Volume 12, No.9, September 2025

Journal of Global Research in Mathematical Archives

ISSN 2320 – 5822

UGC Approved Journal

RESEARCH PAPER

Available online at http://www.jgrma.info

A SURVEY OF NATURAL LANGUAGE AND COMPUTER VISION METHODS FOR PRESENTATION ENHANCEMENT

Mr. Deepak Mehta[1]

[1] Assistant Professor, Mandsaur University, Mandsaur, Department of Computer Sciences and Applications
deepak.mehta@meu.edu.in

Abstract: Natural Language Processing (NLP) and Computer Vision (CV) are revolutionizing the presentation user experience for both designers and presenters. This systematic review examines how NLP and CV are used together to increase the quality of presentation, their interactivity, and engagement with the audience. NLP enables features such as on-demand speech recognition, translation, content summarization, and communication with chatbots, allowing presenters to interact effectively with diverse audiences. At the same time, "CV capabilities" support gesture understanding, tracking of facial expression, eye-tracking and augmented reality display, allowing the person presenting their material to assess the interest of the audience and adjust presentation accordingly. The combination of the technologies involved facilitates the creation of smart presentation aid and reactive systems that dynamically fit and customize the presentation process according to the live response. This review classifies and critically analyzes the recent works' contributions in the two fields, which find application in education, business and conjecture situations in the field of public speaking. It also highlights the increasing need to design multimodal AI systems to combine linguistic and visual types of communication and interaction. The study concludes by delineating critical methodologies, tools and system architectures behind these developments, which point to the revolutionary impact NLP-CV unification may have in the future of intelligent, user-focused presentation systems.

Keywords: Natural Language Processing (NLP), Computer Vision (CV), Presentation Enhancement, Adaptive Systems, Intelligent Assistant, Human–Computer Interaction (HCI).

1 INTRODUCTION

One of the most important segments of academic and professional communication is taken up by conference presentations. They provide an avenue for researchers to exchange existing knowledge, enter into intellectual dialogue, and nurture partnerships. Academic conferences are greatly appreciated irrespective of the related requirements in regards to preparation, travelling, and budget, one of the reasons being due to the currency of the information being presented, which is usually before journal publications are accessible to them through peer reviews, which may take months or years to see print [1]. The use of Artificial Intelligence (AI) in its specific components such as Computer Vision (CV) and Natural Language Processing (NLP) has been exploited over the past couple of years as a way of revamping the presentation experience [2]. It is changing the way of creating, shipping presentations, and the way they are thought of appearing on the scene, as it allows in real-time feedback, intelligence creation of content, and multimodal interaction.

Computer vision, as one of the fundamental subdomains of AI, helps to interpret and process visual data through images and video materials [3]. It has been developing with the success of DL and NN and has suffered from breakthroughs in image recognition, object detection, and gesture analysis. The features are currently used in ways other than state-of-the-art in areas like healthcare, automotive, and security to improve human-computer interaction in a real-time presentation setting [4]. CV methods support gesture recognition and analysis, facial tracking, eye-tracking, and visual content enhancement so that a presenter can see and react to audience attention in real time [5].

NLP has grown considerably, and models can now can-do speech recognition, summarization, machine translation and even audience interaction through the use of chatbots. NLP interprets human language and comes up with results that can be understood by human and machines [6]. This allows real-time subplot typography, content individualization and multilingual access when making live presentations. The development of monolingual as well as multilingual NLP technologies has also increased the range of inclusive communication, particularly in terms of the diverse global and educational contexts [7].

The intersection of CV and Natural NLP allows the creation of intelligent presentation systems that aid speakers, engage the audience and even automate the generation of material. Such integration is transforming the field of communication in academic, business, and public spaces, offering a dynamic and personal experience. Previous research at the intersection of NLP and CV is reviewed, and methodologies are classified; use cases are analyzed, and tools and system architectures compared. It also highlights

the growing integration of multimodal machines with artificial intelligence, transforming conventional presentations into interactive user experiences.

1.1 Structure of the paper

The paper is organized as follows: Section II explores NLP-based presentation enhancements like language translation, preprocessing, and virtual assistants. Section III discusses CV techniques such as image enhancement, audience sentiment analysis, and object detection. Section IV covers the integration of NLP and CV in multimodal systems for real-time feedback and adaptive presentations. Section V reviews related literature, and Section VI concludes with key findings and future directions.

