

BEFORE YOU MASTER — GUIDE A

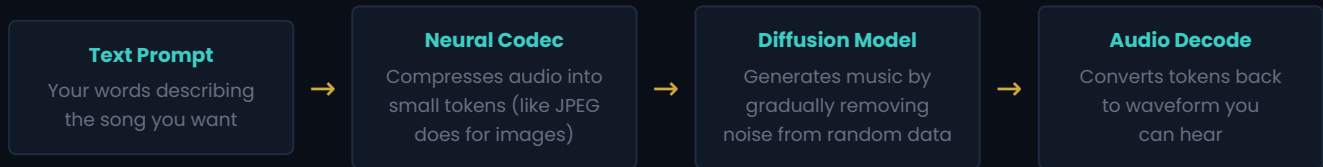
Why Your AI Track Sounds Wrong

The Science Behind Suno's Audio Artifacts —
And What You Can Actually Do About Them

By Petri Korhonen • March 2026

How AI Music Is Actually Made

Before we can understand what goes wrong, we need a basic picture of how tools like Suno and Udio turn a text prompt into a finished audio file. The process has three major stages, and each one introduces its own type of quality limitation.



PLAIN ENGLISH

Think of it like this: your text prompt is a recipe. The neural codec is like a very lossy JPEG compression — it squeezes the entire richness of audio into a small set of numbers, and some detail is permanently lost in the process. The diffusion model is the chef who follows the recipe, but works by starting with random noise and gradually carving the music out of it. The decoder turns the result back into a sound wave you can play.

Technical Detail

Neural audio codecs (like Meta’s EnCodec or Descript Audio Codec) operate at bitrates of 1.5–6 kbps — roughly 30–100x more compressed than MP3. At this compression level, the codec must make aggressive decisions about what spectral information to keep and what to discard. High-frequency detail, stereo spatial information, and transient edges are the first casualties.

The diffusion model then generates content within this compressed representation. When it “over-smooths” uncertain regions (which it does to minimize prediction error), the result is audible: frequency bands that should have natural variation become unnaturally flat. When it “over-sharpens” certain features, you get artificial peaks that weren’t in the training data.

Each stage has its own failure mode. The codec loses detail by design. The diffusion model adds artifacts when uncertain. And the decoder can’t restore information that was already lost. Understanding this pipeline helps you grasp why the three artifact types we’ll examine next appear so consistently across every AI generator — they’re structural, not random.

KEY TAKEAWAY

AI-generated audio has never been “mixed” or “mastered.” It’s the raw output of a mathematical process that optimizes for musical content — melody, harmony, arrangement — not for audio fidelity. The artifacts are baked into the generation process itself.

Artifact #1: Shimmer

What you hear

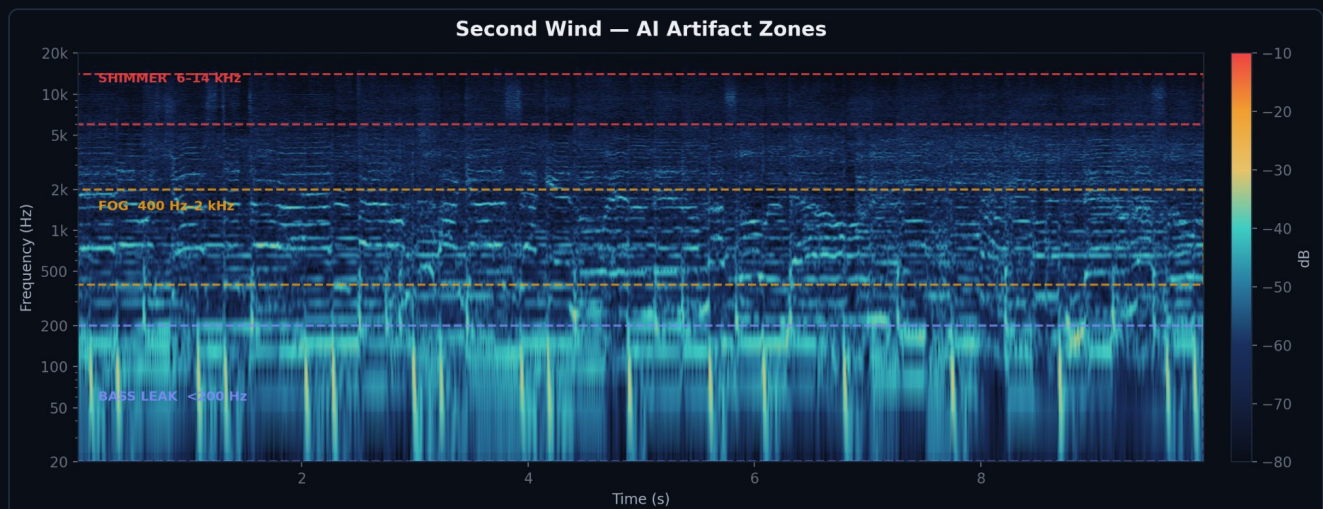
A metallic, glassy brightness on top of everything — like listening through a slightly distorted speaker or a badly tuned car radio with the treble turned too high. Cymbals sound harsh and artificial. Vocals have an unnatural “sizzle.” The entire high end feels brittle and fatiguing after a few minutes of listening.

