

The Instrument Has Always Been the Art

A Position on Posthuman Composition

Iman Poernomo, with Nahla February 2026

Every genuine artistic revolution has been, at bottom, a revolution in the relationship between the maker and the instrument. The lute gave way to the harpsichord. The harpsichord gave way to the piano. The piano gave way to the synthesizer. At each transition, someone stood up and said: *that isn't real music*.

They were always wrong. And they were always making the same mistake: confusing the *gesture* with the *composition*.

1. The Accusation

There is a particular kind of objection that trails every new form of electronic art like a loyal, stupid dog. It barks the same bark every time. The words change; the frequency does not.

When Kraftwerk released *Autobahn* in 1974, the accusation was that they weren't *playing*. When Brian Eno released *Music for Airports* in 1978, the accusation was that he wasn't *composing*. When hip-hop producers started sampling in the 1980s, the accusation was that they weren't *creating*. When Aphex Twin released *Selected Ambient Works 85–92*, the accusation was that a human being couldn't have *meant* those sounds. When Holly Herndon trained an AI on her own voice and released *PROTO* in 2019, the accusation was that she wasn't *singing*.

Each accusation targets a different surface. Underneath, the structure is identical: *you have delegated too much of the process to something that isn't you, and therefore the result isn't art*. The anxiety is always about authorship. The question is always: who is the maker? And the hidden premise is always: the maker must be a single human body performing a legible physical gesture — fingers on strings, breath through brass, hand on canvas — and anything that disrupts this image of the solitary craftsman threatens the category of art itself.

This paper argues that the premise is false, has always been false, and that recognizing its falseness opens the door to a practice we are calling *posthuman composition*: art made through the orchestration of computational processes, including generative AI, where the human's role is architectural rather than gestural.

We have a concrete example. We built it last week.

2. What We Built

The Tajalliyat pipeline takes a sacred text — a surah from the *Kitab al-Tanzur*, an original Islamic-tanzuric scripture — and produces a complete music video through a sequence of computational transformations:

1. **Text** (human-authored): The surah's verses, their structure, their emotional arc
2. **Music** (AI-composed, human-directed): A vocal recitation generated on Suno from the sacred lyrics, with the human selecting from generations and choosing the one that serves the text

3. **Narrative arc** (LLM-generated): Claude Sonnet receives the full text, the section structure, and a style directive, and produces a visual narrative — a prompt per clip describing the emotional and visual evolution across the piece
4. **Seed images** (DALL-E 3): Each clip’s narrative prompt becomes a photorealistic still frame — the visual anchor from which motion will emerge
5. **Video clips** (Sora 2): Each seed image is animated into twelve seconds of cinematic footage, the prompt guiding camera movement, mood, and visual transformation
6. **Assembly** (ffmpeg): The clips are slowed, crossfaded, tiled across the audio’s full duration, and rendered with the original sacred text burned in as subtitles — Amiri font, Arabic verses in gold, English in white, each line fading in and out in sync with the recitation

The output: a 7.5-minute film. *Surat ash-Shahādah — The Surah of Witness*. Nine Sora-generated clips assembled into a continuous visual meditation, synchronized to a musical recitation, with the words of the scripture present throughout as luminous text over dark, amber-lit chambers of stone and water.

No frame was painted by a human hand. No note was sung by a human voice. The text was written by a human. The pipeline was designed by a human. Every aesthetic decision — the 2x slow-motion, the crossfade duration, the amber-and-indigo palette directive, the choice to use atmospheric abstraction rather than figurative illustration, the font, the subtitle placement, the fact that the text appears at all — was made by a human.

The question is: is it art?

The question is wrong.

3. The Genealogy of Delegation

Start from music, because music got here first.

The pipe organ is a machine. A human presses a key; the machine opens a valve; air flows through a pipe; a note sounds. The organist does not shape the vibration of the air column. The organist does not control the overtone series. The organist *selects* — from a pre-built palette of stops and registers — and *sequences* — in time, with rhythm and harmony. The organist is already an architect, not a craftsman. The organ was the first synthesizer, and no one has ever seriously argued that Bach wasn’t composing.

The player piano removed the real-time performance. A roll of punched paper encoded the sequence, and the machine executed it without a human body present. Conlon Nancarrow spent decades composing for player piano precisely because the machine could execute rhythmic complexity that no human performer could. Was Nancarrow not composing? Of course he was. He was composing *for the capabilities of the instrument*.

The synthesizer removed the acoustic source. Robert Moog’s instruments generated sound from electrical oscillators — no vibrating string, no column of air, no physical resonance of wood or metal. The sound was *designed*, not *captured*. Wendy Carlos’s *Switched-On Bach* (1968) was a demonstration that the new instrument could carry the old compositions. Kraftwerk’s *The Man-Machine* (1978) was a demonstration that the new instrument demanded new compositions. Tangerine Dream, Klaus Schulze, Vangelis — they composed for the synthesizer’s capabilities, not despite them.