2 NLP TECHNIQUES FOR PRESENTATION ENHANCEMENT

Effective presentation involves more than just delivering content it requires engaging the audience, managing emotions, and conveying messages clearly. Natural Language Processing (NLP) techniques can significantly enhance these aspects by helping presenters structure their communication, adapt to audience responses, and convey ideas more persuasively[8]. Below are key NLP-based strategies that contribute to improving presentation effectiveness[9]:

- A meta model is a set of targeted meta-questions for identifying limiting beliefs. Studying language models to identify unconscious limitations, violations, and generalizations is fundamental. Each of the twelve models uses a unique set of questions to tease out limiting beliefs.
- Negative words about a person or thing can be removed through reframing, which is also known as changing the frame. To help the other person reframe their perception of a certain situation, this method is frequently employed.
- The presenter has a better chance of swiftly engaging the audience and winning their sympathy and support if he uses this strategy. Because it can force the practitioner out of their "comfort zone," alignment is best done gradually rather than all at once.
- Calibration helps people become more self-conscious and socially aware. When calibrating, it's crucial to monitor both yourself and your opponent, adjusting to your mental and physiological states to effectively read their body language.
- Indirect communication is achieved through the use of stories and metaphors. Finding common answers is where the technique is put into action.

These NLP techniques collectively enhance presentation skills by promoting clearer communication, emotional intelligence, and deeper audience engagement. When integrated thoughtfully, they support a more compelling and persuasive delivery style.

2.1 Techniques for Preprocessing Examinations

A primary focus of preprocessing approaches is cleaning and translating data into a format that is acceptable for further processing [10]. Among these methods are the following: deleting punctuation and on-Unicode characters; removing diagrams and symbols; removing descriptive language before the questions; and removing words with less than three letters. Tokenization (the process of dividing text into smaller pieces), phrase segmentation, stop-word removal, parsing, part-of-speech (POS) tagging, and converting letters to lowercase are further ways. Using lexical and contextual information, most of these strategies deal with the appearance of word formations, while later processing phases address pragmatic and semantic elements. Reducing word form diversity to a consistent style is another goal of normalization. Lemmatization, on the other hand, is superior for question classification since it identifies legitimate root words—essential for extracting semantic information from test questions.

2.2 Language Translation and Multilingual Support

NLP technologies applied to this consumer market data provided immediate insights into the benefits and disadvantages of products. In a short amount of time, and able to acquire valuable knowledge because to the models' vast and easily comprehensible data resources. From 2016 to 2019, the TripAdvisor organization collected over 50,000 reviews from satisfied customers. This data was used to provide insights into the common sources of airline service. When it comes to analyzing different areas of the firm, consumer insights paint a clearer picture [11]. For rapid insights, it offers a complete data repository that is both easy to use and fast. The reviews written by actual customers are crucial for making adjustments and enhancements according to the needs of the target audience. In this case, natural language processing (NLP) is employed to examine all of the data. NLP is a combination of AI and ML that aids in reviewing various feedbacks. As with other problems involving the classification of text in natural language, this study discusses the difficulties of mapping customer verbatim. Utilize sources such as the quality office's transactional system, social media conference transcripts, and other relevant documents.

2.3 Chatbots and Virtual Assistants for Presentations

The individualized help, information retrieval, and task automation provided by chatbots and voice assistants have made them indispensable components of the digital environment. A number of popular voice assistants use NLP to decipher user questions posed in a conversational tone and respond appropriately or carry out specific tasks. Similarly, NLP is used by chatbots on websites, mobile apps, and messaging platforms to carry on conversations with customers, answering questions, and helping with customer

service and transactions [12]. Not only do these technologies make things easier for users, but they also make businesses more efficient. The capacity to understand context, speak many languages, and even exhibit emotional intelligence are all hallmarks of the most current AI advancements.[13]. Consequently, voice assistants and chatbots are becoming proactive, which means they predict the needs of the user. Their combination with other modalities such as computer vision holds promise to be even more immersive and intuitive user interactions in future.

3 COMPUTER VISION TECHNIQUES FOR PRESENTATION ENHANCEMENT

The computer vision would be crucial in enhancing clarity and understanding of images incorporated in the presentation material and mainly in technical matters like the analysis of antenna components [14]. The techniques increase the legibility of images, suppress noise, and highlight important details; thus, they are better used in defect identification and visual conveying.

