Cinefeel: Artificial Intelligence (AI) Driven Multisensory Cinema App for The Visually Impaired

Mrs. Shriyansh Fasate[1], Mrs. Durvesh Hukare[2], Ms. Himani More[3], Dr. S. R. Gupta[4], Ms. Vaibhavi Raibole[5]

[1,2,3,4,5]*Artificial Intelligence & Data Science, Prof. Ram Meghe Institute Technology & Research. Badnera Amravati, India*

Abstract*— Movie viewing is usually a difficult experience to people who cannot see since they are deprived of the exciting and detailed experience that movies are typically associated with. Although there are already solutions to the problem, such as audio description which provide some assistance, they usually have a number of limitations and do not establish a strong emotional connection with films. This paper purpose, an AI based multisensory cinema application that aims to improve the movie experience of individuals with visual disabilities. Based on smart AI applications such as Natural Language Processing, Speech Synthesis and Computer vision, the system generates real time audio descriptions, emotional descriptions and sensory information. In so doing, CineFeel makes users have a better understanding of scenes, character feeling and what is happening on the screen. Test findings indicate that the new systems highly increase the level of interaction and emotional attachment by the users as opposed to working with audio alone. Ultimately, CineFeel desires to make the experience of watching movies more comfortable, entertaining, and pleasant to those with poor eye sight.

Index Terms*— Artificial Intelligence, visually impaired, Multisensory experience, Computer vision, Audio description, Accessibility.

I. INTRODUCTION

The digital entertainment industry has evolved quite well in the past ten years, owing to the improved multimedia technologies, the higher internet connectivity speeds, and the higher number of people having access to the mobile devices. The visual narrative is now a significant part of the contemporary film and the media of the internet. This has not only served to make the experience of the people with eyes better but also to show how the blind and the sightless have been left out by this practice.

The world has a big number of people with visual problems, and they are actively involved in digital media education, entertainment, and socialization. However, as the visual component of human senses is the focal point of cinema, one could not enjoy the benefits of such relevant characteristics, such as facial expression, body language, character motion, and visual symbolism. Accordingly, the whole essence of the characters, the speed of the plot development and the mood of the surrounding people cannot be totally understood and therefore the experience of watching the movie is one-sided and not as interesting.

Screen readers and the traditional audio description services are valuable but not sufficient. Such systems merely provide objective descriptions of the visuals and are not emotional, not space conscious and interactive. This reduces the movie experience to mere passive consumption of information and not an emotional one.

The recent advances in artificial intelligence (AI) and affective computing have made it possible to enhance accessibility in digital storytelling. As a result of the natural language processing, computer vision, and deep learning, AI systems are able to understand the context, emotion, and behavior. However, despite this growth, the use of AI in the movie accessibility is mostly overtaken by implementation of functional support, and less concerned with providing emotional or multi-sensory experiences.

The suggested work and research are CineFeel: an AI-based Multisensory Cinema Application to Visually Impaired users, a novel mobile-based application that will transform the experience of visual storytelling by the representatives of the mentioned category. The app allows a variety of input formats, and it implies that a user can post video content, audio stories, or even the

version created by AI, thereby, offering flexibility and customization. CineFeel involves a multisensory design, emotion sensitive narration, three depersonalizations of space with audio and haptic feedback to create the form of the engaging and emotionally communicative narrative space.

CineFeel is centred around a set of high-level AI models that comprise Gemini, GPT-3.5, BLIP-2, and Whisper and work together to evaluate story structure and cinematic content. Such models enable such system to detect affective tones, make sense of character behaviour, detect the dynamics of a scene and make sense of the surrounding context. The combination of the visual, auditory and language founded information generates the amplified narrative output of CineFeel that goes beyond the literal descriptions as an accent of definition instead of emotional description and research richness.

Emotional narration is generated with the assistance of highly developed Text-to-Speech (TTS) systems such as Eleven Labs, Google Cloud Text-to-Speech, and Azure TTS that have the ability to modulate the voice to convey the emotions of joy, sadness, fear, tension, and excitement. It is a living narration that renders the narration closer to the emotion of the story and the audience as well that makes the narration process more enlivening and humane. In addition, CineFeel uses 3D spatial audio system, such as Dolby.io or Resonance audio SDK, to simulate the directional audio situation. This allows the users to feel the spatial location, movement, and depth of the space with sound hence recreates the visual effect of film space without sight.

To add more immersion, CineFeel also possesses the haptic feedback systems with a vibration-based system. The system also transfers visual and emotional resources into the information of touch through well-organized haptic patterns which make a person more emotional in their interaction with the story. These haptic cues are synchronized with plot points and emotional climax to provide the users with the possibility to feel the cinematics, action scenes or emotion shifts.

Rather than offering an alternative to visual awareness, CineFeel suggests another mode of experiencing the cinema, in which visual representation does not play a crucial role, but, instead, a sort of narration, also of the same importance and engaging with the non-visual representation.

The proposed system is highlighted regarding empathetic empathy emotional competence and humanistic interaction that encompasses technical and gaps in the current accessibility solution of experiential care. As having been identified at the intersection of Artificial Intelligence to the Availability, user Centered Computing, emotion aware Computing, and Assistive device, the study will be beneficial to digital equity and accessible media design. Lastly, CineFeel views the future of film as the blind one where storytelling overcomes the barriers of images and is abundant with the multisensory experience.