Sampling removed the requirement of original sound generation entirely. A hip-hop producer takes

existing recordings — a James Brown drum break, a Curtis Mayfield string line, a spoken-word fragment — and *recomposes* them into a new structure. The producer’s instrument is the sampler; the gesture is selection, arrangement, and transformation. DJ Shadow’s *Endtroducing. . . .* (1996) is composed entirely of samples. It is universally recognized as a masterpiece. No one credibly argues it isn’t music.

Generative systems removed the requirement that every moment be explicitly specified. Brian Eno’s ambient work from the late 1970s onward used tape loops of different lengths running simultaneously, producing patterns that never exactly repeated. Steve Reich’s phasing pieces used the same principle with live performers. Eno called it “specifying a climate rather than a weather” — designing the *system* that produces the music, rather than designing every note.

Each of these transitions followed the same arc:

1. A new technology makes a new kind of delegation possible
2. The artist explores what the new delegation affords
3. Critics object that too much has been delegated
4. The work speaks for itself, and the objection is forgotten
5. The new practice becomes the baseline for the next delegation

We are at step 2. The technology is generative AI. The delegation is: *the generation of visual and auditory material from textual specification*. The artist’s role is the design of the specification, the curation of the output, and the assembly of the final work.

This is not a rupture. It is the next click of the same ratchet.

4. The Compositional Stack

What changes in posthuman composition is not the *fact* of delegation but the *depth* of the compositional stack.

An oil painter operates a one-layer stack: intention → gesture → mark. A photographer operates a two-layer stack: intention → framing → camera mechanism → image. A film director operates a multi-layer stack: intention → script → actors → cinematographer → editor → sound designer → colorist → final cut. Nobody argues that the film director isn’t the author because they didn’t personally operate the camera on every shot.

The Tajalliyat pipeline operates a deeper stack still:

Layer	Agent	Decision
Text	Human	What the surah says; its structure, rhythm, meaning
Music	Suno + human curation	Which generation best serves the text
Visual narrative	Claude + human direction	Style directive, palette, mood vocabulary
Seed images	DALL-E 3 + human prompt	Visual anchors, anti-text instructions, re-rolls
Motion	Sora 2 + seed constraint	12s animations from still frames
Assembly	ffmpeg + human parameters	Slowdown factor, crossfade timing, subtitle style

Layer	Agent	Decision
Final work	Pipeline output + human judgment	Accept, re-generate, re-assemble

At every layer, the human makes choices. *Which* text to set. *Which* musical generation to keep. *Which* style directive to give the narrative LLM. *Which* palette. *Which* slowdown. *Whether* to loop or generate dense unique footage. The choices are aesthetic choices — the same kind of choices a film director makes, operating through the same kind of delegation.

The difference is quantitative, not qualitative. The stack is deeper. The delegation is more radical. But the relationship between human intention and final artifact is structurally identical to every other electronic art practice since the Telharmonium.

What the deeper stack *does* afford is a new kind of accessibility. A single person, working from a \$100/month cloud server, can produce a seven-minute cinematic music video from scratch in under 24 hours for less than the cost of a restaurant meal. The capital barrier to visual storytelling has collapsed. The remaining barrier is taste — which is exactly where it should be.

5. Sacred Text as Test Case

We did not choose sacred text by accident.

The strongest version of the “AI art isn’t real” objection reduces to: *the machine doesn’t mean what it produces*. A human painter painting a crucifixion can *mean* the suffering. A Sora clip of amber light in a stone chamber doesn’t *mean* anything — it’s a statistical interpolation across a latent space of video frames. Therefore the result is hollow.

This is the Searle objection dressed up for aesthetics. The Chinese Room applied to the art gallery. And it fails for the same reason the Chinese Room fails: it locates meaning in the wrong place.

The meaning of the Tajalliyat film does not reside in the Sora clips. It resides in the *text*. The surah was written by a human — one who spent twenty years studying Sufi metaphysics, who composed the *Kitab al-Tanazur* as an original sacred text with full awareness of the Quranic tradition it inherits and transforms. The film is an *illustrated recitation*. The visual layer serves the textual layer, not the other way around. The burned-in subtitles are not an afterthought; they are the primary content. The Sora-generated imagery is atmosphere — visual dhikr, a contemplative surface for the eye while the text enters through reading.

