Towards Intelligent Instruments: From One-Shot Tools to Collaborative Music AI

A Survey and Position

Kevin Bretz, Derya Soydaner, and Aske Plaat
Leiden Institute of Advanced Computer Science (LIACS), Leiden University
Leiden, The Netherlands
k.o.bretz@umail.leidenuniv.nl
d.soydaner@liacs.leidenuniv.nl, a.plaat@liacs.leidenuniv.nl

Full paper, MW:M Research 2026, Berlin. Theme *REWIRE!*, Music Tech strand.

Abstract

Today's generative music systems are one-shot: a prompt goes in, a finished track comes out. Every musical decision in between, from rhythm and structure to phrasing and feel, is inferred by the model, leaving the artist only to accept, reject, or regenerate. Gone too is the artist's subconscious signature, built over years and shaped by muscle memory, personality, and the tools at their disposal, now flattened into the generic. This automates the *product* of music-making while discarding the *process* of real-time, negotiated interplay through which musicians actually create.

It may seem like a bleak time to be a musician, yet history shows us that artists adapt to new technologies by making them their own, as they did with the synthesizer that gave us electronic music. This paper proposes rewiring generative AI from an instrument of automation into a partner in collaboration, the kind you play with rather than prompt. We outline the path from offline, one-shot generation toward online systems that can jam: listening to a musician, staying in time, responding to their ideas, and leaving creative control where it belongs. We identify what this demands technically (low-latency causal generation, synchronization and anticipation, responsiveness to change) as well as conceptually (evaluation that rewards interaction, not fidelity alone).

Beyond assistance, a jamming agent is also a scientific instrument: through self-play, ensembles that improvise together open a new way to study how musical creativity forms.

Addressing the *REWIRE!* theme, we argue the connection most worth patching is between artists and machines that play together. We invite musicians, technologists, and researchers to discuss what such an agent must do to earn a place on stage.

Keywords. human-AI collaboration, interactive music systems, real-time generation, live jamming agents, computational creativity, music technology

1 The State of Our Art

Almost anyone outside a research lab, when asked what artificial intelligence means for music, will describe the same machine. A sentence goes in, a finished song comes out. That machine is what the press covers, what the courtrooms have been busy with since the major labels sued the two leading services in 2024 (Recording Industry Association of America, 2024), and what most working musicians picture. It is also, in the technical sense, the state of the art, where the field's largest models, budgets, and audiences converge.

The people this machine worries have numbers behind their worry. When Goldmedia surveyed more than fifteen thousand members of the collecting societies GEMA and SACEM, 71 percent feared AI could make it impossible for music creators to earn a living. The same study put 27 percent of creators' revenues at risk by 2028 (Goldmedia, 2024). Yet 35 percent of the same creators had already used AI in their work, rising to 51 percent among those under thirty-five (Goldmedia, 2024), and follow-up surveys find concern continuing to rise even as understanding of the technology behind it grows (PRS for Music, 2026). Fear and adoption coexist in the same person, and that coexistence is diagnostic. The anxiety is not about the technology's existence. It is about a design in which the artist's role is condensed into describing and choosing, and about who really owns the final product (Kreminski, 2026). The fear, in one word, is about agency, and the field's own journals ask whom the automation of music creation serves (Morreale, 2021).

This paper presents a survey of where generative music AI stands, drawn along an axis of creative autonomy. At one end sit today's generators, machines that decide everything, prompted rather than played with, delivering finished music with a single input. At the other end sits a machine that plays with the musician, at audio fidelity and without a score, complementing their creative decisions in shared time without diminishing their agency. That far end remains the least developed domain on the axis. It is also, we will argue, the domain most worth expanding, as it is the one design in which the agency anxiety dissolves. A partner that plays with you cannot, by construction, replace the act of you playing.

One disclaimer holds throughout. This is not an argument against generative models or their builders. It is an argument against a monoculture that has collapsed a rich design space onto its least collaborative point, favoring the sellable product over the human process that makes it worth selling.

2 A Fork in the Road

In 1906, John Philip Sousa warned that recorded sound would interfere with live music-making (Sousa, 1906). He was soon proven right. At their late-1920s peak, American movie-theater pits employed some 22,000 musicians, roughly a quarter of the country's professional instrumentalists (Kraft, 1996, 1994). When synchronized sound arrived, that employment collapsed to about 4,100 by 1934, and it collapsed while weekly cinema attendance nearly doubled. That is what pure substitution looks like. The musicians' federation reportedly spent half a million dollars on its Music Defense League campaign in defense of living music, yet the public went to the talkies anyway, and most of the displaced never got those jobs back. When a technology is deployed purely as a substitute, adaptation is not guaranteed. Any honest argument about AI and music has to keep the pit musicians of 1930 in mind.

The synthesizer story ran differently. As synthesizers spread through the 1960s and 1970s, the fear was familiar, string sections and session players replaced by one keyboard (Pinch and Trocco, 2002; Théberge, 1997). In May 1982 the Central London branch of the British Musicians' Union passed a motion to ban synthesizers and drum machines from sessions (Williamson and Cloonan, 2016). The national executive declined to adopt it, the press ridiculed it, and the decade that followed became the most synthesizer-saturated in popular music history. What happened was not substitution but transformation, with new genres, new professions, and eventually entire economies, including the club culture Berlin is famous for, built on machines once dismissed as soulless.

