The Selection Behind the Penalty: Strategic and Mandatory AI Disclosure in Music Streaming

Xi Kang[*], Yahe Tan[*], Minxia Chen, Wanshu Niu, and Liqiang Huang

August 27, 2026

Abstract

Generative AI increasingly assists the creation of cultural goods, confronting creators with a strategic decision: whether to disclose that a work was made with AI. Prior research suggests that audiences discount creative work when they learn it was produced with AI, which makes disclosure costly for creators; yet regulators increasingly require such involvement to be labeled. How do artists decide whether to reveal their AI use, and how much of the performance gap between disclosed and concealed work reflects how audiences respond to the AI label rather than the artists' choice of which work to reveal? We study a leading music-streaming platform whose integrated AI tool let artists choose whether to credit AI, before a platform-wide mandate removed that choice. Under voluntary disclosure, concealed AI songs outdraw the same artist's credited songs roughly threefold, but this gap is not driven by a penalty alone: artists disclose selectively, crediting AI on songs that enter new genres or feature a more visible human contribution and concealing it elsewhere. Exploiting the mandate in a difference-in-differences design, we estimate the causal listener penalty at about five percent of plays, concentrated at discovery and borne mainly by lower-status artists. Much of the voluntary gap therefore reflects selection rather than audience aversion, and the raw gap substantially overstates the market cost of mandated disclosure.

[*]Xi Kang and Yahe Tan contributed equally (co-first authors). Xi Kang: Owen Graduate School of Management, Vanderbilt University, xi.kang@vanderbilt.edu; Yahe Tan: School of Management, Zhejiang University, yahe.tan@zju.edu.cn; Minxia Chen: INSEAD, minxia.chen@insead.edu; Wanshu Niu: Ningbo University, niuwanshu@nbu.edu.cn; Liqiang Huang: School of Management, Zhejiang University, huanglq@zju.edu.cn. The music-streaming platform and its integrated AI tool are referred to throughout by the pseudonyms XMusic and Harmony.ai. We thank J.P. Eggers and Brian Wu for their friendly review and thoughtful feedback. All errors are our own.

1 Introduction

Generative artificial intelligence is increasingly used to assist the creation of cultural goods (Bick et al. 2024), from images (Zhou and Lee 2024) and prose (Doshi and Hauser 2024) to music (International Music Summit 2025), advertising (Grewal et al. 2025), and film (TechCrunch 2025). Recent evidence shows that these tools augment human performance in creative ideation (Boussioux et al. 2024), generate and evaluate early-stage business plans (Csaszar et al. 2024), and deliver sizable productivity gains in knowledge work (Cui et al. 2024, Dell'Acqua et al. 2023), and their use in cultural markets is already substantial: the diffusion of large language models tripled new book releases between 2022 and 2025, with AI-containing titles exceeding half of 2025 releases (Reimers and Waldfogel 2026), and by late 2025 fully AI-generated tracks reached 28 percent of the music uploaded daily to one major streaming service (Deezer 2025a). As creators adopt these tools, they face a strategic decision: whether, and in what form, to communicate to audiences that a work was produced with AI assistance.

The strategic decision to disclose AI assistance is far from straightforward. Audiences increasingly expect transparency about how creative work is made, yet a substantial behavioral literature documents algorithm aversion, whereby audiences devalue creative work once they learn it was made with AI, an effect more pronounced for the subjective, hedonic judgments that characterize cultural goods (Castelo et al. 2019a, Jussupow et al. 2020, Raj et al. 2026, Benegal et al. 2026). This decision is increasingly shaped by regulation: within roughly two years, mandatory AI-content labeling has moved from proposal to law across the largest media markets, requiring greater transparency about AI involvement (Cyberspace Administration of China 2025, European Parliament and Council of the European Union 2024, YouTube 2024, Spotify 2025).[1] These developments motivate the two questions we take up. First, how do artists decide whether to reveal their AI use? Second, when disclosure affects a work's market performance, how much of that effect reflects how audiences respond to the AI label, and how much reflects the creator's strategic decision about which work to reveal?

Two literatures speak to these questions, and each was built to answer a different part. A large behavioral literature establishes, in the laboratory, that audiences devalue creative work once they

[1] These rules are difficult to enforce in full. AI-generated content cannot yet be detected with full accuracy, and existing detection tools produce both false positives and false negatives (Weber-Wulff et al. 2023, Liang et al. 2023, Sadasivan et al. 2023). Mandatory labeling is therefore unlikely to be fully implemented in the near term, leaving creators some latitude over what audiences learn. OpenAI, for instance, withdrew its own AI-text classifier in 2023 for its low rate of accuracy (OpenAI 2023).

are told it involved AI, an effect driven by perceived authenticity and difficult to reframe away (Raj et al. 2026, Benegal et al. 2026). By assigning disclosure at random to a fixed work, these experiments identify the audience's causal response to a label with a cleanness no field study can match; they answer, in the laboratory, whether disclosing AI assistance costs a creator. That same randomization, by design, holds constant the selection that operates when disclosure is chosen rather than assigned, so it sets aside the strategic question of which work a creator would have revealed.

A newer field literature studies real disclosure decisions in live markets, where the lab's clean assignment gives way to creators' own choices. Studies of crowdfunding and digital games observe creators' stated positions on AI or their disclosure choices and document the performance gap between disclosed and concealed work (Heshmati and Wang 2026, Li 2026). This evidence captures behavior the laboratory cannot: how disclosure plays out where creators decide for themselves and audiences act with real stakes. It speaks to whether disclosing AI costs artists in the field, but only in part, because without a counterfactual on the same work its cross-sectional gaps combine the audience's response to a label with the selection of which work creators choose to reveal, as these authors themselves note. Each literature thus answers the question it was built for, and together they set up ours: in a real market, the response and the selection have not yet been told apart.

Our setting shows the same field pattern, and lets us see why it is hard to read. When artists are free to choose, those who disclose AI do worse: on the large music-streaming platform we study, a song that used AI but concealed it outdraws the same artist's AI songs that credited the tool by roughly a threefold margin. That gap is the penalty a disclosure mandate is feared to impose. But it is difficult to interpret, because the observed gap between disclosed and concealed work blends two very different things: how listeners respond to a label, and which songs artists choose to reveal in the first place. Separating the two is a first-order question for theory and policy alike, and it is difficult: it requires observing an artist's true AI use, including what was concealed, together with a change in disclosure that does not depend on the artist's choice.

We take on this challenge by studying XMusic, a major music-streaming platform with an integrated generative-AI music tool, Harmony.ai (Figure 1). For more than two years, disclosure was voluntary: an artist who made a song with Harmony.ai could credit the tool to listeners or stay silent, and many stayed silent. Then, on September 18, 2025, the platform began attaching a mandatory "AI-assisted" label to eligible AI songs, removing the choice. Two features of the setting make the separation possible. First, we observe ground-truth AI use from the platform's back end, including songs that used AI but were never disclosed to listeners, so we see concealment directly

rather than inferring it from what artists say. Second, we can hold the work fixed: the same audio, the same artist, the same song, observed both before and after a label is imposed on it.

The two regimes do different jobs. The voluntary regime reveals the selection: we model which AI songs artists choose to credit, and we measure the raw conceal-versus-credit performance gap within each artist's own catalog. The mandatory regime supplies the missing counterfactual. Because the label applied only to AI songs published on or after an April 2025 eligibility cutoff, otherwise-similar concealed songs published just before the cutoff were never labeled, while those published just after were labeled at the September shock. Comparing the two cohorts through the shock in a difference-in-differences design recovers the causal listener response with the song held fixed. This is the field counterpart of the laboratory estimate, now with the selection margin observed rather than assumed away.

Four findings follow. First, adoption is consequential: artists who adopt Harmony.ai sharply increase their output relative to matched non-adopters (Section 4), confirming that the tool materially changes production and that the population making disclosure choices is large. Second, disclosure is strategic rather than idiosyncratic: crediting tracks the song and the artist in interpretable ways, with songs that enter a genre new to the artist, carry a larger production team, or feature a more visible human contribution more likely to credit the AI, while songs in the artist's habitual genres are concealed (Table 3). Third, the raw voluntary gap is large: within an artist's own catalog, concealed AI songs outperform credited ones by roughly a threefold margin in plays (Table 4). Fourth, when the mandate forces disclosure onto a previously concealed song, the effect is confined to discovery: fewer listeners begin the song and its plays fall by about five percent, while completion and post-listening engagement are unchanged (Table 5). This penalty falls unevenly, concentrated among lower- and middle-tier artists (Table 6).

Setting the two regimes side by side answers the question in the paper's title. Even taking the causal estimate at its largest and the voluntary gap at face value, the genuine listener response accounts for at most a few percent of the observed conceal-credit gap, and the overwhelming remainder is selection (Section 7). The raw voluntary comparison overstates the causal penalty of disclosure by more than an order of magnitude.

Our primary contribution is to open up the strategic disclosure decision and its market consequences, a margin that prior work has had to hold fixed or leave folded into a cross-sectional gap. By observing true and concealed AI use alongside an exogenous labeling shock on the same work, we separate, in a real market, the audience's response to a label from the artists' selection of what

to reveal, and show that the selection dominates. This complements the evidence we build on rather than competing with it: the laboratory studies identify the clean causal response by holding the work fixed and assigning disclosure at random, and the field studies document how disclosure choices play out in live markets. Our design adds the piece their questions did not require, the concealment itself together with a within-work label, and in doing so shows how the disclosure penalties documented across this fast-growing literature can be read together, as compounds of listener response and artist selection whose mix each design is positioned to reveal in part.

A second, conceptual contribution follows from connecting the economics of voluntary disclosure (Grossman 1981, Milgrom 1981) to the behavioral evidence on AI aversion (Raj et al. 2026, Benegal et al. 2026): we show why the classic prediction of unraveling toward full disclosure fails, and reverses, for AI authorship, so that silence rather than revelation is the equilibrium and voluntary disclosure sorts songs rather than informing listeners. These results also carry a direct policy reading. Mandatory AI labeling is debated as though it levied a heavy tax on the artists who use these tools, a presumption the raw disclosure gap appears to license; our estimates suggest the cost of a mandatory label is modest rather than ruinous, that it operates at discovery rather than in how listeners engage with a song they have chosen to play, and that it is regressive, falling on the newer and smaller artists whose audiences have the least other information and near zero for the established. As transparency rules take effect across the largest media markets, they should be judged against the penalty disclosure actually imposes, not the far larger gap that artists were avoiding.

The paper proceeds as follows. Section 2 develops the conceptual framework that links disclosure economics to the AI-authenticity penalty. Section 3 describes the platform, the tool, and the two regulatory regimes, and Section 4 the data, samples, and the adoption fact. Section 5 analyzes disclosure choice and the raw performance gap under the voluntary regime, and Section 6 estimates the causal label penalty under the mandatory regime. Section 7 bridges the two to quantify how much of the gap is selection, and Section 8 concludes.

2 Related Literature and Conceptual Framework

Our study sits at the intersection of two literatures that rarely meet: the behavioral work on how audiences react to artificial intelligence in creative markets, and the economics of voluntary and mandated disclosure. The first tells us how consumers respond when they *learn* that a creative

good involved AI; the second tells us when producers *choose* to reveal private information, and when a regulator can improve on that choice by compelling it. Read together they yield the paper's central conceptual claim. Because the audience reaction the first literature documents is both real and difficult to escape, the disclosure decision the second literature studies becomes a problem of strategic concealment; and the disclosure penalty one observes in any field market is therefore a compound of a genuine audience response and the selection of who chooses to disclose. Prying those two apart is the task our two-regime design is built for, and organizing why they must be pried apart is the task of this section.

2.1 The Artificial Intelligence Disclosure Penalty and Its Mechanism

A fast-growing experimental literature documents an *AI disclosure penalty*: audiences evaluate a creative work less favorably when told it was produced by, or with the help of, artificial intelligence, even when the work itself is held fixed. The most comprehensive demonstration is Raj et al. (2026), who across sixteen preregistered experiments (N = 27,491) find that disclosing AI involvement consistently lowers evaluations of creative writing, with a meta-analytic effect of roughly a third of a standard deviation. The pattern is not specific to text: comparable penalties appear for paintings and visual art (Bellaiche et al. 2023, Horton et al. 2023) and for AI-labeled artistic production more broadly (Jago 2019, Tigre Moura et al. 2023), and it arises even though people frequently cannot tell AI- from human-authored work apart when they are not told which is which (Köbis and Mossink 2021). The penalty is thus a reaction to the *label*, not to any detectable deficiency in the work.

The two leading experimental studies converge on a single mechanism: perceived authenticity. Raj et al. (2026) find that diminished authenticity *fully* mediates the penalty—once it is controlled, the direct effect of disclosure is statistically indistinguishable from zero—and Benegal et al. (2026), whose stimulus is an AI-disclosed musical composition, reach the same conclusion in a design close to our own setting. This convergence matters for us twice over. It locates the penalty in a construct with deep roots in cultural markets, where audiences are known to prize the belief that a work issues from genuine human expression and to discount works that seem to violate it (Newman and Bloom 2012, Jago 2019, Askin and Mauskapf 2017). And because authenticity, as these authors operationalize it—"genuine, internally driven" creative intent—subsumes the creator's own agency, it absorbs the most natural rival account, that audiences dock AI-assisted work because the human contribution is diluted rather than because of the label as such.

The penalty is one instance of the broader phenomenon of algorithm aversion, the tendency to

discount algorithmic outputs relative to human ones (Dietvorst et al. 2015, Longoni et al. 2019). But aversion is not universal, and its scope conditions tell us where a disclosure penalty should be largest. For objective, utilitarian judgments people often display the opposite tendency—algorithm *appreciation*, preferring algorithmic to human advice (Logg et al. 2019)—and aversion sharpens as tasks become more subjective, more human, and more hedonic (Castelo et al. 2019b, Longoni and Cian 2022). Music is the paradigmatic hedonic, authenticity-laden good: its consumption is affective and experiential, and a song's value rests heavily on the listener's sense of its human origins. Our setting is therefore the case in which the disclosure penalty should be *most* pronounced—a scope condition we return to when interpreting the size of our estimates.

A striking feature of this evidence is how *sticky* the penalty is. Raj et al. (2026) deploy, and fail to neutralize the penalty with, essentially the full menu of interventions known to soften algorithm aversion elsewhere: humanizing or anthropomorphizing the tool, boosting beliefs about its capabilities, and— most relevant here—framing the work as human–AI *collaboration* rather than pure automation. Collaboration labels offered no relief; once AI involvement was disclosed, audiences discounted the work regardless of how the human contribution was emphasized. The implication for producer behavior is direct and consequential: if the penalty cannot be talked down through framing, the only reliable way for a creator to avoid it is not to disclose at all.

