Machine Vision as Environmental Narrator: A Study of Seeing Ai and Google Lookout

Dr. M. Kannadhasan[1], Ms. M. Sruthi Sriee[2]

[1]*Assistant Professor, Department of English, Thiruvalluvar University, Vellore, Tamil Nadu*
[2]*PhD Research Scholar Full–Time, Department of English, Thiruvalluvar University, Vellore, Tamil Nadu*

Abstract—**Recent advances in computer vision have transformed assistive technologies for visually impaired users by enabling real-time narration of surrounding environments. Applications such as Seeing AI and Google Lookout employ machine vision and natural language generation to identify objects, people, text, currency, and spatial relationships, thereby translating visual data into spoken descriptions. This paper examines how these applications function as environmental narrators, mediating human–environment interaction through algorithmic perception. By analysing their scene description capabilities, narrative structures, and representational choices, the study explores how machine vision constructs meaning from everyday environments.**

Using a qualitative comparative methodology, the paper evaluates selected features of Seeing AI and Google Lookout, including object recognition, scene coherence, temporal immediacy, and contextual framing. Particular attention is given to how these systems prioritize certain elements within complex scenes and how such prioritization shapes users' understanding of space, activity, and social presence. The analysis reveals that assistive AI does not merely translate visual information but actively curates' environmental narratives based on data training, probabilistic inference, and linguistic conventions.

The paper argues that these AI-generated narratives reconfigure traditional notions of perception by positioning machines as interpretive intermediaries rather than passive tools. While these technologies significantly enhance accessibility and autonomy, they also raise critical questions regarding accuracy, bias, environmental reductionism, and ethical representation.

Index Terms—**Assistive Artificial Intelligence, Computer Vision, Environmental Narratives, Machine Perception, Accessibility Technologies, Scene Description, Visual-to-Verbal Translation**

Recent advances in computer vision and assistive artificial intelligence have significantly reshaped the landscape of accessibility technologies for individuals with visual impairments.

Many artificial intelligence tasks can be solved by designing the right set of features to extract for that task, then providing these features to a simple machine learning algorithm. For example, a useful feature for speaker identification from sound is an estimate of the size of speaker's vocal tract. It therefore gives a strong clue as to whether the speaker is a man, woman, or child (Goodfellow, Bengio, and Courville 3).

One of the most notable outcomes of these advancements is the emergence of real-time scene narration through applications such as *Seeing AI* and *Google Lookout*. Designed to support visually impaired users, these applications provide continuous verbal feedback about objects, people, text, currency, and spatial relationships within a given environment. "Seeing AI is a Microsoft research project that brings together the power of the cloud and AI to deliver an intelligent app designed to help you navigate your day. Point your phone's camera, select a channel, and hear a description of what the AI has recognized around you." (Microsoft; Google). By offering dynamic descriptions rather than fragmented cues, assistive AI enables users to navigate everyday spaces with increased autonomy and situational awareness.

This paper frames machine vision as an environmental narrator to emphasize the interpretive role these technologies play in mediating perception. Machine-generated descriptions do not simply reproduce visual reality; instead, they select,

prioritize, and linguistically structure environmental elements based on probabilistic models and training data.

The primary objective of the paper is to examine how *Seeing AI* and *Google Lookout* function as environmental narrators through their scene description capabilities. The paper addresses the following research questions: How do these systems construct coherent environmental narratives? What visual elements are prioritized or omitted in real-time narration? What are the implications of such algorithmic mediation for accessibility and perception?

By addressing these questions, the study contributes to broader scholarly discussions on assistive technology, machine perception, and inclusive design, foregrounding both the transformative potential and the ethical challenges of AI-mediated environmental narration.

The concept of *environmental narration* emerges from broader discussions on mediated perception and meaning-making. Perception, particularly in technologically mediated contexts, is not a direct or neutral apprehension of reality but a process shaped by representational systems.

