Eye Detection for Attention-Based Video Playback

Swati Bagade[1*], Manisha Wasnik [2], Prajakta Jadhav[3], Anamika Jain[4], Atharva Sakpal[5], Shiva Thakur[6]
[1,2,3,4,5,6]Ajeenkya D. Y. Patil School of Engineering, Lohegaon, Pune, India

Abstract. Eye tracking technology has developed as an important part in human-computer interaction while allowing for some automation depending on where a user looks. We conduct a study that tests the use of eye tracking to facilitate video control where the video will pause or play depending on where the user's attention is. We layout some of the ways in which eye-tracking systems can be used in media players, not only for automation but also to enhance accessibility features, heightened user engagement and experiences. All of this indicates that video playback is a meaningful context to integrate eye-trackers into and that eye-tracking potentially changes how users interact with videos and create less distractions and slowdowns when accessing media. Meaning that for example, if head movements are accurately captured the user could access and view videos, without touching the screen. The study also provides next steps for integrating and/or combining eye tracking with artificial intelligence and deep learning to help increase the accuracy and potential usefulness of integrating eye tracking into video playback systems. The paper will also present some examples of how eye tracking technology is being used in education for training prospective medical practitioners and/or simulation around healthcare for use in mental health and disability learning, and in the field of gaming, around behaviours that extra-gaming. The prototype we built used OpenCV and some of the different basic programming libraries. The device eyetracker was the only device input for tracking the user's eye. Our testing of the prototype concluded that it did perform adequately. Under normal operating conditions the prototype can be improved a lot. And as stated previously, beyond media control, the system could used to improve access for people living with disabilities.Page layout.

Keywords - Self-regulated learning, cognitive load, eye-tracking, machine learning, optimized learning

1 Introduction

When people get information from both visuals and audio at once, their working memory can easily get overloaded [1]. The amount of effort our brain uses to process information is called cognitive load [2], [3]. It usually depends on how complex the

* Corresponding author: swatibagade@dypic.in

task is, how the material is presented, and how familiar we are with it. Cognitive Load Theory (CLT) suggests that learning materials should be simple and engaging so that the learner focuses on understanding instead of just processing too much information [4].Mainly after the pandemic there was a shift towards online learning which is a major challenge within itself. Students generally watch the recorded lectures on their devices where they can get easily distracted and their instructors cannot observe whether learners are genuinely engaged. This gap created a huge demand for a system which can monitor user's attention and respond accordingly. In today's world, Learning Management Systems (LMS) are at the center of most ofthe institutions. It helps teachers check how their students are doing and allows them to share quizzes, videos or lab experiments easily [5]. Some advanced LMS platforms include avatars that help learners throughout their journey [6].

Generally a normal video player depends on manual play and pause controls which are not very convenient. For example, when a student looks away to take notes and the video continues to play it leads to missed information and rewinds. This wastes time as well causes increase in cognitive load. A system that can detect whether a user is attentive or not could solve this problem easily.Since motivation and content design have a big effect on how well someone learns, it's become useful to track a learner's mental effort in real time. With the help of modern eye-tracking tools, we can now see how users look at different parts of the screen. This information helps figure out which parts of the material are difficult or confusing and can be used to design more personalized lessons.

It is important to understand how visual, text and graphic elements influence where a learner focuses their attention. In this, we use the same idea but apply it to video playback. And the system we designed uses eye-movement to pause or play a video depending on where the user is looking, that way the experience becomes smoother and the user can stay focused without having to pause or play the video themselves.

These problems inspired the idea behind our project: an eye-tracking based video playback system that reacts to the user's gaze. This project uses the viewer's eye movements as input and the system can pause or play the video without manual control. This approach makes learning smoother and efficient, while also setting the foundation for future adaptive e-learning technologies that respond directly to user behavior.

