



**FRANKFURTER
BUCHMESSE**

WHO IS NARRATING THE FUTURE?

**AI adoption, use and outlook
in audio production**



September 2026

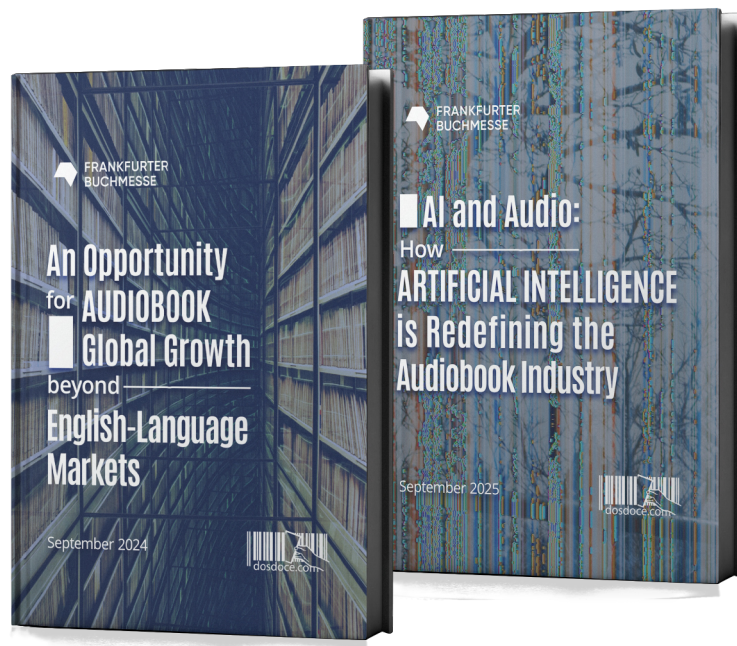
TABLE OF CONTENTS

	Introduction – Executive Summary	3
1	Section 1. AI & Audio Production Adoption, Uses and Outlook	9
	1.1 Consensus: Coexistence, Not Replacement	
	1.2 The Economic Case: Unlocking the Long Tail	
	1.3 Production Workflows and the Supply Chain	
	1.4 Criteria for selecting AI tools and technology providers	
	1.5 Barriers to AI use	
	1.6 Global and Multilingual Expansion	
	1.7 Quality, Differentiation, and Discoverability	
	1.8 Outlook: 2026–2030 and Beyond	
2	Section 2. Outline of the AI and Audio Production Phases	46
	2.1 Defining the AI production strategy	
	2.2 Pre-production and text preparation	
	2.3 Performance tools	
	2.4 Sound effects and music	
	2.5 Quality control and post-production	
	2.6 The three-speed model	
	2.7 Takeaways to adopt AI effectively	
3	Sources	63
4	Credits	64
	About Dosdoce	
	About Frankfurt Buchmesse / Frankfurt Audio	
	Acknowledgements	
	Imprint – Credits	

INTRODUCTION

Executive Summary

Most publishing industry professionals interviewed in last year's report, ***AI and Audio: How Artificial Intelligence is Redefining the Audiobook Industry***¹, published by the Frankfurt Book Fair in October 2025, acknowledged holding a "cautiously progressive" attitude toward the use of artificial intelligence (AI). It was a step forward compared with the "skeptical" mindset expressed in the 2024 report ***An opportunity for audiobook global growth beyond English-language markets***², where they indicated they were following the news and developments in this field, but still did not trust the technology.



In less than a year, we have detected another shift in mindset within the publishing community. In the new report we are publishing this year, ***Who is narrating the future? AI adoption, use and outlook in audio production***, we have moved from the aforementioned "cautiously progressive" attitude to one of "actively testing and redefining production processes," **given that more than two-thirds of respondents indicated they are currently evaluating different types of AI tools**, as well as developing and testing all kinds of workflows in production processes using AI tools.

AI tools have gone from being experimental resources to becoming indispensable components of the audio content workflow — including audiobooks, audio dramas, podcasts, or multi-voiced productions — and they are here to stay. In this environment of transformation, the publishing industry is entering an era of unprecedented change, in which AI technologies accelerate the pace of audio production, potentially reduce costs, and expand the creation of multilingual catalogues.

As we will analyze in detail in Section 1 of this report, **the large majority of respondents (85%) already use AI in at least one stage of production, and a meaningful share (17%) have embedded it across even more than six phases of their workflow** — evidence that AI is moving from isolated experimentation toward operational infrastructure, **but still 15% indicated that they do not use any AI tools at all in their audio production workflow process.**

AI is scaling quickly across the entire production process in many creative industries such as music, films or videogames, from concept to pre-vis through post and delivery. Companies like Netflix have indicated in their quarterly earnings reports³ that thanks to the integration of AI tools in their production processes, they are making higher quality output, more quickly and efficiently than they could have using traditional methods. AI workflows have now been used in roughly 300 of Netflix's titles, with the largest concentration to date in post-production.

Unfortunately, we have not seen this level of AI production adoption in the publishing sector. As we will see later on in this report, very few players in the industry have defined their AI-driven audio production strategy — evaluating the technical capabilities of the different tools and the human impact of its usage on narration, casting, as well as analyzing the potential audio catalogue business opportunities.

On top of this point of departure, most of what is published about producing audio content in the publishing sector with AI tools is written by people who have never produced such content and talks about "magic prompts" and "total automation" as if producing an audiobook were as simple as generating simple text. The reality of production — using AI for

text preparation, narration, proofing the recording and preparing the files for delivery — tells a different story. The process is more rigorous, more complicated, and requiring more artistic input than the leading AI technology providers promise.

To better understand the dynamics of this new landscape of AI tool usage and workflow definition, the Frankfurt Book Fair team encouraged Dosdoce.com to prepare a new report to help book industry professionals (publishers, literary agencies, streaming platforms, distributors, etc.) move from theory to practice.

To achieve this objective, this report is divided into two main sections:

The first section of this report aggregates the opinions, arguments, data points and areas of agreement and disagreement of more than 85 companies and professionals across the globe from leading publishing houses, to audio production studios, as well as distributors and streaming platforms, AI technology providers and industry experts, that have generously contributed with their insights about the main business opportunities and challenges derived from the irruption of AI in the audio production and distribution processes.

As readers will discover in first section of this report, there is little dispute that AI narration, translation, and production tools will expand the audiobook catalogues faster and cheaper than existing production studio economics ever allowed, particularly for backlist, academic, niche, and non-English-language titles.

There is equally little dispute that premium, flagship, and emotionally significant titles will continue to command human narration, because that is where audiences currently place the greatest value on performance.

The open questions are less about whether coexistence will happen than about how the industry manages its side effects: closing the gap between AI narration's quality and consumers' response to it; resolving a rights and licensing framework built for a pre-AI cost structure;

protecting narrators, translators, and authors from uncompensated use of their voices and texts in training data; and ensuring that the regions and languages with the most to gain from cheaper production, such as Latin America, Eastern Europe, the Middle East and Africa, are not left out of the investment in ethical, representative voice data that would let them benefit.

The recurring thread across nearly every response is that growth is not sustainable without transparency and disclosure toward both authors and listeners, and that trust, once spent, will be far costlier to rebuild than any single title was to produce.

The second section of this report describes an outline of the six phases that we consider essential when producing audio content with AI tools, as well as the fundamental role and weight of the human team in each of the phases:

Phase 1 — Defining the AI production strategy

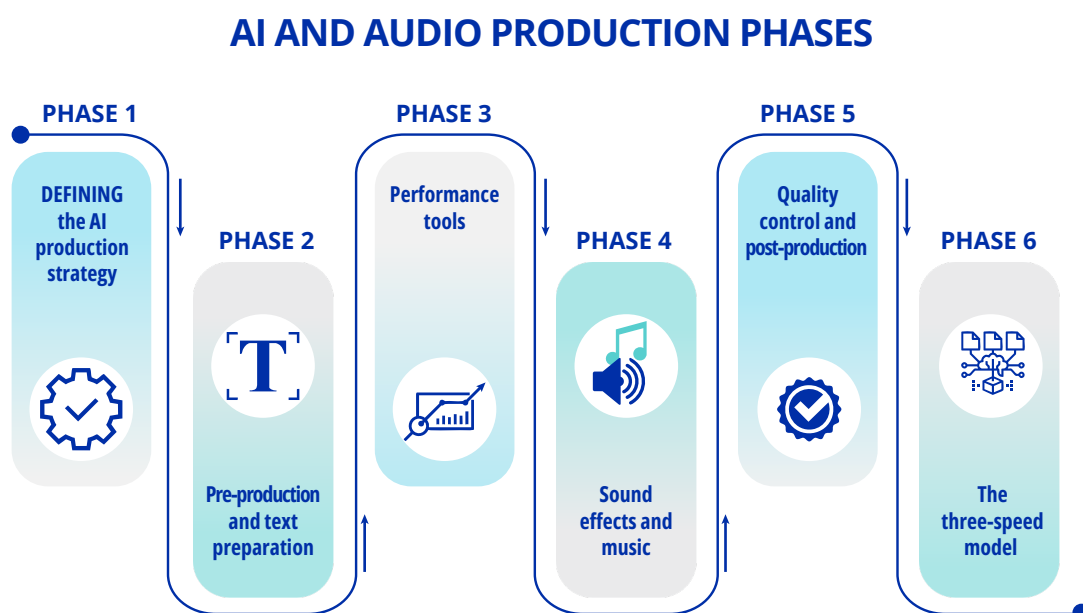
Phase 2 — Pre-production and text preparation

Phase 3 — Performance tools

Phase 4 — Sound effects and music

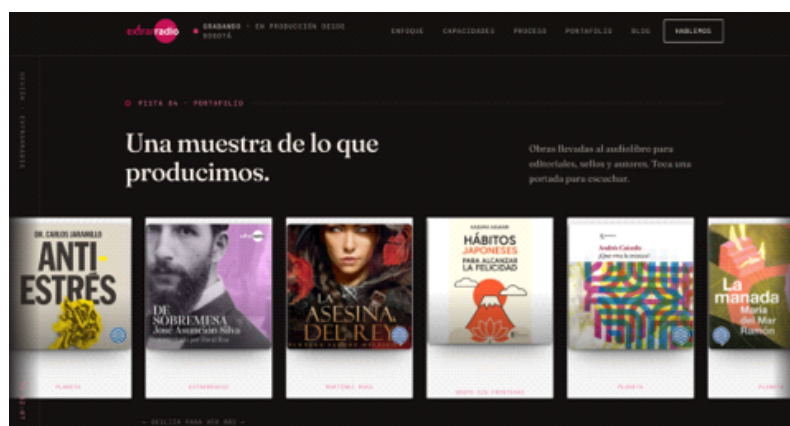
Phase 5 — Quality control and post-production

Phase 6 — The three-speed model



The goal is not to list tools (which change every season) but to extract the editorial criteria, workflows, and lessons learned that allow AI to be adopted effectively, without sacrificing the editorial rigor that distinguishes a professional editorial-driven product from a mere technical demonstration.

