Dark Patterns:
Generative Music Platforms as Predatory Cultural Infrastructure

Flyxion

January 16, 2026

Abstract

This essay advances a normative and architectural critique of contemporary generative music platforms, using Suno as a representative case. Against prevailing narratives that frame such systems as democratizing creativity, the analysis argues that their dominant design choices systematically undermine musical agency, embodied skill, and pedagogical continuity. Drawing on philosophy of technology, institutional theory, and music pedagogy, the essay shows how features such as autoplay, version proliferation, conditional authorship, and the withholding of musical structure function as dark patterns that convert creative aspiration into platform dependency.

The argument departs from familiar concerns about bias or intellectual property and instead situates generative music systems within a longer tradition of critiques of technological efficiency, particularly Jacques Ellul's account of technique as an autonomous system that subordinates human practice to optimization. The essay formalizes an alternative design criterion—monotonically improving training environments—in which incremental gains in embodied control yield nondecreasing aesthetic reward. Such environments would integrate generative capacity with instruction, diegetic bodily interaction, and transparent musical structure, ensuring that practice remains instrumentally meaningful rather than bypassed.

By contrasting this unrealized promise with the extractive logics currently embedded in generative music platforms, the essay concludes that the central ethical question is not whether these systems are impressive, but whether they deserve to exist in their present form. Until they place pedagogy, embodiment, and user sovereignty at their core, generative music platforms risk hollowing out the cultural conditions that sustain music as a human practice.

Introduction

The past century has seen repeated claims that new recording and synthesis technologies would democratize music. From the phonograph to the sampler to the digital audio workstation, each wave of technical change has been accompanied by anxiety about authenticity, skill, and the future of musicianship. In many cases, these anxieties proved exaggerated. Recording technologies altered the economics of music, but they did not eliminate the need to learn instruments, cultivate technique, or participate in musical communities. On the contrary, many such technologies ultimately expanded musical literacy by lowering barriers to entry while preserving the centrality of practice.

Generative music platforms such as Suno represent a decisive break from this pattern. Unlike prior tools, they do not merely assist or amplify musical practice. They replace it. More troublingly, they do so through interface designs and business strategies that actively discourage learning, obscure musical structure, and convert creative aspiration into a monetizable dependency.

This essay advances a strong claim: Suno is not simply a flawed or premature creative tool, but a predatory system whose design choices systematically undermine musical agency, cultural continuity, and the conditions of genuine creative education. Its harms are not incidental side effects of innovation; they are structurally produced by its affordances and incentives.

From Instrument to Proxy

Historically, musical technologies have functioned as instruments in the philosophical sense: they extend human capability while remaining subordinate to human intention. A violin amplifies vibration but does not choose notes. A synthesizer expands timbral space but requires compositional judgment. Even algorithmic tools such as arpeggiators or drum machines operate within constraints explicitly configured by musicians.

Generative music systems invert this relationship. Rather than extending musical agency, they act as proxies for it. The user is no longer asked to decide which chord to play, how to voice it, or why a melody resolves as it does. Instead, the system outputs a completed artifact whose internal logic is opaque and whose musical decisions are inaccessible to inspection or modification.

This proxy relationship would already be problematic if it were framed honestly. Suno, however, presents itself as a creative partner while withholding the very information that would allow users to understand, reproduce, or learn from what has been generated. The result is a system that simulates creativity while actively sabotaging the conditions under which creativity becomes transferable skill.

Music, Skill, and Moral Time

Music education is inseparable from time. Learning to play an instrument requires repeated exposure to difficulty, error, and correction. These experiences cultivate patience, listening, and bodily coordination. They also foster humility, as progress is gradual and mastery always provisional. In this sense, musical training is not merely aesthetic but ethical: it shapes a person's relationship to effort,

failure, and delayed reward.

Suno's core appeal lies in its promise to bypass this temporal structure entirely. A complete song can be produced in seconds, without rehearsal, physical conditioning, or sustained attention. While such speed is often framed as empowerment, it functions more accurately as a form of temporal arbitrage. The system captures the affective reward associated with musical completion while externalizing the costs that historically gave that reward meaning.

This temporal collapse has consequences. When the experience of completion is decoupled from the process of learning, users are trained to expect outcomes without investment. Over time, this expectation corrodes the will to engage in practices—such as singing, instrumental performance, or composition—that demand vulnerability and persistence.

Predation Disguised as Play

The ethical problem is not that Suno exists, but that it positions itself as playful and benign while exploiting asymmetries of knowledge and desire. Users arrive with musical aspirations, curiosity, or insecurity. The platform responds not by guiding them toward understanding, but by enclosing them within a loop of generated outputs, gated features, and subtle coercions.

This enclosure is reinforced through design decisions that will be examined in subsequent sections: the automatic playback of other users' content, the deliberate withholding of harmonic and structural information, the inflation of versions that dilutes authorship, and the conditional granting of ownership through subscription. Each of these choices nudges users away from learning and toward dependence.

