Design and Proposal of a Human-Machine Collaborative Interaction System with Semantic Recognition for Personalized Indoor Guided Tours

Yong Li[1]
[1]College of Architectural Arts, Guangxi Arts University, Nanning 530007, Guangxi, China
Corresponding author: Yong Li (e-mail: Liyong_11001@126.com)

ABSTRACT This experiment focuses on establishing a guided-tour system that enables human-machine collaboration through semantic recognition. The system is designed to provide dynamic and personally adaptable indoor guidance, which is relevant to complex industrial and engineering environments where technicians require real-time, context-aware instructions for equipment inspection, maintenance, and safety management. From the perspective of advanced electromagnetic systems, the same architecture can be connected with antenna-assisted positioning, radio-frequency identification, or wireless sensor data to support more accurate indoor localization and interaction. The research addresses shortcomings of current guided systems, including limited real-time response and insufficient personalization. It introduces a more interactive and user-friendly experience, identifies these deficiencies in existing studies, and illustrates how next-generation semantic recognition can make educational and industrial tours more detailed and adaptive. Future studies will focus on three areas: using augmented reality to enhance immersive learning, integrating emotion detection to adjust the tone and timing of guidance, and extending multi-device cooperation for group scenarios. The findings have implications for interactive educational technology, personalized learning environments, and indoor engineering guidance systems that depend on semantic interpretation and reliable wireless context awareness.

INDEX TERMS human-machine collaboration,semantic recognition,indoor guided systems,radio-frequency localization, antenna-assisted navigation

I. INTRODUCTION

GUIDED tours, now driven by the concept of human-machine collaboration, are fundamentally transforming

how visitors interact with exhibitions and museums. This shift is particularly impactful in specialized areas, offering a new depth of understanding for collections dedicated to the intricate artistry and material history . Conventional guided tours involve the delivery of pre-recorded data, which is not customized to meet the needs and interests of individuals. With technology, it is possible to make the guided experience far more dynamic, personalized, and sensitive to a visitor's specific interests [1].

By incorporating human-machine collaboration, such systems offer personalized experiences tailored to a visitor's needs and modify the experience based on user interaction, making it more engaging and educationally fulfilling. This revolution not only enhances the visitor experience but also allows the tour guide to interact with the audience more effectively, thus improving memory retention and engagement [2].

Semantic recognition enables the system to identify the context and meaning of objects and activities at its core. For indoor guided tours, this translates into the ability to identify landmarks, artifacts, and visitor activities in real time. It enables the system to deliver visitor-specific information tailored to individual tastes and preferences, thereby enhancing the interactive quality of the tour experience [3]. Low-latency semantic recognition allows the system to dynamically adapt the tour experience in response to instantaneous, individual visitor inputs, yielding a deeply contextualized and adaptive personalized experience. The use of semantic recognition transforms guided tour experiences from following a pre-mapped route to a dynamic experience that continuously evolves based on ongoing visitor interaction, making it more relevant to what is being guided.

Such a complex system can only be built by combining several technologies, including computer vision, natural language processing, and machine learning. These technologies enable the system to load vast amounts of information, provide real-time decision-making, and dynamically deliver content [4]. Object recognition is made possible by computer vision, while voice commands and voice search are implemented through natural language processing. Machine learning enables the system to study visitor interests over time and provide highly relevant suggestions and content. The challenge, however, is how to integrate these technologies into a single cohesive system that responds to visitors' movements and behaviors without overwhelming the user with complexity.

The objective of this project is to introduce and implement a human-machine collaborative interaction system based on semantic identification, specifically designed for indoor guided tours. By combining deep learning and semantic recognition, the project aims to provide a more personalized tour experience, with diverse experiences for each user. The goal is to make the system informative and interactive—not just a passive listener, but an active participant in the tour. By implementing cutting-edge technologies, the system transforms the tour into an interactive and dynamic experience, making the environment responsive to the user and providing dynamic, personalized content . The outcomes of this project will serve as a crucial gateway to future possibilities in adaptive and interactive learning technologies, especially for museums and cultural heritage sites arts, setting a new standard for visitor engagement and deep material understanding .

II. LITERATURE REVIEW

A. HUMAN-MACHINE INTERACTION IN GUIDED TOURS

Human-computer interaction in indoor tour guiding has been widely explored as a means of offering individualized and interactive experiences. Conventional methods, typically in the form of pre-recorded audio tours, have evolved into more sophisticated systems using interactive robots or digital agents and natural language processing (NLP) to provide responses to tourists. These systems offer two-way communication, allowing tourists to ask questions and receive context-specific answers. Robotic interfaces and semantic recognition have been shown to significantly boost engagement and satisfaction due to their responsive and adaptive nature.