Figure 1: Conceptual Framework

II. LITERATURE SURVEY

According to the proposal of the research work by K.Patil, A. Kharat, P. Chaudhary, S. Bidgar and R. Gavhane, it was proposed that the research work should aim at enhancing the lives of the low vision users by making them less dependent on others. Low vision users tend to experience the difficulty in executing normal tasks like the ability to detect objects, comprehend their environment, read or recognize the people around them. As a solution to these challenges, the study will suggest the creation of a wearable assistive device that will act as a personal virtual assistant to blind people. The primary objective of this system is to allow the visually impaired individuals to perform simple everyday tasks on their own with the help of a convenient and user-friendly interface.

The device suggested is the combination of various smart components into one wearable device. The features of these sections include environment awareness, object detection, mood-sensitive face recognition, and text reading. The interface with the system is primarily sound, enabling the users to issue commands and get a sound answer, which is necessary to the blind users. The system was developed using Python programming and various deep learning methods and libraries to ensure proper identification and interpretation. Overall, the project is all about simplicity, usability, and usefulness, as it will make daily routines simpler and easier to manage by the low vision users to enable them to perform their own tasks and trust in their ability to do them on their own. [1].

The concept of immersive spatial sound of mobile multimedia: the article titled Immersive spatial sound for mobile multimedia, by V. R. Algazi and R. O. Duda suggested that the development of a novel approach to enabling individuals to experience immersive multimedia on handheld computers is achievable, despite the fact that mobile screens are not large enough to immersivity present information to the individual user. The use of sound instead of visuals in the study is meant to provide the listeners with the real-life feel of space and movement through headphones. The researchers designed a system that is referred to as MTB or motion tracked 3D sound and take the advantage of how individuals experience sound in the real world. The system also uses a head tracker to shift the sound based on the movement of the listeners head and this creates the illusion that the sound is emitted in various directions around them. This approach does not only create a natural experience of listening, but also creates possibilities of the mobile applications, which can be entirely immersive without having to use big displays. The paper outlines the key concepts behind this technology, the difficulties encountered in implementing it, and some of the applications of this technology in practice by using the MTB 3D sound. Last of all, it talks of the future opportunities of mobile multimedia, where one can add 3D sound, head-tracking and even small-sized visual displays, to make interesting experiences that are both natural and involving [2].

According to I. Hwang, H. Lee and S. Choi, in their article titled as Real-Time Dual-Band Haptic Music Player for Mobile Devices, a new kind of haptic music player that could be used in mobile devices to enable one to feel the music via the vibrations was proposed. Unlike the old system where the deep sounds are merely produced by vibration, the new player has the capacity of forming dual-band vibrations which means that the deep sounds are produced with high pitched vibrations. The original component of this system is a small two-mode actuator will be incorporated with an algorithm in real-time vibration generation. The algorithm is fed with a music file, receives vibration patterns, and translates them into touch sensations by use of a so-called haptic equalizer, which correlates vibrations with the un of a human being thus making the experience natural and lifelike.

The authors also conducted user study to understand the experience of the people using the system considering the aspects such as accuracy, peace, enjoyment and personal preference. They conduct a comparison of the new dual-band haptic playback and the traditional one that provides the bass vibrations using just a single-frequency actuator. The findings indicated that the dual-band vibrations provide more enjoyable experience and are better than the old style of bass-only method. The paper also provides information on the way such systems can be enhanced further. On the whole, these works indicate that the development of sophisticated actuators, real-time vibration algorithms, and perceptual information may contribute greatly to the multimedia experience on cellular phones and make music not only what you can hear but something you can touch with your body, as well [3].

The article titled "Calliope: Automatic Visual Data Story Generation with a Spreadsheet," by D. Shi, X. Xu, F. Sun, Y. Shi and N. Cao suggested that it is easier to help people comprehend complex graph or network data by providing simpler methods of data reporting where it is essential to know the relationship between various items in the data. Graph data may be used to demonstrate the relationship among people, organizations, or concepts but it is commonly complex and difficult to comprehend. It is time-consuming and difficult to analyze these graphs by hand, find interesting facts, and use them to create meaningful topics, which is typically required of data reporter or analysts. The researchers developed an automatic system named Calliope-Net to cope with this problem. There are three key components of this system: a discovery of facts module, which identifies key facts within the graph, a facts organization component,

which organizes the facts discovered into a meaningful way, and a visualization component, which generates the information in the form of labelled node-link diagrams.

The algorithm of creating novel layouts is also incorporated into Calliope-Net as the algorithm has been developed to ensure that the labelled graphs are not only informative but pleasing to the eye so that the users could easily comprehend the structure and the most important facts of the information presented in the data. To determine the effectiveness of the system, the researchers tested the system using two case studies and Peran in-lab user study. The findings revealed that Calliope-Net assists users in finding vital facts and interpretation of graph data in a more efficient way and a presentable and comprehendible graph view format. On the whole, the research demonstrates that an automated generation and visualization of graphs can significantly enhance the effect and clarity of dealing with complex network data [4].

