challenges or who learn at different speed.

E. LIMITATIONS AND FUTURE WORK

While the study on adaptive learning system using eye-tracking technology offers valuable insights, several limitations were identified, each presenting opportunities for future research directions.

1) SAMPLE SIZE AND REPRESENTATION

One primary limitation of this study is the limited sample size (N=22). This constraint could affect the generalizability of our results to broader populations. The challenge lies in ensuring a sufficiently large and diverse sample that accurately represents various learning styles, backgrounds, and cognitive abilities.

2) CONTENT AND REAL-TIME ENGAGEMENT DETECTION

The specific content or tasks used in the study might not be equally engaging or challenging for all participants, potentially skewing the results. The task of creating uniformly suitable content across different skill levels and interests poses a considerable challenge. The accuracy and responsiveness of the system in detecting low engagement in real-time may be limited due to the utilization of just gaze data but could have better accuracy by integrating multimodal data including facial cues and physiological data. To develop algorithms that can reliably interpret eye movements as indicators of engagement across different individuals remains a complex task.

3) ADAPTIVE INTERVENTION STRATEGY AND LONG-TERM EFFECTS

The effectiveness of summaries as the only form of intervention for low engagement may be limited. The design of interventions that are universally effective across different learning styles and preferences is challenging. Moreover, the study mainly focus on immediate effects without considering long-term learning outcomes. The challenge lies in designing longitudinal studies that can assess the lasting effects of these interventions on knowledge retention and application, while controlling for other variables that may influence learning over time.

4) LLM BASED SUMMARIES

The use of ChatGPT for generating summaries presents its own set of limitations. While AI-generated summaries can be efficient, they may lack the nuanced understanding of complex topics that a human expert would provide. There's a risk of oversimplification or occasional inaccuracies in the summaries. The AI's output is based on its training data, which may not always include the most up-to-date information or specialized knowledge relevant to the study content. The challenge lies in ensuring the quality, accuracy, and relevance of these AI-generated summaries, possibly through a system of expert review or by combining AI-generated content with human-curated information.

5) INDIVIDUAL DIFFERENCES AND PARTICIPANT RESPONSES

A significant limitation of our study is the potential oversight of individual differences in cognitive processing speeds and styles. The participants may vary greatly in how they process information, which could impact the effectiveness of the adaptive interventions. The challenge is to develop a system flexible enough to adapt to a wide range of individual cognitive characteristics while still providing meaningful and comparable data across the study population.

Future work should address several key areas to enhance the robustness and applicability of this adaptive learning system. The main focus would be on expanding the sample size and diversity of participants to improve the generalizability of findings across various learning styles and backgrounds. The refinement of the real-time engagement detection algorithms by incorporating multimodal data, including facial cues and physiological signals alongside gaze data, could significantly improve the accuracy and responsiveness of the system. Additionally, implementing multimodal content adaptation based on learner engagement levels including interactive visualizations, audio explanations, and video demonstrations would provide more engaging and effective learning experiences. The development of personalized learning paths that automatically adapt based on learner progress and preferences would ensure optimal learning outcomes.

Furthermore, future studies should explore a wider range of adaptive intervention strategies beyond text summaries, tailoring them to individual learning preferences and styles. Conducting longitudinal studies will be essential to assess the long-term effects of these interventions on knowledge retention and application. Improving the quality and relevance of AI-generated summaries through expert review or hybrid AI-human curation systems will be crucial for ensuring the effectiveness of the interventions. These improvements will help create a better learning system that adapts in real-time to learner needs, while ensuring high-quality educational content that works for different types of learners.

