UC Riverside
Conference Papers

Title
Beyond AI-Generated Music: Rethinking Artificial Intelligence as a Co-Creative Partner in the Composer's Workflow

Permalink
https://escholarship.org/uc/item/09535009

Author
Cao, Wanying

Publication Date
2026-06-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

# Beyond AI-Generated Music: Rethinking Artificial Intelligence as a Co-Creative Partner in the Composer's Workflow

Wanying Cao
Department of Music, University of California, Riverside
wcao030@ucr.edu

## Abstract

Current discussions of artificial intelligence in music often focus on AI-generated artifacts and whether machines can compose autonomously. This paper shifts attention from finished outputs to the broader creative workflow, asking how generative AI can participate in audiovisual composition without directly generating the final score. Through two practice-based case studies, this paper examines AI across conceptual development, prompt mediation, material generation, research, evaluation, and reflection. In the first case, conversational AI supported concept and prompt development, while an image-to-video system produced visual material that required substantial human editing before independently composed music was added. The second case examines a Dunhuang-inspired dance film in which AI supported cultural research, theoretical inquiry, and audiovisual ideation, while verification, composition, orchestration, synchronization, and aesthetic judgment remained human-directed. Drawing on human–AI co-creativity research, Vilém Flusser's theory of technical apparatuses and play, and Michel Chion's and Nicholas Cook's theories of audiovisual relations, the paper proposes a shifting-role co-creative workflow. It argues that AI can shape composition before a single note is written, while artistic judgment, cultural responsibility, and final creative accountability remain human.

**Keywords**: human–AI co-creativity; audiovisual composition; creative workflow; human agency; generative AI

## 1. Introduction

Much of the public and scholarly discussion surrounding artificial intelligence and music begins with the finished artifact. The recurring questions are familiar: Can a machine compose? Can an AI-generated piece be judged as creative? At what point might a computational system be considered an author rather than a tool? These questions have shaped a substantial body of work in computational creativity and music generation, where the production and evaluation of musical outputs remain central concerns (Carnovalini & Rodà, 2020, pp. 1–2). Yet they describe only part of what composers actually do. Composition also begins in activities that may leave no direct trace in the final score: gathering references, testing associations, framing an aesthetic problem, listening comparatively, rejecting materials, revising a concept, and deciding what a work should not become.

This paper shifts the unit of analysis from the AI-generated artifact to the creative workflow. It asks what changes when generative AI enters those preparatory, exploratory, and evaluative stages, even when no AI-generated music appears in the finished work. Such a shift does not require treating the machine as an autonomous composer. Earlier work on creativity support already distinguished between computational systems that organize or stimulate human creative activity and systems that contribute directly to the creative act (Lubart, 2005, pp. 366–368). More recent research on human-computer co-creativity has made the relation more explicit, describing situations in which human and computational agents alternate, divide

tasks, or contribute differently to a shared creative process (Kantosalo & Toivonen, 2016, pp. 77–78). What matters in these accounts is not simply whether a machine produces material, but how initiative, evaluation, and creative responsibility are distributed.

The question is especially relevant to audiovisual composition. A composer working with moving images is rarely responding to sound alone. Decisions are made through a network of visual, narrative, cultural, and technical constraints. Michel Chion's account of "added value" demonstrates how sound can alter the perceived meaning, temporality, and materiality of an image (Chion, 2019, p. 5), while Nicholas Cook's analysis of musical multimedia treats meaning as something produced through relations among media rather than contained in any one component (Cook, 1998, pp. 98–99). In this context, an AI system may affect composition without ever generating a note: it may alter how a visual sequence is interpreted, suggest an unexpected conceptual connection, mediate between verbal intention and a generative image system, or produce material that subsequently changes the composer's musical response.

Two practice-based case studies provide the basis for this inquiry. The first follows an experimental audiovisual study in which a conversational AI was used to develop a concept and translate visual intentions into prompts for an image-to-video system. The generated outputs did not fully realize the intended transformations, and manual editing became necessary before the soundtrack was composed independently. The second examines a Dunhuang-inspired dance film in which generative AI was used primarily for cultural research, conceptual exploration, and the articulation of sound-image questions; factual verification, musical composition, orchestration, synchronization, and final aesthetic judgment remained human-directed.

Accordingly, this study asks how generative AI can function as a co-creative partner across an audiovisual composer's workflow without replacing human artistic agency or directly generating the final musical work. It considers the changing roles AI may assume across research, conceptualization, prompt mediation, material generation, evaluation, and reflection; the redistribution of agency and control when machine outputs depart from artistic intention; the forms of judgment that remain dependent on the human composer; and, ultimately, whether these interactions are better understood through a shifting-role model of co-creation than through a binary distinction between "human-created" and "AI-generated" music.

Taken together, the cases suggest that "AI-assisted" and "AI-generated" should not be treated as interchangeable categories. The argument advanced here is that creative agency can shift across stages of a workflow without being transferred wholesale from composer to machine. AI may expand a field of possibilities, introduce productive uncertainty, or redirect attention, while human judgment continues to determine what is credible, culturally appropriate, musically coherent, and artistically necessary. The paper therefore proposes a shifting-role model of co-creation in which the relevant question is not simply whether AI can compose, but how the composer's role changes when research, ideation, generation, evaluation, and revision are distributed across human and computational actors.

## 2. From Generation to Co-Creation: Theoretical Framework

The distinction between generation and co-creation matters because they describe different objects of study. Computational creativity has often focused on whether systems can produce artifacts or behavior that would be judged creative. In music, this has produced extensive work on algorithmic composition, musical metacreation, and music-generation systems.

Carnovalini and Rodà note, however, that computational creativity may also aim to model human creativity or support artists in expressing their own creativity, not only to produce autonomous machine creativity (Carnovalini & Rodà, 2020, pp. 1–3). This broader view already weakens any simple equation between creativity and autonomous generation.

