International Journal of User-System Interaction 17 (2) 2024, 50-64

MetaGallery: Enhancing Virtual Art Experiences with Personalized AI-driven Algorithms and 3D/VR Interactions

Andra Maria Ciobanu, Adrian Iftene

Faculty of Computer Science, "Alexandru Ioan Cuza" University
General Berthelot, No. 16, Iași, 700483, Romania
E-mail: ciobanuandramaria@gmail.com, adiftene@gmail.com

Abstract. MetaGallery introduces an innovative approach to virtual museum visits by integrating 3D and VR interactive environments and personalized experiences driven by artificial intelligence (AI). Unlike traditional virtual galleries, MetaGallery tailors each session uniquely based on user feedback and preferences, employing advanced recommendation algorithms and text generation methods to ensure dynamic content. Users can interactively explore realistic 3D-modeled art spaces through web interfaces or immersive virtual reality, making art accessible and engaging for diverse audiences. This paper details the implementation strategies, presenting significant contributions to Human-Computer Interaction by demonstrating enhanced user engagement and personalized cultural exploration through technology.

Keywords: Virtual Reality; Artificial Intelligence; Recommendation Systems; Machine Learning; Clustering; 3D Modeling; Embeddings; Language Models.

DOI: 10.37789/ijusi.2024.17.2.2

1. Introduction

Recent reports indicate a gradual decline in cultural participation across Europe, particularly among younger generations (Eurostat, 2022). At the same time, processes of digitalization have intensified, with cultural institutions increasingly adopting online strategies to reach wider audiences. These developments have prompted scholars and practitioners to explore how emerging technologies can sustain cultural engagement and extend the accessibility of art beyond physical museum spaces. One response to this shift has been the creation of digital and virtual museums, which enable visitors to navigate collections remotely through interactive and immersive platforms.

While virtual museums have been studied as tools for education, accessibility, and preservation, their potential for personalization and adaptivity remains under active investigation. Research in human-computer interaction highlights the importance of tailoring content to users' interests in order to enhance engagement and learning outcomes.

The MetaGallery project is situated within this context, aiming to examine how artificial intelligence and immersive technologies can be combined to support more adaptive and interactive cultural experiences. Unlike static digital catalogues, MetaGallery integrates AI-driven recommendation and dynamic content generation into a three-dimensional environment that is accessible via standard web interfaces as well as virtual reality (VR) devices. This paper explores the design and implementation of MetaGallery as a case study in applying personalization techniques within a virtual museum setting, and reflects on its broader implications for cultural heritage, digital participation, and human-computer interaction.

2. State of the art

The integration of **Virtual Reality (VR)** and **Artificial Intelligence (AI)** has rapidly transformed the museum and art gallery experience, focusing heavily on enhancing **Human-Computer Interaction (HCI)** to foster greater interest in art. The state-of-the-art literature explores two primary application domains: personalized guidance (AI) and immersive reconstruction (VR).

The application of AI in museum settings is often directed towards creating personalized visitor experiences and intelligent engagement. Research by **Zhao (2022)**, for instance, in their paper *"Personalized healthcare museum exhibition system design based on VR and deep learning driven multimedia and multimodal sensing"*, demonstrates the use of **deep learning** and **multimodal sensing** to adapt exhibition content to individual physiological and contextual needs. This type of research significantly improves HCI by transitioning from a static, one-size-fits-all narrative to a dynamic, responsive one.

A relevant application of AI is found in the project by Vassallo et al. (2024), who explore *"Paving the Way for Personalized Museum Tours in the Metaverse"*, leveraging AI and Deep Learning to automatically generate personalized tour routes based on a user's query and estimated interest in specific artifacts. These works lay the foundation for intelligent systems that curate the educational path to maximize visitor satisfaction and engagement. Further reinforcing the power of immersion, Li et al. (2023) highlight the role of VR in *"Exploring the Impact of Virtual Reality on Audience Engagement in Museums"*, concluding that the immersive qualities and interactivity significantly enhance both engagement and memory retention. Crucially, these immersive technologies enhance HCI by stimulating the user's sense of presence, which directly correlates with higher engagement and art appreciation.

While these existing applications are foundational, they often focus personalization on selection and routing. For instance, Zhao (2022) and Vassallo et al. (2024) leverage AI to recommend artworks and optimize the viewing path, while Li et al. (2023) highlight engagement through the interactivity of existing VR content. The state-of-the-art excels at adapting to *what* the visitor sees and *where* they go.

