

Article

From Alignment to Evocation: On the Capability Boundaries and Collaborative Paths of AI Art Creation—A Framework Based on the Neuroaesthetic “Ring Scale” and Prompt Engineering

Xianqun Yi ^{1,*} and Hongsheng Li ²¹ Independent Researcher, Nanning 530000, China² Independent Researcher, Shengzhen 518000, China; 1243905313@qq.com

* Correspondence: 849676229@qq.com

Abstract

Recent generative art outputs across music, literature, painting and moving-image media have attracted extensive scholarly and public interest, yet evaluations of their creative capacities are mostly limited to informal observational accounts. Drawing on neuroaesthetic reasoning, this paper puts forward a dual-layer analytical framework that differentiates two distinct modes of aesthetic reception: Alignment, defined as statistical template matching, and evocation, referring to the novel association of scattered embodied memory fragments. Building on this binary categorization, the study introduces the tentative Ring Scale taxonomy—a figurative target-shooting metaphor rather than quantitative metric—as a purely descriptive tool for stratifying relative aesthetic evocation intensity. This framework further unpacks the neurocognitive underpinnings of auditory, visual and textual aesthetic pathways, alongside their combined multimodal interactions within film and television works. It tentatively accounts for why generative systems tend to deliver more cohesive aesthetic outcomes within the auditory domain, and hypothesises a present functional limitation of current large models: these systems perform comparatively well within Alignment-driven aesthetic effects, while layered high-order evocation remains constrained by inherent structural limitations of statistical training architectures. From this diagnostic observation, three directional paradigm shifts for human–AI collaborative creation are outlined: shifting from human substitution to human–machine complementarity, shifting from exhaustive template imagery generation to targeted latent fragment elicitation, and shifting from optimising figurative Ring-tier descriptive labels to pursuing transformative aesthetic fission effects. The study frames imaginative cognition as the central driving force behind fruitful human–AI co-creation, and positions prompt engineering as the actionable operational bridge connecting human imaginative thought to machine-executable generative parameters. Three tentative prompt design tactics are then elaborated: physiological arousal framing, multisensory scenario simulation prompts, and intentional strategic blank-leaving. Additionally, this work discusses the plausible constructive functions of model hallucination phenomena when viewed through the lens of high-tier aesthetic evocation, rather than merely framing such outputs as technical errors. All judgments and tier comparisons raised throughout the paper are framed as unvalidated observational hypotheses open to empirical testing. To facilitate follow-up empirical scrutiny, the paper collates a full set of testable hypotheses derived from its theoretical reasoning and outlines feasible experimental validation pipelines, with an open

Academic Editor: Elena Sidorova

Received: 11 June 2026

Revised: 24 July 2026

Accepted: 29 July 2026

Published: 3 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

call for controlled empirical research to corroborate or refine the proposed qualitative framework.

Keywords: AI art creation; neuroaesthetics; Alignment and evocation; Ring Scale; imagination; prompt engineering; AI hallucination; testable hypotheses

1. Introduction: A Question Demanding Clarification

Between 2025 and 2026, AI-generated content—especially in the domain of music—has stirred strong public response. Notable examples include AI-rendered versions of Jay Chou’s *Hair Like Snow* reimagined in the aesthetic style of the Tang Anxi Frontier Command, and *Lonely Sandbank Cold* infused with revolutionary sentimental undertones. The performance, arrangement and emotional expressiveness of such AI pieces deliver a distinct visceral feeling of being “struck” to mass audiences. Meanwhile, AI literary and visual works have also made visible technical strides, yet they generally yield less widespread emotional resonance and public engagement compared to AI music. Since late 2025, AI filmmaking has gradually taken shape, ranging from standalone AI short films to AI-augmented full-length narrative features. Several representative works achieve sophisticated visual texture, narrative rhythm and emotional expression, representing a critical transition from single-modal generation to integrated multimodal AI creation.

This observable disparity brings forward a core research question: What underlying factors lead to uneven aesthetic performance of generative AI across different artistic media? Does this gap originate from unbalanced technical iteration progress, or inherent neurocognitive distinctions among art forms? Beyond merely explaining the existing divergence, if researchers and creators intend to elevate the expressive depth of AI art, what cohesive theoretical tool can be adopted, and what actionable operational paths can be constructed?

This paper aims to tentatively outline a systematic interpretative framework grounded in neuroaesthetic theories to fill this research gap. As a hypothesis-driven conceptual theoretical paper without original empirical experiments, this work attempts to offer unified neurocognitive explanations for cross-modal AI aesthetic differences, alongside a set of testable collaborative creation strategies and prompt engineering guidelines for human–AI co-creation.

2. Glossary

This glossary clarifies all core interdisciplinary terminology adopted throughout this manuscript, distinguishes empirical neuroscientific terms, cross-cultural aesthetic concepts, and tentative theoretical constructs proposed by the present paper. Brief in-text citations (Author, Year) are attached for disciplinary grounding; full bibliographic information is consolidated uniformly in the References section without duplication.

2.1. Alignment

Definition 1. *A low-level aesthetic matching mechanism driven by statistical fitting. It refers to the cognitive pleasure triggered when art’s audio-visual/textual features match readers’ pre-stored labeled experiential templates, centered on recognition rather than novel emotional awakening. Alignment can be regarded as one core prerequisite for most aesthetic reception processes (Hu 2026).*

2.2. Evocation

Definition 2. *A high-level aesthetic experience based on the recombination of scattered embodied memory fragments. Art stimuli activate dormant multimodal memory fragments in audiences and build unprecedented associative links between these fragments, generating ineffable immersive shock. It serves as the phenomenological carrier of Yijing in Chinese aesthetics. This term integrates Western neurocognitive memory theory and Chinese classical aesthetic thought (Vessel and Rubin 2010).*

2.3. Yijing (Aesthetic Evocative Resonance)

Definition 3. *Indigenous core concept of Chinese aesthetics standardized for international interdisciplinary communication. It denotes an integrated experiential field formed by the fusion of imagery, bodily sensation, implicit sentiment and blank implication. Different from Western aesthetic theories focusing on formal beauty or direct emotional expression, Yijing relies on Liubai (blank-leaving) to invite audiences to complete meaning construction through imagination. While this term is not constantly foregrounded in every subsection, its logic of reserved interpretive space parallels Wolfgang Iser's "textual gaps" in Western literary theory; discussions of non-objective painting, void-vitality interplay and viewer's imaginative immersion deliver comparative aesthetic reasoning aligned with the core connotation of Yijing (Jullien and Todd 2009).*