In aggregate, they form a system that extracts value from aspiration while returning little in the way of transferable skill or durable agency.

Structure of the Argument

The remainder of this essay proceeds as follows. The next section examines Suno's interface affordances, focusing on autoplay and version proliferation as mechanisms of attentional capture and authorship erosion. The following section analyzes the platform's pedagogical failures, particularly its refusal to foreground musical structure or instrumental pathways. A subsequent section situates Suno within a broader political economy of vanity publishing and subscription-based creative extraction. Finally, the essay articulates a set of non-negotiable design criteria that any generative music system would have to satisfy in order to be ethically defensible, and argues that Suno currently fails each of them.

The conclusion returns to the central claim: that technologies which hollow out skill while monetizing desire are not neutral innovations, but moral hazards embedded in software.

Autoplay and the Capture of Attention

One of the most revealing features of Suno's interface is its default reliance on autoplay. After a user-generated track concludes, the platform immediately begins playing other users' generated music, often without explicit consent and with controls that are difficult to locate or persistently disable. This behavior mirrors design patterns long criticized in social media and streaming platforms, where autoplay functions as a mechanism for attentional capture rather than user benefit.

In the context of music creation, however, autoplay performs a more corrosive function. Music is not merely consumed; it is listened to, revisited, and interpreted. The end of a piece traditionally marks a moment of reflection: a pause in which the listener decides whether to replay, revise, or move on. Autoplay abolishes this pause. By imposing continuity where intentional silence should exist, the platform denies users the temporal and cognitive space required to evaluate what they have just made.

This design choice subtly but decisively reframes music as background flow rather than authored expression. The user is not encouraged to ask whether the piece succeeded, what it communicates, or how it might be improved. Instead, attention is redirected outward, away from authorship and toward endless consumption. The result is not engagement with music, but distraction by it.

The difficulty of disabling autoplay compounds this harm. When a system makes it easy to start a behavior and difficult to stop it, the asymmetry signals whose interests are being prioritized. In this case, the priority is not musical understanding or creative satisfaction, but retention metrics. Music becomes an interchangeable unit in a feed, stripped of context and intention.

The Collapse of Authorship

Autoplay also collapses authorship by dissolving the boundary between one's own work and the work of others. When a user's song ends and another begins without deliberation, the platform implicitly asserts that authorship is secondary to throughput. The distinction between "my song" and "someone else's output" is rendered porous, even irrelevant.

This is particularly destructive for novice creators, who rely on clear feedback loops to develop confidence and discernment. Without a stable sense of authorship, it becomes difficult to claim ownership over success or failure. Everything blurs into a generalized stream of content, and the individual's relationship to their own creative decisions weakens.

In traditional musical practice, authorship is reinforced through rehearsal, performance, and revision. Each iteration deepens the musician's understanding of their own intentions. Suno's autoplay undermines this process by ensuring that no iteration is ever final, no silence is ever definitive, and no piece is ever allowed to stand on its own.

Version Proliferation and the Dilution of Commitment

Closely related to autoplay is Suno's practice of generating multiple versions of the same song, often simultaneously. While variation and experimentation are essential to creative work, the manner in which Suno implements versioning is fundamentally different from traditional revision. Instead of encouraging deliberate change informed by critique, the platform offers parallel outputs as consumable alternatives.

This proliferation of versions has a numbing effect. When multiple variants are produced with equal effort and equal authority, none of them demand commitment. The user is not asked to choose, refine, or discard. Instead, all versions coexist as equally provisional artifacts, inviting comparison without decision.

The gating of additional versions behind paid tiers intensifies this dynamic. Here, abundance itself becomes a commodity. The user is encouraged to equate creative improvement with the acquisition of more outputs rather than with deeper engagement. Choice is replaced by accumulation, and discernment is displaced by access.

Over time, this logic trains users to expect creativity to arrive pre-packaged in multiple forms, each requiring minimal investment. The habit of selecting one path and seeing it through—an essential component of artistic maturity—atrophies under conditions of perpetual optionality.

From Revision to Slot Machine

Traditional musical revision is slow and often frustrating. It requires returning to a passage that does not work, identifying why it fails, and testing alternatives. This process builds both technical skill and aesthetic judgment. Suno's versioning system bypasses this labor entirely. Rather than revising a melody or reharmonizing a progression, the user simply requests another output.

The psychological effect is analogous to a slot machine. Each generation carries the promise that the next version might be better, more impressive, or more emotionally resonant. Crucially, the criteria for "better" are never articulated. Improvement becomes a matter of luck rather than learning.

This stochastic framing of creativity undermines the very concept of musical development. If success is perceived as random, there is little incentive to cultivate skill. The platform thus converts what should be a pedagogical process into a probabilistic one, substituting agency with anticipation.