Todd Lubart's earlier classification of computers as creative partners makes a similar distinction. His categories range from management and communication functions to creativity enhancement and integrated human-computer cooperation (Lubart, 2005, p. 368). The value of this taxonomy for the present study lies less in its specific labels than in its separation of contribution from finished output. Nicholas Davis develops this process-oriented view more explicitly through human-computer co-creativity, treating interaction itself as part of the creative problem rather than judging the system only by what it produces (Davis, 2013, pp. 9–10). A computational system may therefore matter because it changes how a work develops, even when it does not generate the final artifact.

Kantosalo and Toivonen sharpen this distinction through alternating and task-divided co-creativity. In the former, human and computational agents take turns modifying a shared artifact or concept; in the latter, they perform different creative subtasks (Kantosalo & Toivonen, 2016, pp. 77–78). Their framework is particularly useful for workflows in which an AI system changes function over time. A conversational model may organize references, suggest associations, or support evaluation without participating directly in musical writing. Co-creativity, in this sense, is better understood by tracing what contribution occurs at a given stage and who retains the authority to evaluate it.

Rezwana and Maher's Co-Creative Framework for Interaction Design (COFI) extends this emphasis by foregrounding participation style, contribution type, communication, and turn-taking rather than AI capability alone (Rezwana & Maher, 2023, pp. 67:1–67:3). This distinction is important because a sequence of prompts and responses does not automatically constitute meaningful collaboration. A sophisticated output may have little effect on a creator's decisions, while a modest suggestion may substantially redirect them. The present study therefore focuses on interaction trajectories: who initiates a move, how that move changes the next decision, and whether the result is accepted, revised, ignored, or resisted.

Large language models complicate this picture further because their influence may occur before a recognizable artifact exists. Wan and colleagues treat LLM-supported prewriting as a co-creative process in which initiative shifts between human and model across ideation, illumination, and implementation (Wan et al., 2024, pp. 84:2–84:3). Participants also verified outputs and integrated only what they considered useful (Wan et al., 2024, pp. 84:15, 84:19–84:20). Their study further suggests that imperfect or unexpected responses can become productive by opening directions that were not initially intended.

This paper extends that observation to what may be called pre-compositional co-creativity. Before a note is written, a project may already have been shaped through research, reference gathering, verbal description, audiovisual analysis, and the testing of aesthetic possibilities. In such cases, AI influence is indirect but consequential: a conversation reframes the artistic question; a suggested reference leads to verification and then to a timbral decision; or an unsuccessful visual generation prompts manual intervention that changes the later soundtrack. The creative consequence lies in the altered decision space.

For audiovisual composition, this relational emphasis also aligns with established sound-image theory. Chion's concept of added value explains how sound can shape the perception of an image as though the resulting expressive quality had already been visually present

(Chion, 2019, p. 5). Cook likewise treats multimedia meaning as emerging through changing relations among media rather than through simple illustration (Cook, 1998, pp. 98–99). These theories do not explain human-AI co-creativity, but they clarify why changes in visual interpretation can lead to changes in musical form, timbre, synchronization, and expectation.

The framework adopted here therefore treats co-creation as a shifting distribution of roles rather than a binary choice between human authorship and machine autonomy. Contribution and control are not identical; initiative can move between participants; and breakdown can become part of the creative process when technical limits, uncertain information, or irrelevant suggestions require rejection, verification, or manual reconstruction. The case studies that follow examine where AI enters an audiovisual composer's workflow, what role it assumes at each stage, and where human judgment remains decisive.

## 3. Playing with the Apparatus: Flusser and Creative Resistance

Human–AI co-creation can be described in terms of roles, initiative, and division of labor, but these categories do not fully account for another recurring feature of creative interaction: the system does not always do what the artist intends. Generative outputs may be technically limited, aesthetically unconvincing, factually uncertain, or simply different from what a prompt appears to request. Such moments are often discussed as failures of the technology. In creative practice, however, they can also change the direction of the work. Vilém Flusser's account of technical apparatuses offers a useful framework for understanding this tension between intention and system behavior.

Flusser wrote *Into the Universe of Technical Images* decades before contemporary generative AI, and his apparatus should therefore not be treated as a direct theoretical description of large language models or generative media systems. Nevertheless, his account of creation as the production of new information through recombination is strikingly relevant to the interaction examined here. In "To Play," Flusser rejects the myth of creation *ex nihilo* and instead describes new information as emerging through the synthesis of information that already exists (Flusser, 2011, pp. 87–89). Creativity, in this formulation, is less an act of producing something from nothing than a process of entering a field of possibilities and searching for improbable combinations.

This emphasis on possibility also changes the relationship between artist and machine. Elsewhere in the book, Flusser describes the producer of technical images as working from within an apparatus whose internal operations are not fully accessible to the user. The apparatus automatically performs operations that the user neither constructs nor necessarily understands; the creative gesture lies partly in attempting to draw an improbable result from the possibilities permitted by its program (Flusser, 2011, pp. 34–38). The resulting relationship is therefore neither one of complete authorship nor one of passive automation. The user sets intentions and makes selections, but those intentions encounter a programmed space with its own boundaries.

This tension becomes particularly important when an apparatus resists the desired result. Flusser associates a new form of human freedom with the ability to program an apparatus toward a desired form and to stop the process when that form appears. At the same time, he warns that automated systems can exceed the user's capacity to control their possible outcomes, allowing the logic of the program to move away from the intention that initiated it (Flusser, 2011, pp. 73–74). Although contemporary generative systems differ substantially from the technologies Flusser addressed, the tension he identifies remains analytically

productive: creative control is exercised through an apparatus but is never identical with control over all of its operations.

This paper uses the term creative resistance to describe what occurs when that discrepancy becomes productive. Creative resistance is not a concept proposed by Flusser, nor does it imply that the computational system possesses intention. It refers instead to moments in which the difference between an artist's requested outcome and a system's actual output requires a consequential human response. The response may involve revising a prompt, rejecting a generation, manually reconstructing material, verifying information through external sources, or changing the original concept itself. In each case, the system's inadequacy becomes part of the causal history of the artwork.

