MUSIC.AI

Edge AI for Audio

Trends, Use Cases, and Predictions

May 2025

Edge AI for Audio: Trends, Use Cases, and Predictions

By Music AI

1. Introduction: Trends in Artificial Intelligence and Audio Processing

The Changing Landscape of Sound

Artificial intelligence has already transformed the way that sound is generated, processed, and organized, both for audio professionals and ordinary consumers. Features such as stem separation, automatic transcription, and voice cloning services have helped deliver improvements in efficiency, accessibility, and customer satisfaction.[1] Recording companies are already managing their catalogs more effectively with smart metadata tagging from tools like Cyanite.

Mixing and mastering has gotten faster and better than ever before with tools like RoEx. Artists are finding clever new ways to hone their craft with tools like our own Moises platform.

Historically, the majority of AI audio processing has happened on the cloud, with more than 70% of total market share being in that segment.[2] Recently, however, the technology has arrived at a crucial turning point.

The Limits of Cloud-Based AI

While its impact is undeniable, the cloud-based AI solutions that enabled many of these innovations also come with limitations.[3]

Latency:	Privacy:	Cost:
Cloud-based solutions come with increased latency, which makes some use cases untenable.	Data privacy and security can be especially challenging with services provided over the internet.	Frequent communication with cloud-based models can be prohibitively expensive and consume large amounts of bandwidth and energy.[3]

In order to achieve the outcomes many customers are looking for, it is essential to run at least some AI models locally. This allows processes to happen much closer to real-time, helps protect customer data, and improves overall efficiency.[4]

Two Sides of Edge AI: Embedded vs On-Device

While embedded use cases could reasonably be considered a form of on-device implementation, the unique hardware limitations and specialized operating systems involved make it useful to distinguish embedded AI as a separate category.

For example, the considerations are fundamentally different when running an AI model on a microcontroller inside an IoT device than when running it as part of a software application on a laptop. For this reason, this paper will use the term “on-device” exclusively to refer to models running on consumer electronics, such as mobile phones, computers, and tablets. The term “embedded AI” will refer to deployments running closer to the hardware on lower-powered devices.

The table below should provide a sense for the differences.

	On-Device Laptop, Mobile, Tablet	Embedded Hardware
	On-Device AI	**Embedded AI**
Definition	AI models run locally on consumer devices (phones, laptops)	AI models run on specialized hardware
Hardware	Smartphones, laptops, tablets	Audio processors, microcontrollers, sensors, smart appliances
OS / Environment	Full OS (Android, iOS, Linux, Windows)	RTOS, bare-metal, embedded Linux
Model Constraints	Medium to large models; uses NPU/GPU acceleration	Highly optimized, quantized, minimal footprint
Updates	App store or OTA	Firmware update, SDK-level flash
Experience Delivered	App-based intelligence (e.g. live audio effects)	Device-integrated AI functionality

The Growing Embedded Market

Recent improvements in both specialized hardware and the models themselves have made it increasingly feasible[5] to deploy AI models on embedded systems, which (along with the benefits outlined above) has led to its growing popularity.

The market was already valued at $9.13 billion in 2023, and it is projected to grow to $29.26 billion by 2032.[6] We are in the early stages of a major realignment, and the opportunities for adoption are substantial.

Considering a Hybrid Solution

While edge AI comes with numerous benefits, there are still plenty of use cases where cloud-based AI might be ideal.[3] For example, running AI models on the cloud allows for the highest possible performance for tasks where latency might be a lower priority than performance. Cloud-based services can also seamlessly add in all of the latest model updates, while edge solutions are either limited to their original versions or receive updates much more manually.

Due to these trade-offs, some products may benefit from a hybrid solution combining both cloud-based and edge AI. To take an example from the audio industry, software developers working on a DAW might choose to run workloads where latency is especially important on-device, while running other, possibly more computationally expensive, workloads on the cloud.

For embedded systems, the device might need to connect to the cloud-based model as little as possible, in order to share data or potentially handle tasks that would otherwise be impossible on such limited hardware. These types of use cases are one reason why product teams should consider working with AI providers who have experience building for both edge and cloud environments.

2. Introducing Embedded, AI-Powered Audio Features

The Music AI Embedded SDK

The Music AI Software Development Kit (SDK) is a collection of artificial intelligence models designed to run stem separation and other AI-powered audio services locally, including on embedded systems. This allows for customers to run AI-powered audio features in near real time, with or without an internet connection. The SDK has support for C/C++, and is compatible with Windows, Mac, and Linux (including embedded Linux), with ports for Android and RTOS now in development.