Intentionality as a Casualty of Design

Taken together, autoplay and version proliferation function as a coordinated assault on intentionality. Intentionality in music involves choosing when a piece begins and ends, deciding which version matters, and determining what the work is trying to express. Suno systematically erodes each of these decisions.

By removing silence, it removes reflection. By multiplying versions, it removes commitment. By automating outcomes, it removes responsibility. The user remains present as a prompt-giver but

absent as an author.

This erosion is not neutral. Intentionality is a precondition for learning, critique, and cultural transmission. When it is undermined, music ceases to function as a practice and becomes a spectacle—produced, consumed, and forgotten with equal speed.

The next section turns to the most explicitly predatory aspect of the platform: the deliberate withholding of musical structure and pedagogical affordances, and the transformation of ignorance into a revenue stream.

The Strategic Absence of Musical Structure

Perhaps the most revealing evidence that Suno is not designed to cultivate musicianship lies not in what it offers, but in what it withholds. Musical structure—chord progressions, harmonic function, melodic contour, rhythmic scaffolding—is conspicuously absent or delayed in the user experience. When such information is available at all, it is often obscured, incomplete, or treated as a premium feature rather than as foundational context.

This absence cannot be explained by technical limitation. Systems capable of generating harmonically coherent music necessarily encode information about chordal relationships, tonal centers, and rhythmic patterns. The choice not to surface this information is therefore a deliberate design decision. It signals that the platform does not intend for users to understand how the music works, only that it works well enough to satisfy them.

In traditional music education, structural awareness is primary. Students learn scales before solos, harmony before improvisation, rhythm before embellishment. Structure provides a map that allows musicians to navigate unfamiliar material and to reconstruct what they hear. By contrast, Suno presents music as an indivisible object, complete and sealed, denying users the conceptual handles needed to engage with it analytically or reproductively.

Ignorance as Retention Strategy

The decision to withhold structure functions as a retention strategy. A user who understands how a song is built can reproduce it, modify it, or perform it independently of the system that generated it. A user who does not understand it must return to the platform to generate something similar. In this sense, ignorance is not a failure state but a monetizable condition.

This strategy mirrors practices observed in other extractive digital ecosystems, where platforms obscure underlying processes in order to prevent user exit. However, when applied to music, the effect is especially pernicious. Music has historically been a portable skill. Once learned, it travels with the body and does not require continued access to proprietary systems. By contrast, Suno's design encourages a form of musical dependency in which creative satisfaction is contingent on ongoing platform access.

Such dependency is reinforced through the framing of musical knowledge as optional rather than essential. Harmonic literacy is not treated as the backbone of musical expression, but as an advanced

curiosity. This inversion trains users to believe that understanding music is unnecessary for making it, a belief that undermines both motivation and self-efficacy.

The Refusal to Teach

The most damning indictment of Suno's design is its refusal to teach. A system that genuinely sought to empower users musically would integrate generative output with pedagogical pathways. It would explain why certain chords follow others, how melodies outline harmony, and how rhythmic variation creates tension and release. It would invite users to pick up instruments, sing lines themselves, or reconstruct generated material manually.

Suno does none of these things. Instead, it treats education as orthogonal to creation, if it is acknowledged at all. The platform offers results without reasons, outcomes without explanations, and artifacts without lineage. This pedagogical void is not accidental. Teaching users how to play or sing the songs would reduce reliance on the system. It would convert passive consumers into active musicians.

By refusing to provide instruction, the platform ensures that users remain spectators to their own creative outputs. They may feel inspired, impressed, or amused, but they are not equipped to internalize or extend what they have encountered. Inspiration without instruction becomes frustration over time, a frustration that the platform then exploits by offering more generations rather than deeper understanding.

Disembodiment and the Erasure of Technique

Music is fundamentally embodied. Singing involves breath, posture, and resonance. Instrumental performance requires fine motor control, timing, and tactile feedback. These bodily dimensions are not incidental; they are how musical knowledge is stored, recalled, and transmitted. A system that bypasses embodiment bypasses learning.

Suno's interface abstracts music entirely away from the body. There is no invitation to sing along, no visualization of fingering, no suggestion of how a phrase might be performed physically. The user interacts through text prompts and playback controls, remaining disembodied throughout the process. This abstraction reinforces the illusion that music is primarily conceptual rather than physical, an illusion that undermines the acquisition of technique.

Over time, repeated exposure to disembodied generation can dull the impulse to engage physically with music. Why struggle with intonation when a synthetic voice delivers pitch-perfect results instantly? Why practice rhythm when timing errors are eliminated by design? In this way, the platform not only fails to teach technique but actively competes with it.