Flusser's discussion of creation also helps distinguish distributed production from the disappearance of human judgment. In "To Create," he questions the traditional image of the isolated author and describes information increasingly as the product of dialogue and recombination rather than individual origination (Flusser, 2011, pp. 95–103). This offers a useful way to understand AI-assisted workflows without concluding that every participant possesses equivalent authority. A work may emerge from interactions among artists, software, archives, generative systems, and existing cultural materials while still depending on human decisions about what should be retained, corrected, contextualized, or discarded.

Within the framework of this study, playing with the apparatus therefore involves both exploration and resistance. Generative AI can enlarge a composer's field of possible decisions, but its creative significance is not limited to successful outputs. Its errors, constraints, and unexpected responses may also redirect the process. This perspective provides the bridge between the interaction models discussed above and the two practice-based cases that follow: in both, creative agency becomes most visible not when the machine perfectly fulfills an instruction, but when the human creator must decide what to do when it does not.

## 4. Methodology: Practice-Based Comparative Case Study

This study adopts a **practice-based comparative case-study methodology** to examine how generative AI enters different stages of an audiovisual composer's creative workflow. Practice-based research is particularly appropriate to this inquiry because the creative process and its resulting artifacts are not treated merely as illustrations of an external theory; they constitute part of the material through which knowledge about the process is produced. Linda Candy and Ernest Edmonds define practice-based research as original investigation in which new knowledge is gained partly through practice and its outcomes, while emphasizing that practice and research remain distinct but interdependent activities (Candy & Edmonds, 2018, pp. 63–64). The purpose of the present study is therefore not simply to document the use of AI in two projects, but to subject those creative processes to comparative and theoretically informed reflection.

The research examines two cases created within the same compositional practice but involving substantially different forms of AI participation. Case I is an experimental audiovisual study in which conversational AI contributed to conceptual development and prompt formulation, followed by AI-based visual generation and extensive human editing and music composition. Case II is a Dunhuang-inspired audiovisual dance film in which AI participated primarily in cultural research, theoretical exploration, and early conceptual development rather than in the generation of the final musical materials.

The materials considered include project briefs, conversational AI exchanges and prompt histories where available, generated and rejected audiovisual outputs, editing decisions, compositional drafts, digital audio workstation sessions, completed audiovisual materials, and contemporaneous or retrospective reflections on the decision-making process. In this respect, documentation functions not as a substitute for practice but as a means of making tacit creative decisions available for critical examination. Robin Nelson's model of Practice as Research similarly places practical "know-how" in dialogue with critical reflection and conceptual knowledge, enabling artistic activity to become the subject of explicit research inquiry (Nelson, 2013, p. 37).

Comparison across the two cases is organized through five analytical questions. First, the **role** asks what function AI performs at a given stage—for example, research assistant, dialogic partner, prompt mediator, material generator, or reflective aid. Second, **initiative** considers whether a particular creative move originates with the human creator, the computational system, or an iterative exchange between them. Third, **control** examines who evaluates, selects, rejects, or modifies the resulting material. Fourth, **creative consequence** asks whether an AI contribution materially changes a later artistic decision rather than merely appearing within the workflow. Finally, **resistance** identifies moments when technical limitations, unreliable information, or unsatisfactory outputs require human correction, verification, or conceptual revision.

The comparison is qualitative and interpretive rather than statistical. Its purpose is not to claim that two cases represent all AI-assisted composition practices, but to use close examination of contrasting creative situations to identify transferable analytical distinctions. Case-study research is particularly useful for producing this kind of context-dependent knowledge and for examining processes whose significance may be lost when separated from the circumstances in which decisions occur (Flyvbjerg, 2006, pp. 221–223).

No audience survey or experimental participant data are used in this study. Claims concerning possible listener or viewer responses are therefore treated as compositional and perceptual questions, not as empirically demonstrated audience effects. Likewise, collaborators are discussed only insofar as necessary to explain the creative workflow; private communications with collaborators are not analyzed as research data. This methodological boundary allows the study to focus specifically on the distribution of role, initiative, control, judgment, and responsibility between the composer and AI systems.

## 5. Case Study I: AI as Generator, Mediator, and Resistant Apparatus

Case I originated as a course-based experimental audiovisual composition exercise in which the composer-researcher was asked to create a short work using artificial intelligence as an active participant in the creative process. The exercise focused on the relations among presence, absence, randomness, and pattern, while also asking how sound and image might interact through synchronization, divergence, layering, and transformation. Within these parameters, the composer-researcher chose to construct a gradual audiovisual movement from recognizable visual stability toward instability, fragmentation, and disappearance.

The resulting study combined four distinct stages of production. A conversational AI system was first used to help translate the visual concept into prompting language. Runway was then selected as an image-to-video generator, producing six short transformations from a single source image. Because these generations did not sufficiently realize several aspects of the intended visual trajectory, particularly severe fragmentation, shaking movement, and

progressive disintegration, the clips were subsequently assembled and reworked manually in CapCut. The completed visual sequence was then imported into Logic Pro, where the soundtrack was composed by the composer-researcher using synthesizers, loop-based sonic materials, and manually constructed sound–image relationships.

The case is therefore not presented simply as an example of AI-generated video. Its analytical value lies in the sequence through which the work developed:

**human intention → AI-mediated prompting → machine generation → human evaluation → manual intervention → human composition.**

Within this relatively compact experiment, the function of AI changed repeatedly, while human control did not remain constant either. The case makes these shifts unusually visible.

## 5.1. Human Artistic Intention

Before either AI system entered the process, the composer-researcher had already established the conceptual trajectory of the work:

**human control → AI intervention → instability → fragmentation → disappearance.**

A single source image was selected as the visual point of departure so that successive changes could be perceived as transformations of the same material rather than as unrelated images. The desired emotional development followed a similar trajectory. The opening would retain a comparatively stable and recognizable visual state; later stages would gradually weaken that stability until the image approached abstraction and disappearance.

This sequence matters because the generative systems did not originate from the artistic problem. The source material, the overall direction of transformation, and the desired progression had already been established. AI entered an existing compositional intention and was asked to participate in realizing it.

