

Article

Exploring Multiverses: Generative AI and Neuroaesthetic Perspectives

Maurizio Forte ^{1,2}

¹ Department of Classical Studies, Art, Art History and Visual Studies, Duke University, Durham, NC 27708, USA; maurizio.forte@duke.edu

² Department of State, Advisor Office of Cultural Heritage—Neuroaesthetics and Cultural Diplomacy, Washington, DC 22209, USA

Abstract: This paper examines the transformative potential of generative artificial intelligence (AI) and neuroaesthetic methodologies in archaeology, museum collections and art history. It introduces the concept of the AI multiverse, which allows archaeologists and social scientists to construct multiple plausible reconstructions of ancient environments and cultural practices, addressing the inherent uncertainties in archaeological data. Generative AI tools create simulations and visualizations that redefine traditional archaeological frameworks by incorporating multivocal and dynamic interpretations. The study also integrates visual thinking strategies (VTSs), eye tracking and saliency map analyses to investigate how structured observation enhances cognitive and emotional engagement with visual artifacts. A case study involving the painting *My Mother, She Fell From the Sky* highlights the impact of VTS on guiding viewers' gaze and improving interpretive depth, as evidenced by heatmaps and saliency distribution.

Keywords: multiverse; generative AI; neuroaesthetics; saliency maps; eye-tracking; vision; human perception



Academic Editors: Paolo Rosati,
Marco Raoul Marini and
Augusto Palombini

Received: 28 January 2025

Revised: 2 March 2025

Accepted: 5 March 2025

Published: 10 March 2025

Citation: Forte, M. Exploring Multiverses: Generative AI and Neuroaesthetic Perspectives. *Heritage* **2025**, *8*, 102. <https://doi.org/10.3390/heritage8030102>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The application of generative artificial intelligence (AI) in archaeology, art and the humanities is revolutionizing how researchers interpret, visualize and reconstruct material culture, from the past to the present [1]. Generative AI [2] refers to systems that create new content or data by learning patterns from existing information, making it particularly well-suited to tackling the challenges of incomplete or fragmented archaeological records. The world of digital archaeology is marked by various definitions, milestones and methodological revolutions: virtual archaeology in 1996, cyberarchaeology in 2008, and (generative) AI archaeology in 2022.

If virtual archaeology was mainly focused on computer graphic reconstructions [3], and cyberarchaeology [4] on virtual reality and simulations AI archaeology is centered on the idea of the past as a multiverse.

Virtual archaeology mostly relied on a unidirectional approach to address a degree of uncertainty: a photorealistic reconstruction founded on precise and validated facts. The downside of this approach was to imagine the reconstruction of the past as a single digital world. Computer graphics and photorealism were often very convincing in the validation of these kinds of digital data. Cyber-archaeology was predominantly accessible to various reconstructions and simulations, with the interaction of models being crucial. AI-archaeology is inherently amenable to infinite iterations of visualizations, reconstructions, and simulations. This situation addresses the methodological issue of uncertainty [5] through

exponentially expanding potential knowledge. The concept of the multiverse refers to a hyperinformative realm without limitations for the development and comparison of ideas, theories, and visions. This multiplication of content [6] coming from machine learning [7] can reframe previous views and open new perspectives (Figure 1). The Venn diagram in Figure 1 shows the evolution of virtual and cyberarchaeology in AI-archaeology [8] and the potential of neural networks in the simulation-reconstruction process. The shared use of digital tools connects virtual archaeology and cyberarchaeology, highlighting common technologies like 3D modeling and photogrammetry. The overlap between virtual archaeology and AI archaeology introduces innovative historical concepts, where AI predicts missing elements and generates alternative reconstructions. The intersection of cyberarchaeology and AI archaeology focuses on advanced virtual simulations, integrating AI-driven dynamic interactions within immersive environments. At the center, the convergence of all three fields forms future-oriented reconstruction models, where digital tools, immersive simulations, and AI-driven insights merge to create dynamic, interactive, and predictive archaeological reconstructions that allow for multiple interpretations of the past.

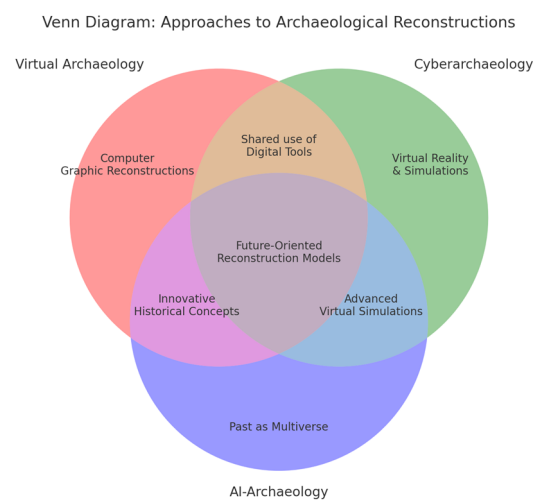


Figure 1. Venn diagram: intersections and goals of virtual archaeology, cyberarchaeology and AI archaeology.

We can imagine the AI multiverse as a space of increased knowledge, rather than the simple result of a validation process [9]. Multiverse AI paves the way for multifactorial analyses by combining different ontologies of data and models. The more we generate AI models, the more we learn and interpret [10]. It is a different approach in comparison with the usual bottom-up process in archaeology; data recording-single hypothesis-validation. The link between data intake and output in generative AI is subject to multiple validations: in theory endless processes, in practice the final ones are selected by specific scholarships and consistency between research questions and final results. This methodology contradicts the concept of reconstructing the past as a singular “snapshot” due to the dynamic and developing characteristics of time and space; it is unfeasible to “freeze” the past, a city, a site, or a landscape, as they perpetually coevolve into something new.

Generative ideas stimulate new and more advanced research questions, and this aspect is extremely powerful when we deal with simulations of ancient societies, particularly in relation to human activities. I believe that the “human factor”—the interaction between human activities, minds, artifacts, and built and natural environments—can be properly investigated by AI [11].

Generative AI operates through machine learning models, such as neural networks, that analyze vast datasets—including excavation records, artifact photographs, and environmental data—to identify patterns and relationships. Once trained, these models can

produce simulations, visualizations, and predictions that help fill gaps in the archaeological record. One notable application of generative AI in archaeology is its ability to create detailed 3D models of ancient structures and artifacts. By training on datasets of similar objects or architectural styles, AI can generate reconstructions that align with known archaeological and cultural contexts.

In this paper, the research focus will be mainly on the use of AI in the interpretation and perception of artifacts, images and visual information. Human vision is, in fact, multimodal [12] and it is based on ranking, segmentation and discretization of the content [13]. Multimodality in vision refers to the integration of various sensory inputs and cognitive processes that help humans perceive and interpret visual information. Vision does not operate in isolation; it interacts with other modalities such as touch, sound, and even memory. Human vision prioritizes certain elements of a scene based on saliency and relevance. This involves ranking objects or features in terms of importance for survival, decision-making, or attention. The visual system breaks down a complex scene into distinct regions or objects to make sense of it. It prioritizes certain elements of a scene through ranking, focusing attention on features like brightness, contrast, and emotional significance, which are most relevant to survival or decision-making. Simultaneously, it employs segmentation to break down complex visual scenes into distinct objects or regions by identifying edges, contrasts, and spatial boundaries [14]. Depth perception further aids this process by distinguishing the foreground from the background, enabling spatial awareness. Through discretization, the brain isolates and categorizes visual features, organizing continuous input into manageable units for interpretation. These processes collectively allow humans to navigate, understand, and interact with their surroundings effectively. Saliency maps and neuroaesthetic experiments show that it is sufficient to analyze just a percentage of an image in order to create and memorize a visual narrative. Additionally, research on the primary visual cortex (V1) suggests [15] that V1 generates a saliency map to guide attention, allowing observers to focus on the most informative parts of a scene. This mechanism enables efficient processing and memory retention by prioritizing salient regions over less conspicuous areas.

2. The Multiverse

The idea of the multiverse—a concept suggesting the existence of multiple, parallel realities [16]—has traditionally been the domain of physics [17] and speculative fiction. The concept of the multiverse, which posits the existence of multiple, parallel realities or universes, has traditionally been the domain of theoretical physics and speculative fiction. In physics, it emerges from theories such as quantum mechanics, string theory, and cosmic inflation, suggesting that our universe might be just one of many coexisting in a vast and complex multiverse [18]. Beyond physics, the multiverse has captured the imagination of writers and creators, serving as a rich narrative framework for exploring themes of identity, choice, and existence. Recently, the multiverse concept has transcended its original boundaries, entering philosophical discourse, popular culture, and even discussions in metaphysics and theology, as it invites profound questions about the nature of reality and our place within it [19].

However, its principles can be fruitfully applied to the realm of artificial intelligence (AI) in reconstructing and simulating ancient environments, societies and human lives. By leveraging the multiverse framework, archaeologists can develop a richer, multidimensional understanding of the past, where each reconstructed scenario or simulation represents a distinct “universe” or a parallel-plausible interpretation. Instead of targeting the view of a “single” past, the multiverse vision embraces the idea of “multiple” past where the interpretation comes at the intersection of different views and hypotheses. This

approach offers an innovative way to embrace the uncertainties inherent in archaeological data while providing new opportunities for visualization, hypothesis testing, and public engagement.

