



Intelligent Guide Systems for Digital Heritage: A Systematic Review of AR-AI Integration in the Era of Generative Cognition

Malak Alaa¹, Islam Gamal², Amr Ahmed³, Omar Ahmed⁴, Omar Wagih⁵, Ali Ashraf⁶, Rawan Hesham⁷

^{1,2,3,4,5,6,7} Department of Artificial Intelligence, Faculty of Computers and Artificial Intelligence, Capital University, Cairo, Egypt

ARTICLE INFO

KEYWORDS

Generative Cognition; Grand Egyptian Museum Case Study; Adaptive AR-AI Framework; Digital Heritage Impact; Egyptian Antiquities; Augmented Reality; Artificial Intelligence; Adaptive Deep Learning

HOW TO CITE

Ashrafi, A., Ahmed, A., Alaa, M., Ahmed, O., Wagih, O., & Hesham, R. (2026). Intelligent Guide Systems for Digital Heritage: A Systematic Review of AR-AI Integration in the Era of Generative Cognition. *International Journal of Automation and Digital Transformation*, 5(2), 43-55.
<https://doi.org/10.54878/oqzpsr75>

ABSTRACT

Digital transformation in cultural heritage institutions has increasingly prioritized immersive technologies to enhance visitor engagement and knowledge delivery [8]. However, traditional Augmented Reality (AR) solutions remain constrained by the scalability bottlenecks of marker-based tracking and the pedagogical limitations of static content delivery [1]. This work uses the Grand Egyptian Museum (GEM) as a primary case study to validate a consolidated system-level perspective of recent advancements in the integration of Artificial Intelligence (AI) with AR [6]. By synthesizing existing research through the lens of GEM's large-scale operational challenges—such as an expected annual volume of 5–8 million visitors—this study identifies recurring architectural patterns. It evaluates how robust perception capabilities, achieved through mobile-optimized Convolutional Neural Networks such as YOLOv8 and Efficient Net [2, 15], and intelligent cognition, powered by Large Language Models (LLMs) grounded in historical Knowledge Graphs [4, 19], address critical technical challenges including lighting sensitivity and occlusion [9]. The findings indicate that the convergence of adaptive vision and generative storytelling transforms the museum visit from passive observation into an active, dialogue-driven exploration [25]. The study concludes that applying this framework to the Grand Egyptian Museum establishes a new benchmark for 'Generative Cognition' in high-traffic heritage sites, providing a scalable model for transforming visitor interaction globally.



Double-blind peer review by independent reviewers.

Received: 01 Jul 2026, Accepted: 12 Nov 2026, Published: 02 Dec 2026

*Corresponding author: loka.lolo2004@gmail.com

Copyright: © 2026 by the author. Licensee Emirates Scholar Center for Research & Studies, UAE.

Open access under the Creative Commons Attribution (CC BY) license

<https://creativecommons.org/licenses/by/4.0>

Introduction

Museums and cultural heritage institutions are undergoing a significant evolution, transforming from static repositories of artifacts into dynamic, interactive environments. In this era of digital transformation, Augmented Reality (AR) has emerged as a cornerstone of "smart tourism," capable of bridging the gap between historical context and modern engagement. However, despite its potential, the widespread adoption of AR in complex museum settings remains hindered by significant technical limitations. Traditional implementations—primarily relying on marker-based tracking—face critical scalability issues, requiring manual image uploads and struggling under variable lighting or the oblique viewing angles typical of crowded galleries [1]. Furthermore, current content delivery remains largely static, offering pre-scripted narratives that fail to adapt to a visitor's specific interests, thereby limiting educational impact.

Recent advances in Artificial Intelligence (AI) have fundamentally shifted these paradigms. Adaptive deep learning methods now enable robust object recognition and scene understanding, while generative models allow for dynamic narrative construction and personalized content delivery. By reviewing existing research and market data—with a specific focus on the Egyptian heritage landscape and institutions like the Grand Egyptian Museum (GEM)—this paper identifies a clear opportunity for the convergence of adaptive vision and generative storytelling to transform passive viewing into active, dialogue-driven exploration.

Despite the proliferation of studies in this field, existing review papers remain fragmented. Many surveys focus narrowly on isolated components—such as computer vision algorithms or user interface design—without analyzing how these elements are integrated at the system level. Others highlight application cases but lack a comparative framework that connects perception, cognition, and evaluation strategies. Moreover, recent developments in generative AI and adaptive learning are often discussed in isolation, leaving their broader architectural implications insufficiently synthesized.

This research does not propose a new system or algorithm. Instead, it provides a systematic analytical review of existing studies to identify design patterns, research trends, and open challenges that are not explicitly consolidated in previous reviews. By organizing prior research into a structured framework, this study highlights unresolved technical and experiential challenges. Ultimately, this review serves as a

consolidated reference to support both academic research and the practical design of future intelligent cultural heritage applications.

Literature Review

AR in Museums and Cultural Heritage

Last advancements in augmented reality (AR) have significantly transformed museum experiences, enhanced user engagement and interaction. In earlier decades tourist's experiences and engagements in the museums were characterized by traditional and static exhibition methods. However, it lacked interactivity and engagement. Consequently, AR increased engagement by creating more interactive experience for the tourists instead of the passive experience.