2.4. Latent Memory Network

Definition 4. *Composite theoretical construct proposed in this paper combining embodied cognition and memory neuroscience. It describes distributed long-term storage of unlabeled, subthreshold multimodal sensory fragments across brain sensory regions and hippocampal circuits. These fragments remain unconscious until large-scale neural synchronization triggered by aesthetic stimuli. It differs from explicit semantic and working memory defined in mainstream neuroscience (Nikolaev and van Leeuwen 2018).*

2.5. Ring Scale

Definition 5. *Exploratory qualitative taxonomy proposed in this paper for grading the intensity of aesthetic evocation, explicitly not a psychometrically validated quantitative measurement scale. It classifies aesthetic experience layers according to the range of synchronized memory fragments and cross-brain neural coordination degree. All ring numerical labels draw on a target-shooting figurative metaphor only, functioning merely as descriptive categorical markers rather than quantifiable metrics. This taxonomy lacks standardized scoring criteria and inter-rater reliability verification; full behavioral and physiological experimental protocols for subsequent empirical validation are elaborated in later sections (Yi 2026).*

2.6. Default Mode Network (DMN)

Definition 6. *Mature empirical neuroscientific concept referring to a resting-state brain network including medial prefrontal cortex, posterior cingulate cortex and bilateral inferior parietal lobules. It undertakes self-reflection, autobiographical memory extraction and imaginative contemplation. High-tier A/A+ evocation requires functional coupling between DMN and limbic-sensory pathways. Nearly all neuroimaging claims within this manuscript draw upon published empirical data, while simplified circuit descriptions are adopted for illustrative metaphorical purposes (Vessel et al. 2012).*

2.7. Neural Synchronization

Definition 7. *Standard neuroimaging terminology describing coordinated oscillatory activities across distributed brain regions induced by unified aesthetic stimuli. Large-scale cross-modal synchronization corresponds to deep Yijing experience, while localized weak synchronization only generates mild alignment pleasure. The term is used literally in this paper without figurative extension (Sachs et al. 2020).*

2.8. Amygdala Emotional Pathway (Auditory Short-Circuit Pathway)

Definition 8. *Simplified affective neuroscience circuit deployed to interpret music's unique emotional impact. Certain auditory signals may skip exhaustive high-level cortical semantic processing and trigger fast limbic responses within the amygdala, generating preliminary physiological affective arousal. This paper only invokes this neural mechanism once as an explanatory metaphor, without ignoring multi-cortical regulatory factors shaped by culture and personal experience (Damasio 1999).*

3. Core Framework: Alignment and Evocation

Imagination arguably serves an indispensable foundational role across all artistic media. Whether in music, painting, literature, or film, creators first construct novel mental scenario simulations: they visualise unseen scenes, hear unwritten melodies, and embody the inner sensations of fictional characters. This ability to disassemble and recombine scattered experiential fragments constitutes the core driving force of artistic creation. Once a work is finished and presented to audiences, imaginative activity transfers from creator to recipient. Viewers fill the work's interpretive gaps with their own stored fragments of lived experience, reconstructing immersive inner scenes and even projecting their own identities into the depicted world. The full vitality of art emerges through this two-way circulation of imagination between creator and audience.

Western aesthetic theories elaborate the cognitive logic of imaginative recombination. Koestler's (1964) *The Act of Creation* puts forward the theory of bisociation, holding that all creative acts—artistic, literary, humorous and scientific alike—stem from the unexpected juxtaposition of ideas belonging to separate thought matrices. This reasoning shares structural isomorphism with the mechanism of “category displacement” and “imagery collision” analysed later in this paper. Martindale's (1992) *The Clockwork Muse* advances the habituation/dishabituation framework, arguing that the fundamental motivation of artistic innovation lies in pursuing novel sensory arousal to break rigid aesthetic paradigms. As a product of random deviation from training data distributions, AI hallucination provides a natural computational counterpart to this artistic dishabituation mechanism.

3.1. Alignment: Aesthetics as Statistical Fitting

“Alignment” describes the matching state where an artwork's imagery and emotional patterns align with the audience's pre-existing experiential templates. Typical audience feedback under this mode includes judgements such as “this feels authentic” or “this matches my expectations.” It constitutes the basic entry-level tier of aesthetic reception.

Neurologically, Alignment arises when external stimuli successfully activate labelled memory templates stored in the brain; aesthetic pleasure emerges when the prediction error between input and stored templates falls within an optimal range. Its technical foundation lies in statistical fitting. Trained on massive human artwork datasets, generative AI can map different imagery and acoustic combinations to corresponding audience affective

responses, then conduct optimised sampling within this mapping space. Contemporary models deliver comparatively polished Alignment-oriented outputs, with musical creation standing as the most mature field of application.

3.2. *Evocation: Aesthetic Experience at the Fragment Recombination Level*

“Evocation” occurs when artistic imagery activates scattered, unintegrated latent memory fragments within audiences and establishes brand-new associative connections between them. Unlike the recognitive satisfaction of Alignment, evocation brings an inef-fable immersive impact, often described by audiences as “this struck me” or “it expresses feelings I could never put into words.” This layered immersive experience corresponds to the Chinese aesthetic concept of *Yijing*.

Its neural basis hinges on the latent memory network formed by a lifetime of multi-modal embodied experience. Countless subtle sensory fragments—the chill of rain accompanying a throat tightness during parting, the faint ache in the neck while watching a crescent at dusk—remain stored below conscious awareness, unlabelled and rarely retrieved in daily cognition. These fragments lie dormant until unified artistic stimuli trigger coordinated neural activation. Rather than supplying brand-new objective information, powerful evocation awakens these scattered fragments and builds interconnections between them, forming an integrated holistic perceptual field: this is the aesthetic state of *Yijing*. Such profound resonance relies on three intertwined conditions: accumulated embodied lived experience, subthreshold latent memory storage, and the imaginative capacity to synchronously activate discrete sensory fragments.

3.3. *The Relationship Between Alignment and Evocation*

Alignment acts as a necessary preliminary condition for most aesthetic engagement; audiences rarely enter deep aesthetic perception without basic template recognition, which serves as the baseline for aesthetic access.

Evocation exists along a continuous depth spectrum:

Preliminary evocation (B+): activates a small set of memory fragments, generating mild vague emotional resonance;

Large-scale evocation (A): synchronises multimodal memory across distinct brain regions, producing the visceral “being struck” sensation;

Extreme evocation (A+): triggers wide-ranging functional synchronisation between the default mode network and task-related neural circuits, temporarily blurring self-boundaries and potentially producing long-term shifts in perceptual habits after appreciation.