One device encompasses the mechanism best. The TR-808 arrived in 1980 as an affordable programmable rhythm machine, effectively a direct replacement for drummers. Judged on those terms it failed. Its sounds were perceived as toy-like next to the sample-based machines of the time, and production ended around 1983 after reportedly fewer than twelve thousand units were sold (Werner, 2015). Then the used market made it cheap, and it landed with producers who had no interest in imitating a drummer. The circuits that failed as a substitute succeeded beyond anyone's intention as an instrument, becoming a staple in electro, hip-hop, house, and

Table 1: Three regimes of generative music AI, organized by interaction rather than fidelity.

	Regime I, the deus ex machina	**Regime II, the co-producer**	**Regime III, the bandmate**
Time	Not shared, fully offline	Turn-based, offline between turns	Shared, real time
Artist's channel	A single prompt	Edits, controls, curves, gestures, prompts	The performance itself
Machine's role	Renders a finished piece	Iteratively recommends and renders	Listens, plays, improvises, adapts
Musical decisions	Almost all inferred by the model	Interleaved between artist and model	Negotiated continuously
Representative systems	Jukebox, MusicLM, MusicGen, Stable Audio Open, consumer services	Bach Doodle, Piano Genie, Wekinator, DDSP, Cococo, Music ControlNet, MusiConGen, Amuse	Voyager, Continuator, OMax, BachDuet, ACCompanion, ReaLchords, ReaLJam, jam_bot, streaming accompaniment, LiveBand, Magenta RealTime
Typical evaluation	Audio fidelity, prompt adherence	Task success, user studies	Largely unsettled (Sec. 4.2)

techno. Its once unpopular sounds became foundational, and its bass drum circuit is now literally studied in the academic literature (Werner et al., 2014).

The lesson is more nuanced than "artists adapt". The same technology faced two destinies, and what decided between them was not capability but design, interface, price, and access. Replacements are bought instead of musicians, and their story ends in a cost saving. Instruments are bought by musicians, and their story ends in genres, scenes, and professions nobody could have predicted. Technology does not choose which it becomes. People do.

Today's one-shot generators are, precisely, the 808 as advertised, a stand-in for hired musicians aimed at buyers of finished music. The open question is what the 808 as adopted looks like. What is the playable form of this technology?

3 Of Duets and Deities

The technical landscape becomes cleanly segmented once one asks these four questions: What is the role of the machine? How is time treated? What is the artist's mode of input? And where do the musical decisions live? Asked this way, three regimes emerge, supporting previous classifications of musical agents and generation systems (Rowe, 1993; Tatar and Pasquier, 2019; Herremans et al., 2017) and the design space Kim et al. (2026) recently mapped across 184 live systems. Table 1 summarizes them.

3.1 Deus ex Machina

The lineage runs from Jukebox (Dhariwal et al., 2020) through text-conditioned MusicLM and MusicGen (Agostinelli et al., 2023; Copet et al., 2023) to open diffusion systems (Evans et al., 2025) and the consumer services. A description goes in, a mixed and mastered piece comes out, and interaction is reduced to the input of a vague idea in the form of a single text prompt. No musician works this way, hence the name, a deity lowered in to resolve the plot. The regime still deserves recognition. It gives people without instrumental training a way to make something that sounds finished, and it effectively produces functional music at a low price. Plenty of trained musicians chop and resample its output, and the engineering is undoubtedly ingenious. Nonetheless, it remains a "one-shot pony".

The pitfalls of automated music generation become apparent when we look at what the design discards. First, language is a lossy channel, because creative expertise is tacit and embodied and cannot be fully articulated through prompts (Casacuberta and Guersenzvaig, 2025). Music is the extreme case, groove, timbre, phrasing, and feel do not compress into adjectives, and what cannot be said cannot be asked for. Second, the author becomes ambiguous. When the model fills every unspecified decision with the statistically probable choice, outputs regress toward the genre mean and carry more of the tool's signature than the user's (Kreminski, 2026). Controlled studies confirm the pattern, individual output improves while collective diversity narrows (Doshi and Hauser, 2024; Anderson et al., 2024). Third, the process is skipped entirely, and for most musicians the process is the point. Musicology insists that music is an activity before it is an object (Small, 1998), its creative decisions living in real-time interplay (Berliner, 1994; Sawyer, 2003), and practicing musicians state outright that they make music because they love the creative act of making it (Krol et al., 2025). Fourth, the economics of substitution hit hardest where music is bought as product rather than process, in stock, library, jingle, and background music. There, generative systems are projected to displace the majority of music-library revenues by 2028 (PMP Strategy and CISAC, 2024).

The data from streaming services supports this sentiment. By mid-2026 more than half of the tracks uploaded daily to one major streaming service were fully AI-generated, yet they drew only a small percentage of listeners, much of which is fraudulent bot traffic the service demonetized (Deezer, 2026). This becomes more problematic when we take into account that today's uploads become tomorrow's training data, and models retrained on their own output demonstrably collapse (Shumailov et al., 2024). And since 97 percent of listeners could not reliably tell fully generated tracks from human recordings in a blind test (Deezer and Ipsos, 2025), the finished product no longer equates to human presence.

3.2 Co-Producer

The middle regime provides multiple decision points. Its workflow is the producer's in any DAW: adjust, render, listen, adjust again. Coconet's counterpoint infilling reached a global audience as the Bach Doodle, gathering more than 55 million harmonization requests in three days (Huang et al., 2019). Piano Genie compressed eight buttons into full-range piano in real time (Donahue et al., 2019), the Wekinator let performers teach mappings by demonstration (Fiebrink et al., 2009), and DDSP made learned timbre a functional parameter inside a signal chain (Engel et al., 2020). The regime has since produced a control library for large models, realizing time-varying melody, dynamics, rhythm, and chord controls (Wu et al., 2024a; Lan et al., 2024), and steering interfaces that let novices nudge a model rather than re-roll it (Louie et al., 2020). A neighboring family generates a fitting stem for recorded context, though still offline (Donahue et al., 2023; Parker et al., 2024; Nistal et al., 2024).