## 5.2. Conversational AI as Prompt Mediator

The first AI system did not generate any of the audiovisual material that appeared in the final study. Instead, a conversational model was used to explore how the visual concept might be expressed in a form legible to an image-to-video generator.

The interaction included practical discussion of image-to-video generation and the development of six progressively changing groups of descriptive terms. Vocabulary associated with stability, instability, distortion, fragmentation, randomness, and disappearance was proposed, selected, recombined, and revised by the composer-researcher before being transferred into Runway.

What matters here is not that an AI system “wrote the prompts.” The more significant function was one of **mediation**. A visual idea existed first in perceptual and conceptual terms; it then had to be reformulated as natural language capable of affecting the behavior of another generative system. Natural language consequently became an interface between the composer-researcher’s imagination and the image generator.

This stage illustrates a form of AI participation that occurs before material generation itself. The conversational model influenced the conditions of generation without producing the visual output directly.

### 5.3. Runway as Material Generator

The revised prompting language was then applied to a single source image using Runway. Six short generations were produced, each intended to advance the image further along the predefined trajectory. These clips were not conceived as six finished audiovisual works. They functioned as intermediate visual materials that would later be selected, arranged, altered, and scored.

The generations produced varying degrees of movement, abstraction, distortion, and loss of visual continuity. Because the same source image remained perceptible across multiple stages, the sequence maintained a degree of pattern even as generative variation introduced increasing irregularity. The experiment therefore produced a visual relationship between repetition and unpredictability rather than replacing one with the other.

Table 1 records this process by comparing the intended transformation at each stage with the resulting generative output and the composer-researcher's subsequent response.

| Stage | Intended Transformation | Generative Outcome | Human Evaluation and Intervention |
|---|---|---|---|
| 1 | **Stability / recognizable presence** Maintain a clear and stable representation of the source image as the initial state. | The source image remained clearly recognizable, with relatively limited transformation and stable visual continuity. | The result was retained as the initial point of reference for the progressive transformation. |
| 2 | **Subtle instability** Introduce slight instability and movement while preserving the overall structure and recognizability. | Minor movement and visual fluctuation appeared, while the underlying image remained largely intact. | Prompt language was adjusted to encourage greater instability without immediately losing visual continuity. |
| 3 | **Increasing distortion** Increase abstraction and distortion, reducing visual stability and coherence. | The generated image became more abstract and less stable, but the structural disruption remained milder than intended. | Additional vocabulary emphasizing distortion, irregularity, and unstable motion was introduced in subsequent prompts. |
| 4 | **Fragmentation / tearing** Break the image into fragments and introduce tearing or structural rupture. | The system introduced further visual variation but did not produce the intended degree of tearing or fragmentation. | The output was retained as generative source material rather than accepted as a finished realization of the concept. |
| 5 | **Strong instability / shaking movement** Create strong instability and pronounced shaking or erratic movement. | Generated motion suggested instability but did not achieve the intensity or camera-shaking quality originally envisioned. | Additional shaking, irregular movement, and temporal disruption were created manually during editing in CapCut. |
| 6 | **Dissolution / reduced recognizability** Move toward dissolution, with the image losing recognizable form and coherence. | The final AI-generated variation increased abstraction and weakened the visual stability of the original image, but did not independently complete the intended movement toward disappearance. | The generated clips were reordered, layered, and further manipulated through manual editing to construct the final visual trajectory before music composition. |

Table 1. Intended Transformations, Generative Outcomes, and Human Interventions in Case I

*Source: Compiled by the researcher from prompt iterations, Runway-generated outputs, and audiovisual editing records from Case I.*

At this stage, AI had shifted from mediator to material generator. Yet the production of material did not transfer evaluative authority to the system. As Kantosalo and Toivonen note, generation and evaluation may be distributed across different participants within a co-creative process; the production of alternatives does not itself determine which alternatives should be retained (Kantosalo & Toivonen, 2016, p. 83).

## 5.4. Resistance and Manual Intervention

The most analytically significant moment occurred when the generated material failed to coincide with the intended visual development.

The issue was not that the clips were entirely unusable. Several contained movements, textures, or distortions that could contribute to the final work. **The problem was more specific: a persistent mismatch emerged between artistic intention and generative realization.**

Effects involving pronounced tearing, forceful shaking, unstable camera-like movement, and progressive visual disintegration remained more limited or qualitatively different from what had been envisioned.

The generated outputs were therefore not accepted as finished results. Instead, the six clips were imported into CapCut and treated as source material for further construction. Shaking, irregular movement, temporal disruption, and other destabilizing effects were added manually. Clips were reordered and adjusted so that the overall sequence more closely approached the intended progression from stability toward fragmentation and disappearance.

This moment is central to the concept of **creative resistance** developed earlier in this paper. The generative apparatus did not simply execute an artistic instruction, but neither did its limitation end the collaboration. The limitation itself created the next creative problem.

From the perspective of co-creative interaction, control consequently shifted again. Rezwana and Maher's COFI framework emphasizes that participation, initiative, contribution, and interaction can change dynamically within human–AI collaboration rather than remaining assigned permanently to one participant (Rezwana & Maher, 2023, pp. 67:2, 67:8–67:9). In Case I, the machine generated possibilities, but human dissatisfaction with those possibilities became the basis for renewed intervention.

The failure of the generative system therefore did not terminate collaboration; **it redistributed control**.

Flusser's concept of the apparatus provides another useful perspective. The creator operates through a programmed system whose available possibilities cannot be reduced to the creator's intention. Creative practice consequently involves negotiating, testing, and sometimes working against what the apparatus makes possible (Flusser, 2011, pp. 34–38, 73–74). The manual reconstruction undertaken in Case I was not external to the AI-assisted process. It emerged because of that encounter with the limits of the apparatus.