L. P. Khan, V. Gupta, S. Bedi and A. Singhal proposed in their article- "StoryGenAI: An Automatic Genre-Keyword Based Story Generation," that using generative artificial intelligence and more advanced visual creation technologies can help revolutionize the process of automatic story creation and presentation. Even though today large language models (LLM) are quite capable of producing clear, meaningful, emotionally compelling scripts, the task of transforming said written scripts into precise and valuable visuals remains a major challenge. This issue is referred to as the semantic gap, which arises after the visual information does not entirely conform to the meaning, emotions, or intent of the text. To eliminate this problem, the paper presents an entire framework that relates AI-based story writing to high-quality visual storytelling. The authors mention contemporary approaches to narrative creation, strategies that allow to externalize and maintain the primary message of the narrative, and effective visual making systems which rely on diffusion models, as well as multimodal AI systems. The proposed structure aims at maintaining the visual and textual correspondence, maintaining the fluent development throughout the storyline, and consistency of the characters, scenes, and plot development. By conducting experiments and close case studies across various categories, the findings indicate that the system can produce readable and visually homogeneous stories. On the whole, this piece of work establishes a solid foundation of future automated storytelling systems and provides an idea of how this can be used in the future in fields like entertainment, education, and interactive media where rich, AI-based narratives can significantly enhance user experience [5].

In "Multi-Sensory Interaction to Blind and Visually Impaired People" by Cho, J.-D., it was proposed that multi-sensory interaction would be important in enhancing the process of learning, inclusion, and collaboration, particularly because people receive information in different ways and using various senses. It describes the natural ability of the human brain to combine data of other senses, including sight, sound, touch and movement (proprioception) to create a whole and significant experience. This is referred to as multi-sensory integration that is most effective when information in one sense has commonality with information in another sense. Another concept introduced by the introduction is that of cross-modality (the interaction of different sensory channels with one another) and cross-modal correspondence (the natural way people form associations between features in different senses, e.g., sounds with shapes or textures with colors) despite their non-relationship. The introduction has made it clear that the conventional systems of human-computer interaction are highly reliant on visual information and therefore, it is very hard to get visually impaired individuals to interact with the system. In spite of the way other sensory modalities, like haptic feedback, sound, and even scents, have been used to counteract the absence of vision, none of the other modalities still rely on the richness and power of the vision. Thus, some key areas of research have been identified in the paper that aim to enhance accessibility and inclusion such as universal access in HCI, assistive technologies, haptic and tactile interface, ambient assistive intelligence, and multi-sensory color coding. To sum it up, introduction gives a background to the research which is aimed to create human-centered multi-sensory systems that are more accommodative to the visually impaired and more inclusive interactive experiences [6].

The article in question, A comprehensive review on NUI, multi-sensory interfaces and UX design on

applications and devices to visually impaired users, in turn, by L. Arora, A. Choudhary, M. Bhatt, J. Kaliappan and K. Srinivasan, suggested that the given paper will be a review of the way modern technologies (gadgets, devices and apps) are designed to be used by partially sighted or people with partial visual impairment/ no eyesight at all! Undoubtedly, the technology has been introduced comprehensively to our life, it served numerous people but as mentioned in this abstract, it still has much to offer to the blind or vision impaired face in terms of their proper use of these technologies. The issues are in the fact that many apps and gadgets do not suit their needs. This paper will particularly focus on the review of the Natural User Interfaces (NUI), the multi-sensory interface, and user experience design (UXD) or NMUD in reference to CVI and PVI users. It also quickly makes us remember how important the idea of an accessible and human-centered design is - to help technology to be applied to the available abilities, limitations and daily experiences of blind users. In this paper, we have offered a taxonomy of the significant elements of design which developers consider when designing assistive applications.

Moreover, the abstract also refers to several difficulties of technologies which are already presented to the visually impaired people. These can be connected to usability, low cost, and access or usability on the go; it can also have a practical limitation such as short battery life, system latency (delays), no user control and low functionality. Finally, potential areas of research are also indicated in which the user-centered design needs to be followed more harshly [7].

According to Z. Xie, Z. Li, Y. Zhang, J. Zhang, F. Liu, and W. Chen in the article titled A Multi-Sensory Guidance System for the Visually Impaired Using YOLO and ORB-SLAM, the topic of the guiding system of the visually impaired has gained popularity over the last few years. The current guidance systems present in the market generally apply auxiliary instruments and techniques like GPS, UWB, or a plain white cane which uses the one of the senses (tactile or auditory) of the user. The guidance methodologies may be inadequate in a complicated indoor setting. The paper suggests a multi-sensory guidance system to the vision loss users capable of establishing the tactile and auditory guidance based on the ORB-SLAM and YOLO methods. The local obstacle avoidance system is achieved based on an RGB-D camera, and the tactile level through point cloud filtering that can notify the user by a vibrating motor. With the coordinate transformation, our suggested approach will be able to produce a dense navigation map to execute global challenge avoidance and path planning on behalf of the user. The hearing level also includes real-time target detection and voice-prompt system which is based on YOLO. Our bike was the suggested system in the form of a smart cane. There are four test scenarios on which experiments are conducted. In practice, it was observed that the obstacles of the walking path can be reliably detected and categorized in real-time. Our suggested system may serve as a powerful assistant that allows people with sight loss see their way safely by combining YOLO with ORB-SLAM [8].