Choice from the availability of modes opens a quite other facility. Modes have differing 'affordances', different potentials for making meaning. The logic of time – of modes that entail movement: speech, dance, action, and gesture – offers different potentials epistemologically and ontologically to modes bound by the logic of space: (still) image, writing (in some ways), colour, layout, 3D objects, etc. (Kress and Selander)

In the context of assistive AI, computer vision systems translate visual environments into verbal descriptions, thereby transforming sensory data into narrative structures that mediate how environments are understood and experienced. Rather than functioning as neutral tools, AI systems operate as interpretive agents that actively structure perception. Philosophers of technology emphasize that technologies do not merely extend human capacities but reshape them. Peter-Paul Verbeek notes that technologies "mediate human experiences and interpretations of the world" (Verbeek 13). Applied to assistive AI, this mediation involves algorithmic

decisions about relevance, salience, and description, positioning machine vision as an active narrator rather than a passive conduit of visual information.

Conceptualizing machine vision within human–environment interaction highlights the relational nature of perception. Environmental understanding emerges through an interaction between users, technological systems, and spatial contexts. As Lucy Suchman observes, technological action is always "…articulating the differences within particular human–machine configurations, expanding our unit of analysis to include extended networks of social and material production, and recognizing the agencies, and attendant responsibilities, involved in the inevitable cuts through which bounded socio material entities are made" (Suchman 6-7). Scene narration produced by assistive AI is therefore contingent on contextual factors such as lighting, movement, training data, and linguistic conventions, reinforcing the idea that machine vision participates in shaping environmental experience.

Ethical and representational concerns further complicate the role of AI as an environmental narrator. AI-generated descriptions may reproduce biases embedded in datasets or reduce complex environments to simplified representations. Kate Crawford cautions that "AI systems are neither neutral nor objective but shaped by the values and assumptions of their creators" (Crawford 8). In assistive contexts, such biases can influence how people, spaces, and activities are described, raising questions about accuracy, inclusivity, and ethical responsibility. This framework thus foregrounds machine vision as a mediating, interpretive force within contemporary assistive technologies.

This paper adopts a qualitative comparative research design to examine how assistive AI applications function as environmental narrators through real-time scene description. Qualitative analysis is particularly suited to this investigation, as it allows for close attention to meaning, representation, and interpretive patterns rather than numerical measurement alone. As Creswell notes, qualitative research is effective when the goal is to understand "how people make sense of their experiences" (Creswell 42).

Seeing AI and *Google Lookout* were selected as case studies due to their widespread adoption, similar core functionalities, and shared objective of supporting visually impaired users through computer vision–based narration. Case study methodology enables an in-depth examination of contemporary phenomena within real-life contexts. According to Yin, case studies are appropriate when "the boundaries between phenomenon and context are not clearly evident" (Yin 18). These applications provide an ideal basis for comparison, as both operate in everyday environments while employing distinct design and narration strategies.

Data collection involved generating and documenting scene description outputs across varied environments, including indoor spaces, public areas, and dynamic social settings. The narrated outputs were treated as textual data for analysis, reflecting how visual information is translated into spoken language. The approach aligns with Kress's assertion that all representations are shaped by the interests of the maker and the affordances of the mode (Kress).

The analysis focused on four key parameters, they are: accuracy, narrative coherence, prioritization, and contextual framing. Accuracy refers to the correctness of object and scene identification; narrative coherence examines how descriptions are structured into meaningful sequences; prioritization analyses which elements are foregrounded or omitted; and context assesses the system's ability to situate objects within broader environmental relationships. Together, these parameters enable a systematic evaluation of environmental narration.

The paper acknowledges certain limitations. The analysis is limited to two applications and does not incorporate direct user interviews or large-scale usability testing. Additionally, scene descriptions may vary based on device hardware, environmental conditions, and software updates.

Assistive AI applications such as *Seeing AI* and *Google Lookout* construct environmental understanding by segmenting complex visual fields into identifiable elements and arranging them into verbal descriptions. Rather than presenting exhaustive detail, these systems generate selective narratives that aim for coherence and usability. As Kress observes, representation is always shaped by

'selection and organization' rather than total reproduction of reality (Kress). Scene narration thus functions as a structured account in which objects, actions, and spatial relations are linguistically ordered to produce intelligible environments.

By using artificial intelligence in inclusive ELT, teachers can overcome conventional limitations and provide special support to students with disabilities. For example, hearing-impaired students can benefit from improved pronunciation and listening skills thanks to AI-based speech recognition and synthesis systems. Similarly, AI-powered adaptive learning plat forms can provide personalized learning materials and pathways that suit the unique requirements of students with different language proficiency levels or learning preferences (M. Kannadhasan 17).