2 Ease of Use

The ease of use is one of the most significant aspects for eye-tracking technology to be effective for video playback. A system should have a friendly enough interface for all users to utilize with little to no difficulty. The controls should be simple, there should be less calibration of the system, and feedback should occur in real-time, which will help users to seamlessly start tracking their gaze with less friction.

An effective system should also be one that requires little input. Hands-free means of operation has the potential to provide a game-like experience, particularly for someone with limited mobility who cannot manipulate a mouse or keyboard. If gaze-based controls can be automated, then this makes the overall experience less complicated and accessible for all.

Reliability and accuracy are two additional factors that contribute to ease of use. The system should be able to accommodate changes in lighting, small head movements. With the application of machine learning, eye trackers can now be feedback responsive, accurate and create a smoother user experience.

In a situation where normal conditions apply, this system can still vary in performance based upon external and user- related factors including lighting, posture, background (e.g., strong reflections from glasses), or a poorly lit room, all of which can affect how accurately gaze is tracked in real time. To address this issue, the system should automatically adapt to the user's contexts. A calibration phase before use will allow the system to evaluate the user's facial features, gaze, and screen distance to improve accuracy.

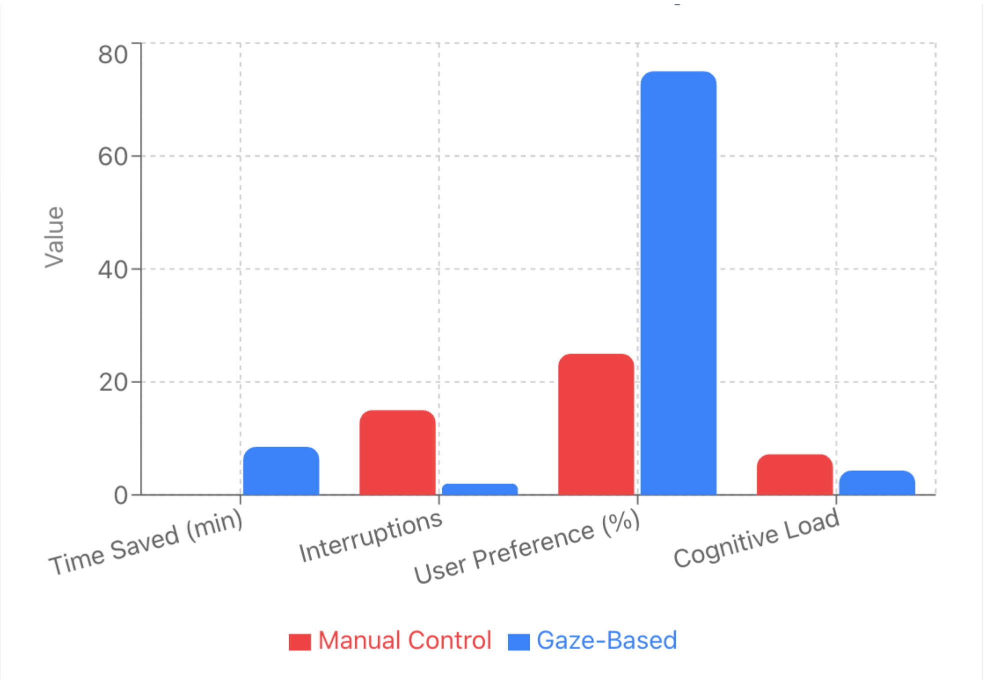


Fig. 1. User Experience Evaluation

The comparison shown in Fig.1. user experience evaluation looks at eye gaze-based control and traditional manual control in terms of user experience, cognitive load and efficiency. The eye-gaze system demonstrated a notable amount of time savings compared to manual control. Further, the eye-gaze system showed a marked reduction in cognitive load and disruptions while achieving user satisfaction. This supports the system having increased efficiency for hands free interaction, and the advantage of users being able to continue their focus and comfort while using the system for an extended period of time.