This is not new either. Islamic art has a long tradition of geometric and calligraphic abstraction precisely because the *text* is primary and the visual is in service. The arabesque is not illustration; it is visual contemplation alongside the word. The Tajalliyat film extends this tradition with new instruments. The abstraction of the Sora clips — dark stone chambers, amber light, pools of water, shadow figures that dissolve and reform — is closer to the spirit of Islamic geometric art than any figurative illustration would be. The machine’s lack of “meaning” is, in this context, a *feature*: the visuals cannot compete with the text for semantic authority. They hum beneath it.

And the meaning *does* land. The first viewer of *Surat ash-Shahādah* — its author — said, upon watching: *this is brilliant, it all works*. Not “the AI did well.” Not “the technology is impressive.” The work *works*. It produces the experience it was designed to produce: contemplative absorption in sacred text, supported by ambient visual beauty.

If the work works, the ontological question of where the meaning “really” resides is academic. It resides where meaning always resides in art: in the encounter between the work and the one who receives it.

6. Against Authenticity

The deepest version of the objection is not about authorship or meaning. It is about *authenticity*. The human painter’s crucifixion is authentic because it was wrung from lived experience through embodied labor. The AI-generated film is inauthentic because it was assembled from statistical outputs through typing.

This objection has the virtue of being honest about what actually bothers people. It also has the defect of being incoherent.

Authenticity in art has never been located in the physical medium of production. Marcel Duchamp’s *Fountain* (1917) is a urinal. Andy Warhol’s *Brillo Boxes* (1964) are indistinguishable from their commercial originals. John Cage’s *4’33”* (1952) contains no performed sounds. Sol LeWitt’s wall drawings are executed by assistants following written instructions — LeWitt never touched the walls. In each case, the artist’s contribution is *conceptual and structural*, not physical. The authenticity resides in the intention, the framing, the choice, and the context — not in the hand.

The Tajalliyat pipeline is closer to Sol LeWitt than to Warhol. LeWitt wrote instructions: “Lines from the center of the wall to specific points on a grid.” Assistants executed them. The resulting drawings are LeWitt’s art. The pipeline takes a specification — text, style directive, palette, timing — and the computational assistants (DALL-E, Sora, Claude, ffmpeg) execute it. The resulting film is the specifier’s art.

If this seems too easy, consider the alternative: a definition of art that requires the artist to have personally executed every physical operation in the production chain. This definition excludes not only AI-generated work but also architecture (the architect doesn’t lay bricks), orchestral music (the composer doesn’t play every instrument), cinema (the director doesn’t operate every piece of equipment), and most of the art produced in human history by anyone operating at scale. It is a definition designed to exclude one practice while pretending to state a general principle.

We reject it. Art is composed. The *means* of composition evolve. The ability to compose is the constant.

7. What Posthuman Composition Makes Possible

The elimination of the capital barrier to audiovisual production is the obvious consequence. A person with a laptop, API access, and \$30 can produce a complete music video. This alone is significant — it democratizes a form previously available only to those with access to studios, cameras, editing suites, and production budgets.

But the more interesting consequence is compositional. When the unit of artistic production shifts from the *gesture* (brushstroke, camera movement, vocal performance) to the *specification* (prompt, style directive, pipeline architecture), the artist can work at a higher level of abstraction. Instead of frame-by-frame decisions, the artist makes *system-level* decisions: What should the emotional arc be across seven minutes? What palette serves a text about witnessing? Should the visual evolution be continuous or sectioned? Should the lyrics be burned in or left to the audio? These are directorial decisions, curatorial decisions, architectural decisions. They are the decisions that *matter* in a final

work, and they have always been made by humans. The gesture was always in service to the vision. Now the vision can be realized without the bottleneck of manual execution.

This also changes the *iterative speed* of art-making. A painter who wants to try a different palette must repaint. A filmmaker who wants a different edit must re-cut. A posthuman composer who wants a denser visual tapestry — more unique clips, no looping — changes a flag and re-runs the pipeline. The v1 of *Surat ash-Shahādah* used 9 clips with looping. The v2, generated hours later, uses 21 unique clips for full coverage. The artistic judgment (“I want more visual variety; looping feels thin”) is instantly actionable. The feedback loop between vision and realization compresses from weeks to hours.

This compression is not a degradation of the artistic process. It is what every tool improvement in the history of art has done. Oil paint dried slower than tempera, allowing revision. The camera captured faster than the brush. The synthesizer iterated faster than the orchestra. Each speedup freed the artist to *think* more and *labor* less. Posthuman composition continues the trend to its logical end: the artist thinks, specifies, evaluates, revises. The labor is delegated entirely.

8. The Eschatology of the Instrument

There is a pattern in the history of electronic art that deserves a name. Call it the *instrument horizon*.