Younger generations are less interacted with static tourism. According to Cho, Hyejin [7], AR solved this problem by turning the tourist from passive viewer to active viewer. When the museums use interactive narratives and AR instead of just showing static exhibits, they see a significant increase in dwell time and interest among younger generations. Who ignored it before. Furthermore, passive text labels and static exhibits are frequently forgotten. Based on Choi, Lee Jun, Piao Gui Cheng, and Rehman Ullah Khan [8] study, AR and deep learning solved this problem by leveraging multimodal learning combining visual overlays and audio narration. Which enables the brain to process visual and verbal information simultaneously. Consequently, this increases cognitive engagement and memory retention. and improving learning outcomes more than interacting with passive text alone. Moreover, AR with AI including CNN and LLM can handle real time context. Based on Qorbani, Mahya, Kamran Paynabar, and Mohsen Moghaddam [19] study, this study presents a context-aware large language model (LLM) assistant that integrates diverse data modalities, such as hand actions, task steps, and dialogue history, into a unified framework for real-time question answering. Study highlights the potential of LLM-driven multimodal integration to enable adaptive, intuitive assistance for AR. High interactive AR systems create a state of being immersed in historical context instead of standing in a museum. This immersive state identifies tourist satisfaction and deciding to revisit the same place. According to Lu, Yage, et al [16] study the Terracotta Warriors Museum, it proves AR Visual attractiveness and Entertainment are critical. In addition, supporting heritage site preservation, for example in Terracotta Warriors Museum, AR involves using smartphones apps to enhance viewing fragile artifacts by allowing them to recreate a detailed version,

you can rotate and zoom in without causing any damage to the original artifacts.

There are many global museums using AR in innovative ways to improve user experience. for example:

France: the louvre museum launched “Egypt Augmented” project that uses AR for certain artifacts like the Naos of Amasis and the Zodiac of Dendera to restore their original colors and animations.

USA: the Smithsonian offers AR experience They have an AR that brings dinosaurs to life, and a variety of 3D scanned objects that can be viewed in AR through their website.

UK: the British museum offers gamified augmented reality (AR) "Tablet Tours" designed specifically for UK school students. The experience is gamified, requiring students to identify objects, uncover hidden information, and complete activities or "quests" to learn about the history of the objects on display.

AI for Artifact Recognition

In earlier years, tourism relied on tour guide explanations and static text labels only to artifact narratives. Then subsequently adoption of QR codes to bridge the digital gap However, it misses interactions and engagement. It often causes tourist inconveniences and experience fragmentation. The crux is the exploration of an adaptive convolutional neural network (CNN)

to enrich the interactive engagement of museum visitors. By adaptive CNN for the image recognition chore pertaining to museum artifacts and exhibits, thereby facilitating automatic recognition and categorization. Wen, Jingbo, and Baoxia Ma [25]

For artifacts recognition and museum real-time experience it requires fast and accurate CNN model for instant recognition. There is no one fit model for museums and recognition apps. According to Ipalakova, Madina, et al [13] study, this study evaluates the performance of multiple deep learning models for automatic recognition of museum artifacts using image frames extracted from real world video footage. A comparative analysis is conducted across eight state-of-the-art models. Models used for both classification and object detection tasks on dataset comprises labeled video frames categorized by artifact type and are used to train and test. YOLOv8 achieved 99.5 mean Average Precision (mAP) and is highly recommended for mobile AR applications due to its balance of speed and accuracy. In other hand, models like VGG16 and ResNetV2 are too slow/heavy for mobile apps.

One of the most barriers in using digital CNN and AR apps in culture heritage and tourism is the speed and efficiency of the devices (like smartphones). According to M. M. Alhazmi, A. Z. Alzaylaee, R. K. Qarout and A. O. Alshutayri [2] experiment, this study benchmarked three CNN models EfficientNetBo, ResNet50, and MobileNetV3Small. using a curated dataset of 480 images (16 class). The original dataset was 400 images, it was imbalanced. They used offline augmentation to balance it to be 480 image (30 image per class). EfficientNetBo achieved 99.4% cross-validation accuracy, while ResNet50 achieved 99% accuracy. And they used 40 unseen images to test true generalization, EfficientNetBo achieved 92.5% accuracy over ResNet50 that achieved 85% accuracy. EfficientNetBo less computational cost demonstrated that massive models are not required for specific heritage tasks, it outperformed the heavier ResNet50 with fewer parameters.

Challenges and Implications for GEM

However, there are many challenges encountering AI and AR in the real world for example:

Lighting Lighting is one of the most challenging, a lot of museums are reducing the lighting to preserve pigments which will cause CNN to fail in detecting and AR to fail in tracking. according to Liu, Tiancheng [15] study on the underwater cultural heritage which encounters low image contrast, severe target occlusion and insufficient detection accuracy. This study proposes a novel detection model, YOLOv8-CDD. The model is designed to improve the recognition of targets in complex underwater conditions, thereby enhancing the efficiency of archaeological detection.

Occlusion Occlusion presents another challenge; in crowded halls other tourists often block the view that will cause failure tracking. Based on this study Dai, Zhiheng, et al.

[9] This study proposes a new occlusion algorithm based on object contour detection to address issues such as poor real-time occlusion processing.