Two caveats bound this literature, and both point toward the gap our design fills. First, the penalty is contingent rather than universal: a minority of studies find that AI disclosure can *raise* some evaluations—by signaling technological modernity, or by lifting engagement even as perceived quality falls—so its sign depends on the outcome and the domain. We take music consumption to be squarely in the unfavorable, authenticity- dominant case, and treat the point as a scope condition rather than a threat. Second, and more fundamentally, nearly all of this evidence is experimental, with disclosure *randomly assigned* to a fixed stimulus judged by an uninvolved evaluator. Random assignment is exactly what gives these studies their clean identification of the audience's causal response to a label; it is also exactly what renders them silent on the question a real market forces—what happens when disclosure is not assigned but *chosen*. In this precise sense Raj et al. (2026) is the laboratory twin of our own mandatory-regime estimate: both isolate the pure listener response to a label with content held fixed, one by randomizing disclosure and the other by an administrative rule that imposes it. The difference is that a randomized design cannot, even in principle, observe the selection that governs disclosure when it is voluntary.

2.2 Disclosure, Unraveling, and Strategic Concealment

When disclosure is a choice, the natural benchmark is the economics of voluntary disclosure, whose classic prediction is *unraveling*. If a producer holds private information about a verifiable attribute and disclosure is costless, full revelation is the unique equilibrium (Grossman 1981, Milgrom 1981). The logic is iterative: the highest type discloses to separate itself, its departure lowers the audience's estimate of those who stay silent, that inference induces the next-highest type to disclose, and so on until silence itself is damning and even mediocre types reveal. Later work shows the result is fragile, breaking down when disclosure is costly, when the audience is unsure whether the sender is even informed, or when the attribute is not verifiable (Dye 1985, Verrecchia 2001, Healy and Palepu 2001). AI authorship violates these conditions in three mutually reinforcing ways, and—crucially—all three push in the same direction, away from disclosure.

First, AI involvement is *not verifiable* to the audience. Listeners typically cannot detect it unaided (Köbis and Mossink 2021), so a creator who stays silent is not reliably exposed, and audiences often cannot even condition their beliefs on the fact of non-disclosure. Second, and more fundamentally, the valence of the private information is inverted relative to the canonical model. There, disclosure conveys *good* news that high types are eager to certify; here, as Section 2.1 establishes, disclosure conveys *bad* news—a sticky penalty that no AI-using type escapes and none benefits from volunteering. With no type that gains by revealing, the engine that drives unraveling never turns over. Third, the disclosed attribute is a stigmatizing binary, not a continuous quality index a producer might use to separate from a marginally worse rival. The equilibrium these forces select is not revelation but *concealment*: creators who used AI stay silent by default, and disclosure occurs only among the subset with an offsetting private reason to reveal.

It follows that, under a voluntary regime, disclosure is a *selection device*. The creators who credit AI are systematically different from those who conceal it—in the songs they choose to credit, in why they credit them, and in the audiences they expect to reach. Consequently the disclosure penalty one reads off field data—the raw gap between the performance of disclosed and concealed AI work—is *not* the audience's causal response to the label. It is that response compounded with the selection of who discloses what. This selection margin is, by design, outside the reach of the experimental literature, whose randomization delivers a clean causal response precisely by holding selection constant. It also sits just beyond the emerging field literature: contemporary studies of AI disclosure in crowdfunding (Heshmati and Wang 2026) and in digital games (Li 2026) observe real market behavior the lab

cannot, yet they measure creators' *stated* positions or disclosure choices rather than their true, potentially concealed, use, and, without randomization or an observed counterfactual on the same work, they report disciplined associations that combine response with selection—as their authors carefully note. The two literatures are thus complementary: the behavioral work identifies the causal response with content fixed, the field work captures disclosure choices in live markets, and each answers the question it was designed to answer. What neither is positioned to observe—and what our two-regime design adds—is the concealment itself, together with an exogenous label on the same work, the combination that decomposing a real market's penalty requires.

This is where mandated disclosure enters, and where its role in our design departs from its usual one. In the standard treatment a mandate matters because it supplies information the market failed to reveal, thereby shifting beliefs and behavior (Jin and Leslie 2003, Mathios 2000, Granja 2018). In our setting it does that too, but its analytical value is identification: by compelling disclosure onto works that were previously concealed, holding those works fixed, a mandate removes the selection margin and exposes the audience's causal response *in the field*—the market counterpart of the experimental estimate. The same logic carries a distributional prediction. Because a mandate binds only where information was being withheld and audiences had no other way to infer it, its effect should concentrate where listeners' prior information about the creator is thinnest, and fade where an established reputation already substitutes for the missing signal.

2.3 Conceptual Framework: Separating Response from Selection

These threads combine into a simple framework that organizes the empirical analysis; we state it verbally rather than formally. Consider an artist who has used AI on a given song and privately knows both the song's attributes—its quality, and how novel or experimental it is—and her own standing with her audience. If she credits the AI she incurs the authenticity penalty of Section 2.1, which she cannot reframe away. She therefore credits only when a private benefit of disclosure outweighs that penalty—when crediting plausibly signals novelty or experimentation to an audience that rewards it, say, or when her reputation is deep enough to absorb the hit—and conceals otherwise. Two implications follow immediately. Disclosure is *selected*: whether a song is credited co-varies systematically with the song's and the artist's attributes. And the observed voluntary penalty *overstates* the causal one, by the magnitude of that selection. A disclosure mandate, by removing the choice on the songs it covers, identifies the causal response with content held fixed.

The framework yields three claims that structure the analysis. First, disclosure is *strategic*

rather than idiosyncratic: crediting should track song- and artist-level attributes in interpretable ways—more likely where AI use can be read as novelty, or where a substantial and visible human contribution cushions the authenticity hit, and concealment concentrating where the human hand is least evident (Section 5.1). Second, the causal penalty is *far smaller* than the raw voluntary gap, and the wedge between them is selection rather than listener response (Sections 5.2, 6, and 7). Third, the causal penalty is concentrated where information is the binding channel: felt at *discovery*, where a listener decides whether to begin a song, rather than in post-consumption engagement, and appears concentrated among lower- and middle-tier artists, whose audiences hold the thinnest priors (Sections 6 and 6.3).

That last claim also marks where our field evidence and the laboratory part company, in a way the two settings themselves predict. The experimental work finds no reputational buffer—reputation does not soften the penalty in a vignette (Benegal et al. 2026)— whereas we expect, and find suggestive evidence, that it does. The difference is informational: a vignette gives the evaluator nothing to weigh but the disclosure itself, while a listener choosing among familiar artists brings a catalog and a set of priors to the decision, so a label moves behavior only where those priors are thin. We develop this reconciliation, and what it implies for a field-and-lab literature that has reached conflicting conclusions, in Section 7. Throughout, it bears repeating that music is the domain in which the authenticity penalty should be *largest*; a causal effect that turns out modest even here is, if anything, an upper bound on what disclosure mandates cost creators in less authenticity-sensitive markets. Our question is deliberately not whether AI makes better or worse songs—a separate and active debate (Doshi and Hauser 2024, Zhou and Lee 2024)—but what happens when its involvement is revealed, holding the song fixed.

3 Empirical Setting

3.1 The Platform and the Integrated AI Tool

Music streaming platform. We partnered with XMusic, a publicly listed music streaming platform in China, to study artists' AI disclosure behavior and its performance implications. XMusic hosted more than 200 million monthly active listeners and over 40 million paying subscribers as of the end of 2023, with a catalog exceeding 60 million licensed tracks. What distinguishes XMusic from its primary competitor is its position as China's largest community for independent artists: by the end of 2022, prior to the platform's integration of AI music creation capabilities, more than

611,000 platform-certified independent artists had uploaded original work to the platform, a figure that grew to approximately 684,000 artists contributing over 3.1 million tracks by the end of 2023. This artist-centric character makes XMusic a uniquely appropriate site to study how AI tools reshape artists' creative production strategies.

Generative AI system. In July 2023, XMusic integrated a proprietary generative AI system, which we refer to as Harmony.ai, directly into its artist creation and publishing workflow, making it available to all platform-certified independent artists. Two features of Harmony.ai's design and timing make XMusic a clean setting for studying AI-assisted creative production.

First, Harmony.ai is an augmentation tool built for working artists rather than a consumer-facing application aimed at casual listeners. Unlike end-to-end generators that produce a finished song from a single text prompt, Harmony.ai supports the full production pipeline—lyric writing, melody composition, chord arrangement, instrument selection, AI vocal synthesis, and final mixing—and exposes fine-grained controls (chord progressions, key, tempo, and compositional input supplied as text, mood keywords, lyrics, chord notation, or melody MIDI) that presuppose musical literacy.[2] Crucially, Harmony.ai is built for iterative collaboration: an artist can work back and forth with the system, revising a specific element such as a chord, a vocal line, or an arrangement choice while holding the rest of the composition fixed. This stands in contrast to end-to-end text-to-music models such as Suno, which struggle to make targeted changes to a particular aspect of a track in response to follow-up instructions. Harmony.ai is therefore a productivity and creativity tool that rewards musical expertise rather than one that substitutes for it.

Second, although Harmony.ai is not the only augmentation tool of its kind—comparable systems such as Google's Lyria have since emerged, which helps our findings generalize beyond a single product—it was the dominant tool available to artists on XMusic during our first observation window. The professional toolset central to our study became available to artists within the XMusic creation workflow on July 1, 2023,[3] predating the public release of the leading general-purpose music generators.[4] Because experienced Chinese artists had no widely available substitute during the

[2]Harmony.ai offers more than 450 arrangement templates spanning diverse genres and moods, multi-track export, and editable pitch, vocal style, and arrangement parameters. These professional-grade affordances also impose a meaningful barrier to entry: producing a polished track requires the user to make compositional decisions a trained artist would recognize, in contrast to one-click consumer tools where the user supplies only a natural-language prompt.

[3]Harmony.ai first reached the public earlier, in January 2022, as a lightweight mini-program that produced one-tap personalized songs (for example, customized New Year greetings) for a general consumer audience. That earlier debut differed substantially in both functionality and intended users from the mid-2023 release studied here, which embedded a full professional composition toolset directly into the artist creation workflow on XMusic.

[4]Suno, the most prominent consumer text-to-music system, ran only as a semi-private beta in mid-to-late 2023

voluntary regime, and because Harmony.ai usage is recorded directly in XMusic's back end, we can cleanly observe which voluntary-regime songs were AI-assisted and how artists chose to disclose that fact to listeners, with little risk that the analysis is contaminated by AI music produced through other systems. Under the mandatory regime, by contrast, the analysis incorporates the full range of AI tools then available to artists.

Finally, Harmony.ai is embedded within XMusic's publishing infrastructure: songs composed with the tool flow through the same creation-to-monetization pathway as human-composed songs, distributed through the artist dashboard, earning streaming royalties, and receiving platform copyright registration.

3.2 Two Regulatory Regimes: From Voluntary Credit to Mandatory Label

Our identification rests on a shift between two disclosure regimes on the same platform, separated by a national regulation requiring that AI-assisted online content be labeled. Under the first, *voluntary* regime—opened by the July 2023 integration of Harmony.ai and running until September 17, 2025, the day before XMusic complied with the new rule—artists decided for themselves whether to credit AI involvement to listeners. Under the second, *mandatory* regime, XMusic complied by displaying a public "AI-assisted" label on qualifying songs beginning September 18, 2025, removing that choice.

A distinction that runs through the entire setting is between an artist's *platform-side declaration*—what the artist tells XMusic's back end about a song—and *listener-facing disclosure*—what appears in the song's public metadata that listeners actually see. During the voluntary regime, XMusic displayed no AI tag of any kind. Listeners could infer AI involvement only from the public credit list: songs that listed Harmony.ai as a credited collaborator signaled AI to the many users who recognized the heavily promoted tool, whereas songs that omitted the credit did not. The strategic listener-facing choice was therefore whether to include the Harmony.ai credit. We classify an AI-assisted song as *credited* if its public credit named Harmony.ai and *concealed* otherwise.[5] Separately, artists had a private incentive to declare Harmony.ai use to the platform: doing so secured full copyright and a 100% revenue share on the track. Among the 21,718 AI-assisted songs released

and did not reach general public availability until its web application launched on December 20, 2023; Udio, its closest competitor, launched in April 2024; and Google's Lyria, the closest augmentation-oriented analog, was announced on November 16, 2023. All postdate Harmony.ai's mid-2023 integration into XMusic, and the Western systems additionally faced accessibility constraints for users in China.

[5]Because platform-side declaration and listener-facing disclosure are distinct, we observe *single-concealed* songs—AI use declared to the platform but withheld from listeners—but not *double-concealed* songs, where AI use is hidden from both. Our concealment measure is thus a lower bound on true concealment; Section 8 discusses the implication.

by Harmony.ai adopters (Section 4.1), about 61% were concealed and 39% credited—a split with ample variation on both sides.

The first change came on April 15, 2025. On that date the platform upgraded its back-end declaration channel: at upload, artists were now required to report whether the song was made with *any* AI tool—not only Harmony.ai—and, if so, to name the tool used.[6] This was a *supply-side* change. It altered what artists had to do when publishing a song, but listeners saw nothing new; the declaration was recorded internally and remained invisible on song pages. Appendix A details the upload-interface change, how we date it, and the back-end evidence that artists declared truthfully.

The second change—the event our causal design exploits—came on September 18, 2025, when XMusic complied with the labeling regulation by publicly displaying the "AI-assisted" label on song pages. This was a *demand-side* change: the songs themselves were unchanged; only their public labeling was. The regulation was enforced prospectively. Any song reported to the platform's back end as AI-assisted and *published on or after* April 15, 2025 received the "AI-assisted" label when public display began on September 18, 2025; AI-assisted songs published *before* April 15, 2025 did not. Publication date relative to April 15, 2025 is therefore the single feature that determines whether an otherwise-comparable song was labeled. We exploit this threshold, comparing songs published in the roughly six weeks on either side of it (March 1–April 14 versus April 15–May 31, 2025).