In the study of ancient civilizations, the available evidence—whether material, stratigraphic, or paleoenvironmental—is often fragmentary and subject to interpretation. This incomplete nature of data makes it challenging to construct a singular, definitive model of ancient lifeways. The multiverse concept offers a solution by allowing for the coexistence of multiple plausible reconstructions, each grounded in a different interpretation of the data. Instead of forcing a single narrative, archaeologists can explore various “what-if” scenarios, effectively creating a multiverse of past environments and practices. The more we explore, the more we learn. Artificial intelligence is particularly adept at implementing this multiverse structure. Machine learning algorithms can evaluate extensive archaeological data, discern trends, and produce simulations based on varying input parameters [20]. For instance, in reconstructing an ancient settlement, AI could produce multiple models by varying assumptions about spatial organization, resource distribution, or climatic conditions. Each model becomes a parallel universe within the multiverse, representing a possible iteration of past human behavior and environmental interaction.

In the case of the reconstruction of Etruscan sacred spaces, for example, limited architectural remains and conflicting interpretations make it difficult to determine their original layouts and symbolic functions. By applying AI, archaeologists can create multiple reconstructions based on varying interpretations of fragmentary remains, landscape features, and historical contexts. These reconstructions might include divergent alignments, material compositions, or decorative elements. Each simulation represents a distinct “universe” within the multiverse, offering insights into the range of possibilities for how these spaces might have been used and experienced.

Similarly, AI can simulate the dynamics of ancient communities by modeling interactions between variables such as agricultural practices, resource availability, and social hierarchies. In studying the development of settlement patterns, for instance, researchers could generate parallel scenarios where different factors—such as population pressure, trade networks, or environmental change—play varying roles in shaping spatial organization. These simulations enable to test hypotheses and evaluate the relative plausibility of competing interpretations.

The concept of the multiverse aligns well with the probabilistic nature of archaeological research. Data from excavations and surveys are inherently uncertain, often comprising incomplete structures, ambiguous stratigraphy, or degraded organic remains. Traditional approaches to reconstruction risk oversimplifying this complexity by presenting a single, deterministic model.

AI, combined with the multiverse framework, allows archaeologists to embrace and even celebrate this uncertainty. By generating multiple plausible reconstructions, AI provides a platform for exploring the range of possibilities inherent in the data. This approach shifts the focus from seeking definitive answers to understanding the spectrum of potential realities that could have existed.

Beyond academic research, the multiverse approach to AI-driven simulations has significant potential for education and public engagement. By presenting multiple reconstructions of ancient environments, these simulations can foster a deeper appreciation for the complexity and richness of past human experiences. It is important to emphasize that the multiverse vision tends to eradicate the idea that the best interpretation should come mainly from recognized scholarships. A multiverse opens the gate to other multivocal interpretations.

Public engagement with the multiverse framework can also inspire a more inclusive understanding of the past. Traditional reconstructions often prioritize dominant narratives, marginalizing alternative perspectives and interpretations. By generating multiple reconstructions, AI allows for the inclusion of diverse viewpoints, ensuring that the archaeological record is represented as a complex, multifaceted tapestry rather than a monolithic story.

The integration of AI and the multiverse framework in archaeology is still in its early stages, but the potential for growth is immense. Advances in machine learning, data processing, and simulation technologies will enable increasingly detailed and accurate reconstructions of ancient environments and cultural practices.

One promising area of development is the use of generative AI models, such as neural networks, to create highly realistic visualizations of ancient landscapes. These models can combine data from multiple sources—including excavation records, paleoenvironmental reconstructions, and artifact analyses—to produce immersive, multidimensional simulations [21]. By iterating across different parameters, these models can generate a vast array of scenarios, enriching our understanding of the archaeological record.

Another exciting prospect is the application of AI to “counterfactual archaeology”, where researchers explore how different environmental or cultural variables might have altered the development of ancient societies. For example, what if certain Etruscan settlements had adapted differently to changing climatic conditions? AI-driven multiverse simulations could model alternate scenarios, offering insights into the resilience and adaptability of past human communities.

By leveraging the power of AI, we can move beyond static, singular reconstructions to a richer, more dynamic understanding of ancient environments and practice a multiverse of the past waiting to be explored.

AI Validation and Multiverses

The AI multiverse approach is going to challenge the traditional way archaeologists deal with data input, data analysis and representation. In other words, it will determine a multiplication of content and hypotheses thanks to multifactorial simulations. This entails great research opportunities and some risks as well.

In practice, an AI multiverse approach involves the integration of diverse datasets—ranging, for instance, from paleoenvironmental data to remote sensing and 3D models, into a unified environment that various models can analyze simultaneously. Because these datasets often vary in quality and completeness, specialized AI models may focus on particular evidence types, such as high-fidelity 3D artifact scans or textual inscriptions, while a meta-model synthesizes their outputs to highlight which lines of evidence are most relevant to each hypothesis. One of the transformative potentials lies in the capacity of machine learning algorithms to detect patterns across multiple analyses. These systems can draw unexpected connections—such as uncovering how the distribution of certain artifacts might correlate with distinctive ceramic traditions, thereby revealing possible trade links or cultural interactions. Beyond supporting human-led interpretations, AI can also propose novel hypotheses, inspiring archaeologists to explore questions they might not have identified through conventional research methods. When dozens or hundreds of potential scenarios are running in parallel, researchers, students, and even the public may need new platforms and techniques for making sense of these expanded possibilities. This also broadens public engagement by demonstrating the contingent nature of archaeological knowledge, although it raises ethical questions about which stories or reconstructions are given prominence and how best to communicate uncertainty.

The combination of big-data analytics and open access to evidence means hypotheses can evolve more rapidly, transforming them into living entities that shift and refine themselves as new information comes to light. Rather than working toward static, single-threaded narratives, the AI multiverse approach underscores the probabilistic nature of interpreting the past, challenging researchers to remain open to multiple, evolving models while striving for clarity and responsibility in how these are presented.

However, the implementation of AI in a multiverse framework for archaeology also raises significant ethical and epistemological concerns [22]. How is it possible to validate AI scenarios and hypotheses?

One primary challenge is the risk of bias, mistakes and misrepresentation—AI models are only as good as the data they are trained on [23], and historical and archaeological records (and their interpretations) might reflect the use of incorrect data, potentially marginalizing alternative narratives. Additionally, the proliferation of multiple plausible reconstructions necessitates a rigorous validation process to ensure that the scenarios generated are not speculative to the point of distortion. Ethical considerations must also address the ownership and authority over interpretations of the past—who decides which scenarios are presented, and how do we prevent AI-driven archaeology from being misused for political or ideological ends [24]? Finally, an epistemological framework must be developed to differentiate between plausible reconstructions and purely imaginative speculations, ensuring that AI-enhanced archaeology remains a tool for knowledge production rather than fiction. These risks highlight the need for transparency in AI methodologies, interdisciplinary collaboration, and continuous critical evaluation of AI-generated narratives.

AI-generated reconstructions rely on training data that may contain biases, gaps, or errors. If not carefully chosen, these models can produce unsubstantiated or misleading information that reinforces historical misconceptions and presents a skewed image of the past. When such reconstructions are offered as authoritative without thorough scholarly corroboration, the risks become even more pronounced. Another challenge arises from the tendency of generative AI models to produce “hallucinations” by generating convincing, yet entirely fabricated reconstructions based on flawed correlations in their datasets. Without a robust critical framework to verify outputs against actual archaeological data, there is a danger of wrongly legitimizing speculative or erroneous accounts of history.

Further concerns emerge when AI-produced reconstructions overshadow vital material evidence, such as stratigraphic studies, inscriptions, or artifacts, leading to conclusions that lack a firm empirical foundation.

In museum and digital heritage contexts, there is a risk that the public might regard AI-generated depictions as indisputable historical truths if not provided with clear disclaimers and sufficient context. This can undermine the inherently interpretive nature of archaeological work by suggesting a certainty that does not exist.

I believe that the key to tracking a process of validation in and by AI is the correct use of metadata. Metadata play a critical role in ensuring the reliability and accuracy of AI-driven archaeological reconstructions by embedding valuable contextual information that can be used to verify the final outputs. In digital reconstructions of ancient sites, the underlying AI processes often operate on large datasets pulled from varied sources, such as high-resolution imaging, geospatial scans, and historical documentation. Without robust metadata that detail precisely how these sources were generated, where they originated, and under what conditions they were captured, it becomes challenging to assess whether the AI models are accurately interpreting the input data or inadvertently distorting significant historical–archaeological elements.

When researchers collect data from archaeological digs, museums, or archival documents, they record details such as the exact location of an artifact, the orientation of

architectural remains, the materials used, and even environmental factors that might affect how structures have changed over time. These metadata are then fed into AI algorithms alongside the primary visuals or textual datasets. This allows scholars to trace the lineage of each data point, identify potential discrepancies, and cross-verify results with the historical or archaeological record. As a result, metadata effectively serve as a sort of “audit trail” for the AI reconstruction process, making it possible to pinpoint the origins of decisions the model makes when generating 3D visualizations or predictive simulations.

Equally important is how metadata can be employed to refine AI models over time. Each time an AI-assisted reconstruction is proposed, researchers can compare it against the metadata-enriched archive of previous findings and scholarly interpretations, noting any divergences or anomalies. Patterns that emerge, such as consistent misinterpretations in the rendered geometry of a building façade or systematic discrepancies in certain color or material classifications, can then be addressed by adjusting the training datasets or refining algorithms. This iterative loop of verifying AI outputs with metadata-driven contextual checks not only bolsters the credibility of the end result but also helps the algorithm learn from its mistakes and improve with each new project.