The coherence of these narratives depends on how effectively individual elements are connected into meaningful sequences. When successful, AI narration resembles a situational story, foregrounding actors, actions, and settings in ways that support comprehension and navigation. This narrative structuring underscores the interpretive role of machine vision in mediating environmental perception.

A key feature of AI-generated narration is prioritization—the decision regarding what to describe first and what to omit. Assistive AI often foregrounds people, obstacles, and readable text, reflecting assumptions about functional relevance. Such prioritization aligns with Verbeek's claim that technologies "actively shape what can be perceived and acted upon" (Verbeek 14). The framing of environments is therefore not neutral but guided by design objectives, safety considerations, and training data.

The selective framing influences how users understand and engage with their surroundings. By emphasizing certain elements over others, AI narration establishes a hierarchy of environmental importance, shaping users' situational awareness and interaction. The real-time nature of scene narration introduces a temporal dimension to environmental mediation. Continuous updates allow users to respond dynamically to changing surroundings, reinforcing a sense of immediacy. However, this constant flow of narration also fragments perception into successive descriptive moments. Real-time

mediation thus produces environments that are continuously reconstructed through algorithmic perception. While this enhances responsiveness and safety, it also means that environmental understanding remains provisional, dependent on moment-to-moment computational interpretation.

From an accessibility perspective, AI-generated environmental narration significantly enhances autonomy, mobility, and inclusion. By translating visual information into speech, assistive AI enables users to independently navigate spaces that might otherwise be inaccessible. According to WHO frameworks on disability, accessibility technologies play a crucial role in reducing environmental barriers (World Health Organization).

At the same time, user experience is shaped by the clarity, relevance, and reliability of narration. Overly detailed descriptions may overwhelm users, while insufficient detail may hinder navigation. Effective environmental narration therefore requires a careful balance between informativeness and cognitive load.

Despite their benefits, assistive AI systems raise ethical concerns related to bias and representational reductionism. AI-generated descriptions may misidentify people, oversimplify cultural contexts, or omit socially significant details. Crawford warns that AI systems "encode the values, assumptions, and priorities of the institutions that build them" (Crawford 9). In assistive contexts, such biases can shape how environments and social interactions are narrated and understood.

Moreover, the reduction of complex environments into simplified verbal summaries risks flattening lived experience. These limitations highlight the ethical responsibility of developers and researchers to critically examine how assistive AI represents people, spaces, and activities, particularly in diverse cultural contexts.

A comparative analysis of *Seeing AI* and *Google Lookout* highlights both convergences and divergences in environmental narration, design philosophy, and functional capabilities. Both applications employ computer vision and natural language generation to provide real-time descriptions of objects, people, and activities. As Goodfellow, Bengio, and Courville note, modern AI systems rely on deep learning architectures to recognize patterns and infer meaning from complex datasets (Goodfellow, Bengio, and Courville). In this respect, both apps perform the fundamental task of translating visual input into intelligible narratives for users with visual impairments.

Despite these shared capabilities, the applications differ in narrative strategy and user experience design. *Seeing AI* is structured around discrete "channels," each optimized for specific functions such as short text reading, scene description, and facial recognition, offering modular but highly focused narration (Microsoft). In contrast, *Google Lookout* emphasizes continuous scene scanning and integrated narration, allowing a fluid, context-aware interpretation of complex environments (Google). This design difference reflects Verbeek's assertion that "technologies mediate experience differently depending on the affordances and intentions embedded by designers" (Verbeek 14).

Each system also exhibits distinct strengths and constraints. *Seeing AI* provides highly detailed descriptions in controlled contexts but can be slower in dynamic environments, whereas *Google Lookout* excels in real-time contextual awareness but occasionally sacrifices granularity or accuracy in dense scenes. These trade-offs underscore that AI-based environmental narration is not neutral; it embodies design priorities, computational limitations, and assumptions about user needs.

In sum, the comparative analysis demonstrates that while both apps enhance accessibility through environmental narration, their differing design philosophies and operational strategies shape the scope, coherence, and interpretive framing of the narratives they produce.

The findings of this study carry important implications for the design and development of assistive technologies. Understanding how *Seeing AI* and *Google Lookout* function as environmental narrators emphasizes the need for design strategies that balance accuracy, coherence, and contextual sensitivity. As Norman observes, technology should not only function efficiently but also "support the user's mental model and interpretive needs" (Norman 47). Designers of assistive AI must therefore consider how algorithmic decisions about prioritization, scene

segmentation, and narration influence user comprehension and autonomy.