Both the treatment and control groups consist of AI-assisted songs whose artists reported the AI assistance to the platform's back end but did not credit it to listeners—that is, concealed AI songs. The **treatment group** comprises 5,593 such Harmony.ai songs published in the six weeks after the threshold (April 15, 2025–May 31, 2025); all received the public label on September 18, 2025. The **control group** comprises 4,246 Harmony.ai songs published in the six weeks before it (March 1–April 14, 2025); because they predate the threshold, they were never eligible for the label and remained unlabeled throughout. We restrict both groups to Harmony.ai songs so that treatment and control are the same kind of AI-assisted song, differing only in publication timing.[7]

The control group's value rests on the exogeneity of its membership. A song belongs to it purely

[6]The platform warned artists that a song its internal AI-detection algorithm found to contain undeclared AI assistance could be withheld from publication pending further verification, with repeat or serious cases exposing the account to suspension. This deterrent underlies the high declaration compliance documented in Appendix A.

[7]The back-end declaration introduced on April 15, 2025 covers all AI tools, and the labeled AI songs on the platform were produced with many different systems. We restrict the estimation sample to Harmony.ai songs to hold the production technology fixed across treatment and control. Songs that both reported AI assistance to the platform and credited it to listeners—credited AI songs—in the same publication windows are the basis of our falsification test (Section 6).

by virtue of having been published before April 15, 2025—a platform-wide calendar cutoff imposed by the regulation, not a boundary any artist chose. Treatment and control artists are otherwise behaviorally alike: both affirmatively chose to create with Harmony.ai. The one systematic difference is lifecycle—treatment songs are somewhat younger when public labeling begins—which we absorb with song fixed effects and a publication-age control (Section 6). Appendix A records the back-end detail that distinguishes the two groups.

3.3 Why the Recommender Is Not the Treatment

A natural concern in any streaming setting is that estimated effects reflect the platform's recommendation algorithm rather than listener behavior—that a labeled or credited AI song is consumed less because the recommender surfaces it less. This concern does not apply here, in either regime. Per the platform's engineering team, XMusic's ranking policy is *disclosure-neutral* throughout our sample.[8] During the voluntary regime, the recommender applied no differential discount to AI songs on the basis of disclosure: credited and concealed AI songs were ranked alike. During the mandatory regime, the recommender applied a *uniform* discount to all AI songs regardless of their public label. The platform introduced this discount on January 1, 2025, well before public labeling began, and left it unchanged both before and after September 18, 2025. In neither period does the ranking policy respond to the disclosure margin we study.

Two implications follow. First, the voluntary-regime concealed–credited performance gap (Section 5.2) contains no differential-algorithm channel; it reflects listener response and artist selection only. Second, because the recommender treats labeled and unlabeled AI songs identically before and after September 18, 2025, the September 18, 2025 difference-in-differences isolates the listener-information channel—the effect of listeners *seeing* the label. This is corroborated within our own estimates: a falsification test on songs that were already publicly crediting AI before the mandate finds no effect of the label (Section 6), which is inconsistent with a differential algorithmic discount switching on at September 18, 2025.

We flag one caveat, developed in Section 8: policy-neutrality is not the same as outcome-neutrality. An engagement-optimizing recommender can still *amplify* an initial listener response through feedback, even without a disclosure-specific rule. Our mandatory-regime estimates price this in—the penalty deepens gradually over the weeks following the label rather than jumping discretely—while for the voluntary regime it folds into a broader equilibrium caveat on interpreting

[8] Based on direct communication with the collaborating firm's engineering team, 2026-07-20.

that gap.

4 Data, Samples, and the Voluntary-Regime Adoption Fact

Our data come from XMusic's internal records, observed through a snapshot taken in January 2026 and covering the full catalog of songs released on the platform together with the artists who created them and listeners' interactions with each song. From this corpus we draw two analysis samples, one per regime, constructed differently because each regime answers a different question through a different source of identifying variation. This section describes the two samples and the outcomes, presents summary statistics, and establishes one descriptive fact that motivates everything that follows: once artists gain access to Harmony.ai, the market fills with AI-assisted songs.

4.1 The Voluntary-Regime Sample

This paper is about how artists decide whether to credit AI and how that choice shapes a song's performance. Assessing performance requires a benchmark: how do AI-assisted songs—whether or not the AI is credited—fare relative to comparable non-AI songs? Our primary benchmark is *within artist*: we compare each adopter's AI-assisted songs against the non-AI songs released by the same adopter, holding fixed everything about the artist—skill, following, and audience—so that the comparison isolates AI and its disclosure rather than the kind of artist who uses AI. We therefore build the voluntary-regime sample around the population of Harmony.ai adopters and the songs they release. As a robustness check we also construct a matched sample of comparable non-adopters and benchmark adopters' AI songs against *their* non-AI songs (Appendix B); we keep this across-artist comparison out of the main analysis to avoid confronting selection into adoption, a question we take up separately in future work.

This benchmark comparison is meaningful only in the voluntary regime. As Section 3.3 explains, the platform's recommender applied no discount to AI-assisted songs during the voluntary-regime window, so their performance can be compared directly against non-AI songs. By the mandatory regime the recommender discounted all AI-assisted songs uniformly, regardless of disclosure, which would confound any comparison against non-AI songs—and is why our mandatory-regime design contrasts AI songs with one another, labeled versus unlabeled, rather than against non-AI songs.

Adopters. We begin from the platform's roster of certified independent artists and isolate those who adopted Harmony.ai after its July 2023 launch. We exclude virtual artists and the platform's own AI accounts, artists who had released AI-assisted songs before the launch date, and strategic manipulators and pure experimenters, and we require more than two released songs as of the launch date. To keep the adoption event clean, we retain only artists whose first adoption falls between July 2023 and March 2024, dropping later adopters to avoid the structural break created by the spring-2024 public release of Suno's third-generation model, which sharply lowered the barrier to AI music creation and drew in a different cohort of users. Finally, we drop adopters who released fewer than five songs before adopting or who had no releases in the three months immediately preceding it, retaining artists with sustained creative activity rather than one-off or experimental users. This yields 1,932 Harmony.ai adopters, the population for our within-artist analysis.

The AI-assisted songs. The disclosure analyses focus on the AI-assisted songs the adopters released. These artists released 21,718 AI-assisted songs in total (Table 1), about 61% concealed from listeners and 39% credited, following the credit-based classification of Section 3.2. Measuring event time as the number of 28-day months between a song's release and its artist's adoption date, and restricting to the window from seven months before adoption to six months after, the estimation panel contains 16,663 of these AI-assisted songs.These AI songs, together with the same adopters' non-AI songs that serve as their within-artist benchmark, are the basis for the disclosure analyses of Sections 5.1 and 5.2. The matched non-adopter sample used for the across-artist robustness check and for the adoption-output result of Section 4.6 is constructed in Appendix B.

4.2 The Mandatory-Regime Sample

The mandatory regime lets us ask whether the AI label *causally* changes how listeners consume a song. Identifying this effect requires comparing songs that received the label to comparable songs that did not, with label assignment driven by something other than the songs' appeal. The September 18, 2025 rollout supplies exactly this comparison. As established in Section 3.2, a song received the label if and only if it was declared AI-assisted at upload and published on or after the April 15, 2025 threshold; publication timing relative to that platform-imposed cutoff, rather than any property of the song, determines treatment. We assemble a song-by-week panel around the rollout that exploits this threshold.

The treatment group, which we denote T, consists of AI-assisted songs created with Harmony.ai,

declared to the platform but not credited to listeners, and published in the six weeks after the threshold (April 15, 2025–May 31, 2025); these received the public label on September 18, 2025. The control group, C, consists of Harmony.ai songs meeting the same conditions but published in the six weeks before the threshold (March 1–April 14, 2025); because they predate it, they were never eligible for the label and remained unlabeled throughout. The treatment group comprises 5,593 songs from 3,539 artists, and the control group 4,246 songs from 1,141 artists. Control-group membership is exogenous to artist behavior—the Harmony.ai upload channel was force-closed by the platform on April 15, 2025, so assignment turns only on the calendar (Section 3.2). For the headline estimate both groups are restricted to their concealed subset, the songs for which the label adds genuinely new listener information; the credited songs supply the falsification test of Section 6. We observe each song over a symmetric window of calendar weeks spanning the rollout; the estimating equation appears in Section 6.

4.3 Outcomes

Our outcomes capture how listeners respond to a song along two distinct dimensions: *consumption*—whether and how thoroughly listeners play the song—and *engagement*—whether they actively endorse or propagate it. Each outcome is a count of events (a *volume* measure); every volume measure has a paired *reach* measure counting the distinct listeners who generated those events. We treat the volume measure as primary and report the reach measure as a robustness check.

Consumption. We use four consumption measures, ordered here from the least to the most demanding. *Plays*, the number of times a song was played, is our headline outcome—the broadest measure of listener demand and the least sensitive to platform-internal definitions. *Effective plays* counts only those plays whose listening duration exceeds a platform-defined minimum, distinguishing genuine listening from an incidental or accidental start.[9] *Full plays*, the number of plays heard through to completion, is the most demanding consumption measure, proxying for listener satisfaction conditional on starting the song. Finally, *unique listeners* counts the distinct accounts that played the song.

Engagement. We use five engagement measures, which escalate in the cost of the action they require of a listener. *Likes* count the times a listener marked the song with a "like"; *collects*, the times

[9]The exact minimum-duration threshold is set by the platform; we confirm its value against XMusic's data dictionary.

a listener saved it to a personal playlist or library, a stronger signal of intent to return; and *comments*, the number of comments posted on the song. *Shares* count the times listeners redistributed the song outside the platform, and *downloads*—available in the voluntary regime only—the times listeners saved it for offline play.

Discovery channel. In the mandatory regime we additionally decompose effective plays by how the listener reached the song, because a disclosure label may act on discovery differently than on deliberate demand. *Search effective plays* originate from a listener's search query, indicating that the listener deliberately sought the song or artist; *initiative effective plays* are plays the listener actively chose to start, as distinct from plays served automatically by the recommender.

Measurement window and transformation. The two regimes measure the same underlying behaviors and differ only in the aggregation window. In the voluntary regime, each outcome is summed over a fixed 28-day window following the song's release. In the mandatory regime, each outcome is observed per song-week: the panel is at the song-by-calendar-week level, so every observation aggregates the outcome over a seven-day window. All outcomes are winsorized at the 0.1st and 99.9th percentiles and enter the regressions in logs; in the panel specifications, song-weeks with no recorded activity are filled with zeros to keep the panel balanced. Plays is the headline outcome throughout; secondary outcomes are reported with Benjamini–Hochberg false-discovery-rate corrections. Summary-statistics tables report untransformed levels for interpretability.

4.4 Song and Artist Characteristics

Our analyses also use a small set of song- and artist-level characteristics. These enter the voluntary-regime disclosure analysis (Section 5.1) as predictors of the concealment choice and serve throughout as dimensions of heterogeneity. We measure four song-level traits. *Genre novelty* is one minus the share of the artist's previous releases that fall in the current song's genre, so it equals one for a genre entirely new to the artist and approaches zero for a genre the artist releases in habitually. *Role novelty* analogously captures how unusual the artist's creative role on the song is relative to their history, measured as the Euclidean distance between the song's role vector (across singer, writer, composer, and arranger) and the artist's historical role frequencies. *Team size* is the number of distinct human collaborators credited across those four roles, excluding the AI tool and virtual performers, and *own-role share* is the fraction of the four roles performed by the focal

artist, so that higher values denote a more nearly solo effort. At the artist level, we sort artists into *tiers* by pre-adoption follower count (bottom, middle, and top terciles), which we use for the heterogeneity splits.

4.5 Summary Statistics

Table 1 reports summary statistics for the voluntary-regime sample, partitioning the adopters' songs into all AI songs, those concealed from listeners, those credited, and the non-AI songs that serve as the within-artist benchmark.mandatory-regime sample, reporting the treatment and control groups' artists and songs and, in a four-cell panel, the pre-shock means of the concealed and credited contrasts that the label design exploits. The tables report baseline comparability, not treatment effects; the systematic differences across the disclosure columns—explored formally in Section 5.1—are the first sign that disclosure is a chosen, not a random, act.

4.6 Adoption and Production: AI Fills the Market

Before turning to disclosure, we document one piece of context. Comparing adopters to their matched non-adopters (constructed in Appendix B) in a staggered difference-in-differences design on the artist-month panel, we find that gaining access to Harmony.ai sharply raises an artist's monthly output.[10] Total output rises by 0.305 log points, roughly a 36% increase; the effect survives the Oster (2019) test for omitted-variable bias ($\beta^* = 0.245$). The increase is overwhelmingly composed of AI-assisted songs: output of non-AI songs rises only modestly (0.051 log points, about 5%) and that smaller effect is fragile to the same test ($\beta^* = -0.004$), as we report honestly.The event-study path is flat before adoption and jumps sharply at it. Because this result is context rather than a contribution, we report the estimates and the event study in Appendix C (Table A2).

This adoption effect is context, not the paper's contribution. It tells us that the voluntary regime we study is one in which AI-assisted songs quickly came to constitute a large and growing share of new releases—which is precisely why how artists chose to disclose that AI use, and how listeners responded, is worth understanding. We turn to disclosure next.

[10] Adoption timing is staggered across the sixteen monthly cohorts, so we report both the two-way fixed-effects estimate and, in Appendix C, the Callaway and Sant'Anna (2021) estimator and Rambachan and Roth (2023) sensitivity bounds; the pre-adoption coefficients are economically small and the conclusion is robust under both.

5 The Voluntary Regime: Strategic Disclosure and the Performance Gap

Under voluntary disclosure, two descriptive patterns set up the causal analysis that follows. First, disclosure is not random: artists systematically credit AI on some songs and conceal it on others. Second, concealed and credited songs perform very differently. Because the first pattern implies the second is confounded by selection, neither can be read as the causal effect of disclosure, and it is this distinction between correlation and causation that motivates the mandatory-label experiment of Section 6. The analyses in this section are descriptive and correlational throughout, and we draw no causal inferences from them. Both use the within-artist benchmark, comparing each adopter's AI songs against the non-AI songs the same adopter released.

5.1 Who Discloses?

We first ask which song and artist characteristics are associated with an artist's choice to credit AI assistance to listeners rather than conceal it. Two features of the setting make this a genuinely strategic choice. First, whether a song is disclosed as AI-assisted is entirely the artist's decision: listeners cannot independently verify how a song was made, and learn of AI involvement only if the artist credits the AI tool in the song's credit list. A listener may suspect AI after hearing a song, but the platform provides no way to confirm that suspicion unless the artist has named the tool.[11] Second, crediting AI changes nothing else about the song: it does not alter the recording, impose any fee, or reduce the artist's royalties. The full-royalty incentive described in Section 3.2 attached to the separate platform-side declaration, not to the listener-facing credit. Because the credit is therefore a free and unverifiable choice, the systematic patterns documented below cannot be artifacts of a technical rule that attaches the credit to particular kinds of songs; they reflect deliberate decisions by artists.