By mandating the inclusion and preservation of rich metadata from the onset, professionals in archaeology and heritage preservation can create a robust framework for scrutinizing and validating AI-generated reconstructions. In doing so, they ensure that models remain anchored in archaeological accuracy rather than drifting toward plausible but historically unfounded speculations. This systematic validation process, powered by metadata, upholds rigorous academic standards while opening the door to more advanced and nuanced reconstructions in the future.

A successful example of the archaeological validation process is the exhibition “AI Rethinks the Past” organized at Duke University in 2024 [25]. In this case, paleobotany data coming from specific archaeological excavations constitute a solid and empirically validated data-entry/taxonomy for the AI prompts. A paleobotany team later verified all the produced photos and videos in connection with several criteria including scientific consistency, geolocation, seasons and correctness. In this situation the study protocol is rigorous and the AI simulations of paleoenvironments provide a rather good depiction of flora, geomorphology and cultivations.

3. AI and Neuroaesthetics

Neuroaesthetics and AI represent a significant intersection between the study of human perception, cognitive processes, and the role of advanced technology in understanding and enhancing artistic experiences. This exploration bridges the disciplines of neuroscience, psychology, and artificial intelligence, aiming to uncover how art can influence human emotions and cognition and how AI can assist in decoding and even augmenting these effects.

The integration of neurometrics and advanced tools such as EEG and eye-tracking technologies has revolutionized the field of neuroaesthetics [26]. These tools allow researchers to measure brain activity, eye movement, and emotional responses in real time. For instance, using devices like head-mounted displays and eye-tracking recorders, scientists can analyze mental states during the observation of art. Neurometric indices, such as attention, emotional intensity, and cognitive workload, reveal profound insights into how individuals interact with visual stimuli. A case study on “The Sarcophagus of the Spouses” demonstrated significant differences in cognitive and emotional engagement between viewing the artifact in a museum versus a virtual reality (VR) environment. While attention remained stable in the VR setting, the physical museum experience elicited higher emotional engagement, particularly in the initial moments of observation [27]. This under-

scores the unique power of physical artifacts in evoking emotional responses, even as VR provides a controlled and immersive alternative.

Empathy [14] plays a central role in art perception, particularly in the context of faces and expressions depicted in sculptures and paintings. The fusiform face area's specialization in face recognition [28] underscores the importance of human representation in art. Sculptures like "The Sarcophagus of the Spouses" not only depict human features but also evoke deep emotional connections by mirroring real-life expressions of happiness, sadness, or tranquility. This empathetic engagement is further supported by neuroimaging studies that reveal activation in the limbic system when individuals view such artworks. The emotional resonance elicited by these pieces highlights art's ability to transcend time and culture, fostering universal connections through shared human experiences.

Advancements in AI have enabled researchers to delve even deeper into these neuroaesthetic phenomena. By combining machine learning algorithms with eye-tracking and EEG data, AI can analyze complex patterns and predict emotional and cognitive responses to art. This technology also offers the potential to create personalized artistic experiences, tailoring content to individual preferences and emotional states. In museum settings, AI-driven systems can guide visitors through curated pathways that align with their interests and cognitive profiles, enhancing both engagement and understanding. Moreover, AI-generated art itself raises intriguing questions about creativity and the nature of aesthetic appreciation. As machines produce works that evoke genuine emotional responses, the boundaries between human and artificial creativity become increasingly blurred.

The comparative analysis of VR and physical art experiences provides valuable insights into the strengths and limitations of each medium. While VR offers a controlled environment for studying cognitive and emotional processes, physical settings retain a unique ability to elicit strong emotional connections. The initial moments of museum visits often evoke heightened attention and emotional intensity, reflecting the sensory richness and authenticity of the experience. In contrast, VR excels in accessibility and replicability, making it an invaluable tool for education and outreach. By understanding these differences, researchers and practitioners can leverage the strengths of both mediums to create complementary experiences that cater to diverse audiences.

Cognitive processes associated with art perception involve specific brain regions, including the parietal, frontal, and temporal lobes. The parietal lobe plays a critical role in spatial awareness and the analysis of somatosensory stimuli, which is essential for understanding the physical context of an artwork [29]. The frontal lobe contributes to decision-making and planning, particularly when interpreting complex artistic compositions. The temporal lobe, especially the fusiform face area, is crucial for recognizing and responding to human faces depicted in art [30]. This cortical region is uniquely sensitive to faces, triggering emotional responses mediated by the limbic system and amygdala. These responses can mirror the emotions elicited by real human expressions, highlighting the profound empathetic connections between viewers and artistic representations of humanity.

The application of neurometrics extends beyond the laboratory to practical settings like museums and cultural heritage sites. By integrating eye-tracking data with cognitive and emotional indices [31], researchers can observe mental states in real time, correlating these states with specific visual stimuli. This capability allows for a deeper understanding of how individuals engage with art and provides a basis for designing experiences to enhance emotional and cognitive impact. For example, heat maps generated from eye-tracking studies reveal gender-based differences in viewing patterns and dwelling times. Similarly, background skills and expertise influence how individuals interact with art, offering insights into the role of cultural and educational contexts in shaping aesthetic experiences.

The future of neuroaesthetics and AI lies in fostering interdisciplinary collaboration and expanding the scope of research. By bringing together neuroscientists, artists, technologists, and psychologists, the field can develop holistic approaches to understanding and enhancing aesthetic experiences. This collaboration extends to practical applications in education, cultural preservation, and public engagement. For example, interactive exhibits that combine AI, neuroimaging, and traditional art forms can provide visitors with immersive and educational experiences, deepening their appreciation for both art and science.

In conclusion, the intersection of neuroaesthetics and AI offers a rich and dynamic field of exploration, bridging the gaps between art, science, and technology. By studying how humans perceive and respond to art, and by leveraging AI to analyze and augment these experiences, researchers can unlock new dimensions of understanding and creativity. This interdisciplinary approach not only advances scientific knowledge but also enriches our cultural and emotional lives, affirming the enduring power of art in the human experience.

4. AI, Eye-Tracking and Visual Thinking Strategies

During the Fall semester of 2023, in my undergraduate class “Why Art” at Duke University, I had the chance to set up an experiment involving eye-tracking, visual thinking strategies and AI with the scope to understand the mechanics of visual observation of museum’s paintings [32]. The subject of this research test was *My Mother, She Fell From the Sky*, by Liên Trương, 2021, an oil, silk, acrylic, copper pigment, and enamel on canvas. It was from the collection of the Nasher Museum of Art at Duke University; 72 × 96 inches (182.88 × 243.84 cm, Figure 2). The first part of the experiment consisted of measuring eye-tracking from a distance of 3 m from the artifact and from the same position. The eye-tracking device was Pupil Invisible [33], a lightweight, wearable eye-tracking device developed by Pupil Labs. It has inward-facing infrared cameras for eye tracking and an outward-facing scene camera (1080p, 30 fps) for recording the user’s field of view.



Figure 2. Liên Trương, *My Mother, She Fell From the Sky*, 2021. Oil, silk, acrylic, copper pigment, and enamel on canvas. Collection of the Nasher Museum of Art at Duke University.

The class of 18 students (18 and 19 years old, 10 females and 8 males) was split into two groups: the first group was asked to spend 2 min to describe the visual narrative of the artifact by specific questions (by Visual Thinking Strategies Protocol, VTS) [34] and then to start the eye-tracking experiment, while the second group started the eye-tracking experiment immediately (without taking any additional time in the observation). They viewed the artwork for 30 s without any prompts (therefore, no VTS) while their eye movements were recorded. Data collection involved fixation duration, scan path, heatmaps of gaze concentration and areas of interest (AOIs) naturally attracting attention.

The questions for the visual narrative were based on the visual thinking strategy (VTS) protocol:

- *What's going on in this image/artwork/object?*
- *What do you see that makes you say that?*
- *What more can we find?*

The goal of the experiment was to evaluate initial and post-VTS eye-tracking data, shifts in interpretative statements indicating deeper engagement or changed focus and eye-tracking metrics. Heat maps compare baseline and post-VTS viewing, analyzing AOIs to track changes in gaze distribution and fixation duration. Qualitative analysis involves thematic coding of verbal responses, assessing shifts from descriptive to analytical engagement.

Visual thinking strategies (VTSs) [35] provide a foundational approach to understanding how individuals perceive and interpret art. VTS emphasizes observation without requiring prior knowledge of the artifact, making it an accessible and inclusive method. By comparing groups exposed to this technique with those who are not, researchers can isolate its impact on perception and interpretation.

The heat maps generated by each group of observers show different results of the cumulative eye-tracking. In fact, the two heat maps illustrate how eye-tracking patterns change when students engage with an artwork with and without the guidance of visual thinking strategies (VTSs).

Heat Map 1 (Figure 3): Without Visual Thinking. The first heat map shows that the participants' gaze is more diffuse, with attention spread across the entire canvas. There are no strong, concentrated areas of fixation. This suggests that students are observing the painting in a more casual or unstructured manner, without honing in on specific features. The scattered heat zones imply that viewers are exploring the painting without a clear framework or prompts to guide their interpretation.



Figure 3. Heat maps of the eye-tracking experiment concerning the first group of students observing the picture without VTS. The color palette (red, highest, light blue lowest) show the intensity of visual focal areas.