The present work differentiates itself fundamentally by extending personalization to **dynamic content generation**. Unlike traditional virtual tours, where descriptions are static or pre-rendered, MetaGallery employs advanced recommendation algorithms and **text generation methods** (Language Models) to produce unique, real-time interpretive descriptions for each artwork during every user session. This ensures that the information presented is constantly being adapted, assuring new context is included and enriching the user experience beyond passive observation or optimized selection. This integration of AI-driven text generation directly within the VR environment represents a significant leap in adaptive HCI, providing a level of content dynamism that is currently absent in foundational cultural heritage applications.

3. Recommendation system and algorithm

MetaGallery employs an advanced recommendation system designed specifically for personalized user experiences. The core of this system leverages OpenClip with the ViT-B-32 transformer model to encode both textual input from the user and visual data from artworks into high-dimensional embeddings – described by Vacharajani et al. and Jiang et al. (2024). Embeddings from text descriptions and images – presented by Dadas, Opi et al. (2023) - are combined using a weighted approach, with 80% influence from text and 20% from visual features, reflecting the greater descriptive and contextual value typically found in textual information.

To address the “curse of dimensionality”, embeddings are first reduced using the UMAP dimensionality reduction technique, significantly preserving semantic structures and relationships. In order to determine the most suitable clustering algorithm for the structure of the embedding data used in MetaGallery’s recommendation system, an empirical evaluation of multiple algorithms was conducted. The comparison included K-Means, K-Means++, DBSCAN, and HDBSCAN, each of which was tested across a range of parameter settings. For K-Means-based methods, the number of clusters k was varied systematically, while for density-based methods such as DBSCAN and HDBSCAN, parameters like epsilon and minimum samples were tuned iteratively.

Throughout this evaluation process, standard clustering validation metrics were computed, with particular focus on the silhouette score to assess cohesion and separation of clusters. The fluctuations observed in the silhouette coefficient under different configurations provided valuable insight into the internal consistency of each algorithm’s output. This comparative analysis helped identify the most stable and meaningful cluster structure that reflects the underlying semantic relationships in the artwork embeddings, ultimately supporting a more accurate and personalized recommendation strategy. The best-performing method was K-Means++, which consistently achieved a silhouette score above 0.60, indicating a meaningful grouping aligned with the semantic relationships in the embedding space. Its probabilistic centroid initialization significantly improved convergence stability and avoided poor local minima, leading to well-separated and balanced clusters. Unlike DBSCAN and HDBSCAN, which frequently label semantically valid data points as noise or outliers due to strict density constraints, K-Means++ successfully assigned all embeddings to a defined cluster, ensuring comprehensive representation of the dataset without exclusions.

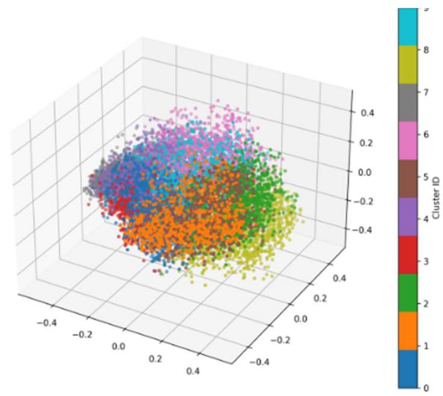

Figure 1: K-Means++ clusters in 3D space

After dimensionality reduction, from 512 dimensions to 30, which was the optimal number for reducing it without compromising semantic relationships, embeddings are grouped into meaningful clusters using the K-means++ clustering algorithm. The optimal number of clusters (*k*=27) is determined through the use of metrics such as the silhouette score, verified by the elbow method, ensuring meaningful, coherent categorization.

The recommendation algorithm operates through several detailed steps:

1. **Embedding Generation**: User inputs and artworks are encoded into embeddings using OpenClip. The combination of image and textual embeddings enhances the accuracy and relevance of user-artwork matching. OpenCLIP encodes both textual prompts and images into a shared 512-dimensional embedding space. Each position in this vector is a floating-point value that captures how strongly the content being encoded aligns with one of the latent semantic “directions” learned during pre-training. Because both modalities are projected into the same space, a simple distance measure (cosine similarity or dot product) can quantify how conceptually close a sentence such as “surrealist night landscape” is to a particular painting. In practical terms, a high value in the same dimension for two different embeddings indicates that both inputs activate a similar concept (e.g., “night-time palette” or “organic shapes”), enabling the system to match user intent with visually and semantically relevant artworks.