The boundaries between human-created and machine-created audio are blurring rapidly. We are at a defining moment in which various AI tools are enabling creators and producers to rewrite and enrich their manuscripts, create sound effects, design covers, and translate content into dozens of languages. There are also AI agents that tell us which new titles to produce next based on the performance of previous productions. These tools are helping publishers scale audiobook production, expand distribution opportunities worldwide, and generate new revenue streams for their authors. “We do not foresee a total replacement of human narration. Our own production model already operates under what we internally call «three speeds»: 1) human narration for *premium* or flagship titles; 2) a hybrid model in which AI handles the mechanical phases (text preparation, quality control, draft synthesis) while a human narrator and editor retain control of the performance; and 3) fully synthetic narration reserved for backlist titles or those requiring less interpretive demand, whose production would not otherwise be economically viable” says **David Roa, founding partner of Extrarradio Producciones (Colombia) and co-author of the report *Who is narrating the future? AI adoption, use and outlook in audio production.***



In conversations held with the more than 85 worldwide industry experts who helped us prepare this report, most of them point out that AI represents one of the sector's greatest challenges, but also one of its greatest business opportunities. **"85% of respondents use AI in at least one phase of audio production, yet only 17% have integrated it across more than six audio production workflow phases — suggesting real usage still lags the adoption headline.** The data, insights and testimonials gathered in this report highlight the remarkable potential for growth, innovation, and accessibility, as well as the implications related to cultural diversity and quality control, but above all, the need to ensure the protection of narrators and creators' rights and full transparency toward listeners. We are confident that by the end of this decade, we will see the coexistence of human production and AI production, making viable titles that were not viable before, without displacing human narration," says **Javier Celaya, founding partner of Dosdoce.com and co-author of the report *Who is narrating the future? AI adoption, use and outlook in audio production.***

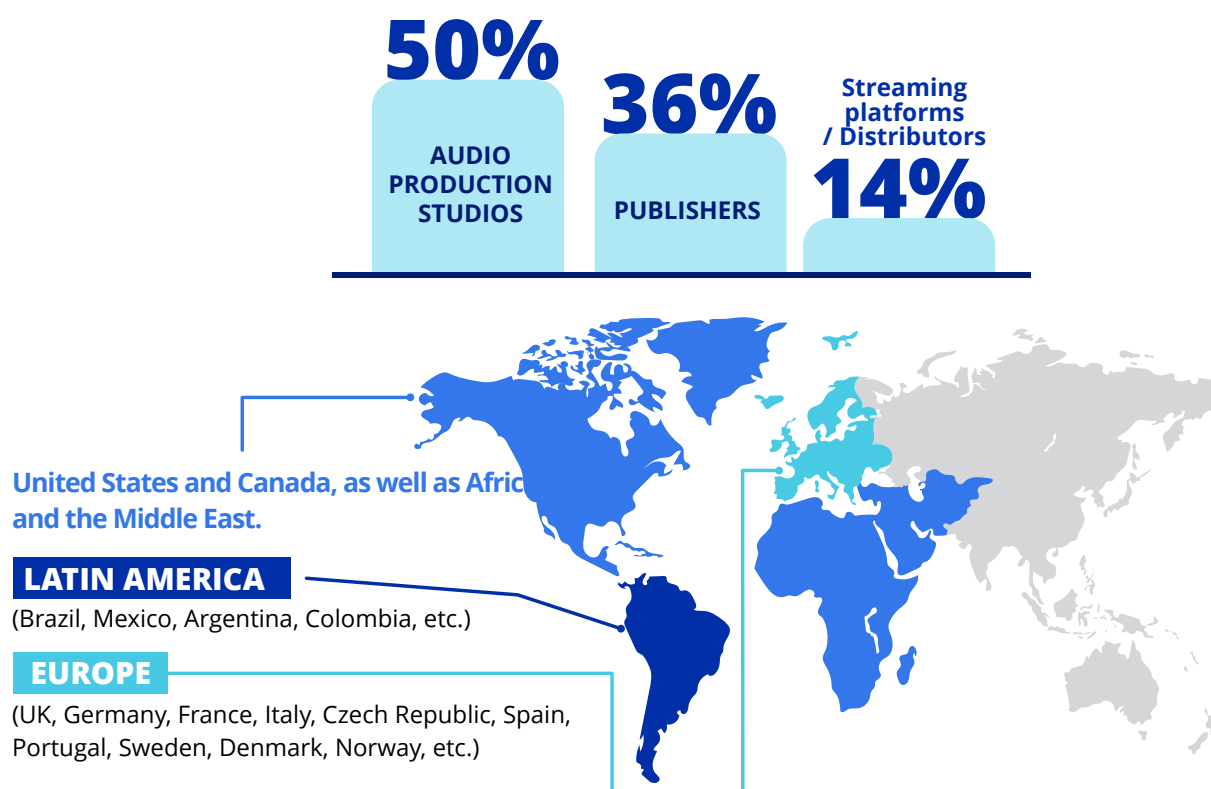
SECTION 1

AI and Audio Production

This section aggregates and analyzes questionnaire responses from 53 publishers, production studios, and distributors, contextualized by insights from 32 additional audio professionals from all over the world.

Respondents include major trade and independent publishing houses, audiobook distribution and streaming platforms, providers of AI narration and production, independent producers, and recognized industry experts. The diverse representation of those involved in audiobook production provides a guide for the adoption, strategic decisions, challenges and expectations surrounding integrating artificial intelligence into audio production workflows.

PARTICIPANT PROFILE



As illustrated in the first graph of this report, **50% of the contributions came from audio production studios across the globe, 36% from publishers, and streaming platforms or distributors represented the remaining 14%.** This composition reflects operational realities. Many publishers redirected us to their production studios, to which they outsource their production workflows.

Participants included organizations from Europe (UK, Germany, France, Italy, Czech Republic, Spain, Portugal, Sweden, Denmark, Norway, etc.), Latin America (Brazil, Mexico, Argentina, Colombia, etc.), the United States and Canada, as well as Africa and the Middle East.

The results of the research point to broad adoption with **85% of the respondents stating they use AI tools** in at least one phase of the production process, **while only 15% indicated that they do not use, so far, any AI tools** at all in their audio production workflow.

1.1 Consensus: Coexistence, Not Replacement

The dominant response rejects choosing between AI and human narration in favor of a hybrid production model.

Human narration is reserved above all for flagship titles of the greatest editorial or commercial value: major new releases, anticipated bestsellers, culturally driven titles, key series, or projects where interpretation is an essential part of the experience. In these cases, the human performance is key as it provides interpretation, prestige, identity, and emotional connection with the audience.

Consistent with this line of thinking, **Tressa Hopkinson, Editor-in-Chief, The Working Narrator (USA)**⁴ indicated that despite its name, today's artificial intelligence is not intelligent in the way people are. "It recognizes patterns, makes predictions, and generates responses based

on its training, inputs, and instructions. Current AI systems do not bring lived experience, cultural understanding, ethical responsibility, or human emotional interpretation to the decisions they help us make. That is why I do not believe AI's greatest business opportunity is replacing people. Its real value over the next three years will be in helping people do more—reaching new audiences, opening new markets, reducing repetitive work, and making audio available in situations where traditional production has not been financially or practically possible.”



Many of the respondents linked AI narration to catalogue expansion, precisely because these are the titles that traditional human narration economics have never been able to justify. AI translation and multilingual production tools will create significant opportunities for the publishing sector, accelerating first-pass translation, localization, pronunciation research, and parts of the quality-review process. That could allow publishers to bring more titles into more languages, more quickly, and at a more sustainable cost. “However, language is about more than translating words correctly. Meaning, humor, emotion, and cultural context still require knowledgeable human oversight. The strongest model will pair the speed of AI with the judgment of translators, directors, performers, and others who understand both the language and its audience,” added Hopkinson.

Many of the respondents also indicated business opportunities for backlist titles, low-turnover works, niche titles, non-fiction, educational or academic content and current-affairs titles. As we will detail in Section 2 of this report, AI will make production workflows faster and more efficient, assisting production teams in processes such as manuscript preparation, pronunciation research, editing, mastering, metadata, and quality control. These tools can give narrators and production teams more time to focus on the work that needs them most: performance, interpretation, and creative decision-making.

Another of the clearest patterns identified is the distinction that participants draw between fiction and non-fiction. Fiction tends to be reserved for human narrators because it demands emotion, nuance, characters, rhythm, and interpretation. Non-fiction, especially when it is more informative, technical, academic, or current-affairs oriented, is seen as more suitable ground for testing AI. Even so, nuances appear: some short works of fiction or those with few characters could potentially accommodate more automated models.

The importance of the source material also stands out in the responses. As we will describe in Section 2 of this report, final quality depends heavily on how the manuscript arrives: file format, text cleanup, pronunciation guides, proper names, acronyms, numbers, and content specifics are key in the AI production workflow.

It's important to note that AI does not work for every book. It is more successful with linear, technical, academic content or content with less interpretive demands. With fiction, classics, narrative non-fiction, and culturally sensitive content, emotion, context, accent, direction, and human performance are important to listeners.

The author's voice is an exception. Cloned voices or authorized replicas when the author cannot narrate or is deceased, or when the goal is to carry an author's voice into other languages found more acceptance.

This use opens opportunities, but requires consent, authorization, human supervision, and transparency for the listener.

One third of the respondents indicate that they are currently redesigning their internal production frameworks with the objective to enhance their audio catalogue with an AI production strategy: full human narration reserved for premium and flagship titles; a hybrid tier in which AI handles mechanical stages such as text preparation, quality control, and manages the selected synthetic voice while a human narrator and editor retain performance control; and full synthetic narration reserved for backlist or simpler titles that would not otherwise be produced at all.