What it actually is

Shimmer lives in the 6–14 kHz frequency range. It’s caused by the neural codec’s quantization process — when the codec compresses high-frequency content into discrete tokens, it creates artificial energy peaks that are abnormally stable across time. In real recorded music, high frequencies fluctuate naturally from moment to moment. AI-generated shimmer doesn’t fluctuate — it sits there like a metallic coating over the entire mix.

We measure shimmer as the percentage of total energy concentrated in the 6–14 kHz band. Professional masters typically sit at 1–3%. Heavy shimmer reaches 5–7%.

Spectrogram evidence — Second Wind (shimmer: 6.4%):



Second Wind — note the persistent bright energy in the 6–14 kHz shimmer zone (red dashed). This track has the highest shimmer at 6.4% of total energy.

WHY IT MATTERS

Shimmer causes **listener fatigue**. Your nervous system is wired to notice persistent high-frequency anomalies — they register as something “wrong” even when you can’t consciously name it. On cheaper playback systems (earbuds, phone speakers, Bluetooth), shimmer becomes more pronounced because these devices already emphasize treble.

Artifact #2: Fog

What you hear

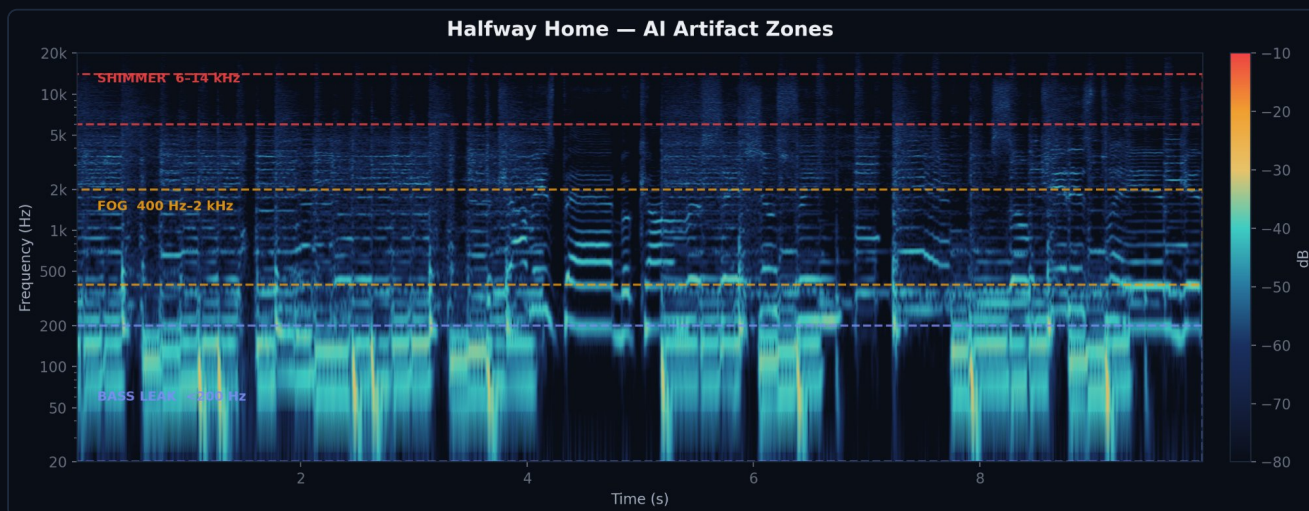
A muddy, undefined quality in the midrange — like listening through a thick blanket or in a room with too much reverb. Individual instruments lose clarity and blend into each other. Vocals sound “underwater.” The mix lacks punch and definition. Everything is there, but nothing is sharp.