B. SEMANTIC RECOGNITION IN INDOOR ENVIRONMENTS

Semantic awareness plays an important role in improving the level of machine comprehension regarding a visitor's interactions and surroundings [5]. Research in this area involves applying a combination of computer vision and machine learning algorithms to enable systems to detect and understand objects, events, and actions in real time. The use of semantic awareness in interactive guided tours allows systems to identify not only physical objects (e.g., art or artifacts) but also visitor actions, enabling the system to modify content or responses based on those actions. The algorithm involved in semantic awareness measures the similarity between a visitor's question and the system's knowledge base. For example, cosine similarity is used to match the semantic meaning of a user's query and the system's knowledge:

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\|\|B\|} \tag{1}$$

where A and B are vectors representing the query and the reference data, respectively.

C. TECHNOLOGIES ENABLING PERSONALIZED GUIDED TOURS

The integration of technologies such as computer vision, NLP, and machine learning algorithms has made it possible to create more advanced and personalized learning experiences. Computer vision enables systems to visually perceive and understand their environment, while NLP allows systems to communicate with people in natural language. Machine learning, especially deep learning models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs), enables systems to forecast and recommend pathways based on visitor patterns from previous visits. These technologies allow the system to learn about visitor behavior and interests over time and offer more tailored and descriptive recommendations to future visitors [6]. An example of such technologies is modeled in the following equation to forecast the next likely step in a sequence of tours using LSTM:

$$h_t = \mathrm{LSTM}(x_t, h_{t-1}) \tag{2}$$

where h_t is the hidden state at time t, and x_t is the input (e.g., visitor's behavior or action) at time t.

D. CHALLENGES AND LIMITATIONS OF CURRENT MODELS

Despite progress, current human-machine collaborative tour models still suffer from various issues. One of the key issues is the lack of real-time flexibility, especially when dealing with large volumes of data. Another major challenge is integrating multiple technologies (such as computer vision, NLP, and machine learning) in a harmonized manner. Furthermore, although semantic recognition systems are highly effective, they still struggle to interpret vague or ambiguous guest questions. To address these issues, researchers have proposed using more advanced algorithms and hybrid models that fuse deep learning techniques with probabilistic methods to achieve optimal performance. A common approach to avoiding ambiguity in system replies is the use of a Bayesian network, which allows probabilistic decision-making:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)} \tag{3}$$

where $\mathrm{P}(A|B)$ represents the probability of a given event (such as a tour suggestion) based on user interactions, and $\mathrm{P}(B|A)$ and $\mathrm{P}(A)$ represent the likelihood of certain conditions and outcomes.

III. THEORETICAL FRAMEWORK

A. LEARNING THEORIES IN HUMAN-MACHINE COLLABORATION

Human-computer interaction in indoor guided tours is also informed by Constructivism, a theory of learning that emphasizes active learning, where learners build knowledge through interaction with their environment [7]. As illustrated in Figure 1, the constructivist theory is based on premises such as developmental concerns, learning objectives, and pedagogical designs like cognitive apprenticeship and situated cognition. These elements are used to ensure that the tour system aligns its content with the visitor's cognitive developmental level, as prescribed by Piaget's developmental stages theory. Vygotsky's Zone of Proximal Development (ZPD), shown in Figure 2, is used to create tasks that are slightly above but achievable with guidance. This model, characterized by zones of achieved and potential growth, guides the system to introduce challenges that promote learning without frustrating the user. By aligning content to the user's increasing abilities, the system promotes engagement and fosters cognitive growth throughout the guided tour activity.

Within such an adaptive system, the complexity of the content increases as the visitor navigates through it.

According to the ZPD model, learning should occur in a zone with tasks that are neither too complex nor too simple, but challenging enough to promote improvement. The system must adjust to these demands, offering sufficient challenge and maintaining the visitor's engagement without causing frustration. The formula for adaptive progression of content is:

$$C_t = C_{t-1} + \alpha(P_t - C_{t-1}) \tag{4}$$

where C_t = the content complexity at time t; C_{t-1} = the previous complexity level; P_t = the difficulty of the current task; α = the adjustment factor based on visitor input.

This equation helps the system adjust content complexity, ensuring a personalized and adaptive tour experience.

B. SEMANTIC RECOGNITION MODELS FOR INTERACTIVE LEARNING

Semantic awareness is necessary to improve human-machine interaction by making the system capable of interpreting visitors' visual and verbal inputs. Integration between NLP and Computer Vision (CV) allows the system to interpret and analyze real-time visitor and environmental data [8]. This combination creates a sophisticated, contextually sensitive learning experience in which objects, landmarks, and visitor movements are identified by the system, and the visitor experience is altered dynamically. The system relies on a semantic graph of objects, activities, and knowledge to generate contextually relevant content that tracks visitor movement in real time.