VII. CONCLUSION

This study demonstrates the effectiveness of integrating real-time gaze tracking with AI-driven adaptive learning interventions to enhance student engagement and learning outcomes. Through an experimental study with 22 university students, we validated that AI-generated adaptive interventions, triggered by detected low engagement, led to significantly improved learning outcomes and engagement levels. As additional validation of our approach, we implemented two deep learning models - an InceptionTime and a Transformer-based network - to predict post-reading engagement and comprehension levels from gaze patterns. The Transformer model demonstrated superior performance in predicting both engagement and comprehension levels based on user survey responses. The effectiveness of our adaptive system was further evidenced by significantly higher objective comprehension scores in the experimental group, who showed improved comprehension assessments compared to those who did not receive adaptive interventions. Our research makes three key contributions to educational technology: (1) demonstrating the feasibility of using gaze data for real-time engagement prediction, (2) validating the effectiveness of AI-generated adaptive interventions in improving learning outcomes, and (3) a framework for integrating eye-tracking technology with adaptive learning systems. These findings advance the field by providing empirical evidence for the benefits of real-time adaptive interventions and offering a practical approach to implementing personalized learning experiences. The potential of this system opens new possibilities for developing more responsive and effective educational technologies that can dynamically adapt to individual learner needs.

REFERENCES

[1] E. F. Barkley and C. H. Major, *Student Engagement Techniques: A Handbook for College Faculty*. Hoboken, NJ, USA: Wiley, 2020.

[2] D. J. Shernoff, *Optimal Learning Environments to Promote Student Engagement*. New York, NY, USA: Springer, 2013.

[3] J. Parsons and L. Taylor, ''Improving student engagement,'' *Current Issues Educ.*, vol. 14, no. 1, pp. 1–33, 2011.

[4] S. Harvey and A. Goudvis, *Strategies That Work: Teaching Comprehension for Understanding and Engagement*. Portland, ME, USA: Stenhouse, 2007.

[5] D. Brassell and T. Rasinski, *Comprehension That Works: Taking Students Beyond Ordinary Understanding To Deep Comprehension*. Huntington Beach, CA, USA: Teacher Created Materials, 2008.

[6] Kinshuk, N.-S. Chen, I.-L. Cheng, and S. W. Chew, ''Evolution is not enough: Revolutionizing current learning environments to smart learning environments,'' *Int. J. Artif. Intell. Educ.*, vol. 26, no. 2, pp. 561–581, Jun. 2016.

[7] R. G. Aghaei, A. Arya, and R. Biddle, ''A dashboard for affective e-learning: Data visualization for monitoring online learner emotions,'' in *Proc. EdMedia+ Innovate Learn.*, 2016, pp. 1536–1543.

[8] M. Liu and D. Yu, ''Towards intelligent e-learning systems,'' *Educ. Inf. Technol.*, vol. 28, no. 7, pp. 7845–7876, 2023.

[9] H. A. El-Sabagh, ''Adaptive e-learning environment based on learning styles and its impact on development students' engagement,'' *Int. J. Educ. Technol. Higher Educ.*, vol. 18, no. 1, p. 53, Dec. 2021.

[10] Y. Deng, Z. Ren, A. Zhang, W. Lei, and T.-S. Chua, ''Towards goal-oriented intelligent tutoring systems in online education,'' 2023, *arXiv:2312.10053*.

[11] E. Yang, J. Pan, X. Wang, H. Yu, L. Shen, X. Chen, L. Xiao, J. Jiang, and G. Guo, ''AdaTask: A task-aware adaptive learning rate approach to multi-task learning,'' in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 10745–10753.

[12] F. M. Fahid, J. Rowe, Y. Kim, S. Srivastava, and J. Lester, ''Online reinforcement learning-based pedagogical planning for narrative-centered learning environments,'' in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, 2024, pp. 23191–23199.

[13] V. Shute and B. Towle, ''Adaptive e-learning,'' in *Aptitude*. Evanston, IL, USA: Routledge, 2018, pp. 105–114.

[14] D. Kelly and B. Tangney, ''Adapting to intelligence profile in an adaptive educational system,'' *Interacting Comput.*, vol. 18, no. 3, pp. 385–409, May 2006.

[15] H. Hills, *Individual Preferences in E-Learning*. Evanston, IL, USA: Routledge, 2017.

[16] N. B. A. Normadhi, L. Shuib, H. N. M. Nasir, A. Bimba, N. Idris, and V. Balakrishnan, ''Identification of personal traits in adaptive learning environment: Systematic literature review,'' *Comput. Educ.*, vol. 130, pp. 168–190, Mar. 2019.

[17] X. Wu, H. Wang, Y. Zhang, B. Zou, and H. Hong, ''A tutorial-generating method for autonomous online learning,'' *IEEE Trans. Learn. Technol.*, vol. 17, pp. 1558–1567, 2024.