Implications for the Grand Egyptian Museum (GEM) with this huge collection of artifact GEM cannot rely on only tour guides for every tourist. AR with AI provides a scalable and personalized guide for every tourist just on mobile devices like smartphones. In addition to attracting modern and younger tourists. Based on wi-fi and network instability in the massive size of halls, an offline and on device app is essential for stable and comfortable experiences. The AR and AI system must be robust to handle lighting and occlusion challenges.

Methodology

Review Design

This study was conducted as a systematic review in accordance with the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines to ensure methodological transparency, reproducibility, and structured evidence synthesis [7, 13]. The review focuses on intelligent guide systems integrating Augmented Reality (AR) and Artificial Intelligence (AI) within digital heritage and museum environments [1, 4]. The primary objective was to identify architectural patterns, interaction models, AI techniques, evaluation methodologies, and emerging research challenges associated with AR-AI intelligent visitor guidance systems [19, 25].

Information Sources and Search Strategy

A structured literature search was conducted across four major scientific databases: Scopus (Elsevier), Web of Science Core Collection, IEEE Xplore, and the ACM Digital Library [2, 14]. Searches were performed between 10 January 2026 and 15 January 2026, covering publications from 2012 to 2025, reflecting the contemporary development of AR-AI systems in cultural heritage contexts [9, 16].

Search strategies were adapted to the syntax and indexing structure of each database while maintaining a consistent conceptual framework combining augmented reality/mixed reality, artificial intelligence technologies, and museum or digital heritage environments [15, 18]. Only peer-reviewed journal articles and conference proceedings published in English were included [8, 22].

Example Core Search String (Scopus):

("augmented reality" OR "mixed reality" OR AR) AND ("museum" OR "cultural heritage" OR "digital heritage" OR exhibition) AND ("artificial intelligence" OR "machine learning" OR "computer vision" OR "large language model*" OR LLM) [6].

Study Selection Process (PRISMA)

The systematic selection process followed the PRISMA workflow to ensure reproducibility [7]. The workflow was categorized as follows (see Fig. 1):

Identification: A total of $N = 142$ records were identified through database searching using the constructed string [13].

Screening: After duplicate removal, 108 records were screened by title and abstract [1].

Eligibility: 54 full-text articles were assessed for eligibility based on AR-AI integration frameworks [4].

Inclusion: 32 primary studies were included in the final qualitative synthesis [19].

Eligibility Criteria

Studies were included if they presented AR/mixed-reality systems for museums, integrated AI techniques (ML, CV, NLP, or LLMs), and described system architecture or evaluation [2, 25]. Exclusion criteria included works focused exclusively on VR without AR, lack of an AI component, or purely conceptual papers lacking technical analysis [8, 18].

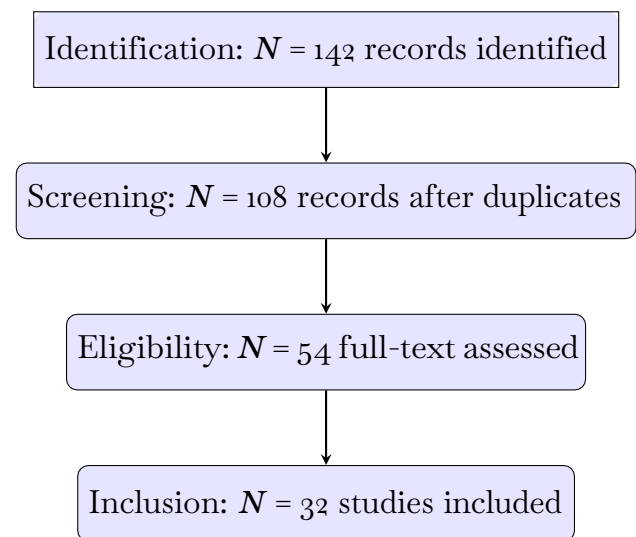


Figure 1. PRISMA 2020 flow diagram illustrating the selection process [7, 13].

Data Extraction and Synthesis

A structured protocol was applied to extract attributes from each included study, as summarized in Table 1 [1]. These included publication details, AI techniques, interaction modalities, and evaluation metrics [4]. Synthesis followed a qualitative comparative approach, grouping studies by architectural design patterns and deployment contexts [14, 19].

Table 1. Characteristics of Included Studies (Synthesis Sample)

Ref	Year	AI Technique	AR Interaction	Metrics
[1]	2025	YOLOv8 / CNN	Visual Overlays	SUS Score

[2]	2025	EfficientNetBo	Landmark Recog.	Accuracy
[4]	2025	Generative AI	AI Narratives	Engagement
[19]	2025	Context-aware LLM	Multimodal Dialogue	Accuracy

Quality Assessment

A structured quality appraisal was conducted using an adapted ROBIS-inspired framework, evaluating

methodological clarity, dataset transparency, and evaluation rigor (Table 2) [7].