This regime supplies the clearest evidence that agency is a design variable. Novices given steering tools reported more trust, control, comprehension, and ownership of the composition relative to the AI (Louie et al., 2020). A recent experiment shows that, as automation rises, psychological ownership and sense of agency fall, even more so for experts (He, 2026). AI Song Contest teams treated their models as collaborators, with their primary complaint being steerability (Huang et al., 2020). Co-creative systems research concentrates on generative ability while neglecting interaction design (Rezwana and Maher, 2023), and an interface built around musicians' own inspirations rather than the model earned a best paper award at the flagship venue (Kim et al., 2025).

This regime is healthy, growing, and already commercially viable, the natural intersection for working musicians. But one property separates it from a partner. Time is not shared. The artist acts, the model renders, the artist reacts, and however tight the loop, the model remains an external device, because nothing musical happens while it thinks.

3.3 Bandmate

The third regime is defined by a single hard constraint. The machine must perceive, decide, and generate within the musician's time frame, committing to each moment irrevocably and recovering from its mistakes in public, like everyone else on stage.

The regime is older than deep learning, reaching back to the score-following accompanists Dannenberg and Vercoe presented in 1984 (Dannenberg, 1984; Vercoe, 1984), with probabilistic successors that learned a soloist's timing from rehearsal (Raphael, 2010) and made anticipation itself a design principle (Cont, 2008). George Lewis's *Voyager* treated the computer as an independent improviser in "a nonhierarchical, improvisational, subject-subject model of discourse" (Lewis, 2000). The Continuator mirrored a player's style in call-and-response (Pachet, 2003), OMax improvised from a live-built corpus of the performer's own material (Assayag et al., 2006), scenario systems guided improvisation along a predefined structure (Nika et al., 2017), reflexive loopers built a band around a soloist (Pachet et al., 2013), and robotic musicianship gave the regime a body (Hoffman and Weinberg, 2011). BachDuet held up one voice of real-time interplay (Benetatos et al., 2020), and the ACCompanion combined reactivity, robustness, and expressivity in a piano accompanist, though still bound to a score (Cancino-Chacón et al., 2023). These systems established the regime's vocabulary of listening, anticipation, and initiative, decades before a functional system was considered realizable.

What is new is the attempt to make such agents learned, score-free, and audio-native, and in the last two years the regime has stopped being hypothetical. ReaLchords showed that reinforcement learning instills anticipatory, adaptive accompaniment that imitation alone does not. The imitation-trained baseline never recovers from mid-song key changes the tuned agent rides out (Wu et al., 2024b). ReaLJam put the result in front of live musicians over a predict-and-commit interface (Scarlatos et al., 2025), and a music language model has held its own in free improvisation with a virtuoso keyboardist (Blanchard et al., 2025). Moving from symbols to sound, streaming accompaniment models generate audio chunk by chunk, exposing the hard latency-quality trade-off and the insufficiency of naive likelihood training when the future is withheld (Wu et al., 2025). Adversarial training has since pushed audio accompaniment to real-time rates on a consumer GPU (Pasini et al., 2026). An open-weights live music model now streams continuously under real-time user control (Lyria Team, 2025), and streaming diffusion has been deployed live as a generative delay effect on consumer laptops (Novack et al., 2026). Real-time neural synthesis is already ordinary concert equipment (Caillon and Esling, 2021). The wider real-time turn points the same way, with full-duplex spoken dialogue at 200 milliseconds (Défossez et al., 2024) and playable neural game engines (Valevski et al., 2025). Music simply sets a higher bar. A conversation tolerates a pause. A groove tolerates tens of milliseconds.

Here is another asymmetry worth naming. Regime I demos in a shared link. Regime III demos require a musician, an instrument, a room, and a whole live performance setup. The regime which addresses the public's loudest objection to AI music thus commands a fraction of the investment the one-shot regime absorbs even as its feasibility is demonstrated lab by lab. That is a wiring fault, and wiring faults can be fixed.

4 The Bar for a Bandmate

Naming the destination is easy. The tricky part is getting there, defining what the machine must do, and how we would know it is doing it.

4.1 Playing in Time

Survive a hard real-time budget. Instrument builders have long held that a playable instrument should answer within about ten milliseconds with a millisecond of jitter (Wessel and Wright, 2002). Performer studies find jittery ten-millisecond latency as damaging as a fixed twenty (Jack et al., 2016). Ensembles are slightly more forgiving, with workable one-way budgets near 25 milliseconds and a stability sweet spot set by natural acoustics (Rottondi et al., 2016; Chafe et al., 2010). The entire loop, capture, inference, synthesis, must fit inside that budget. That defuses the field's favorite lever, since one cannot scale the model when the latency budget limits inference cost (Wu et al., 2025).

Commit under causality. A live agent generates each moment before hearing what it will accompany. This forces a choice offline models never face, wait for context and answer late, or commit early and risk compounding its own mistakes. The literature names this trade-off logical latency versus exposure bias (Wang et al., 2022). The failure mode is audible, as imitation-trained online models drift into their own progression and ignore the partner (Wu et al., 2024b). Remedies are emerging, reinforcement learning that rewards adaptation, and adversarial training whose rollout matches streaming inference exactly (Pasini et al., 2026). Symbolic systems are for the most part protected by their quantized temporal grid, as the beat-quantized commit protocol of Scarlatos et al. (2025) illustrates. Raw audio enjoys no such protection, which is why the audio-native version of this regime is the frontier.