## 5.5. Human Composition after AI Visual Generation

Only after the visual sequence had been assembled and manually revised was it imported into Logic Pro for scoring. The music was composed by the composer-researcher using synthesizers, loop-based materials, layered textures, and manually determined points of synchronization between sonic and visual events. Changes in visual density, movement, and instability informed decisions concerning musical pacing, texture, repetition, and fragmentation.

No generative music system composed the soundtrack. Yet describing the musical component as entirely unaffected by AI would also be inaccurate. The composer was scoring a visual

object whose form had emerged through dialogue with one AI system, generation by another, and subsequent human responses to the limitations of those systems.

The distinction is important. **AI did not generate the musical material, but it helped construct the audiovisual conditions to which the music responded.**

Case I therefore demonstrates why the binary distinction between "human-made" and "AI-generated" material is insufficient for describing the workflow examined in this study. The composer-researcher remained responsible for the original concept, the evaluation of outputs, manual editing, musical composition, and final structure. At the same time, the generative systems introduced materials and constraints that materially altered what happened next.

Within a single experiment, AI consequently moved from **prompt mediator**, to **material generator**, to **resistant apparatus**, while the human participant moved from **initiator**, to **evaluator**, to **editor**, and finally to **composer**. The significance of the case lies precisely in these changes of role.

## 6. Case Study II: AI Before the First Note

The second case concerns a Dunhuang-inspired audiovisual dance film. In contrast to Case I, generative AI did not produce the principal audiovisual materials of this project. Its participation occurred primarily before and around the act of composition: during research, conceptual development, theoretical questioning, and reflection. The case therefore offers a different test of the central argument of this study. If AI does not generate the score, can it nevertheless participate meaningfully in the conditions through which a score is conceived?

### 6.1. Dunhuang and the Artistic Question

The project takes Dunhuang as both a visual reference and a field of cultural memory. Its desert landscape, Buddhist mural traditions, painted figures, and representations of movement provided the visual and conceptual environment for a contemporary dance film. Rather than attempting to reconstruct a historically authentic Dunhuang sound world, the musical approach treated this material as the starting point for a contemporary audiovisual interpretation.

The compositional process began with a question formulated before AI entered the workflow:

**What would it sound like if the figures within the Dunhuang murals came alive?**

The question was deliberately open. It did not specify instrumentation, musical style, harmony, or historical period. Instead, it established a relationship between stillness and movement, historical image and contemporary body, and visual memory and imagined sound. As in Case I, the initial artistic problem was therefore human-defined. AI entered only after the conceptual direction had been established.

### 6.2. AI as Research Expander

A conversational AI system was used during the early research stage to expand possible lines of inquiry. Dialogues addressed Dunhuang music history, Silk Road cultural exchange, Buddhist performance traditions, historical instruments, and contemporary interpretations of Dunhuang-related aesthetics. The value of these exchanges was not that the system supplied a

coherent historical account. Rather, it produced associations, terms, and possible research directions that the composer-researcher might not otherwise have pursued immediately.

This distinction was methodologically important. **AI-generated statements were treated as hypotheses and research prompts rather than evidence.** Historical or cultural claims that appeared potentially relevant were subsequently checked against academic literature, recordings, museum materials, archival resources, and other historical sources. Suggestions that could not be adequately supported, or that proved irrelevant to the artistic problem, were discarded.

**The process can therefore be understood as a form of epistemic responsibility within human–AI co-creation.** Human agency was exercised not only through artistic selection but also through determining what could legitimately be accepted as knowledge. The system accelerated discovery, but the responsibility for verification remained with the researcher. In this case, resistance did not primarily take the technical form encountered in Case I. It appeared instead as epistemic uncertainty: an AI response could be suggestive enough to redirect attention while remaining too unreliable to serve as a factual foundation.

### 6.3. AI as Theoretical Interlocutor: Chion

The conversational system also became useful in articulating questions about the relationship between sound and image. These questions were subsequently examined through established audiovisual theory rather than answered by the AI itself. Michel Chion's distinction among causal, codal, and reduced listening provides one useful entry point. Causal listening attends to the perceived source of a sound; codal listening interprets sound through codes or language; reduced listening directs attention toward sonic qualities such as timbre, texture, and pitch independent of source or meaning (Chion, 2019, pp. 22–29).

For the composer-researcher, however, Chion's concept of added value became particularly relevant. Chion argues that sound can contribute expressive or informational qualities to an image in such a way that the viewer retrospectively experiences those qualities as if they had already belonged to the image (Chion, 2019, p. 5). This raised a concrete compositional problem in the film's opening.

The early visual sequence is dominated by gray and earth-yellow tones and presents the dancer in an environment associated with the mural imagery. One possible reading of these images is restraint, distance, or even oppression. Yet the compositional question was whether such an effect resides in the image itself, or whether the soundtrack can produce an added value that makes the visual material seem retrospectively more oppressive than it would appear in silence or with another score.

This distinction affected musical decisions. The opening does not immediately settle into a conventionally tonal or overtly historical musical language. Instead, Chinese-associated timbral references are placed alongside electronic and experimental textures. The purpose was not to use theory as a formula for scoring, but to remain attentive to how sound might reorganize the perception of the image. Conversational AI helped formulate the question; Chion supplied the conceptual vocabulary through which the question could be examined.

### 6.4. Cook and Cultural Expectation

Nicholas Cook's account of multimedia relations offered a second way of approaching the same material. Cook distinguishes among conformance, complementation, and contest: media may reinforce a similar meaning, contribute different but mutually productive information, or establish a more resistant and contradictory relationship (Cook, 1998, pp. 98–100). The categories are useful here precisely because they do not assume that music should simply illustrate what is visible.

A second compositional question therefore concerned cultural expectation. If an image presents Dunhuang-associated murals, muted desert colors, and a dancer apparently emerging from this visual world, does that imagery encourage an expectation of a tonal, traditional Chinese, or otherwise culturally recognizable musical language? Because the present study contains no audience experiment, this is not advanced as a claim about what viewers in fact expect. It is instead treated as a compositional problem concerning the possible relationship between visual coding and musical convention.