From observational comparison of existing generative works, we tentatively hypothesise that current large language and multimodal models achieve stable, high-quality Alignment effects via statistical matching, and can reach preliminary B+-grade evocative effects. Nevertheless, inherent statistical training architectures appear to restrict their capacity to independently generate A-level large-scale evocation and A-level extreme transformative aesthetic resonance.

4. Neural Pathways: The Neurological Basis of Three Art Forms

4.1. *The Auditory-Emotional Direct Pathway (Music)*

Auditory stimuli can trigger rapid subcortical affective responses, yet this subcortical circuit is only adopted as a simplified explanatory metaphor in this paper. Partial auditory signals may bypass full layers of cortical semantic analysis and activate the amygdala to generate immediate physiological shifts such as altered heart rate and skin conductance, as outlined in Damasio (1999). Crucially, musical emotion relies on distributed whole-

brain interactions shaped by individual memory and cultural background, rather than this single isolated shortcut.

Sound-triggered limbic priming lays the neural foundation for preliminary evocation. Unlike text and visual imagery that demand deliberate semantic decoding, musical tones can activate scattered latent memory fragments without conscious analytical effort, readily generating mild resonant experiences corresponding to the 4–6 figurative Ring tiers defined in the qualitative taxonomy.

This auditory pathway boasts highly tunable acoustic parameters: tempo, harmonic tension, timbral brightness and dynamic envelopes can all be precisely calibrated within large parameter spaces, giving current generative AI a comparative advantage in searching for emotion-optimised acoustic combinations that are difficult for human creators to exhaustively test.

Even cross-regional neural synchronisation linked to aesthetic resonance is not a fixed universal neural pattern. Long-term cultural environments and personal life histories profoundly shape audience neural responses (Yang et al. 2019; Lee 2025). Cross-cultural neuroimaging research confirms consistent aesthetic bias: audiences exhibit stronger large-scale neural synchronisation when appreciating artworks matching their native cultural schemas, while identical auditory stimuli trigger divergent latent memory retrieval across Eastern and Western observers (Bao et al. 2013). Such cultural moderators add layered complexity to simplified subcortical models and demonstrate that no single neural circuit can fully account for cross-cultural aesthetic resonance.

4.2. *The Visual-Semantic Construction Pathway (Painting and Visual Art)*

Visual signals travel through the retina, lateral geniculate nucleus and occipital cortex before entering the ventral stream for object recognition and meaning construction. This processing pipeline operates slower than the auditory pathway and heavily depends on prefrontal and parietal world models that encode spatial logic, material tactile expectations and physical light–shadow rules. From an evolutionary perspective, the human visual system has evolved into a refined perceptual matching system specialised for real-world scene identification.

Visual aesthetic perception is constrained by two intertwined thresholds: reality matching and Yijing evocation. Current generative models carry two inherent structural limitations here: they lack embodied three-dimensional spatial cognitive frameworks and have no storage of lived multimodal sensory memory traces accumulated through real-world bodily experience.

4.3. *The Multimodal Scenario Simulation Pathway (Literature)*

Reading text is a visual act, yet literary aesthetic experience hinges on constructing full-body multimodal mental simulation. Take Han Yu's exile poem line "Snow blocks the Blue Pass, the horse cannot advance" as an example. Seven Chinese characters activate layered perceptions: bone-chilling cold, rugged mountain terrain, personal frustration, and shared existential predicaments familiar to many readers from their own life trajectories. It is this simultaneous cross-modal activation of scattered memory fragments that enables the poem to produce enduring resonant shock across centuries.

Ezra Pound's Imagist verse *In a Station of the Metro* follows the same core mechanism of juxtaposing unrelated imagery to stir audience imagination:

"The apparition of these faces in the crowd;
/Petals on a wet, black bough."

Compared with Han's writing rooted in Chinese historical context, Pound's work focuses on instantaneous visual contrast, yet both rely on unmediated imagery collision to spark immersive aesthetic feeling.

Literary creation imposes higher activation thresholds for aesthetic resonance. To achieve layered Yijing experience, artistic language must simultaneously mobilise sensory fragments across multiple modalities and forge novel associative links between them. Imagination acts as an indispensable medium on both creation and reception sides: writers compress complete multimodal mental scenes into text, while readers rely on their own embodied memory to reconstruct immersive scenarios while reading.

4.4. *The Unique Position of Multimodal Integrated Art (Film)*

Film integrates auditory, visual and literary aesthetic pathways into a unified temporal framework, making it a rigorous comprehensive test bed for generative AI's overall creative capacity. Its aesthetic potential is multiplicative rather than additive: strengths or defects within any single sensory channel will be amplified by the other two modalities.

Genuine cinematic evocation demands consistent cross-modal resonance. When image, score and dialogue converge on a unified emotional tone, they create amplified immersive impact; mismatched emotional logic across channels creates perceptual fissures that pull audiences out of the immersive state.

The opening sequence of *Dune* (2021) illustrates this logic clearly. Hans Zimmer crafted original custom instruments to generate unprecedented timbres matching the film's alien desert atmosphere, a form of embodied imaginative creation unreachable through pure statistical template matching. By contrast, when tasked with scoring identical desert footage, present AI models will only sample existing "desert-orchestral" acoustic templates from training datasets. This comparative case lends tentative observational support to this paper's core tentative inference: within Alignment tasks, AI can exhaust existing artistic paradigms, yet wholly original timbral and multimodal conceptions rooted in human embodied imagination remain difficult for autonomous generative workflows to produce under current statistical architectures.

Although multiple representative AI short and feature-length films have emerged since late 2025, the most successful cases still require deep human imaginative intervention across the full creation pipeline—global story architecture, core emotional beat layout, and control of synchronised multimodal resonant rhythm. These stages demand holistic cross-modal foresight, which positions human imaginative cognition as the primary generative driving force within human–AI film co-creation.

5. The Ring Scale: A Qualitative Taxonomy for Grading Aesthetic Evocation Intensity

This qualitative taxonomy borrows target-shooting ring scores purely as an intuitive figurative device, rather than a rigorously validated, mathematically precise psychometric measuring tool. Numerical labels ranging from 1 to 10 act only as symbolic tier markers to separate distinct depths of aesthetic resonance. The "target" of this metaphor corresponds to the audience's scattered latent embodied memory fragments that artistic works seek to awaken.