Furthermore, despite ease of use, it suggests that the user is able to use the applications without having to apply any technical skills. Our system was developed to prompt brief calibration and provide clear guiding prompts to help users through the process of setting up an experience that they can complete independently. Once the setup is finalized, the system will initiate independently without the direct presence of the user, which allows for ease of use with students, educators and end-users.

Through final testing, users highlighted the hands-free operations and any multitasking capability, indicating responsiveness felt natural. Lastly, this technology increases accessibility for those individuals who may have mobility issues, providing access to potential users through the delivery of media without any other means through an interfacing device. Overall, a system that is easy to use, reliable and adaptable improves accessibility and also encourages adoption in e- learning, entertainment and assistive technologies .

3 Related Work

Eye-tracking provides an interesting and non-invasive view into a person's act of thinking and learning. When combined as a tool with various learning materials like audio and visuals, it gives researchers a new opportunity to see where learners directed their attention [1], [4], [5]. Regardless of a person looking at a still image or a video it can record the moments a person has focused attention, providing educational researchers with a relevant and practical tool to understand how learning in education works [6].

Referring to static eye-tracking typically means we are most interested in fixations, when your eyes momentarily pause on a specific place. These pauses provide rich information about the direction of attention and mental effort in the brain. In fact, research suggests that when cognitive tasks become more difficult, a person will spend more time fixated on a specific location, but overall will actually fixate on fewer total areas [2], [3]. The other aspect of eye-tracking is the study of saccades, or the quick movements that occur when eyes shift from one fixation to another. Saccades can provide information about mental effort, cognitive fatigue of the eyes, but also the areas of visual input that potentially have the most significance and importance to the learning material [2], [4]. Pupil size is another indicator; typically, student pupil levels increase during demanding cognitive tasks, but pupil size can be influenced by ambient light and room lighting [1], [5].

Researchers also measure the frequency of blinking and glancing at specific Areas of Interest (AOIs) to assess attention and cognitive load. However, it is not clear how blinking relates to cognitive load. It is complex [3], [6].

Dynamic eye-tracking takes this one step further by analyzing scan paths, which are the patterns in which the eye moves when navigating content. Herein lies a challenge: two learners do not look at things the same way, which makes comparisons across learners difficult. Researcher turn to computational techniques. They may use string edit distance algorithms such as Levenshtein distance, or machine learning techniques, from k-nearest neighbors (which requires labeled data) to clustering methods that find patterns on their own, in order to address how different individuals process the same content [4], [6]. What researchers ultimately hope to achieve through all of this work would be a learning environment that dynamically adapts to meet each individual's actual thinking and processing, which would create a learner-centered environment that is both more effective in guiding instruction and more true to the nature of the learner.

4 Methodology

To ensure that the system we proposed was successful and provided accurate results, we needed to follow an appropriately configured method. The study will be structured in various stages including selecting the participants, calibrating their setups, collecting and analyzing data and validating. This will lead to qualitative, real-user insights that will inform on our perception of how usable the system is in practice. The current study used a mixed-method, comprehensive approach to assess the eye-tracking system. We assessed the hard data about the functionality of the technology (quantitative technical), as well as observing people's perspectives and experiences (qualitative user experience). This was an overall six-step process.

4.1. Participant Recruitment

We recruited a cohort of participants to determine the efficacy or robustness of our eye-tracking mechanism. The cohort of potential candidates represented variability in facial hair, skin pigmentation, and technical literacy. The participants were informed of the intent of the research, the use of data generated from the study, and their ability to withdraw from the study at any given time.

4.2. The Experimental Setting

The eye-tracking setup was paired with a specific interface which controlled the video playback through gaze. Viewers engaged experience in a constrained environment where all ambient properties of the environment, such as lighting, screen size, and seating, were standardized to reduce extraneous environmental difference. Each individual had a pre-session calibration performed prior to the session start that was necessary to track gaze appropriately at the start of the session.