Definition: Stem separation is a term in audio production for when a track is divided into its individual parts. For example, a professional audio engineer might choose to separate the bass, vocals, and drums, so that they can make adjustments to each part separately. These capabilities go well beyond music, however. For example, cinematic stem separation could also allow filmmakers to lower the volume of background noise in a scene, so that the audience can better hear the characters' voices.

3. Why Music AI?

Customizable configuration

Developed with the specific challenges of edge AI in mind, the embedded SDK includes models with a flexible range of specifications, allowing developers to optimize towards their chosen combination of quality, speed, latency, and hardware limitations

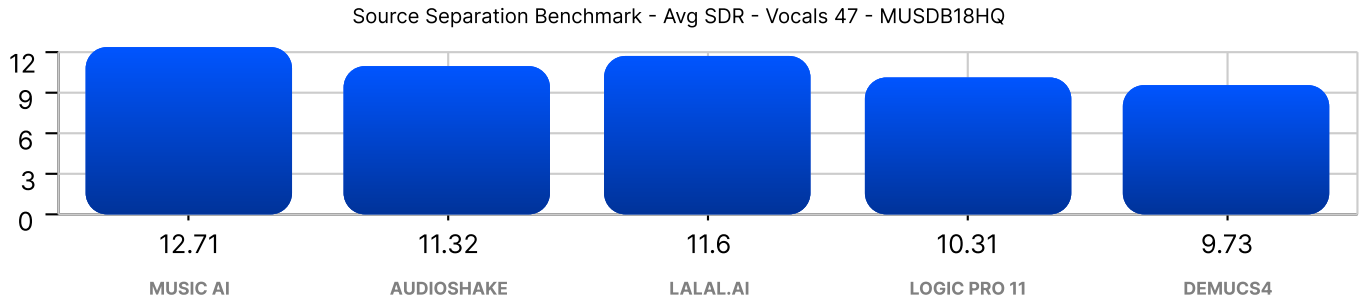
EXAMPLES:

- A desktop developer working on an application for professional audio engineers could select a model that prioritizes sound quality.
- An automotive safety team could select a model that prioritizes real-time performance, with low latency.
- A product team working on a feature for a new set of earbuds could prioritize meeting stringent hardware requirements while still meeting other specifications.

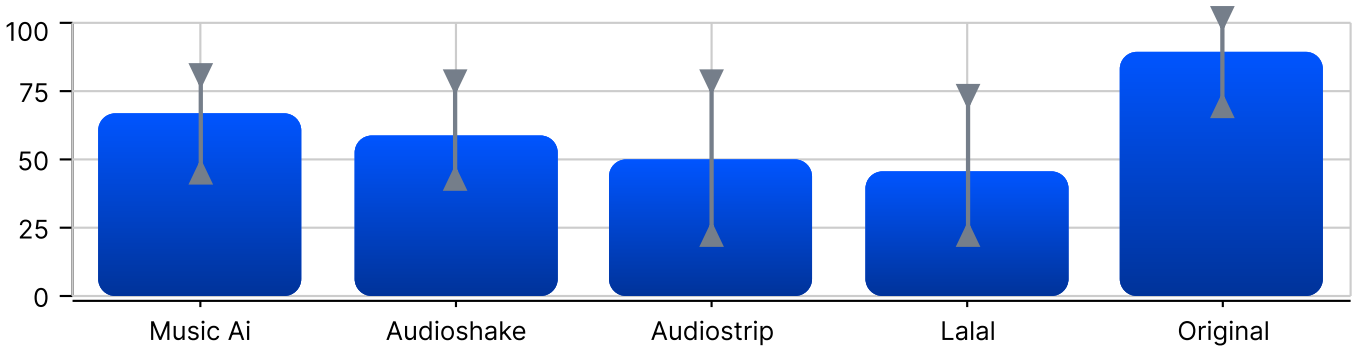
Superior performance

Music AI stem separation technology outperforms competing solutions, with an average SDR that is 15.8% higher than its nearest competitor.[7]

Definition: SDR is an objective measure widely accepted across the audio research community for consistently evaluating separation quality. It measures the power ratio between the intended source and the distortion artifacts introduced during separation. A higher SDR indicates a superior quality of separation and translates to cleaner, more accurate audio stems.



Additionally, in a blind test of trained musicians, participants rated Music AI's stem separation models higher than the alternatives in the vast majority of cases, validating the results of the SDR testing.[8]



Lightweight model options

With the limitations of some embedded devices, we understand the benefits of delivering a model that is lightweight and efficient. We have already offered model implementations as small as 1 MB, so your team will have the flexibility to bring these solutions to a wide range of hardware.