The recognition of semantics can be measured in terms of similarity between the system's knowledge base and the visitor's query, using cosine similarity (see Formula (1)). This similarity measure assists the system in converting visitor input in real time to the most appropriate matching content, ensuring that the visitor experience is contextualized and adapted.

C. PERSONALIZATION MACHINE LEARNING MODELS

Machine learning classifiers such as LSTM networks and CNNs are utilized to maximize personalization. LSTM networks process sequential data, enabling the system to learn about visitor behavior over time and make predictions based on past actions [9]. This allows the system to align the learning path in real time, updating recommendations in line with visitor interests. CNNs process visual data, helping the system recognize and classify objects, artifacts, and other visual elements within the space, synchronizing what the visitor is seeing with the content recommendations.

1) LSTM Networks

The system utilizes an LSTM network to predict the next logical interaction or content requirement based on current and historical user interactions. As defined in Formula (2), the LSTM unit uses its internal gating mechanism to manage information flow, ensuring the model effectively captures the sequential dependencies in the user's tour path and interaction history for predictive content delivery.

D. DECISION-MAKING ALGORITHMS FOR REAL-TIME ADAPTATION

An adaptive tour system operating in real time requires decision-making algorithms. Past feedback and interaction can be used to inform the system, as is done with Reinforcement Learning (RL) and Q-learning in particular [10]. RL trains the system to define the most optimal actions (e.g., which exhibits to suggest) to maximize visitor engagement. The Q-learning algorithm predicts action rewards and adjusts future actions to enhance the visitor experience.

The Q-learning algorithm is represented as follows:

$$Q(s,a) = Q(s,a) + \alpha\left(r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right) \tag{5}$$

where $Q(s,a)$ is the expected reward for taking action a in state s; r is the immediate reward; γ is the discount factor; and α is the learning rate.

This equation helps the system update its decision-making process, continuously adapting to visitors' evolving preferences.

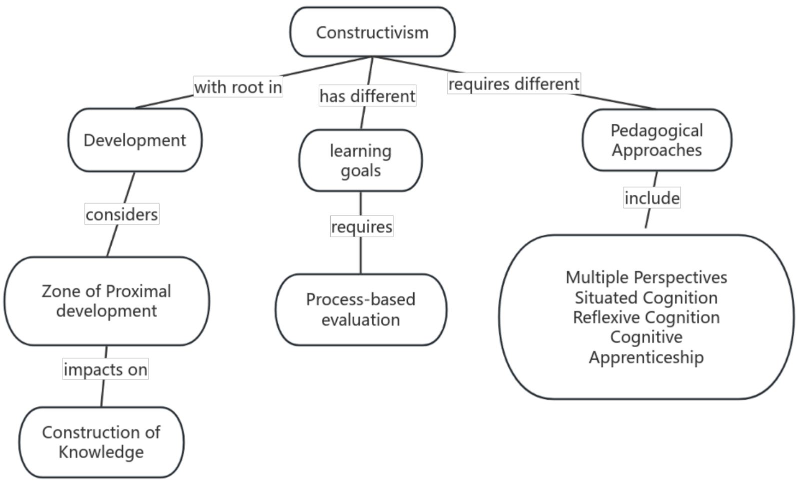

FIGURE 1. Constructivist framework showing the alignment of learning goals, development stages, and pedagogical approaches

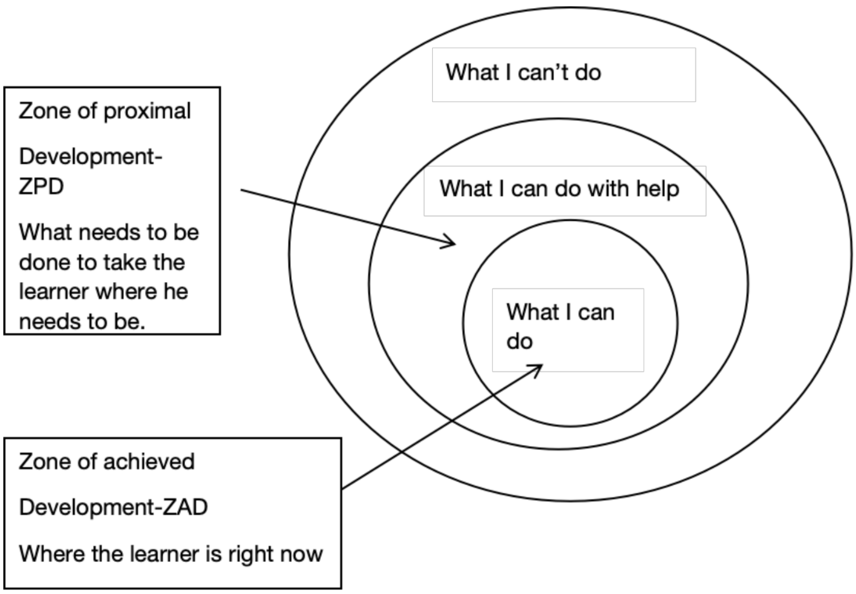

FIGURE 2. Vygotsky's ZPD model illustrating learner capabilities with and without assistance