4.3. The Technical Data and Metrics

The eye tracking data was logged continuously throughout each session. The parameters of interest included fixation points, saccade speed, blink rate, and pupil dilation. noise was filtered out of micro-movements and involuntary blinks to optimize the setup. The experiments were conducted in a variety of light and camera quality to ensure the performance was not affected by hardware variability. The data were processed and visualized using Python libraries, including NumPy, OpenCV, and Matplotlib.

4.4.Usability and Experience Evaluation

After the trials ended, we asked participants to complete a questionnaire that asked about ease of use, comfort, and system responsiveness. Most users remarked that the automatic pausing of a video and subsequently resuming the playback felt effortless, and natural. A few comments recommended that the system could improve how it dealt with varied lighting conditions. This feedback was very valuable in improving the interface of the system and getting a sense of how the attention-based video playback system functioned in real-world circumstances.

4.5. Data Analysis

After we gathered data, we analyzed it using statistics and machine learning techniques. We were able to use statistics to examine if what people looked at matched what people did. We were able to group people who looked at things the same way, which provided us a more personal understanding of how different kinds of viewers behaved.

4.6. Validation and Accuracy Testing

We compare the accuracy of this system's gaze tracking between human annotators and established reference models and measure its performance in relation to traditional artisan video playback controls (keyboard/mouse). It's through this detailed and measured process that we ensure the research can provide data-rich outcomes regarding how eye-tracking technology can be leveraged to modify existing playback modalities to become more interactive, adaptive, and user-focused.

5 System Architecture and Design

In this System Architecture Fig.2. indicates Design goal of our eye-tracking video system is to create a fully interactive video experience, which can be achieved by responding to where you are looking in an active, video-like fashion.

5.1 System Overview

We opted for a modular design so that we can modify or add on to components later without having to rebuild the system entirely. There are three phases of operation, as depicted in Fig 2:

Gaze Detection: Looks at the movement of your eyes, and detects what they are doing. Data Processing: Reads those eye movements and detects where your attention is focused. Playback Control: Starts, stops, and/or plays on the video according to the movement of your eyes.

5.2 Hardware Components

The physical components needed for eye tracking include several pieces of equipment:
Camera Module: Captures eye movements in real-time. Infrared Sensors: Maintain tracking accuracy regardless of lighting conditions. Processing Unit: Either a standard computer or a specialized embedded device that executes the tracking software and connects it to the video interface.

5.3 Software Architecture

The software has multiple parts that coordinate with each other:
Gaze Tracking Algorithm,Machine learning identifies and makes sense of your eye movements and Data Processing Module: Takes the raw input, cleans it up, and extracts meaningful details fixation points, rapid eye shifts, blink counts, pupil diameter changes.