Treat shared time as an input, not an inference. Human bandmates do not infer the metric pulse on every beat. They follow shared cues, a count-in, a drummer, a click, and joint-action psychology shows how much coordination relies on such common ground (Keller et al., 2014). Live agents should be handed what bandmates already share. Our current work, *Silent Metronome*, makes this move, supplying metrical phase to a streaming accompaniment model as a per-frame conditioning signal. It is the machine's equivalent of the click track most stages already run. Under strictly causal operation this more than triples beat alignment over the causal baseline, surpassing even a non-causal reference granted a full second of look-ahead (Bretz et al., 2026). It is one instance of a generalizable design choice where common ground is given, not guessed.

Anticipate, adapt, and recover. An accompanist who reacts only after hearing arrives late by definition, so a partner must predict where the musician is going in order to arrive together. Also, no digital system runs in zero time, which forces an agent to generate at least its own latency ahead simply to sound simultaneous, and musical anticipation stacks on top of that. The requirement runs from classical accompaniment systems (Cancino-Chacón et al., 2023) to agents distilled from teachers that saw the future (Wu et al., 2024b). Meanwhile the human adapts to the machine in return, making the interaction a two-way, nonstationary coordination problem no fixed corpus fully specifies, a natural opening for reinforcement learning. And because commitment is irrevocable, graceful failure is a key feature. On stage there are no wrong notes that matter, only what comes next.

4.2 Judging a Partner, Not a Product

The field's default tools evaluate using the wrong metrics. Fréchet Audio Distance (Kilgour et al., 2019; Gui et al., 2024) and pairwise listening preferences score the system. A jamming agent can score beautifully on both while ignoring its partner completely, because what they cannot see is the relationship. The rewiring this regime needs is an evaluation protocol in which interaction is the measure.

Concretely, that means adherence and interplay balance, or does the agent follow when led and lead when space opens? When perturbations are introduced, such as tempo, key, or dynamics changes mid-take, does it adapt coherently, and how fast? Is the output stable over long horizons across minutes rather than ten-second clips? Does it recover after encountering errors? And, indispensably, musician-in-the-loop studies where the people the system is for are the measurement instrument, ideally ones that follow a system into real practice (Sturm et al., 2019). The early results are instructive because they are mixed. In ReaLJam's live sessions musicians rated enjoyment at 4.3 of 5 but their sense of control at 2.7 (Scarlatos et al., 2025), the gap an interaction-first agenda exists to close. Musicians also report greater engagement with systems that communicate their internal state (McCormack et al., 2019). These constructs, trust, control, communication, do not exist for an offline system. Fidelity was the right metric for the product regime. Interaction is the metric for the partner regime.

5 Beyond Art

Once machines can jam with people, they can jam with each other, and that turns a tool into a social experiment.

Music has no win condition, so self-play in this domain does not converge onto a single champion (Silver et al., 2017). What do ensembles of improvising agents settle into? Turn-taking, call-and-response, a shared pulse as a reference of mutual prediction? These are the phenomena the psychology of joint action studies in human ensembles (Keller et al., 2014; Repp and Su, 2013), but there the experimenter's hands are tied. One cannot ablate a drummer's beat perception, restrict a bassist's prediction horizon, or rerun a take under a different reward. With agent ensembles, every one of those interventions is a config file. Since reinforcement learning evokes interactive behaviors that imitation of corpora does not (Wu et al., 2024b), multi-agent training could discover conventions present in no dataset. The formation of musical coordination, and perhaps of musical culture, becomes an experimental objective. It moves us a step closer to computational creativity's fundamental target, systems that realize creative expression rather than produce passable combinations (Colton and Wiggins, 2012). This section is deliberately speculative. But it changes the implications, because every improvement that makes the agent a better partner makes it a better instrument for studying partnership.

6 Dissonances

Despite our conviction towards AI as a bandmate, the skepticism it draws does not fall on deaf ears, so we answer the six most common counterarguments, and address what we can.

Prompting is itself a creative act. Curation, taste, and prompt-craft are real skills, and nothing here argues against them. The issue is not prompting as an option, but prompting as the only option. A sentence can describe music, but it cannot carry timing, phrasing, or the continuous listening and reacting that playing consists of. Wherever the prompt is the sole interface, those signals are turned into a stale, averaged variant. The argument is for more interfaces alongside the prompt.

The market has spoken. Markets choose among interfaces that exist, and demand for instruments that do not yet exist is invisible. Nobody's 1980 market research predicted that a failed drum machine would create whole genres, that demand was born only after the interface met the right hands. The steering and co-design studies above, where users given control report ownership and musicians describe what they want built (Louie et al., 2020; Krol et al., 2025), suggest demand for controllable interfaces is growing.

Latency is a hardware problem that will solve itself. Yes, actual inference latency will continue to steadily drop as hardware development progresses, but part of the budget is physics, sound crossing a stage, buffer boundaries, human motor delay. Ensemble coupling has an acoustic sweet spot silicon cannot move (Chafe et al., 2010). More fundamentally, even a zero-latency system still has to solve the anticipation problem, and anticipation is modeling, not clock speed.