Heat Map 2 (Figure 4): After Visual Thinking Concentrated Gaze. The second heat map has two distinct focal points where the gaze is heavily concentrated (red areas). This suggests that participants are now focusing on specific features of the painting. These focused areas likely correspond to features or elements of the painting that were highlighted or emphasized during the VTS process. The more structured pattern implies that participants are engaging with the painting more deeply, possibly influenced by questions or discussions that directed their attention to specific areas.



Figure 4. Heat maps of the eye-tracking experiment concerning the group of students observing the picture after VTS. The color palette (red, highest, light blue lowest) show the intensity of visual focal areas.

It is interesting to observe that in the first NVTS image, the percentage of red heat maps is 44.13%, while in the second one is 77.86%. This means also a more extensive visual focus after VTS.

Figure 5 shows the visual comparison between the cumulative heat maps of the non-VTS group (left) and the VTS group. The lack of the original background of the image helps to better understand that the gaze after VTS becomes much more focused and extended in the main regions of interest.

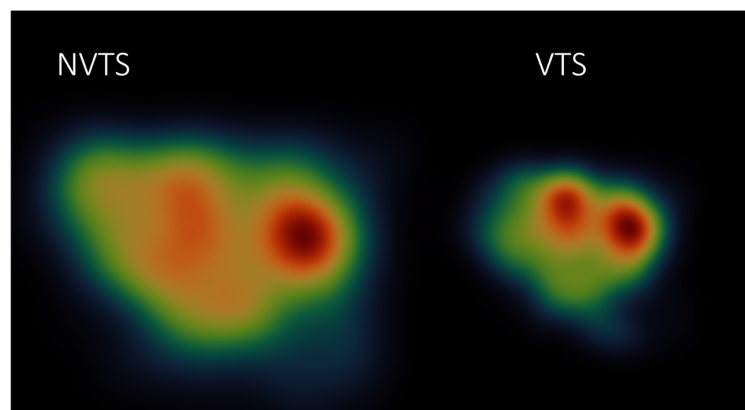


Figure 5. Heat maps comparison: on the (left) eye-tracking without VTS, on the (right) with VTS (no background). The color palette (red, highest, light blue lowest) show the intensity of visual focal areas.

The guided VTS process significantly changed how participants engaged with the artwork. Without VTS, observation is random and exploratory; after VTS, observation is targeted and deliberate. The concentrated gaze points in the second heat map may indicate participants are not only looking but interpreting and analyzing specific elements of the painting. The shift from diffuse to focused gaze patterns demonstrates how VTS can foster a more thoughtful and analytical engagement with visual material. This comparison highlights how structured approaches like VTS can transform casual observation into a more intentional and meaningful interaction with art.

Figure 6 bar chart compares the percentage of relevant areas in the two heat maps. It visually highlights the difference in the proportion of the most focused regions based on eye-tracking data. The second heat map shows a significantly larger coverage of relevant

features compared to the first one, demonstrating the increased attention and focus on specific areas.

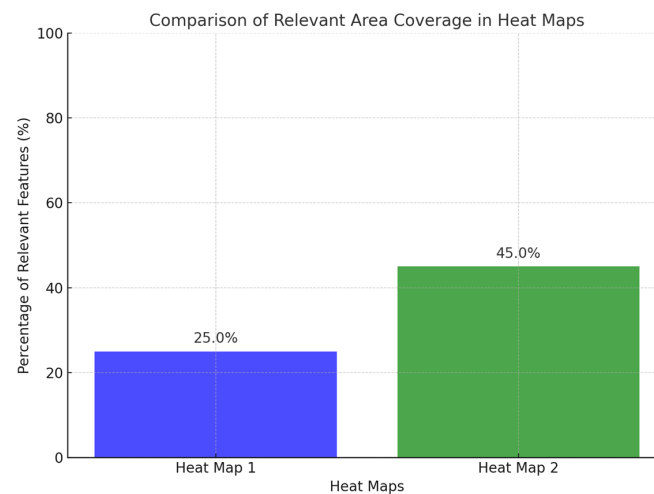


Figure 6. Comparison of the relevant features in the first and second heat maps.

All the students involved in the VTS experiment wrote a specific narrative concerning their own interpretation of the original image (Figures 7 and 8). The narrative of each student was used as a prompt for generating AI images with the same style as the original one. For this experiment, we use FolloFox AI Distillery [36], an open-source text-to-image generator. Figures 7 and 8 show a collection of AI images generated by the VTS prompts of 16 students: each gallery shows slightly different subjects. Each student's textual prompt was then processed through text-to-image generation, where AI converted descriptions into visual outputs while maintaining stylistic coherence. To achieve that, we used two techniques: one (ControlNet) that makes the AI consider the original art piece as a canvas for the imagination of the students, and another (IPAdapter) that essentially allows the AI to consider the art piece as part of the imagination itself, as if the memory of the art piece was influencing the generative process. In this way, and leveraging on its intrinsic artistic capabilities, the AI model was able to apply learned artistic elements from the original painting onto the newly generated images, ensuring consistency in brushwork and lighting effects.

The last research question of the experiment was to evaluate and compare the consistency of the textual prompt of the VTS with the original image and the AI-generated image. In this case, it was chosen the case of a student whose VTS narrative described the original painting in this way: "A ritual is occurring where a person is being sacrificed in the center of a fire. There seems to be an orange and yellow patch at the center with human hand figures on it. There are ghostly figures that may represent souls, floating above the center pit. These seem to be spirits who are making contact with the people on the ground below. There are also red shadows around the people on the ground which could represent violence". Therefore, this statement became the main prompt of the AI text-to-image processing. In this case, Distillery was trained with the colors and style of the original image in order to avoid discrepancies in the general design of the scene.

The last step of the process was to compare the AI-generated image, based on one of the students' prompts with the original painting in order to evaluate the consistency of the prompt and the fidelity of the AI interpretation. This analysis helps also to understand the relationship between art and AI art/machine learning and, particularly, to visualize how different observers can interpret a cultural subject in several ways. The interaction between the artist and its creation and the feedback of the public generate always new content and different symbolic meanings. This part of AI analysis involved ChatGBT 4/o in combination with Python 3.13 (<https://www.python.org/>, accessed on 15 January 2025).



Figure 7. AI-generated images (Distillery) concerning the VTS prompt of a group of eight students of the class Why Art, Duke University.



Figure 8. AI-generated images (Distillery) concerning the VTS prompt of a second group of eight students of the class Why Art, Duke University.

The first processing was the edge detection/structural comparison (Figure 9) of the original picture with the AI-generated one. The edges in the AI-generated image are clearer and more defined, with a focus on distinct objects such as the firepit, human figures, and the surrounding environment. The composition is more structured and less layered, with identifiable shapes and minimal abstraction. The AI image prioritizes clarity and narrative focus, reflecting the key elements described in the prompt (e.g., the central firepit and human participants). The reduced edge complexity highlights the AI's tendency to simplify and emphasize specific aspects, at the expense of broader symbolic representation.

The second analysis of the heatmaps (Figure 10) provides a visualization of three metrics (SSIM, MSE, and similarity) comparing the original image and the AI-generated image. The structural similarity index (SSIM) value indicates a low-to-moderate level of structural similarity between the two images. The edges and overall structural components in the AI-generated image align only partially with those in the original, reflecting a focus on certain features (e.g., the firepit) rather than the broader abstract composition. The mean squared error (MSE) score, relatively low, suggests that the pixel-level differences between the original and AI-generated images are not extreme. However, the AI-generated image simplifies and emphasizes certain areas (e.g., human figures, firepit), leading to fewer nuanced variations found in the original painting.

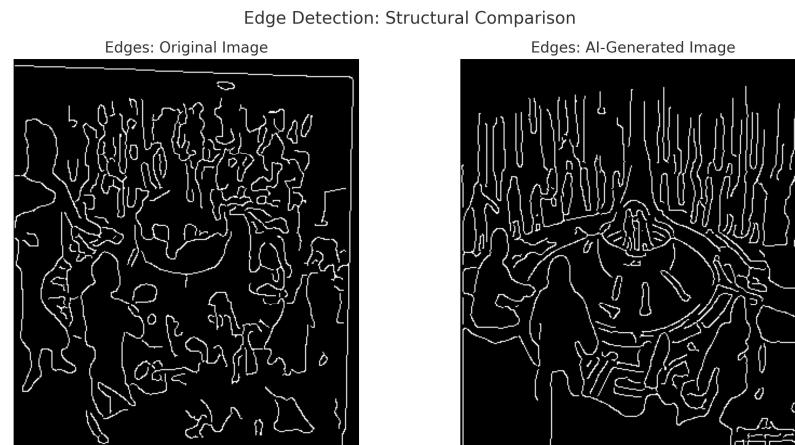


Figure 9. Edge detection of the main scene of the painting in the original picture (left) and in the AI-generated image (right). ChatGBT 4/o in combination with Python 3.8.x.