[18] Z. A. Pardos, M. Tang, I. Anastasopoulos, S. K. Sheel, and E. Zhang, ''OATutor: An open-source adaptive tutoring system and curated content library for learning sciences research,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2023, pp. 1–17.

[19] A. Almusaed, A. Almssad, I. Yitmen, and R. Z. Homod, ''Enhancing student engagement: Harnessing,'' *Educ. Sci.*, vol. 13, no. 7, p. 632, 2023.

[20] E. Bonner, K. Garvey, M. Miner, S. Godin, and H. Reinders, ''Measuring real-time learner engagement in the Japanese EFL classroom,'' *Innov. Lang. Learn. Teaching*, vol. 17, no. 2, pp. 254–264, Mar. 2023.

[21] C. R. Henrie, L. R. Halverson, and C. R. Graham, ''Measuring student engagement in technology-mediated learning: A review,'' *Comput. Educ.*, vol. 90, pp. 36–53, Dec. 2015.

[22] K. Watanabe, T. Sathyanarayana, A. Dengel, and S. Ishimaru, ''EnGauge: Engagement gauge of meeting participants estimated by facial expression and deep neural network,'' *IEEE Access*, vol. 11, pp. 52886–52898, 2023.

[23] S. D'Mello and A. Graesser, ''Dynamics of affective states during complex learning,'' *Learn. Instruct.*, vol. 22, no. 2, pp. 145–157, Apr. 2012.

[24] N. Bergdahl, ''Engagement and disengagement in online learning,'' *Comput. Educ.*, vol. 188, Oct. 2022, Art. no. 104561.

[25] M. T. Alshammari and A. Qtaish, ''Effective adaptive E-Learning systems according to learning style and knowledge level,'' *J. Inf. Technol. Educ., Res.*, vol. 18, pp. 529–547, Aug. 2019.

[26] S. Jacob, S. Ishimaru, S. S. Bukhari, and A. Dengel, ''Gaze-based interest detection on newspaper articles,'' in *Proc. 7th Workshop Pervasive Eye Tracking Mobile Eye-Based Interact.*, Jun. 2018, pp. 1–7.

[27] J. Santhosh, D. Dzsotjan, and S. Ishimaru, ''Multimodal assessment of interest levels in reading: Integrating eye-tracking and physiological sensing,'' *IEEE Access*, vol. 11, pp. 93994–94008, 2023.

[28] A. P. Pai, J. Santhosh, and S. Ishimaru, ''Real-time feedback on Reader's engagement and emotion estimated by eye-tracking and physiological sensing,'' in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2022, pp. 97–98.

[29] J. Santhosh, A. P. Pai, and S. Ishimaru, ''Toward an interactive reading experience: Deep learning insights and visual narratives of engagement and emotion,'' *IEEE Access*, vol. 12, pp. 6001–6016, 2024.

[30] N. Bosch, ''Detecting student engagement: Human versus machine,'' in *Proc. Conf. User Model. Adaptation Personalization*, Jul. 2016, pp. 317–320.

[31] C. Thomas and D. B. Jayagopi, ''Predicting student engagement in classrooms using facial behavioral cues,'' in *Proc. 1st ACM SIGCHI Int. Workshop Multimodal Interact. Educ.*, Nov. 2017, pp. 33–40.

[32] K. Joshi, R. Kumar, S. Bharany, D. K. J. B. Saini, R. Kumar, A. O. Ibrahim, A. Abdelmaboud, W. Nagmeldin, and M. A. Medani, ''Exploring the connectivity between education 4.0 and classroom 4.0: Technologies, student perspectives, and engagement in the digital era,'' *IEEE Access*, vol. 12, pp. 24179–24204, 2024.

[33] T. T. Anannya, S. R. Mahmud, S. R. Khan, S. R. Zaman, M. S. Jalal, and S. Sharmin, ''From blackboards to byteboards: Adapting to the digital age in undergraduate education on campus,'' in *Proc. 17th Int. Conf. Signal-Image Technol. & Internet-Based Syst. (SITIS)*, Nov. 2023, pp. 54–61.