Live agents raise the barrier to entry. A jamming partner needs someone who can already play, this is true. However, playing is itself designable, as demonstrated by Piano Genie, which turned eight buttons into full-range pianism in minutes (Donahue et al., 2019), and a live agent could likewise meet a beginner at gesture level and grow with them. In fact, such a system could become a core tool for learning to play. Accessibility and agency are not opposed.

A bandmate machine can still replace bandmates. It can, and the lineage began with exactly that ambition (Vercoe, 1984). The narrower claim survives. A design that requires a playing musician cannot remove the act of playing, and who else is on stage remains the musician's choice, not the model's.

It is still trained on musicians' work. Fully conceded, and the interactive domain reformulates the problem rather than sidestepping it. An audit of ten years of music-generation research found that 42.6 percent of training datasets were scraped without permission from the musicians involved (Morreale et al., 2023). The one-shot regime is squeezed on both sides, as it trains using artists' work and then goes on to compete with the same artists. Regime III dissolves most of the output side, because a jamming agent's product is the musician's own performance rather than a competing product. The input side remains but becomes more tractable, as the players themselves can choose to consent, license, and contribute data, analogous to how instrument communities share patches and presets. Consent-based training is no longer hypothetical. Certification schemes exist (Fairly Trained, 2024), artist-led models trained on volunteered voices have been staged as artworks (Cotton and Tatar, 2023), and European law now requires public training-content summaries (European Parliament and Council, 2024; European Commission AI Office, 2025). The first European judgment has held that memorization of protected works in model weights is itself infringement (Landgericht München I, 2025). In July 2026 the same court extended that holding to one of the consumer services (Landgericht München I, 2026). Ownership favors the same design. In the United States purely machine-generated material is unprotectable and prompts alone, on current technology, do not establish authorship. The registrations that succeed demonstrate fine-grained iterative human control (U.S. Copyright Office, 2025; U.S. Court of Appeals for the D.C. Circuit, 2025), which an instrument provides continuously and a text box does not. Even the industry's endgame points here. Two of the three majors struck late 2025 settlements for licensed, fingerprinted generation platforms (Universal Music Group and Udio, 2025; Warner Music Group and Suno, 2025), repairing provenance and compensation while leaving agency exactly where it found it. The remaining suits grind on. A jamming agent built on unconsenting data would be a partner with stolen hands, and no one should want it on stage. Built on consenting data, it is the rare design in which the copyright interest and the creative interest point the same way.

7 How to Rewire

This year's call describes its theme as patching new connections between people, machines, and established structures. Take the metaphor literally and draw today's signal chain. It flows one way, artists' recordings enter as training data at one end, finished products exit toward content buyers at the other. The artist appears only once in that chain, as raw material, and nowhere in it does anyone play. The connection most worth patching runs between artist and machine, in shared time, carrying signal both ways. Making it real is not one community's job, so we close with what it asks of each.

Researchers. Interaction deserves to be a research priority, not an optional demo accessory, a reframing colleagues at this venue have begun to argue for (Smith, 2025). Shared interactive benchmarks can grow around the protocols above, with musicians as co-designers and co-authors rather than as subjects recruited at the end. Their domain expertise is the ground truth.

Builders and industry. The instrument business is the durable business, and no city knows this better than Berlin, whose software instruments and controllers sit on stages everywhere. The natural homes are the places musicians already live, plugins in a DAW, boxes on a pedalboard, devices that speak the clocks and sync protocols the ecosystem already shares, with latency published like a spec-sheet line. A one-shot service meets each customer once. An instrument that grows with its player builds the loyalty this industry was founded on.

Artists. These systems will be trained and designed with or without artists' input. With it, they become instruments. Without it, they default to substitutes. Claiming that seat runs through participatory design, residencies pairing players with research groups, and artist-run evaluation sessions in which the musicians' judgment is the benchmark.

Funders and policy. The surveys show creators are not asking for prohibition, since 95 percent of Goldmedia respondents demand training-data transparency (Goldmedia, 2024). The cultural-funding logic that sustains venues, ensembles, and instrument building extends to artist-in-the-loop AI, and a city that funds club culture can fund the instruments of its next genre. With European transparency obligations in force, the policy scaffolding for consent, licensing, and provenance is already standing (European Parliament and Council, 2024).

What would earn a machine its place on stage is, in the end, what earns anyone their place on stage. It keeps time without being reminded. It listens more than it plays. It follows when led and offers when invited. It recovers from its mistakes gracefully, in public, without taking the band down with it. And it defers, because the music remains the musician's. That is the bar, and we should make no compromises in enforcing it.

8 The Count-In

Read one way, the state of the art names a leaderboard position, the best score, the biggest model, the most complete-sounding audio. Read the other way, it names a condition, the state our art is in, given what we chose to build. This paper has argued that the two readings have come apart. The leaderboard's corner of the design space automates the product of music-making. The corner that would collaborate on the process, live agents that listen, keep time, anticipate, and defer, is demonstrably within reach, and structurally the answer to the fear that dominates the public conversation about AI and music.

History's fork is patiently waiting. The 808 failed as the stand-in it was sold as and changed the world as the instrument nobody planned, the difference decided not by circuitry but by design and hands. The tool always outlasts the argument about it, but who benefits is decided while the argument is still raging. The state of the art is not a static fact about models. It is the current patch, and patches are meant to be rewired.

What must such an agent do to earn a place on stage?

References

Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., Sharifi, M., Zeghidour, N., and Frank, C. (2023). MusicLM: Generating music from text. arXiv preprint arXiv:2301.11325.

Anderson, B. R., Shah, J. H., and Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th ACM Conference on Creativity & Cognition*, pages 413–425.