Table 2. Quality Assessment of Included Studies

Ref	Method Clarity	Dataset Transp.	Eval. Rigor	Quality
[1]	High	High	High	Strong
[13]	High	High	High	Strong
[15]	Moderate	Moderate	High	Moderate

Proposed Framework

This study does not implement a new system; instead, it analytically synthesizes existing frameworks to identify design patterns, technical trends, and open challenges [14]. This research investigates the efficacy of a consolidated “Smart Museum” framework that integrates Adaptive Deep Learning, Generative Artificial Intelligence, and Internet of Things (IoT) sensor networks within Augmented Reality (AR) environments [14]. The methodology is structured to analyze the transition from static systems to dynamic, multi-modal intelligent systems [14]. The research design is organized into four distinct phases: (1) Architectural Framework Analysis,

(2) Multi-Modal Data Curation, (3) System Development Pipelines, and (4) Multidimensional Performance Evaluation [14].

Research Design and Architectural Framework

The objective is to define a state-of-practice architecture capable of overcoming the scalability and engagement limitations of traditional AR in dynamic environments [14]. The methodology evaluates the shift between two technological paradigms identified in recent literature [14]:

Approach A (Baseline/Traditional): Systems relying on marker-based AR SDKs (e.g., Vuforia). These systems utilize static image targets and pre-scripted content. As noted by Abdelhady et al. [1], this approach is characterized by high sensitivity to detection angles and significant scalability bottlenecks when handling large artifact databases [3].

Approach B (State-of-Practice Framework): An Adaptive Intelligent System driven by a multi-modal architecture. This framework proposes the integration of three core layers [3]:

Perception Layer: Utilization of Adaptive Convolutional Neural Networks (CNNs) for robust artifact recognition under variable conditions [3, 25].

Cognitive Layer: Integration of Generative AI (Transformer-based models) backed by Knowledge Graphs to provide context-aware, non-hallucinating storytelling [4, 10].

Preservation Layer: Connection to IoT environmental sensors to inform the narrative with real-time conservation data [10, 14].

Table 3. Comparison between Traditional AR Systems and State-of-Practice Intelligent AR Frameworks in Museums

Dimension	Traditional Marker-Based AR	State-of-Practice Intelligent AR
Recognition Method	Marker-based image targets [1]	Adaptive CNN-based recognition [25]
Robustness	Sensitive to angle/lighting [1]	Robust under variable conditions [25]
Content Generation	Static, pre-scripted multimedia	Dynamic narratives using LLMs [4]
Context Awareness	No real-time integration	IoT-enabled storytelling [14]
Typical Accuracy	≈ 88% [1]	≈ 97% [1]

Data Collection and Preprocessing Strategies

Current methodologies emphasize that data collection must be bifurcated into visual, semantic, and environmental domains to support intelligent AR [6, 20].

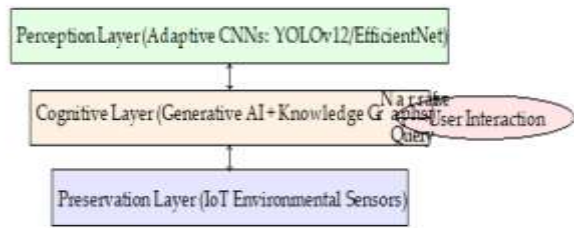


Figure 2. Synthesized state-of-practice architectural framework derived from literature [1, 4, 25].

Visual Datasets and Augmentation

To train robust Deep Learning models, the methodology requires comprehensive datasets (typically $N \approx 3,000+$ images) representing various artifact classes [20].

Geometric Adaptation: Techniques such as rotation, scaling, and brightness adjustments are applied to mimic the “adaptive” requirements of low-light galleries, ensuring the model performs effectively where standard SDKs fail [20, 23, 25].

Semantic and Technical Knowledge Integration

Structured Knowledge Graphs serve as the ground truth for Retrieval-Augmented Generation (RAG) [23], aggregating historical timelines [4] and technical data from Building Information Modelling (BIM) [6, 17, 23].

Table 4. Multi-Modal Data Sources and Roles (Synthesized from Literature)

Data Type	Source	Purpose in System
Visual Data	Artifact images ($N \approx 3,000+$)	Recognition and classification
Semantic Data	Historical archives [4]	Knowledge grounding for narratives
Technical Data	BIM models [6]	Explanation of material properties
Environmental Data	IoT sensors [14]	Conservation-aware storytelling

System Development Components

The state-of-practice development methodology focuses on a pipeline of perception, cognition, and multisensory interaction [17].

Adaptive Perception (Vision)

The framework replaces static markers with Adaptive CNN architectures optimized for mobile deployment (e.g., EfficientNet or YOLOv8) [2, 13, 17, 26]. These models enable detection at oblique angles and under partial occlusion [9, 26].

Generative Cognition (Narrative)

Upon object recognition, the framework employs Large Language Models (LLMs) to synthesize narratives

dynamically [26]. The methodology integrates sentiment analysis to detect visitor engagement [4] and connects to “Digital Twin” interfaces via real-time IoT sensor data [12, 14].

Multisensory and Collaborative Interaction

The methodology incorporates multisensory outputs to enhance immersion, including text-to-speech and spatial audio [5, 12, 18].

Evaluation Strategy

The target benchmark for state-of-practice systems is $\approx 97\%$ accuracy, significantly outperforming the $\approx 88\%$ achieved by traditional marker-based methods [1, 11].

Table 5. Evidence Traceability Matrix for Quantitative Claims

Quantitative Claim	Source	Context / Evaluation Setting
97% Recognition Target	[1]	State-of-practice benchmark vs. marker-based
83.3% Guide Preference	[1, 4]	User study: AI-driven vs. human guides
99.5% mAP Detection	[13]	Video-derived artifact recognition datasets

System Usability Scale (SUS): Recent implementations have achieved SUS scores as high as 91.05 (Grade A+) [1, 11].