To document these patterns, we estimate a linear probability model of the disclosure choice on the AI-assisted songs of the voluntary-regime sample:

$$\begin{aligned}\text{Credit}_{ij} = \beta_1 \,\text{GenreNovelty}_{ij} + \beta_2 \,\text{RoleNovelty}_{ij} + \beta_3 \,\text{TeamSize}_{ij} \\ + \beta_4 \,\text{OwnRoleShare}_{ij} + \mathbf{x}_{ij}'\theta + \gamma_{g(i)} + \delta_{t(i)} + \alpha_j + \varepsilon_{ij}.\end{aligned} \tag{1}$$

[11] A growing body of evidence, from blind listening tests and consumer surveys, finds that ordinary listeners cannot reliably distinguish AI-generated from human-composed music (Figueiredo et al. 2025, Deezer and Ipsos 2025). Consistent with this, journalistic investigations have documented AI-generated tracks accumulating millions of streams without listeners recognizing their origin (Le Monde 2026).

The dependent variable Credit_{ij} equals one if AI song i by artist j credits the AI tool to listeners and zero if the AI use is concealed. The four regressors of interest are genre novelty, role novelty, team size, and own-role share, defined in Section 4.4; $\mathbf{x}_{ij}$ collects additional song controls. We include genre fixed effects $\gamma_{g(i)}$ and month fixed effects $\delta_{t(i)}$ throughout and artist fixed effects α_j where indicated (Table 3, column 2), and we cluster standard errors by artist. The estimation sample comprises the 13,338 AI-assisted songs with sufficient pre-adoption release history (Table 1). These coefficients are descriptive associations, not causal effects; we report them to show that disclosure is systematically related to observable song and artist traits.

Table 3 reports the estimates, and three associations stand out. First, novelty co-moves with disclosure. When an artist releases a song in a genre new to their catalog, they are more likely to credit the AI: a higher value of the genre-novelty measure of Section 4.4 is associated with a higher probability of crediting (Table 3, column 1: coefficient 0.097, $p < 0.05$). Second, both a larger production team and a larger own-role share are associated with crediting rather than concealing: the coefficient on team size is 0.307 (column 1, $p < 0.01$) and on own-role share 0.458 (column 1, $p < 0.01$), so songs with more substantial and more visible human involvement are the ones that carry the AI credit. Third, the novelty association is concentrated among lower-tier artists. Estimated separately by pre-adoption follower tercile, the genre-novelty coefficient is 0.216 for bottom-tier artists (Table 3, column 3, $p < 0.01$) but an insignificant 0.042 for top-tier artists (column 5). Taken together, these patterns show that the disclosure choice co-moves with the song's novelty and with the visibility of the artist's own contribution. For the purposes of what follows, the relevant point is narrower: concealed and credited songs differ systematically on traits that also predict performance, so a raw comparison of their performance cannot be read as the effect of disclosure.

5.2 The Conditional Performance Gap

Given that disclosure is systematically chosen, we next describe how concealed and credited AI songs differ in performance.

We make this comparison within artist, so that it holds fixed everything about the artist, including skill, following, and audience, and reflects differences across the artist's own songs. Its interpretation is bounded in one important way. Comparing within artist removes artist-level selection, such as more popular artists concealing more often, but it does not remove song-level selection: which of an artist's songs are concealed rather than credited is itself the strategic choice of Sec-

tion 5.1. The comparison is therefore descriptive, and its magnitude combines any listener response to disclosure with the sorting of songs into concealment. We return to this point in Section 5.3.

We estimate the following specification on the song-month panel of the adopters' catalog:

$$Y_{it} = \alpha_{a(i)} + \gamma_m + \beta_C \,\text{Concealed}_i + \beta_R \,\text{Credited}_i + \varepsilon_{it}. \tag{2}$$

The outcome Y_{it} is a log performance measure of song i in month t. The indicators Concealed_i and Credited_i mark the disclosure type of the artist's AI songs; the omitted reference category is the artist's own non-AI songs, so that each coefficient measures that type's average performance relative to the same artist's non-AI catalog, net of artist fixed effects $\alpha_{a(i)}$ and relative-month fixed effects γ_m (months since the artist's adoption); standard errors are clustered by artist. The object of interest is the concealed-minus-credited gap $\beta_C - \beta_R$, which we test as a linear contrast.[12]

Table 4 reports the estimates, and the ordering is consistent across outcomes. For plays (row "ln(1+plays)"), the concealed coefficient is 0.880 (column 1, $p < 0.01$) and the credited coefficient is -0.277 (column 2, $p < 0.05$): relative to the same artist's non-AI songs, concealed AI songs perform well above baseline and credited AI songs below it. The concealed-minus-credited gap is 1.157 log points (column 3, $p < 0.01$), about a threefold difference in plays. The same ordering holds across effective plays, full plays, likes, and collections, each with a positive concealed coefficient and a negative credited coefficient and gaps between roughly 0.7 and 1.3 log points (Table 4).

The gap is concentrated in discovery rather than in engagement depth. For play duration, the concealed coefficient is large and positive (1.582, $p < 0.01$) while the credited coefficient is negative and only marginally significant (-0.295, $p < 0.10$), consistent with the association operating on whether listeners begin a song rather than on how thoroughly they listen once they do.

5.3 The Gap Is Mostly Selection

Two further patterns indicate that this raw difference overstates any effect one might attribute to disclosure itself. First, a song-level selection model offers no rescue. Absent a credible exclusion restriction for the concealment choice, the implied selection correction is highly unstable, changing sign and magnitude across specifications. We therefore place no weight on it, relying instead on inverse-probability weighting for observable selection and on the policy experiment for design-based identification. Second, the concealed-song advantage does not survive aggregation: it is present

[12] Because concealed and credited AI songs do not exist before adoption, their song-level pre-trends are mechanically uninformative; pre-trend validity for the adopters rests on the artist-level event study of Appendix C.

across songs but disappears at the artist-month level, which indicates that it reflects which songs artists choose to conceal rather than a property of concealment itself.

These descriptive patterns motivate the experiment of the next section, which removes the disclosure choice and identifies the causal listener response to an AI label.

6 Mandatory Regime: The Causal Effect of the AI Label

The voluntary regime shows that disclosure is chosen strategically and that the raw performance gap between concealed and credited songs is mostly selection. Neither fact tells us what disclosure does to a song's audience, because in the voluntary regime the artist decides both whether to disclose and, implicitly, which songs to disclose. The mandatory regime removes that choice. On September 18, 2025, the platform began displaying a public "AI-assisted" label on every song its records marked as AI-assisted, whether or not the artist had chosen to credit the tool to listeners. For a song that had concealed its AI origin, the label is new information delivered to listeners by an administrative rule rather than by the artist. This section uses that rule to identify the causal effect of AI disclosure on listener behavior.

6.1 Design and Identification

Consider a song that listeners had only ever encountered as a human creation, with nothing on its page to signal any AI involvement. On September 18, 2025, the platform's compliance with the labeling regulation attaches a public "AI-assisted" label to that song. How do listeners respond? Answering this question is the central goal of the paper.

We want to estimate the causal effect of this AI-transparency policy on how listeners engage with a song. The ideal comparison would hold a song fixed and vary only whether listeners can see the label, contrasting two otherwise-identical songs, one labeled and one not. Our setting provides a close approximation. As Section 3.2 describes, the platform enforced the regulation prospectively: a song received the September 18, 2025 label if and only if it had reported AI assistance to the platform and was published on or after the April 15, 2025 registration threshold. Publication date relative to that threshold is therefore the single feature that determines whether an otherwise-comparable AI song was labeled. We compare AI songs published just after the threshold, which were labeled, against AI songs published just before it, which were not.

Both groups are the same kind of song: Harmony.ai songs whose artists reported the AI assistance

to the platform but did not credit it to listeners, that is, concealed AI songs (Section 3.2). We call the songs published in the roughly six weeks after the threshold (April 15–May 31, 2025), all of which were labeled on September 18, 2025, the *treated songs*, and the songs published in the roughly six weeks before it (March 1–April 14, 2025), none of which were ever labeled, the *control songs*. The treated songs number 5,593 across 3,539 artists, and the control songs number 4,246 across 1,141 artists (Table 2; Section 4.2 describes both samples). Treatment here is a property of the individual song rather than of the artist: because assignment turns only on a song's publication date, the same artist can have some songs labeled and others not, and our identification comes from comparing labeled against unlabeled songs, not labeled against unlabeled artists.

Two features of this design make the comparison credible. First, a song's assignment to the treated or control group is not a disclosure choice. Assignment turns entirely on whether the song was published before or after the April 15, 2025 threshold, a platform-wide calendar cutoff imposed by the regulation rather than a boundary any artist chose (Section 3.2); the treated and control artists are otherwise behaviorally alike, having all chosen to conceal their use of Harmony.ai. The selection problem that made the voluntary-regime gap uninterpretable is therefore absent. Second, the two groups face the same recommendation system. As Section 3.3 describes, since January 1, 2025 the platform has applied a uniform ranking discount to all AI songs regardless of their public label, and this discount was unchanged around September 18, 2025; the contrast we estimate is therefore not a differential-algorithm effect but the listener response to the label itself.

One difference between the two groups remains, and we address it directly. Because the treated songs were published later than the control songs, they are somewhat younger when public labeling begins, when both are four to six months old, and a song's play count declines naturally as it ages. A raw comparison would confound the label with this lifecycle gap. We absorb every fixed song characteristic with song fixed effects and control for a song's age in days, so that the estimate is identified from the change a labeled song undergoes relative to a comparable unlabeled song of the same vintage.

All songs share the same September 18, 2025 shock date, so we estimate a song-by-week difference-in-differences with song and week fixed effects:

$$Y_{it} = \alpha_i + \gamma_t + \beta\,(\mathrm{Treat}_i \times \mathrm{Post}_t) + \lambda\,\mathrm{SongAge}_{it} + \varepsilon_{it}, \tag{3}$$

where Y_{it} is a log weekly performance measure of song i in calendar week t, Treat_i equals one for the

treated songs, and Post_t equals one for weeks from September 18, 2025 onward. Song fixed effects α_i absorb the Treat_i main effect and every fixed song characteristic; week fixed effects γ_t absorb the Post_t main effect and any platform-wide weekly shock common to the two groups; SongAge_{it} controls the lifecycle gap. The coefficient of interest is β: the within-song change in performance that a treated song experiences after it is labeled, over and above the change experienced over the same weeks by a comparable control song. We cluster standard errors by song. We estimate the specification on the concealed AI songs described above, for which the label is genuinely new information; the songs that had already credited Harmony.ai to listeners before the shock supply the falsification test of Section 6.3.

6.2 The Causal Label Penalty

Table 5 reports the estimates. Labeling a previously-concealed song as AI-assisted reduces its weekly plays: the coefficient on Treat × Post is -0.047 (Table 5, Panel A, column 1, $p < 0.01$), a decline of about 4.6 percent. The effect on effective plays, which count only listens that pass a minimum-duration threshold, is smaller but of the same sign, -0.029 (Panel A, column 2, $p < 0.05$), about 2.9 percent. These are modest magnitudes, but they are precisely estimated and emerge cleanly at the shock. Figure 2 plots the week-by-week coefficients: the pre-September 18, 2025 coefficients are individually indistinguishable from zero and show no systematic trend, so the two groups were on parallel play trajectories before the label, and the divergence begins only once the label appears.[13] The penalty is present from the first post-label weeks and, if anything, drifts modestly deeper over the weeks that follow rather than reverting, consistent with the label's effect compounding through the platform's engagement-driven recommendation loop rather than fading as a one-time novelty.

The headline specification in Panel A includes calendar-week fixed effects, and we prefer it for a specific reason: with song and week fixed effects, Equation (3) is the standard two-way fixed-effects difference-in-differences, and the week effects absorb any platform-wide movement common to both groups in a given week, such as seasonal listening patterns, sitewide promotions, or the ongoing uniform discount on AI songs. Dropping the week fixed effects, so that identification comes only from the within-song pre-post change net of song age, yields a smaller and more conservative penalty of -0.020 for plays ($p < 0.05$) and -0.017 for effective plays ($p < 0.10$; Table 5, Panel B), about

[13] A joint test of the pre-label coefficients and a window-narrowing (bandwidth) robustness check confirm that the effect's sign and magnitude remain stable as the publication window is tightened around the threshold.

2.0 and 1.7 percent. The estimate therefore ranges from a conservative 2.0 percent without week effects to a preferred 4.6 percent with them; both are negative and statistically significant, and we treat the two-way specification as the headline.

The penalty is confined to discovery. The play-based outcomes, which measure whether listeners begin a song at all, move together and negatively, whereas the outcomes that measure what listeners do once they have started a song do not fall. Full plays, likes, collections, and comments are all statistically indistinguishable from zero in the headline specification (Table 5, Panel A, columns 3–6). The natural reading is that the label operates at the point of discovery: it makes some listeners less likely to click on or be served a song they now see is AI-assisted, but it does not sour the experience of those who do listen. This discovery interpretation also fits the near-absence of any vocal reaction in the songs' comment threads, where explicit mentions of AI remain rare after the label and show no sign of organized backlash.[14]

6.3 Falsification and Who Bears the Penalty

If the estimate in Table 5 is the effect of new information, it should vanish for songs whose AI use listeners could already see. We test this by re-estimating Equation (3) on the previously-credited pair, treated and control songs that had voluntarily credited the AI tool before September 18, 2025. For these songs the mandatory label tells listeners nothing they did not already know, so the predicted effect is zero. The estimate is statistically indistinguishable from zero, with and without the song-age control (Table 6, Panel A), which is consistent with the no-new-information logic. We are cautious about how much weight this test can bear, however, because the previously-credited pair is small: relatively few songs credited the generative-AI tool before the mandate, so the coefficient is imprecise, with a point estimate that is negative but a confidence interval wide enough to include both zero and an effect the size of the headline. The test fails to reject a null effect rather than establishing one. The sharper argument against a differential-algorithm explanation is the design itself: because the recommender treats labeled and unlabeled AI songs identically before and after the mandate (Section 3.3), a hidden ranking penalty switching on at September 18, 2025 is ruled out by construction, not merely by this low-powered contrast.