[34] A. Abedi and S. S. Khan, ''Improving state-of-the-art in detecting student engagement with resnet and TCN hybrid network,'' in *Proc. 18th Conf. Robots Vis. (CRV)*, May 2021, pp. 151–157.

[35] K. Ahuja, D. Kim, F. Xhakaj, V. Varga, A. Xie, S. Zhang, J. E. Townsend, C. Harrison, A. Ogan, and Y. Agarwal, ''EduSense: Practical classroom sensing at scale,'' *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, Sep. 2019.

[36] L. R. Harris, ''A phenomenographic investigation of teacher conceptions of student engagement in learning,'' *Austral. Educ. Researcher*, vol. 35, no. 1, pp. 57–79, Apr. 2008.

[37] K. Benabbes, K. Housni, B. Hmedna, A. Zellou, and A. E. Mezouary, ''A new hybrid approach to detect and track learner's engagement in e-Learning,'' *IEEE Access*, vol. 11, pp. 70912–70929, 2023.

[38] S. Ishimaru, S. Jacob, A. Roy, S. S. Bukhari, C. Heisel, N. Großmann, M. Thees, J. Kuhn, and A. Dengel, ''Cognitive state measurement on learning materials by utilizing eye tracker and thermal camera,'' in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 8, Nov. 2017, pp. 32–36.