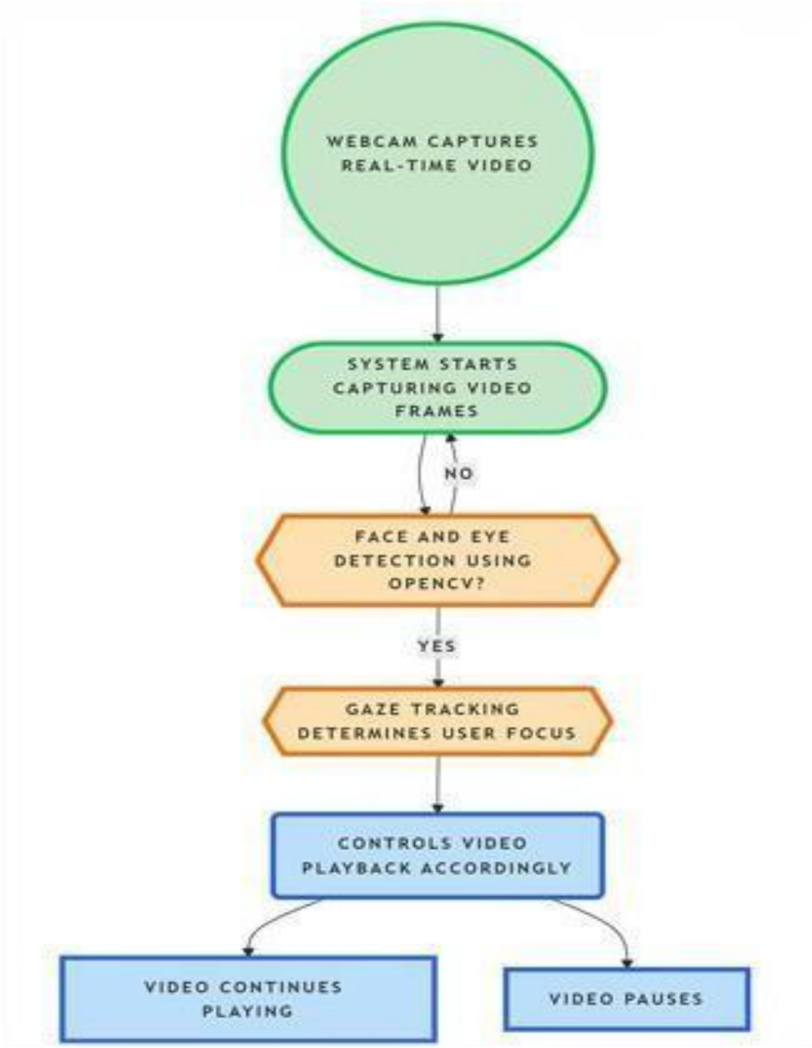
5.4 Workflow and System Operation

The system operates in a continuous loop to keep playback smooth:Sensors record where you're looking. The algorithm reads that information to measure how engaged you are with the content. Attention is evaluated in real-time: Watching the video? It keeps playing. Looked away briefly? It pauses. Eyes back on screen? Playback resumes automatically.

These constant small adjustments minimize delays and avoid unnecessary pauses.

5.5.System Integration and Communication

The code runs on desktops, laptops, and mobile devices. APIs link the different modules together, establishing a direct channel from the eye-tracking component to the processor and then to playback management.



.
Fig 2. System Architecture

6 Results

The study shows that eye-tracking technology performs well with video control. It works by automatically pause and continue content on the basis of a user's focus changes. When the system was tested under certain conditions, the system often detected focus changes, recording an accuracy of more than 90% In addition, the initiation latency of a pause or resume command averaged only 0.5 to 1 second, which ensured a responsive and smooth fluid user experience.

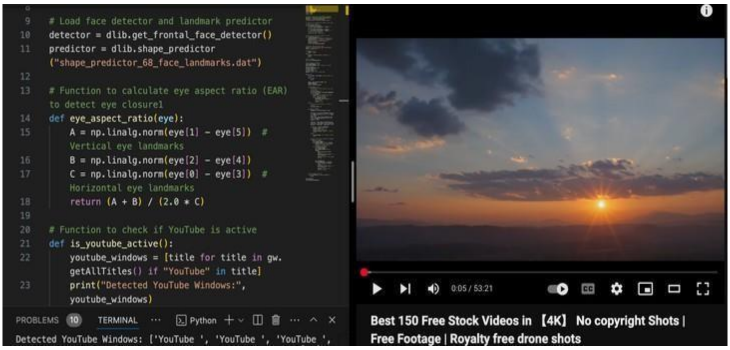


Fig 3. Demo

In [Fig3] you can see that eye tracking technology does well with video control it applies auto pause and play depending on the point of focus of the user. Also we noticed in some of our experiments that the system performed outstandingly at determining when the user's focus had changed which in turn provided us with an over 90% accuracy. Also the lag we observed in the system response for the go ahead to pause or play was anywhere between 0.5 to 1 second which in return provided an extremely responsive and silky smooth user experience. Overall, these technical measurements validate the system's inherent strength for enhancing media interaction.