The penalty is not borne evenly, and here we read the evidence cautiously. Splitting the head-

[14] An exploratory difference-in-differences on the share of comments mentioning AI does not survive reading the underlying comments: post-label AI mentions are sparse, concentrated in a single controversial song, and mix artist self-disclosure with listener detection and spam. We therefore treat the comment record as evidence that the label prompted no broad listener outcry, consistent with a quiet discovery-side response rather than expressed disapproval, and we do not rest any estimate on it.

line sample by the artist's pre-period follower tercile, the cleanest result is the middle tier: a plays penalty of -0.044 (Table 6, Panel B, $p < 0.01$), about 4.3 percent, resting on parallel pre-trends (joint pre-trend test $p = 0.20$). Bottom-tier artists show a similar-sized penalty, -0.043 ($p < 0.05$, about 4.2 percent), but it does not pass the pre-trend test ($p = 0.04$), so we treat it as directionally consistent rather than established. Top-tier artists show no effect (-0.004, not significant). Because all AI songs face the same recommendation treatment regardless of tier (Section 3.3), the surviving pattern is a statement about listeners rather than about the algorithm: the label moves behavior where listeners hold weak priors about the artist and have little else to weigh, and not for established artists whose audiences decide on other grounds. We note one likely source of the pre-trend violations. The treated and control songs sit on either side of the April 15, 2025 registration cutoff, and the treated window coincides with a sharp influx of newly-registered artists in May 2025, so the two groups draw on somewhat different populations of artists, most visibly among lower-follower artists. We therefore report the tier pattern as suggestive; narrowing the publication window around the cutoff—which brings the treated and control cohorts into closer compositional alignment—flattens the pre-trends while leaving the headline effect's sign and magnitude broadly intact.

6.4 Betrayal, Not Buffer: The Role of Prior Exposure

A natural question is whether a song's history with its audience cushions or sharpens the label's effect. Two hypotheses point in opposite directions. Prior exposure could buffer the penalty, if listeners who already know and like a song discount the new label; or it could deepen it, if the label reads as a belated admission that a song they had taken to be human was not, so that the disclosure lands as a betrayal. We test this by interacting the mandatory disclosure effect with a song's cumulative plays before the shock, purged of its mechanical correlation with song age. The evidence points to betrayal: songs with larger pre-label audiences suffer the larger penalty. We lead with the conservative estimate, which measures the pre-label audience at a stale snapshot roughly seven weeks before the shock (August 1, 2025) to limit contamination from the outcome's own recent trajectory; there the interaction is -0.011 ($p < 0.01$), and measuring the audience at the eve of the shock (September 17) widens it to -0.028 as an upper bound.

We flag this result as suggestive rather than decisive. Part of the pattern is mechanical: songs with high recent play counts tend to revert regardless of any label, and a placebo regression on the never-labeled control group recovers a reversion of -0.014 ($p < 0.01$) that is present at both measurement dates. Its value is interpretive: the penalty is largest exactly where the conceal-

ment mattered most, among songs whose audiences had already formed an attachment under the impression that the music was human-made.

6.5 Robustness

Several specification checks stand behind the headline. The estimate is robust to including or excluding week fixed effects (the Panel A and Panel B contrast above) and to the battery of secondary outcomes, whose p-values we adjust for multiple testing using the Benjamini–Hochberg procedure; the discovery outcomes survive the adjustment and the engagement outcomes remain null. We do not report an Oster bounding exercise here, because the within-song design already absorbs so much variation that the stability test is degenerate.[15]

Taken together, the mandatory regime delivers what the voluntary regime could not: a clean estimate of the causal effect of AI disclosure. Forcing a previously-concealed song's AI origin into view costs it roughly five percent of its plays, the loss is concentrated at discovery rather than in engagement, it falls on artists whose audiences hold weak priors, and it lands hardest on the songs whose listeners had the most to unlearn. The next section connects this causal magnitude back to the voluntary regime, asking how much of the strategic concealment we documented there the label penalty can rationalize.

7 Selection versus Causation: Bridging the Two Regimes

The two regimes measure different things. The voluntary regime shows that concealed AI songs outperform credited ones by a wide margin, but that gap reflects both how listeners respond to disclosure and which songs artists choose to conceal, the strategic concealment the framework of Section 2.3 anticipates. The mandatory regime isolates the first channel on its own. Setting the two side by side answers the question in the paper's title: how much of the disclosure penalty is a genuine listener response, and how much is the selection behind it?

7.1 How Much of the Gap Is Selection?

Placing the two estimates side by side settles the question. In the voluntary regime, concealed AI songs outperform the same artist's credited AI songs by 1.157 log points in plays (Table 4, column

[15] The high fitted R^2 of Equation (3), driven by the song fixed effects, makes the Oster δ=1 bound uninformative for this design; the test is better suited to the artist-month adoption regressions of Appendix C, where we do report it.

3), roughly a threefold difference. Read at face value, that gap invites the conclusion that crediting AI is ruinous for a song, and it is exactly the comparison an artist weighing whether to credit the tool would have in view. The mandatory regime shows that conclusion to be almost entirely an artifact of selection. When an administrative rule forces disclosure onto a previously-concealed song, holding the song itself fixed, its plays fall by roughly five percent (Table 5), nowhere near the threefold raw gap.

The two estimates come from different regimes, samples, and outcome windows, so we treat them as an order-of-magnitude bridge rather than a formal decomposition. Even under that caution the conclusion is unambiguous. Taking the causal estimate at its largest and the voluntary gap at face value, the genuine listener response accounts for at most a few percent of the observed concealed-credited gap; the rest reflects which songs artists chose to credit in the first place. Neither row contains a differential-algorithm channel, because the recommender is disclosure-neutral in both regimes (Section 3.3), so what the causal estimate leaves unexplained is selection, not ranking. We therefore treat the selection share as a bound rather than a point estimate: pinning it down precisely would require harmonizing the two outcome windows and margins, which the order-of-magnitude gap makes unnecessary for our purpose. The raw voluntary comparison overstates the causal penalty of disclosure by more than an order of magnitude, and that overstatement is the selection the paper's title names.

7.2 Field and Lab Evidence as Complements

Our estimates help show how a literature whose findings can look conflicting—about whether disclosing AI helps or hurts creative work—in fact fits together. On one side is a body of laboratory evidence, most comprehensively Raj et al. (2026), who assign AI disclosure at random to a fixed piece of writing and document a robust penalty mediated by perceived authenticity; a companion vignette study using a musical stimulus reaches the same conclusion and adds that the artist's reputation does not soften it (Benegal et al. 2026). On the other side is a body of field evidence: on Kickstarter, disclosing AI use lowers funding success while an explicitly anti-AI stance raises it (Heshmati and Wang 2026), and a mandatory AI label depresses outcomes for digital games (Li 2026). Each body sees something real, and each is built to see one part of the whole. The randomized experiments identify the audience's response to a label and, by holding the work fixed, deliberately set aside the selection that operates when disclosure is a choice; the field studies observe those choices in live markets but, without a counterfactual on the same work, do not separate them

from the response. The two are complements, and our two-regime design recovers both parts, which is why our streaming estimates sit between the field's large raw penalties and the experiments' clean causal ones, and why the apparent disagreement in this literature is better read as a difference in what each design is built to identify.

Three distinctions do the reconciling. First, stated versus actual behavior. The Kickstarter evidence observes artists' public *statements* about AI rather than their actual use, which its question does not require it to separate from the selection of which artists disclose; our design adds true AI use and true concealment, and shows that separation to matter enormously. Second, the evaluator and the setting. A crowdfunding backer is a patron who may actively reward a human-first stance, and a vignette respondent is asked to render an authenticity judgment, whereas a streaming listener is mostly deciding whether to press play. The penalty we identify is a discovery effect, a modest reduction in whether listeners begin a labeled song, not a verdict on its artistic worth. Third, causal identification. The cross-sectional gaps in the prior field work combine listener response with selection, the confound our two-regime design is built to break apart; once broken apart, the causal penalty is small.

The sharpest contrast is over reputation. The lab vignette finds no reputational buffer, whereas our field data point to one: the penalty appears concentrated on lower- and middle-tier artists, with no detectable effect for the top tier (Section 6.3). The difference is again about information. A vignette gives the evaluator nothing to weigh but the disclosure itself, while a streaming listener choosing among established artists brings priors, a familiar catalog, and a reputation to the decision, so the label shifts behavior only where those priors are thin.

7.3 The Collapse of Voluntary Disclosure

A final pattern in the data is worth naming, if only to set it aside. Voluntary crediting of AI did not persist up to the mandate and then stop; the share of AI songs carrying a voluntary AI credit fell sharply over the winter of 2024–2025, months before the September 18, 2025 label. The timing rules out the tempting reading that artists abandoned disclosure because they had learned the raw penalty was mostly selection. The decline instead coincides with the January 2025 change in the platform's recommender, which began de-prioritizing all AI songs regardless of whether their artists credited the tool (Section 3.3): once crediting AI no longer distinguished a song in the ranking, the private case for volunteering it weakened. We record this as a descriptive pattern rather than a result, and leave its causes, which our design is not built to identify, to future work.

8 Conclusion and Discussion

As platforms and regulators move to require the disclosure of AI in cultural markets, the animating fear is that transparency will punish the artists who comply. The raw data appear to confirm it: within an artist's own catalog, songs that conceal their AI involvement outdraw those that credit it by roughly a threefold margin. Across two disclosure regimes on a major music-streaming platform, we show that almost none of this gap is an audience response to disclosure. When a labeling mandate forced revelation onto previously-concealed songs while holding the song fixed, plays fell by only about five percent, an order of magnitude less than the raw comparison implies. What the raw gap mostly captures is selection: artists conceal exactly the songs that disclosure would cost the most, so a cross-section of concealed against credited songs describes their choice far more than it measures the label's effect.

8.1 Contributions

Our contribution is, first, conceptual. Two literatures speak to AI disclosure without speaking to each other. A behavioral literature shows that audiences penalize creative work once they learn AI was involved, robustly and through a single channel—perceived authenticity—that resists nearly every attempt to reframe it away (Raj et al. 2026, Benegal et al. 2026). A separate economics of disclosure asks when producers voluntarily reveal private information, and predicts, in its classic form, that they unravel toward full disclosure (Grossman 1981, Milgrom 1981). Joining the two shows that each implies the other's blind spot. Precisely because the authenticity penalty is real and cannot be talked down, and because a song's AI origins are neither verifiable to listeners nor a source of good news that any artist would volunteer, the unraveling logic runs in reverse: the equilibrium is concealment, not revelation. Disclosure under a voluntary regime is thus an act of selection rather than of information, and the penalty an observer reads off a cross-section of disclosed against concealed work is not the audience's response to a label but that response compounded with the choice of who discloses. Naming that compound, and the conditions that generate it, reframes how the disclosure penalty documented across this literature ought to be read.

The contribution is, second, one of identification. That compound can be decomposed only by a design that observes both true—potentially concealed—AI use and an exogenous source of disclosure on the same work, and each strand of prior evidence supplies one of these halves rather than both. The experimental studies recover the causal response cleanly, by randomly assigning disclosure to

a fixed stimulus; that same randomization holds selection constant, so the choice that dominates any real market falls outside what a vignette is built to observe (Raj et al. 2026). The emerging field studies capture that choice in live markets—artists' stated positions on AI in crowdfunding (Heshmati and Wang 2026), or their disclosure of AI in digital games (Li 2026)—and, without an observed counterfactual on the same work, report associations that combine response with selection, as their authors carefully note. Our two-regime design brings the two halves together: the voluntary regime exposes the selection, and a labeling mandate imposed on already-concealed songs recovers the causal response in the field, the market counterpart of the experimental estimate. Setting the two side by side is what lets us attribute the raw gap almost entirely to selection, and it clarifies what the penalty *is*: an effect that operates at discovery, dampening whether listeners begin a labeled song rather than how they judge it once they do, and that appears regressive, concentrated on lower- and middle-tier artists whose audiences bring their own priors. Disclosure moves behavior only where information about the artist is otherwise scarce. Read this way, a literature that has seemed to disagree about whether disclosing AI helps or hurts is better understood as differing in how cleanly each design separates the audience's response from the choices around it, and in how authenticity-sensitive each domain is (Longoni and Cian 2022, Logg et al. 2019).

8.2 Policy Implications

More broadly, our findings speak to a policy debate that moved from proposal to law during the very window we study. Mandatory AI-content disclosure is arriving across the world's largest media markets at once. China's Measures for Labeling AI-Generated Synthetic Content, together with a mandatory national standard, took effect on September 1, 2025 (Cyberspace Administration of China 2025, Standardization Administration of China 2025); the European Union's AI Act, in force since August 2024, requires that AI-generated and synthetic content be disclosed under Article 50 from August 2, 2026 (European Parliament and Council of the European Union 2024); and although the United States has not enacted a federal labeling mandate—bills such as the AI Disclosure Act of 2023 and the AI Labeling Act of 2023 stalled in committee (U.S. Congress 2023a,b), with transparency requirements surfacing instead through state law such as California's AI Transparency Act (California State Legislature 2024)—the major platforms have moved on their own. YouTube and TikTok began requiring disclosure or automatic labeling of synthetic content in 2024 (YouTube 2024, TikTok 2024), and the music-streaming services followed: Deezer's AI tagging system in June 2025 (Deezer 2025b), Spotify's AI-use disclosures in music credits in September 2025

(Spotify 2025), and Apple Music's transparency tags in March 2026 (Stassen 2026). Much of the debate around these measures has been conducted in the abstract, and it is animated by a specific worry: that compelling artists to reveal AI involvement imposes a heavy penalty on those who comply, a presumption the large raw gap between concealed and disclosed AI work appears to license.

Our estimates recast that worry. The causal cost of a label is modest rather than ruinous; it falls at discovery, in whether listeners begin a song, rather than in how they engage with one they have chosen to play; and it appears close to zero for established artists. The penalty looks large only in cross-section, because the cross-section is governed by the selection of who discloses rather than by any listener response. Two implications follow for platforms and regulators. First, a disclosure mandate is unlikely to devastate compliant artists on average, because the headline tax the debate fears is largely a statistical artifact of concealment rather than a verdict listeners render on labeled work. Second, the cost that does exist appears regressive, falling hardest on the newer and smaller artists whose audiences have the least other information to go on and lightest on the established artists who need protection least. The right question for a transparency rule is therefore not whether it imposes the outsized penalty implied by comparing disclosed and concealed work directly, but whether its genuine, smaller, and unevenly distributed cost is justified by the benefits of disclosure, and whether that cost can be cushioned where listener information is thinnest.