Assayag, G., Bloch, G., Chemillier, M., Cont, A., and Dubnov, S. (2006). OMax brothers: A dynamic topology of agents for improvization learning. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing for Multimedia (AMCMM)*, pages 125–132, Santa Barbara, CA.

Benetatos, C., VanderStel, J., and Duan, Z. (2020). BachDuet: A deep learning system for human-machine counterpoint improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 635–640, Birmingham, UK.

Berliner, P. F. (1994). *Thinking in Jazz: The Infinite Art of Improvisation.* University of Chicago Press, Chicago, IL.

Blanchard, L., Naseck, P., Brade, S., Lecamwasam, K., Rudess, J., Huang, C.-Z. A., and Paradiso, J. (2025). The jam_bot, a real-time system for collaborative free improvisation with music language models. In *Proceedings of the 26th International Society for Music Information Retrieval Conference (ISMIR).*

Bretz, K., Soydaner, D., and Plaat, A. (2026). Silent Metronome: Rhythmic grounding for live music accompaniment. Unpublished manuscript, project page at `https://kevin-bretz.github.io/projects/silentmetronome/`.

Caillon, A. and Esling, P. (2021). RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. arXiv preprint arXiv:2111.05011.

Cancino-Chacón, C., Peter, S., Hu, P., Karystinaios, E., Henkel, F., Foscarin, F., Varga, N., and Widmer, G. (2023). The ACCompanion: Combining reactivity, robustness, and musical expressivity in an automatic piano accompanist. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5779–5787, Macao.

Casacuberta, D. and Guersenzvaig, A. (2025). Disembodied creativity in generative AI: Prima facie challenges and limitations of prompting in creative practice. *Frontiers in Artificial Intelligence*, 8(1651354).

Chafe, C., Cáceres, J.-P., and Gurevich, M. (2010). Effect of temporal separation on synchronization in rhythmic performance. *Perception*, 39(7):982–992.

Colton, S. and Wiggins, G. A. (2012). Computational creativity: The final frontier? In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI)*, pages 21–26.

Cont, A. (2008). ANTESCOFO: Anticipatory synchronization and control of interactive parameters in computer music. In *Proceedings of the International Computer Music Conference (ICMC)*, Belfast, UK.

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., and Défossez, A. (2023). Simple and controllable music generation. In *Advances in Neural Information Processing Systems*, volume 36.

Cotton, K. and Tatar, K. (2023). Caring trouble and musical AI: Considerations towards a feminist musical AI. In *Proceedings of the International Conference on AI and Musical Creativity (AIMC).*

Dannenberg, R. B. (1984). An on-line algorithm for real-time accompaniment. In *Proceedings of the 1984 International Computer Music Conference*, pages 193–198, Paris, France. Computer Music Association.

Deezer (2026). AI music tops 50 percent of daily uploads on Deezer. Press release, July 21, 2026. `https://newsroom-deezer.com/2026/07/ai-music-exceeds-50-percent-daily-uploads-deezer/`.

Deezer and Ipsos (2025). Deezer and Ipsos study: AI fools 97 percent of listeners. Press release, November 12, 2025. Ipsos survey of 9,000 respondents in eight countries.

Défossez, A., Mazaré, L., Orsini, M., Royer, A., Pérez, P., Jégou, H., Grave, E., and Zeghidour, N. (2024). Moshi: A speech-text foundation model for real-time dialogue. arXiv preprint arXiv:2410.00037.

Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., and Sutskever, I. (2020). Jukebox: A generative model for music. arXiv preprint arXiv:2005.00341.

Donahue, C., Caillon, A., Roberts, A., Manilow, E., Esling, P., Agostinelli, A., Verzetti, M., Simon, I., Pietquin, O., Zeghidour, N., and Engel, J. (2023). SingSong: Generating musical accompaniments from singing. arXiv preprint arXiv:2301.12662.

Donahue, C., Simon, I., and Dieleman, S. (2019). Piano Genie. In *Proceedings of the 24th ACM International Conference on Intelligent User Interfaces (IUI)*, pages 160–164.

Doshi, A. R. and Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):eadn5290.

Engel, J., Hantrakul, L., Gu, C., and Roberts, A. (2020). DDSP: Differentiable digital signal processing. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

European Commission AI Office (2025). Explanatory notice and template for the public summary of training content for general-purpose AI models. Published July 24, 2025.

European Parliament and Council (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act). Article 53 obligations for general-purpose AI providers applicable from August 2, 2025.

Evans, Z., Parker, J. D., Carr, C., Zukowski, Z., Taylor, J., and Pons, J. (2025). Stable Audio Open. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Fairly Trained (2024). Fairly Trained launches certification for generative AI companies that respect creators' rights. Press release, January 17, 2024. `https://www.fairlytrained.org`.

Fiebrink, R., Trueman, D., and Cook, P. R. (2009). A meta-instrument for interactive, on-the-fly machine learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 280–285, Pittsburgh, PA.

Goldmedia (2024). AI and music: Market development of generative AI in the music sector. Technical report, Goldmedia. Study commissioned by GEMA and SACEM. `https://www.gema.de/en/news/ai-study`.

Gui, A., Gamper, H., Braun, S., and Emmanouilidou, D. (2024). Adapting fréchet audio distance for generative music evaluation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

He, W. (2026). Subjective task load and psychological ownership in generative AI collaborative music creation: Mechanisms shaping creators' state sense of agency. *Frontiers in Psychology*, 17(1835406).

Herremans, D., Chuan, C.-H., and Chew, E. (2017). A functional taxonomy of music generation systems. *ACM Computing Surveys*, 50(5):69:1–69:30.