- **Image Enhancement:** The aim of image enhancement techniques is to improve the visual quality of photographs of antenna components by adjusting their contrast, brightness, and other visual attributes.
- **Noise Reduction:** The methods of reducing the noise in antenna component images aim at increasing the signal-to-noise ratio, which may help in preventing unnecessary obstruction of adequate fault detection and classification.
- **Image Enhancement and Noise Reduction:** Image preprocessing methods, including as picture improvement and noise reduction, are crucial for capturing high-quality, clear images of antenna components. To improve an image's suitability for analysis and interpretation, image enhancement techniques are employed.
- **Image Segmentation for Region-of-Interest Extraction:** Using image segmentation algorithms, it is possible to locate and extract ROIs from pictures of antenna components. To effectively identify and categorise defects, it is crucial to narrow the scope of the study.

3.1 Architecture of Audience Feedback Analysis

A combination of text, images, and audio is used in audience feedback analysis to measure engagement and sentiment. NLP and computer vision (CV) extract important features, which are then processed by ML models to give real-time insights that aid in making presentations more effective. In Figure 1, can see the project's structural blueprint. The following is a detailed outline of the suggested method's procedure [15]:

- The application receives either a recorded video file or a live video stream.
- The frame extraction program extracts frames from the input video based on the administrator's requirements. The administrator can define the time interval between each frame. In addition, the administrator can set a maximum number of frames for analysis and assessment purposes.

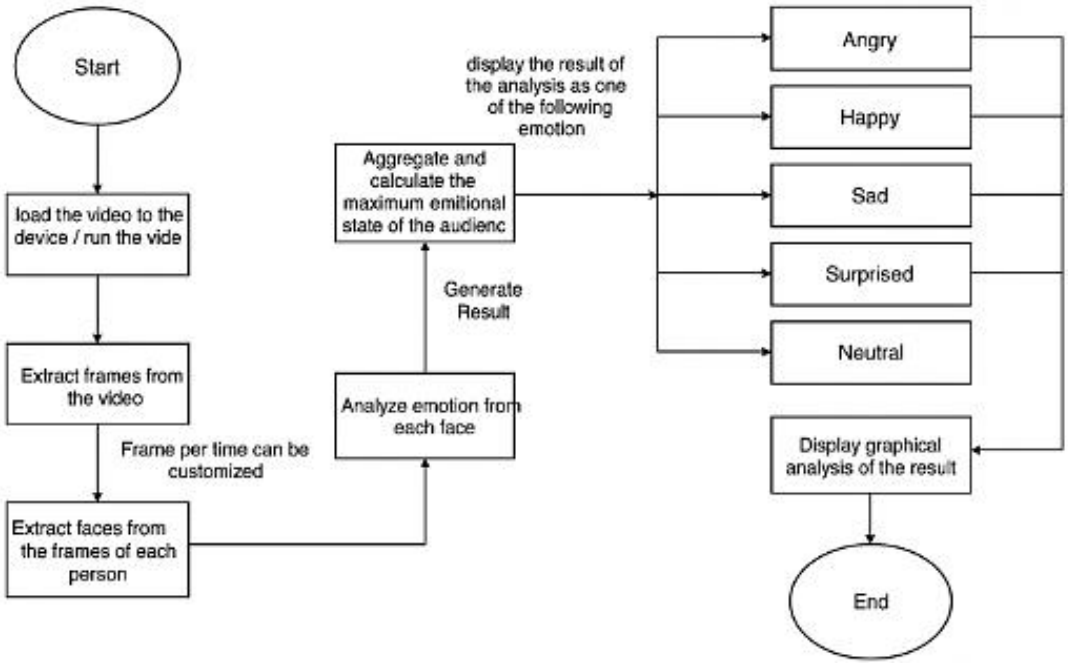

Figure 1: Structural Architecture of Audience Feedback Analysis

- The faces of every single person are cropped or eliminated from every single frame that is extracted. The cropping of a person's face happens regardless of the frame it appears in; it happens in every frame.
- Emotions are labelled and saved in a distinct evaluation area for every clipped face.

- The Project's output is shown to the Administrator based on the Calculation, which indicates the most common or average emotion of the gathering.