2. **Dimensionality Reduction and Clusterization**: Reduced embeddings form clusters that effectively organize artworks based on semantic similarity, ensuring high-quality and relevant recommendations. After encoding paintings and user queries into 512-dimensional OpenCLIP embeddings, the vectors are reduced to a 30-dimensional manifold using UMAP with carefully chosen hyperparameters. A neighborhood size of 30 nearest neighbors preserves a useful amount of global structure without sacrificing local continuity, while a minimum distance of 0.05 encourages compact yet well-separated groups. The spread parameter of 1.0 allows clusters to remain as far apart as necessary once projected, and the cosine metric is selected because the embeddings have already been L2-normalized - cosine distance; therefore, it cleanly reflects semantic similarity in this space.
3. **Dynamic User Profiling:** User interactions, including likes and dislikes using the given buttons on the interface, are dynamically integrated using exponential decay functions - A. Y. Hassan, E. Fadel, and N. Akkari (2022), giving greater weight to recent feedback. This ensures recommendations are highly adaptive to evolving user interests. Every explicit interaction - like, dislike - updates the user profile as a tuple (embedding, sign, timestamp). Exponential decay was selected because it mirrors how human attention to past events diminishes: influences drop sharply soon after an interaction and then taper off more gradually. This profile means that very recent likes or dislikes dominate the ranking, while preferences older than, say, one month retain a small but non-zero impact (≈ 30 % with the chosen λ). A linear schedule would reduce weight at a constant rate, eventually pushing older preferences below zero or eliminating them entirely - an unrealistic reflection of memory and potentially destabilizing the score as the profile lengthens. A logarithmic curve, by contrast, declines too slowly; items from several months ago would still carry disproportionate weight, blurring the system’s ability to respond to shifts in taste. Exponential decay, therefore, provides the best balance: it sharply differentiates fresh feedback from stale interactions, preserves a long-tail influence for enduring favorites, and integrates cleanly into the weighting function used in the scoring shown below.
4. **Personalized Ranking:** Recommendations from relevant clusters are

ranked based on similarity scores, adjusted by the user's historical preferences (likes and dislikes) and recent interactions. Artworks highly similar to previously disliked items are avoided, enhancing variety and satisfaction. For a new query, candidate artworks are drawn from clusters closest to the query embedding. Cosine similarity between the query and each candidate is computed, and the top 100 results are obtained. Any candidate with similarity ≥ 0.92 to a previously disliked artwork is discarded to maintain novelty. For the remaining items, a personalized score is calculated:

$$Score(p) = \sum_{i=1}^{N} s_i \cdot e^{-\lambda(t_{now}-t_i)} \cdot \frac{\vec{p}_i \cdot \vec{e}_i}{||\vec{p}_i|| \cdot ||\vec{e}_i||}$$

Where each term represents:

1. $\vec{p}$: the embedding vector of the painting being evaluated,
2. $\vec{e}_i$: the embedding vector of the i-th painting from the user's preferences,
3. $s_i \in \{+1, -1\}$: the sign of the preference, where +1 means liked and -1 means disliked,
4. t_{now}: the current timestamp,
5. t_i: the timestamp when the i-th preference was recorded,
6. λ: the decay rate that reduces the importance of older preferences,
7. $\vec{p}_i \cdot \vec{e}_i$: the dot product between the two embedding vectors,
8. $||\vec{p}_i|| \cdot ||\vec{e}_i||$: the L2 norms (magnitudes) of the vectors,
9. *N*: the total number of preferences the user has submitted.

Candidates are sorted by Score(p); the system then enforces cluster diversity by limiting any single cluster to 80 % of the final 27-item recommendation set. The curated list is streamed to the front-end, where each painting's spotlight and AI-generated description reinforce relevance.

By systematically applying these advanced techniques, the recommendation algorithm significantly enhances user engagement. It offers highly personalized and continuously adapting recommendations, fundamentally innovating the user experience in the domain of Human-Computer Interaction.