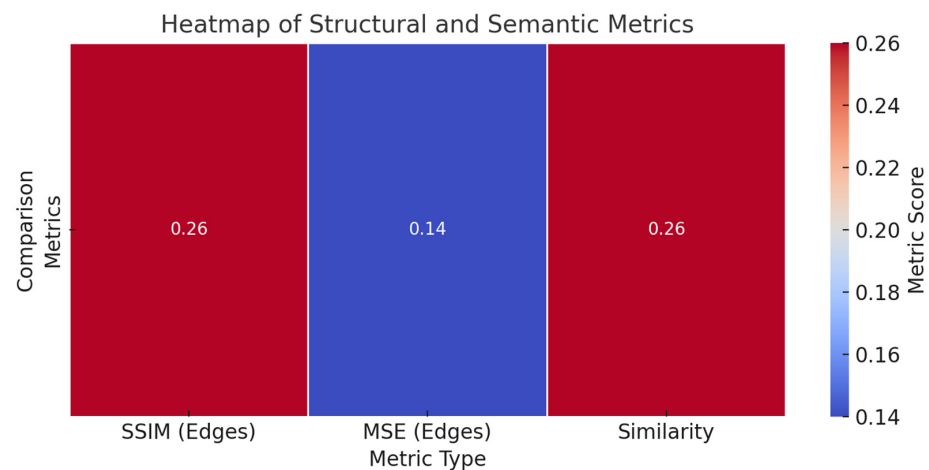


Figure 10. Comparison metrics concerning the original image and the AI-generated one based on SSIM, MSE (Pixel-level intensity) and similarity (conceptual alignment using semantic analysis). ChatGBT 4/o in combination with Python.

Finally, the semantic similarity reflects how well the AI-generated image captures the conceptual meaning or narrative of the original. The score suggests that while the AI image aligns with the general theme (e.g., ritual, firepit, human figures), it diverges in the subtleties and complexity of the abstract symbolism. The AI-generated image captures the narrative core (firepit and figures. figure) but simplifies the broader context, as reflected by low-to-moderate SSIM and semantic similarity scores. The MSE indicates that pixel-level differences are subtle, highlighting the AI's ability to retain some visual harmony. The AI seems to prioritize clarity and focus on specific narrative elements, sacrificing the nuanced abstraction present in the original.

This bar chart illustrates the semantic similarity scores between the original image and the AI-generated image when compared with the provided narrative prompt. The relatively low score reflects the abstract and symbolic nature of the original image. While the original image may align conceptually with the narrative prompt, its abstract elements and lack of clearly defined structures reduce its semantic alignment with explicit descriptions. This suggests that the original image's meaning is open to interpretation and not tightly bound to a single narrative.

The AI-generated image (Figure 11) has a slightly higher semantic similarity score compared to the original image. This increase in similarity is due to the AI's clear depiction of key elements from the prompt, such as a central firepit. Human figures surrounding the

ritual. A structured scene that matches the described ritualistic activity. The AI-generated image prioritizes a direct narrative alignment with the prompt over artistic abstraction.



Figure 11. AI-generated image (Distillery) based on a student's prompt during the VTS experiment. ChatGBT 4/o in combination with Python.

The AI-generated image focuses on literal and recognizable elements, leading to a higher semantic similarity with the narrative prompt. The original image, by contrast, emphasizes abstraction and symbolic representation, resulting in a lower semantic score. The original image's abstraction allows for multiple interpretations but reduces alignment with a specific narrative. The AI image sacrifices abstraction for clarity, ensuring stronger alignment with the explicit elements of the prompt. The score difference (0.0632) is small, indicating that both images reflect the prompt to some extent but in vastly different ways—one abstract, the other literal.

This analysis highlights how AI-generated art tends to prioritize narrative clarity and prompt-specific alignment over abstraction and symbolism. The original image's abstraction may make it more engaging for interpretation, but it aligns less closely with the specific semantic content of the prompt.

Figure 12 compares the semantic alignment of the original and AI-generated images with the narrative prompt: Original Image: Semantic similarity score is 0.2698, reflecting its abstract and ambiguous representation. AI-Generated Image: The semantic similarity score is 0.3330, showing a stronger alignment with the prompt's described scene.

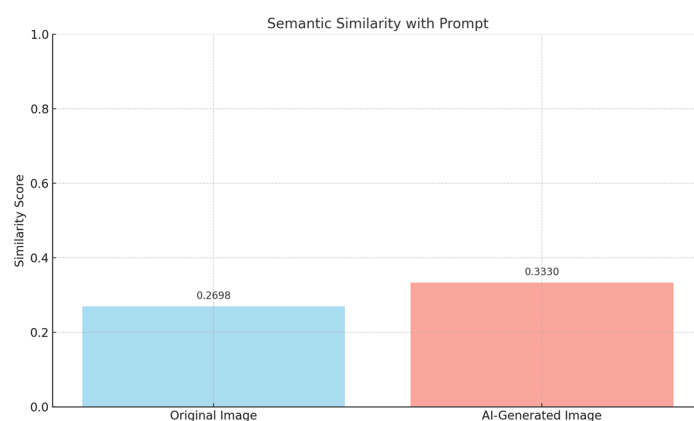


Figure 12. Semantic similarity between the original image and the AI-generated image. ChatGBT 4/o in combination with Python.

Global Variance, Figure 13. The AI-generated image has a higher global variance (3622.66) compared to the original image (2904.91), indicating greater overall contrast in pixel intensities. The variance for the center region of the original painting (3102.74) is similar to its global variance, showing consistent intensity distribution. AI Firepit: The variance for the AI firepit region is significantly lower (870.60), suggesting that the firepit has a more uniform and focused intensity distribution compared to the abstract center in the original painting.

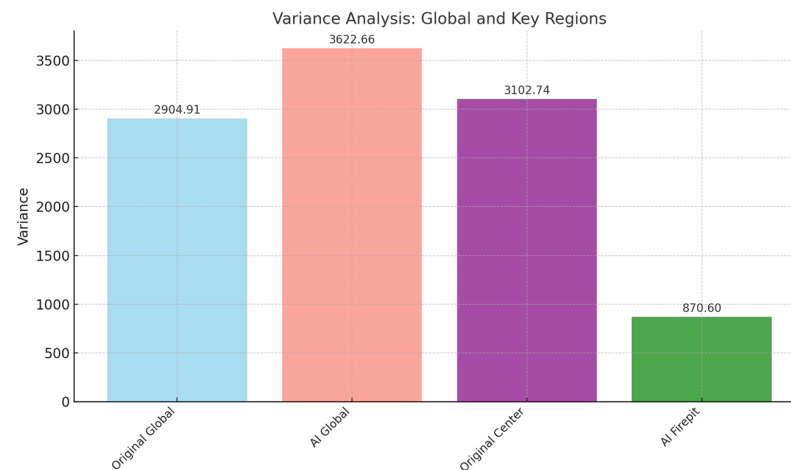


Figure 13. Variance analysis, global and key regions. ChatGBT 4/o in combination with Python.

In the semantic and symbolic comparison (Figure 14), the AI interpretation highlights a more structured and socially grounded scene, emphasizing communal participation around a central fire. The arrangement of figures suggests hierarchy and intentional roles within the ritual, contrasting the ambiguous roles in the original image. The central fire is the focal point, symbolizing transformation, energy, or purification. This element reinforces the theme of ritualistic change, perhaps linked to rebirth or spiritual transcendence.

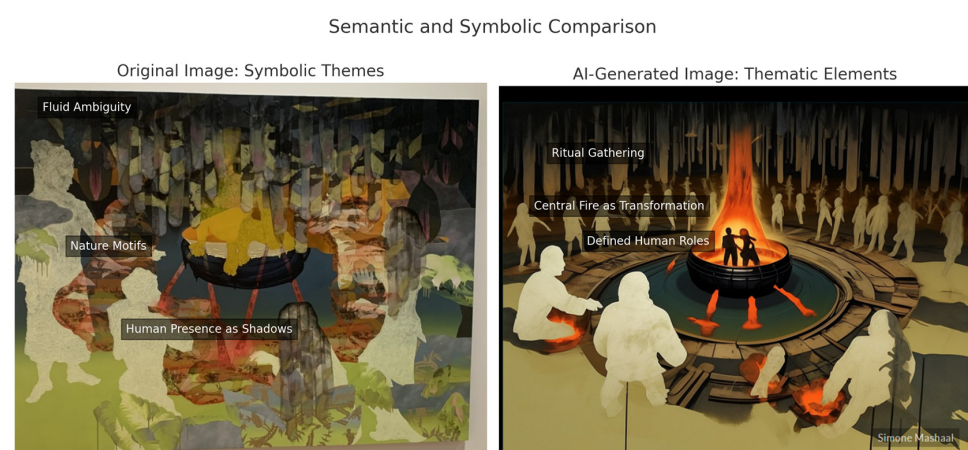


Figure 14. Semantic and symbolic comparison between the original painting and the AI-prompt-based generated image. ChatGBT 4/o in combination with Python.

Human figures are distinct and actively engaged, portraying clear roles within the ritual (e.g., offering, watching, or conducting the ceremony). This structured representation contrasts with the shadows in the original image, where individual roles are obscured.

The original image uses abstraction to evoke mystery and fluidity, suitable for an interpretive, symbolic exploration of ritual. In contrast, the AI-generated image provides a concrete, structured depiction, making the ritual more accessible and straightforward.

ward. It emphasizes a symbiotic relationship with nature, integrating its motifs as part of the ritual. The AI-generated image shifts the focus to human agency and the fire as a transformative medium.

The shadows in the original image convey universality and inclusiveness, aligning with rituals that transcend individual identity. The AI-generated image emphasizes distinct roles and interactions, reflecting a more hierarchical or role-based ritual structure.

5. Saliency Maps

One of the key factors in the human interpretation of images [37] is information encoding, which means how much we need to capture for interpreting and transmitting visual content. What kinds of cultural and contextual elements can influence this process and what methods we can adopt for this kind of research?