3.2 Application of CV in Ensuring Personnel Safety During Mining

Safety of workers is very important in mines. Also, having someone monitor the situation can help ensure that manufacturing is safe. The continuity of the entire production process is severely compromised in the event of a personnel safety mishap, which would lead to a six-month to one-year suspension of production [16]. In the past, designated safety officers were required to check on employees around the clock, ensuring they were safe on the job and properly equipped. But as computing power increases, CVs can take over most of the duties of safety inspectors, with better monitoring and continuous monitoring making a significant impact on safe production [17]. Researchers have found that the likelihood of an accident increases with rising fatigue levels; consequently, they have developed a system to monitor employees' weariness, combining CV algorithms with traditional fatigue detection methods (FMD). Chen et al., for instance, used a marine predator algorithm-optimized LSSVM to increase the model's accuracy by 18.68%.

3.3 Object Detection and Augmented Reality Overlays

Applications of augmented reality, like scalable item identification in real-time and assembly guidance, are vital. Diverse Deep Neural Network (DNN) models have already tackled the problem of real-time object recognition from RGB photos. Most AR/MR systems can currently understand their 3D surroundings, but they still struggle to perceive and categorize complex real-world items. To incorporate these characteristics, Deep Convolutional Neural Networks (CNNs) are necessary; yet, it is still challenging to run big networks on mobile devices [18]. With such demanding requirements for zero end-to-end latency and pinpoint precision, relying on the cloud or edge for object detection is an intimidating prospect. For embedded devices to be better able to recognize objects, it suggests an object detector.

4 INTEGRATION OF NLP AND CV: MULTIMODAL SYSTEMS

The integration of NLP and Computer Vision (CV) has emerged as a powerful approach for achieving deeper and more holistic content understanding. While NLP excels in extracting meaning from text and CV in interpreting visual data, their combination enables systems to process and reason across both modalities simultaneously[19]. Multi-modal transformer frameworks, especially, have proven to be very beneficial in comparison to single-modality models [20]. Such systems can interpret subtle connections between text and imagery, such as correlating textual statements with corroborating or conflicting images, which is crucial for sophisticated activities like misinformation identification, content summary, and adaptive output production. This cross between modalities draws the AI systems a step closer to human-level understandings, with more contextual and semantically rich interpretation possible. Consequently, it provides novel directions to the creation of smarter apps in spheres such as education, media and human to computer interaction, where each visual and linguistic perception is essential.

4.1 Computer Vision in Education

Computer vision technology is critical to real-time student engagement and monitoring of behaviour. The computer vision systems track the emotions and focus of students during the educational session by analyzing facial expressions, patterns of body postures, and eye movement. Figure 2 shows the functionality of eye-tracking technology as a constituent part of a CV-based system, since this technology makes use of camera measurements of pupil and cornea reflections to ascertain the precise direction of the gaze [21]. The data collected assists the system on discriminating whether the students are concentrating on what they are learning or are distracted. Through consistent monitoring, the system adjusts instructional content and pace to reflect the learner's new levels of engagement. This system is sensitive to the distraction or fatigue of students and provides interactive learning content or suggests taking short breaks to maintain attention and motivation.

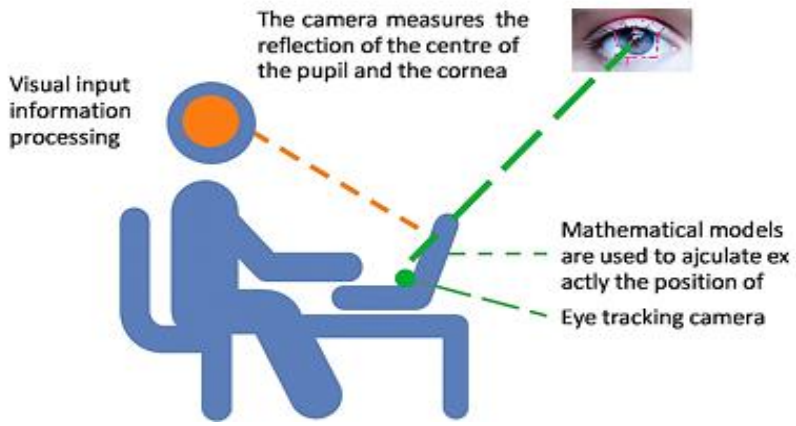

Figure 2: Real-Time Eye-Tracking System for Monitoring Student Engagement in Online Learning