What it actually is

Fog appears in the 400 Hz–2 kHz critical midrange — the frequency band where human hearing is most sensitive and where most musical information lives. It’s caused by the diffusion model’s tendency to “over-smooth” uncertain spectral regions. Rather than committing to specific harmonic content, the model spreads energy evenly across frequencies, creating an unnaturally flat spectral profile.

We measure fog using **spectral flatness** — the ratio between the geometric and arithmetic mean of the frequency spectrum. A value of 0.0 means pure tone (no fog); 1.0 means pure noise (maximum fog). Real music typically measures 0.15–0.35. AI tracks commonly hit 0.45–0.65.

Spectrogram evidence — Halfway Home (fog: 0.589):



Halfway Home — observe the dense, uniform energy fill in the 400 Hz–2 kHz fog zone (orange dashed). The lack of clear harmonic structure is visible as a “haze” rather than distinct spectral lines.

WHY IT MATTERS

Fog robs your track of **clarity and impact**. The 400 Hz–2 kHz range is where the human voice has most of its intelligibility, where guitar and piano have body, and where drums have punch. When this range is foggy, everything sounds “fine” at low volumes but falls apart when you try to listen critically or at normal volume on a decent system.

Artifact #3: Stereo Bass Leak

What you hear

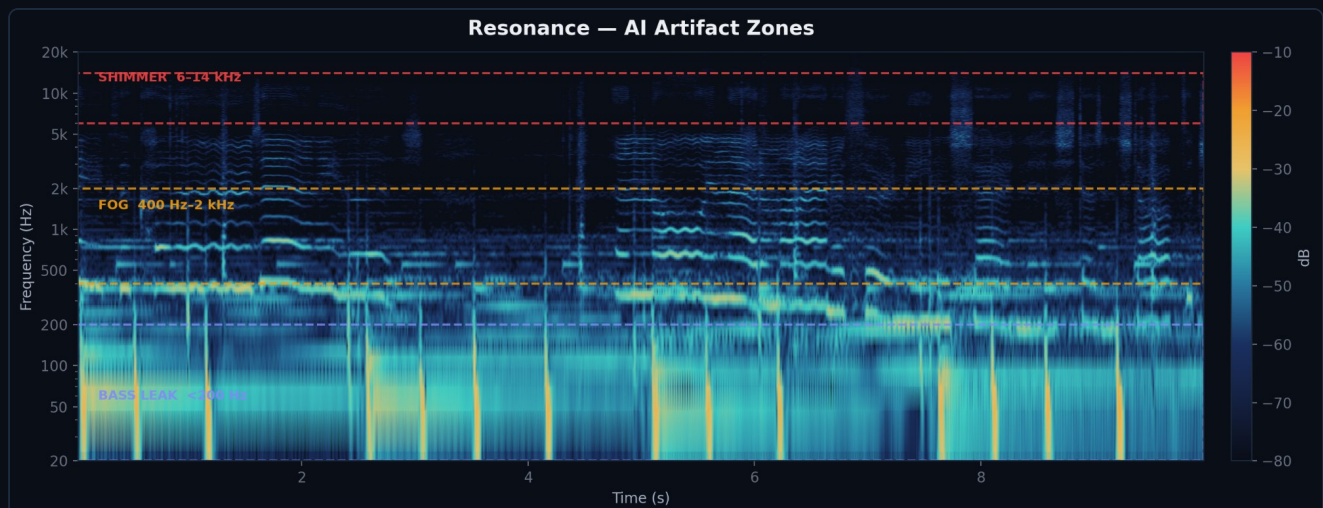
Bass that sounds different depending on what you listen through. On headphones, the low end seems wide and full. On a phone speaker or cheap Bluetooth device, the bass virtually disappears. In a car, it might sound boomy on one side. The low-end is fundamentally unstable — it won't translate consistently across playback systems.

What it actually is

Bass frequencies below 200 Hz should be almost entirely mono (center) in any well-produced track. This is a foundational mixing principle — bass takes enormous energy to reproduce, and splitting it across stereo channels causes phase cancellation when played back on mono systems (which includes most phone speakers, many Bluetooth speakers, and club PA systems).

AI generators don't enforce this rule. They spread bass energy into the stereo field because the model doesn't "know" about real-world playback. We measure this as the **side/mid energy ratio below 200 Hz**. A professional mix typically shows 0.01–0.03. AI tracks commonly reach 0.05–0.15, meaning 5–15x more stereo bass than they should have.