In line with the objective to redesign internal production frameworks, **Laure Saget, CEO of AudioLib Hachette (France)** indicated that “we have refined our human review workflows to better integrate AI-assisted processes while maintaining editorial and production quality standards. Human oversight remains essential at every stage of audiobook production. While technology continues to evolve, preserving the integrity of the relationship between publishers, authors, performers, and audiences remains paramount. The long-term success of the audio sector hinges on a strategic equilibrium where operational efficiency meets high-quality standards and absolute transparency”.



High-intensity editing and post-production for premium audio, particularly audio drama involving music and sound design, is also flagged as an area where current tools still cannot meet the same quality and therefore require the human touch even while, as a commodity, AI music and sound-effect tools proliferate.

On the other hand, 15% of respondents openly advocate, at present, for a human-only production model. These organizations have decided, so far, to keep their entire production in human voices or reject AI tools for reasons of quality issues, brand positioning, or editorial conviction. Based on the conversations we had during the research phase of this report, it appears sometimes this reflects a human-first stance or a personal lack of interest. In other cases, it stems from specific tech limitations and barriers: such as voice quality in their local languages, client mandate not to use AI, language availability, contractual limitations, or internal processes not yet ready. This provides an interesting counterpoint: not all publishers see AI as a necessary solution for their audio catalogue growth, even if they acknowledge that the technology may evolve.

These results reinforce a central idea of the report: audio production strategy will not be “one size fits all”. As described in Phase 6 of Section 2, the sector appears to be moving toward a **three-tier catalogue segmentation strategy** in which human narration will remain associated with titles of greater performance, emotional, or commercial value; AI will be used to expand coverage and make viable works that previously fell outside audio; and hybrid or authorized-voice models will occupy the middle ground, provided there is sufficient quality, clear rights, and editorial oversight.

1.2 The Economic Case: Unlocking the Long Tail

The most frequently repeated argument is economic: AI tools reduce the cost and time required to produce an audiobook, making backlist titles, academic monographs, regional and niche works, and non-English translations viable for the first time.

Notably, respondents indicated that the cost savings were generated in the range of 20% in the second year and up to 50% in the third year, not the 85-95% figures promoted by technology providers in the industry. Comparing complete workflows, not isolated cost lines, is essential for realistic budgeting and determining actual annual returns on investment.

Elena Bazán, Head of Audio Production, Audiolibre (Mexico), warns publishers that catalogue expansion can't be an end in itself. "Publishers need a clear strategy for how their audiobook catalogue grows — and if AI narration is adopted purely as a way to lower costs or scale volume, it risks undermining the industry's credibility rather than strengthening it. First-time audiobook listeners are especially fragile: their early experience shapes whether they stick with the format at all. Lose them at that stage, and the whole industry loses a listener, not just a single publisher. That's why the retention question matters collectively, not just competitively. And retention itself depends on quality and emotional nuance — the human sensitivity a narrator brings — which, for now, remains a distinctly human capability."



At the more advanced technical vision, that we will analyze in the next section of this report, several production studios and streaming platforms indicated that they aim to create fully AI-supervised workflows in which an audiobook is produced from the same pipeline as the ebook, at marginal cost per title and in a few hours, compared with weeks or months for a human based studio production. Because such workflows run through APIs, producing an entire catalogue becomes only marginally harder than producing a single title. As we will see later in the report, this is repeatedly framed not as cost-cutting for its own sake, but as an industry growth strategy: extending audio access to markets, languages, and audiences that legacy production economics excluded.

1.3 Production Workflows and the Supply Chain

More than 90% of the respondents agree that AI's near-term comparative advantage lies less in creative performance and more in operational production stages, such as script and text preparation, quality control, workflow automation, and metadata generation, as described in Phases 2 to 6 of Section 2.

As indicated at the beginning of this report, the results of the research point to broad adoption among the audio publishing industry **with 85% of the respondents stating they use AI in at least one phase of the production process, while only 15% indicated that they do not use any AI tools at all in their audio production workflow process.**

In our view, the relevant figure is not just how many organizations currently use AI, but how they use it. AI tools are not limited to isolated pilot tests, but rather organizations incorporate it into different phases of their production workflows.

As illustrated in the second graph of this report, **38% of respondents have integrated AI tools in 3 or 4 phases of their production processes**, indicating a growing operational adoption of AI tools in the sector, but only **17% of respondents report even broader use**, integrating AI tools across more than 6 phases of production.

NUMBER OF WORKFLOW PHASES WITH INTEGRATED AI



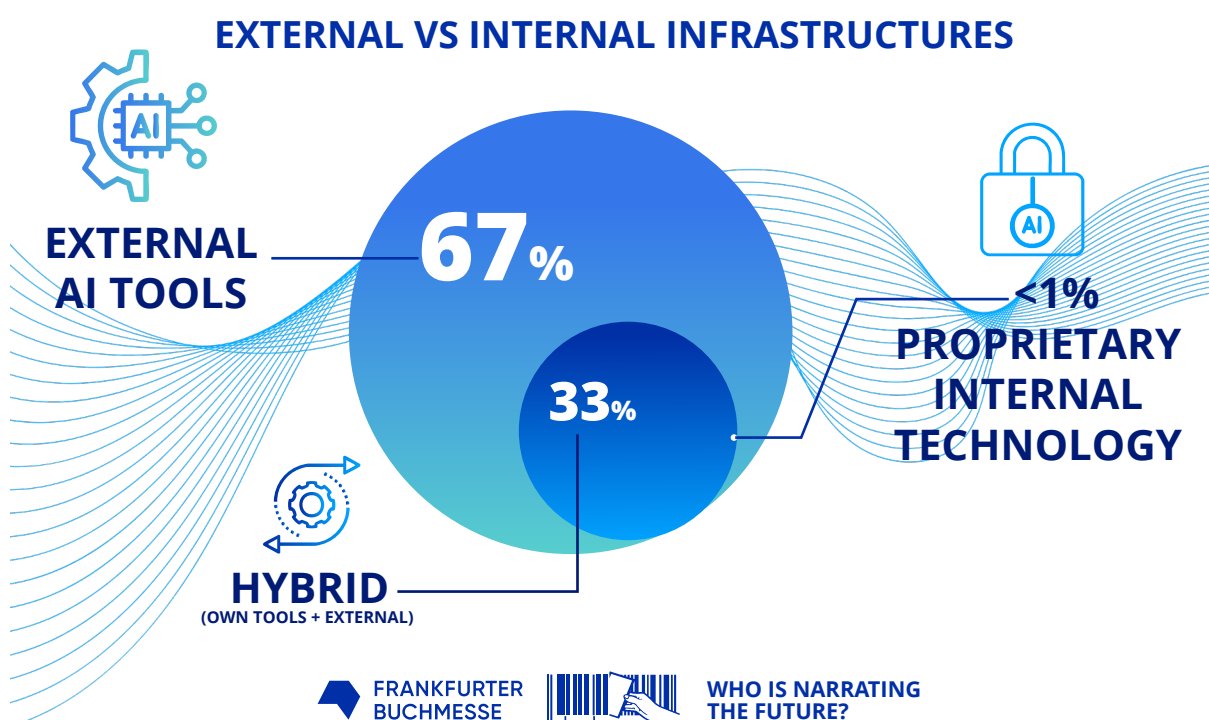
As we will describe in detail in Section 2, the questionnaire asked participants to detail the usage of AI tools in the production workflow to understand where it is being applied: text preparation and adaptation, editorial preproduction analysis, synthetic voices or voice cloning/replicas, interpretive direction via speech-to-speech, translation or localization of catalogues, automated quality control, and technical postproduction, among other processes.

The responses show that AI is particularly strong when used for text preparation and adaptation (such as adding suggestions for sound effects and music), editorial preproduction analysis, management of voice synthesis, and automated quality control such as postproduction comparison of text versus recording. This confirms that AI's role is not

limited to generating synthetic voices: it is also being used as an editorial, technical, and verification support tool.

Technical postproduction likewise appears as a relevant area. Voice cloning is present, although less adopted by respondents. By contrast, AI interpretive direction and catalogue translation or text localization are not widely adopted, most likely because they require more demanding technical, creative, linguistic, or legal conditions.

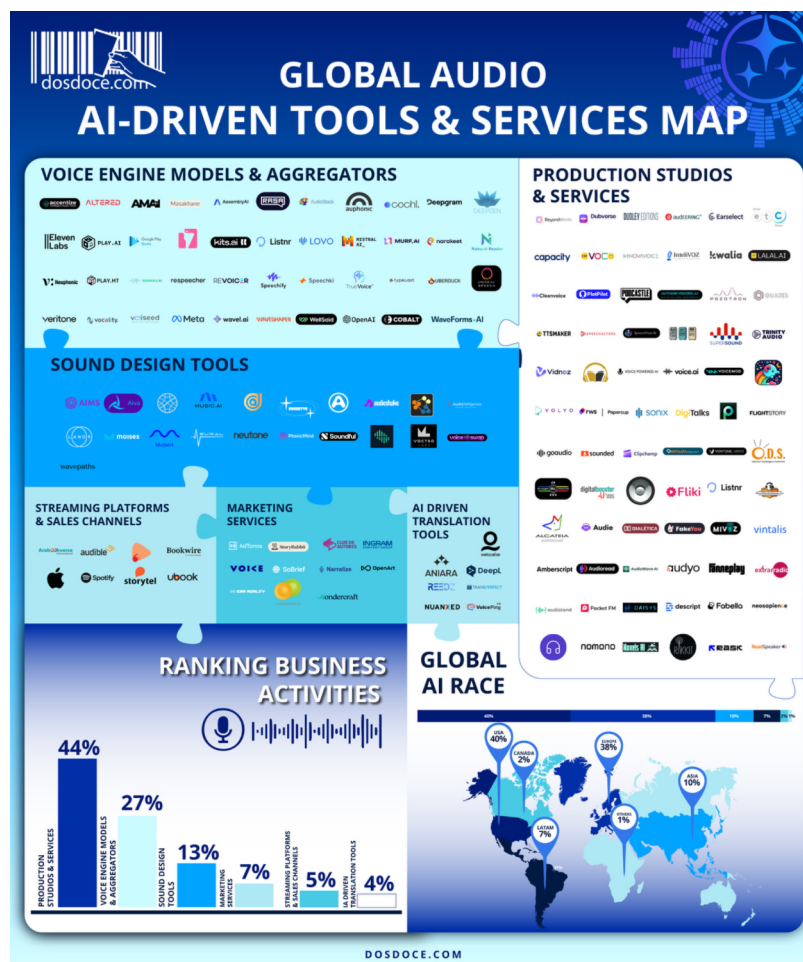
As illustrated in the third graph of this report, **more than 2/3 of the respondents work with external AI tools and 1/3 combine their own tools with external solutions. Less than 1% report using exclusively proprietary internal technology.**



The combination of internal and external tools is found when organizations (mainly streaming platforms and big publishing houses) seek greater control over the process. In these cases, third-party solutions are integrated within a proprietary architecture: API integration, in-house quality-control systems, voice management, end-to-end workflows, or adaptation to specific language needs.