The article titled Sight-to-Sound Human-Machine Interface to Guide and Navigate Visually Impaired People, by G. Yang and J. Saniie is dedicated to the issues that visually impaired individuals struggle with when exploring the complex environment. Navigation is inherently a challenge to them since they, unlike the sighted who relies on visual clues to form mental images or cognitive maps of his or her environment, are deprived of this visual feedback and have to resort to other forms of sensory inputs. To cope with such a gap, the researchers have suggested a new system, the Sight-to-Sound Human-Machine Interface (STS-HMI). This system involves machine vision to scan the visual data of the surrounding and then converts them into dual ear audio signals that are 3D audio clues that replicate directional and distance hints. Listening to these audio clues, the user who has lost their sight can create a mental map of the things that are around him, and as a result of this, they can be guided more easily and effectively.

The experiment involves a sequence of experiments carried out in complicated navigation scenarios to test the efficiency of the STS-HMI system. The findings indicate that the system enables the vision loss users to navigate through difficult environments with minimum efforts hence increasing their speed and safety. In a nutshell, the STS-HMI helps in connecting the gap between sight and hearing since it transforms visual data into meaningful sound patterns, which provides a flexible, user-friendly, and immediate navigation solution. On the whole, this study

demonstrates the possibilities of human-machine platforms in enhancing mobility and independence of users with vision loss and demonstrates how technology can complement natural sensory constraints and enable them to move around more confidently in their everyday life [9].

In the article Audio-Vision Substitution for Blind Individuals by D. J. Brown and M. J. Proulx the authors are asking the question to what extent can human beings understand and process information where hearing replaces the sight of their bodies through a sensory substitution device with the primary purpose to solve the problem of the limitation of the human information processing capacity. It is a matter of two things, how much auditory experience a person can handle, versus whether that same experience can be converted to visual and the other way around, and how it influences the performance of people in tasks, and one example she mentioned is object recognition. The researchers begin by examining past researches. On what vision-touch exchange and sound-vision substitution can be on such devices and also reflect on theories of restricting the amount of human information processing particularly using auditor input [10].

The article, Environmental sound classification with convolutional neural networks by K. J. Piczak suggested that with the emergence of deep learning in the fields of Computer Vision and Natural Language Processing, deep learning has now found its way in the field of Music Information Retrieval. Transfer learning on deep learnings is a common practice in order to maximize deep learning, and also, learn with few data. By so doing, a previously trained model of a neural network may be utilized as the new learning task input. The common assumption is, in the case the first and new tasks seem to be connected and are founded on the same input data (e.g., music audio), then deep representation of the data is also probable to inform a new task. Nonetheless, given that all most of the networks used in training a single learning task, it is possible that they are not general enough to be applicable to all future tasks. This paper presents the result of our empirical experiment on what is more relevant to generate deep representations to data and learning tasks in music. We can address this question by conducting a large-scale empirical study: we apply a variety of learning sources, and we apply various levels of sharing among the sources, and we learn music representations by a variety of deep neural network architectures. We then validate these representations using a number of target datasets to test them [11].

Application of Psychophysical Techniques to Haptic Research, by L. A. Jones and H. Z. Tan, had suggested that this paper is a review of psychophysical techniques of studying human haptic perception. As indicated in the paper most researchers choose their psychophysical methodology depending on tradition and what other researchers are doing instead of any distinct reason as to what method suits the question being investigated. This is the reason why we make our best to ensure that the review describes and compares various approaches in a clear manner to enable the researchers to understand it better.

On the one hand, the paper covers classical psychophysical techniques commonly used to measure individual sensory thresholds (e.g., the lowest strength of a stimulus that can be detected or the minimal difference in strength between two stimuli that can be detected) with some of the more modern methodology such as adaptive procedures and methods based on the signal-detection theory. It gives an account of how these methods should be carried out so that it can be able to find out how far the thresholds can be determined and the difference thresholds as well as the points that may influence the responses that the participant gives including attention and expectation or bias.

Besides threshold measurements, the paper outlines the ways of evaluating the response of the human observer to suprathreshold stimuli (stimuli that are not only visible but also detectable). They are scaling procedures, which facilitate the degree of measured quality of segmentations perceived. The review shows the importance of stimulus- and response-related biases on results and emphasizes that it is important to be mindful of these effects when analyzing the perceptual processing [12].

In the article "Artificial Intelligence-Powered Mobile Application to Help Visually Impaired People," by S. Jagadish, M. Lalitendra, K. Nikhita, P. Narsingarao, and Lolla Sreedhar, the authors explain that a mobile application based on the AI was created to assist visually impaired people in accomplishing their daily tasks to ensure that they become more independent, safe, and accessible. The most important aspect that can enable users to experience their world is

integration of different cognitive features (such as recognition of an object, color identification, detecting a currency ad bill, reading a barcode and text etc.). Due to the fact that the system provides the user with the accurate audio feedback, it contributes to the practicality and reliability of using the app when commuting within a low-tether zone, identifying familiar objects and accessing printed materials with the help of the app, without involving the help of relatives or friends. The app has an artificial intelligence that is seamlessly and accurately accurate in its functions. In general, the paper highlights the importance of AI-driven assistive technologies to promote the quality of life, independence and inclusiveness to visually impaired individuals [13].