This figurative numerical framework was never constructed for universal standardised quantitative measurement, nor can it generate stable, replicable objective scores without follow-up-controlled laboratory experiments. Any ring-tier assessments of existing AI artworks offered in this paper are tentative observational summaries drawn from publicly available generative art examples, not conclusions proven via systematic empirical testing. Future empirical research may adjust or reconstruct this tiered metaphor by adopting the subjective rating and physiological detection experimental protocols elaborated later in this article.

Drawing on the dual theoretical framework of Alignment and evocation derived in prior chapters, this paper puts forward a provisional descriptive categorization system named the Ring Scale to distinguish diverse levels of aesthetic resonance and immersive audience experience. We explicitly clarify from the outset that this taxonomy exists solely as an exploratory qualitative analytical instrument, not a formally validated psychometric scale. It lacks uniform scoring rules, inter-rater reliability validation, and normative experimental benchmarks. Every ring tier operates merely as a descriptive categorical label for sorting different depths of audience aesthetic perception, without backing from large-scale systematic quantitative data.

Based on the foregoing framework, this paper proposes the figurative “Ring Scale” to qualitatively stratify layers of aesthetic impact and resonant emotional intensity (Table 1).

Table 1. Qualitative Taxonomy of Aesthetic Evocation Intensity (Provisional Ring Framework).

Figurative Ring Tier (Metaphorical Marker Only)	Subjective Experiential Description	Corresponding Neural Performance	Core Mechanism	Evocation Grade
Miss	No positive aesthetic response; dissonance or aversion may arise when stimuli mismatch all familiar internal templates	Prediction error deviates drastically from the acceptable range for template matching	Complete failure of Alignment matching	N/A
1–3 Rings	Superficial recognitive comfort: The artwork feels familiar and predictable to viewers	Activation is limited to isolated stored perceptual templates; no novel cross-regional neural associations are formed	Pure Alignment matching, no latent memory mobilisation	N/A
4–6 Rings	Mild subtle resonance: The piece feels intriguing and faintly affecting	A small set of scattered latent memory fragments are weakly activated, without broad cross-brain neural coordination	Alignment remains dominant, with tentative fragment recombination (only preliminary resonance)	B+ Preliminary evocation
7–8 Rings	Powerful immersive affective impact, often accompanied by somatic sensations (e.g., chills, tearfulness)	Synchronous co-activation of distributed multimodal embodied memory traces, constructing an integrated Yijing perceptual field	Evocation surpasses Alignment as the dominant aesthetic effect	A Large-scale evocation
9 Rings	Potential lasting perceptual shift after aesthetic engagement	Prolonged large-scale functional coupling between the default mode network and limbic-sensory circuits, transient blurring of self-perceptual boundaries	Deep, reconstructive evocative resonance	A+ Extreme evocation
10 Rings	Rare, ineffable peak aesthetic epiphany (extremely uncommon in existing AI works)	Sustained wide-ranging cross-cortical neural synchronisation, with diminished subjective perception of temporal flow	Ultimate form of holistic aesthetic evocation	A+ Extreme evocation

Note: All ring numbers function exclusively as figurative markers originating from target-shooting scoring logic, not empirically validated quantitative metrics. No unified scoring rubric or cross-observer reliability test has been established for this taxonomy. All tier divisions remain open to revision through subsequent empirical studies.

This qualitative framework theoretically helps separate superficial matching satisfaction derived from Alignment and profound spiritual resonance brought by evocation; it further distinguishes transient emotional impact from lasting cognitive adjustment. Drawing only on informal observational comparisons of publicly released generative artworks and adopting the target-shooting metaphor laid out above as a descriptive frame, we tentatively hypothesise that artistic outputs falling below the 7-Ring tier mainly rely on statistical matching of visual or acoustic features, an area where generative AI has achieved relatively mature performance at present. Tiers 7–8 mark the emergence of holistic Yijing resonance, while Tier 9 and above correspond to transformative aesthetic experiences that require activation of deep latent memory reserves. Under this tentative observational conjecture, generative AI may produce works that reach preliminary evocative effects within musical creation, yet multimodal creation forms such as literary writing and film tend to only attain low-to-medium evocation levels. The comparative constraint visible across existing multimodal AI outputs, as we tentatively infer, may lie in the absence of global imaginative cognition to support unified cross-modal evocation, a gap that human–AI collaborative creation might compensate for.

It would be overly one-sided to treat firsthand embodied personal experience as an indispensable threshold for generating high-evocation, empathy-rich artistic works. Human creators frequently produce deeply touching, resonant art concerning trauma, loss and transcendence without corresponding personal suffering; they reconstruct coherent emotional fragments from shared cultural archives, historical records and literary accumulation to realise authentic high-tier evocation effects (Coëgnarts 2025). This counterexample mitigates rigid binary reasoning that directly equates physical lived experience with aesthetic depth. Though current generative large models lack native embodied sensory perception, their massive training corpus integrates aggregated human affective memory data, enabling partial reconstruction of cross-modal memory fragments to achieve mild-to-moderate evocation within the 4–6 Ring tier (Plate and Hutson 2025). Rather than positing an unbridgeable divide between human and machine aesthetic evocation, this paper frames embodied sensory experience as a major amplifier of extreme A+ peak resonance rather than an indispensable prerequisite for resonant aesthetic experience.

Supplementary Empirical Verification Scheme

To respond to the potential criticism that this figurative tiered classification may appear arbitrary without empirical grounding, this paper designs a complete multi-dimensional experimental verification roadmap for follow-up empirical research, covering subjective evaluation, physiological signal detection, and cross-cultural control groups.

Subjective rating experimental design: Recruit volunteer participants covering diverse age groups, educational backgrounds, and degrees of artistic exposure. Standardised artistic samples spanning music, painting, literature and film will be provided. Participants will complete two sets of subjective evaluation indicators: familiarity scoring for Alignment effects, and open descriptive items recording the intensity of emotional shock and resonance, which can be mapped onto the taxonomy's figurative ring tiers.

Objective physiological indicator matching test: Collect real-time skin conductance response (SCR) and heart rate variability (HRV) data while participants appreciate artworks. Fluctuations in physiological signal amplitude can be compared across different

evocation tiers, to tentatively explore whether high-tier resonant experience, linked to large-scale neural synchronisation, generates significantly stronger physiological arousal than low-tier Alignment-oriented works.