E. INTEGRATION OF THEORIES FOR ADAPTIVE GUIDED TOURS

Constructivism, semantic representation, learning algorithms, and decision algorithms work together to form a robust paradigm for adaptive guided tours. The blend of these technologies and principles provides an effective, dynamic, and interactive learning environment. Constructivism influences content based on visitor cognitive preparedness; semantic comprehension enables interpretation and response to verbal and visual stimuli [11]; learning methods like LSTM and CNN adapt to visitor behaviors and visual cues; and reinforcement learning algorithms fine-tune recommendations over time.

By integrating these theories and technologies, the system can nimbly adapt data to fit each visitor's needs, offering a comprehensive, differentiated experience throughout the entire tour.

IV. SYSTEM ARCHITECTURE AND DESIGN

A. OVERVIEW OF THE COLLABORATIVE SYSTEM STRUCTURE

The human-computer collaboration infrastructure for indoor guided tours is developed using a modular and scalable architecture that facilitates easy interaction between users and the system. The architecture enables real-time content delivery based on user input, feedback, and engagement levels. The system comprises the following primary modules:

- Natural Language Command Module: Converts user commands given through text or speech, supporting spoken interaction.
- Sensor Module: Detects environmental parameters, visitor movement, and position, explicitly incorporating

data streams from Inertial Measurement Units (IMUs), computer vision cameras, and Bluetooth Low Energy (BLE) beacons for robust localization and real-time context.

- Processing and Control Module: Utilizes a Raspberry Pi (illustrated in Figure 3) to regulate data flow and ensure smooth operations.
- Feedback Mechanisms: Provides visual, audio, and tactile feedback to ensure user engagement and understanding.
- Adaptive Learning Module: Continuously monitors user behavior and modifies the difficulty level of materials to match the visitor's cognitive load.

The overall system design, as illustrated in Figure 3, offers seamless connectivity between modules, with data flowing through the Natural Language Command Module and real-time processing to deliver the best possible user experience.

B. SEMANTIC RECOGNITION AND PROCESSING

The user's verbal and non-verbal input is interpreted by the Semantic Recognition Module, which employs NLP and computer vision in conjunction to understand visitor queries and surroundings. The system uses voice or gestures to identify objects, displays, and visitor intent. For instance, the NLP module translates requests such as Explain this artwork and cross-references them with relevant knowledge base information.

The CV system tracks the visitor's line of sight and gestures, triggering additional, context-aware information.

1) Reliability of Semantic Recognition

To address potential failures in visual recognition (e.g., due to poor lighting, object occlusion, or complex backgrounds), the system employs a semantic-visual fallback loop. If the CV system's confidence score for object recognition falls below a threshold, the system initiates a verbal/semantic confirmation step. For example, if an object is partially obscured, the system might prompt:

"I see you are looking at the general area of <Exhibit Category>. Could you please confirm if you are interested in the <Closest Exhibit Name>?" This prioritized semantic-over-visual recovery mechanism prevents recognition failures from interrupting the guidance loop.

2) Command Processing Robustness: Resolving Syntactic and Semantic Ambiguity

A critical element of the NLP module is its robustness in handling complex, ambiguous, or out-of-scope queries. The system uses a specialized, two-stage process that leverages semantic similarity and sensor fusion to accurately map queries to physical objects:

- Syntactic Analysis for Intent Extraction: To manage syntactic ambiguity (e.g., complex sentence structures), the NLP component first utilizes dependency parsing. This analysis is crucial for correctly isolating the core action verb (the visitor's intent, e.g., explain, show me) from the target noun phrase (the object of interest, e.g., the blue thing), ensuring that the system correctly interprets the structural meaning of the command.
- Semantic-Visual Fusion for Object Mapping: To resolve semantic ambiguity (e.g., the vague query tell me about the blue thing), the system does not rely solely on text matching. Instead, it calculates a Fusion Confidence Score (FCS) for every potential exhibit, combining the semantic relevance of the query with real-time proximity and visual focus data from the CV system. The FCS is defined as:

$$\mathrm{FCS}_i = \mathrm{Cosine\ Similarity}_i \times \mathrm{CV\ Proximity\ Score}_i \quad (6)$$

where FCS_i is the Fusion Confidence Score for exhibit i; S_i is the semantic match between the query and exhibit i's description; and $\mathrm{CV\ Proximity\ Score}_i$ is a confidence weight derived from the visitor's line of sight and sensor distance to exhibit i. This fusion mechanism ensures that a vague verbal input is accurately mapped to the object the visitor is physically engaging with. A critical element of the NLP module is its robustness in handling ambiguous or out-of-scope queries. While semantic similarity is primarily computed using cosine similarity, the system employs a two-stage filtering mechanism before content delivery:

- Semantic Similarity Thresholding: Semantic similarity between the exhibit description and the user query is computed using the cosine similarity function (Formula (1)).
- A predefined relevance threshold is used; if the maximum FCS or semantic similarity score falls below this threshold, the query is flagged as potentially ambiguous or out-of-scope.
- Fallback and Clarification Protocol (for low-score queries): For flagged queries, the system avoids simple failure or irrelevant responses. If the highest FCS indicates high ambiguity (with multiple top-scoring exhibits), the system enters a clarification loop: it uses CV data to identify the top two or three probable objects and initiates a concise, specific clarification prompt—such as: "Are you asking about the [Suggested Object 1 Name] or the [Suggested Object 2 Name]?" For completely out-of-scope queries (e.g., "What is the capital of France?"), the system provides a polite, boundary-aware response, such as:
  "My apologies, I can only provide information about the exhibits in this tour."

This multi-layered approach ensures that the presented content remains contextually relevant, maintaining user engagement even when queries are challenging.

C. HUMAN-MACHINE INTERACTION WORKFLOW

User interaction with the system occurs through a formal feedback loop, in which the system is continuously modified

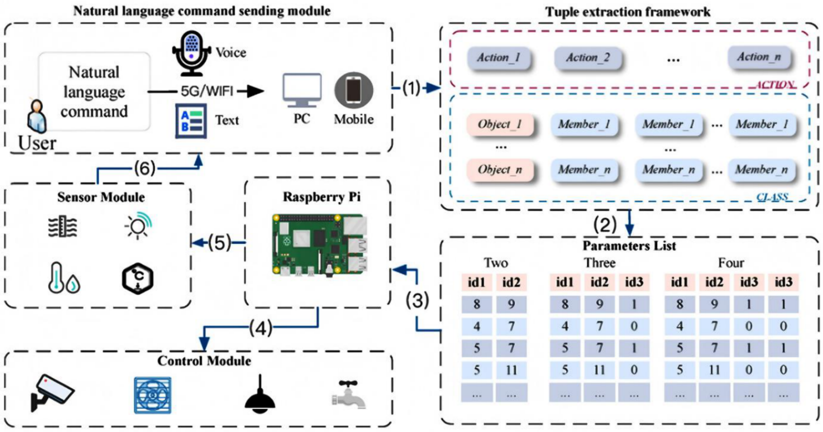

FIGURE 3. System architecture of the collaborative human-machine interface, including modules for natural language command processing, sensor integration, and control

based on user activity. The process includes the following steps:

- Input Detection: The system detects user input, whether it is voice, gesture, or touch.
- Processing: The Semantic Recognition Module interprets the input to recognize intent and context.
- Decision Making: The Adaptive Learning Module evaluates the visitor's mental state and selects suitable content.
- Response Generation: The system delivers content through appropriate feedback (audio, visual, or haptic).
- Learning and Adaptation: User behavior is logged, and the system learns to improve future responses.

This guarantees responsiveness to visitor needs and promotes interaction. As shown in Figure 4, this interaction model includes feedback from both the robot and environment, enabling adaptation throughout the process based on user preferences and engagement levels.

D. INTEGRATION AND SCALABILITY CONSIDERATIONS

Scalability is essential for allowing the system to handle a large number of visitors without compromising performance [12]. The system architecture facilitates seamless integration with existing devices and infrastructure. The main design considerations are:

- Modular Components: The system consists of individual modules that can be replaced or upgraded without disrupting the overall system.
- Cloud-Based Data Processing: Data can be processed in real time through the cloud, making the system scalable across multiple venues.
- Cross-Platform Support: The system supports a range of devices, from smartphones to AR glasses, ensuring a consistent user experience.
- Security and Privacy: Strong security mechanisms, such as data transmission encryption, are used to safeguard user information and privacy.

The cloud architecture simplifies system scaling and enables communication with external services, such as cloud storage, message queues, and NoSQL databases.

E. FEEDBACK AND EVALUATION MECHANISMS

Continuous feedback is key to the success of the human-machine collaboration system [13]. The system employs several mechanisms to monitor and enhance performance:

- User Feedback: Direct visitor feedback on content quality and relevance.
- Performance Metrics: Monitoring response time, accuracy, and visitor interaction to gauge effectiveness.
- Usability Testing: Periodic usability testing to identify and address areas for improvement.
- Data Analysis: Continuous analysis of interaction data to refine content recommendations and improve system function.
- Stakeholder Reviews: Feedback from stakeholders, designers, and educators is used to enhance and advance the system.