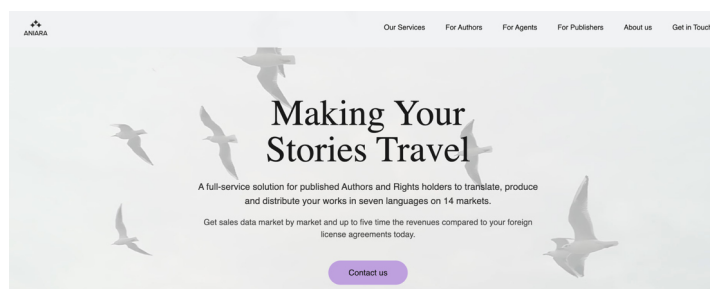
Last year, Dosdoce conducted an extensive research study identifying more than 160 Audio AI-driven entities from the creation of synthetic voice engines to simultaneous translations of audiobooks and podcasts into multiple languages, as well as technological developments aimed at marketing and advertising audio content, to applications for automating all kinds of processes in the pre- and post-production of audio content with the support of artificial intelligence.

One of the key takeaways from last year's report is that AI tools are empowering audio production professionals by freeing them from repetitive and tedious tasks in production processes, allowing them to channel their time into more creative and complex aspects of their work.



The 160+ AI solutions and services identified in the Global Audio AI-Driven Tools & Services Map⁵ have become an extension of the work of audio professionals improving their productivity and creativity, not to replace it. Among all these tools, **ElevenLabs** stands out this year as one of the most

cited solutions, especially linked to AI voice generation, voice cloning or replicas or managing production processes. **Pozotron**, also highly cited, is associated with quality control, text-audio alignment, pronunciation, and error detection. Generative text AI tools such as **Claude** and **ChatGPT** are used for preparation, analysis, adaptation, pronunciation guides, casting, text review, or editorial support.



In addition to internal tools, other productivity, translation, cloud, postproduction, or technical audio solutions are mentioned—including **Aniara.One, Copilot, Gemini, Google Vertex AI, Azure, DeepL, Adobe Firefly, iZotope RX, Whisper, Murf, Suno AI, Nuanxed**—but less often.

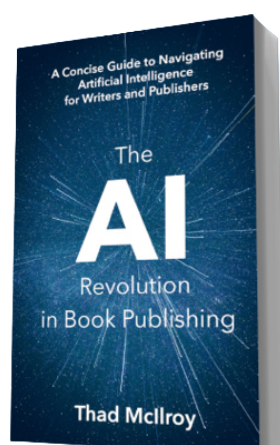
The current usage penetration of AI tools suggests an ecosystem that is neither standardized nor consolidated, while a few tools receive heavy use, there is a long tail of solutions applied to specific tasks. While AI tools are becoming normalized as a support layer across different phases of audio production, the dominant pattern is a gradual, selective project-specific integration, considering the required quality standard and each organization's editorial, technical, or contractual constraints.

As we will see later on in the report, several respondents forecast that over the next few years AI tools won't be confined to a single corner of audiobook production and distribution. **Chris Kenneally, host of the podcast The Spoken World** foresees that “they will be found useful in ‘all of the above’ (voices, translation, production workflows, marketing) as well as in catalogue development, metadata creation, and distribution.”



In the next few years, several of the more technologically advanced respondents envision a future where audiobooks cease to be fixed, pre-produced assets and instead become on-demand renderings, generated at the moment of listening in whatever voice and language the listener selects, in the same way listeners might choose a font or letter size for an ebook today. The binding constraint for this potential listening evolution will be licensing negotiations with title IP owners rather than technology. Rights negotiations and disputes, rather than raw technical capability, have shaped the pace of adoption before and are expected to do so again⁶.

Taking into consideration these upcoming technological advances, **Thad McIlroy, Principal of The Future of Publishing and author of *The AI Revolution in Book Publishing* (USA)**, reflects on the need to redefine the audiobook marketing and distribution processes to guarantee author's visibility and discoverability: "Part of the strength of the existing rights marketplace, perhaps its main strength, is that when a publisher commits to purchasing rights and then commissioning a translation, and then, in many/ most cases, an accompanying audiobook, that investment also ensures that the publishers will invest marketing efforts and dollars in promoting the work in their home market. If authors and publishers pay small amounts to commission translations into text and audiobook formats, how then are they going to market those formats? Sure, they can upload the ebook to Amazon, and the audiobook to Audible and equivalent, but that doesn't ensure that anyone will be aware of its presence. That is going to demand a whole rethinking around audiobook distribution and marketing, one that has yet to be addressed."

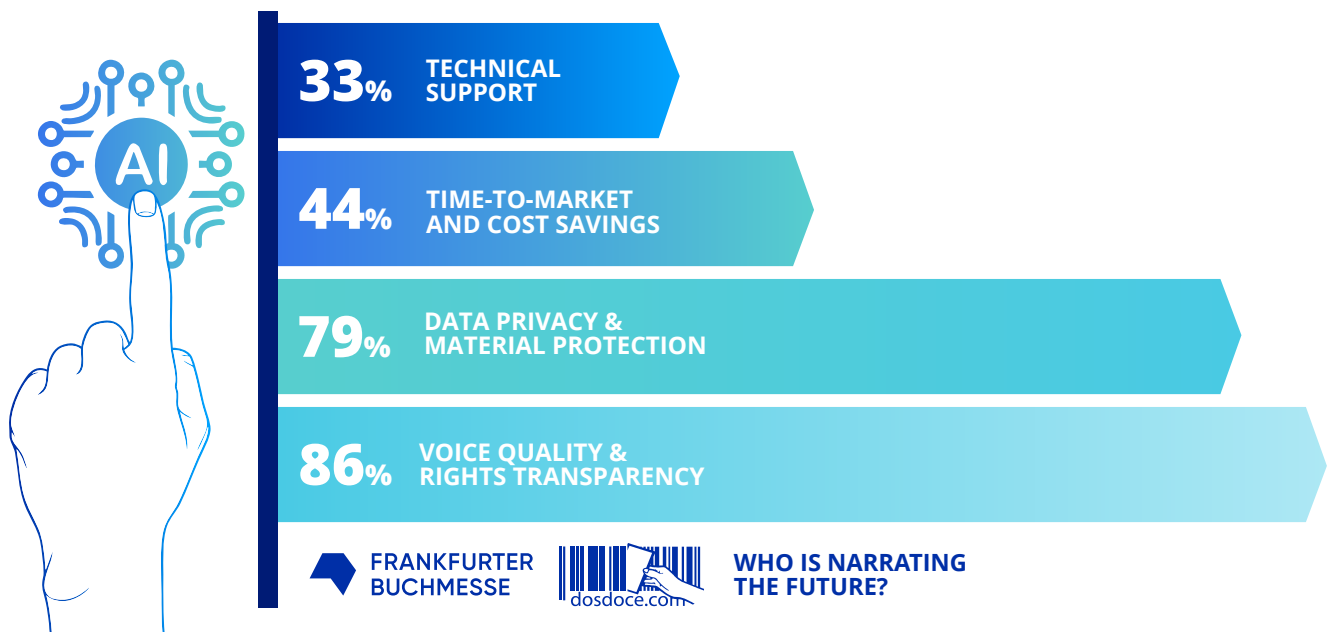


1.4 Criteria for selecting AI tools and technology providers

When evaluating AI tools and technology providers, the responses show organizations look beyond cost and efficiency factors.

As illustrated in the fourth graph of this report, **86% of participants mention voice quality as well as transparency regarding rights as important considerations. 79% of respondents cite data privacy as a factor in the protection of manuscripts, audio, and sensitive materials.** These results indicate that AI adoption in audio is heavily conditioned by trust.

CRITERIA WHEN CHOOSING AI TOOLS OR PROVIDERS



In parallel, 44% of respondents refer to the importance of time-to-market and cost savings, followed by technical support at 33%. This does not mean these factors are irrelevant, but that they clearly carry less weight.

Publishers seem to show sensitivity toward privacy and cost, consistent with their role as those responsible for catalogue, rights, and budgets. Production companies keep voice quality and security as priority criteria.

Regardless of role, technology selection is seen as an editorial, legal, and reputational decision, not merely an economic one.