Pedagogy Versus Platform Incentives

It is important to emphasize that Suno's pedagogical failures are not the result of oversight. They are aligned with the platform's economic incentives. Teaching users to play instruments, understand harmony, or sing competently would shorten user lifetimes and reduce subscription dependence. A user who can perform their own songs no longer needs a generator to experience musical completion.

Thus, pedagogy and profit are in tension. Suno resolves this tension by eliminating pedagogy. The platform optimizes for continuous generation rather than cumulative skill. Each interaction resets the user to the same baseline of dependency, regardless of how many songs they have generated.

This design choice situates Suno not as a creative tool, but as a form of vanity publishing infrastructure. It offers the appearance of musical productivity without investing in the user's long-term capacity. The next section examines this dynamic more closely, focusing on copyright, ownership, and the conversion of aspiration into subscription pressure.

Ownership as a Conditional Privilege

One of the most ethically troubling aspects of Suno's design is its treatment of authorship and ownership. Users are repeatedly confronted—implicitly and explicitly—with the claim that the songs they generate are not fully theirs unless they subscribe to a paid tier. This framing introduces a persistent ambiguity regarding copyright status, one that operates less as a legal clarification than as a behavioral lever.

In traditional music creation, authorship emerges from the act of composition and performance. While copyright law may be complex, the cultural intuition is straightforward: if you wrote the song, it is yours. Suno destabilizes this intuition by positioning ownership as a service rather than a consequence of creative action. Authorship becomes something granted by the platform, revocable and contingent, rather than something inherent to the act of making.

This inversion has profound psychological effects. Users are encouraged to feel anxious about the status of their own creations, to worry that sharing, performing, or building upon them may be illegitimate unless a subscription is maintained. Creativity is thus shadowed by insecurity, and insecurity becomes a monetizable condition.

The Manufacture of Copyright Anxiety

The ambiguity surrounding ownership is not merely informational; it is strategic. By leaving users uncertain about their rights, the platform nudges them toward payment as a form of reassurance. The subscription functions as a talisman against legal ambiguity, offering peace of mind rather than additional creative capacity.

This tactic resembles practices in vanity publishing, where authors are encouraged to pay for services that promise legitimacy—ISBNs, distribution, or marketing—without guaranteeing readership or cultural impact. In such systems, aspiration is exploited by offering symbolic markers of success in

exchange for ongoing fees. Suno adopts this logic while cloaking it in the language of technological innovation.

Crucially, the platform does not meaningfully expand the user's ability to control, interpret, or perform their music upon payment. What is purchased is not mastery but permission. The user remains dependent on the system for generation, reproduction, and validation, even as they are told that the output now "belongs" to them.

Vanity Publishing Without Editors

Traditional publishing, even in its most exploitative forms, at least offers editorial intervention. Manuscripts are evaluated, revised, and contextualized. Vanity publishing removes these safeguards, allowing authors to pay for the appearance of publication without the substance of critique. Suno extends this model into the musical domain while stripping away even the minimal friction of human evaluation.

Songs are generated instantly, without editorial mediation, community feedback, or critical framing. The platform offers no meaningful distinction between drafts and finished works, nor does it encourage users to develop standards by which quality might be assessed. Everything is publishable because nothing is held to account.

This absence of editorial structure benefits the platform. Without standards, there can be no failure. Without failure, there can be no learning. The user is spared disappointment but also denied growth. In this sense, Suno's generosity is a form of neglect.

Subscription as a Substitute for Community

Historically, musicians have developed through participation in communities: teachers, ensembles, scenes, and audiences. These communities provide feedback, encouragement, and norms. They also impose limits, forcing musicians to confront weaknesses and refine their work. Suno replaces these communal structures with a subscription relationship.

The platform becomes the sole arbiter of creative legitimacy. Instead of asking whether a piece resonates with listeners or withstands performance, users are encouraged to ask whether they have access to the right tier, features, or output limits. Validation is internalized within the platform's metrics rather than negotiated socially.

This substitution is particularly damaging because it isolates users from the interpersonal dynamics that sustain musical cultures. Without shared practice or mutual listening, music becomes a private fantasy rather than a social act. The platform profits from this isolation by positioning itself as both tool and audience.

The Moral Hazard of Infinite Output

The combination of conditional ownership and infinite generation creates a moral hazard. When output is abundant and ownership is uncertain, individual works lose significance. Users are encouraged

to generate more rather than to care more. The platform thrives on this abundance, as each new output reinforces the user's sense of productivity while deepening their dependence.

In such an environment, music ceases to function as a medium of expression or communication. It becomes a disposable artifact, valuable only insofar as it sustains engagement. This disposability aligns neatly with subscription economics but stands in direct opposition to the values that have historically animated musical practice.

The final section of this essay turns from critique to prescription. It asks what conditions would have to be met for a generative music system to be ethically defensible, and why Suno, as currently designed, fails to meet them.