"AI Based App for Blind People." According to the authors of the abstract as explained by T. Qureshi, M. Rajbhar, Y. Pisat, Vijay Bhosale and the student, we only managed to determine that the research took into consideration the global importance and growing demand regarding the assistive technology of the visually impaired, because the World Health Organization (WHO) revealed that there are about 1.3 billion individuals in the world with some degree of vision impairment. [1]. The article brings to fore that even with the technological advances that are occurring at a phenomenal rate among the visually impaired, affordable and easily accessible assistive technology remains in high demand [11].

The article in question was authored by M. Daraban, K. Lengyel, A. Faluvegi and C. Amariei, and it was titled Real-time Spatialized Sound Generator on Embedded Wearable Platform, in which the authors suggested the idea of making the blind less physically reliant on other people through a mobile application called Trinetra App, and this will act as their digital companion to guide them through their daily life. The paper begins by stating that the blind people tend to rely on others to perform simple chores such as steering their way through the world and taking part in daily routines, and argues on the plausible enough that they should be able to be as self-reliant as other people. They can only improve their daily life by constantly communicating with the user and giving them real time information of what is around them.

It uses state of the art features such as computer vision, speech recognition, image processing and optical character recognition (OCR) to identify objects, read the world and identify text, and the surrounding world. The app is built on flutter and can be used on both Android and iOS devices - as such, it has that going on. It is able to identify objects in front of the user and the recognition written text to be read [15].

Mobile App for Enhancing Accessibility Among the Visually Impaired, V. Pujari, K. Madnal and D. Premchandran, suggested that the study paper is concerned with trying to fix what is left about accessibility to mobile applications by persons with visual impairments even after the substantial advances in technologies. The abstract mentions that, although mobile applications have already provided high quality of life to visually impaired persons, the majority of the existing methods are not yet satisfied with low accuracy rate, high latency time and low usability issues until they can also be used in an actual situation. This paper introduces a new mobile application, which is specially designed to meet the needs of the blind users. The app assists users to be more effective in engagement with their environment and ease in their daily activities. It is constructed on the basis of the AI-powered features, such as real-time object recognition, real-time text extraction, and real-time currency detection that allow users to get instant responses they can trust. The app accelerates independence with faster, more precise and convenient access to the safety [16].

Restorative Neuroscience and Neurology. By Abboud S, Hanassy S, Levy-Tzedek S, Maidenbaum S, Amedi A. EyeMusic: proposed that the article is about the improvement of sensory substitution devices (SSDs) that help the blind to perceive what they see through sound or tactile sense. As the paper has pointed out, several of the earlier SSDs have serious flaws such as giving users unpleasant auditory experiences and failing to represent any color, thus, undermining the usability and acceptance in the long term. This is an introduction of a new visual-to-auditory sensory substitution system, the EyeMusic.

EyeMusic is created to encode the visual information into musical notes and reduce the shape and color of objects encoding to a less cumbersome and more direct and fluid way. Pentatonic scale musical notes are taken as input and the sounds of natural instruments which are modulated by the inputs of the users are received hence making the auditory output pleasant compared to the traditional SSDs. The blind was induced into a short behaviour training plan to get

familiar with the sounds played by the machine. The results showed that: a) the participants were capable of identifying the two kinds of parties, and b) with proper priming, they could define them with naming terms of simple shape and color within 2-3 hours, which implied that the system is user-friendly.

In addition to the performance tests, it also involved the researchers to measure the opinion of sound quality by the users. The findings indicated that the respondents could accept and feel comfortable with the auditory stimuli and therefore were free to utilize the system in the long term without stress levels. Overall, this research concluded that the EyeMusic approach is a good user-friendly tool that can allow blind and visually impaired listeners to manipulate visual information [17].

The article, Haptic interfaces and devices, by V. Hayward, O. R. Astley, M. Cruz-Hernandez, D. Grant, and G. Robles-De-La-Torre, led me to believe that the article was about how haptic interfaces can be used to promote direct and tactile communication between human beings and machines, based on the response to the activity of a person in physical space. One of the most peculiar and the most important aspects of haptic interfaces raised by the paper is the two-way information flow it enables because the user feeds the machine with the information by moving, and the machine reacts to the feedback with touch. It simply makes the interaction more human and interactive as compared to a sight or sound-based interface.

Also, the abstract makes an emphasis on the fact that haptic interfaces, as any display technology, is developed with the consideration of the shortcomings and strengths of the human perception so that a tactile experience can be achieved and be comprehended. The paper also explains the fundamental aspects and concepts of operation of haptic systems in that the systems have base operations which are supposed to create the touch experiences of real objects. It also encompasses other new and emerging uses of the haptic interfaces in other fields. [18]

III. METHEDOLOGY

CineFeel system workflow is designed to transform the conventional visual and aural medium to the multisensory experience of the theater. In the sequence, the architecture has a speech-to-text processing and affects the computing coupled with a spatial audio rendering and haptic feedback generation. This pipeline will guarantee that the meaning of the scene is correct and that there is the proper synchronisation of the multisensory outputs. It is broken down into four processes i.e., Data Ingestion, Semantic and Emotional Analysis, Multisensory Synthesis and Output Delivery.