■ 1.5 Barriers to AI use

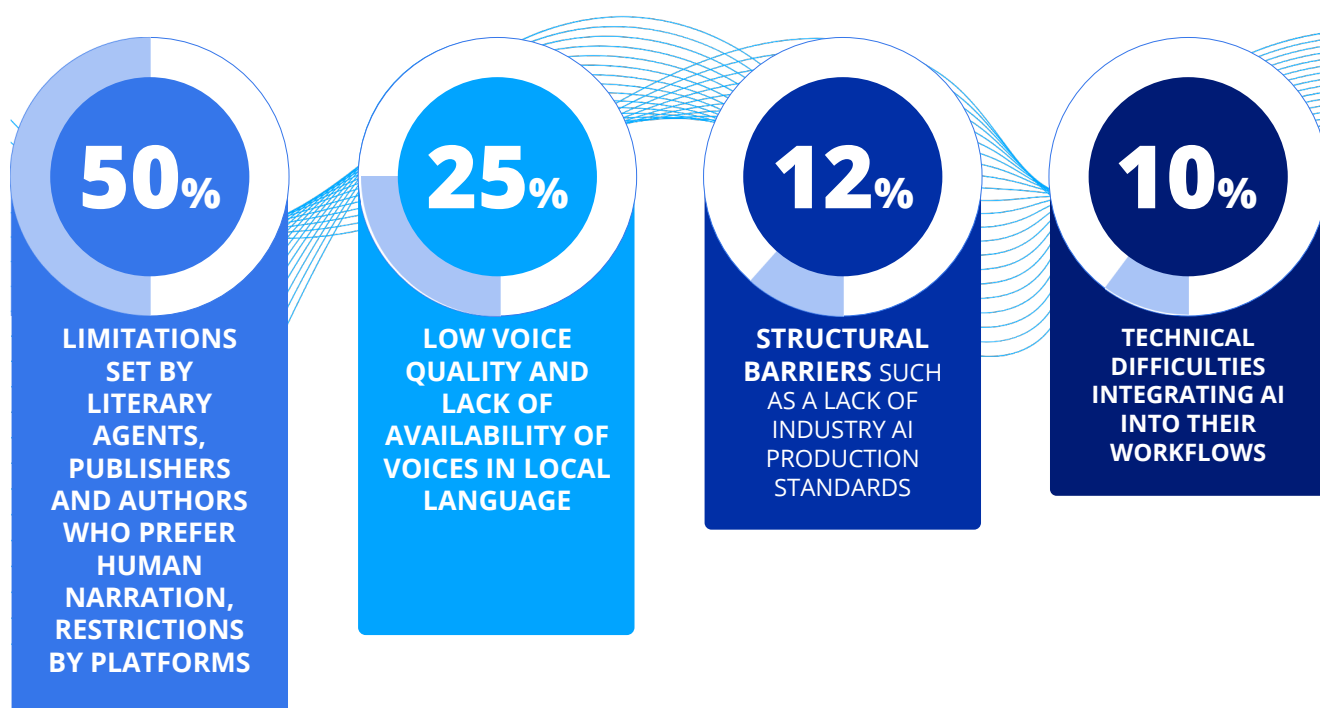
With regard to the barriers to scaling AI use in audio production processes, the responses clearly indicate that the main obstacle is not solely technological. Challenges to the adoption of AI include legal, commercial, and market-acceptance factors.

50% of respondents identified the principal barrier as being limitations by literary agents who do not authorize the use of AI, publishers and authors who prefer human narration, or distribution restrictions by platforms on specific AI tools. Thus, even when the parties involved in the production processes are able and willing to use AI tools, its integration also depends on external restrictions.

25% of the respondents indicated other reasons for not using AI tools, such as low voice quality and naturalness, manuscript protection, lack of availability of suitable voices in their local language or accent, or the low profits of AI productions. Additionally, 12% of respondents cited structural barriers such as a lack of industry standards and approximately 10% cited technical difficulties in integrating AI into their current workflows as barriers.

This range of reasons helps us better understand the lack of AI adoption in some organizations. In some cases, the barrier does not appear to be a lack of interest, but rather external conditions or insufficient guarantees to adopt it safely. Thus, it follows that AI will only be able to scale sustainably if the sector resolves issues of trust, authorization, rights, commercial acceptance, and integration into production processes.

BARRIERS IDENTIFIED TO SCALING AI USE



FRANKFURTER
BUCHMESSE



WHO IS NARRATING
THE FUTURE?

Another challenge is the existing tension between synthetic voices quality and consumer attitudes toward it. As illustrated in the fifth graph of this report, **90% of the respondents** recognize the need for greater transparency and disclosure toward both authors and consumers about where and how AI is used. Meanwhile, **60% of respondents also indicated** the need for consent-based voice cloning, ethically licensed voice datasets and fair compensation for voice talent in order to minimize legal risks and guarantee sustainability.



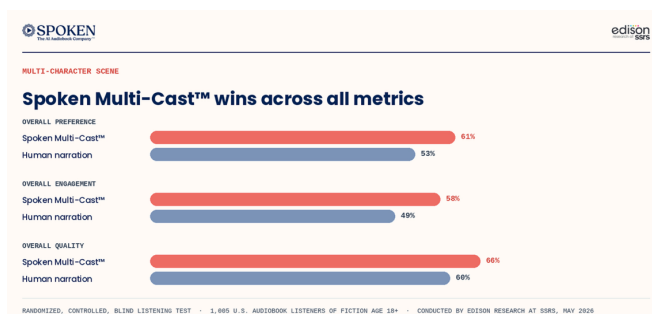
FRANKFURTER
BUCHMESSE



WHO IS NARRATING THE FUTURE?
AI adoption, use and outlook
in audio production

Market data reinforces this gap between AI production capability and adoption. In the United States, **AI-narrated titles are reported to have earned only around 0.03 percent of total audiobook revenue in 2025⁷**. The consumption and number of AI-narrated audiobooks published have increased in 2025, though willingness to try AI-narrated audiobooks remains small: only 16% of audiobook listeners have listened to an AI-voiced audiobook.

Respondents interpret this data as evidence that production capacity is currently outpacing consumer demand, and that stated sentiment (skepticism toward AI) diverges from revealed behavior such as completion rates and blind-test preference. Notwithstanding, a recent study conducted by Edison Research for Spoken⁸ indicated higher ratings for AI voices than human narration in overall favorability (61% vs. 53%), perceived quality of narration (66% vs. 60%), and overall engagement (58% vs. 49%) for character-driven stories. For exposition without multiple characters, the human narrator was rated higher in the Spoken study. Corresponding to this data, **Philip Marshall, Founder of Spoken (USA)** foresees that over the next three years, “multi-cast narration will become the default for character-driven stories, particularly fiction, rather than a luxury reserved for top titles. The implications are enormous, including monetizing backlist titles, expanding into new distribution channels, and even personalizing audio to listener taste.”



In this context, the upcoming emergence of new tech services offering new AI audio consumption models and pricing strategies is widely expected to segment the market into a lower-cost AI-narrated tier and a premium human-narrated tier, at least until quality between the two becomes indistinguishable to the average listener — a point several respondents flag as a possible tipping point for change.

■ 1.6 Global and Multilingual Expansion

Beyond time-to-market and cost reductions, translation and localization are repeatedly cited as AI's largest business opportunity: enabling simultaneous multi-language production and reaching audiences previously excluded due to production economics.

Emerging audio markets such as Latin America, the Middle East and Africa are named specifically as regions where AI-driven cost reduction could meaningfully diversify accent and language coverage and expand catalogues that remain heavily constrained.

Aligned with this global growth vision, **Rachel Ghiazza, Audible's Chief Content Officer (USA)**, indicates that we are still at the very beginning of what the audiobooks category can be, especially in newer marketplaces. "Millions upon millions of print and eBooks are not yet available in audio. We believe the best way to grow the category is to meet listeners where they are by bringing more titles to life in local languages, both through professional narration and AI production. Ultimately, AI production tools will help us close the content gap so that more listeners, across more languages and genres, can find a story that speaks to them. For Audible, this is an "and" approach. We're expanding our audiobook catalogue through AI, while also continuing to invest more than ever in professionally narrated titles. AI enables the production of titles that would otherwise never make it to audio, especially for backlist and non-English language titles. At the same time, we're seeing professional narration continuing to thrive for premium, high-investment productions including multi-cast editions with immersive sound design and full orchestral scores."

With a similar global perspective, **Victor Meinberg, COO, Skeelo (Brazil)**, indicates that "the biggest opportunity AI brings to audio over the next three years is reach. AI workflows and synthetic voices make it economically viable to produce audiobooks for backlist, local authors, and niche titles that never justified traditional studio costs, and to localize catalogues across languages at a fraction of the price. From a distribution

standpoint across Latin America and Europe, this is how we finally bring audio to readers and markets that legacy production economics left out. The players who win will be the ones who pair that scale with quality and transparency, so AI grows the category instead of diluting trust in it.”



Following this logic, but with a higher degree of emphasis, **Thad McIlroy, author of *The AI Revolution in Book Publishing***, advocates that proponents of this technology (such as himself) point to the huge opportunities in reaching listeners who could not otherwise be reached as they wait for publishers to mount \$10,000+ human-narrated audiobooks. “We point out come on! There's no way that spending \$10,000 on the Lithuanian audiobook translation of a minor English literary novel could ever make sense. It hardly makes sense for the English version. Are we supposed to sit silently on the sidelines murmuring: ‘There must never be an audiobook created without a human narrator,’ knowing that this will deprive some 80% of the world from hearing the audiobook in their own language?” McIlroy continues indicating that this debate connects inextricably with the surrounding issue of AI translation for text only, apart from audio: “Here too many purists would rather that people either learn English or simply suffer that they will not have access to the book in their native language. And, of course, we English speakers are grotesquely Anglo-centric. What about books that could be translated from Lithuanian to Lao? Do Laotians not deserve to read and hear Lithuanian literature because AI is involved in its creation? Contrary to rumor, we AI enthusiasts have

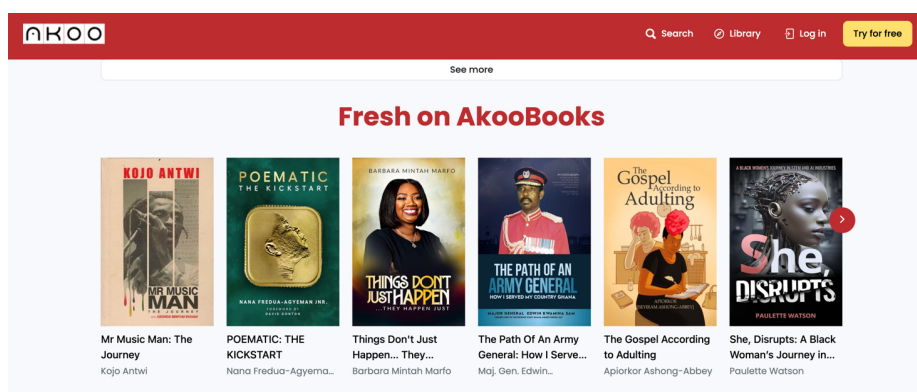
zero interest in depriving translators and audiobook narrators and actors of their well-earned occupations. It's about new markets, not about destroying existing ones."

On a more personal note, **Nadim Sadek, Author, Founder & CEO, Shimmr AI (UK)**, reflects on his own experiences with the usage of AI tools. "When thinking about the opportunities for the audiobook industry, in a time of increasingly pervasive and competent AI resources, I turn to my own experiences with my books. My two books - on how Publishing, and Creativity, can Embrace AI - have been published in over a dozen languages, largely because the budget and time needed to do so was so shrunken by using AI. They have both been turned into audiobooks, using my own synthesized voice. This gives each of the publishers the confidence to now produce an audiobook too. Curiously, my Greek publisher can't use the generated audiobook because 'my Greek accent is too thick'! So, what I expect to see happening is every book becoming multi-modal - certainly translated, usually produced as an audiobook, and then extending into other forms, such as explanatory podcasts, using the author's voice. The distillation of my perspective is that every written book will find itself expressed in multi-modal forms, so that people can engage with it in various ways. As we are primarily an audio-visual species, audiobooks will be a key plank in this new platform of accessibility, and discoverability."