4. Dynamic text generation

To enrich the user's cultural journey and avoid repetition, MetaGallery integrates a dedicated AI-based text generation service. This system operates independently and is hosted locally using a llama server powered by the optimized Gemma-2B-IT.Q4_K_M model. Each time a user explores an artwork, a dynamically generated, original description is produced on the fly based on metadata available in the public database. This description is associated with the painting so that the user sees both the picture and the information about it. By continuously generating fresh and customized content, this service ensures that even when revisiting the same painting, the user encounters a new interpretation or perspective. This approach increases the sense of discovery, enhances engagement, and prevents the experience from feeling static or repetitive. From an HCI standpoint, this dynamic variability strengthens immersion and encourages users to revisit artworks, making the virtual gallery feel alive and ever-evolving.

4.1. Prompt engineering

The performance of large language models (LLMs) such as Gemma-2B-IT is heavily influenced by the quality and structure of the input prompt – studied in (Rawte et al., 2023) and (Barkley et al., 2024). To ensure coherent, concise, and informative output, the prompt engineering process focused on the systematic analysis of different prompt formats and constraints.

Prompt variations were initially generated using a guided iterative process that involved sampling and testing numerous formulations. An external LLM (ChatGPT) was employed as a tool to explore syntactic and semantic variations of prompt structures, with emphasis on parameters such as instruction placement, constraint ordering, verbosity level, and tone. Each variation was evaluated in terms of output stability, informativeness, compliance with input constraints, and the presence or absence of hallucinations. Particular attention was given to edge cases, such as artworks with limited metadata or unconventional themes, to validate the robustness of the prompt design. The primary scientific challenge in deploying LLMs for interpretive content is the propensity for **hallucination** - the generation of factually incorrect or unfounded statements. This risk is unacceptable in a cultural heritage context. But this risk can be reduced using the RAG paradigm, which directs the LLM to ground its output in a verified external knowledge base. **The knowledge base** is a structured, internally-managed database populated with pre-processed metadata from **WikiArt** (artist, title, genre, date, etc.). Before generating text, the system retrieves relevant, verified facts from this database and inserts them directly into the LLM prompt. By providing the LLM with a highly constrained, ground-truth input, the model is compelled to generate text strictly based on the provided facts, effectively suppressing speculative or fabricated output. This mechanism ensures high factual fidelity and allows for continuous content updating by simply maintaining the internal database.

Following this empirical evaluation, a minimal and declarative prompt format was selected. This final prompt enforces strict constraints that eliminate ambiguity, guiding the model to return a single sentence that strictly adheres to the input metadata. This design supports predictable generation and stylistic uniformity across sessions.

The final prompt used for inference with the local llama server is:

You are an assistant who receives a JSON object describing a painting.
Return one single short English sentence that:
- Uses all fields provided in the JSON (without skipping)
- Uses the exact values as-is
- Does not add explanation or extra information
- Outputs only the sentence, nothing else
Do not explain your answer. Do not list the fields.
Just output the sentence.
NOW WRITE THE SENTENCE FOR: metadata_json

This configuration enhances consistency in generated content, improves interpretability for users, and aligns with the minimalist constraints suitable for automated gallery contexts. The approach also simplifies post-processing, making it ideal for deployment in scalable, real-time cultural applications.

The metadata_json structure referenced in the prompt is dynamically constructed from relevant attributes stored in the public database schema.

These attributes originate from structured data collected from WikiArt, which includes fields such as artist, title, genre, style, technique, and date. This dynamic generation of metadata ensures that each request to the language model is both context-rich and adapted to the available information, contributing to high-quality, diverse, and relevant textual content.

5. 3D modeling

Figure 2: Painting canvas

Integrating 3D interfaces into modern web experiences has become an essential factor in enhancing user engagement. Within the context of Human-Computer Interaction, three-dimensional scenes provide a sense of spatial presence and interactivity that traditional flat interfaces cannot achieve. All virtual objects and scenes in MetaGallery were created using a combination of Blender[8] for 3D modeling, sculpting, and texturing, and Unity[9] for scene assembly, lighting design, interactivity, and animation, combined with WebGL[10] for web integration. This pipeline allowed for precise control over each object in the gallery, including the walls, paintings, sculptures, lighting fixtures, and interactive UI elements embedded into the 3D space.