4.2 Real-time Feedback and Adaptive Presentations

The NLP and CV are applied in adaptive presentation systems and feedback systems to determine the engagement of the audience in a real-time manner and therefore adjust the content presented [22]. CV is used to determine audience interaction with concepts such as facial expressions, eye movement and sentiment along with NLP for determining audience interaction with speech patterns and questions in real-time, through adaptive presentation systems. Monitoring such user signals, the presentation could be adapted by a presenter (e.g., accelerating, decelerating or paying more visual attention to one thing or another). Adaptive presentations help to improve interaction and understanding of the user and customize the activities according to the audience outlook. Real time feedback is the reason that the presenters can respond immediately to the cues the audience is putting forward, and encourages interactive and productive communication. Further development enhance precision, support integrity of heterogeneous data, and make the process of interaction between humans and machines smooth.

4.3 Intelligent Presentation Assistants

Intelligent presentation assistants are AI-driven devices that help deliver presentations and make the delivery of these presentations easier, and accomplish some operations, as well as optimize content delivery. With the NLP and CV there are many things that intelligent assistants are capable of doing with speech transcription, automated slides, summarization, and computer vision to track the audience engagement. They are also capable of making contextual recommendations and visual cues to the presenter, and also present non-real time analytics to the presenter to capture improved efficiency in presentation. Intelligent assistants are capable of offering voice-controlled initiation without the use of hands by the use of gestures and responsive feedback. Such tools decrease mental strains on presenters and provide a more engaging audience through personalization experience [23]. With the ongoing development of the technology, intelligent presentation assistants probably offer more multimodal insight and foreseeing to enhance the quality of the presentation and user satisfaction in the long run.

5 LITERATURE OF REVIEW

This literature Summary focuses on variously progress in XR, AI-driven recommender, generative AI, relation extraction, NLP, and vision-based transformers, with a focus on methodological solutions, discoveries, and open issues, and contemporary limitations as well as future steps to achieve enhancements and optimizations of intelligent, interactive, and domain-specific applications.

Sharma *et al.* (2025) explore the game-changing impact of deep learning on AI-powered interview platforms. As a means of better assessing prospective pupils, it is focused on crucial techniques including Natural Language Processing (NLP), Computer Vision, and Speech Recognition. In spite of all these technological advancements, there is still a long way to go before we can fully trust them to understand human speech, provide adequate context, and act ethically. Both natural language processing (NLP) and computer vision (CV) can decipher candidates' intentions and thoughts from their body language and important statements. An essential function of speech recognition technology is its ability to accurately capture candidates' speech patterns. Nevertheless, to take full advantage of these systems have to solve current issues, which include, easy integration of these technologies, protect the privacy of the user, and have fairness and transparency in AI [24].

Supriyono *et al.* (2024) explores the new developments in NLP and the related implications, issues, and further perspectives. Owing to the ever-growing amount of text data being produced on a daily basis using various diverse sources, identifying useful and relevant information is getting more complicated. Traditional methods of reception and analysis of written information, which are based on manual work, can be long and prone to errors, which highlights the need to find a successful automated variant. Recent progress in NLP, that is, the transformer-based models and the deep learning approach, has been shown to hold great potential to enhance the accuracy and stability of a wide array of NLP tools. This work also introduces a new technique that takes a systematic review approach and combines it together with advanced methods in NLP to make the NLP systems more efficient, in general [25].

Hussain (2023) a detailed insight into the vast universe of Artificial Intelligence, paying much attention to the fields of generative AI and computer vision. aim to give businesses the finer and deeper insights on these essential AI subfields. Through this, organizations are equipped with the knowledge and strategies of how to adopt generative AI and computer vision technologies in their organizations. The end result want to achieve with that is the provision of knowledge and insight to businesses that can then use the potential of these frontiers of AI both to be innovative and to strengthen their competitive advantage in an ever more technological world [26].

Sonbol *et al.* (2022) performed a survey as a systematic literature mapping (classification) to identify the representations employed in RE tasks literature, to evaluate the trend of relevant research, to identify research directions in this field, to identify gaps in these works, and to determine future directions. The total number of papers found was 104, down from 2,227 at the outset, after a number of inclusion and exclusion criteria were applied. study has been moving in the direction of using lexical features and syntax to advanced embedding methods, particularly in the last two years, according to the survey. This is because embedding advanced representations have been demonstrated to be useful in most RE activities [27].