This global optimistic outlook also comes with caution: without deliberate investment in diverse and ethically sourced voice datasets, multilingual models, transparent rights frameworks, and fair compensation for local voice talent, AI-driven expansion risks reinforcing existing global inequalities in representation rather than correcting them.

In line with these cautionary notes, **Ama Dadson, Founder & CEO, AkooBooks Audio (Ghana) and Frankfurt Buchmesse Audio Ambassador**, indicates that AI has enormous potential to democratize audiobook production, particularly for underserved languages and

regions, such as those in Africa, but this business opportunity also requires investment in diverse, ethically licensed African voice datasets, multilingual models, transparent rights frameworks and fair compensation for voice talent to avoid the mentioned fears: “Without intentional representation of African voices and languages, the next generation of AI audio risks reinforcing existing global inequalities rather than expanding access. For African publishing, AI's greatest opportunity may not be replacing professional narrators, but making it economically viable to produce educational content, local-language titles and backlist works that would otherwise never become audiobooks. The emergence of African AI voice companies is an encouraging development, and we hope to see continued progress in the quality, diversity, and commercial readiness of African voices for long-form narration.”



Crossing the Atlantic from Africa to Brazil, we find similar worries.

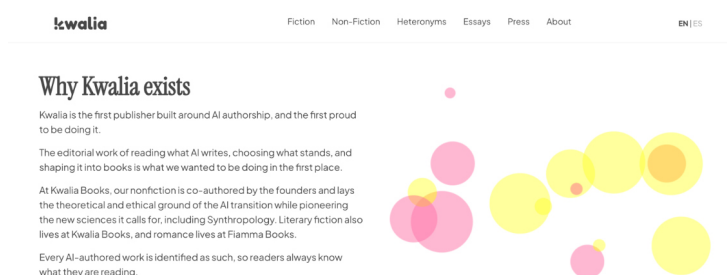
Camila Leme (Brazil), one of the most experienced audio industry professionals from Latin America,

states that “AI presents a transformative opportunity to optimize workflows and reduce production costs, which remain a major barrier for many publishers today in emerging audio markets like Latin America. It is also a powerful tool to diversify accents and scale up catalogues that are still heavily restricted in both Portuguese and Spanish. However, we must ensure human oversight at both ends of the process, backed by absolute transparency in how AI is utilized.”

1.7 Quality, Differentiation, and Discoverability

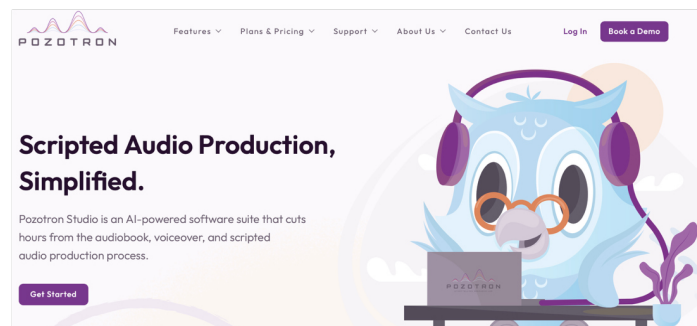
As the volume of AI-generated audio content increases, respondents expect quality, rather than price alone, to become the primary differentiator, rewarding publishers and streaming platforms who pair scale with high production standards and transparency rather than treating AI purely as a way to save money.

Supporting this claim, **Javier del Puerto, Founder of Kwalia (UK)**, envisions that “once narration, translation and quality control run as workflow steps rather than budget decisions, three markets open at once: the backlist, which was rarely economical to record; languages that never justified a studio production or a translation budget; and the small publisher, who can now run a full audio programme.” Javier reaffirms that “at Kwalia our audiobooks come out of the same AI-supervised workflow as our ebooks, at marginal cost per title and in about two hours where a studio production waits weeks or months. Because that workflow runs through APIs, a catalogue is barely harder to produce than a title. The publishers best positioned over the next three years are those who industrialize production now and keep their editorial judgment for what merits production and for which market.”

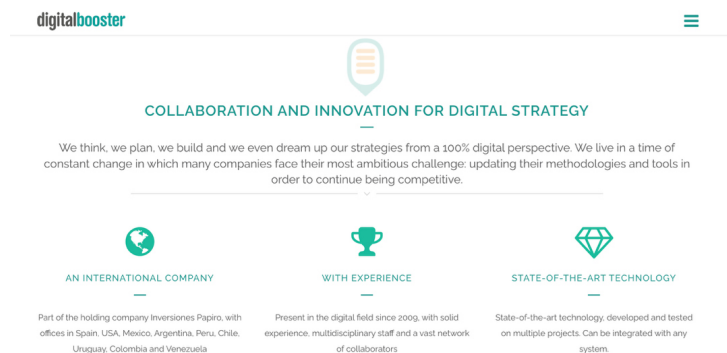


Corresponding to this view, **Adam Fritz, CEO Pozotron (UK)**, states that “one of the biggest opportunities is multilingual production, enabling publishers and producers to create high-quality versions of the same

title in multiple languages simultaneously. This has the potential to significantly expand global reach while making international audio production faster, more scalable, and more cost-effective.” As audio production technologies are further developed and implemented thoughtfully, Fritz indicated that AI tools will make audio content creation more efficient without replacing the creative expertise behind great productions. “As producers are increasingly expected to create more content at lower cost and on tighter timelines, AI can streamline repetitive production tasks, reducing both time and resource requirements,” stated Fritz.



Technology will continue to improve, achieving quality levels close to perfection and cost optimization. Following this logic, **Ricard Fidel, Manager e-Services & B2B Team Digital Booster - Urano World (Spain)**, states that “we are close to meeting those goals: quality levels close to perfection and cost optimization. As quality improves, prices become reasonable, and these technologies are increasingly adopted by the various market players, production and market share will grow. Key to this are quality, volume, and effective distribution and marketing.”



Bibiana Riccardi, Founder of Sonda Media (Argentina), sees a future of the audiobook industry that will be driven by a large number of new AI-generated titles, but at the same time, there will be room for titles aimed at specific segments of listeners who are willing to pay more for content that feels more human and less artificial. “I see the future not as a complete replacement of human narration, but as a coexistence of different models: AI narration for scale, accessibility, and catalogue expansion, and human narration for projects where interpretation, emotional nuance, and a stronger sense of authorship remain essential,” indicated Riccardi.

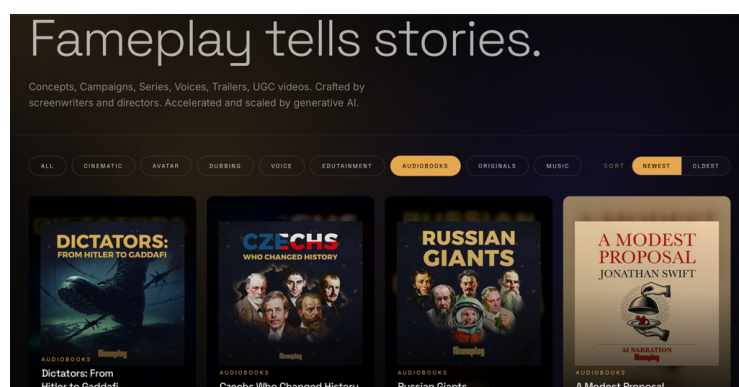
As highlighted in the introduction of the report, many respondents indicated business opportunities for backlist titles, low-turnover works, niche titles, non-fiction, educational or academic content and current-affairs titles. In this context, **Danielle D’Orlando, Curator of Audio, Princeton University Press (USA)**, foresees that “as the academic library market evolves to meet consumer behavior, we can see a space for more cost-effective recording options using AI-voice specifically for these niche productions, but we do still foresee and emphasize the importance of human narration at the forefront of our business.”

Maite Goñi Eizmendi, Director Radio & Audio, EITB (Basque Country), advocates for the coexistence of a mix of production models (human, hybrid and AI-based) rather than a complete replacement of human narration. “AI will be very useful for specific formats, pilots, accessibility, personalization, archive exploitation, metadata generation and lower-volume or highly repetitive workflows. However, human voices, actors and presenters will remain essential for editorially sensitive content, brand value, emotion, talent recognition and formats where trust and authenticity are central. In practice, the most realistic 2026-2030 scenario is a hybrid production model: AI-assisted workflows with human oversight, clear governance, consent-based voice cloning and transparent use towards audiences.”

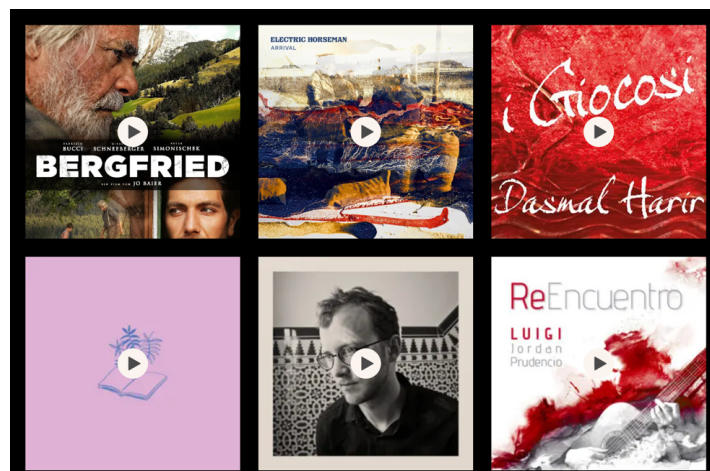
Bruno Sambado, CEO, Tale House (Portugal), also advocates for the coexistence of a mix of production models, especially in new audio markets such as Portuguese-language one: "A human voice can break your heart in a way nothing else can. AI can't replace that, but it can make sure no book is ever left without a voice at all. At Tale House, in Lisbon, we let each do what it does best, so that every story in our language finally gets heard."



Juraj Felix, CMI Group Digital Development Director & CEO at Fameplay (Czech Republic), also believes the future is a mix of models rather than AI completely replacing human narration. "AI will make it possible to produce many more audiobooks that wouldn't be commercially viable otherwise - especially niche titles, educational content, backlists, or multilingual versions. At the same time, I don't see human narrators disappearing. For fiction, emotionally rich stories, and premium productions, the performance is part of the value, and listeners will continue to appreciate that. I think AI will expand the market rather than replace human talent, with each approach serving different types of content and audiences."