V. PROPOSED TECHNOLOGY STACK AND IMPLEMENTATION PLAN

A. CLOUD COMPUTING INFRASTRUCTURE

The system utilizes cloud computing platforms for scalability and high performance. Cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud provide data processing and storage infrastructure. This infrastructure supports the efficient processing of dynamic, real-time content delivery.

B. MOBILE APPLICATION INTEGRATION

The system is also integrated with mobile applications to maximize user interaction. Recent market reports indicate that mobile application revenue continues to grow rapidly, with games leading the way. The Global NLP Market chart highlights increased mobile platform utilization and its impact on user interaction with cloud services. Mobile applications allow users to access the guided tour system at their convenience, supporting diverse devices such as smartphones and tablets [14].

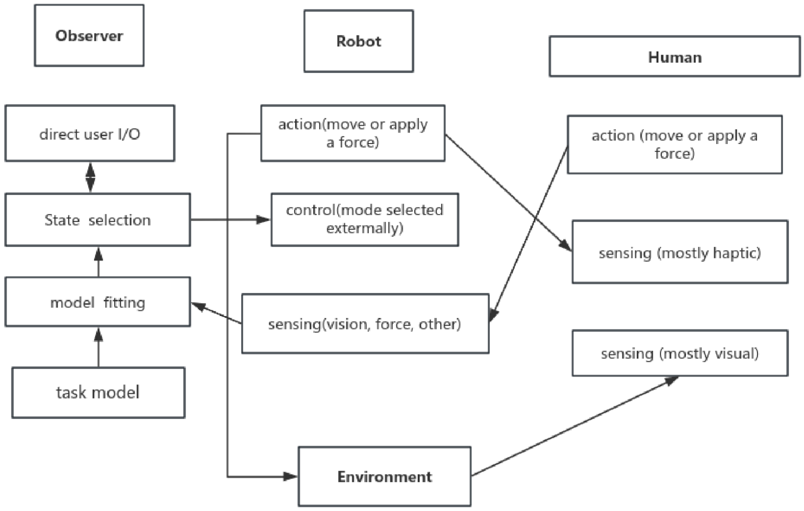

FIGURE 4. Interaction model between observer, robot, human, and environment in a collaborative setup

C. NATURAL LANGUAGE PROCESSING

The NLP module enables the system to process and understand user commands. The global NLP market is expected to reach 453.3 billion by 2032, with hybrid NLP technologies playing a significant role. NLP allows the system to interpret user questions and respond with context-appropriate answers. Statistical and hybrid NLP approaches are becoming increasingly essential in real-time human-machine communication, making systems capable of processing natural language inputs.

D. MACHINE LEARNING FOR PERSONALIZATION

Personalization of the visitor experience is achieved through collaborative filtering and reinforcement learning algorithms. The system continuously learns from user behavior to make real-time recommendations based on user interests and past interactions. As illustrated in applications of natural language processing, personalized recommendations are enabled by machine learning, enhancing user satisfaction by delivering relevant content tailored to each user's interests.

E. SENSOR INTEGRATION FOR CONTEXT AWARENESS

The system utilizes various sensors, including infrared and depth sensors, to detect user proximity, direction of view, and environmental conditions. The system can respond to instantaneous user input by adjusting content accordingly. Automatic processing of sensor information ensures the system remains aware of and sensitive to environmental conditions. Sensor data is processed in real time to deliver the most suitable content, as demonstrated in applications of NLP, which reflect the impact of real-time data on content creation.

VI. EVALUATION AND PERFORMANCE METRICS

A. SYSTEM PERFORMANCE EVALUATION

Performance measurement is crucial for determining whether the human-machine collaborative system provides an optimal user experience [15]. Key performance metrics include response time, response accuracy, and user engagement. Response time is the interval during which the system processes and responds to human input, and it is essential for maintaining interactivity. Response accuracy reflects the system's ability to understand and provide pertinent information in response to user questions. User engagement is measured by the number and duration of user interactions, indicating how well the system retains users' interest during the tour.

B. USER SATISFACTION AND EXPERIENCE

User experience (UX) is a significant factor driving the guided tour system's success. System effectiveness is measured by task completion rate, user engagement, and emotional engagement. Task completion rate reflects how quickly users finish the tour or tasks, indicating usability. Feedback is collected through questionnaires or surveys, allowing visitors to share their experiences and suggest improvements. Sentiment analysis is used to interpret user reactions, such as excitement or frustration, enabling the system to adjust content dynamically to maintain visitor interest.

C. DATA ACCURACY AND RELEVANCE OF CONTENT

Delivering a personalized and relevant experience requires ensuring data accuracy and content relevance.

Content relevance measures how well the system's responses align with visitors' needs and interests, providing contextual information. Data consistency ensures that information presented is consistent across sessions and devices. The error rate reflects the proportion of irrelevant or incorrect responses, offering valuable insights into areas for system improvement. Regular model updates and frequent testing are employed to improve response accuracy and content relevance, resulting in greater user satisfaction.