Hoffman, G. and Weinberg, G. (2011). Interactive improvisation with a robotic marimba player. *Autonomous Robots*, 31(2–3):133–153.

Huang, C.-Z. A., Hawthorne, C., Roberts, A., Dinculescu, M., Wexler, J., Hong, L., and Howcroft, J. (2019). The Bach Doodle: Approachable music composition with machine learning at scale. In *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR)*.

Huang, C.-Z. A., Koops, H. V., Newton-Rex, E., Dinculescu, M., and Cai, C. J. (2020). AI Song Contest: Human-AI co-creation in songwriting. In *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, pages 708–716.

Jack, R. H., Stockman, T., and McPherson, A. (2016). Effect of latency on performer interaction and subjective quality assessment of a digital musical instrument. In *Proceedings of the Audio Mostly Conference*, pages 116–123.

Keller, P. E., Novembre, G., and Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B*, 369(1658):20130394.

Kilgour, K., Zuluaga, M., Roblek, D., and Sharifi, M. (2019). Fréchet audio distance: A reference-free metric for evaluating music enhancement algorithms. In *Proceedings of Interspeech*, pages 2350–2354.

Kim, Y., Brade, S., Wang, A., Zhou, D., Kim, H., Wang, B., Lee, S.-J., Flores Garcia, H., Huang, C.-Z. A., and Donahue, C. (2026). A design space for live music agents. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, Barcelona, Spain.

Kim, Y., Lee, S.-J., and Donahue, C. (2025). Amuse: Human-AI collaborative songwriting with multimodal inspirations. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan.

Kraft, J. P. (1994). The "pit" musicians: Mechanization in the movie theaters, 1926–1934. *Labor History*, 35(1):66–89.

Kraft, J. P. (1996). *Stage to Studio: Musicians and the Sound Revolution, 1890–1950*. Johns Hopkins University Press, Baltimore, MD.

Kreminski, M. (2026). Endless forms most similar: The dearth of the author in AI-supported art. *AI & Society*, 41(3):1655–1669.

Krol, S. J., Llano Rodriguez, M. T., and Loor Paredes, M. (2025). Exploring the needs of practising musicians in co-creative AI through co-design. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan.

Lan, Y.-H., Hsiao, W.-Y., Cheng, H.-C., and Yang, Y.-H. (2024). MusiConGen: Rhythm and chord control for transformer-based text-to-music generation. In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, pages 311–318.

Landgericht München I (2025). GEMA v. OpenAI, Judgment of November 11, 2025, Case 42 O 14139/24. Memorization of protected lyrics in model weights held to constitute reproduction. Appeal pending as of July 2026.

Landgericht München I (2026). GEMA v. Suno, Judgment of July 31, 2026, Case 42 O 763/25. Memorization of protected songs in a music generation model held to infringe the reproduction right, text and data mining exception rejected, injunction, disclosure, and damages granted. Not yet final.

Lewis, G. E. (2000). Too many notes: Computers, complexity and culture in *Voyager*. *Leonardo Music Journal*, 10:33–39.

Louie, R., Coenen, A., Huang, C.-Z. A., Terry, M., and Cai, C. J. (2020). Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Lyria Team (2025). Live music models. Google DeepMind. arXiv preprint arXiv:2508.04651.

McCormack, J., Gifford, T., Hutchings, P., Llano Rodriguez, M. T., Yee-King, M., and d'Inverno, M. (2019). In a Silent Way: Communication between AI and improvising musicians beyond sound. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pages 1–11.

Morreale, F. (2021). Where does the buck stop? Ethical and political issues with AI in music creation. *Transactions of the International Society for Music Information Retrieval*, 4(1):105–113.

Morreale, F., Sharma, M., and Wei, I.-C. (2023). Data collection in music generation training sets: A critical analysis. In *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR)*, pages 37–46, Milan, Italy.

Nika, J., Chemillier, M., and Assayag, G. (2017). ImproteK: Introducing scenarios into human-computer music improvisation. *ACM Computers in Entertainment*, 14(2):1–27.

Nistal, J., Pasini, M., Aouameur, C., Grachten, M., and Lattner, S. (2024). Diff-A-Riff: Musical accompaniment co-creation via latent diffusion models. In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, pages 272–280.

Novack, Z., Brade, S., Kim, H., Flores Garcia, H., Shikarpur, N., Talegaonkar, C., Kim, S., Chen, V. K., McAuley, J., Berg-Kirkpatrick, T., and Huang, C.-Z. A. (2026). Live music diffusion models. arXiv preprint arXiv:2605.22717.

Pachet, F. (2003). The Continuator: Musical interaction with style. *Journal of New Music Research*, 32(3):333–341.

Pachet, F., Roy, P., Moreira, J., and d'Inverno, M. (2013). Reflexive loopers for solo musical improvisation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 2205–2208.

Parker, J. D., Spijkervet, J., Kosta, K., Yesiler, F., Kuznetsov, B., Wang, J.-C., Avent, M., Chen, J., and Le, D. (2024). StemGen: A music generation model that listens. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1116–1120.

Pasini, M., Nistal, J., Hayes, B., Bjare, M. R., Lattner, S., and Fazekas, G. (2026). LiveBand: Live accompaniment generation in the audio domain. arXiv preprint arXiv:2606.03803.

Pinch, T. and Trocco, F. (2002). *Analog Days: The Invention and Impact of the Moog Synthesizer*. Harvard University Press, Cambridge, MA.