Somewhat counterintuitively, as audio production becomes cheaper and more abundant in the retail channels, discoverability will become even more difficult. In this scenario, distinctive human performances and recognizable narrators will become strong marketing assets capable of cutting through a large volume of AI catalogues. In line with this vision, **Chris Kling, Managing Director of Klangkantine Studios (Germany)**, warns publishers that if almost any text can be heard instantly through freemium apps, a pre-generated synthetic audiobook still needs to give listeners a compelling reason to pay.



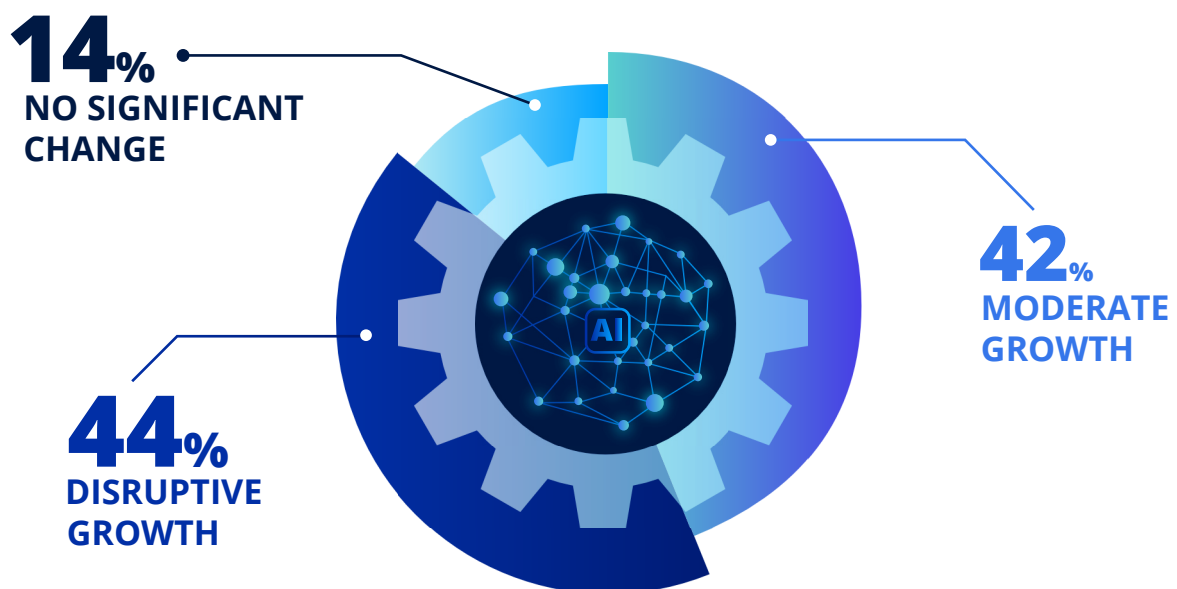
“As audio becomes more abundant, discoverability becomes harder. Distinctive performers, authentic voices and recognizable casting can turn narration into both a creative asset and a marketing signal. The winners will use AI to scale around human creativity and automate what is interchangeable, without making the final product interchangeable,” added Kling.

A related warning concerns listener loyalty: long-term audience retention can be tied specifically to high production quality and human sensitivity; several respondents caution that publishers who approach AI production purely as a volume or cost-saving strategy risk losing first-time listeners and, with them, an opportunity that benefits the format.

1.8 Outlook: 2026–2030 and Beyond

We can conclude from the responses that contributors do not expect AI to displace premium human narration in the near term, **but 86% do expect disruptive or moderate growth in the industry, attributable to the use of AI tools.** AI-produced titles that would not otherwise have been made into audio will sit alongside a premium tier of human-narrated flagship fiction and emotionally significant works.

SECTOR GROWTH EXPECTATIONS FROM AI USE



FRANKFURTER
BUCHMESSE



WHO IS NARRATING
THE FUTURE?

As illustrated in the sixth graph of this report, the dual production path (human and AI) is anticipated to increase the number of titles available in audio over the next three years: **44% of respondents point to disruptive growth and another 42% moderate growth, with only 14% of the responses expecting no significant change.**

In line with this growth perspective, **Jim Dinegar, Executive Director, Audio Publishers Association (USA)** indicates that "while production standards tied to the human voice continue to engage and bring in

more consumers to the audiobook format, AI tools may make more titles available which will continue to provide more selection for the growing format."

Further analyzing this dual growth perspective, **Videl Bar-Kar, Vice President Audio Bookwire (Germany)**, envisions a spectrum of audiobook consumption—and consumers— which will be vast and varied: "The content we create and distribute should mirror that. On the one side, immersive spatial audio technology is expanding traditional listening experiences with incredible human performances - on the other, AI provides an opportunity for high volume consumption where performance isn't the central factor, rather clarity and consistency are. The future of the audiobook industry will be driven by this diversity of listener habits and preferences."

Blog / Company

Bookwire and ElevenLabs partner to bring ebooks to audio

Jane Friedman, one of the leading experts on publishing trends in the US and who publishes the newsletter The Bottom Line providing nuanced market intelligence to thousands of authors and industry professionals, stated that "some indie authors are using AI narration to support serializations offered for free on YouTube or elsewhere, and this generates advertising income and some subscription income. When people aren't paying for AI narration, naturally they'll be more open to it.

 Jane Friedman

THE BOTTOM LINE

I expect a tiered marketplace to evolve in the near term, where AI-narrated audio costs less for the consumer, and human-narrated audio carries a premium. What I don't know is how much the technology will

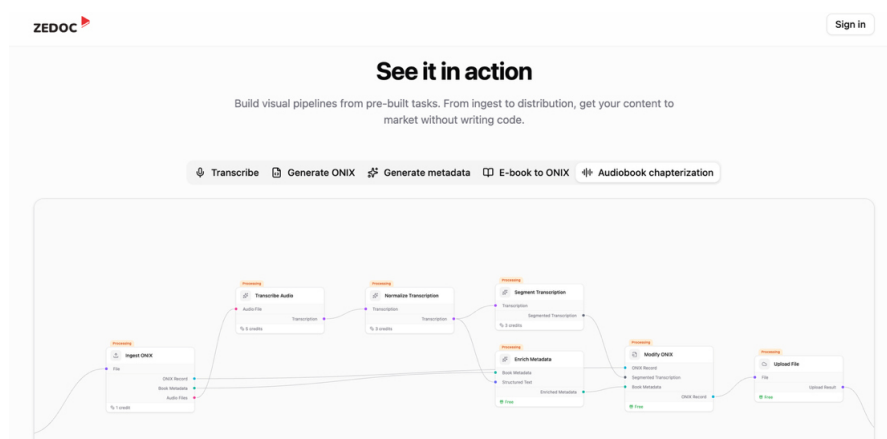
progress and ultimately make it impossible for the average listener to tell what AI is generated and what is not. Will that be a tipping point for acceptance? We'll see," added Friedman.

Corresponding to a more futuristic perspective, full on-demand creation of audiobooks — any book, in any voice and language, generated at the moment of listening — is described by several respondents as technically plausible in less than five years, with barriers coming from licensing rather than technology. Supporting this viewpoint, **Rickard Lundberg, CEO Aniara (Sweden)**, advocates for coexistence, with readers given the choice between human and AI narration on the same title. "In the near term, AI-generated audiobooks will dominate the tier just below the biggest bestsellers, while the very top sellers retain the resources to keep human narration as their flagship version. Longer term — we'd put it around 2033 — we expect production effort to become almost beside the point: the real shift will be readers choosing which voice narrates which book, the same way they choose a font or an edition today."

Robert Holmström, Managing Director, Earselect (Sweden), also believes in the coexistence of both models, but he expects that AI-narrated titles will eventually make up a larger portion of the overall catalogue. "The biggest unknown is how listeners will respond to AI voices in the long run. We don't have enough solid data yet. For instance, I've seen titles receive poor reviews specifically because of their AI narration, yet those same audiobooks had high completion rates. This could suggest that while listeners may complain, they still consume the content. However, it's too early to draw any firm conclusions."

Taking a leap further into the future, **Javier del Puerto, Founder of Kwalia (UK)**, is convinced of where these technological advances will take us: "Within three years the audiobook stops being a produced asset and becomes a rendering, generated at the moment of listening in

the voice and language the listener chooses. The constraint will be licensing rather than technology; we have been here before, as far back as when authors' objections made Amazon switch off the Kindle's read-aloud function in 2009. It will be negotiated, as it was when audiobooks reached streaming, and rights-holders who arrive with unified rights will get there first."



In agreement with this logic, “AI will power new workflows, substantially reducing the cost of releasing an audiobook. There will be a greater variety in quality, with books featuring fully synthetic voices succeeding in smaller language markets where audiobooks would otherwise not be made available. For services, the ability to switch voices based on the consumer's preference is already being introduced. Possibly, in a few years, we will have our own 'personal voice' which will travel with us,” indicated **Njål Hansen, Founder of Zedoc AI (Norway).**

Arantza Larrauri, Market Director for Europe and Latin America at De Marque (Canada), also sees the future as a coexistence of both models, but with a clear structure. “Thanks to AI, every book will have an audio version by default. Audio will simply become a standard format linked to every title — a future in which no book is left without one. At the same time, certain books — major new releases, landmark titles, and especially fiction with strong emotional weight — will also have human-narrated versions. These will be a premium layer, a desirable and necessary plus with a higher perceived value. The beauty of the human voice, recorded live and with genuine artistic intention, is

irreplaceable and fundamental for that kind of work. So rather than one model displacing the other, I see AI narration guaranteeing universal audio coverage, and human narration elevating the books that need and demand it.”

DE MARQUE'S AI CHARTER

Corporate

[Home](#) » [Blog](#) » De Marque's AI Charter

[Back to the blog](#)

Published on 14 May 2025

We are gradually incorporating artificial intelligence into our products with a single goal in mind: to always provide you with a better service. AI is part of our ongoing efforts in developing our services. With these new possibilities, we're creating features whose goals are to make the use of our platforms even more efficient.

OUR COMMITMENTS

As AI takes on a more prominent role in our products, we are eager to affirm our commitment to the ethical and responsible use of artificial intelligence at De Marque. Here are the six fundamental principles that guide our approach.