8.3 Limitations and Future Research

Several features of our setting bound these findings. Our evidence comes from a single streaming platform, one integrated AI tool, and one national market, and the magnitudes may differ where AI stigma or disclosure norms differ; the qualitative lesson, that a raw disclosure gap is dominated by selection, should nonetheless travel to any market in which artists choose whether to reveal. Music, moreover, is the hedonic, authenticity-dominant case in which the disclosure penalty should be largest (Longoni and Cian 2022, Castelo et al. 2019b), so the modest causal effect we recover here likely bounds from above what a label costs creators in less authenticity-sensitive markets. The mandate we study arrived bundled with a registration regime, so our estimate reflects the disclosure requirement as an institutional package rather than the on-screen label in isolation, and our post-mandate window is short enough that we identify the label's immediate effect on discovery rather than a long-run equilibrium in which artists re-optimize production or listeners habituate. Concealment, finally, is observed only where the platform's detection flags AI that an artist did

not credit, so our concealed sample is a lower bound on true concealment and our estimates speak to detectable AI use. Because the recommender is disclosure-neutral as a matter of policy but not necessarily of outcome (Section 3.3), a discovery penalty that feeds back into ranking could compound over horizons our data do not reach.

Two questions follow. The first concerns the dynamics of voluntary disclosure. Crediting collapsed months before the mandate, in step with a change in the platform's recommender rather than with any learning about the penalty (Section 7.3), and understanding how artists' disclosure choices respond as the private returns to crediting shift, and how those choices settle into market-wide norms, is a natural next step our design is not built to trace. The second concerns what a mandate does beyond audience response. Forced transparency may reshape not only whether artists credit AI but how they use it in the creative process itself, altering the extent, the visibility, or the nature of AI assistance in cultural production. Understanding which disclosure designs preserve the honesty that transparency is meant to produce without discouraging the creation it is meant to inform is an important direction for future research.

References

Askin, N. and M. Mauskapf (2017). What makes popular culture popular? product features and optimal differentiation in music. *American Sociological Review 82*(5), 910–944.

Bellaiche, L., R. Shahi, M. H. Turpin, A. Ragnhildstveit, S. Sprockett, N. Barr, A. Christensen, and P. Seli (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications 8*(1), 42.

Benegal, A., L. C. Vincent, and J. B. Carnevale (2026). Creative or contrived? how AI use shapes the social evaluations of individuals with creative reputations. *Academy of Management Discoveries*. Advance online publication. Two vignette experiments; authenticity mediates the AI-use penalty.

Bick, A., A. Blandin, and D. J. Deming (2024). The rapid adoption of generative AI. Working Paper 32966, National Bureau of Economic Research. Published in *Management Science* (2026).

Boussioux, L., J. N. Lane, M. Zhang, V. Jacimovic, and K. R. Lakhani (2024). The crowdless future? generative AI and creative problem-solving. *Organization Science 35*(5), 1589–1607. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4533642`.

California State Legislature (2024). SB-942 california AI transparency act. `https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942`. Chapter 291, Statutes of 2024; signed by Governor Newsom 19 September 2024; operative date amended to 2 August 2026 by AB 853 (signed 13 October 2025).

Callaway, B. and P. H. C. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Castelo, N., M. W. Bos, and D. R. Lehmann (2019a). Task-dependent algorithm aversion. *Journal of Marketing Research 56*(5), 809–825.

Castelo, N., M. W. Bos, and D. R. Lehmann (2019b). Task-dependent algorithm aversion. *Journal of Marketing Research 56*(5), 809–825.

Csaszar, F. A., H. Ketkar, and H. Kim (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Science*.

Cui, K. Z., M. Demirer, S. Jaffe, L. Musolff, S. Peng, and T. Salz (2024). The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers. *Management Science*. Forthcoming.

Cyberspace Administration of China (2025). Measures for labeling AI-generated synthetic content. `https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm`. Issued 14 March 2025 by the Cyberspace Administration of China with the Ministry of Industry and Information Technology, Ministry of Public Security, and National Radio and Television Administration; effective 1 September 2025.

Deezer (2025a). 28% of all delivered music is now fully AI-generated. `https://newsroom-deezer.com/2025/09/28-fully-ai-generated-music/`. Deezer Newsroom, September 2025.

Deezer (2025b). Deezer launches world's first AI tagging system for music streaming. `https://newsroom-deezer.com/2025/06/deezer-launches-worlds-first-ai-tagging-system-for-music-streaming/`. Announced 20 June 2025.

Deezer and Ipsos (2025, November). AI and music: Survey reveals most listeners cannot distinguish AI-generated tracks. https://newsroom-deezer.com/2025/11/deezer-ipsos-survey-ai-music/. Online survey by Ipsos Digital, fielded 6–10 October 2025; N = 9,000 adults across eight countries. In a blind listening test, 97% could not correctly identify the AI-generated tracks; 71% reported surprise at their inability to distinguish, and 40% would skip AI-generated songs.

Dell'Acqua, F., E. McFowland III, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Working Paper 24-013, Harvard Business School. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General 144*(1), 114–126.

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

Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research 23*(1), 123–145.

European Parliament and Council of the European Union (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (artificial intelligence act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng. Adopted 13 June 2024; published in the Official Journal 12 July 2024; entered into force 1 August 2024. Article 50 transparency obligations apply from 2 August 2026.

Figueiredo, F., G. Martinelli, H. Sousa, P. Rodrigues, F. Pedrosa, and L. N. Ferreira (2025). Echoes of humanity: Exploring the perceived humanness of AI music. Forced-choice blind tests: listeners scored 53% on random AI-vs-human pairs, statistically indistinguishable from chance (N = 308).

Granja, J. (2018). Disclosure regulation in the commercial banking industry: Lessons from the national banking era. *Journal of Accounting Research 56*(1), 173–216.

Grewal, D., C. B. Satornino, T. Davenport, and A. Guha (2025). How generative AI is shaping the future of marketing. *Journal of the Academy of Marketing Science 53*(3), 702–722.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics 24*(3), 461–483.

Healy, P. M. and K. G. Palepu (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics 31*(1–3), 405–440.

Heshmati, M. and M. Wang (2026, January). Market responses to AI statements in creative domains. SSRN Working Paper. University of Washington. Kickstarter AI-disclosure policy, 41,887 campaigns; AI disclosure −11.9pp funding success, anti-AI narratives rewarded.

Horton, Jr., C. B., M. W. White, and S. S. Iyengar (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports 13*(1), 19001.

International Music Summit (2025). 60 million people used AI to create music in 2024, IMS business report 2025 finds. https://djmag.com/news/60-million-people-used-ai-create-music-2024-ims-business-report-2025-finds. *DJ Mag*, reporting the IMS Business Report 2025.

Jago, A. S. (2019). Algorithms and authenticity. *Academy of Management Discoveries 5*(1), 38–56.

Jin, G. Z. and P. Leslie (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics 118*(2), 409–451.

Jussupow, E., I. Benbasat, and A. Heinzl (2020). Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion. In *Proceedings of the 28th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2020_rp/168/.

Köbis, N. and L. D. Mossink (2021). Artificial intelligence versus maya angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior 114*, 106553.

Le Monde (2026, March). Comment les faux artistes générés par l'IA bouleversent l'industrie de la musique. https://www.lemonde.fr/culture/video/2026/03/28/comment-les-faux-artistes-generes-par-ia-bouleversent-l-industrie-de-la-musique_6674845_3246.html. Investigative report, 28 March 2026. Documents AI-generated tracks streamed millions of times without listeners recognizing their origin.

Li, S. (2026). When transparency raises the bar: AI disclosure and post-trial evaluation. Paper accepted to the Academy of Management Annual Meeting 2026 (submission #18422). University of Hong Kong. Unpublished manuscript. Steam indie games; mandatory Jan-2024 AI label; cross-sectional PSM/IPW, penalty at post-trial sentiment, null at sampling.

Liang, W., M. Yuksekgonul, Y. Mao, E. Wu, and J. Zou (2023). GPT detectors are biased against non-native English writers. *Patterns 4*(7), 100779.

Logg, J. M., J. A. Minson, and D. A. Moore (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes 151*, 90–103.

Longoni, C., A. Bonezzi, and C. K. Morewedge (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research 46*(4), 629–650.

Longoni, C. and L. Cian (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The "word-of-machine" effect. *Journal of Marketing 86*(1), 91–108.

Mathios, A. D. (2000). The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market. *Journal of Law and Economics 43*(2), 651–677.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics 12*(2), 380–391.

Newman, G. E. and P. Bloom (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General 141*(3), 558–569.

OpenAI (2023). New AI classifier for indicating AI-written text. https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/. Withdrawn 20 July 2023 for low accuracy.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Raj, M., J. M. Berg, and R. Seamans (2026). The artificial intelligence disclosure penalty: Humans persistently devalue AI-generated creative writing. *Journal of Experimental Psychology: General 155*(4), 896–915.

Rambachan, A. and J. Roth (2023). A more credible approach to parallel trends. *Review of Economic Studies 90*(5), 2555–2591.

Reimers, I. and J. Waldfogel (2026). AI and the quantity and quality of creative products: Have LLMs boosted creation of valuable books? Working Paper 34777, National Bureau of Economic Research.

Sadasivan, V. S., A. Kumar, S. Balasubramanian, W. Wang, and S. Feizi (2023). Can AI-generated text be reliably detected? *arXiv preprint arXiv:2303.11156*. https://arxiv.org/abs/2303.11156.

Spotify (2025). Spotify strengthens AI protections for artists, songwriters, and producers. https://newsroom.spotify.com/2025-09-25/spotify-strengthens-ai-protections/. Announced 25 September 2025.

Standardization Administration of China (2025). Cybersecurity technology—labeling method for content generated by artificial intelligence, GB 45438-2025. https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=F32EA2A561F1886CD8D606513512D547. Mandatory national standard released 28 February 2025; effective 1 September 2025.

Stassen, M. (2026). Apple music launches AI transparency tags—but only if labels and distributors choose to declare them. https://www.musicbusinessworldwide.com/apple-music-launches-ai-transparency-tags-but-only-if-labels-and-distributors-choose-to-decla Music Business Worldwide, 4 March 2026.

TechCrunch (2025). Netflix goes 'all in' on generative AI as entertainment industry remains divided. https://techcrunch.com/2025/10/21/netflix-goes-all-in-on-generative-ai-as-entertainment-industry-remains-divided/. *TechCrunch*, 21 October 2025.

Tigre Moura, F., C. Castrucci, and C. Hindley (2023). Artificial intelligence creates art? an experimental investigation of value and creativity perceptions. *The Journal of Creative Behavior 57*(4), 534–549.

TikTok (2024). Partnering with our industry to advance AI transparency and literacy. https://newsroom.tiktok.com/en-us/partnering-with-our-industry-to-advance-ai-transparency-and-literacy. TikTok Newsroom, 9 May 2024.

U.S. Congress (2023a). AI disclosure act of 2023, H.R. 3831, 118th congress. https://www.govinfo.gov/app/details/BILLS-118hr3831ih. Introduced 5 June 2023 by Rep. Ritchie Torres; referred to committee; not enacted.

U.S. Congress (2023b). AI labeling act of 2023, S. 2691, 118th congress. https://www.govinfo.gov/app/details/BILLS-118s2691is. Introduced 27 July 2023 by Sen. Brian Schatz with Sen. John Kennedy; referred to committee; not enacted.

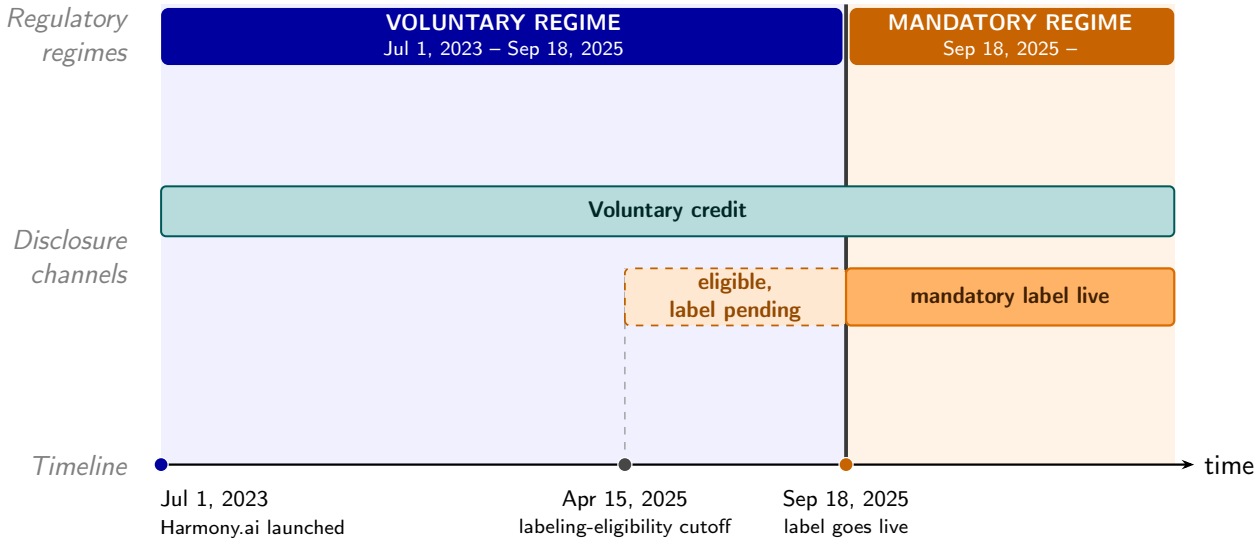
Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics 32*(1–3), 97–180.

Weber-Wulff, D., A. Anohina-Naumeca, S. Bjelobaba, T. Foltýnek, J. Guerrero-Dib, O. Popoola, P. Šigut, and L. Waddington (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity 19*(1), 26.