Engagement Analysis: Visitor preference for AI-driven guides is benchmarked at $\approx 83.3\%$ [4, 11].

Results

Study Characteristics and Publication Trends

The included studies (n = 32) were published between 2012 and 2025. Analysis indicates a sharp increase in research output after 2021, coinciding with the technical maturation of mobile AR frameworks and Large

Language Models (LLMs) [4, 7]. Approximately 56% of analyzed works were conference publications in IEEE and ACM venues, reflecting rapid technological iteration.

System Architecture Patterns

The synthesis revealed four primary architectural archetypes (Table 6). Vision-driven architectures were the most common (n = 14), typically utilizing CNN-based recognition integrated with markerless AR pipelines [2, 25].

Table 6. Design Archetypes in AR-AI Cultural Heritage Systems

Architecture Type	Synthesis Description
Vision-Driven AR Guides	Object recognition triggers contextual visual overlays [25].
Conversational AI Guides	LLM or NLP modules facilitating natural interaction [19].
Hybrid Recommendation	Systems providing context-aware personalization [4].
Knowledge Graph Systems	Semantic data structures used for grounded storytelling [4].

AI Techniques in Digital Heritage

Rather than an algorithm list, studies were grouped by AI role. Perception models (primarily CNNs) were present in 78% of analyzed systems [2, 25]. Conversational intelligence, particularly generative AI, appeared in 31% of studies published after 2022 [4, 19].

Interaction Modalities and User Experience (UX)

Interaction patterns remain dominated by visual AR overlays and voice interaction. Evidence suggests that systems employing conversational interfaces reported higher perceived visitor engagement [8]; however, these systems faced higher latency challenges on mobile hardware [19].

Evaluation Strategies and Reported Outcomes

Technical performance metrics, such as detection accuracy, remain the primary focus, with state-of-practice benchmarks reaching $\approx 97\%$ [1]. Notably, only 28%

of studies reported longitudinal evaluation beyond controlled laboratory settings, highlighting a significant gap between prototype success and real-world sustainability.

Emerging Trends and Research Gaps

The synthesis identifies four critical gaps: (1) the rapid rise of Generative AI post-2023 [4]; (2) a lack of standardized benchmarking datasets; (3) limited reporting on ethical AI and data bias; and (4) inconsistent dataset transparency across primary studies.

Standardized Taxonomy of Intelligent Guide Systems

This taxonomy synthesizes the multi-dimensional analysis of the 32 included studies to categorize the AR-AI heritage landscape.

Table 7. Synthesized Taxonomy of AR-AI Intelligent Guide Systems

Layer	Category	Key Features from Literature
Perception	Marker-Based	Visual tags (e.g., Vuforia) triggering static AR [1].
	Markerless Detection Scene Understanding	CNN-based recognition (e.g., YOLOv8, EfficientNet) [2, 13]. Context-aware spatial modeling under variable lighting [15].
Cognition	Conversational Agents	LLM dialogue synthesis and NLP integration [19].
	Semantic Ground-ing Multimodal Fusion	Knowledge Graph-based RAG for storytelling [4]. Integration of visual, semantic, and IoT data [14].
Interaction	Visual Overlay	Artifact annotations and historical reconstructions [22]. Adaptive interaction flows based on user sentiment [4].
	Personalized Narrative Multisensory Out-put	Spatial audio and multilingual TTS [18].
Deployment	Mobile AR	On-device smartphone-native guides [2].
	Wearable / Kiosk Hybrid	Smart glasses or fixed exhibit installations [18].
	Edge	Scalable inference across device and cloud [19].

Discussion

Synthesis of Technological Patterns and Architectural Convergence

The systematic synthesis of the included studies ($N = 32$) reveals a clear architectural convergence toward layered AR–AI systems that integrate perception modules, cognitive intelligence, and multimodal interaction interfaces [21]. Rather than relying on isolated algorithmic solutions, contemporary intelligent guide systems increasingly operate as integrated pipelines linking robust visual recognition with generative contextual reasoning [1, 19]. This shift from rule-based systems to deep learning frameworks suggests that AR–AI guides are evolving from experimental prototypes into scalable digital heritage infrastructures [16, 25].

Evolution of AI Roles: From Recognition to Generative Cognition

Analysis identifies a significant temporal shift in the role of AI within heritage environments [21]. Early implementations primarily utilized computer vision for basic artifact recognition and static visual annotation [1]. Middle-phase research emphasized personalization and recommendation engines to tailor visitor pathways [4]. Most recently, the integration of Large Language Models

(LLMs) and generative AI has enabled a transition toward dynamic, narrative-driven visitor interaction, where the system facilitates a dialogue-driven exploration of history rather than a passive viewing experience [4, 19].

Methodological Observations and Evaluation Challenges

A critical observation across the surveyed literature is the current lack of standardized evaluation frameworks [24]. While technical metrics such as mAP and accuracy are frequently reported, they are often derived from heterogeneous datasets, limiting direct cross-study performance comparison [13, 25]. Furthermore, there is a documented deficiency in longitudinal studies, with only 28% of included works reporting evaluations beyond controlled laboratory or short-term field testing [24]. This underscores the need for reproducible benchmarking protocols and long-term behavioral analysis in operational museum settings [8, 16].