The experimental language of the opening may consequently be understood as moving between complementation and contest. The music shares the atmosphere and cultural field of the image without simply reproducing its most obvious stylistic associations. As the visual world becomes more colorful and the score later moves toward more recognizable tonal and Chinese-associated materials, the audiovisual relationship can move closer to conformance. The usefulness of Cook's framework lies not in assigning each passage permanently to a category, but in recognizing that the relationship between media may change over time.

Again, **AI did not determine which relationship was preferable. Its contribution was to help articulate a question that could then be evaluated through compositional practice and audiovisual theory.**

### 6.5. Human Musical Judgment

Once musical sketching began, the process became predominantly human-directed. Decisions about Chinese instrumental timbres, synthesizers, electronic textures, orchestration, pacing, and synchronization responded to the evolving image, movement, and narrative structure. Visual and musical revisions therefore remained mutually influential.

The central distinction of Case II lies in judgment. AI could suggest historical connections, conceptual directions, and questions about audiovisual meaning, but it could not determine whether a timbre was culturally appropriate, an experimental texture productively ambiguous, a synchronization overly literal, or a tonal transition emotionally effective. Such decisions depended on situated artistic judgment.

AI thus shaped the conditions of composition rather than composing the music itself. Its strongest contribution occurred before and around musical writing—through research, association, questioning, and reflection—while verification, interpretation, musical construction, and final aesthetic decisions remained human-directed. **Case II therefore demonstrates how pre-compositional co-creativity can become compositionally consequential without producing any AI-generated notes.**

# 7. Cross-Case Discussion: Where Is the Composer?

The two cases differ substantially in medium, scale, and mode of AI participation. In Case I, generative systems entered directly into the production of visual material. In Case II, AI remained largely outside the finished audiovisual artifact, participating instead in research, conceptual development, and reflection. Considered together, however, the cases suggest that asking whether AI "created" a work is too coarse a question to describe what occurred. **A more useful inquiry concerns where creative roles, initiative, evaluation, and responsibility were located at different moments in the process.**

## 7.1. One AI Ecosystem, Multiple Roles

Across the two cases, AI did not occupy a single position that could be consistently labeled assistant, generator, or collaborator. Its function changed according to the stage of the project and the problem being addressed.

| Role | Function within the Cases |
|---|---|
| **Research Assistant** | Expands and organizes possible lines of inquiry |
| **Dialogic Partner** | Produces questions, associations, and responses that sustain exploration |
| **Conceptual Provocateur** | Introduces possibilities not initially considered by the composer-researcher |
| **Prompt Mediator** | Translates artistic concepts into language intended for another generative system |
| **Material Generator** | Produces audiovisual source material that enters the developing work |
| **Resistant Apparatus** | Produces insufficient, unintended, or difficult-to-use results that demand further decisions |
| **Reflective Partner** | Assists in articulating and reconsidering decisions after or between creative stages |

Table 2. Shifting AI roles observed across the two case studies.

*Source: Cross-case synthesis by the researcher.*

These roles should not be understood as a new fixed taxonomy. Kantosalo and Toivonen distinguish alternating from task-divided co-creativity and identify functions such as concept definition, generation, and evaluation, while also acknowledging that actual co-creative systems rarely conform neatly to rigid categories (Kantosalo & Toivonen, 2016, p. 80). The present cases extend that observation from the level of system classification to the temporal structure of a single creative process. A system may function as an assistant at one moment, a generator at another, and a source of resistance shortly afterward.

COFI (the Co-Creative Framework for the Interaction design) likewise shifts attention from isolated computational capability toward interaction: participation style, distribution of tasks, timing of initiative, communication, and contributions to the shared product all affect the character of co-creation (Rezwana & Maher, 2023, pp. 67:8–67:11). The cases examined here suggest that these relations can change repeatedly within one workflow. In Case I, initiative moved from human conceptualization to conversational mediation, generative production, human evaluation, manual intervention, and finally human musical composition. In Case II, AI expanded research possibilities but ceded epistemic and aesthetic decisions to the composer-researcher. What remained consistent was not the role of AI, but the movement of roles.

## 7.2. Technical Limitation and Epistemic Uncertainty

The two cases reveal different forms of AI limitation. In Case I, the problem was primarily technical: the image-to-video system could not produce the required degree of tearing, instability, and fragmentation. Rather than abandoning its outputs, the composer retained them as source material and manually reconstructed their movement and organization. This intervention demonstrates what this study terms technical agency: the capacity to act creatively where a system cannot adequately realize an artistic intention.

Case II required a different response. Conversational AI could rapidly propose historical and cultural associations, but relevance did not guarantee accuracy. Claims concerning Dunhuang, historical instruments, religious performance, and cultural exchange therefore required independent verification. This constitutes epistemic agency: the responsibility to distinguish generative suggestion from knowledge that can legitimately inform research and artistic interpretation.

These cases show why “AI failure” is too broad a category. Technical inadequacy demands practical intervention; epistemic uncertainty demands evidentiary judgment. Neither necessarily ends collaboration. An incomplete or unreliable output may instead provoke correction, resistance, redirection, or verification. The relevant analytical question is therefore not simply whether an AI output succeeds, but what it causes the human participant to do next.

## 7.3. Distributed Agency, Unequal Responsibility

Neither case supports an image of the composer as the isolated source of every creative decision. The works emerged through interactions among collaborators, source materials, software, generative systems, audiovisual structures, and compositional choices. This recalls

Flusser's movement from the isolated creator toward dialogic and networked processes in which new information emerges through relations among human and technical participants (Flusser, 2011, pp. 95–103).

Distributed agency, however, does not imply equal responsibility. AI could introduce an association, unexpected image, or new direction of inquiry, but it could not assume responsibility for the reliability of a historical claim, the cultural implications of an association, or the suitability of a musical decision. Those judgments remained human.

The distinction is therefore crucial: creative agency may be distributed, while artistic responsibility remains asymmetrically human. As AI participation expands, the composer's role shifts toward defining problems, evaluating outputs, verifying claims, negotiating limitations, integrating materials, and taking responsibility for the finished work.