Saliency maps are an essential tool in both neuroscience and artificial intelligence (AI), offering a means to visualize and quantify the most important features of an image or scene that capture human attention or drive AI model decisions. In the context of art, saliency maps provide fascinating insights into how people perceive and interact with visual works, revealing the underlying patterns and elements that guide focus and evoke emotional or cognitive responses.

Saliency maps are computational or visual representations [38] that highlight the regions of an image that are most likely to draw attention. They are rooted in the concept of saliency—the quality that makes certain aspects of a visual scene stand out. Saliency is influenced by low-level features, such as color, contrast, and brightness, and high-level cognitive factors, including context, meaning, and prior knowledge. In computational terms, saliency maps are often generated using algorithms that model human visual attention. These algorithms analyze an image to identify regions of interest based on feature contrasts. For example, bright, saturated colors in an otherwise muted scene, or sharp edges in a soft, blurry context, are likely to appear as salient.

In neuroscience, saliency maps are connected to studies of the visual cortex and attention mechanisms in the brain [39]. Eye-tracking studies frequently use saliency maps to correlate gaze patterns with specific visual stimuli, providing data on how people process visual information in real time.

Art has always been a powerful medium for engaging visual attention, often deliberately playing with elements like color, composition, and texture to guide the viewer's gaze. Saliency maps offer a quantitative approach to studying these effects, revealing how artists manipulate visual elements to create focal points or evoke specific emotional responses. One of the most direct applications of saliency maps in art is through eye-tracking studies. These studies track the movement of viewers' eyes as they explore a painting or sculpture, generating data that can be transformed into saliency maps. The resulting heatmaps highlight areas where viewers spend the most time looking, as well as the sequence of their gaze.

Eye-tracking studies have also shown that saliency in art is not solely driven by low-level visual features. High-level factors, such as cultural context, personal experience, and the narrative embedded in the work, significantly influence gaze patterns.

In applying a saliency map analysis to the above-mentioned case study, we can better evaluate the difference between an empirical experiment, the eye-tracking, and the AI simulation of a saliency map over the same subject (Figure 15).

The saliency map (on the right) indicates regions of the image that are most visually or semantically prominent. The central area of the original image, which likely contains significant features (such as the bright yellow-orange area), is highlighted strongly in red. This suggests that this region is the most attention-grabbing. The outer regions of the image, such as the edges and corners, show low saliency (yellow or white areas). This indicates

that these areas contain less visually significant information and are less likely to attract attention. Bright colors, such as yellows and oranges, and high contrast in the original image have a strong influence on the saliency. The heatmap reflects this by assigning these areas higher attention weights. The saliency map also captures contextual elements, such as the figures on the sides of the original image, albeit with lower intensity. This implies that while these figures are part of the composition, they do not dominate attention in the way the central elements do.



Figure 15. Saliency map (right) of the original painting (on the left). ChatGBT 4/o in combination with Python 3.8.x. The red colors show the main regions of visual interest of the painting according to an AI simulation.

Saliency maps generated by AI models offer a unique perspective on how machines “see” art [40]. These models, often trained on large datasets, use neural networks to predict the most salient regions of an image. Comparing AI-generated saliency maps with human gaze patterns provides insights into both human and machine perception [41].

While human saliency is influenced by emotional, cultural, and contextual factors, AI saliency is typically based on algorithmic rules and data patterns. This distinction can lead to interesting divergences in interpretation. For example, an AI model might focus on fine details or high-contrast areas that humans might overlook in favor of more emotionally or contextually relevant regions. These differences highlight the challenges and opportunities in teaching AI systems to better understand human aesthetics.

In Figure 16, similar to the previous saliency map, the strongest highlights (in red) are concentrated in the central area, specifically around the fire and the human figures in the center of the composition. The flame and the figures directly below it dominate the visual field, indicating their critical role in capturing attention. The saliency map also picks up on some peripheral figures, such as those on the left and right of the central scene. However, their importance is relatively diminished compared to the central flame and figures.

The saliency map reflects vertical attention, with the fire extending upwards and maintaining a visually significant streak in the middle. The bright vertical saliency suggests that viewers’ attention might naturally follow the fire’s upward trajectory. The contrast between the fire’s bright orange glow and the dark background plays a key role in determining saliency. The map effectively captures this contrast, emphasizing regions where there is a stark difference in brightness. This saliency map aligns well with the likely narrative intention of the artwork; focusing on the central ritualistic scene while maintaining peripheral awareness of the surrounding figures. The central focus supports the visual hierarchy, leading viewers to the most crucial elements of the composition first.



Figure 16. Comparison between the AI image generated (on the left, created by a student's prompt) and its saliency map. ChatGBT 4/o in combination with Python 3.8.x. The red colors show the main regions of visual interest of the painting according to an AI simulation.

The balance of saliency here is better distributed compared to the previous example, as secondary attention is given to surrounding elements, making the overall scene more dynamic. The visual weight reinforces a theatrical composition, emphasizing the ritual and its participants. Overlaying this saliency map onto the original image could provide further clarity on how well the artwork communicates its intended narrative. Additional analyses could examine whether the surrounding figures could draw slightly more attention to enhance balance.

The first saliency map reflects a complex and distributed visual structure, where attention is spread across multiple elements. This indicates that the original image invites exploration and encourages the viewer to construct meaning through interaction with different parts of the composition. The second saliency map reveals a hierarchical and centralized focus, emphasizing the ritual as the core of the narrative. This suggests a simpler visual interaction that guides the viewer to a specific interpretation.

In the chart of Figure 17, both maps display nearly identical distributions of saliency between the central and peripheral regions. This might suggest a similar spatial distribution of salient features in the two saliency maps. The high peripheral saliency could mean that the visual elements or features in the images being analyzed are distributed toward the edges rather than concentrated in the center. If Map 1 (Figure 17, 20.802298% central, 79.197702 peripheral) represents the original artwork and Map 2 (Figure 17, 19.358028%, central 80.641972%, peripheral) represents an AI-generated interpretation, this result implies that the AI-generated image mimics the spatial saliency distribution of the original. This is an additional confirmation that the student's prompt was deeply influenced by the observation during the VTS experiment. On top of this, the two images recall the main spatial structure of the scene.

In visual design or art, saliency in the peripheral regions can engage viewers by encouraging exploration beyond the center. If the purpose of the AI-generated image was to replicate the visual structure of the original, this similarity in saliency distribution might indicate that it successfully captured the spatial dynamics of the original.

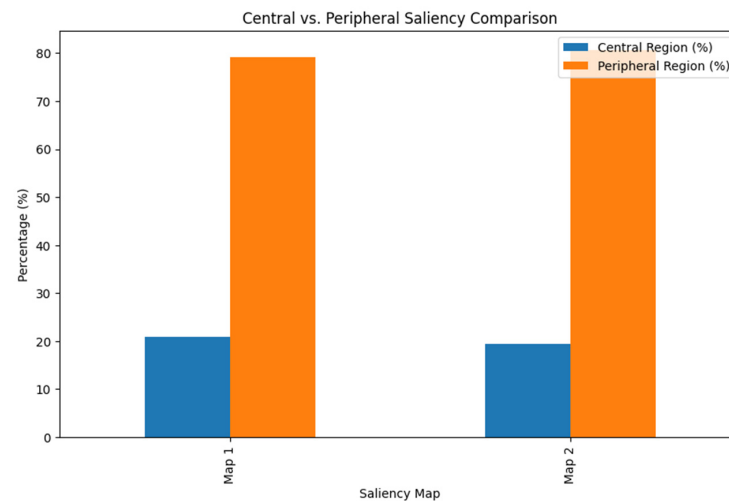


Figure 17. Comparison between central and peripheral saliency of the original and AI-generated images. ChatGBT 4/o in combination with Python.

Saliency Maps and AI

As discussed before, saliency denotes the capacity of some visual features to inherently capture attention owing to their contrast, prominence, or significance within a particular context. In art [42] and artifacts, saliency functions on two levels. It illustrates the physical attributes of an artwork, including edges, color contrasts, and symbolic areas that automatically capture the viewer's attention [43]. Conversely, computationally created saliency maps are visual depictions produced by algorithms that replicate human attention, providing a method to examine viewer interaction with visual compositions. Saliency maps offer a means to evaluate human perception and interaction with art. These maps delineate areas of significant visual prominence within an image, with luminous zones signifying components that capture attention owing to pronounced edges, contrasts, or geometric configurations. In an architectural picture, elements such as obelisks and domes are prominent due to their structural uniqueness and symmetry. In contrast, darker areas on a saliency map indicate regions of diminished visual importance, such as untextured backgrounds or uniform hues. This approach corresponds with human inclinations to concentrate on structured patterns, symmetry, and details that have aesthetic or cultural importance. In artworks like frescoes, saliency maps often highlight the outlines of human figures, complex patterns, or components that direct attention and enhance the narrative or emotional resonance of the piece.

By incorporating saliency analysis into these systems [38], artists can guide AI models to produce works that align with human aesthetic preferences [44]. For example, an AI model might be programmed to generate abstract art that maximizes saliency in certain regions, creating compositions that naturally draw the viewer's eye [45]. Alternatively, artists can use saliency maps to evaluate and curate AI-generated pieces [46], selecting those that achieve the desired balance of attention and impact.

The interplay between saliency maps and art raises important questions about aesthetics, creativity, and the nature of human attention [47]. As AI systems become more sophisticated, they not only analyze but also influence the way we perceive and create art. This dynamic has both exciting possibilities and potential challenges. Saliency maps can be used to enhance viewer engagement with art in museums and galleries [48]. For example, interactive exhibits could display real-time saliency maps based on visitors' gaze patterns, offering insights into how different people perceive the same work. This could spark discussions about the subjective nature of art and the diverse factors that influence visual attention.