[39] L. Wang, "Attention decrease detection based on video analysis in e-learning," in *Transactions on Edutainment XIV*. Berlin, Germany: Springer, 2018, pp. 166–179.
[40] D. Bikowski and E. Casal, *Interactive Digital Textbooks and Engagement: A Learning Strategies Framework*. Honolulu, HI, USA: University of Hawaii National Foreign Language Resource Center, 2018.
[41] A. K. Hashim and V. W. Vongkulluksn, "E-reader apps and reading engagement: A descriptive case study," *Comput. Educ.*, vol. 125, pp. 358–375, Oct. 2018.
[42] K. F. Hew, W. Huang, J. Du, and C. Jia, "Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions," *J. Comput. Higher Educ.*, vol. 35, no. 1, pp. 40–68, Apr. 2023.
[43] J. Krasodomska and J. Godawska, "E-learning in accounting education: The influence of students' characteristics on their engagement and performance," *Accounting Educ.*, vol. 30, no. 1, pp. 22–41, Jan. 2021.
[44] H. Keser and A. Semerci, "Technology trends, education 4.0 and beyond," *Contemp. Educ. Researches J.*, vol. 9, no. 3, pp. 39–49, Aug. 2019.
[45] M. Imran and N. Almusharraf, "Digital learning demand and applicability of quality 4.0 for future education: A systematic review," *Int. J. Eng. Pedagogy (iJEP)*, vol. 14, no. 4, pp. 38–53, May 2024.
[46] S. Akshay, P. K. Bijith, S. Sanjana, and J. Amudha, "Ibehave: Behaviour analysis using eye gaze metrices," in *Proc. Int. Conf. Pattern Recognit. Mach. Intell.* Cham, Switzerland: Springer, 2023, pp. 260–269.
[47] R. Zhu, L. Shi, Y. Song, and Z. Cai, "Integrating gaze and mouse via joint cross-attention fusion net for students' activity recognition in E-learning," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 3, pp. 1–35, Sep. 2023.
[48] G. Lan, T. Scargill, and M. Gorlatova, "EyeSyn: Psychology-inspired eye movement synthesis for gaze-based activity recognition," in *Proc. 21st ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, May 2022, pp. 233–246.
[49] S. Ghosh, A. Dhall, M. Hayat, J. Knibbe, and Q. Ji, "Automatic gaze analysis: A survey of deep learning based approaches," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 1, pp. 61–84, Jan. 2024.
[50] Y. Wang, S. Lu, and D. Harter, "Multi-sensor eye-tracking systems and tools for capturing Student attention and understanding engagement in learning: A review," *IEEE Sensors J.*, vol. 21, no. 20, pp. 22402–22413, Oct. 2021.
[51] P. Vadiraja, J. Santhosh, H. Moulay, A. Dengel, and S. Ishimaru, "Effects of counting seconds in the mind while reading," in *Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, Sep. 2021, pp. 486–490.
[52] J. Bidwell and H. Fuchs, "Classroom analytics: Measuring student engagement with automated gaze tracking," *Behav. Res. Methods*, vol. 49, no. 113, pp. 1–17, 2011.
[53] B. D. Carolis, F. D'Errico, N. Macchiarulo, and G. Palestra, "'Engaged faces': Measuring and monitoring student engagement from face and gaze behavior," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.-Companion Vol.*, Jun. 2019, pp. 80–85.
[54] Y. Chen, J. Zhou, Q. Gao, J. Gao, and W. Zhang, "MDNN: Predicting student engagement via gaze direction and facial expression in collaborative learning," *Comput. Model. Eng. Sci.*, vol. 136, no. 1, pp. 381–401, 2023.
[55] P. Sharma, S. Joshi, S. Gautam, S. Maharjan, S. R. Khanal, M. C. Reis, J. Barroso, and V. M. de Jesus Filipe, "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning," in *Proc. Int. Conf. Technol. Innov. Learn., Teach. Educ.* Cham, Switzerland: Springer, 2022, pp. 52–68.
[56] A. M. Mathew, A. A. Khan, T. Khalid, and R. Souissi, "GESCAM: A dataset and method on gaze estimation for classroom attention measurement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 636–645.
[57] S. Amin, M. I. Uddin, A. A. Alarood, W. K. Mashwani, A. Alzahrani, and A. O. Alzahrani, "Smart e-learning framework for personalized adaptive learning and sequential path recommendations using reinforcement learning," *IEEE Access*, vol. 11, pp. 89769–89790, 2023.
[58] F. Z. Lhafra and O. Abdoun, "Towards an adaptive learning process using artificial intelligence technologies," in *Proc. Int. Conf. Digit. Technol. Appl.* Cham, Switzerland: Springer, 2023, pp. 23–32.
[59] I. Gligorea, M. Cioca, R. Oancea, A.-T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: A literature review," *Educ. Sci.*, vol. 13, no. 12, p. 1216, Dec. 2023.
[60] W. Kaiss, K. Mansouri, and F. Poirier, "Effectiveness of an adaptive learning chatbot on students' learning outcomes based on learning styles," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 18, no. 13, pp. 250–261, Jul. 2023.
[61] S. Dutta, S. Ranjan, S. Mishra, V. Sharma, P. Hewage, and C. Iwendi, "Enhancing educational adaptability: A review and analysis of AI-driven adaptive learning platforms," in *Proc. 4th Int. Conf. Innov. Practices Technol. Manage. (ICIPTM)*, Feb. 2024, pp. 1–5.
[62] A. Ezzaim, A. Dahbi, A. Haidine, and A. Aqqal, "Enhancing academic outcomes through an adaptive learning framework utilizing a novel machine learning-based performance prediction method," *Data Metadata*, vol. 2, p. 164, Dec. 2023.
[63] T. Hussain, L. Yu, M. Asim, A. Ahmed, and M. A. Wani, "Enhancing e-learning adaptability with automated learning style identification and sentiment analysis: A hybrid deep learning approach for smart education," *Information*, vol. 15, no. 5, p. 277, May 2024.
[64] W. S. Sayed, A. M. Noeman, A. Abdellatif, M. Abdelrazek, M. G. Badawy, A. Hamed, and S. El-Tantawy, "AI-based adaptive personalized content presentation and exercises navigation for an effective and engaging e-learning platform," *Multimedia Tools Appl.*, vol. 82, no. 3, pp. 3303–3333, Jan. 2023.
[65] H. Wei, A.-N. Moldovan, and C. H. Muntean, "Sensing learner interest through eye tracking," in *Proc. 9th IT & T Conf.* Dublin, Ireland: Dublin Institute of Technology, 2009, pp. 1–8.
[66] C. M. Tyng, H. U. Amin, M. N. M. Saad, and A. S. Malik, "The influences of emotion on learning and memory," *Frontiers Psychol.*, vol. 8, Aug. 2017, Art. no. 235933.
[67] T. S. Moghadam, A. Darejeh, M. Delaramifar, and S. Mashayekh, "Toward an artificial intelligence-based decision framework for developing adaptive e-learning systems to impact learners' emotions," *Interact. Learn. Environ.*, vol. 2023, pp. 1–21, Mar. 2023.
[68] A. Omar, K. Hassane, and K. Mohamed, "Towards a new generation of intelligent, adaptive e-learning platforms," *Int. J. Comput. Digit. Syst.*, vol. 17, no. 1, pp. 1–10, Jan. 2024.
[69] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review," *IEEE Access*, vol. 11, pp. 48392–48409, 2023.
[70] W. Xin Zhao et al., "A survey of large language models," 2023, *arXiv:2303.18223*.
[71] M. Perkins, "Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond," *J. Univ. Teach. Learn. Pract.*, vol. 20, no. 2, pp. 1–26, Jan. 2023.
[72] J. Jeon and S. Lee, "Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT," *Educ. Inf. Technol.*, vol. 28, no. 12, pp. 15873–15892, Dec. 2023.
[73] G.-A. Morosanu, L.-A. Rata, and M. Geru, "Outcomes of large language models and artificial intelligence in education," *Didactica Danubiensis*, vol. 3, no. 1, pp. 30–52, 2023.
[74] H. Gabbay and A. Cohen, "Combining LLM-generated and test-based feedback in a MOOC for programming," in *Proc. 11th ACM Conf. Learn. Scale*, Jul. 2024, pp. 177–187.
[75] I. Azaiz, N. Kiesler, and S. Strickroth, "Feedback-generation for programming exercises with GPT-4," 2024, *arXiv:2403.04449*.
[76] J. Meyer, T. Jansen, R. Schiller, L. W. Liebenow, M. Steinbach, A. Horbach, and J. Fleckenstein, "Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions," *Comput. Educ., Artif. Intell.*, vol. 6, Jun. 2024, Art. no. 100199.
[77] A. Xu, W. Monroe, and K. Bicknell, "Large language model augmented exercise retrieval for personalized language learning," in *Proc. 14th Learn. Anal. Knowl. Conf.*, Mar. 2024, pp. 284–294.
[78] A. Garg, A. Raina, A. Gupta, J. Singh, M. Saini, P. Iiitd, R. Mehta, R. Oberoi, S. Sharma, S. Jain, S. Tyagi, U. Arora, and D. Kumar, "Analyzing LLM usage in an advanced computing class in India," 2024, *arXiv:2404.04603*.
[79] H. Abu-Rasheed, C. Weber, and M. Fathi, "Knowledge graphs as context sources for LLM-based explanations of learning recommendations," 2024, *arXiv:2403.03008*.
[80] M. Farrokhnia, S. K. Banihashem, O. Noroozi, and A. Wals, "A SWOT analysis of ChatGPT: Implications for educational practice and research," *Innov. Educ. Teach. Int.*, vol. 61, no. 3, pp. 460–474, May 2024.