The composer is thus most clearly present at moments when possibilities become commitments: when one result is chosen over another, information is accepted or rejected, and technical possibilities are given cultural, musical, and audiovisual meaning.

## 8. Proposed Model: The Shifting-Role Co-Creative Workflow

The comparison of the two cases suggests that AI-assisted audiovisual composition is difficult to describe as a fixed division of labor between a human creator and an artificial collaborator. Roles change as the work develops; initiative may move between participants; and moments of failure or uncertainty can alter the direction of the process. To account for these movements, this study proposes the **Shifting-Role Co-Creative Workflow (SRCW),** a practice-derived analytical model for AI-assisted audiovisual composition.

The SRCW is not intended as a universal theory or a prescriptive sequence that every AI-assisted project should follow. Rather, it provides a vocabulary for tracing how creative roles and forms of responsibility may shift across different stages of a compositional process. Its emphasis on changing roles develops from existing accounts of alternating and task-divided co-creativity, while extending those distinctions toward a more explicitly temporal model in which the same system may perform different functions within a single project (Kantosalo & Toivonen, 2016, pp. 77–80).

The workflow consists of eight phases. **Phase 1, Artistic Intention**, is human-led: the artist defines the initial problem, conceptual trajectory, or aesthetic question. **Phase 2, Dialogic Expansion**, introduces reciprocal interaction between human and AI through questions, references, associations, and alternative possibilities. **Phase 3, Verification**, returns authority more decisively to the human participant, who evaluates factual, cultural, conceptual, or aesthetic claims before allowing them to influence the work. **Phase 4, Translation and Generation**, involves the transformation of artistic intentions into prompts, media operations, or generative alternatives through human–AI exchange.

At **Phase 5, Resistance / Breakdown**, the system ceases to function as a transparent extension of intention. Outputs may be technically inadequate, unexpected, culturally reductive, unverifiable, or simply aesthetically unconvincing. Such moments recall Flusser's account of the artist working within the possibilities and constraints of a programmed apparatus rather than exercising complete control over its operations (Flusser, 2011, pp. 34–38, 73–74). Resistance therefore becomes analytically significant because it reveals the limits of the system and makes subsequent human decisions especially visible.

**Phase 6, Selection and Intervention**, is again predominantly human-led. Generated materials may be rejected, edited, reorganized, manually reconstructed, or subjected to further research.

**Phase 7, Composition**, refers to the stage in which musical form, orchestration, timbre, timing, synchronization, and other audiovisual decisions are integrated into the developing work. In the two cases examined here, this phase remained primarily human-directed even when earlier stages had been substantially shaped by AI.

Finally, **Phase 8, Reflection and Re-entry**, recognizes that the process does not necessarily end with a finished artifact. Reflection may expose a new artistic question, reveal a problem in an earlier assumption, or reopen dialogue with an AI system. Research on human–AI co-creativity has similarly emphasized interaction dynamics and shifting participation rather than a simple one-way transfer from human instruction to machine output (Rezwana & Maher, 2023, pp. 67:2–67:3). Studies of large language models in prewriting also suggest that initiative can move back and forth before an artifact is produced, with AI affecting the framing and development of ideas rather than merely generating finished content (Wan et al., 2024, pp. 84:4, 84:19–84:20).

For this reason, the SRCW should be understood as a recursive system rather than a linear pipeline. A project may move from generation back to dialogic expansion, from verification back to artistic intention, or from resistance directly into a new compositional question. Not every phase must appear with equal weight, and some may recur several times.