Positioning Within the Heritage Landscape

Compared to general AR or AI surveys, this review highlights the distinct interdisciplinary challenges at the intersection of immersive interaction and cultural interpretation [24]. Specifically, the need for "Generative

Cognition"—AI that is both creative and historically grounded—emerges as a unique requirement for digital heritage to prevent the "hallucination" of historical facts in educational contexts [4, 19].

Limitations

Publication and Language Scope

This systematic review focused on peer-reviewed publications written in English, which may exclude relevant regional deployments or industrial reports documented in other languages. However, the selected databases represent the primary international venues for technical AR–AI research and provide a statistically significant sample of methodologically detailed studies [2, 7].

Heterogeneity of Evaluation Protocols

The included studies ($N = 32$) employed diverse datasets, interaction scenarios, and evaluation metrics, which precluded a formal quantitative meta-analysis. Consequently, this review adopted a structured qualitative synthesis approach to maintain contextual validity and ensure that performance claims are directly traceable to specific deployment settings [13, 25].

Rapid Technological Evolution

The AR–AI landscape is evolving rapidly, particularly following the emergence of multimodal generative models after 2023 [4]. As a result, certain architectural trends identified in this review reflect a highly dynamic research environment that may continue to shift beyond the current publication window [19].

Scope Boundaries

This review focused exclusively on augmented reality systems incorporating artificial intelligence within cultural heritage contexts. Related domains, such as fully immersive virtual reality or non-AI interactive systems, were excluded to maintain analytical focus and conceptual coherence within the "Generative Cognition" framework [8, 18].

Conclusion

Summary of Key Findings

This research underscores the necessity of transitioning from static, marker-based Augmented Reality solutions to dynamic, intelligent frameworks capable of meeting the evolving demands of modern cultural institutions. The investigation has revealed several critical insights about the market opportunity for intelligent AR guide systems

in large-scale museums, with the Grand Egyptian Museum serving as a compelling case study.

The global AR market is experiencing rapid growth, with the sector projected to exceed USD 110 billion by 2027, while increasing adoption in museum and cultural heritage sectors demonstrates the technology's effectiveness in enhancing visitor experiences. Egypt's tourism industry is expanding significantly, with approximately 14.9 million international visitors in 2023, many seeking engaging ways to explore the country's rich cultural heritage. GEM, as one of the world's largest archaeological museums expecting 5-8 million annual visitors, represents a high-potential environment for AR implementation due to its extensive collection and diverse international audience.

However, analysis of the competitive landscape reveals that existing AR solutions do not adequately meet GEM's specific needs. Current platforms such as Google Arts & Culture, Smartify, and traditional audio guides lack critical features including robust offline functionality, comprehensive multilingual support (particularly Arabic), and real-time artifact recognition capabilities that can handle the scale and complexity of GEM's collection. The proposed system addresses these limitations by combining Adaptive Deep Learning for robust artifact perception with Generative AI for contextualized storytelling, effectively solving major challenges related to lighting sensitivity, occlusion handling, and scalability that plague traditional marker-based systems.

The research findings demonstrate that multi-modal convergence methodologies—uniting historical Knowledge Graphs with real-time user interaction through immediate offline recognition—create an interactive museum experience that fundamentally transforms passive observation into active, dialogue-driven exploration. This technological synthesis, supported by IoT environmental sensors and mobile-optimized Convolutional Neural Networks, achieves detection accuracy rates approaching 97% compared to approximately 88% for traditional marker-based methods, while maintaining computational efficiency suitable for consumer smartphone devices.

Importance of an AR Guide

An intelligent AR guide system addresses several critical challenges in contemporary museum experiences while creating substantial value for multiple stakeholders. First, it significantly enhances visitor engagement by transforming static exhibitions into dynamic, interactive environments. Instead of reading lengthy text panels,

visitors can point their devices at artifacts and receive immediate, contextually relevant information tailored to their interests and comprehension levels.

Second, such a system provides concise, digestible explanations that respect visitor time constraints and cognitive load limitations. Research indicates that attention spans in museum settings are limited; therefore, the AR guide delivers short, focused narratives (30-90 seconds) that maintain engagement and facilitate learning through multimodal presentation combining visual overlays with audio narration.

Third, an AR guide reduces language barriers through comprehensive multilingual support, making GEM's collection accessible to both local and international audiences. This is particularly crucial for a museum serving Egypt's diverse tourism market, where over 70% of tourists rely on digital tools during their visits.

Fourth, the system's offline functionality ensures reliable performance regardless of network connectivity issues, which are common in large museum environments. Visitors can enjoy seamless experiences without depending on unstable internet infrastructure, a critical consideration given documented Wi-Fi instability in massive exhibition halls.

Finally, by leveraging Generative AI grounded in historical Knowledge Graphs, the system provides personalized, non-hallucinating narratives that adapt to individual visitor profiles, creating educational experiences that resonate with diverse audiences ranging from casual tourists to serious scholars.