Licensing & Annotation

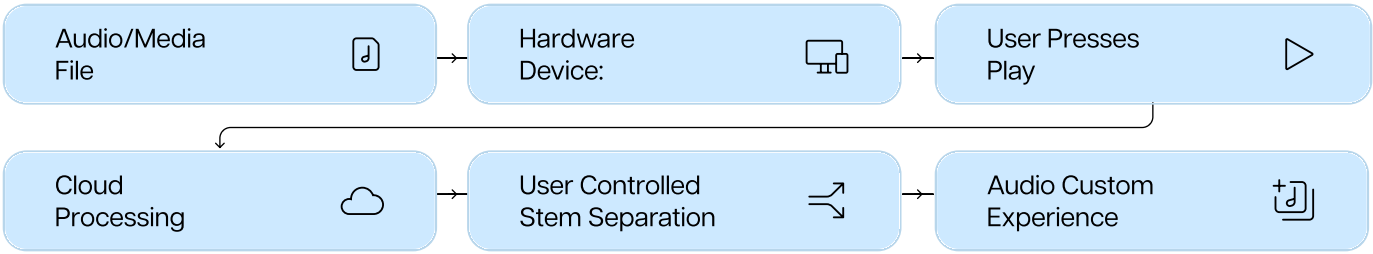
Through licensing deals, in-house recordings, and commissions with professional musicians, Music AI maintains a vast library of the highest-quality data, which we annotate meticulously using our in-house software, as well as the expertise of both data scientists and musicologists. Our intense, proprietary training and data preparation methodology helps us deliver the unmatched sound quality outlined elsewhere in this paper.

Expert Implementation for Embedded Devices

For customers building products on more specialized embedded hardware, Music AI has the engineering and implementation expertise to work with your team to fine-tune our models to match your specifications and hardware requirements.

Hybrid AI Solutions

Music AI is a leading provider of cloud-based AI audio tools, and we are able to work with your team to offer a hybrid solution that integrates the benefits of both cloud-based and edge AI. For example, you can use our API to run workloads demanding the highest possible quality, along with our on-device solution for features that might require a real-time component.



4. Use Cases: Stem Separation on the Edge

On-Device: AI Smart Volume by Moises Live

An early use case for Music AI's embedded SDK was to enable our own Moises Live solution. The AI Smart Volume® feature in Moises Live allows customers to take granular control over the volume of any media running on their computer, giving them the ability to adjust specific sounds separately without changing the overall volume. For example, a customer streaming a TV show can use this tool to raise the volume of the dialogue in real-time without also raising the volume of any background noises. The AI Smart Volume® feature is also totally source agnostic. It works the same for any audio playing through your device, whether that's music, videos, or even live media.

Watch a demonstration.

This level of real-time volume control is only possible because the models are running on-device. A cloud-based solution would experience higher latency and could only run while connected to the internet. This product also provides an excellent example of our SDK's hardware flexibility. Moises Live was initially optimized for Qualcomm NPUs, but it now runs on all modern PCs, with support for Mac and mobile devices also in development.

Qualcomm Partnership: Optimizing for Snapdragon X series

Our team partnered with Qualcomm to take advantage of the advanced neuroprocessing unit (NPU) built into the Snapdragon X series processors. Moises Live runs on additional chipsets, but it still runs with as much as 35x faster processing[9] on the hardware it was optimized for than on a traditional CPU. This success demonstrates both the excellence of the Qualcomm NPU and our team's ability to design solutions targeting specific hardware.

Embedded Use Cases: Smart Speakers and Soundbars

Other great use cases include embedding stem separation technology into the hardware of a smart speaker or soundbar. Similar to the on-device use case, there could be very low latency, which could create an experience that feels like real-time operation. Additionally, the customer could rest assured knowing that any private data remained safely on their device, since there would be no need for any communication over the internet.

This technology would allow customers to enjoy numerous exciting new activities. For example, on the smart speaker, musicians could remove a chosen instrument from a recording, allowing them to more easily practice along with the music.

On the soundbar, cinephiles could replace the single volume bar with personalized control of audio categories like dialogue, sound effects, and background music.

Our SDK could enable these types of product integrations because of its compatibility with embedded Linux and RTOs (in development), as well as its optimized performance on embedded hardware. This is what would allow customers to take advantage of AI-powered audio features directly on a speaker, without running their audio through a PC or mobile device.