El-Komy *et al.* (2022) concentrate on using natural language processing to decipher words using computer vision. A lot of robotics applications employ this idea. With the suggested integrated application, people with visual impairments can be guided by robots that can perceive their environments using natural gestures and spoken languages, even though robots should be able to perceive their environments through other means of interaction. Because sight is fundamental to human survival, finding a substitute means of guidance for the visually impaired is crucial. This paper aims to address this by utilising a blind person's attached smartphone—which possesses vision, language, and intelligence—to take pictures of their environment. These pictures are then sent to a central server that uses a F-RCNN to identify objects in the images, both to inform the person about them and to help them avoid obstacle [28].

Nagarhalli *et al,* (2021), learning methods have been essential in making NLP more efficient. The study of AI and its related areas, including as ML, DW, and NLP, has recently shown phenomenal expansion in terms of both funding and quality of research. The broad range of applications and the availability of powerful computer devices at lower costs have sparked a new wave of enthusiasm in these areas of study. As a whole, artificial intelligence and its subfields have had a favorable effect on almost every industry in the modern economy [29].

Table 1 summarizes key studies on AI-driven language and vision techniques, highlighting methodologies, findings, challenges, and future directions relevant to enhancing interactive, intelligent presentation systems across diverse applications.

Table 1: Summary of a Study on Natural Language and Computer Vision Techniques for Presentation Enhancement

Author	Study On	Approach	Key Findings	Challenges	Future Directions
Sharma et al. (2025)	Application of Deep Learning to AI-Powered Interview Bots	Assessment of prospective students using a combination of NLP, CV, and speech recognition	NLP detects emotional tone and key points- Computer Vision interprets body language. Speech Recognition captures verbal communication accurately	Ethics and transparency, Accurate understanding of human communication, Integration of multi-modal technologies	Better integration of AI components- Enhanced ethical guideline. Improved privacy, fairness, and explainability in decision-making systems
Supriyono et al. (2024)	Implications of recent developments in natural language processing	Systematic review combined with transformer-based models and deep learning techniques to improve NLP applications	NLP advancements improve precision and consistency. Automation helps address complexity and volume of modern text data	Managing large-scale, unstructured data. Limitations of conventional NLP approaches	Adoption of more scalable transformer model. Continued development of efficient and accurate automated systems
Hussain, (2023)	Business-Oriented Generative AI and Computer Vision	Exploratory review focused on business applications	Empowers strategic adoption; supports innovation and competitiveness	Limited understanding of domain-specific needs	Design tailored solutions combining generative AI and CV for enterprise use
Sonbol et al., (2022)	Relation Extraction (RE) Techniques	Systematic mapping; 104 papers post-filtering from 2,227	Shift from lexical features to embedding-based techniques	Need for domain-specific embeddings; lack of benchmark datasets	Develop domain-adaptive RE models with interpretable embeddings
El-Komy et al. (2022)	Providing assistance to the visually challenged by combining computer vision with natural language processing	Smartphone-based assistive system using F-RCNN for object detection, capturing real-world images and providing audio feedback	Successful real-time object detection and description. NLP and vision systems improve mobility and environmental awareness for blind users	Real-time performance constraints. Reliability of detection in varying environments	Enhancing robustness of the system- Broader implementation in wearable devices. Further refinement of AI-human interaction for assistive technology
Bi et al., (2021)	Vision-based Transformers	Survey of 15 articles; comparison of object detection, tracking, etc.	Transformers excel in CV tasks; validated by experiments	High computational cost; model interpretability	Optimize vision transformers for real-time tasks; improve

					efficiency and interpretability

6 CONCLUSION AND FUTURE WORK

Natural Language Processing (NLP) and Computer Vision (CV) are revolutionizing presentation delivery through intelligent automation, real-time viewer feedback, and adaptive content interaction. NLP features such as speech recognition and multilingual translation enhance communication and editor engagement. Concurrently, CV techniques—like gesture recognition, facial expression analysis, gaze tracking, and AR overlays—enable presenters to adjust content based on audience response. Together, NLP and CV contribute to intelligent presentation assistants that improve user experience, cognitive efficiency, and communication effectiveness. These multimodal systems are increasingly applied in education, business, and public speaking, offering more interactive and immersive presentations. Despite advancements, challenges persist. High computational demands limit use on low-resource and mobile devices, and there is a lack of standardized performance benchmarks. Additionally, concerns around data privacy, informed consent, and algorithmic bias hinder large-scale deployment and consumer trust.