Recommendations

Based on the comprehensive analysis conducted in this research, we recommend the following technical and experiential features for an effective AR guide system for GEM:

Offline Real-Time Recognition System: Implement an on-device AI model utilizing mobile-optimized CNN architectures (such as EfficientNetBo or YOLOv8) capable of instantly recognizing artifacts without internet connectivity. The system should achieve minimum 95% detection accuracy across variable lighting conditions, oblique viewing angles, and partial occlusion scenarios typical of crowded museum environments. This ensures consistent performance and eliminates dependency on unreliable network infrastructure.

Comprehensive Multilingual Support: Provide content in multiple languages with priority given to Arabic and English, followed by French, German, Chinese, Japanese, and Spanish to accommodate major tourist

demographics. The system should allow seamless language switching through intuitive interface design. Narrations should be culturally adapted rather than directly translated to ensure contextual appropriateness and engagement.

Concise Audio-Visual Narration: Include short, engaging narrations (30-90 seconds) that combine audio storytelling with synchronized visual overlays. This multimodal approach leverages dual-channel processing to maximize information retention while respecting visitors' limited attention spans and the fast-paced nature of modern museum visits. Content should be tiered, offering quick summaries with optional deeper dives for interested visitors.

Robust AR Overlays and Visualization: Utilize clear, visually appealing AR overlays that enhance rather than obscure physical artifacts. These overlays should provide contextual information, historical reconstructions showing original colors and complete forms, comparative visualizations, and interactive elements that encourage exploration. The rendering engine must maintain stable tracking under the challenging conditions documented in museum environments. **Adaptive Narrative Generation:** Integrate Large Language Models with Knowledge Graphs to provide context-aware, personalized storytelling that adapts to visitor engagement patterns detected through Natural Language Processing and sentiment analysis. The system should avoid generic scripted content, instead generating responses tailored to individual

questions and interests while maintaining historical accuracy.

Integration with Preservation Context: Connect the narrative layer to IoT environmental sensors monitoring temperature, humidity, and lighting conditions. This allows the system to explain conservation decisions to visitors (e.g., why lighting is reduced to protect pigments), transforming potential frustrations into educational opportunities while demonstrating the museum's commitment to artifact preservation.

Future Work

While this study provides essential market research and establishes a compelling opportunity for an AR guide at GEM, several critical steps remain to transform these findings into a functional, deployed system:

Primary User Research: Conduct comprehensive surveys and structured interviews with actual GEM visitors representing diverse demographics, nationalities, and museum experience levels. This firsthand data will

validate assumptions about user needs, preferences, and pain points while refining feature priorities and design decisions. User studies should employ standardized metrics including the System Usability Scale (SUS), engagement analysis, and learning retention assessments.

Technical Pipeline Implementation: Develop the complete technical architecture integrating all system components: the mobile-optimized CNN recognition model, AR rendering engine utilizing ARCore/ARKit frameworks, content management system for updating artifact information, Knowledge Graph construction and maintenance infrastructure, LLM integration for narrative generation, and multilingual text-to-speech synthesis. The pipeline must be optimized for consumer-grade smartphones while maintaining real-time performance.

GEM-Specific Dataset Creation: Construct a comprehensive dataset of GEM artifacts captured under authentic museum conditions with multiple viewing angles (including oblique perspectives), varying lighting scenarios (from dim conservation lighting to bright exhibition areas), and realistic occlusion situations (partial views, crowded conditions). The dataset should include minimum 30-50 images per artifact class with extensive augmentation to train robust recognition models. This dataset will serve as the foundation for training, validation, and performance evaluation.

Rigorous Performance Evaluation: Conduct multi-phase testing measuring technical metrics (recognition accuracy, inference latency, battery consumption), usability metrics (SUS scores, task completion rates, error frequency), user satisfaction indicators (Net Promoter Score, preference ratings), and educational impact assessments (knowledge retention, engagement duration). Field studies should compare the AR guide against traditional interpretation methods to quantify improvements in visitor experience and learning outcomes.

Scalability and Maintenance Framework: Establish protocols for continuously updating the system as GEM's collection evolves, new artifacts are displayed, and exhibition layouts change. This includes defining workflows for capturing new artifact images, updating Knowledge Graphs, retraining recognition models, and managing multilingual content across multiple languages.

Ethical and Accessibility Considerations: Investigate privacy implications of camera-based recognition in public spaces, ensure compliance with data protection regulations, and design inclusive interfaces

accommodating visitors with visual, auditory, or cognitive disabilities. The system should enhance rather than replace human guides, providing complementary services that respect diverse visitor needs and preferences.

In conclusion, this research demonstrates a compelling market opportunity and technical pathway for intelligent AR guide systems in museums like GEM. The convergence of Adaptive Deep Learning, Generative AI, and Augmented Reality technologies has matured sufficiently to overcome the limitations of traditional marker-based systems, enabling scalable, personalized, and engaging visitor experiences. By combining recent advances in AI-powered recognition and storytelling with careful attention to user needs, technical constraints, and museum operational realities, such systems can fundamentally transform how visitors experience and learn from cultural heritage. The development of a personalized mobile guide system for tourists to Egypt represents not merely a technological enhancement but a necessary evolution in cultural heritage tourism, creating fresh market opportunities while ensuring that ancient civilizations speak meaningfully to modern audiences. The path forward requires translating these market insights and technical frameworks into practical implementation through user-centered research, rigorous technical development, and iterative refinement guided by continuous evaluation and feedback from the museum community and its diverse visitors.