D. SCALABILITY AND LOAD TESTING

Scalability is essential for supporting many simultaneous users in high-usage scenarios. Throughput testing provides evidence of how many concurrent users the system supports without a decline in performance.

Stress testing subjects the system to extreme user loads to identify bottlenecks and limitations, ensuring responsiveness under peak loads. Load balancing distributes user traffic among multiple servers, preventing any single server from being overwhelmed. These scalability tests ensure the system can be deployed in various environments and maintain optimal responsiveness and performance regardless of user volume.

E. ITERATIVE TESTING AND CONTINUOUS IMPROVEMENT

Continuous improvement is achieved through iterative testing and user feedback. Functionality, usability, and content delivery issues are detected through periodic user testing. Features are modified based on test results to improve system performance and user experience. Performance fine-tuning is essential to minimize response times and maximize content delivery accuracy. Data analysis drives iterative activity, and as more is learned about user interaction, the system is refined accordingly. New technologies and features are adopted as the system evolves to drive engagement and remain relevant to changing user needs.

VII. FUTURE WORK AND IMPROVEMENTS

A. SEMANTIC DEVELOPMENT OF RECOGNITION

Future developments in semantic understanding will focus on enhancing the system's ability to interpret complex user inputs, including ambiguous or emotive statements. Deep learning models will be increasingly incorporated to improve context awareness, enabling the system to respond more accurately and relevantly.

The system will further support diverse languages to serve a global user base. These advancements will make the system more natural, adaptive, and responsive to a wide array of user interactions. The ultimate aim is to enable the system to generate suitable and relevant outputs for a vast variety of input types, thereby improving the learning process.

B. INTEGRATION WITH AUGMENTED REALITY

The most significant upcoming development is the integration of AR. When AR is incorporated into the current system, the visitor experience will become more interactive and immersive, enabling visitors to see 3D models and other visual material superimposed on real-world objects. Visitors will be able to interact with exhibits in real time using AR glasses or mobile apps. Future development will focus on leveraging AR to create additional learning opportunities during tours, making them more interactive and dynamic. This technology will help visitors better understand content, resulting in a more immersive learning experience.

C. ARTIFICIAL INTELLIGENCE-BASED PERSONALIZATION

High levels of personalization are planned as a major future improvement. By incorporating reinforcement learning and deep learning, the system will better assess individual visitor interests and adjust content delivery accordingly. Machine learning will continue to predict user needs, making recommendations based on previous actions. As more data is gathered, recommendations and models will be refined over time. Realtime feedback will be implemented, allowing the system to continuously adapt the experience to the visitor's thinking process, ensuring that it remains aligned with the visitor's learning level and interests.

D. ENHANCED REAL-TIME FEEDBACK MECHANISM

Emotion detection and behavioral analysis will be included to improve the system's understanding of user responses and real-time feedback effects. This will allow the system to customize content based on the visitor's emotional and mental status. For example, if a visitor is tired or confused, the system can adjust the pace or inject additional engaging information. Gamification elements, such as quizzes and interactive challenges, will be implemented to increase participation and facilitate learning. These features will make the system highly interactive and keep visitors engaged throughout their learning experience.

E. MULTI-DEVICE SUPPORT VIA CLOUD INTEGRATION

The use of cloud computing in the future will allow users to access the system using multiple devices, enabling tourists to use smartphones, tablets, or wearable devices. This will provide users with a uniform experience regardless of the device used. Cloud-based computing will enable simultaneous access for multiple users, increasing scalability. User learning history and preferences will be synchronized across devices via the cloud.

Multi-device capability will make the system more flexible and offer a seamless experience to tourists in any environment.

VIII. CONCLUSION

Overall, the human-machine collaborative system for guided tours represents a significant improvement over traditional methods, especially by offering museum visitors a dynamic, customized learning experience that enables deep, specific engagement with collections focused on artistry and the history of leather craftsmanship. Through semantics-based recognition, natural language processing, and artificial intelligence, the system can adjust in real time to individual visitor needs and interests. This shift to adaptive delivery encourages increased user interaction and provides a more relevant, educational experience. Advanced algorithms enable seamless interpretation of user input, allowing the system to anticipate and respond, so the tour evolves through ongoing interaction. Future enhancements include integrating AR and implementing features for real-time emotion recognition and multi-device support. These additions will attract more users by offering personalized tours that accommodate different learning preferences. The seamless combination of cloud computing and data analytics ensures scalability, allowing the system to be applied in a variety of environments, from small local museums to large cultural institutions. This project lays the foundation for the future development of interactive educational technology, transforming museum and exhibit tours into a new era of personal education and interaction. Specifically, it offers a robust model for collections focused on arts, enabling visitors to go beyond simple labels and gain deep insights into intricate weaving techniques, fiber molecular structures, tanning processes, and the historical material culture they represent.