Cross-cultural controlled experiment: Set participant groups from distinct Eastern and Western cultural and linguistic backgrounds. Comparative analysis will be conducted on the distribution of figurative Ring tier ratings assigned to identical artwork stimuli to explore how cultural embodied memory reserves may affect the depth of evocation, and clarify the boundary conditions for applying this qualitative framework.

If subsequent controlled laboratory experiments can identify stable correlational relationships between subjective perceptual experience, physiological indicators and the Ring's figurative tier division, this categorization system could be further optimised into an auxiliary qualitative evaluation tool for aesthetic research. On the contrary, if experimental data present obvious inconsistencies with the tier logic proposed here, the dividing criteria of this provisional taxonomy would need to be revised and adjusted accordingly. This fully testable experimental design demonstrates that the Ring Scale is not merely a subjective arbitrary classification, but a tentative theoretical metaphor open to future empirical falsification and iterative improvement.

6. The Capability Boundaries of AI

The core operational mechanism of contemporary generative artistic models lies in statistical learning. These algorithms map correlations between sensory features and audience affective responses from large-scale human artwork corpora, then perform targeted sampling from the established mapping matrix. Their generative workflow adopts a backward decomposition logic: holistic Yijing aesthetic resonance is split into discrete visual and acoustic imagery units before recombination. While AI can reverse-engineer the link between formal features and audience feedback from massive artistic datasets, this decomposition procedure inherently severs the holistic associative bonds that constitute genuine evocative resonance.

Yijing aesthetic resonance cannot be reduced to a simple aggregate of isolated imagery fragments. It emerges from novel cross-modal connections activated synchronously across scattered embodied memory traces. Artworks reconstructed solely through statistical recombination of pre-separated formal elements can only reproduce superficial visual or acoustic features, rather than reconstruct the integrated experiential field of Yijing, because the inter-fragment associations central to deep resonance are fractured during model training and decomposition.

Taking musical creation as an illustrative case based on observed generative outputs: within Alignment tasks, AI can match or even outperform human creators across many parameter combinations; at the preliminary B+ tier of evocation, precise tuning of acoustic parameters enables mild resonant effects. Nevertheless, when aiming for A and A+ high-tier evocation, current statistical architectures present structural constraints rooted in the absence of cumulative embodied sensory experience and lifelong latent memory reservoirs, which complicate the synchronised activation of scattered experiential fragments required for profound aesthetic impact.

This contrast highlights the divergent core drivers of the two aesthetic modes. Alignment performance is primarily constrained by dataset scale and computational capacity, while imaginative cognition emerges as the decisive factor for high-tier evocation. Imagination refers to the capacity to construct entirely new multimodal mental simulations, reorganise disjoint experiential fragments, and identify interpretive potential within reserved blank spaces. Existing generative systems lack this holistic imaginative faculty; their outputs remain rearrangements of pre-recorded training data, rather than original cross-sensory associations untethered to archived human artistic works. Within human–

AI collaborative pipelines, this tentative observational inference suggests human creators serve as the primary source of holistic imaginative framing, while AI functions as an amplifying auxiliary tool instead of an independent originator of novel resonant aesthetic logic.

Crucially, the foregoing analysis of AI's creative limits constitutes only a tentative inference grounded on currently available research and generative examples, rather than empirically verified conclusive evidence. Its falsifiability defines a clear empirical research agenda for follow-up work: if future researchers develop embodied AI systems that accumulate long-term multimodal interactive sensory data via robotic platforms, and controlled comparative experiments demonstrate that such systems produce equivalent A/A+ evocative effects to human masterpieces, this core hypothesis would be fully overturned. This falsifiable experimental blueprint forms a central strand of the empirical research agenda proposed throughout this paper.

7. From Description to Guidance: Three Strategic Shifts

7.1. Shift One: From Substituting Humans to Complementing Humans

Viewed from the human–AI collaborative perspective, positioning generative AI as a supplementary auxiliary tool delivers the most valuable creative outputs. Such models excel at exhaustive traversal and optimisation of quantifiable parameter combinations, yet they lack the capacity to introduce unquantifiable imaginative layers into creation. A clear division of collaborative labour can be outlined as follows: human creators bear responsibility for holistic imaginative construction, which means conceiving complete multimodal scenario simulations in advance; generative AI undertakes the exhaustive screening of optimal expressive schemes within parameter spaces corresponding to that imaginative framework.

Specific allocation across artistic domains:

Music: Humans define the core emotional tone and overall structural layout, while AI optimises arrangement and acoustic parameters.

Literature: Humans supply embodied multi-sensory scene details, while AI organises textual language and enriches imagery combinations.

Painting: Humans clarify the intended Yijing orientation and core visual fragments, while AI iterates visual presentations and stylistic adaptations.

Film: Humans formulate holistic story logic and key emotional beats, while AI handles shot design, soundtrack matching and cross-modal temporal synchronisation.

7.2. Shift Two: From Exhausting Known Imagery to Detecting Individual Fragments

The mainstream generative logic of current AI systems operates in a closed loop of existing imagery databases, merely recombining archived visual and acoustic materials. Genuine high-tier evocation relies on activating audiences' unmarked, subthreshold latent memory fragments that cannot be categorised via fixed semantic labels. This creates a clear developmental direction for prompt interaction: shift from generalised statistical matching toward personalised fragment mining.

Two feasible interactive mechanisms can be designed for follow-up practice:

Individual fragment detection protocol: Through multi-round iterative communication with a single creator or audience, AI gradually maps the unique distribution of that user's latent memory fragments;

Yijing feedback loop: AI generates draft works iteratively, humans feedback subjective feelings of resonant impact, and the model adjusts generative parameters continuously until matching the user's exclusive latent fragment system.

7.3. Shift Three: From Figurative Tier Markers to Transformative Aesthetic Value

Most prevailing evaluation criteria for AI creation prioritise public popularity indicators such as playback volume, likes and share counts, which conform entirely to the aesthetic logic of Alignment. Works capable of reaching the highest figurative Ring tiers centred on cognitive transformation rarely gain instant mass recognition; they often resonate deeply with a small group of recipients at first and spread gradually through subtle aesthetic influence.

Thus, the assessment system for AI art ought to be supplemented with qualitative indicators measuring transformative perceptual impacts, instead of solely relying on quantitative popularity data. The core evaluation standard should shift from broad public reception to lasting cognitive reshaping brought by aesthetic experience. A unique strength of generative AI lies in its capacity to produce massive niche customised drafts, each tailored to distinct combinations of human latent memory fragments.