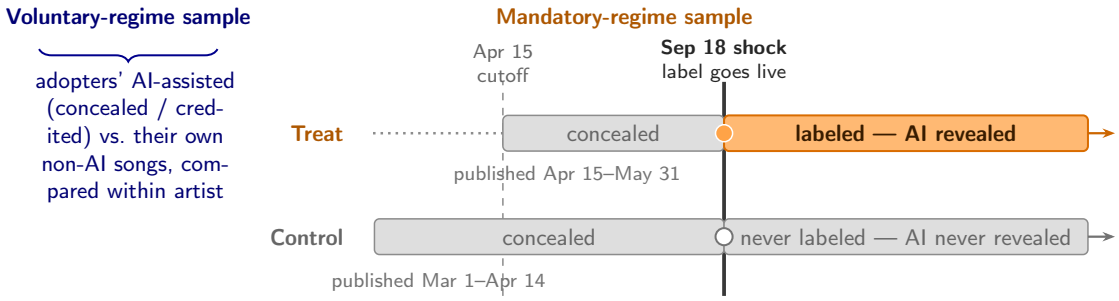
YouTube (2024). How we're helping creators disclose altered or synthetic content. `https://blog.youtube/news-and-events/disclosing-ai-generated-content/`. YouTube Official Blog, 18 March 2024.

Zhou, E. and D. Lee (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus 3*(3), pgae052.

Figures



(a) Regulatory regimes, disclosure channels, and the event timeline



(b) Analysis samples.

Figure 1: Two Disclosure Regimes, Artists' Disclosure Channels, and the Analysis Samples

Note. The timeline is schematic and not drawn to scale; panels (a) and (b) share the same horizontal scale, so the samples in (b) line up with the dates in (a). **Regimes.** The bold vertical line marks the regime boundary. Under the *voluntary* regime (Harmony.ai's launch on July 1, 2023 through September 17, 2025), artists alone decided whether to disclose a song's AI involvement. Under the *mandatory* regime (from September 18, 2025), XMusic displays a public "AI-assisted" label. **Disclosure channels.** A song's AI use can surface in two independent ways. First, the artist may *voluntarily credit* the AI tool in the song's credit list—available throughout both regimes. Second, the platform may add a *mandatory label*; this applies only to AI songs published on or after the April 15, 2025 labeling-eligibility cutoff (the "eligible, label pending" segment), and the labels became visible to listeners at the September 18, 2025 boundary ("mandatory label live"). After that date a single song can be revealed through both channels at once. **Samples.** The voluntary-regime sample compares each Harmony.ai adopter's AI-assisted songs—whether concealed from or credited to listeners—against the same adopter's own non-AI songs, within artist (adoptions July 2023–March 2024). The mandatory-regime sample is a calendar-aligned cohort of *concealed* AI songs around the September 18 shock: the *Treat* group was published in the six weeks on/after the April 15 cutoff (April 15–May 31, 2025) and received the label; the *Control* group was published in the six weeks just before it (March 1–April 14, 2025) and was never labeled.

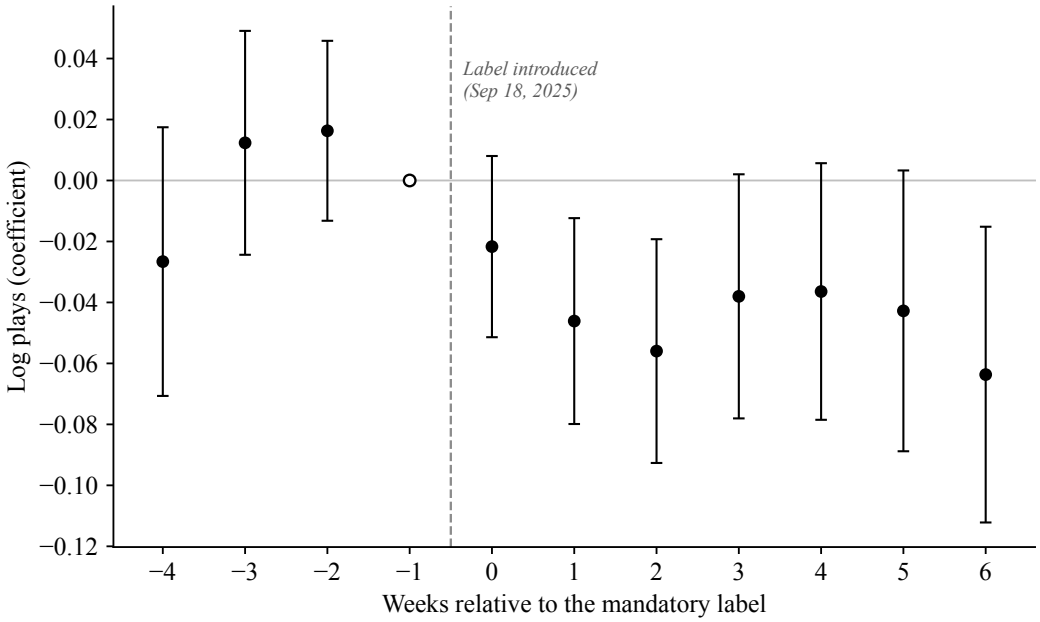


Figure 2: The Causal Label Penalty: Event Study

Note. Each point is the coefficient on the interaction of Treat with an indicator for the given week relative to the September 18, 2025 mandatory label, from the event-study form of Equation (3): the within-song change in log weekly plays for treated (labeled) songs relative to control (unlabeled) songs of the same vintage. Week -1 is the omitted reference (hollow marker). The specification absorbs song and calendar-week fixed effects and controls for song age; the sample is the previously-concealed mandatory-regime sample of Table 2 (N = 132,440 song-weeks). Bars are 95% confidence intervals from standard errors clustered by song. The pre-label coefficients are individually indistinguishable from zero and show no systematic trend; the play penalty emerges once the label appears and, if anything, deepens over subsequent weeks.

Tables

Table 1: Summary Statistics: Voluntary-Regime Sample

Panel A: Artists who adopted Harmony.ai

	Mean	SD	Median	Min	Max	*N*
Female[a]	0.146	0.353	0.000	0.000	1.000	1,932
Catalog size (songs)	126.2	336.1	50.5	3	7,374	1,932
Followers	2,398.9	11,110.8	135.5	0	267,566	1,932
Artist age (years)	23.0	6.3	22	14	43	1,912
Platform tenure (days)	789.0	617.3	635	2	3,271	1,932
Artist level	1.32	0.50	1	1	5	1,932
Pre-adoption monthly songs	0.508	0.260	0.462	0.050	1.000	1,932

Panel B: Songs published by artists who adopted Harmony.ai, by type

	AI songs			Non-AI songs
	All	Concealed	Credited	
Plays (28d)	5,859.0	8,307.1	1,975.9	9,510.3
Effective plays (28d)	4,168.2	5,812.5	1,560.3	6,329.3
Likes (28d)	139.3	224.2	4.5	206.7
Collections (28d)	211.3	313.8	48.8	299.4
Genre novelty	0.650	0.621	0.697	0.528
Team size	1.540	1.190	2.096	1.650
Songs[b]	21,718	13,320	8,398	164,634

Note. Panel A reports summary statistics—mean, standard deviation, median, minimum, maximum, and number of artists—for the artists who adopted Harmony.ai, the voluntary-regime sample. Panel B reports song-level means for these artists' catalog, grouped into their AI songs (all, and split by the artist's voluntary disclosure choice into concealed and credited) and their non-AI songs, which serve as the within-artist benchmark for the conditional comparison of Table 4. Crediting means the AI tool appears in the song's credit list (Section 3.2). All variables are reported in raw, untransformed units. Covariate balance between adopters and matched non-adopters, used only for the robustness and adoption-context results, is reported in Appendix B.

[a] Female is an indicator equal to one if the artist's platform gender field is female. Gender is undefined for band accounts; the field distribution is 0.034 band or unverifiable, 0.819 male, and 0.146 female.

[b] The 21,718 AI songs (13,320 concealed, 8,398 credited) belong to the 1,932 adopting artists; the disclosure-choice model (Table 3) further restricts to the 13,338 with sufficient pre-adoption release history.

Table 2: Summary Statistics: Mandatory-Regime Sample

Panel A: Artists behind treated and control songs

	Labeled (T)	Never labeled (C)	Diff.	p
Followers	129.3	488.1	−358.9	<0.001
Platform tenure (days)	29.1	520.7	−491.6	<0.001
Artist level	0.008	0.076	−0.068	<0.001
Catalog size (songs)	16.3	95.4	−79.1	<0.001
Sample songs per artist	1.6	3.7		
Artists	3,539	1,141		

Panel B: Songs — pre-shock means by analysis cell

	Previously concealed			Previously credited		
	T	C	Diff. [p]	T	C	Diff. [p]
Plays	2.589	4.063	−1.474 [0.056]	4.121	1.623	2.498 [0.354]
Effective plays	2.286	3.154	−0.868 [0.164]	3.172	1.530	1.642 [0.521]
Collections	0.021	0.040	−0.019 [0.221]	0.184	0.041	0.143 [0.404]
Song age (days)	121.3	178.8	−57.5 [<0.001]	121.3	180.9	−59.5 [<0.001]
Cumulative pre-shock plays	799.1	4,628.3	−3,829.2 [<0.001]	395.5	1,409.6	−1,014.1 [0.011]
Songs	5,337	4,123		133	123	

Note. This table summarizes the mandatory-regime sample. Panel A reports artist-level characteristics for the artists behind the treated and control songs; Panel B reports song-level means over the pre-shock window. Treated songs (T) were uploaded between April 15 and May 31, 2025 under the mandatory registration interface and received the public "AI-assisted" label on September 18, 2025; control songs (C) were uploaded between March 1 and April 14, 2025, before the April 15 registration threshold, and so were never eligible for the label. In total there are 5,593 treated and 4,246 control songs. Panel B partitions the songs into two analysis cells: the *previously concealed* pair, whose treated–control contrast is the headline estimate of Table 5, and the *previously credited* pair (songs whose AI use had already been voluntarily credited to listeners), whose contrast is the falsification test of Table 6; the previously-credited cells are small, so that test is correspondingly imprecise. Within each pair, the Diff. column reports the treated-minus-control difference with its p-value in brackets. Song age is measured at September 18, 2025 and cumulative plays at September 17, 2025; all outcomes are reported in raw, untransformed units. The treated and control songs come from systematically different artists—treated artists are newer, smaller, and lower-tenure (Panel A)—and their pre-shock song-level values differ accordingly (treated songs are also younger and have accumulated fewer plays at the shock). Song fixed effects absorb these level differences; identification in Tables 5 and 6 instead rests on parallel pre-trends.

Table 3: The Disclosure Choice: Which AI Songs Credit the AI Tool?

	Dependent variable: Credited = 1 (AI songs only)				
	(1)	(2)	By artist tier (no artist FE)		
			(3) Bottom	(4) Middle	(5) Top
Genre novelty	0.097**	0.016	0.216***	−0.009	0.042
	(0.046)	(0.047)	(0.048)	(0.072)	(0.073)
Team size	0.307***	0.283***	0.449***	0.290***	0.212***
	(0.026)	(0.038)	(0.050)	(0.042)	(0.030)
Own role share	0.458***	0.244***	0.427***	0.492***	0.520***
	(0.046)	(0.054)	(0.071)	(0.083)	(0.061)
Role novelty (Euclidean)	0.017*	−0.004	0.009	0.023	0.005
	(0.009)	(0.004)	(0.014)	(0.019)	(0.012)
Artist FE	No	Yes	No	No	No
Genre, month FE	Yes	Yes	Yes	Yes	Yes
Observations	13,338	12,638	3,321	3,965	6,039
R^2	0.527	0.818	0.644	0.527	0.525

Note. Linear probability model of the voluntary disclosure choice among Period 1 AI songs (Table 1, Panel B): the dependent variable equals one if the song credits the AI tool to listeners (credited) and zero if the AI use is concealed. This is a descriptive choice model, not a causal design; it documents systematic selection into disclosure, which motivates the policy experiment in Table 5. Songs in a genre new to the artist, songs with larger production teams, and songs in which the artist holds a larger own-role share are each more likely to credit the AI; the genre-novelty association is concentrated among bottom-tier artists (column 3). Tiers defined by pre-adoption follower terciles. Standard errors clustered by artist in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Song Performance Under Voluntary Disclosure: A Conditional Comparison

Panel A: Within-artist performance coefficients, by song type	(1) Concealed AI	(2) Credited AI	(1)−(2) Gap
Plays	0.880^{***}	-0.277^{**}	1.157^{***}
	(0.143)	(0.116)	(0.142)
Effective plays	0.749^{***}	-0.533^{***}	1.281^{***}
	(0.115)	(0.109)	(0.127)
Full plays	0.575^{***}	-0.599^{***}	1.174^{***}
	(0.104)	(0.101)	(0.119)
Likes	0.241^{***}	-0.482^{***}	0.722^{***}
	(0.056)	(0.062)	(0.071)
Collections	0.376^{***}	-0.429^{***}	0.804^{***}
	(0.075)	(0.073)	(0.086)
Play duration	1.582^{***}	-0.295^{*}	1.877^{***}
	(0.254)	(0.162)	(0.223)
Panel B: Audio-content differences, concealed vs. credited AI songs (feature blocks)	Dims	Mean |std. diff.|	Joint p
Mood & affect	12	0.41	<0.001
Acoustic instruments	11	0.22	<0.001
Electronic & rhythmic instruments	13	0.27	<0.001
Other acoustic instruments	15	0.23	<0.001
Tempo & tonality	2	0.19	<0.001
Vocal presence	3	0.25	<0.001
All features (joint)	56	0.27	<0.001
Artist FE	Yes	Yes	Yes
Relative-month FE	Yes	Yes	Yes
Observations (song–month)	95,929	95,929	95,929

Note. **This table reports descriptive, within-artist conditional comparisons, not causal estimates:** Table 3 shows disclosure is strategically chosen, so these gaps combine any listener response with selection. The causal estimate is in Table 5. Panel A estimates the within-artist specification of equation (2) on the adopters' song–month panel, with artist fixed effects and relative-month (months-since-adoption) fixed effects; the omitted reference category is the artist's own non-AI songs. Columns 1–2 report β_C and β_R, each song type's average performance relative to the same artist's non-AI catalog; column 3 reports the concealed−credited gap $\beta_C - \beta_R$ (plays: 1.16 log points, ≈3×), estimated as a linear contrast (`lincom`) with its own standard error in parentheses and significance stars. The duration contrast (large concealed premium, no credited penalty) indicates the gap operates through discovery rather than engagement depth. The matched-pair difference-in-differences version, which additionally uses the non-adopter controls, is reported in Appendix D. Panel B summarizes audio-content differences between concealed and credited AI songs over the feature space of Section 4, grouped into blocks; per block we report the number of dimensions, the mean absolute standardized difference (Cohen's d) between the two groups, and the p-value of a joint equality test (Hotelling T^2). The purpose is to characterize whether the performance gap reflects differences in measured musical content; block differences are controlled in the conditional specification and do not track the direction of the plays gap. Each outcome is the natural logarithm of one plus its 28-day count, winsorized at (0.1, 99.9); standard errors clustered by artist in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: The Causal Effect of Mandatory AI Labeling: Difference-in-Differences Around September 18, 2025