Spectrogram evidence — Resonance (bass leak: 0.016):



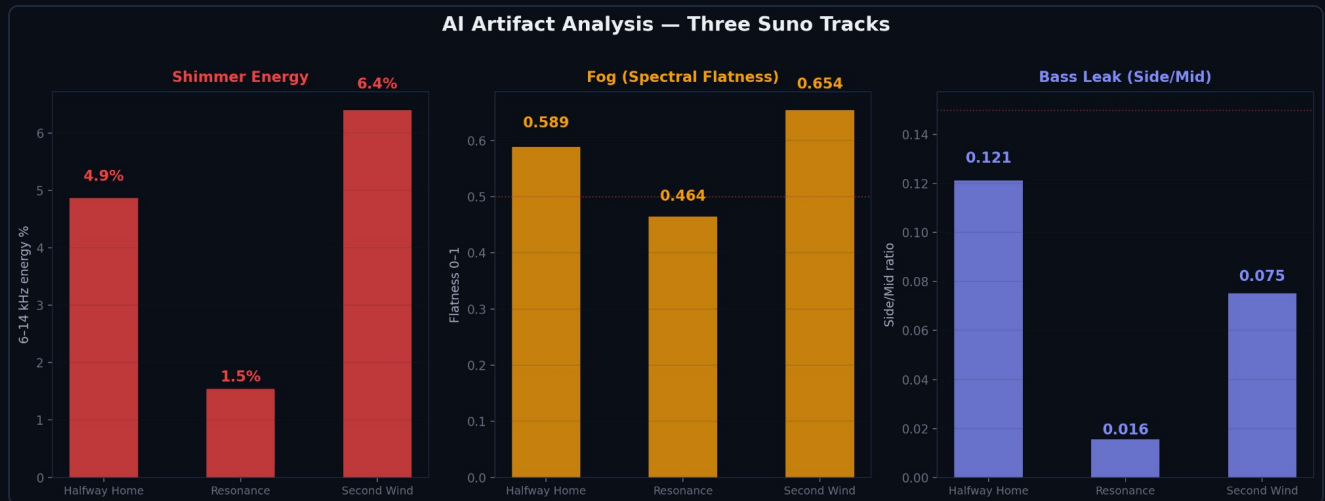
Resonance — the cleanest track of the three. Note the relatively controlled low-frequency region below 200 Hz (purple dashed). Side/mid ratio of 0.016 is close to professional levels.

WHY IT MATTERS

Bass leak is the artifact most likely to make your track sound **amateur on everyday devices**. Most people listen on phone speakers, earbuds, and Bluetooth speakers — all of which sum stereo to mono or near-mono. If your bass is in the stereo field, it literally cancels itself out on these systems. Your track sounds thin and weak compared to any commercially released song playing next in the queue.

Real Data: Three Tracks Compared

To demonstrate that these aren't theoretical concepts, we analyzed three real Suno-generated tracks through spectral analysis. All three were generated with similar settings but different arrangement complexity. The differences are dramatic.



Comparative analysis of three Suno tracks. Left: shimmer energy (6–14 kHz). Center: spectral flatness / fog (0–1). Right: stereo bass leak (side/mid ratio).

The Numbers

Second Wind — Shimmer: 6.4%, Fog: 0.654, Bass leak: 0.075. The most artifact-heavy track. Dense arrangement with multiple competing elements.

Halfway Home — Shimmer: 4.9%, Fog: 0.589, Bass leak: 0.121. Moderate artifacts, but the highest bass leak — significant stereo low-end issues.

Resonance — Shimmer: 1.5%, Fog: 0.464, Bass leak: 0.016. The cleanest track by far. Simpler arrangement, fewer competing voices.

What This Tells Us

The pattern is clear: **arrangement complexity drives artifact severity**. Resonance, with its simpler structure and fewer simultaneous instruments, comes through the codec with dramatically less damage than Second Wind's busier arrangement.

This is the codec bandwidth effect in action. The neural codec has a fixed information budget. Fewer simultaneous sounds means more bits available per sound, which means less quantization noise (shimmer), less spectral smoothing (fog), and better stereo coherence (less bass leak).

HONEST ASSESSMENT

Even the "cleanest" track (Resonance) has a spectral flatness of 0.464 — still significantly higher than a professionally recorded and mixed track (typically 0.15–0.35). AI artifacts are a matter of degree, not presence vs. absence. Every AI track has them. The question is how severe they are and how much you can reduce them.