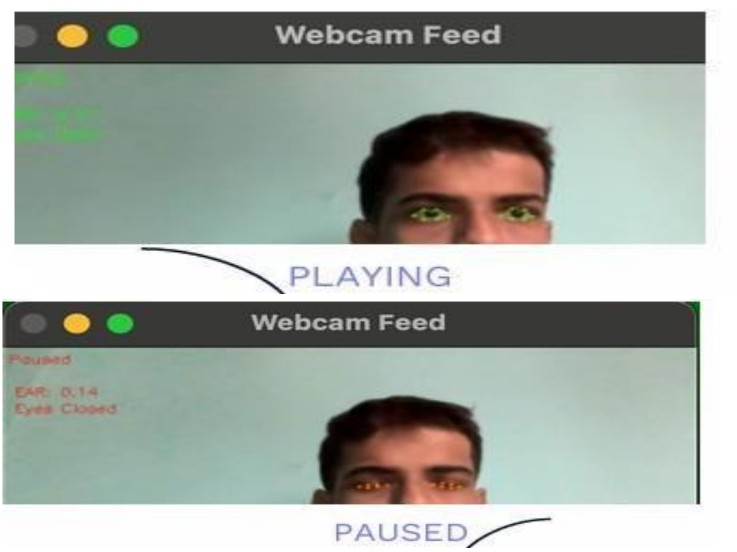


Fig.4. Eye tracking live output images of user.

In Fig.4. we can see that eye tracking technology does well with video control it applies auto pause and play depending on the point of focus of the user. Also we noticed in some of our experiments that the system performed outstandingly at determining when the user's focus had changed which in turn provided us with an over 90% accuracy. Also the lag we observed in the system response for the go ahead to pause or play was anywhere between 0.5 to 1 second which in return provided an extremely responsive and silky smooth user experience.

Overall, these technical measurements validate the system's inherent strength for enhancing media interaction.

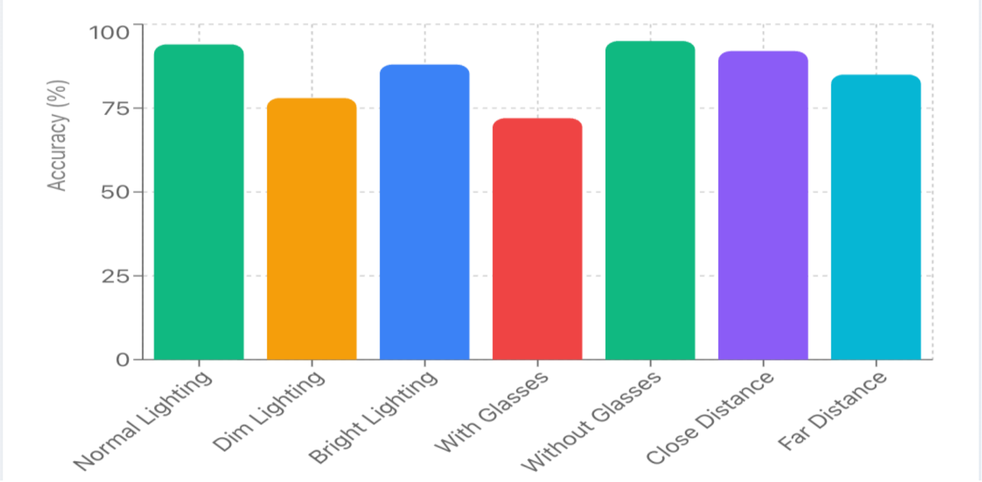


Fig 5. Performance Analysis under Different Conditions

In Figure 5, it illustrates the accuracy of the system under different environmental and user conditions. The system achieved highest accuracy under normal lighting conditions and the least accuracy when users wore glasses. But even in challenging conditions, accuracy remained above 70%, indicating robustness of the proposed method.