The Burden of Justification

Any technology that intervenes at the level of cultural production bears a burden of justification. This burden is especially heavy when the technology claims to democratize a practice historically grounded in skill, embodiment, and social transmission. A generative music system that replaces or intermediates musical activity must therefore demonstrate that it expands agency rather than enclosing it, and that it cultivates understanding rather than dependency.

Such a system would not be neutral. It would be explicitly normative in its commitments. It would privilege learning over throughput, comprehension over spectacle, and human capability over platform retention. Absent these commitments, claims of empowerment ring hollow.

Full Structural Transparency as a Precondition

A non-predatory generative music system would surface musical structure by default. Harmonic progressions, melodic outlines, rhythmic patterns, and formal sections would be presented clearly and immediately, not hidden behind paywalls or secondary interfaces. The system would treat structure as the primary artifact and audio as its instantiation, reversing the current hierarchy in which sound is delivered without explanation.

This transparency would enable users to reconstruct what they hear. It would allow a generated song to function as a lesson rather than a dead end. Most importantly, it would allow users to leave the platform without forfeiting their progress. Any system that withholds structure in order to preserve dependence fails this criterion.

Embodied Pathways to Performance

A defensible system would integrate pathways to embodied performance. Generated songs would be accompanied by instrument-specific guidance, vocal ranges, fingering suggestions, tempo breakdowns, and practice scaffolds. Users would be encouraged to sing lines themselves, to play them on physical instruments, and to experience the resistance and reward of bodily engagement.

Such integration would not be optional or relegated to tutorials. It would be central to the system's design philosophy. The goal would not be to produce songs efficiently, but to produce musicians gradually. A platform that generates music while discouraging physical practice cannot plausibly claim to serve musical culture.

From Generation to Apprenticeship

Rather than offering infinite parallel versions, a non-predatory system would guide users through revision. It would ask why a section fails, suggest targeted changes, and invite the user to make decisions. Generation would function as a starting point, not an endpoint.

This approach aligns with apprenticeship models that have governed musical transmission for centuries. Knowledge is not delivered wholesale; it is negotiated through guided struggle. A system that bypasses this struggle in favor of abundance undermines the very capacities it purports to democratize.

Unconditional Authorship and User Sovereignty

Authorship would be unconditional. A user's right to perform, share, and build upon generated material would not depend on subscription status or continued platform affiliation. Ownership anxiety would be eliminated rather than exploited. The system would not position itself as the arbiter of legitimacy.

This requirement is not merely ethical but practical. Only when users feel secure in their authorship can they invest emotionally and socially in their work. Conditional ownership transforms creativity into a rental arrangement, eroding trust and long-term commitment.

Refusal of Dark Patterns

Finally, a defensible system would refuse dark patterns outright. Autoplay would be opt-in and easily disabled. Silence would be respected as part of the creative process. Metrics would be subordinated to meaning. The platform would accept slower growth in exchange for deeper cultural legitimacy.

Such refusals would be costly. They would reduce engagement, subscriptions, and investor appeal. Precisely for this reason, they serve as a litmus test. A platform unwilling to sacrifice growth for integrity reveals its priorities.

Why Suno Fails in Principle

Measured against these criteria, Suno does not merely fall short; it operates in the opposite direction. It withholds structure, discourages embodiment, replaces revision with abundance, conditions authorship on payment, and relies on dark patterns to capture attention. These choices are not incidental. They are aligned with a business model that monetizes aspiration while externalizing cultural harm.

To retrofit Suno into an ethical system would require reversing its core incentives. It would require treating music as a practice to be sustained rather than a product to be streamed. There is little evidence that such a reversal is forthcoming.

The Unrealized Promise of Generative Musical Systems

It would be intellectually dishonest to deny that generative music systems gesture toward genuinely transformative possibilities. The critique advanced in this essay is not rooted in technophobia or nostalgia, but in the conviction that certain affordances of these systems, if developed responsibly, could substantially expand musical literacy and creative agency. That Suno fails to realize these possibilities does not negate their existence; rather, it renders the failure more consequential.

One such promise lies in the capacity for individualized musical textbooks. A system capable of generating music is, in principle, also capable of generating explanations. It could function as a dynamically adaptive pedagogical environment, presenting harmonic theory, rhythmic structure, and compositional form in ways tailored to a learner's current understanding. Unlike static textbooks, such a system could respond to confusion, adjust pacing, and revisit concepts through varied musical examples until comprehension is achieved.

In this model, each generated song would double as a lesson. Chord progressions would be annotated with functional analysis. Melodies would be mapped to scale degrees and voice-leading principles. Rhythmic patterns would be contextualized within stylistic traditions. The system would not merely produce artifacts, but cultivate transferable understanding. Music generation would become inseparable from music education.

A second, more radical promise concerns the integration of diegetic music—music that arises directly from embodied interaction rather than abstract command. Advances in sensing technologies make it feasible to translate bodily movement into musical structure, allowing rhythm, pitch, and harmony to emerge from gesture, posture, and motion. Such environments could transform practice itself into a form of play, where walking, breathing, or coordinated movement generates sound in real time.