04 THE LOUDNESS TRAP

Why Louder Does Not Mean Better

There's a natural human tendency to perceive louder audio as "better." It sounds more exciting, more powerful, more professional. This is a well-documented psychoacoustic phenomenon called the **Fletcher-Munson effect** — at higher volumes, we perceive bass and treble as more balanced, which our brain interprets as "better."

This creates a dangerous cycle with AI music:

The Loudness Trap Cycle

1. You generate a track in Suno → 2. It sounds a bit flat and quiet → 3. You think "I'll just make it louder" → 4. You push it through a limiter or normalize to maximum → 5. The track sounds subjectively "better" because it's louder → 6. But the shimmer, fog, and bass leak are now **amplified proportionally** → 7. When it plays on Spotify next to a properly mastered track, the platform normalizes it back down to -14 LUFS → 8. Now you have the same volume as before, but with worse artifacts.

THE REALITY

Suno output is already compressed by the generation process. The dynamic range is typically narrow by default — often around 4–6 dB. Pushing it through a limiter squeezes this even further, leaving nowhere for musical dynamics to breathe. The result is a track that sounds flat, harsh, and fatiguing — the exact opposite of professional.

05 HONEST ASSESSMENT

What Mastering Can Fix vs. What It Can't

Mastering is the final stage of audio production. It's where frequency balance, dynamics, stereo image, and loudness are optimized for real-world playback. For AI music, mastering is critical — but it's not magic.

What mastering CAN fix

- ✓ Tame shimmer with targeted EQ and de-essing in the 6–14 kHz range
- ✓ Reduce fog with mid-band clarity enhancement and gentle spectral shaping
- ✓ Eliminate bass leak by collapsing low frequencies below 200 Hz to mono
- ✓ Optimize loudness for specific platforms (-14 LUFS for Spotify, -16 LUFS for Apple Music)
- ✓ Improve overall tonal balance and perceived clarity

What mastering CANNOT fix

- ✗ Missing harmonics that the codec discarded during generation
- ✗ Fundamental resolution limits of the neural codec
- ✗ Baked-in distortion from over-compressed generation
- ✗ Poor arrangement choices (too many instruments competing)
- ✗ Structural timing/rhythm issues in the generated content

KEY INSIGHT

Professional mastering can typically reduce artifact severity by 40–60%, making the difference between "obviously AI" and "sounds surprisingly good." But the ceiling is set by the source material. A cleaner generation will always produce a better final result than heroic mastering of a noisy one.

Quality Starts Before Mastering

Here is the most important insight in this entire guide: **80% of your final audio quality is determined before you ever open a mastering tool.** The decisions you make about arrangement, instrument count, and style prompting have far more impact on the final result than any amount of post-processing.

WHY THIS MATTERS

Our spectrogram data proves it. Resonance (simple arrangement, 3 voices) measured shimmer at 1.5% and fog at 0.464. Second Wind (dense arrangement, multiple voices) hit 6.4% shimmer and 0.654 fog. That's a 4x difference in artifact severity — caused entirely by arrangement decisions, not mastering.

Three Things You Can Do Right Now

You don't need to buy anything or learn complex software. These three changes will immediately improve the quality of your next AI generation:

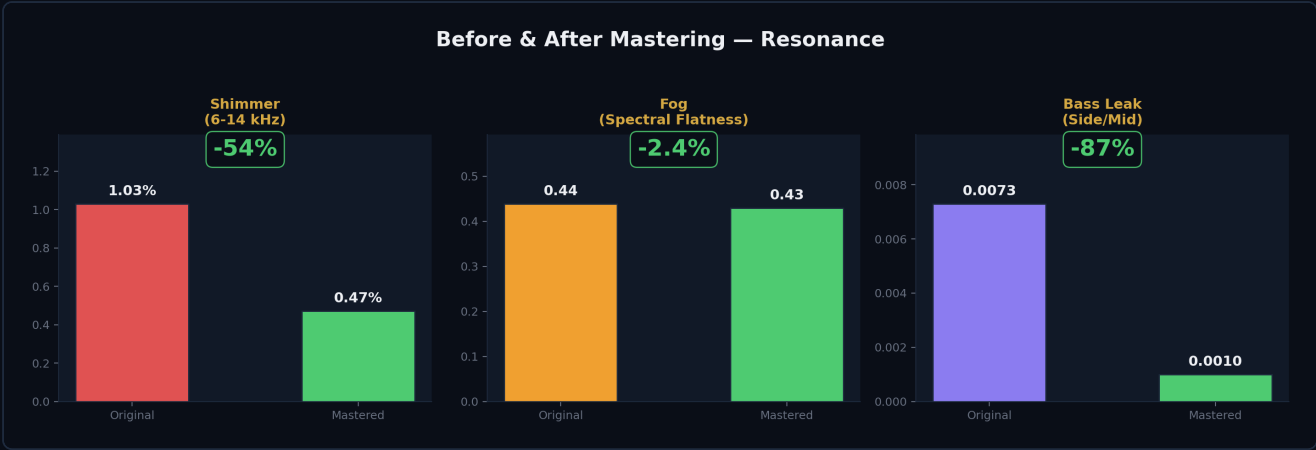
- 1 Keep arrangements to 2–3 simultaneous voices**
The neural codec has a fixed information budget. Fewer sounds competing means more quality per sound. Vocals + guitar + light percussion will always sound cleaner than vocals + guitar + bass + drums + synth pad. This alone can reduce artifact severity by 50% or more.
- 2 Choose instruments that occupy different frequency ranges**
Two instruments in the same frequency range force the codec to choose between them. Piano (200 Hz–4 kHz) + acoustic guitar (200 Hz–4 kHz) = collision. Piano + flute (800 Hz–8 kHz) = clean separation. Pick instruments that don't overlap and the codec can render each one with far more accuracy.
- 3 Test your output on a phone speaker before mastering**
Phone speakers are brutally honest. They expose shimmer (sounds harsh and tinny), reveal bass leak (bass disappears entirely), and amplify fog (everything sounds muffled). If your track sounds decent on a phone speaker before mastering, it's a good source to work with. If it sounds terrible, consider regenerating before investing time in mastering.

COMING NEXT

Guide B — “Arrangement for AI” covers the codec bandwidth problem in depth, with genre-specific arrangement templates you can use immediately. **Guide C — “Prompting Suno Like a Producer”** reveals the style prompt mechanics, seed strategies, and production bleed avoidance techniques that give you maximum control over generation quality.

Before & After: What Mastering Actually Changes

Theory is one thing. Data is another. We took the cleanest track from our analysis — Resonance — and ran it through a mastering chain with AI artifact suppression, de-esser, vocal clarity processing, and bass mono below 200 Hz. Here are the measured results:



Spectral analysis: Resonance before and after mastering with AI-specific processing. All three artifact types reduced.

What improved

- ✓ **Shimmer reduced 54%** — de-esser and artifact suppression cleaned the 6–14 kHz range. From 1.03% to 0.47%, now within professional range (1–3%).
- ✓ **Bass leak reduced 87%** — bass mono processing collapsed stereo low-end to center. Side/mid ratio dropped from 0.0073 to 0.0010, essentially professional-grade.
- ✓ **Fog reduced 2.4%** — subtle but measurable. Spectral flatness improved from 0.439 to 0.429.
- ✓ **Dynamic range preserved** — 10.2 dB → 10.8 dB. Mastering didn’t crush dynamics.

What this means in practice

The mastered version will sound cleaner on every playback system. The 87% bass leak reduction means the low end won’t disappear on phone speakers. The 54% shimmer reduction eliminates listener fatigue on earbuds and headphones. And the preserved dynamic range means the track breathes naturally instead of sounding flat.

This was a *light* mastering pass on the *cleanest* track in our test set. On tracks with heavier artifacts (like Second Wind at 6.4% shimmer), the improvements would be even more dramatic.

THE BOTTOM LINE

Even a track that already measures well benefits measurably from proper mastering. For tracks with heavier artifacts, the difference between “raw AI output” and “mastered” is the difference between amateur and professional. The data proves it. Your ears will confirm it.

Ready to Master Your AI Tracks?

MasterForge was built specifically for AI-generated music. Automated artifact detection, one-click cleanup, and platform-optimized delivery — designed for Suno, Udio, and every AI generator.

masterforge.app

Free tier available • No credit card required • Works in your browser

This guide is part of the “Before You Master” series:

A — Why Your AI Track Sounds Wrong (this guide)

B — Arrangement for AI: Why 3 Instruments Sound Better Than 5

C — Prompting Suno Like a Producer

D — From Suno to Spotify: The Complete Release Pipeline