3.1. Data Processing and Data Consumption.

The user interface will serve as the start of the workflow in which the user will add a movie, video story, or audio file as the material.

- Standardization of the inputs: The system can accept non-regular input. The media is uploaded and scanned to establish the presence of audio track or not.
- Audio Extraction and Generation: It is broken in a video file with audio and the audio track is stripped in order to process it. In cases where the input is not of audio nature, or it does not have enough audio data, the system executes an AI-based narration system to produce descriptive audio to provide the visually impaired consumers with access. It does so that it will transform all the inputs into a unified audio stream that can be further analysed at the downstream.

3.2. Semantic and Emotional Analysis.

After the audio has been standardized, the system further uses more advanced Natural Language Processing (NLP) and computer vision algorithms to shred the story into pieces.

- Speech-to-Text (STT) Conversion:

The processing voice feedback is sent to the Whisper AI system in the automatic speech recognition (ASR). The given module is used to transform a verbal communication and narration into written transcript that becomes a semantic basis of communication.

- Scene Segmentation and Metadata Generation:

The system is doing a dual-stream analysis. It is a textual transcript analysis to know form of conversation and narrative progress. It is a process step in video frame that is used to determine video transition, video cuts and time variation. This results in creation of scene-specific information such as the timestamps of boundaries and description of context.

- Emotion and Sentiment Detection:

With the help of the created metadata, an Emotion Detection AI can identify the special affective patterns of a scene (e.g., happiness, sadness, tension, fear, excitement). At the same time, the haptic and sound effects are generated in accordance with the intensity of the scene, i.e., between action scenes and the scenes with pause and the narrative, which is examined by the system.

3.3. Multisensory Synthesis

This is where the semantic and emotional information is trailed to the physical and acoustic results.

- Emotion-Aware Narration: Text-to-Speech (TTS) can be implemented with Coqui TTS or XTTS and it can enable the system to alter the tone, the degree of pitch and the speed depending on the emotion on which the scene is classified.
- Spatial Audio Rendering Audiostream is rendered with the help of a Spatial Audio Engine (developed based on such solutions as Dolby.io or Resonance Audio). In this module, the audio objects are located in 3D coordinate system which is a fake creation of depth, direction and movement of the characters, in an attempt to build a real acoustical scene.
- Haptic Feedback Generation A Feedback Engine of Emotion Converts Emotional Power and Scene Dynamics into Haptic Feedback. It creates low intensity vibration pattern of soft emotional sequences and high intensity pattern of suspense or action-oriented sequences in the system.

3.4. Synchronization and Output

The last step makes sure that the time fit of all the sensors modalities generated is considered.

- Multisensory Synchronization Multisensory Synchronization Multisensory Synchronization Multisensory synchronization can be done with a synchronization engine that synchronizes the spatial audio, emotion sensitive narration pattern and haptic feedback pattern to the original media stream with temporal accuracy.
- Delivery and Logging: Multisensory output is delivered to the consumer through an audio enabled and haptic enabled device. Simultaneously, with the help of the system, the session metadata, emotional profile, and playback analytics are logged into the key database. This type of information is kept to keep track of the system as a means of improving the system performance and also to provide future customization of the system users.

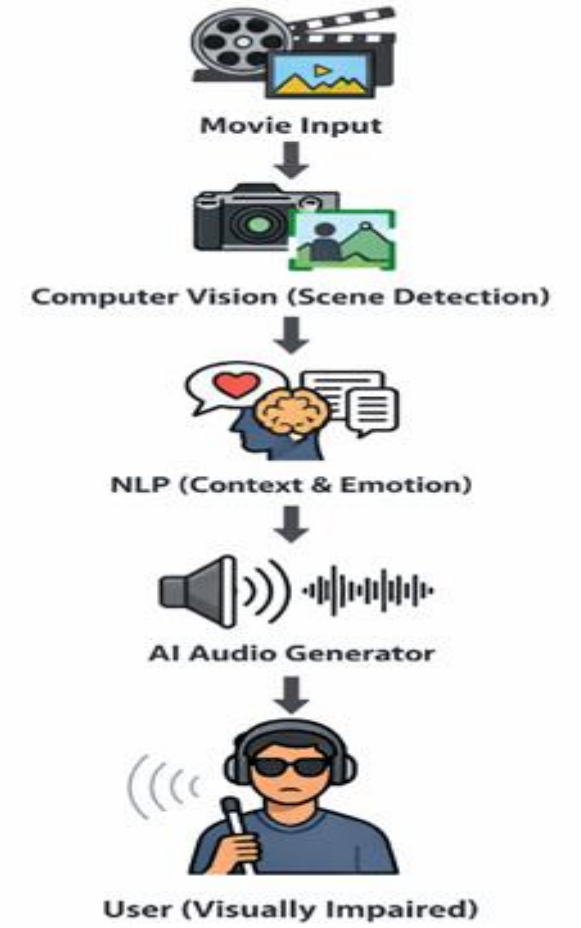

Figure 2: System Workflow

IV. DISCUSSION

The proposed CineFeel system illustrates how artificial intelligence is going to make the entertainment accessible and more entertaining, especially to the visually challenged. The system can analyze visual movie content through computer vision, natural language processing (NLP), and audio generation applications using AI in order to transform them into meaningful audio experience so that users may have a clearer understanding and enjoy films.