Moreover, saliency maps could guide the curation of exhibitions, helping curators design layouts that optimize viewer engagement. By analyzing gaze patterns and attention flows, curators can position artworks in ways that encourage exploration and discovery.

The use of saliency maps in art also challenges traditional notions of aesthetics and artistic intention. If saliency maps reveal that viewers consistently focus on unintended elements of a work, does this undermine the artist's original vision? Or does it highlight the dynamic and participatory nature of art, where meaning is co-created by the artist and the viewer? In the context of AI-generated art, saliency maps further blur the boundaries between creator and audience. When an AI system generates a piece based on saliency principles, is the resulting work a product of the machine's "vision", the programmer's intent, or the viewer's response? These questions challenge us to rethink the relationship between technology, creativity, and human experience.

In the art world, the use of saliency maps to optimize viewer engagement could lead to a homogenization of aesthetic experiences, where works are designed to appeal to predictable patterns of attention rather than fostering genuine creativity and diversity. Balancing the benefits of saliency analysis with the need to preserve artistic integrity and authenticity is an ongoing challenge.

Human figures, for instance, often dominate saliency maps due to their distinct edges, contrasts in clothing, and expressive gestures. Simultaneously, backgrounds with minimal texture or contrast are de-emphasized, showcasing the ability of saliency analysis to suppress visually inactive areas. In works where details like drapery, musical instruments, or symbolic gestures are prevalent, saliency is often distributed evenly across these intricate regions, reflecting their shared importance in guiding the viewer's gaze. This analysis aligns with cognitive research showing that human attention gravitates toward areas with significant transitions, contrasts, or culturally resonant symbols. In terms of practical applications, saliency maps serve as tools for evaluating and interpreting visual focus in both art and artifacts. They provide valuable insights into artistic techniques, revealing how creators manipulate visual elements to guide attention or evoke specific responses. In neuromarketing, saliency maps are used to assess how people engage with visual stimuli, offering metrics that can inform design strategies for advertisements, branding, and user interfaces. They are also increasingly applied in the creation and evaluation of algorithmic or generative art. By identifying regions that align with human aesthetic preferences, saliency maps help refine and curate AI-generated works to enhance their visual impact. Furthermore, when paired with eye-tracking data, these maps validate and complement computational predictions, allowing researchers to compare algorithmic models with real-world gaze patterns. The examples presented illustrate how saliency maps are applied to analyze both architectural and artistic compositions.

Saliency maps also provoke deeper questions about perception, interpretation, and artistic intention. They highlight a dynamic interaction between the viewer and the artwork, revealing how certain elements capture attention while others recede into the background. This raises intriguing questions about whether saliency maps reflect the artist's intended focal points or simply the natural tendencies of human perception. The comparison between human and algorithmic attention further enriches this dialogue, revealing areas of overlap and divergence that highlight the challenges of teaching AI to fully understand human aesthetics.

In conclusion, saliency maps offer a powerful lens through which to explore the interplay between art and attention. Whether used to analyze existing works, guide new creations, or enhance viewer engagement, they provide valuable insights into the mechanisms of visual perception and the dynamics of aesthetic experience.

6. Conclusions

This paper demonstrated the transformative potential of integrating generative AI, neuroaesthetic tools, and methodologies like Visual Thinking Strategies (VTSs) in archaeological and art historical research. By leveraging cutting-edge technologies such as AI-generated simulations, eye-tracking experiments, and saliency map analyses, it was possible to explore how these approaches contribute to understanding human interaction with cultural heritage [49] and reconstructing multiple narratives of the past.

Generative AI proved to be an invaluable tool for simulating alternative interpretations of archaeological contexts and art, enabling the conceptualization of the past as a multiverse rather than a fixed timeline. The AI-generated visualizations highlighted multiple plausible scenarios based on human input and descriptive prompts. These simulations captured the complexity of cultural models by offering nuanced visual interpretations that are rooted in data-driven creativity. This “multiverse approach” opens pathways for engaging with uncertainties and contradictions inherent in material cultures, fostering a more inclusive and diverse understanding of ancient societies.

Eye-tracking experiments provided critical insights into human cognitive engagement with visual stimuli, including original artworks and AI-generated interpretations. The results revealed that VTS significantly influences how individuals visually explore art, guiding attention toward specific features and fostering a deeper understanding of symbolic and structural elements. Before applying VTS, viewers exhibited diffuse patterns of attention, while post-VTS results demonstrated more focused and rich visual engagement. This underscores the pedagogical value of VTS in enhancing observational and interpretive skills, as well as the importance of guided frameworks in art and archaeological education.

The experiment involving AI, VTS, and eye-tracking provided key insights into how structured observation affects human perception and interpretation of artworks. The study demonstrated that pre-VTS eye-tracking data exhibited scattered gaze patterns with limited focal engagement. However, after employing VTS, participants’ gaze became more concentrated on specific features, emphasizing deeper interpretive engagement and cognitive mapping of the artwork. This result supports the effectiveness of structured visual strategies in enhancing interpretative depth and viewer engagement.

Furthermore, the AI-generated visualizations based on VTS-driven textual prompts revealed the potential of AI in transforming verbal descriptions into dynamic visual outputs. The comparison between human perception, as mapped through eye-tracking and saliency maps, and AI-generated outputs, based on textual prompts, demonstrated an intriguing alignment in focal points. AI models consistently highlighted key compositional elements that corresponded to the most salient areas of human gaze fixation. This suggests that AI can effectively translate textual narratives into meaningful visual representations, providing an innovative approach to exploring visual storytelling in art. However, discrepancies between AI-generated and original artworks also highlight the limitations of current AI models in capturing abstract and emotionally resonant artistic elements.

Saliency map analysis complemented the eye-tracking results by quantifying the visual impact of central and peripheral regions in both original and AI-generated imagery. The comparison between central and peripheral saliency distribution highlighted differences in how human-created and AI-generated images emphasize key visual elements. AI interpretations often concentrated saliency in symbolically charged areas, such as the central fire or ritual objects, while the original images showed more diffuse saliency, reflecting the complexity of human artistic intention. This analysis illustrates how saliency maps can bridge the gap between computational and humanistic approaches to visual studies, providing actionable insights into visual storytelling and composition.

In summary, the integration of these tools raises important implications for archaeology, art history, and neuroaesthetics:

- **Enhanced Multimodality:** Combining AI simulations with human-centric tools like eye-tracking and VTS creates a multimodal framework for exploring cultural artifacts, blending quantitative precision with qualitative interpretation.
- **Cognitive Engagement:** Eye-tracking and saliency maps underscore the importance of understanding human perception and attention when observing and interpreting artifacts, both to academic audiences and the public.
- **AI as a Visual Translator:** The ability of AI to generate imagery based on text descriptions opens new pathways for digital reconstruction, enabling the visualization of lost or imagined historical narratives.
- **Ethical Responsibility:** While generative AI offers creative and interpretive potential, its use necessitates transparency in methodologies and careful consideration of biases introduced by prompts, datasets, and algorithms.

Additionally, the role of AI in shaping museum experiences and cultural heritage education cannot be overstated. The intersection of AI and neuroaesthetics offers new perspectives on how people engage with and interpret visual information. As demonstrated in this study, AI-driven generative models can produce insightful reconstructions, yet these need to be critically analyzed through interdisciplinary methods. The ability of AI to generate alternative reconstructions of the past necessitates careful validation processes to ensure that these outputs are not speculative distortions but meaningful contributions to our understanding of historical contexts.

The concept of an AI-driven multiverse, which allows for multiple interpretations and reconstructions, represents a fundamental shift in archaeology and cultural heritage studies. Rather than seeking singular, definitive narratives, this approach embraces complexity, uncertainty, and subjectivity. By integrating AI-driven simulations with human cognitive studies, this research suggests a future where technology and human creativity work in tandem to reimagine history as a dynamic and evolving discourse.

However, with this transformation comes the responsibility to critically assess the role of AI in shaping historical and artistic narratives. Questions of authorship, authenticity, and interpretive bias remain central. Future research must continue to refine methodologies for integrating AI with cognitive science, ensuring that the resulting interpretations remain scientifically sound and ethically responsible. The findings from this study lay the groundwork for a broader, more nuanced approach to AI-assisted heritage studies, emphasizing the importance of collaborative, interdisciplinary frameworks.

This paradigm shift ultimately enriches not only our understanding of the past but also our collective cultural imagination.

Funding: This research received no external funding.

Data Availability Statement: Part of the data might be available on request.