	(1) Plays	(2) Effective	(3) Full plays	(4) Likes	(5) Collections	(6) Comments
Panel A: Headline specification						
Treat × Post	−0.047***	−0.029**	0.002	0.001	0.001	0.000
	(0.015)	(0.015)	(0.013)	(0.002)	(0.003)	(0.001)
Song FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Song-age control	Yes	Yes	Yes	Yes	Yes	Yes
Observations (song–week)	132,440	132,440	132,440	132,440	132,440	132,440
Panel B: Without calendar-week fixed effects						
Treat × Post	−0.020**	−0.017*	0.000	0.001	0.007***	0.000
	(0.010)	(0.009)	(0.008)	(0.001)	(0.002)	(0.000)
Song FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar-week FE	No	No	No	No	No	No
Song-age control	Yes	Yes	Yes	Yes	Yes	Yes
Observations (song–week)	132,440	132,440	132,440	132,440	132,440	132,440
Songs (treated / control)			5,593 / 4,246			
Control-group mean (raw)	20.780	16.217	13.296	0.071	0.150	0.001
Control-group SD (raw)	181.609	147.125	119.267	0.816	1.510	0.037

Note. This table presents song-by-week difference-in-differences estimates of how the mandatory "AI-assisted" label, displayed platform-wide beginning September 18, 2025, affected the performance of AI songs whose AI use had been concealed from listeners. The estimation sample is the mandatory-regime sample summarized in Table 2, Panel B: AI-assisted songs whose artists reported their AI use to the platform but did not credit it to listeners, published in the roughly six weeks on either side of the April 15, 2025 registration threshold (control songs March 1–April 14, treated songs April 15–May 31, 2025). Treated songs, published after the threshold, received the public label when it was introduced on September 18, 2025; control songs, published before the threshold, remained unlabeled throughout. Each outcome is measured as the natural logarithm of one plus its weekly count. Panel A reports the headline specification, which absorbs song and calendar-week fixed effects and controls for song age; Panel B omits the calendar-week fixed effects. The estimated penalty is confined to the play-based (discovery) outcomes, while the engagement outcomes—full plays, likes, collections, and comments—are statistically indistinguishable from zero in the headline specification. Because the platform's recommendation system treats labeled and unlabeled AI songs identically throughout (Section 3.3), these estimates isolate the listener response to the label. No pre-label coefficient differs significantly from zero (Figure 2); the falsification test and heterogeneity by artist tier appear in Table 6. The control-group mean and standard deviation report the raw (un-logged) weekly count of each outcome among control songs during the pre-shock period. Standard errors clustered by song appear in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Falsification Test and Heterogeneity by Artist Tier

	(1) Plays	(2) Effective	(3) Full plays	(4) Likes	(5) Collections	(6) Comments
Panel A: Falsification — songs already crediting AI before the label						
Treat × Post	−0.061	−0.084	−0.037	0.010	0.018*	0.001
	(0.054)	(0.053)	(0.046)	(0.006)	(0.010)	(0.001)
Song FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Song-age control	Yes	Yes	Yes	Yes	Yes	Yes
Observations (song–week)	3,584	3,584	3,584	3,584	3,584	3,584
Panel B: Headline effect by artist tier (sample splits)						
Bottom tier	−0.043**	−0.047**	−0.005	0.001	−0.000	−0.000
	(0.019)	(0.018)	(0.014)	(0.001)	(0.002)	(0.000)
Observations (song–week)	44,450	44,450	44,450	44,450	44,450	44,450
Middle tier	−0.044***	−0.040***	−0.027**	−0.000	−0.000	0.000
	(0.014)	(0.014)	(0.013)	(0.001)	(0.001)	(0.000)
Observations (song–week)	44,058	44,058	44,058	44,058	44,058	44,058
Top tier	−0.004	0.000	0.011	0.002	0.003	0.000
	(0.016)	(0.016)	(0.014)	(0.002)	(0.003)	(0.000)
Observations (song–week)	43,932	43,932	43,932	43,932	43,932	43,932
Song FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Song-age control	Yes	Yes	Yes	Yes	Yes	Yes

Note. This table reports two analyses that accompany the headline mandatory disclosure effect of Table 5, estimated on the same mandatory-regime sample and specification (song and calendar-week fixed effects and a song-age control, with standard errors clustered by song). Panel A is a falsification test that re-estimates the design on the previously-credited pair: AI songs whose artists had voluntarily credited the AI tool before the label, so that the September 18, 2025 label conveyed no new information to listeners. The effect is statistically indistinguishable from zero, consistent with this logic, though the previously-credited pair is small and the estimate is correspondingly imprecise. Panel B splits the previously-concealed (headline) sample by the artist's pre-period follower tercile, with no higher-order interactions. The penalty is clearest for the middle tier, similar in size but less cleanly identified for the bottom tier, and absent for the top tier; because the recommendation system treats all AI songs alike (Section 3.3), the pattern points to listeners rather than the algorithm, with the label mattering most where audiences hold weak priors about the artist. Each outcome is the natural logarithm of one plus its weekly count. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix

Contents

A The Platform's AI-Declaration and Labeling Mechanics

This appendix records the back-end mechanics summarized in Section 3.2: how artists declared AI use to the platform, how that declaration changed on April 15, 2025, and how the "AI-assisted" label was triggered on September 18, 2025.

The pre-April 15, 2025 upload interface. During the voluntary regime, the upload form carried a single optional checkbox—"Harmony.ai-assisted creation," under a "creation support (optional)" heading. Helper text framed it as copyright protection: an artist who checked it was assured full ownership and a 100% revenue share on the track. Checking the box was the only way a song entered Harmony.ai's dedicated back-end record. The form offered no generic "is this an AI song?" control and no field for naming a non-Harmony.ai tool, so artists using other AI systems could only disclose, if at all, through the public credit list.

The April 15, 2025 redesign. On April 15, 2025 the platform replaced the optional Harmony.ai checkbox with a mandatory toggle—"is this an AI song?"—and a free-text field for naming the AI tool(s) used, and simultaneously closed the dedicated Harmony.ai upload channel. Declarations were thereafter recorded in the platform-wide song-tag table as artist-submitted AI declarations rather than in Harmony.ai's separate record.[16]

Dating the change. The platform retained no official go-live date for the April 15, 2025 redesign. We date it from the first large-scale appearance of artist-submitted AI declarations in the back-end data, cross-validated across the Harmony.ai record and the platform-wide song-tag table. The date is therefore empirically inferred rather than publicly announced.

The label-trigger rule and compliance. When public display began on September 18, 2025, the "AI-assisted" label fired under a single rule: a song is labeled if and only if an artist manually declared

[16] In the platform's schema the post-April 15, 2025 declaration is stored as `ai_source = artist_manual` with a positive `is_ai_song` flag; songs published earlier and detected as AI by the platform's library algorithm carry `ai_source = rep_alg`. The public "AI-assisted" label is triggered only by the former.

it as AI at upload. Songs identified as AI through other channels—the platform's library or artist-side detection algorithms—were not labeled, regardless of how long they had been on the platform. Two pieces of back-end evidence indicate that artists declared truthfully once declaration became mandatory. First, the channel through which artists' songs had previously been algorithmically tagged as *not* AI collapsed from a stock of roughly 8.8 million songs to about 2,800 newly tagged songs after April 15, 2025. Second, we manually sampled two songs from every combination of AI-source channel and AI flag on the live platform and confirmed that the "AI-assisted" label displayed exclusively on manually declared AI songs.

B The Matched Non-Adopter Sample

Our main analysis is within artist: each adopter's AI-assisted songs are benchmarked against the non-AI songs the same adopter released (Section 4.1). This appendix describes the matched sample of non-adopters used for two secondary purposes—the across-artist robustness benchmark and the adoption-output result of Section 4.6—and reports covariate balance.

Construction. The comparison group consists of non-adopting artists—those who never used Harmony.ai—drawn from a pool of 78,900 artists who had released more than two songs by the July 2023 launch and were active in the preceding year. We match each adopter to one non-adopter by one-to-one nearest-neighbor propensity-score matching with replacement and a caliper of 0.01, on six artist characteristics: age, gender, cumulative song count, follower count, platform tenure, and platform-assigned artist level.[17] Because an artist's characteristics evolve, we match *month by month*: for each of the sixteen adoption cohorts from July 2023 to October 2024 we measure covariates as of the first of the adoption month, match within that cohort, and assign each matched non-adopter the adoption date of its paired adopter. This yields 5,991 adopters and 5,637 matched non-adopters, balanced on all six covariates.

Covariate balance. Table A1 reports covariate means for adopters and their matched non-adopters. The main analysis is within artist and does not rely on this match; balance supports only the across-artist robustness comparison (Appendix D) and the adoption-output result (Appendix C).

[17] Age is winsorized at 1%; song count and follower count enter in logs given their skew. Controls matched to more than one adopter—a consequence of matching with replacement—are duplicated with unique identifiers so that each adopter–control pair is a distinct observation.

Table A1: Covariate Balance Across Monthly Matching Rounds

	Balance-test p-value						
Matching round	ln(Catalog)	ln(1+Fans)	Tenure	Level	Female	Age	N (T/C)
2023-07	0.715	0.473	0.464	0.270	0.778	0.102	152 / 152
2023-08	0.244	0.892	0.472	0.642	0.228	0.602	670 / 665
2023-09	0.101	0.869	0.967	0.173	0.036	0.359	480 / 477
2023-10	0.311	0.818	0.385	0.616	0.276	0.118	457 / 455
2023-11	0.980	0.264	0.658	0.886	0.303	0.380	371 / 371
2023-12	0.290	0.315	0.501	0.446	0.501	0.573	397 / 393
2024-01	0.348	0.976	0.132	0.136	0.602	0.495	362 / 360
2024-02	0.919	0.868	0.861	0.222	0.731	0.888	317 / 317
2024-03	0.797	0.681	0.650	0.216	0.538	0.668	390 / 389
All rounds	Adopters vs. matched non-adopters						5,991 / 5,637

Note. Each cell is the p-value from a test of equal means between AI adopters and their propensity-score-matched non-adopters, computed within the monthly matching round shown (1:1 nearest-neighbor with replacement, caliper 0.01, on the six covariates listed). The nine rounds span the adopter window (July 2023–March 2024; Section 4.1). Balance is maintained round by round: nearly every p-value exceeds conventional thresholds, with a single exception (Female, 2023-09). The final row reports the pooled sample sizes. Covariates: ln(catalog size), ln(1+followers), platform tenure, artist level, an indicator for female, and artist age.

C Adoption and Production

Section 4.6 summarizes the effect of Harmony.ai adoption on an artist's output. Table A2 reports the staggered difference-in-differences estimates, comparing adopters to their matched non-adopters (Appendix B). This result is context for the paper, not one of its contributions.

D Across-Artist Benchmark for the Conditional Gap

As a robustness check on the within-artist comparison of Table 4, we re-estimate the conditional gap as a matched-pair difference-in-differences that adds the non-adopter controls (Appendix B) as a counterfactual: we interact Treat × Post with the concealed, credited, and non-AI song-type indicators on the matched adopter–control panel, absorbing artist and month fixed effects and clustering by artist. Table A3 reports the estimates, which are consistent with the within-artist results of Table 4.

Table A2: AI Tool Adoption and Music Production: Matched Staggered Difference-in-Differences

	ln(1 + songs released), artist–month	
	(1) Total	(2) Non-AI songs
Adopt × Post	0.305***	0.051***
	(0.019)	(0.019)
Oster $\beta^*(\delta{=}1)$	0.245	−0.004
Artist FE	Yes	Yes
Relative-month FE	Yes	Yes
Matching weights	Yes	Yes
Observations	37,238	37,238
R^2	0.531	0.519

Note. Staggered difference-in-differences on the matched artist–month panel (Table 1, Panel A); controls inherit their matched adopter's adoption date, and relative-month fixed effects are calendar months measured relative to the artist's first AI adoption. Both outcomes are defined before and after adoption, so their difference-in-differences measure genuine changes: the rise in total output alongside essentially unchanged non-AI output implies that adopters add AI songs on top of their human output rather than displacing it. We do not report a difference-in-differences on AI-song counts, because AI songs do not exist before adoption and that outcome therefore has no pre-period against which an effect could be identified. Outcomes winsorized at (0.1, 99.9) and log-transformed; specifications include time-varying controls for catalog size and followers. Standard errors clustered by artist in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Conditional Gap: Matched-Pair Difference-in-Differences

	(1) Concealed AI	(2) Credited AI	(3) Non-AI	(1)−(2) Gap
Plays	0.755^{***}	-0.401^{**}	-0.124	1.156^{***}
	(0.189)	(0.179)	(0.156)	(0.142)
Effective plays	0.684^{***}	-0.597^{***}	-0.064	1.281^{***}
	(0.173)	(0.174)	(0.155)	(0.127)
Full plays	0.469^{***}	-0.704^{***}	-0.105	1.173^{***}
	(0.164)	(0.165)	(0.148)	(0.119)
Likes	0.240^{***}	-0.485^{***}	0.003	0.724^{***}
	(0.085)	(0.090)	(0.073)	(0.072)
Collections	0.348^{***}	-0.457^{***}	-0.024	0.805^{***}
	(0.112)	(0.113)	(0.098)	(0.086)
Play duration	1.348^{***}	-0.528^{*}	-0.232	1.876^{***}
	(0.329)	(0.291)	(0.272)	(0.223)
Artist FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Non-adopter controls	Yes	Yes	Yes	Yes
Observations (song–month)	116,804	116,804	116,804	116,804

Note. **This table reports descriptive estimates, not causal ones.** This is the matched-pair difference-in-differences version of the conditional gap: it estimates Treat × Post × {concealed, credited, non-AI} on the matched panel of adopters and their non-adopter controls (Appendix B), with artist and month fixed effects, clustered by artist. It parallels the within-artist Table 4 but uses the non-adopter controls as the counterfactual. Column 4 is the concealed−credited gap ($\beta_C - \beta_R$) from a linear contrast (`lincom`), with its standard error in parentheses and significance stars. Each outcome is the natural logarithm of one plus its 28-day count, winsorized at (0.1, 99.9). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.