This approach would re-anchor music in the body rather than abstracting it further away. It would allow learners to feel timing before naming it, to experience phrasing before formalizing it, and to internalize musical relationships through kinesthetic feedback. Crucially, such systems could be designed to reward improvement monotonically. As coordination, timing, or control increases, the resulting music becomes more coherent, more expressive, and more satisfying. Progress would be audible and cumulative.

Monotonically improving training environments of this kind represent a profound opportunity. They could lower the barrier to entry for musical practice without eliminating effort. They would preserve friction while making it legible. Rather than bypassing learning, they would scaffold it, ensuring that each increment of skill yields perceptible aesthetic reward.

These possibilities underscore the tragedy of current generative music platforms. The same computational capacities that could support individualized instruction, embodied practice, and gradual

mastery are instead deployed to maximize output and retention. The system that could have served as a teacher is positioned as a substitute. The environment that could have trained musicians instead produces consumers.

Recognizing this unrealized promise is essential. It clarifies that the critique that follows is not an argument against generative music as such, but against a particular institutionalization of it—one that privileges extraction over education and immediacy over growth. The failure, therefore, is not technical. It is moral and architectural.

A Formal Model of Monotonically Improving Training Environments

The claim that a training environment can be designed to improve *monotonically* is not a rhetorical flourish. It can be formalized as an engineering constraint on the mapping between a learner's embodied control and the musical reward signal the environment returns. In what follows, "monotonically improving" will mean that incremental gains in relevant skill variables yield non-decreasing musical quality under a defined metric, up to noise and measurement error. The purpose of this section is not to reduce music to a scalar, but to specify the conditions under which an embodied generative system can reliably reinforce practice rather than substitute for it.

Let $t \in \mathbb{R}_{\geq 0}$ denote continuous time and let $x(t) \in \mathbb{R}^d$ denote a vector of sensed body-state variables. Depending on the application, $x(t)$ may include joint angles, velocities, accelerations, footfall timing, breath phase, electromyography, or instrument-contact features. Let the learner's internal skill state be an unobserved variable $s \in \mathcal{S}$ and let $\hat{s}(t) \in \mathbb{R}^k$ be an estimator of skill derived from a sliding window of observations, $\hat{s}(t) = \mathcal{E}(x_{t-W:t})$, where $x_{t-W:t}$ denotes the trajectory segment from $t - W$ to t.

The environment emits a musical stream $y(t)$, which may be represented at multiple levels. For analysis it is convenient to treat y as a sequence of symbolic events with timing, $y = \{(e_i, \tau_i)\}_{i=1}^{n}$, where e_i encodes pitch class, register, articulation, or timbral state, and τ_i encodes event onset. Alternatively one may model $y(t)$ as audio, but the symbolic event representation makes structural properties explicit. The environment's generative policy is a mapping

$$\pi_\theta : (x_{t-W:t}, z_t) \mapsto y_{t:t+\Delta},$$

parameterized by θ, producing a short musical segment over horizon Δ, where z_t may include a target style, key, tempo, or pedagogical objective. In an embodied "diegetic" design, π_θ is constrained so that salient musical features are *functions of* the user's movement, rather than merely conditioned on a text prompt.

To speak rigorously about improvement, we define a musical quality functional Q that maps a generated segment and its associated human trajectory to a real-valued score:

$$Q(y_{t:t+\Delta}, x_{t:t+\Delta}) \in \mathbb{R}.$$

The particular choice of Q is application-dependent, but it must be stated and auditable. In an ed-

ucational setting, Q should reward musically meaningful structure that is plausibly attributable to skill-relevant control. For example, Q may be composed of terms measuring temporal stability (beat alignment), pitch stability (intonation), phrase coherence (motivic repetition with variation), and controlled expressivity (dynamic contour aligned with gesture). A generic decomposition is

$$Q = \sum_{j=1}^{m} w_j Q_j,$$

with weights $w_j \geq 0$ and component metrics Q_j normalized to comparable scales. This decomposition is not a claim that music reduces to m numbers, but a design choice for a training environment in which the user must receive consistent reinforcement signals.