Acknowledgments: For the eye-tracking and VTS experiments I am very grateful to all my Duke University classes on “Why Art”. The AI-distillery experiment was supervised by Felipe Infante de Castro while the VTS survey was designed in collaboration with Vincenza Ferrara, Sapienza University. I am very grateful for these very productive collaborations.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Forte, M.; De Castro, F.I. Digital AI for IA: Artificial Intelligence for Interpretative Archaeology, in AI, Cultural Heritage and Art. In *Between Research and Creativity, Workshop Proceedings*; Guidazzoli, A., Liguori, M.C., Eds.; CINECA: Brussels, Belgium, 2024.
2. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
3. Forte, M.; Siliotti, A. *Virtual Archaeology: Re-Creating Ancient Worlds*; H.N. Abrams: New York, NY, USA, 1997.
4. Forte, M. Cyber-Archaeology: Notes on the simulation of the past. *Virtual Archaeol. Rev.* **2011**, *2*, 7–18. [\[CrossRef\]](#)
5. Tenzer, M.; Pistilli, G.; Brandsen, A.; Shenfield, A. Debating AI in Archaeology: Applications, implications, and ethical considerations. *Internet Archaeol.* **2024**, *67*, 8. [\[CrossRef\]](#)
6. Clavert, F.; Gensburger, S. 'Is Artificial Intelligence the Future of Collective Memory? Bridging AI Scholarship and Memory Studies', Call for Papers for the Second Volume of the Memory Studies Review (Brill Publishing). 2024. Available online: https://www.c2dh.uni.lu/sites/default/files/cfp_is_artificial_intelligence_the_future_of_collective_memory_-_2023.pdf (accessed on 2 November 2023).
7. Köpf, A.; Kilher, Y.; von Rütte, D.; Anagnostidis, S.; Tam, Z.-R.; Stevens, K.; Barhoum, A.; Duc, N.M.; Stanley, O.; Nagyfi, R.; et al. OpenAssistant conversations—Democratizing large language model alignment. *arXiv* **2023**, arXiv:2304.07327. [\[CrossRef\]](#)
8. Casini, L.; Marchetti, N.; Montanucci, A.; Orrù, V.; Rocchetti, M. A human-AI collaboration workflow for archaeological sites detection. *Sci. Rep.* **2023**, *13*, 8699. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Lundberg, S.; Lee, S.-I. A unified approach to interpreting model predictions. *arXiv* **2017**, arXiv:1705.07874. [\[CrossRef\]](#)
10. Resler, A.; Yeshurun, R.; Natalio, F.; Giryas, R. A deep-learning model for predictive archaeology and archaeological community detection. *Humanit. Soc. Sci. Commun.* **2021**, *8*, 295. [\[CrossRef\]](#)
11. Ajami, R.A.; Karimi, H.A. Artificial Intelligence: Opportunities and Challenges. *J. Asia Pac. Bus.* **2023**, *24*, 73–75. [\[CrossRef\]](#)
12. Gallese, V. The Multimodal Nature of Visual Perception. *Phenomenol. Cogn. Sci.* **2003**, *2*, 123–143.
13. Gallese, V.; Sinigaglia, C. What Is So Special About Embodied Simulation? *Trends Cogn. Sci.* **2011**, *15*, 512–519. [\[CrossRef\]](#)
14. Gallese, V.; Gattara, S. The Roots of Empathy: The Shared Manifold Hypothesis and the Neural Basis of Intersubjectivity. *Psychopathology* **2006**, *39*, 171–180. [\[CrossRef\]](#)
15. Li, Z. A Saliency Map in Primary Visual Cortex. *Trends Cogn. Sci.* **2002**, *6*, 9–16. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Tegmark, M. Parallel universes. *Sci. Am.* **2003**, *288*, 40–51. [\[CrossRef\]](#)
17. Biamonte, J.; Faccin, M.; De Domenico, M. Complex networks in quantum mechanics. *Commun. Phys.* **2019**, *2*, 53. [\[CrossRef\]](#)
18. Bernard, C. (Ed.) Frontmatter. In *Universe or Multiverse*; Cambridge University Press: Cambridge, MA, USA, 2007.
19. Park, J.; Choi, J.; Kim, S.-L.; Bennis, M. Enabling the Wireless Metaverse via Semantic Multiverse Communication. *arXiv* **2022**, arXiv:2212.06908. [\[CrossRef\]](#)
20. Gattiglia, G. A postphenomenological perspective on digital and algorithmic archaeology. *Archeol. E Calc.* **2022**, *33*, 319–334. [\[CrossRef\]](#)
21. Forte, M. Artificial intelligence rethinks the past: How computers are reconstructing Etruscan and Roman landscapes. In *World Archaeology*; Issue 126; Current Publishing: London, UK, 2024.
22. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bannetot, A.; Tabik, S.; Barbado, A.; Garcia, A.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [\[CrossRef\]](#)
23. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21, Toronto, ON, Canada, 3–10 March 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 610–623. [\[CrossRef\]](#)
24. Pistilli, G. What Lies Behind AGI: Ethical Concerns Related to LLMs. *Revue Ethique et Numérique*, hal-03607808. 2022. Available online: <https://hal.science/hal-03607808> (accessed on 2 December 2024).
25. Available online: <https://rethinkingthepast.org/> (accessed on 2 December 2024).
26. Giorgi, A.; Menicocci, S.; Forte, M.; Ferrara, V.; Mingione, M.; Alaimo Di Loro, P.; Inguscio, B.M.S.; Ferrara, S.; Babiloni, F.; Vozzi, A.; et al. Virtual and Reality: A Neurophysiological Pilot Study of the Sarcophagus of the Spouses. *Brain Sci.* **2023**, *13*, 635. [\[CrossRef\]](#)
27. Forte, M. Perceiving Etruscan Art: AI and Visual Perception. *Humans* **2024**, *4*, 409–429. [\[CrossRef\]](#)
28. Kanwisher, N.; Yovel, G. The fusiform face area: A cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. B Biol. Sci.* **2006**, *361*, 2109–2128. [\[CrossRef\]](#)
29. Chatterjee, A. Prospects for a Cognitive Neuroscience of Visual Aesthetics. *Bull. Psychol. Arts* **2003**, *4*, 55–60.
30. Grill-Spector, K.; Knouf, N.; Kanwisher, N. The fusiform face area subserves face perception, not generic within-category identification. *Nat. Neurosci.* **2004**, *7*, 555–562. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Kang, D.; Youn Kyu, L.; Jongwook, J. Exploring the Potential of Event Camera Imaging for Advancing Remote Pupil-Tracking Techniques. *Appl. Sci.* **2023**, *13*, 10357. [\[CrossRef\]](#)

32. Massaro, D.; Savazzi, F.; Di Dio, C.; Freedberg, D.; Gallese, V.; Gilli, G.; Marchetti, A. When Art Moves the Eyes: A Behavioral and Eye-Tracking Study. *PLoS ONE* **2012**, *7*, e37285. [CrossRef]
33. Pupil Invisible: Eye Tracking for Real-World Applications. Available online: <https://pupil-labs.com/products/invisible> (accessed on 1 December 2024).
34. Ciascai, L.; Haiduc, L. Visual Thinking Strategies—Theory and Applied Areas of Insertion. *Sustainability* **2022**, *14*, 7195. [CrossRef]
35. Yenawine, P. *Visual Thinking Strategies: Using Art to Deepen Learning Across School Disciplines*; Harvard Education Press: Cambridge, MA, USA, 2013; ISBN 978-1-61250-609-8.
36. Available online: <https://followfox.ai/academy> (accessed on 15 January 2025).
37. Borji, A.; Itti, L. State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 185–207. [CrossRef] [PubMed]
38. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [CrossRef]
39. Itti, L.; Koch, C. Computational Modelling of Visual Attention. *Nat. Rev. Neurosci.* **2001**, *2*, 194–203. [CrossRef]
40. Elgammal, A.; Liu, B.; Elhoseiny, M.; Mazzone, M. CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms. *arXiv* **2017**, arXiv:1706.07068. Available online: <https://arxiv.org/abs/1706.07068> (accessed on 25 January 2025).
41. Kümmerer, M.; Wallis, T.S.A.; Bethge, M. DeepGaze II: Reading Fixations from Deep Features Trained on Object Recognition. *arXiv* **2016**, arXiv:1610.01563. Available online: <https://arxiv.org/abs/1610.01563> (accessed on 25 January 2025).
42. Krasovskaya, S.; MacInnes, W.J. Saliency Models: A Computational Cognitive Neuroscience Review. *Vision* **2019**, *3*, 56. [CrossRef]
43. Amini, E.; Javadi, S.; Khatibi, S. Saliency Map Generation Based on Human Level Performance. In Proceedings of the 2024 IEEE Gaming, Entertainment, and Media Conference (GEM), Turin, Italy, 5–7 June 2024; pp. 1–5. [CrossRef]
44. Koide, N.; Kubo, T.; Nishida, S.; Shibata, T.; Ikeda, K. Art expertise reduces influence of visual saliency on fixation in viewing abstract-paintings. *PLoS ONE* **2015**, *10*, e0117696. [CrossRef] [PubMed]
45. Yan, F.; Chen, C.; Xiao, P.; Qi, S.; Wang, Z.; Xiao, R. Review of Visual Saliency Prediction: Development Process from Neurobiological Basis to Deep Models. *Appl. Sci.* **2022**, *12*, 309. [CrossRef]
46. Mancas, M.; Gosselin, B. Computational saliency models for art analysis. *Symmetry* **2017**, *9*, 98.
47. Pu, Y.; Liu, D.; Chen, S.; Zhong, Y. Research Progress on the Aesthetic Quality Assessment of Complex Layout Images Based on Deep Learning. *Appl. Sci.* **2023**, *13*, 9763. [CrossRef]
48. Agosti, M.; Ferro, N.; Silvello, G. AI-enhanced saliency visualization for interactive museum exhibitions. *Heritage* **2019**, *2*, 1520–1535.
49. Kim, D.; Lee, S.; Lee, J. Gaze-Based Visualization for Analyzing Human Interaction with Visual Stimuli in Cultural Heritage Research. *Sensors* **2021**, *21*, 5178. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.