The algorithm relies on convolutional neural networks (CNNs) in scene detection. This aids in identification of items in movie frames and the characters and the surrounding environment. This has the advantage of providing the user with expanded accounts compared to the normal audio narration. Moreover, including the

context and analysis of emotions, the system will not only be capable of communicating what transpires in a scene, but also the emotional touch. This simplifies the process of following the story by the users.

Without the natural language processing, the transformation of visual information into the definite and systematized text descriptions is impossible. The text to speech (TTS) AI-based technology is used to transform these texts into natural speaking sounds. This combination can be said to make the delivery of narration readable and expressive and in sync with the film hence making it a much better movie experience. The multisensory method CineFeel uses enhances the interest of the users too. Haptic feedback and emotion-sensitive audio narration is more immersive than the normal audio description system. Such possibilities can render it extremely convenient and more embracing online entertainment.

However, challenges remain. High-resolution video demands a lot of processing power to do real-time processing thus this can impact the efficiency of the system. The feeling in the film scenes is also hard to realize relying upon the situations of the light, the angles of the camera, and the acting technique. The fact of retaining the narration in correspondence with the movie playback is the next valuable feature that should be addressed with attention.

Regardless of these concerns, CineFeel has good prospects as a cinema help technology. It is possible that the next generation has new capabilities, including better techniques to recognize emotions, numerous languages to speak with, adjustable degree of description, and assistance to multisensory communication. Overall, this system shows that AI-based assistive technologies can be employed to lessen access barriers in the entertainment industry and make the experience of using the system more accommodating to the visually impaired individuals.

4.1. Benefits of the Proposed System.

1. Better Accessibility: The system assists the visually impaired people to enjoy movies by getting a visual content described in detail through audio.
2. Emotion-Based Narration: The emotional context will make the narration interesting and will also allow the users to comprehend the storyline more.
3. Scene Understanding in Real Time.: AI makes it possible to analyze the scenes dynamically, providing timely and appropriate descriptions of the audio.
4. Enhanced User Experience: Expressive audio and a potential haptical feedback increases the cinematic experience due to multisensory input.
5. Customization Options: The users will be able to change the pace of narration, the level of details, and feedback depending on their comfort.

4.2. The Proposed System has the following limitations.

1. High Processing Requirements: Real time video frame analysis takes good computational power.
2. The Problems of Emotion Detection: It may be challenging accurately determine the emotions in films, because of the changes in lighting, camera tricks, and the styles of the actors.
3. Synchronization Issues: Technically, it is difficult to provide the audio descriptions that align with movie scenes.
4. Dependence on Training Data: The quality and diversity of datasets that are used in training is a major determinant of the system performance.
5. Additional Hardware Needs: There are certain multisensory aspects that might need additional tools such as haptic wearables.

V. CONCLUSION

CineFeel system demonstrates the way in which Artificial Intelligence can be significant in enhancing accessibility within the entertainment sector, particularly among the visually impaired persons. The system involves transformation of the visual movie scenes to meaningful audio descriptions by integrating Computer Vision, Natural language Processing, analysis of emotions, and AI-based audio description. This makes movies more accessible to the visually impaired users as they are able to follow them and their storyline, emotions, and general feeling of a cinematic experience. The system also engages the users more than traditional audio description which is a multisensory approach taken by the system. The use of emotion sensitive narration and contextual understanding of the scene gives it a more involving and richer experience. The developments underscore the role of AI in aiding technologies and accessible digital media.

Despite the fact that there are still certain challenges, including real-time processing requirements, complexity of emotion detection, and synchronization problems, the system demonstrates a high potential of further evolution. CineFeel can be further enhanced in the future with such features as multilingual narration, enhanced emotion recognition, and the possibility of combining it with wearable feedback device to become even more effective. Altogether, this paper has shown how the AI-based assistive technologies can be used to ensure the accessibility and enhance the quality of life of the visually impaired people making the movie viewing process more exciting and accommodating.

REFERENCES

[1] K. Patil, A. Kharat, P. Chaudhary, S. Bidgar and R. Gavhane, "Guidance System for Visually Impaired People," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 988-993, doi: 10.1109/ICAIS50930.2021.9395973.

[2] V. R. Algazi and R. O. Duda, "Immersive spatial sound for mobile multimedia," Seventh IEEE International Symposium on Multimedia (ISM'05), Irvine, CA, USA, 2005, pp. 8 pp.-, doi: 10.1109/ISM.2005.69.

[3] I. Hwang, H. Lee and S. Choi, "Real-Time Dual-Band Haptic Music Player for Mobile Devices," in IEEE Transactions on Haptics, vol. 6, no. 3, pp. 340-351, July Sept. 2013, doi: 10.1109/TOH.2013.7.

[4] D. Shi, X. Xu, F. Sun, Y. Shi and N. Cao, "Calliope: Automatic Visual Data Story Generation from a Spreadsheet," in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 453-463, Feb. 2021, doi: 10.1109/TVCG.2020.3030403.