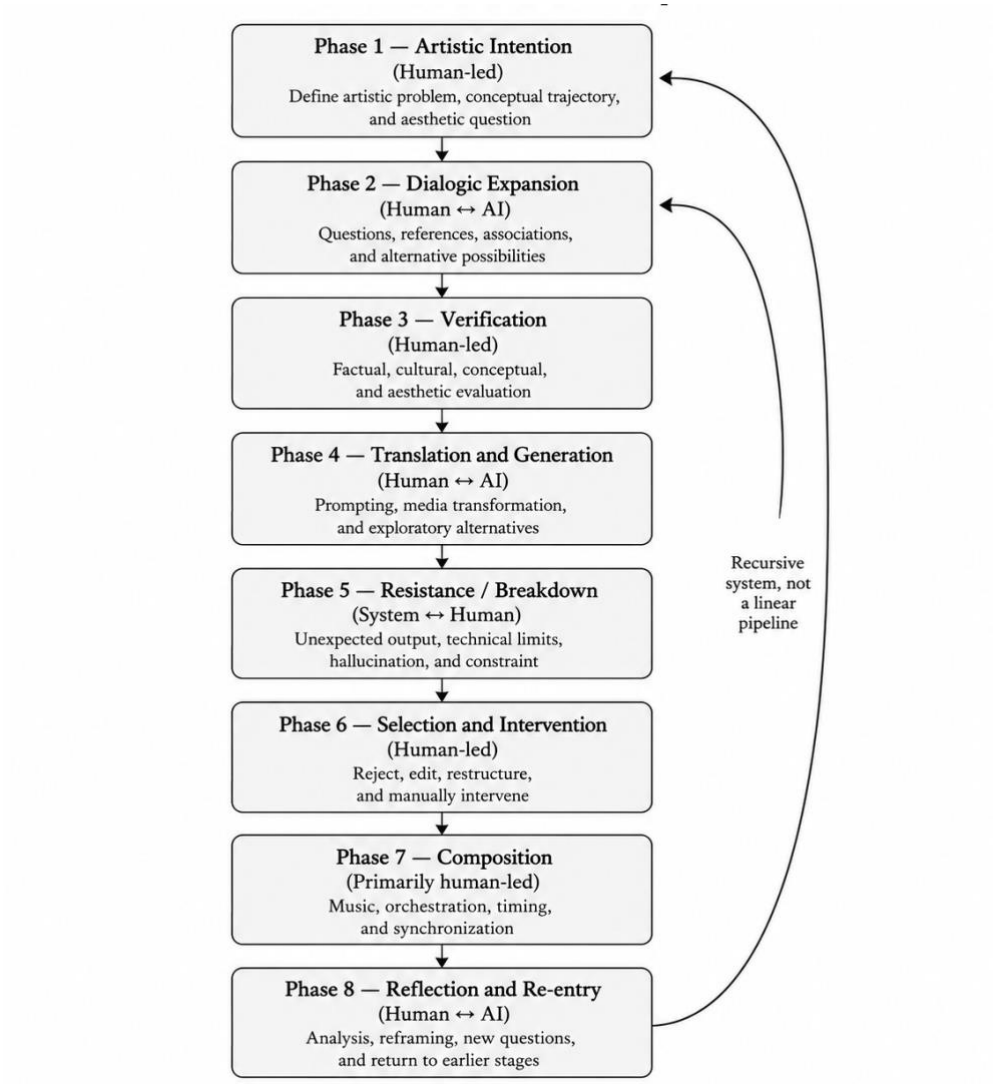



Figure 1. The Shifting-Role Co-Creative Workflow for AI-Assisted Audiovisual Composition.

*Source: Analytical model developed from the comparative case studies in this research; diagram generated with GPT based on the present paper.*

The central implication of the model is that co-creativity is located less in the presence of a particular AI-generated artifact than in the changing relations among **intention, initiative, generation, resistance, judgment, and responsibility**. AI may participate substantially before a note is composed, while the composer remains responsible for deciding which possibilities become part of the work and what those possibilities ultimately mean.

## 9. Conclusion

This study began by asking how generative AI can participate in an audiovisual composer's workflow without replacing human artistic agency or directly generating the final musical work. The two practice-based cases suggest that such participation cannot be understood adequately through the presence or absence of AI-generated musical material alone. AI may enter composition much earlier, through research, conceptual development, linguistic

mediation, visual generation, evaluation, and reflection. Its creative influence can therefore precede the production of any musical note.

The first research question concerned the roles that AI may assume within such a process. Across the two cases, these roles included research assistant, dialogic partner, conceptual provocateur, prompt mediator, material generator, resistant apparatus, and reflective partner. Crucially, these were not stable identities. A system could move between roles as the project developed. The second question concerned the distribution of agency and control. Here, the cases revealed two different forms of limitation. Technical insufficiency in Case I prompted manual intervention, while epistemic uncertainty in Case II required verification. Human agency consequently appeared not only through authorship of musical material, but through the ability to recognize when a system's contribution was inadequate and to determine what should follow.

The third question asked which aspects of audiovisual composition remained dependent on human judgment. In both cases, decisions concerning factual reliability, cultural appropriateness, aesthetic value, musical form, orchestration, synchronization, and final selection remained human-directed. This does not mean that AI had no creative consequence. Rather, it suggests a distinction between distributed creative influence and distributed responsibility. **Creative agency may be distributed, while artistic responsibility remains asymmetrically human.**

The fourth question asked whether a shifting-role model offers a more useful account of AI-assisted composition than the binary distinction between human-created and AI-generated music. The proposed **Shifting-Role Co-Creative Workflow** suggests that it does. The model treats AI-assisted creation as recursive: intention, dialogue, verification, generation, resistance, intervention, composition, and reflection may repeatedly reorganize one another rather than unfold as a fixed sequence.

The distinction between human-created and AI-generated music is therefore insufficient for describing contemporary AI-assisted creative practice. AI may affect the conceptual conditions under which music is composed even when it contributes no audible material to the finished work. The relevant question is no longer simply whether AI can compose, but how the composer's role changes when research, imagination, generation, resistance, judgment, and reflection are distributed across human and computational actors.

**AI does not need to compose the notes in order to participate in composition.**

## 10. References

Candy, L., & Edmonds, E. (2018). Practice-based research in the creative arts: Foundations and futures from the front line. Leonardo, 51(1), 63–69. https://doi.org/10.1162/LEON_a_01471

Carnovalini, F., & Rodà, A. (2020). Computational creativity and music generation systems: An introduction to the state of the art. Frontiers in Artificial Intelligence, 3, Article 14. https://doi.org/10.3389/frai.2020.00014

Chion, M. (2019). Audio-vision: Sound on screen (C. Gorbman, Ed. & Trans.; 2nd ed.). Columbia University Press. https://doi.org/10.7312/chio18588

Cook, N. (1998). Analysing musical multimedia. Clarendon Press.

Davis, N. (2013). Human-computer co-creativity: Blending human and computational creativity. Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 9(6), 9–12. https://doi.org/10.1609/aiide.v9i6.12603

Flusser, V. (2011). Into the universe of technical images (N. A. Roth, Trans.). University of Minnesota Press.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. Qualitative Inquiry, 12(2), 219–245. https://doi.org/10.1177/1077800405284363

Kantosalo, A., & Toivonen, H. (2016). Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In F. Pachet, A. Cardoso, V. Corruble, & F. Ghedini (Eds.), Proceedings of the seventh international conference on computational creativity (pp. 77–84). Sony CSL. https://www.computationalcreativity.net/iccc2016/wp-content/uploads/2016/01/Modes-for-Creative-Human-Computer-Collaboration.pdf

Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the special issue. International Journal of Human-Computer Studies, 63(4–5), 365–369. https://doi.org/10.1016/j.ijhcs.2005.04.002

Nelson, R. (2013). From practitioner to practitioner-researcher. In R. Nelson (Ed.), Practice as research in the arts: Principles, protocols, pedagogies, resistances (pp. 23–47). Palgrave Macmillan. https://doi.org/10.1057/9781137282910_2

Rezwana, J., & Maher, M. L. (2023). Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. ACM Transactions on Computer-Human Interaction, 30(5), Article 67. https://doi.org/10.1145/3519026

Wan, Q., Hu, S., Zhang, Y., Wang, P., Wen, B., & Lu, Z. (2024). “It felt like having a second mind”: Investigating human-AI co-creativity in prewriting with large language models. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), Article 84. https://doi.org/10.1145/3637361