Future development should focus on effective multimodal fusion strategies, lightweight real-time models for edge and mobile deployment, and integration of emotional intelligence, personalized learning profiles, and predictive analytics. Ethical considerations must be built into system design, supported by clear data governance policies. Collaboration among academia, industry, and educators is essential to ensure innovation that is inclusive, reliable, and responsibly deployed.

REFERENCES

[1] S. T. Skidmore, J. R. Slate, and A. J. Onwuegbuzie, “Developing effective presentation skills : evidence-based guidelines,” *Res. Sch.*, vol. 17, no. 2, pp. xxv–xxxvii, 2010.

[2] B. Patel, H. Mallisetty, and K. M. Rao, “Artificial Intelligence Helper Application for Delivering Effective Presentations,” *Publ. by Tech. Discl. Commons*, 2024.

[3] H. Mallisetty, B. Patel, and K. M. Rao, “Artificial Intelligence Assisted Online Interactions,” *Tech. Discl. Commons*, 2023.

[4] S. M. Shelke, I. S. Pathak, A. P. Sangai, and D. V Lunge, “A Review Paper on Computer Vision,” vol. 3, no. 2, pp. 673–677, 2023, doi: 10.48175/IJARSCT-8901.

[5] N. S. Thakur *et al.*, “Advancements in Computer Vision Techniques and Applications,” *J. Emerg. Technol. Innov. Res.*, vol. 12, no. 2, pp. 121–124, 2025.

[6] B. Priya, J. M. Nandhini, and T. Gnanasekaran, “An analysis of the applications of natural language processing in various sectors,” *Adv. Parallel Comput.*, vol. 38, pp. 598–602, 2021, doi: 10.3233/APC210109.

[7] A. R. Jafari, B. Heidary, R. Farahbakhsh, M. Salehi, and N. Crespi, “Language Models for Multi-Lingual Tasks - A Survey,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 6, pp. 1458–1472, 2024, doi: 10.14569/IJACSA.2024.01506146.

[8] S. Rongala, S. A. Pahune, H. Velu, and S. Mathur, “Leveraging Natural Language Processing and Machine Learning for Consumer Insights from Amazon Product Reviews,” in *2025 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*, 2025, pp. 1–6. doi: 10.1109/ICSSES64899.2025.11009528.

[9] D. Ilieva-koleva and R. Vazov, “Neuro-Linguistic Programming Techniques for Perfecting Presentation Skills,” *Sel. Pap. Present. 3Rd M-sph. Int. Conf. Multidiscip. Sci. Bus.*, no. October 2014, pp. 30–37, 2014.

[10] E. A. E. Lukwaro, K. Kalegele, and D. G. Nyambo, “A Review on NLP Techniques and Associated Challenges in Extracting Features from Education Data,” *Int. J. Comput. Digit. Syst.*, vol. 16, no. 1, pp. 961–979, 2024, doi: 10.12785/ijcds/160170.

[11] J. J. Sarvaiya, “Multilingual Text Analysis using Natural Language Processing and Transfer Jinal Jaisukh Sarvaiya National College of Ireland Supervisor : Hicham Rifai,” *Natl. Coll. Irel. Student*, p. 20, 2022.

[12] D. Patel, “AI-Enhanced Natural Language Processing for Improving Web Page Classification Accuracy,” vol. 4, no. 1, pp. 133–140, 2024, doi: 10.56472/25832646/JETA-V4I1P119.

[13] D. N. Yagamurthy, “Advancements in Natural Language Processing (NLP) and Its Applications in Voice Assistants and Chatbots,” *J. Artif. Intell. Cloud Comput.*, vol. 2, no. 4, pp. 1–6, 2023, doi: 10.47363/jaicc/2023(2)326.

[14] Z. Ullah, L. Qi, E. J. S. Pires, A. Reis, and R. R. Nunes, “A Systematic Review of Computer Vision Techniques for Quality Control in End-of-Line Visual Inspection of Antenna Parts,” *Comput. Mater. Contin.*, vol. 80, no. 2, pp. 2387–2421, 2024, doi: 10.32604/cmc.2024.047572.