8. Grounding in Practice: Imagination-Driven Prompt Engineering

The three strategic shifts outlined above must ultimately land on concrete operational tools: prompt engineering. Broadly defined, prompt engineering refers to the practice of externalising and translating internal imaginative thought. Human creators form complete multimodal scenario simulations in their minds, yet such holistic mental imagery cannot be directly transmitted to generative models. The core task of prompt design is to decompose ineffable inner imaginative scenes into machine-readable linguistic instructions. This translation inevitably creates semantic gaps between human conception and machine output; a central practical goal of prompt engineering is to narrow such information loss as much as possible, converting imaginative frameworks into searchable parameter trajectories for generative systems.

It must be clearly stated at the outset that the three prompt tactics elaborated below—physiological arousal framing, multisensory scenario simulation prompts, and intentional strategic blank-leaving—are operational hypotheses logically derived from the preceding neuroaesthetic framework. They possess coherent theoretical grounding but remain unvalidated by rigorous controlled experiments. This section does not aim to claim these tactics are proven effective; instead, it formalises them as testable starting points for subsequent empirical exploration.

8.1. Strategy One: From “Emotion Labels” to “Physiological Arousal Descriptions”

Abstract emotional tags such as “tragic heroism” or “desolation” are high-level semantic markers that generative AI processes only through statistical matching. Physiological descriptions bypass abstract labelling to target embodied emotional responses at the perceptual level. Rather than instructing AI to “write a tragic heroic passage,” creators may describe visceral physical sensations: “let this melody constrict the listener’s chest, leaving them longing to cry out yet unable to voice any sound, with a tangible urge for tears to well up.” This design logic draws on the simplified auditory-emotional metaphor introduced earlier, which facilitates targeted acoustic tuning to elicit mild limbic affective priming.

Operational essentials: Translate abstract emotional goals into specific embodied sensory experiences. Instead of simply requiring a powerful chorus climax, frame the demand as: “The drum entrance at the chorus should deliver a dull impact that makes the audience’s torso jolt involuntarily.” Generally speaking, finer-grained physiological descriptors tend to guide current generative models toward more precise matching of target acoustic characteristics.

8.2. Strategy Two: From “Imagery Accumulation” to “Multimodal Simulation Prompts”

Stacked isolated visual imagery can only activate visual label matching within training datasets, and rarely triggers full-body immersive mental simulation across multiple senses. Without touch, taste, visceral and kinesthetic perception integrated into the scene, audiences cannot form complete, lived imaginative spaces. Multimodal prompts demand creators to inhabit the imagined scene in their own mind first, perceiving wind temperature, scent, bodily posture and muscle tension and then translate these integrated sensations into textual prompts.

A representative illustrative framing: Do not merely request “an old soldier gazing at sunset atop the city wall.” Instead, construct full sensory context: “Picture an old soldier leaning against battlements. Dry wind chaps his lips, and he carries the faint metallic taste of sand in his mouth. He grips a worn cloth pouch rather than a blade—the token his wife pressed into his hand before deployment, its original hues long faded. The writing should dwell not on the sunset or city ramparts, but the soft, frayed texture of that pouch beneath his palm, and the heavy silence weighing over the whole scene.”

Operational essentials: Integrate cross-sensory embodied details covering touch, taste, visceral feeling and bodily tension within prompts. For film shot design, textual guidance ought to simultaneously specify visual texture, ambient sound timbre and characters’ inner physical sensations; all sensory layers jointly build immersive aesthetic resonance.

8.3. Strategy Three: Strategic Omission—From “Pursuing Perfection” to “Preserving Fissures”

Generative AI’s default generative tendency leans toward fully rendered, seamless imagery built on Alignment matching. Yet flawless, exhaustive detail erases interpretive gaps—the necessary imaginative space required for layered Yijing resonance. When every perceptual clue is fully spelled out with no reserved ambiguity, audiences lose space to project their unique latent memory fragments onto the work.

This prompt strategy aligns closely with the Chinese aesthetic concept of Liubai (blank-leaving). Liubai cannot be reduced to empty blank canvas space; it functions as an open invitation for audience imaginative participation, rather than an unfinished artistic flaw. In Chinese landscape painting and calligraphy, blank sections are not neglected unpainted areas—artists achieve expressive power through deliberate omission, letting viewers complete the scene with their own inner vision. The “one-corner, partial landscape” compositions of Ma Yuan and Xia Gui, and sparse depictions of withered branches and lone birds by Bada Shanren all deploy minimal tangible forms to push imaginative activity into reserved voids, turning blankness into the generative field of Yijing.

This aesthetic logic resonates with Wolfgang Iser’s theory of textual gaps in *The Act of Reading* (1978). Iser (1978) argues that indeterminate spaces within literary texts invite readers to participate in constructing meaning, and the imaginative labour of filling such gaps constitutes the core of aesthetic reception. Parallel artistic practices exist across media: musical rests are not interruptions of sound, but sustained reverberation held within the audience’s inner perception; Hemingway’s iceberg theory of literature presents merely one-eighth of explicit narrative, concealing deeper emotional subtext beneath the surface; Francis Bacon’s distorted rendering of *Study after Velázquez’s Portrait of Pope Innocent X* employs blurred, dragging brushwork to create deliberate incompleteness, opening vast psychological imaginative room for viewers—all these practices embody the core logic of Liubai.

Operational essentials: Explicitly guide AI to create controlled, moderate blank-leaving and interpretive fissures within generated works. Concrete operational examples include introducing around 20% subtle instability in the final chorus vocal delivery—breath wavering without full breakdown; allowing the closing electric guitar note to fade slowly

through feedback rather than cutting off cleanly; inserting silent pauses before pivotal dialogue lines, or extending empty static shots in film sequences.

The core adjustable variable lies in the moderate degree of Liubai. Insufficient reserved ambiguity fails to mobilise audience imagination, while uncontrolled, excessive incompleteness disrupts the overall aesthetic coherence of the work. Balancing this spectrum relies on the creator's imaginative judgement and aesthetic intuition to distinguish where full detail is necessary, and where interpretive blankness ought to be retained.

9. AI Hallucination: A Potential Catalyst for Imagination

In the preceding discussion, AI “hallucination”—the model's tendency to generate content inconsistent with training factual constraints—has long been treated as a technical defect to be eliminated. Nevertheless, this property deserves re-examination when situated within artistic creation contexts.

AI hallucination arises from spontaneous divergence from learned data distributions during statistical sampling. Its output imagery combinations fall outside the fixed mapping relationships extracted from human artistic archives. Though such outputs fail to match established aesthetic templates, this unconstrained deviation carries the potential to generate unprecedented sensory pairings and novel associative links unseen in existing creative paradigms.