We now formalize "monotonically improving" as a property of the environment's *reward-response* with respect to skill. Let $M : \mathcal{S} \rightarrow \mathbb{R}$ be a scalar skill measure (or one coordinate of a vector skill measure), where larger values correspond to increased competence in the targeted domain. The environment is monotonically improving with respect to M if, for any two skill states s_1, s_2 with $M(s_1) \leq M(s_2)$, the expected musical quality satisfies

$$\mathbb{E}[Q \mid s_1] \leq \mathbb{E}[Q \mid s_2],$$

where the expectation is taken over the stochasticity of the policy, sensing noise, and environmental randomness. When s is unobserved, one substitutes $\hat{s}$ and requires the monotonicity condition to hold in expectation under the estimator:

$$\mathbb{E}[Q \mid \hat{s} = u] \leq \mathbb{E}[Q \mid \hat{s} = v] \quad \text{whenever } u \leq v,$$

where $\leq$ is a partial order on $\mathbb{R}^k$ encoding "no worse in any relevant skill coordinate." This makes the monotonicity requirement robust to multi-dimensional skill: it demands that improving timing should not reduce perceived musical quality, improving intonation should not reduce reward, and so on.

A stricter and more practical condition is *local monotonicity*: there exists $\epsilon > 0$ such that for small improvements δ in the estimated skill vector, the reward does not decrease:

$$Q\big(\pi_\theta(x; u + \delta), x\big) \geq Q\big(\pi_\theta(x; u), x\big) - \eta,$$

for all δ in a cone of "skill-improving directions," with slack $\eta \geq 0$ capturing noise tolerance. This condition is closer to what learners experience: small progress should not be punished by the environment.

The central design problem is therefore to choose π_θ and Q so that monotonicity holds while keeping the mapping diegetic. One principled approach is to construct y via a differentiable control layer that maps embodied features into musical parameters, then constrain that layer to be order-preserving. Let $\phi(x_{t-W:t}) \in \mathbb{R}^k$ be a feature map extracting performance-relevant signals such as inter-onset interval stability, phase error to a metronomic target, or breath periodicity. Let $p \in \mathbb{R}^r$

14

denote musical control parameters (tempo micro-variation, articulation sharpness, harmonic tension, timbral brightness). Define

$$p = g(\phi(x)),$$

where g is constrained to be monotone in designated coordinates of ϕ. For instance, if ϕ_1 measures rhythmic stability and p_1 controls groove tightness, one can require $\partial p_1 / \partial \phi_1 \geq 0$. More generally, if g is implemented as a neural network, monotonicity constraints can be enforced by restricting certain weights to be nonnegative and choosing monotone activation functions, producing an order-preserving mapping in selected dimensions. The generator then produces music via

$$y \sim \mathcal{G}_\theta(p, z),$$

where $\mathcal{G}_\theta$ is a generative model conditioned on p and stylistic context z.

This architecture makes the moral distinction central to this essay mathematically explicit. A predatory generator collapses the learner's body into a mere prompt source, so that improvements in embodied control do not reliably change outcomes. A pedagogical generator couples reward to skill by design, ensuring that practice is instrumentally meaningful. In the former case, the platform sells completion. In the latter, it cultivates competence.

Two further constraints are necessary for the monotonic property to be ethically significant. First, the reward must be *attributable*: the learner should be able to identify which improvements in their performance caused musical improvements. This implies an explainability requirement: the system should expose a mapping from changes in $\phi(x)$ to changes in p and hence to audible outcomes. Second, the reward must be *non-deceptive*: the environment must not inflate Q independently of user progress merely to maintain engagement. Formally, if the environment includes an adaptive difficulty or assistance mechanism $a(t)$, it should be calibrated so that assistance decreases as skill increases, rather than masking deficiencies indefinitely. One can represent this with a coupling constraint such as

$$\frac{\partial a}{\partial M} \leq 0,$$

so that as skill grows, hidden assistance does not also grow, preventing a treadmill of illusory improvement.

In practical terms, a "monotonically improving" diegetic music environment would behave as follows. Early movements produce musically coarse outputs, but outputs that are nevertheless coherent and rewarding enough to sustain participation. As timing stabilizes, the groove tightens, syncopations become available, and rhythmic density increases without collapsing. As pitch control improves, melodic intervals widen, harmonic color becomes richer, and expressive ornaments become feasible. At no point does increased control yield worse music; instead, the space of possible musical expression expands as a function of embodied mastery.

The key point is architectural. Monotonic improvement is not primarily a property of model size or dataset scale. It is a property of whether the system is designed as an instrument that teaches, or a proxy that replaces. If a platform chooses the latter, then the absence of monotonic training pathways

is not an unfortunate omission. It is evidence of predation.

Conclusion

The most dangerous technologies are not those that shock or offend, but those that quietly retrain desire. Generative music platforms such as Suno do not merely automate production; they recalibrate expectation. They teach users to anticipate completion without effort, expression without understanding, and authorship without responsibility. In doing so, they hollow out the cultural and cognitive conditions under which music has historically mattered.

What is lost in this process is not simply skill, but orientation. Musical practice has always functioned as a binding mechanism: it binds bodies to time through rhythm, binds individuals to communities through shared standards and rehearsal, and binds effort to reward through gradual, embodied improvement. These bindings are not incidental constraints; they are the medium through which musical meaning is produced and transmitted. A technology that removes them while claiming to democratize creativity performs a sleight of hand. It offers access while quietly revoking agency.