[15] Madhura Prakash, Aman Kumar, Anmol Saxena, and Somil Agrawal, Twinkle Singh, “Audience Feedback Analysis using Emotion Recognition,” *Int. J. Eng. Res.*, vol. V9, no. 05, pp. 1090–1094, 2020, doi: 10.17577/ijertv9is050717.

[16] D. Shan, F. Qu, Z. Wang, Y. Ji, and J. Xu, “A Review of the Application of Computer Vision Techniques in Sustainable Engineering of Open Pit Mines,” *Sustain.*, vol. 17, no. 7, pp. 1–16, 2025, doi: 10.3390/su17073051.

[17] N. Malali, “Real-Time Liability Monitoring in Annuities Using Actuarial Dashboards on Streaming Data,” *Asian J. Comput. Sci. Eng.*, vol. 8, no. 1, 2023.

[18] M. M. Nafea, S. Y. Tan, M. A. Jubair, and M. T. Abd, “A Review of Lightweight Object Detection Algorithms for Mobile Augmented Reality,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 11, pp. 536–546, 2022, doi: 10.14569/IJACSA.2022.0131162.

[19] S. B. Shah, S. A. Pahune, V. Prajapati, and M. Menghnani, “Efficient NLP-Based Framework for Tweets Sentiment Analysis via Advanced Machine Learning Model,” in *2025 Global Conference in Emerging Technology (GINOTECH)*, IEEE, May 2025, pp. 1–6. doi: 10.1109/GINOTECH63460.2025.11076911.

[20] Manish Kumar Keshri, "The Integration of NLP and Computer Vision: Advanced Frameworks for Multi-Modal Content Understanding," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 2, pp. 2788–2798, 2025, doi: 10.32628/cseit25112708.

[21] S. Lan, "Intelligent Teaching Assistants Combining Computer Vision and Natural Language Processing: Enhancing Personalized Tutoring in Online Education," *Appl. Comput. Eng.*, vol. 138, no. 1, pp. 133–138, 2025, doi: 10.54254/2755-2721/2025.21383.

[22] A. T. Bimba, N. Idris, A. Al-Hunaiyyan, R. B. Mahmud, and N. L. B. M. Shuib, "Adaptive feedback in computer-based learning environments: a review," *Adapt. Behav.*, vol. 25, no. 5, pp. 217–234, 2017, doi: 10.1177/1059712317727590.

[23] A. de Barcelos Silva *et al.*, "Intelligent personal assistants: A systematic literature review," *Expert Syst. Appl.*, vol. 147, p. 113193, 2020, doi: 10.1016/j.eswa.2020.113193.

[24] T. Sharma, S. Singh, A. Singh, and G. Gupta, "Review of natural language processing and computer vision in AI-Powered interview platforms," 2025, p. 020001. doi: 10.1063/5.0289815.

[25] Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, "Advancements in natural language processing: Implications, challenges, and future directions," *Telemat. Informatics Reports*, vol. 16, p. 100173, Dec. 2024, doi: 10.1016/j.teler.2024.100173.

[26] M. Hussain, "When, Where, and Which?: Navigating the Intersection of Computer Vision and Generative AI for Strategic Business Integration," *IEEE Access*, vol. 11, no. October, pp. 127202–127215, 2023, doi: 10.1109/ACCESS.2023.3332468.

[27] R. Sonbol, G. Rebdawi, and N. Ghneim, "The Use of NLP-Based Text Representation Techniques to Support Requirement Engineering Tasks: A Systematic Mapping Review," *IEEE Access*, vol. 10, pp. 62811–62830, 2022, doi: 10.1109/ACCESS.2022.3182372.

[28] A. El-Komy, O. R. Shahin, R. M. A. El-Aziz, and A. I. Taloba, "Integration of Computer Vision and Natural Language Processing in Multimedia Robotics Application," *Inf. Sci. Lett.*, vol. 11, no. 3, pp. 765–775, May 2022, doi: 10.18576/isl/110309.

[29] T. P. Nagarhalli, V. Vaze, and N. K. Rana, "Impact of Machine Learning in Natural Language Processing: A Review," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2021, pp. 1529–1534. doi: 10.1109/ICICV50876.2021.9388380.