5. Predictions and Recommendations

PREDICTIONS

As both the models and the hardware continue to improve, a growing number of audio workloads will move to embedded systems and other edge devices, allowing for more AI processes to happen closer to real-time. This could lead to aspirational features such as near real-time out-loud translation, stem separation and audio editing for in-person performances, real-time selective noise cancellation for IoT devices in industrial settings, real-time audio-based safety features for driverless cars,[10] and so much more.

As the existing solutions continue to mature, consumers will increasingly demand a higher standard of products, until many AI capabilities become base-line requirements in those industries.

RECOMMENDATIONS

Product teams will need to embrace AI product integration quickly and ambitiously to stay ahead of rising customer expectations, both in audio and in other industries. With barriers to entry so high, and the technology evolving so quickly, attempting to develop models in-house could lead to lower-quality outcomes, higher costs, and longer time-to-market.

In most cases, development teams should instead focus on their core product offerings while partnering with the appropriate AI providers to keep pace with any AI-specific functionalities their products might use. That way, firms can maintain their existing competitive advantages more easily, gain access to the highest-quality models, and build the flexibility needed to deploy services wherever they most benefit customers–whether that's on-device, in the cloud, embedded into specialized hardware, or some combination of the three.

MUSIC.AI

Contact us

If your team is interested in integrating stem separation or other AI audio features into your product, please contact our sales team. The goal of this SDK is to help make your product a success.

About Music AI

Founded by Geraldo Ramos, Eddie Hsu, and Jardson Almeida, Music AI is a leader in AI-driven music and audio technology. With a team spanning across the USA, Brazil, and Europe, the company provides solutions through its B2B Music AI and B2C Moises platforms. Music AI's technologies are used by artists, producers, record labels, agencies, technology firms, hardware manufacturers, and developers. With over 50 million users, our technology processes more than 2.5 million minutes of audio every day.

Sources

[1] D. Stammer, "Generative AI in the Music Business," Popakademie, [Online]. Available: https://www.popakademie.de/media/?file=978_241021-ki-whitepaper-en.pdf&download . [Accessed: May 20, 2025].

[2] Market.us, "The Global AI Audio |Processing Software Market," Nov. 2024, [Online]. Available: https://market.us/report/ai-audio-processing-software-market/#:~:text=The%20Global%20AI%20Audio%20Processing%20Software%20Market,36%\%20share%2C%20holding%20USD%201.6%20Billion%20revenue . [Accessed: May 20, 2025].

[3] D. McCarthy, P. Rutten, and C. Kanaracus, "IDC Whitepaper – Edge AI," NVIDIA, Jan. 2022, [Online]. Available: https://www.nvidia.com/en-us/edge-computing/edge-ai/idc-whitepaper/ [Accessed: May 20, 2025].

[4] K. Cao, Y. Liu, G. Meng, and Q. Sun, "An Overview on Edge Computing Research," IEEE Access, vol. 8, pp. 85714–85728, 2020, doi: 10.1109/ACCESS.2020.2991734, [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2991734 [Accessed: May 20, 2025].

[5] Z. Zhang and J. Li, "A Review of Artificial Intelligence in Embedded Systems," Micromachines, vol. 14, no. 5, art. no. 897, 2023, doi: 10.3390/mi14050897, [Online]. Available: https://doi.org/10.3390/mi14050897 [Accessed: May 20, 2025].

[6] Fortune Business Insights, "Edge AI Market," May 05, 2025, [Online]. Available: https://www.fortunebusinessinsights.com/edge-ai-market-107023 [Accessed: May 20, 2025].

[7] Music.AI, "Source Separation Benchmarks," Sep. 17, 2024, [Online]. Available: https://music.ai/blog/research/source-separation-benchmarks/ [Accessed: May 20, 2025].

[8] Music.AI, "Music AI Subjective Analysis: Source Separation," Apr. 14, 2025, [Online]. Available: https://music.ai/blog/research/Music-AI-Subjective-Analysis-Source-Separation/ [Accessed: May 20, 2025].

[9] R. Husseini, "Coming to Snapdragon X Series: New NPU-Powered Creator Tools and DAW Apps," Qualcomm, Oct. 21, 2024, [Online]. Available: https://www.qualcomm.com/news/onq/2024/10/coming-to-snapdragon-x-series-new-npu-powered-creator-tools-and-daw-apps [Accessed: May 20, 2025].

[10] L. Marchegiani and X. Fafoutis, "How Well Can Driverless Vehicles Hear? An Introduction to Auditory Perception for Autonomous and Smart Vehicles," IEEE Intelligent Transportation Systems Magazine, vol. 14, no. 3, pp. 92–105, 2022, doi: 10.1109/MITS.2021.3049425.