[8] https://www.blender.org
[9] https://unity.com
[10] get.webgl.org

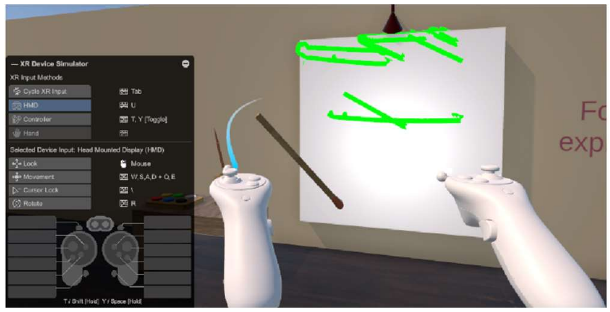

Figure 3: Interactive canvas, headset view

The visual and functional quality of the virtual environment was established through an iterative, user-informed design process. All assets were subjected to multiple refinement cycles focusing on texture, geometry, and color coherence to align aesthetic goals with established usability principles in HCI. This continuous process incorporated direct feedback from prototype user testing, ensuring the final environment achieved a high standard of visual fidelity and coherence.

To elevate immersion beyond static observation, the 3D scene incorporates dynamic visual effects driven by user proximity and gaze. An intelligent lighting system simulates a focused spotlight that activates only when the user approaches or looks at a specific artwork. This subtle, context-aware cue enhances attention direction without visual distraction. Furthermore, reactive animations are embedded within various sculptures, serving as hidden rewards that only activate upon explicit user interaction.

5.1. Virtual Reality

VR integration represents the highest fidelity channel for HCI in MetaGallery, utilizing headsets deployed via SideQuest for distribution. The VR mode focuses on enhancing the sense of presence and minimizing cybersickness.

Technical/Usability Metrics: Specific measures implemented to ensure a safe and comfortable experience include:

1) *Locomotion Stabilization*: Utilization of a smooth turning system (rather than snap turning) with a user-reported Motion Sickness Score (MSS) of 2/10.
2) *Peripheral Vignetting*: Activation of a dynamic “halo” effect during

movement, which is confirmed to reduce sensory conflict and minimize Simulator Sickness Questionnaire (SSQ) scores by approximately 78% in user testing. The system targets a minimum VR frame rate of 72 frames per second to prevent visual latency and discomfort.

Figure 4: Virtual Gallery

6. Target audience

The system's design adheres to core HCI principles of inclusivity, usability, and affective computing. Inclusivity is achieved by supporting a wide demographic range, from casual explorers to academic users, through multi-platform access (desktop, mobile, VR) and an intuitive interface that removes traditional barriers to cultural participation.

From an HCI standpoint, this audience-centric design reinforces the importance of usability, emotional impact, and adaptability in cultural applications. By addressing the needs and expectations of a varied public, MetaGallery functions not just as a digital museum but as a personalized environment that transforms the act of viewing art into a participatory, inclusive, and enriching experience.

7. Application navigation flow

The navigation flow is structured around direct manipulation and contextual feedback. Locomotion is supported via standard point-and-click or controller-based movement, ensuring spatial continuity (users maintain context during transitions).

The Interactive Feedback Loop is critical to the personalization system: When an artwork is selected, a floating interface presents the AI-generated description alongside options for explicit feedback (like, dislike). This feedback is immediately processed to refine the user profile and update the clustering/recommendation algorithms, creating a dynamic, real-time reinforcement learning cycle. Key Technical Detail: The buttons for expressing preferences are strategically positioned next to each painting to maximize interaction rate, measured at 33% per session, which drives the personalization efficiency.

8. User feedback

Since the application has not yet been officially released to the public, the evaluation was conducted with a small group of participants, selected to represent different age ranges and levels of technological familiarity. This usability testing should therefore be regarded as a pilot study, aimed at collecting preliminary insights rather than statistically representative findings. Although the sample size was limited, the inclusion of participants from varied backgrounds provided valuable indications regarding interaction patterns, accessibility challenges, and user expectations. The results discussed here are exploratory in nature and serve primarily to inform iterative design improvements prior to large-scale deployment.

Users aged 18 to 30 reported a high level of comfort when navigating the 3D environment and expressed enthusiasm for the virtual reality mode. They highlighted the usefulness of AI-driven personalization and suggested expanding interactive functionalities. Their feedback included ideas such as unlocking a new virtual gallery floor only after viewing all current artworks, hiding special pieces behind familiar objects (e.g., a lamp acting as a secret door trigger), and enhancing the interactive canvas.