[81] E. Kasneci et al., "Chatgpt for good? On opportunities and challenges of large language models for education," *Learn. Individual Differences*, vol. 103, Apr. 2023, Art. no. 102274.

[82] Y. K. Dwivedi et al., "Opinion paper: 'So what if chatgpt wrote it?' multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, Aug. 2023, Art. no. 102642.

[83] D. Gašević, G. Siemens, and S. Sadiq, "Empowering learners for the age of artificial intelligence," *Comput. Educ., Artif. Intell.*, vol. 4, Aug. 2023, Art. no. 100130.

[84] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," *Multimedia Tools Appl.*, vol. 82, no. 8, pp. 11365–11394, Mar. 2023.

[85] H.-L. Liu, T.-H. Wang, H.-C.-K. Lin, C.-F. Lai, and Y.-M. Huang, "The influence of affective feedback adaptive learning system on learning engagement and self-directed learning," *Frontiers Psychol.*, vol. 13, Apr. 2022, Art. no. 858411.

[86] V. Skaramagkas, G. Giannakakis, E. Ktistakis, D. Manousos, I. Karatzanis, N. S. Tachos, E. Tripoliti, K. Marias, D. I. Fotiadis, and M. Tsiknakis, "Review of eye tracking metrics involved in emotional and cognitive processes," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 260–277, 2023.