Beyond individual user experience, AI-mediated environmental narration also affects broader human perception and interaction with environments. These technologies illustrate how perception is increasingly co-constructed with machines, reshaping the way humans experience and interpret space. As Verbeek asserts, "Artifacts are mediators of human action and perception, not mere instruments" (Verbeek 13). Assistive AI thus transforms users' relationship to their surroundings by selectively framing spatial and social information, highlighting the interpretive role of technology in everyday life.

The paper further insights bear relevance to accessibility policy and inclusive innovation. AI-driven environmental narration has the potential to reduce barriers for visually impaired individuals, contributing to broader goals of social inclusion and equitable access. The World Health Organization emphasizes that accessibility technologies "enable participation and reduce disability-related disadvantages" (World Health Organization). Policymakers and developers must therefore ensure that AI systems are inclusive, culturally sensitive, and ethically designed to meet diverse user needs. The analysis underscores that environmental narration through assistive AI is not merely a technical feature but a mediating force with ethical, social, and practical implications. Thoughtful design, continuous evaluation, and policy support are essential to maximize its benefits and minimize risks of misrepresentation or exclusion.

The study has examined how assistive AI applications, specifically *Seeing AI* and *Google Lookout*, operate as environmental narrators, translating complex visual environments into coherent, real-time verbal descriptions. The analysis revealed that these systems do not merely report visual information but actively structure it, selecting and prioritizing elements based on algorithmic interpretation and design objectives. Scene construction, narrative coherence, and temporal mediation were identified as central mechanisms through which AI shapes user perception and engagement with space.

By comparing the two applications, this study demonstrated that design philosophies and narrative strategies significantly influence the interpretive framing of environments. *Seeing AI* emphasizes modular, function-specific narration, offering detailed insights in controlled contexts, while *Google Lookout* provides fluid, continuous environmental narration that prioritizes context and real-time responsiveness (Microsoft; Google).

The implications of these findings extend beyond individual accessibility, shaping broader understandings of human–machine co-perception and influencing inclusive design, policy, and ethical frameworks. Future research should explore multimodal AI approaches, integrate user-cantered studies to assess practical efficacy, and critically evaluate ethical considerations, including bias, misrepresentation, and cultural inclusivity in environmental narration.

Framing machine vision as an environmental narrator provides a conceptual lens to understand the transformative potential and inherent limitations of AI-assisted perception, emphasizing the active, interpretive role of technology in shaping human engagement with the world.

REFERENCES

[1] Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

[2] Creswell, John W. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. 4th ed., SAGE Publications, 2018.

[3] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[4] Google. *Google Lookout*. Google, 2023, https://www.google.com/accessibility/products-services/lookout/.

[5] Kannadhasan, M. "The Function of Artificial Intelligence in Inclusive English Language Teaching: Strategies for Diverse Learners*." International Journal of Teaching, Learning and Education (IJTLE),* Vol. 3, No. 4, July-Aug. 2024, p. 16, https://dx.doi.org/10.22161/ijtle.3.4.4 s

[6] Kress, Gunther, and Staffan Selander. "Multimodal Design, Learning and Cultures of Recognition." *The Internet and Higher*

Education, vol. 15, no. 4, Dec. 2011, pp. 265–68. https://doi.org/10.1016/j.iheduc.2011.12.003

[7] Microsoft. *Seeing AI*. Microsoft, 2023, https://www.microsoft.com/en-us/ai/seeing-ai.

[8] Norman, Donald A. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books, 2013.

[9] Suchman L. *Human-Machine Reconfigurations: Plans and Situated Actions*. 2nd ed. Cambridge University Press; 2006.

[10] Verbeek, Peter-Paul. *Moralizing Technology: Understanding and Designing the Morality of Things*. University of Chicago Press, 2011.

[11] World Health Organization. *World Report on Disability*. World Health Organization, 2011, https://www.who.int/publications/i/item/9789241564182.

[12] Yin, Robert K. *Case Study Research and Applications: Design and Methods*. 6th ed., SAGE Publications, 2018.

[13] Zhang, Qi, et al. "Advances in Computer Vision for Assistive Technologies: Applications and Challenges." *Journal of Visual Computing*, vol. 39, no. 2, 2021.