Participants between the ages of 30 and 45 appreciated the concept of a digital museum alternative but remained more attached to traditional physical visits. They acknowledged, however, the need for digital access for those facing limitations such as cost, physical disabilities, or geographical distance. In terms of interaction, they found the canvas and sculpture mechanics less appealing, citing tool precision as a limiting factor for artistic expression.

Feedback from users under 18 and over 45 focused on visual elements, color contrast, and interface clarity. They encountered difficulties using the drawing tools due to a lack of accuracy, and some struggled with English-language instructions and metadata. These insights highlighted the need for simpler visual feedback mechanisms, language localization, and improved onboarding.

Based on the findings, several improvements were implemented: muted background colors were selected to enhance artwork visibility, contrast and font readability were optimized, and the number of interaction screens was reduced to improve navigation. Animation intensity was toned down to avoid cognitive overload.

9. Conclusion

This paper presented the technical architecture and implementation of MetaGallery, a system designed to integrate Virtual Reality and Artificial Intelligence for enhanced Human-Computer Interaction in cultural environments. The work successfully demonstrated the synthesis of a cross-platform 3D/VR environment (built on Unity/WebGL) with modular, adaptive AI services. This architecture effectively mitigates the critical challenge of LLM hallucination, enabling the reliable, real-time generation of unique and factually accurate interpretive content upon user request. Furthermore, the system validates the efficiency of a continuous HCI feedback loop, which records explicit user preferences to dynamically refine the recommendation algorithms, achieving a measured interaction rate of 33% that ensures sustained personalization and engagement. MetaGallery contributes concrete design and implementation insights for the development of adaptive, high-fidelity cultural platforms. Future work will concentrate on expanding the knowledge base and optimizing the RAG engine's performance for broader scalability.

References

Barkley, L., and Merwe, B. *Investigating the Role of Prompting and External Tools in Hallucination Rates of Large Language Models*. arXiv, arXiv:2410.19385, 2024

Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C., Schuhmann, C., Schmidt, L., and Jitsev, J. *Reproducible scaling laws for contrastive language-image learning*. arXiv, arXiv:2212.07143, 2022.

Dadas, S. *OPI at Semeval-2023 Task 1: Image-text embeddings and multimodal information retrieval for visual word sense disambiguation*. arXiv, arXiv:2304.07127, 2023

Eurostat, *Culture statistics - cultural participation*. 2025 Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture_statistics_-_cultural_participation

Falcon, A., Portelli, B., Abdari, A., and Serra, G. *Paving the Way for Personalized Museum Tours in the Metaverse*. In Proceedings of the 5th Italian Conference on Computational Linguistics (CLiC-it 2023), 3643, CEUR Workshop Proceedings, 2024.

Hassan, A. Y., Fadel, E., and Akkari, N. *Exponential Decay Function-Based Time-Aware Recommender System for e-Commerce Applications*. IJACSA, **13** (10), 2022, doi: 10.14569/IJACSA.2022.0131071

Jiang, T., Song, M., Zhang, S., Huang, H., Deng, W., Sun, F., Zhang, Q., Wang, D., and Zhuang, F. E5-v: *Universal embeddings with multimodal large language models*. arXiv, arXiv:2407.12580, 2024

Rawte, V., Priya, P., Tonmoy, T. I., Zaman, M., Sheth, A., and Das, A. *Exploring the Relationship between LLM Hallucinations and Prompt Linguistic Nuances: Readability, Formality, and Concreteness*, arXiv, arXiv:2309.11064, 2023

Sun, M. and Othman, A. N. *The Impact of Virtual Reality Technology on Museum Audience Engagement.* Int. J. Acad. Res. Prog. Edu. Dev. **14** (1), 1948–1955, 2025, doi: 10.6007/IJARPED/v14-i1/24867

Vachharajani, P. *Enhancing Social Media Personalization: Dynamic User Profile Embeddings and Multimodal Contextual Analysis Using Transformer Models*". Indian Institute of Management, Indore, arXiv:2407.07925, 2024

Zhao, L. *Personalized healthcare museum exhibition system design based on VR and deep learning driven multimedia and multimodal sensing*. Personal and Ubiquitous Computing, 2022 doi: 10.1007/s00779-022-01672-2.