PMP Strategy and CISAC (2024). Study on the economic impact of generative AI in the music and audiovisual industries. Technical report, International Confederation of Societies of Authors and Composers (CISAC). `https://www.cisac.org/services/reports-and-research/cisacpmp-strategy-ai-study`.

PRS for Music (2026). More creators now worried about AI music competing with human-created music. Press release, February 2, 2026.

Raphael, C. (2010). Music plus one and machine learning. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pages 21–28, Haifa, Israel.

Recording Industry Association of America (2024). Record companies bring landmark cases for responsible AI against Suno and Udio in Boston and New York federal courts, respectively. Press release, June 24, 2024. `https://www.riaa.com/record-companies-bring-landmark-cases-for-responsible-ai-againstsuno-and-udio-in-boston-and-new-york-federal-courts-respectively/`.

Repp, B. H. and Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20(3):403–452.

Rezwana, J. and 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):1–28.

Rottondi, C., Chafe, C., Allocchio, C., and Sarti, A. (2016). An overview on networked music performance technologies. *IEEE Access*, 4:8823–8843.

Rowe, R. (1993). *Interactive Music Systems: Machine Listening and Composing*. MIT Press, Cambridge, MA.

Sawyer, R. K. (2003). *Group Creativity: Music, Theater, Collaboration*. Lawrence Erlbaum Associates, Mahwah, NJ.

Scarlatos, A., Wu, Y., Simon, I., Roberts, A., Cooijmans, T., Jaques, N., Tarakajian, C., and

Huang, C.-Z. A. (2025). ReaLJam: Real-time human-AI music jamming with reinforcement learning-tuned transformers. In *Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., and Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631:755–759.

Silver, D., Schrittwieser, J., Simonyan, K., et al. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359.

Small, C. (1998). *Musicking: The Meanings of Performing and Listening*. Wesleyan University Press, Hanover, NH.

Smith, D. (2025). Beyond the tool: Testing collaborative intelligence in hybrid music performance systems. In *Most Wanted: Music Research (MW:MR 2025)*, Berlin, Germany. `https://mwm.pubpub.org/pub/po0x4gga`.

Sousa, J. P. (1906). The menace of mechanical music. *Appleton's Magazine*, 8:278–284.

Sturm, B. L. T., Ben-Tal, O., Monaghan, Ú., Collins, N., Herremans, D., Chew, E., Hadjeres, G., Deruty, E., and Pachet, F. (2019). Machine learning research that matters for music creation: A case study. *Journal of New Music Research*, 48(1):36–55.

Tatar, K. and Pasquier, P. (2019). Musical agents: A typology and state of the art towards musical metacreation. *Journal of New Music Research*, 48(1):56–105.

Théberge, P. (1997). *Any Sound You Can Imagine: Making Music/Consuming Technology*. Wesleyan University Press, Hanover, NH.

Universal Music Group and Udio (2025). Universal Music Group and Udio Announce Udio's First Strategic Agreements for New Licensed AI Music Creation Platform. Press release, October 29, 2025.

U.S. Copyright Office (2025). Copyright and artificial intelligence, part 2: Copyrightability. Technical report, United States Copyright Office. Released January 29, 2025. `https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf`.

U.S. Court of Appeals for the D.C. Circuit (2025). Thaler v. Perlmutter, No. 23-5233. Decided March 18, 2025. Human authorship required for copyright. Certiorari denied March 2, 2026.

Valevski, D., Leviathan, Y., Arar, M., and Fruchter, S. (2025). Diffusion models are real-time game engines. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Vercoe, B. (1984). The synthetic performer in the context of live performance. In *Proceedings of the 1984 International Computer Music Conference*, pages 199–200, Paris, France. Computer Music Association.

Wang, Z., Liang, Q., Zhang, K., Wang, Y., Zhang, C., Yu, P., Feng, Y., Liu, W., Wang, Y., Bao, Y., and Yang, Y. (2022). SongDriver: Real-time music accompaniment generation without logical latency nor exposure bias. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1057–1067.

Warner Music Group and Suno (2025). Warner Music Group and Suno Settlement and Licensing Agreement. Announced November 25, 2025.

Werner, K. J. (2015). The Roland TR-808 and the tale of the marching anteaters. *Ethnomusicology Review*. Sounding Board, UCLA, November 29, 2015.

Werner, K. J., Abel, J. S., and Smith, J. O. (2014). A physically-informed, circuit-bendable, digital model of the Roland TR-808 bass drum circuit. In *Proceedings of the 17th International Conference on Digital Audio Effects (DAFx-14)*, pages 159–166, Erlangen, Germany.

Wessel, D. and Wright, M. (2002). Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22.

Williamson, J. and Cloonan, M. (2016). *Players' Work Time: A History of the British Musicians' Union, 1893–2013*. Manchester University Press, Manchester, UK.

Wu, S.-L., Donahue, C., Watanabe, S., and Bryan, N. J. (2024a). Music ControlNet: Multiple time-varying controls for music generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:2692–2703.

Wu, Y., Cooijmans, T., Kastner, K., Roberts, A., Simon, I., Scarlatos, A., Donahue, C., Tarakajian, C., Omidshafiei, S., Courville, A., Castro, P. S., Jaques, N., and Huang, C.-Z. A. (2024b). Adaptive accompaniment with ReaLchords. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*.

Wu, Y., Wang, M., Lei, H., Brade, S., Blanchard, L., Wu, S.-L., Courville, A., and Huang, C.-Z. A. (2025). Streaming generation for music accompaniment. arXiv preprint arXiv:2510.22105.