Human creators have long adopted analogous creative techniques to break habitual thought patterns: Surrealist automatic writing, Dalí's paranoiac-critical creation method, and free jazz's unplanned tonal shifts all constitute deliberate, controlled “creative dissociation” to bypass rigid cognitive filtering and unlock unexamined neural associative pathways.

If we frame AI hallucination as an unconstrained imagery generation mechanism free from fixed artistic paradigms, it delivers three core creative merits for human co-creation:

First, it acts as an imagination trigger. No matter how rich a creator's inner world, their associative logic is bounded by personal life experience and ingrained thinking routines. The bizarre, unprecedented composite imagery produced by model hallucination can excavate latent memory connections which creators would never reach through conventional thought, sparking secondary imaginative expansion. AI does not complete creative imagination independently; instead, its unexpected outputs serve as a springboard to push human creators beyond fixed mental paths.

Second, it functions as an accelerator of high-tier evocation. Works built purely on Alignment deliver predictable, comfortable aesthetic feelings without profound shock. By contrast, the unanticipated scenes generated by hallucination may briefly confuse audiences before constructing brand-new cross-memory associations. This cognitive restructuring process mirrors the neural mechanism of A-grade large-scale evocation: both break original memory linkages and establish brand-new experiential connections.

Third, it reinforces the aesthetic effect of Liubai (strategic blank-leaving). Images produced via hallucination often carry hazy, ambiguous, slightly illogical qualities. Such unintended ambiguity naturally generates interpretive fissures within artworks, distinct from blank spaces intentionally designed by humans. Audiences tend to invest stronger imaginative projection when facing these unplanned, accidental aesthetic voids.

An indispensable precondition must be emphasised: the constructive value of hallucination only emerges under human screening and revision. Unfiltered hallucinatory content mostly produces incoherent factual or aesthetic mismatches with no artistic merit. Only when creators leverage their imaginative judgement to pick, refine and reconstruct valuable “accidental outputs” can hallucination transform from a technical flaw into a creative catalyst. This model feature can never replace human holistic imaginative cogni-

tion, yet it provides an extra disruptive stimulus to break creators' rigid associative routines, steering imagination away from repetitive thought loops toward novel creative possibilities.

10. Conclusions

This paper constructs a dual interpretative framework of Alignment and evocation grounded in neuroaesthetic reasoning, alongside a figurative Ring Scale taxonomy deployed as a descriptive target-shooting metaphor rather than a quantitative measuring tool. The text systematically unpacks why generative AI exhibits uneven aesthetic performance across auditory, visual, literary and integrated cinematic multimodal creation. This study advances a tentative, observation-based core inference: modern generative models achieve polished, consistent outcomes within Alignment-driven aesthetic tasks, and statistical matching can yield mild B+-grade preliminary evocation. Nevertheless, inherent architectural constraints hinder autonomous generation of A-tier large-scale resonance and A+ extreme transformative aesthetic experience. Such limitations predominantly arise from the absence of cumulative embodied sensory input and lifelong distributed latent memory reservoirs within model systems, which impedes the simultaneous activation of scattered cross-modal experiential fragments without human imaginative guidance.

This tentative comparative observation yields a redefined framework for human–AI collaborative creation. Within co-creative workflows, imaginative cognition serves as the primary generative engine: human contributors supply holistic imaginative framing, while AI undertakes exhaustive screening and optimisation across available parameter spaces. Prompt engineering acts as the vital translational bridge that converts human mental imaginative scenes into machine-executable generative instructions. The three proposed prompt design tactics—physiological arousal framing, multisensory simulation prompts, and intentional strategic blank-leaving—deliver actionable operational routes corresponding to affective, perceptual and structural aesthetic dimensions respectively. Meanwhile, the long-dismissed phenomenon of AI hallucination carries untapped collaborative value as an imagination trigger, evocation accelerator and reinforcement of Liubai blank-leaving aesthetics. This latent creative potential does not contradict the paper's central tentative hypothesis that contemporary models lack independent holistic imaginative cognition; instead, it expands the collaborative dynamic beyond one-way human-to-machine input, establishing iterative imaginative interplay: humans provide core creative vision, AI delivers randomised disruptive imagery perturbations, and joint iterative refinement unlocks richer aesthetic possibilities.

This paper adopts a balanced analytical stance, refraining from unilaterally glorifying or dismissing generative artistic capacity. Its core contribution lies in reconfiguring the division of creative labour between humans and algorithms. In the current landscape of AI-assisted art, imaginative cognition emerges as an irreplaceable core competency rather than an obsolete skill superseded by computational tools. Human creators' unique reservoir of embodied, scattered sensory fragments now stands more accessible to profound layered aesthetic awakening through targeted human–AI collaborative resonance—subtle, long-dormant inner experiential traces may now be stirred into vivid immersive aesthetic impact via coordinated co-creation.

A companion Chinese-language manuscript *The Serendipity of Imagery: On the Iso-morphism between AI Hallucination and Aesthetic Mechanisms in Chinese Lyric Writing* (forthcoming) extends this framework to the specialised field of ci poetry composition, supplying domain-specific practical case analysis to validate the theoretical system laid out in this article.

11. Hypotheses, Verification, and Call for Empirical Research

This study's core contribution lies in constructing a fully falsifiable neuroaesthetic framework, rather than putting forward definitive universal conclusions or treating its deductions as ultimate objective truth. Scientific inquiry advances through iterative cycles of hypothesis and empirical validation. A rigorous theoretical hypothesis gains its academic value not from instant empirical confirmation, but from clear logical boundaries and feasible experimental testing routes.