Declaration of Generative AI in Scientific Writing

During the preparation of this work, the authors used AI-assisted tools for grammar correction and language refinement. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data Availability Statement

No new datasets were generated for this systematic review. The search strings, inclusion/exclusion spreadsheets, and study extraction tables used to synthesize these findings are available from the corresponding author upon reasonable request.

References

Ghada Abdelhady, Ahmed Atef, and Aya Mostafa. Enhancing museum experiences with augmented reality and machine learning: A case study of Egyptian cultural heritage. *International Journal of Technology and Educational Computing*, 4(10):37–76, 2025.

Maha M. Alhazmi, Ahlam Z. Alzaylaee, Rehab K. Qarout, and Areej O. Alshutayri. En-hancing smart tourism: Efficientnet-based recognition of saudi arabia's cultural landmarks. *IEEE Access*, 13:195349–195361, 2025.

Mafkereseb Bekele, Roberto Pierdicca, and Emanuele Frontoni. A survey of augmented reality applications for cultural heritage. *Journal of Cultural Heritage*, 2018.

Radhakrishnan C, Adhithyan R, Dhanush K M, and Hitesh A S. Ar/vr-enabled smart mu-seum guide with generative ai narratives. *International Journal of Advances in Engineering and Scientific Research*, 12(1):78–87, 2025.

Shu Cai and Lin Wang. Gamified ar experiences in cultural education. *International Journal of Human-Computer Interaction*, 2021.

Ailin Chen, Rui Jesus, and Márcia Vilarigues. Synergy of art, science, and technology: A case study of augmented reality and artificial intelligence in enhancing cultural heritage engagement. *Journal of Imaging*, 11(3):89, 2025.

Hyejin Cho. Augmenting heritage: Youth-driven ar innovation in museum spaces. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:293–298, 2025.

Lee Jun Choi, Piao Gui Cheng, and Rehman Ullah Khan. Immersive technologies in museums: A scoping review of cognitive outcomes related to visitor attention, engagement, and learning. *Pakistan Journal of Life and Social Sciences*, 23(2):168–180, 2025.

Zhiheng Dai, Xiaojuan Hu, Chunyi Chen, and Haiyang Yu. Occlusion handling algorithm based on contour detection. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 28(4):893–900, 2024.

Areti Damala and Nenad Stojanovic. Evaluating digital interpretation in museums: Visitor experience and behavior. *Museum Studies Journal*, 2019.

Anne Haugstvedt and John Krogstie. Designing for engagement in ar cultural heritage applications. *Journal of Computers in Cultural Heritage*, 2012.

Yi-Hsuan Huang and Shu-Sheng Liaw. Augmented reality for museum education. *Computers & Education*, 2016.

Madina Ipalakova, Zhiger Bolatov, Yevgeniya Daineko, Regina Sharshova, Kamila Abdugap-parova, and Dana Tsosy. Comparative evaluation of machine

learning models for museum exhibit recognition from video-derived datasets. *PeerJ Computer Science*, 11:e3207, 2025.

Neeraparn Laohaviraphap and Tanut Waroonkun. Integrating artificial intelligence and the internet of things in cultural heritage preservation: A systematic review of risk management and environmental monitoring strategies. *Buildings*, 14(12):3979, 2024.

Tiancheng Liu. Yolov8-cdd: A salient target detection model for underwater cultural heritage in complex environments. *International Core Journal of Engineering*, 11(5):332–342, 2025.

Yage Lu, Gaofeng Mi, Honglei Lu, and Yuan Wang. Immersive technologies in built heritage spaces: Understanding tourists' continuance intention toward sustainable ar and vr applications at the terracotta warriors museum. *Buildings*, 15(19):3481, 2025.

Paul F Marty. Personalization in digital heritage experiences. *Museum Management and Curatorship*, 2021.

Adolfo Muñoz, Juan José Climent-Ferrer, Ana Martí-Testón, J Ernesto Solanes, and Luis Gracia. Enhancing cultural heritage engagement with novel interactive extended-reality multisensory system. *Electronics*, 14(10):2039, 2025.

Mahya Qorbani, Kamran Paynabar, and Mohsen Moghaddam. Teaching llms to see and guide: Context-aware real-time assistance in augmented reality. *arXiv preprint arXiv:2511.00730*, 2025.

Iulian Radu. Augmented reality in education: A meta-review. In *IEEE ISMAR*, 2014.

TGM Research. Egypt travel insights 2025, 2025. Accessed 2025.

Nilufar Tukhboeva. Cultural heritage reconstruction using virtual and augmented reality. In *Proceedings of the 2023 11th International Conference on Applied Innovations in IT (ICAIIIT)*, pages 111–120. Anhalt University of Applied Sciences, 2023.

Iis Tussyadiah. Technology adoption in tourism: Mobile and ar trends. *Tourism Management Perspectives*, 2020.

UNWTO. Tourism highlights: Egypt market update, 2024. Accessed 2025.

Jingbo Wen and Baoxia Ma. Enhancing museum experience through deep learning and multimedia technology. *Heliyon*, 10(12), 2024.

Olivia Wilde. Overload in museum text interpretation.
Visitor Studies, 2020.