[5] L. P. Khan, V. Gupta, S. Bedi and A. Singhal, "StoryGenAI: An Automatic Genre-Keyword Based Story Generation," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 955-960, doi: 10.1109/CISES58720.2023.10183482.

[6] Cho, J.-D. Multi-Sensory Interaction for Blind and Visually Impaired People. Electronics 2021, 10, 3170. https://doi.org/10.3390/electronics10243170.

[7] L. Arora, A. Choudhary, M. Bhatt, J. Kaliappan, and K. Srinivasan, “A comprehensive review on NUI, multi-sensory interfaces and UX design for applications and devices for visually impaired users,” Frontiers in Public Health, vol. 12, Oct. 2024, doi: https://doi.org/10.3389/fpubh.2024.1357160.

[8] Z. Xie, Z. Li, Y. Zhang, J. Zhang, F. Liu, and W. Chen, “A Multi-Sensory Guidance System for the Visually Impaired Using YOLO and ORB-SLAM,” Information, vol. 13, no. 7, pp. 343–343, Jul. 2022, doi: https://doi.org/10.3390/info13070343.

[9] G. Yang and J. Saniie, “Sight-to-Sound Human-Machine Interface for Guiding and Navigating Visually Impaired People,” IEEE Access, vol. 8, pp. 185416–185428, 2020, doi: https://doi.org/10.1109/access.2020.3029426.

[10] D. J. Brown and M. J. Proulx, “Audio–Vision Substitution for Blind Individuals: Addressing Human Information Processing Capacity Limitations,” IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 5, pp. 924–931, Aug. 2016, doi: https://doi.org/10.1109/jstsp.2016.2543678.

[11] K. J. Piczak, "Environmental sound classification with convolutional neural networks," 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA, USA, 2015, pp. 1-6, doi: 10.1109/MLSP.2015.7324337.

[12] L. A. Jones and H. Z. Tan, "Application of Psychophysical Techniques to Haptic Research," in IEEE Transactions on Haptics, vol. 6, no. 3, pp. 268-284, July-Sept. 2013, doi: 10.1109/TOH.2012.74.

[13] S. Jagadish, M. Lalitendra, K. Nikhita, P. Narsingarao, and Lolla Sreedhar, “Artificial Intelligence-Powered Mobile Application to Help Visually Impaired People,” International journal of innovative science and research technology, vol. 9, no.4, pp.11001105, Apr.2024, doi: https://doi.org/10.38124/ijisrt/ijisrt24apr1577.

[14] T. Qureshi, M. Rajbhar, Y. Pisat, Vijay Bhosale, and Student, “AI Based App for Blind People.” Available:

https://www.irjet.net/archives/V8/i3/IRJET-V8I3554.pdf

[15] M. Daraban, K. Lengyel, A. Faluvegi and C. Amariei, "Real-time Spatialized Sound Generator on Embedded Wearable Platform," 2021 IEEE 27th International Symposium for Design and Technology in Electronic Packaging (SIITME), Timisoara, Romania, 2021, pp. 71-76, doi: 10.1109/SIITME53254.2021.9663671.

[16] V. Pujari, K. Madnal and D. Premchandran, "Mobile App for Enhancing Accessibility Among the Visually Impaired," 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), Wardha, India, 2024, pp. 1-6, doi: 10.1109/IDICAIEI61867.2024.10842749.

[17] Abboud S, Hanassy S, Levy-Tzedek S, Maidenbaum S, Amedi A. EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution. Restorative Neurology and Neuroscience. 2014;32(2):247-257. doi:10.3233/RNN-130338

[18] V. Hayward, O. R. Astley, M. Cruz-Hernandez, D. Grant, and G. Robles-De-La-Torre, “Haptic interfaces and devices,” Sensor Review, vol. 24, no. 1, pp. 16–29, Mar. 2004, doi: https://doi.org/10.1108/02602280410515770.

[19] S. Paneels and J. C. Roberts, "Review of Designs for Haptic Data Visualization," in IEEE Transactions on Haptics, vol. 3, no. 2, pp. 119-137, April-June 2010, doi: 10.1109/TOH.2009.44.

[20] Ruff, R. M., & Perret, E. (1982). Spatial Mapping of Two-Dimensional Sound Patterns Presented Sequentially. Perceptual and Motor Skills, 55(1), 155 163. https://doi.org/10.2466/pms.1982.55.1.155

[21] Roginska, A., & Geluso, P. (Eds.). (2017). Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio (1st ed.). Routledge. https://doi.org/10.4324/9781315707525

[22] Maruricio Lumbreras and Jaime Sánchez. 1999. Interactive 3D sound hyperstories for blind children. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '99). Association for Computing Machinery, New York, NY, USA, 318–325. https://doi.org/10.1145/302979.303101

[23] R. Gunther, R. Kazman, and C. MacGregor, “Using 3D sound as a navigational aid in virtual environments,” Behaviour & Information Technology, vol. 23, no. 6, pp. 435–446, Nov. 2004, doi: https://doi.org/10.1080/01449290410001723364.

[24] N. M. S. Eatesh, V. D. Vemuru, R. R. Muga, N. Dhanush, I. C and N. K, "AI Powered Visual Aid System for the Blind," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RMKMATE59243.2023.10368901.