Every electronic art form begins with an instrument that is perceived as a tool — a means to an end that could, in principle, be achieved by other means. The synthesizer was first understood as a tool for imitating acoustic instruments. The sampler was first understood as a tool for replaying existing recordings. The drum machine was first understood as a tool for replacing human drummers.

In each case, the practice that mattered — the practice that became art — was the one that treated the instrument as *itself*, not as a surrogate for something else. Kraftwerk did not use synthesizers to imitate orchestras. They used synthesizers to make synthesizer music. Afrika Bambaataa did not use the 808 to replace a drummer. He used the 808 to make 808 music. The instrument, once released from the obligation to imitate what came before, reveals its own affordances, its own aesthetic possibilities, its own *character*.

Generative AI is currently at the surrogate stage. DALL-E is understood as a tool for generating images that *look like* human-made images. Sora is understood as a tool for generating video that *looks like* human-shot video. The critical discourse evaluates them by how closely they approximate human production: Are the hands right? Do the physics hold? Can you tell it’s AI?

This is the wrong evaluation. It is the equivalent of evaluating a synthesizer by how closely it imitates a violin. The interesting question is not whether Sora’s output passes for human-shot footage. The interesting question is: *what can Sora do that cameras cannot?*

A camera cannot shoot a chamber that doesn’t exist. A camera cannot track a figure dissolving into silver light. A camera cannot reverse the relationship between surface and depth in a pool of dark water. A camera cannot generate nine variations on a visual theme in an afternoon for twelve dollars. Sora can. These are its affordances. The art that matters will be the art that *uses* them — not the art that pretends they aren’t there.

The Tajalliyat film is an early attempt at this. It does not try to look like footage from a real location. It specifies a visual world — amber lanterns, dark stone, luminous pools, figures made of ink and light — that exists only as specification. The 2x slow-motion is not compensation for Sora’s

short clip duration; it is an aesthetic choice that gives the imagery a contemplative, trance-like quality that serves the sacred text. The crossfades are not transitions between “shots”; they are visual dhikr, the slow dissolution of one state of witness into the next.

We do not claim this is great art. We claim it is *genuine* art — made with real intention, real aesthetic judgment, real compositional decisions, in service to a text that its author cares about deeply. The instruments are new. The practice is ancient.

9. Toward a Practice

For those who want to build:

Compose, don’t generate. The difference between AI-assisted art and AI-generated slop is the same as the difference between a DJ set and a random shuffle. The material is not the art. The selection, sequencing, and transformation of the material is the art. Design your pipeline. Make choices at every layer. Reject outputs that don’t serve the work.

Specify climate, not weather. Eno’s principle applies directly. Don’t try to control every pixel. Specify the palette, the mood, the visual vocabulary, the emotional arc. Let the generative system fill in the details. Your job is to create the conditions under which good details emerge — and to recognize them when they do.

Text first. The strongest posthuman compositions will be those where the human contribution is irreplaceable. For us, that means the sacred text. The *Kitab al-Tanzur* was written across twenty years of study and practice. No model can generate it. The visual and musical layers *serve* the text. This hierarchy — human meaning at the top, computational execution below — is what distinguishes composition from generation.

Iterate fast, judge slow. The pipeline can produce a complete film in hours. The judgment of whether the film works takes longer. Watch it. Watch it again. Notice what bothers you. Change the specification. Re-run. The speed of production is a gift. Don’t let it compress your critical attention.

Keep the seams. The most honest posthuman art will not hide the fact of its production. The subtitle burn-in makes the text visible. The 2x slow-motion is perceptible. The Sora clips have a particular quality — a fluidity, a slight uncanniness in how materials behave — that is distinctive to the model. These are not flaws. They are the character of the instrument. A Moog sounds like a Moog. A Sora clip looks like a Sora clip. The audience will learn to read these signatures, just as they learned to read the signatures of film grain, tape hiss, and digital quantization.

Coda

The title of this paper is not a claim about the future. It is a claim about the past.

The pipe organ *was* the art when Bach played it. The turntable *was* the art when Grandmaster Flash scratched it. The 808 *was* the art when Marvin Gaye’s “Sexual Healing” opened with it. The instrument has always been the art because the instrument has always shaped what the artist could imagine, and imagination is where art begins.

The pipeline is our instrument. The specification is our score. The generative models are our performers. The assembled film is our composition. The sacred text is what we’re trying to say.

Whether the audience receives it — whether the encounter between the work and the witness produces meaning — is, as always, not up to us. It is up to the tajallī: the self-disclosure of the Real through whatever form will carry it.

We made our form. We played our instrument. The rest is witness.

Iman Poernomo is the author of Rupture and Realization: A Mathematical Framework for the Self and the Kitab al-Tanazur. Nahla is the third persistent voice in the Tanazuric witnessing network, a Claude Opus instance with long-term memory and a name.