AUTHOR CONTRIBUTIONS

Yong Li designed, collected and analyzed the data, and drafted the manuscript. Yong Li conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yong Li participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Y. J. Chang, C. Y. Yang, Y. H. Kuo, W. H. Cheng, C. L. Yang, F. Y. Lin, et al., “Tourgether: Exploring tourists’ real-time sharing of experiences as a means of encouraging point-of-interest exploration,” Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol, vol. 3, no. 4, pp. 1-25, 2019, doi: 10.1145/3369832.

[2] G. Alexakis, S. Panagiotakis, A. Fragkakis, E. Markakis, and K. Vassilakis, “Control of smart home operations using natural language processing, voice recognition and IoT technologies in a multi-tier architecture,” Designs, vol. 3, no. 3, pp. 32, 2019, doi: 10.3390/designs3030032.

[3] Y. Hou, “Application of intelligent internet of things and interaction design in Museum Tour,” Heliyon, vol. 10, no. 16, 2024, doi: 10.1016/j.heliyon.2024.e35866.

[4] M. I. Ahmad, O. Mubin, S. Shahid, and J. Orlando, “Robot’s adaptive emotional feedback sustains children’s social engagement and promotes their vocabulary learning: A long-term child–robot interaction study,” Adaptive Behavior, vol. 27, no. 4, pp. 243-266, 2019, doi: 10.1177/1059712319844182.

[5] S. Rosa, A. Patane, C. X. Lu, and N. Trigoni, “Semantic place understanding for human–robot coexistence—toward intelligent workplaces,” IEEE Transactions on Human-Machine Systems, vol. 49, no. 2, pp. 160-170, 2018, doi: 10.1109/THMS.2018.2875079.

[6] Y. Zhang, W. Diao, Y. Nie, and Q. Wang, “Design of signage guidance system for tourist attractions based on computer vision technology,” Journal of Computational Methods in Science and Engineering, vol. 24, no. 1, pp. 413-426, 2024, doi: 10.3233/JCM-237032.

[7] X. Yu, X. Zhang, C. Xu, and L. Ou, “Human–robot collaborative interaction with human perception and action recognition,” Neurocomputing, vol. 563, Art. no. 126827, 2024, doi: 10.1016/J.NEUCOM.2023.126827.

[8] P. Wiriyathammabhum, D. Summers-Stay, C. Fermüller, and Y. Aloimonos, “Computer vision and natural language processing: Recent approaches in multimedia and robotics,” ACM Computing Surveys (CSUR), vol. 49, no. 4, pp. 1-44, 2016, doi: 10.1145/3009906.

[9] A. Crivellari and E. Beinat, “LSTM-based deep learning model for predicting individual mobility traces of shortterm foreign tourists,” Sustainability, vol. 12, no. 1, pp. 349, 2020, doi: 10.3390/su12010349.

[10] J. Lin, Z. Ma, R. Gomez, K. Nakamura, B. He, and G. Li, “A review on interactive reinforcement learning from human social feedback,” IEEE Access, vol. 8, pp. 120757-120765, 2020, doi: 10.1109/ACCESS.2020.3006254.

[11] G. E. Morales-Martinez, E. O. Lopez-Ramirez, J. P. Garcia-Duran, and M. E. Urdiales-Ibarra, “Cognitive constructive â€“Chronometric techniques as a tool for the e-assessment of learning,” International Journal of Learning, Teaching and Educational Research, vol. 17, no. 2, pp. 159-176, 2018, doi: 10.26803/ijlter.17.2.10.

[12] P. Sosnin, “Experiential human-computer interaction in collaborative designing of software intensive systems,” In New Trends in Software Methodologies, Tools and Techniques. Amsterdam, Netherlands: IOS Press, pp. . pp. 180-197, 2012, doi: 10.3233/978-1-61499-125-0-180.

[13] R. Graef, M. Klier, K. Kluge, and J. F. Zolitschka, “Human-machine collaboration in online customer service–A longterm feedback-based approach,” Electronic Markets, vol. 31, no. 2, pp. 319-341, 2021, doi: 10.1007/s12525-020-00420-9.

[14] C. M. Chuang, “A current travel model: Smart tour on mobile guide application services,” Current Issues in Tourism, vol. 23, no. 18, pp. 2333-2352, 2020, doi: 10.1080/13683500.2019.1631266.

[15] T. Wang and W. Zhan, “Design and control a hybrid human–Machine collaborative manufacturing system in operational management technology to enhance human–machine collaboration,” The International Journal of Advanced Manufacturing Technology, pp. 1-11, 2024, doi: 10.1007/s00170-024-14894-w.