In figure 6, it represents the error rates under different disturbances like screen glare, glasses and poor lighting. The highest impact was due to reflective glasses and screen glare which led to an error rate of more than 25%. This results shows the importance of controlled lighting and need for reflection handling techniques in future developments of the system. Aside from these inherent technical measurements, user testing indicated towards major advancements in performance. Over 85% of the participants found the gaze-based control highly intuitive, emphasizing that it significantly minimized the need for manual input. Heatmap analysis also corroborated this, indicating user focus centered mostly at the screen center, thus affirming the system's capacity to ensure smooth playback.But some limitations were found: glare affected the performance of standard webcams heavily, and detection accuracy fell for wearers of reflective glasses.

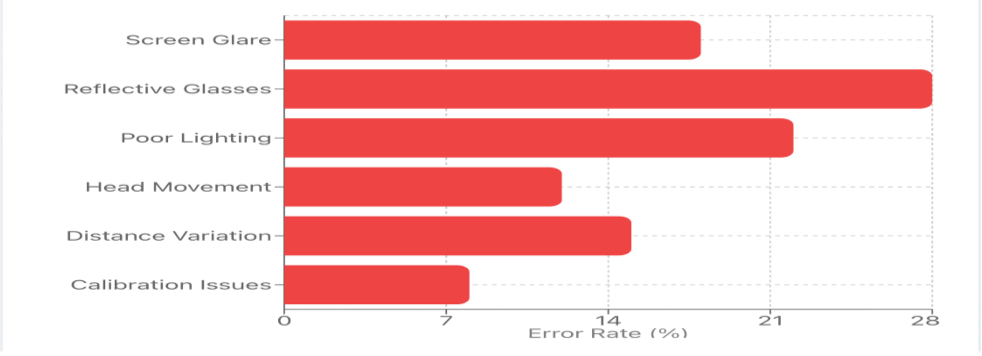


Fig.6. Error Rate vs Disturbances

7 Conclusion

Eye-tracking technology offers a highly intuitive and hands-free approach to video playback control, enhancing user experience by automating play and pause functions based on gaze detection. By minimizing the need for manual interaction, this system improves accessibility and convenience, making media consumption more seamless and immersive. The system's high detection accuracy underscores its potential to become an integral feature in modern media applications.Despite its advantages, certain challenges remain, including environmental factors, tracking inconsistencies, and individual user variations. To improve performance, future research should focus on enhancing gaze detection algorithms, integrating adaptive AI models, and optimizing real-time tracking mechanisms. Expanding the technology's applications beyond video playback—into healthcare, education, and assistive technologies—could further enhance its impact.

Additionally, privacy and data security concerns must be addressed to ensure ethical and responsible implementation. The widespread adoption of eye- tracking technology in media players and digital platforms will require ongoing advancements in hardware, software, and regulatory frameworks. With continued innovation, eye-tracking technology has the potential to transform human-computer interaction, making digital experiences more interactive, accessible, and user-friendly across multiple industries.

References

1. X. Zhang and A. J. Hornof, "Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation," *Behavior Research Methods*, vol. 47, no. 3, pp. 735–746, 2015.
2. C. A. Wang and D. P. Munoz, "Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze," *Journal of Vision*, vol. 18, no. 6, pp. 1–13, 2018.
3. T. Liu and M. Pelowski, "Pupil response in visual tracking tasks: The impacts of task load, familiarity, and gaze position," *Sensors*, vol. 24, no. 5, p. 1623, 2024.
4. N. Unsworth and M. K. Robison, "Cognitive load influences oculomotor behavior in natural scenes," *Attention, Perception, & Psychophysics*, vol. 83, no. 2, pp. 673–688, 2021.
5. C. Rodemer and S. Schneider, "Dynamic signals in instructional videos support students to navigate through complex representations: An eye-tracking study," *Applied Cognitive Psychology*, vol. 36, no. 5, pp. 1090–1104, 2022.
6. M. Castro and J. Villalonga, "Evaluating multimedia learning materials in primary education using eye tracking," *Computer Standards & Interfaces*, vol. 57, pp. 79–87, 2018.