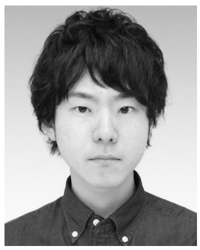
[87] E. Heyd-Metzuyanim, E. S. H. Haataja, M. S. Hannula, and E. G. Moreno-Esteva, "What can eye-tracking, combined with discourse analysis, teach us about the ineffectiveness of a group of students solving a geometric problem?" *Instructional Sci.*, vol. 51, no. 3, pp. 363–396, 2023.

[88] R. da Silva Soares, C. Barreto, and J. R. Sato, "Perspectives in eye-tracking technology for applications in education," *South Afr. J. Childhood Educ.*, vol. 13, no. 1, p. 1204, May 2023.

[89] L. Yan, L. Zhao, V. Echeverria, Y. Jin, R. Alfredo, X. Li, D. Gaševi'c, and R. Martinez-Maldonado, "Vizchat: Enhancing learning analytics dashboards with contextualised explanations using multimodal generative AI chatbots," in *Proc. Int. Conf. Artif. Intell. Educ.* Cham, Swizerland: Springer, 2024, pp. 180–193.

JAYASANKAR SANTHOSH (Member, IEEE) was born in Kerala, India, in 1992. He received the bachelor's degree in computer science from Mahatma Gandhi University, India, in 2014, and the master's degree in computer science from the Technical University of Kaiserslautern, Germany, in 2018. He has been a Ph.D. Researcher with German Research Center for Artificial Intelligence (DFKI), since 2019. He is currently a Working Member of the Immersive Quantified Learning Laboratory (IQL Laboratory) and Smart Data and Knowledge Services (SDS) Department, DFKI Kaiserslautern. In addition, he has been a Teaching Assistant with the University of Kaiserslautern-Landau, since 2022. He was a Research Assistant with DFKI, from 2017 to 2019, and has published papers at IEEE Access, Ubicomp, and IUI conferences. His research interests include deep learning-based affective state recognition, assessing student involvement in e-learning, sensor data analysis, and feedback-based intervention in e-learning. He has been a Professional Member of the Association for Computing Machinery (ACM).

ANDREAS DENGEL received the Diploma degree in CS from TUK and the Ph.D. degree from the University of Stuttgart. He is currently a Scientific Director of DFKI GmbH, Kaiserslautern. In 1993, he became a Professor in computer science with TUK, where he holds the Chair Knowledge-Based Systems. Since 2009, he has been appointed as a Professor (Kyakuin) with the Department of Computer Science and Information Systems, Osaka Prefecture University. He was with IBM, Siemens, and Xerox Parc. He is a member of several international advisory boards, has chaired major international conferences, and founded several successful start-up companies. He is a co-editor of international computer science journals and has written or edited 12 books. He is the author of more than 300 peer-reviewed scientific publications and supervised more than 170 master's and Ph.D. theses. He is a fellow of IAPR and received many prominent international awards. His main scientific emphasis is in the areas of pattern recognition, document understanding, information retrieval, multimedia mining, semantic technologies, and social media.



SHOYA ISHIMARU (Member, IEEE) was born in Ehime, Japan, in 1991. He received the B.E. and M.E. degrees in electrical engineering and information science from Osaka Prefecture University, Japan, in 2014 and 2016, respectively, and the Ph.D. degree (summa cum laude) in engineering from the University of Kaiserslautern, Germany, in 2019. He has been a Junior Professor with the Computer Science Department, University of Kaiserslautern-Landau, Germany, since 2021 (appointed by the University of Kaiserslautern, in 2021, and its campus was merged with the University of Koblenz and Landau, in 2023). In addition, he has been a Senior Researcher with German Research Center for Artificial Intelligence (DFKI), Germany, a Visiting Associate Professor with Osaka Metropolitan University, since 2022, and a Researcher with Keio Media Design Research Institute, since 2014. He was also a Research Associate with the University of Kaiserslautern and a Researcher with DFKI, from 2016 to 2019. He is investigating human–computer interaction, machine learning, and cognitive psychology toward amplifying human intelligence. His awards and honors include the Best Presentation Award at the Asian CHI Symposium 2020, Poster Track Honorable Mention at UbiComp/ISWC 2018, and MITOU Super Creator, which is a title given to outstanding software developers (around ten people per year) by the Ministry of Economy, Trade, and Industry in Japan.