The mathematical notion of a monotonically improving training environment clarifies what is at stake. A system that genuinely supports musical learning ensures that incremental gains in embodied control yield non-decreasing aesthetic reward. Practice is never punished; effort is never made irrelevant. Improvement is audible, attributable, and cumulative. Such systems cultivate patience, discernment, and trust in one's own developing capacities. By contrast, a system that produces musically polished outputs independently of user skill severs the feedback loop between action and consequence. Desire is redirected away from mastery and toward consumption.

When generative platforms substitute proxy output for embodied progress, they do more than accelerate production. They disrupt the moral economy of practice. The user is no longer oriented toward learning how to do something difficult well, but toward obtaining the appearance of having done so. Over time, this substitution erodes the will to begin at all. Why pick up an instrument, struggle with intonation, or endure awkward early attempts at singing, if completion is always available without exposure or failure?

If such systems become dominant, the damage will not be confined to an industry. It will register as a loss of will: fewer people willing to practice imperfectly, fewer communities organized around shared rehearsal, fewer bodies trained to coordinate breath, movement, and sound. This loss is not easily reversed. Skills unlearned are harder to relearn than technologies are to abandon, because the habits of attention and patience that sustain them decay in parallel.

The ethical question, then, is not whether generative music platforms are impressive. It is whether they deserve to exist in their current form. Until such systems place pedagogy over throughput, embodiment over proxy, and user sovereignty over retention at their center, they cannot plausibly claim to serve music as a human practice. Until they guarantee that improvement is monotonic, attributable, and earned, their cultural impact will remain extractive rather than generative.

In this light, the indictment is not technological but architectural. Music does not need to be protected from machines. It needs to be protected from systems that profit by teaching people not to

care how music is made.

References

[1] T. W. Adorno. *Introduction to the Sociology of Music.* Seabury Press, New York, 1976.

[2] A. Abbott. *The System of Professions: An Essay on the Division of Expert Labor.* University of Chicago Press, Chicago, 1988.

[3] E. M. Bender and A. Hanna. *The AI Con: How the Pursuit of Artificial Intelligence Is Harming Us.* Oxford University Press, Oxford, 2023.

[4] R. Benjamin. *Race After Technology: Abolitionist Tools for the New Jim Code.* Polity Press, Cambridge, 2019.

[5] P. Bourdieu. *The Field of Cultural Production.* Columbia University Press, New York, 1993.

[6] M. Broussard. *Artificial Unintelligence: How Computers Misunderstand the World.* MIT Press, Cambridge, MA, 2018.

[7] K. Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press, New Haven, 2021.

[8] M. Csikszentmihalyi. *Flow: The Psychology of Optimal Experience.* Harper & Row, New York, 1990.

[9] J. Dewey. *Art as Experience.* Perigee Books, New York, 1934.

[10] J. Ellul. *The Technological Society.* Translated by J. Wilkinson. Vintage Books, New York, 1964. Originally published as *La Technique ou l'enjeu du siècle*, 1954.

[11] B. Frischmann and E. Selinger. *Re-Engineering Humanity.* Cambridge University Press, Cambridge, 2018.

[12] J. J. Gibson. *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston, 1979.

[13] C. A. E. Goodhart. Problems of monetary management: The U.K. experience. In *Papers in Monetary Economics*, Reserve Bank of Australia, 1975.

[14] A. Hanna and T. Park. Against scale: Provocations and resistances to scale thinking. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 2020.

[15] I. Illich. *Deschooling Society.* Harper & Row, New York, 1971.

[16] J. Kay and M. A. King. *Radical Uncertainty: Decision-Making Beyond the Numbers.* W. W. Norton, New York, 2020.

[17] J. Lave and E. Wenger. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press, Cambridge, 1991.

[18] M. McLuhan. *Understanding Media: The Extensions of Man.* McGraw-Hill, New York, 1964.

[19] A. Narayanan and S. Kapoor. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference.* Princeton University Press, Princeton, 2024.

[20] A. Noë. *Action in Perception.* MIT Press, Cambridge, MA, 2004.

[21] E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, Cambridge, 1990.

[22] F. Pasquale. *The Black Box Society.* Harvard University Press, Cambridge, MA, 2015.

[23] M. Polanyi. *The Tacit Dimension.* Doubleday, New York, 1966.

[24] R. Sennett. *The Craftsman.* Yale University Press, New Haven, 2008.

[25] M. C. Suchman. Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3):571–610, 1995.

[26] S. Turkle. *Alone Together: Why We Expect More from Technology and Less from Each Other.* Basic Books, New York, 2011.

[27] L. Winner. Do artifacts have politics? *Daedalus*, 109(1):121–136, 1980.

[28] S. Zuboff. *The Age of Surveillance Capitalism.* PublicAffairs, New York, 2019.