11.1. List of Testable Hypotheses

All five propositions below are framed as tentative conjectures open to subsequent controlled testing:

Hypothesis 1. *When designing prompts for musical and audio creation, physiological sensory descriptions generate stronger subjective emotional resonance and measurable physiological arousal in audiences compared to generic abstract emotion labels.*

Hypothesis 2. *Multimodal prompts integrating tactile, gustatory and kinesthetic embodied details produce richer immersive mental simulation and deeper Yijing resonance than prompts relying solely on stacked visual imagery, within literary and cinematic creation tasks.*

Hypothesis 3. *Strategic controlled blank-leaving yields superior evocative aesthetic impact compared to fully polished, detail-saturated works. The effect follows an optimal inverted-U gradient: insufficient interpretive gaps fail to activate audience imagination, while excessive uncontrolled incompleteness disrupts overall aesthetic coherence.*

Hypothesis 4. *After human screening and secondary artistic refinement, unconventional imagery combinations generated via AI hallucination act as effective triggers for creators' imaginative expansion and lift the overall evocative tier of finished works.*

Hypothesis 5. *Aesthetic evocation forms a continuous layered spectrum ranging from preliminary B+ resonance to transformative A/A+ peak experience. High-tier evocation relies heavily on audiences' unique pools of latent embodied memory, resulting in pronounced individual differences in perceptual response.*

11.2. Empirical Verification Framework

All hypotheses above can be investigated through standardised randomised controlled trials. The unified experimental logic holds constant confounding variables including AI model type, creative theme and work length, then systematically contrasts audience perceptual outcomes across distinct prompt and generative strategies.

Measurement indicators combine two layers of evidence:

Subjective self-report metrics: self-rated depth of immersive shock and perceived Yijing resonance;

Objective physiological biomarkers: skin conductance response and heart rate variability to comprehensively capture the psychophysical underpinnings of evocative experience.

Targeted experimental designs for separate hypotheses:

For Hypothesis Three (optimal blank-leaving gradient): Set multiple groups with 0–40% graded controlled omission to verify whether evocative effects fit an inverted-U distribution curve.

For Hypothesis Four (hallucination as imaginative catalyst): Conduct between-subject comparison between creators supplied with standard model outputs and creators provided with hallucination-derived imagery drafts.

For Hypothesis Five (latent memory individual differences): Collect measurements of participants' lifetime-embodied experience richness and mental imagery vividness, then examine their correlational relationship with high-tier evocation intensity.

11.3. Call for Follow-Up Empirical Research

The set of hypotheses and corresponding experimental blueprints outlined above forms a complete translational pathway connecting this conceptual framework to empirical aesthetics. All proposed trials are technically executable with existing laboratory equipment, mainstream generative model APIs and standard human subject experimental protocols, requiring no revolutionary new research hardware.

This paper invites scholars across empirical aesthetics, cognitive neuroscience, generative AI and human–computer interaction to conduct targeted experimental validation of the conjectures above.

If the hypotheses receive consistent empirical support, this research can deliver the first set of evidence-based prompt design norms to guide AI art creation, advancing the theoretical shift from Alignment-focused generation toward evocation-centred co-creation into practical artistic workflows.

If any hypothesis is empirically falsified, this outcome still constitutes valuable theoretical progress: eliminating intuitively appealing yet unworkable logic sharpens the overall precision of the Alignment–evocation framework.

Whether confirmed or refuted through controlled testing, the empirical engagement itself represents the most meaningful scholarly response to this hypothesis-driven conceptual paper. The fundamental purpose of putting forward tentative theoretical conjectures is not to demand universal acceptance, but to invite rigorous experimental scrutiny.

Author Contributions: Conceptualization, X.Y. and H.L.; methodology, X.Y. and H.L.; formal analysis, X.Y.; data curation, H.L.; writing—original draft preparation, X.Y.; writing—review and editing, X.Y. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

References

- Bao, Yan, Aneta Szymaszek, Xiaoying Wang, Anna Oron, Ernst Pöppel, and Elzbieta Szlag. 2013. Temporal Order Perception of Auditory Stimuli Is Selectively Modified by Tonal and Non-Tonal Language Environments. *Cognition* 129: 579–85. <https://doi.org/10.1016/j.cognition.2013.05.012>.
- Coëgnarts, Maarten. 2025. Art and embodied knowledge: The bodily basis of aesthetic cognitivism. *The Journal of Aesthetics and Art Criticism* 83: 197–210. <https://doi.org/10.1093/jaac/kpaf025>.
- Damasio, Antonio R. 1999. *The Feeling of What Happens*. New York: Harcourt Brace.
- Hu, Botao Amber. 2026. *Machine Neuroaesthetics Is Advancing Computational Creativity*. Oxford: Oxford University Working Paper.
- Iser, Wolfgang. 1978. *The Act of Reading: A Theory of Aesthetic Response*. Baltimore: Johns Hopkins University Press.
- Jullien, François, and Jane Marie Todd. 2009. *The Great Image Has No Form*. Chicago: University of Chicago Press.
- Koestler, Arthur. 1964. *The Act of Creation*. London: Hutchinson.

- Lee, Harin. 2025. Visual and Auditory Aesthetic Preferences Across Cultures. *arXiv* arXiv:2502.14439.
- Martindale, Colin. 1992. *The Clockwork Muse: The Predictability of Artistic Change*. New York: Basic Books.
- Nikolaev, Andrey R., and Cees van Leeuwen. 2018. Scene Buildup From Latent Memory Representations Across Eye Movements. *Frontiers in Psychology* 9: 2701. <https://doi.org/10.3389/fpsyg.2018.02701>.
- Plate, Daniel, and James Hutson. 2025. Large Language Models as Machines of Beauty: Cognitive Averaging Hutson, Latent Space Geometry, and the Entropic Foundations of Aesthetic Preference. *ISAR Journal of Arts, Humanities and Social Sciences* 5: 55–72. <https://doi.org/10.5281/zenodo.14656205>.
- Sachs, Matthew E., Assal Habibi, Antonio Damasio, and Jonas T. Kaplan. 2020. Dynamic intersubject neural synchronization reflects affective responses to sad music. *NeuroImage* 211: 116512. <https://doi.org/10.1016/j.neuroimage.2019.116512>.
- Vessel, Edward A., and Nava Rubin. 2010. Beauty and the Beholder. *Journal of Vision* 10: 1–14. <https://doi.org/10.1167/10.12.14>.
- Vessel, Edward A., G. Gabrielle Starr, and Nava Rubin. 2012. The Brain on Art: Intense Aesthetic Experience Activates the Default Mode Network. *Frontiers in Human Neuroscience* 6: 66. <https://doi.org/10.3389/fnhum.2012.00066>.
- Yang, Taoxi, Sarita Silveira, Arusu Formuli, Marco Paolini, Ernst Pöppel, Tilmann Sander, and Yan Bao. 2019. Aesthetic Experiences Across Cultures: Neural Correlates When Viewing Traditional Eastern or Western Landscape Paintings. *Frontiers in Psychology* 10: 798. <https://doi.org/10.3389/fpsyg.2019.00798>.
- Yi, Xianqun. 2026. From Alignment to Evocation. *Preprints*. <https://doi.org/10.20944/preprints202606.1630.v1>.

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.