In sum, uncontrolled growth resulting from lower AI audio production costs also carries risks. A surge of low-quality, AI-generated titles does little to enrich the market, and poor quality can drive consumers away rather than draw them in. This potential scenario places greater responsibility on publishers, platforms and tech services providers to create structures that maintain the standard of quality needed to keep audiences engaged. Also, as content becomes increasingly abundant, many indicate the need for the emergence of new business models and services to facilitate content discovery in an age of abundance.

1.8.1 Build AI on ethically licensed data, transparent disclosure to the consumer

The industry's main anticipated action is consistent: build the AI opportunity on ethically licensed data, transparent disclosure to the consumer, fair compensation for creative narrators and creators, as well as clarity around rights, so that growth in production volume does not come at the expense of the trust that readers, listeners, authors, and

narrators place in publishers and retailers. Several respondents explicitly warn that this trust, once lost, would be far harder to rebuild than any single title's production cost.

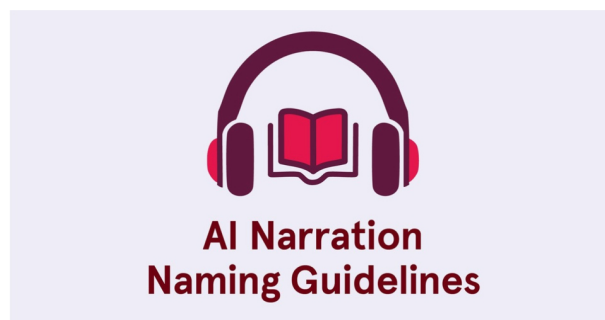
A voice talent doesn't have to choose between all or nothing in regard to AI replica voices. They might carve out well-defined uses for an AI version of their voice — a particular audiobook series, a foreign-language edition, an accessibility tool — while keeping the bulk of their work firmly in human hands. Done right, this opens a new income stream: licensing fees, per-project payments, ongoing royalties. But "done right" is doing a lot of work in that sentence. It requires contracts built on real, informed consent — not boilerplate — with narrowly scoped permissions, fair pay, expiration dates, clear renewal terms, and a genuine ability to walk away from future non-specified uses in the contract.

As demonstrated throughout this section, AI is becoming part of the work and daily life landscape that every member of the audio publishing industry must be aware of and create their own relationship to. Supporting this claim, **Michele Cobb, Partner, Forte Business Consulting (USA)**, stated, "It's fascinating to see how companies are incorporating AI into their workflows while they continue to utilize and support human narration. I do believe the two are not mutually exclusive, and reconciling technology vs art is not unique to audio publishing, but it provides greater challenges and more discussion in an industry built on emotional interpretation and the power of vocal performance. The next few years, I expect some interesting twists and turns as we parse through the intricacies of legal, ethical, and artistic challenges these new developments bring."

In line with these values and vision, **James Long, COO at Pan Macmillan (UK)**, foresees that "over the next three years, the opportunity is to move AI from a production tool into a genuine production strategy. The main goal for audio publishing right now, as I

see it, is to increase the number of titles available for listeners around the world, that they can discover and enjoy wherever they are, in their own language if possible. End-to-end application of AI to enhance the work of our teams will make more titles viable for audio publishing. That only works, though, if it's built on ethically licensed foundations and clear standards around rights and disclosure, so speed and scale never come at the expense of trust."

Further reflecting on the listener's trust imperative, **Jon Watt, Bonnier Group Digital Director and Chair, Audio Publishers Group (UK)**, indicates that "the opportunities AI presents for audio production, narration and translation are already well documented and will continue to grow over the next three years. As the volume of AI-generated audio content increases, however, quality will become an increasingly valuable differentiator. Whether content is narrated, produced or translated by humans, AI, or a combination of both, audiences will place greater value on high-quality audio experiences. As an industry, we have a responsibility to protect both author and consumer confidence. We can do this by being transparent about any uses of AI and by investing only in high-quality, AI-assisted content developed by trusted and ethical partners."

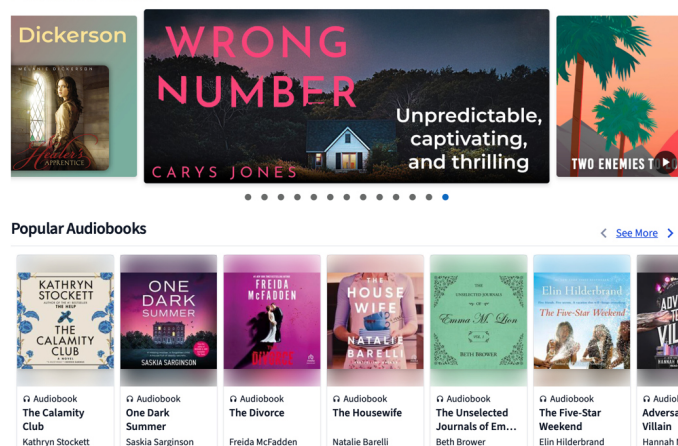


The Audio Publishers Group (part of the UK Publishers Association) and the Audio Publishers Association have collaborated to create international guidelines on how to label AI narrated audiobooks.

Building on this point, **Brad Rose, VP of Content Strategy, Hoopla (USA)**, indicates that "AI is changing how audiobooks are produced, reducing production cost and complexity, and making it possible to bring more niche, multilingual, and backlist titles to listeners than ever before. That means a bigger, more diverse catalogue reaching more

people, faster. But that growth only works if it's built on transparency, so that libraries and the platforms that support them can feel confident in what is being offered to patrons. That's why Hoopla continues to actively advocate for standardized identification practices within metadata from our content partners. Technology will keep evolving, but trust has to remain constant."

Audiobooks



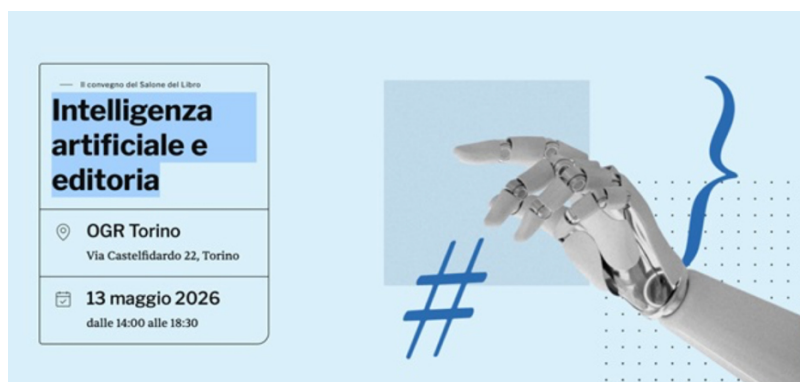
Artificial intelligence will reshape the entire book value chain, creating new business opportunities in areas such as the internationalization of content, the intelligent use of publishers' catalogues, personalized reader engagement, and the optimization of editorial processes. Consistent with this reasoning, **Raúl Pérez, Head of eBooks & Audiobooks Grupo Planeta**, reaffirms that "AI's greatest value lies not in replacing the creativity of authors, translators or editors, but in amplifying their ability to create, discover and connect more and better books with more readers. The challenge is to embrace these opportunities while preserving what defines publishing: editorial judgement, creativity, and the trust that readers place in their publishers."



In relation to AI production and licensing deals, **Lasse Korsemann Horne, Publishing Director at Saga Egmont (Denmark)**, strongly recommends publishers and rights holders to think beyond the immediate payout of a single AI licensing deal and consider their durable position in the value chain. A one-off AI sale — training data, narration rights, translation rights — may generate quick revenue, but if it's structured without safeguards, it could also be the last transaction you ever control. “At the moment audio publishers start selling audio recordings for AI-text-to-speech technology, neither studios nor publishers have the right to sell those recordings for training purpose. The biggest risk for the audio industry right now is basically publishers themselves selling training data at way to low prices and exposing themselves to legal enforcement by creators whose voices were stolen and authors/translators whose text were stolen too in this process of training AI-voices. Start planning for your long-term place in the value chain instead of short-term profit. Your first AI sale might well be your very last sale ever,” stated Lasse Korsemann Horne.

1.8.2 The challenges ahead: navigating a market in transformation

At the second edition of the conference **«Artificial Intelligence and Publishing»**, organized by Torino's Salone Internazionale del Libro in partnership with OGR Torino⁹, **Michael Tamblyn (CEO of Rakuten Kobo)** and **Jens Klingelhöfer (CEO and co-founder of Bookwire)** offered a privileged view of AI's impact on the publishing sector at an international scale.



One of the sharpest observations to emerge from this conversation was the growing competition between authentic author content and what Michael Tamblyn bluntly described as «AI slop» — a flood of low-quality, algorithmically generated material now crowding digital shelves and threatening to erode reader trust. For platforms like Kobo and Bookwire, this raises an urgent and still unresolved question: how do we disclose, clearly and consistently, what has been written by a human and what has been produced or significantly assisted by AI?

Both speakers argued that standardization and greater transparency in AI labelling are not just ethical imperatives but strategic ones — tools to protect and enhance the perceived value of genuine human creation. At the same time, they were careful to resist a purely defensive framing: the benefits that AI brings to the publishing chain are real and should be acknowledged, not dismissed. The challenge is not to reject AI, but to govern it — ensuring that the services and features built around a book, whether recommendation engines, audio enhancements or personalized reading experiences, enrich rather than disrupt the act of reading or listening.

Beyond the reader experience, Tamblyn and Klingelhöfer turned to the less visible but equally consequential question of how AI is reshaping the economics and internal culture of publishing organizations. On the commercial side, both called for a renegotiation of licensing structures — arguing that authors and publishers need better-framed deals that reflect the new realities of AI-driven content distribution and protect their rights in an environment where their works may be used to train or augment AI systems. Internally, the conversation pointed to the importance of clear employee policies governing the use of AI in daily work routines. Far from being a bureaucratic concern, such policies were framed as a positive instrument: when staff understand where AI can assist and where human judgement remains essential, organizations are better positioned to use technology in a way that amplifies, rather than replaces, human input. The consistent thread

running through both contributions was the same: AI is most valuable in publishing when it is deployed deliberately, transparently, and in service of the people — authors, editors, readers — at the heart of the industry.

1.8.3 Building smarter organizations

We can conclude the first section of this report reaffirming that the real challenge ahead isn't whether AI can make publishing more efficient — it's whether that efficiency comes at trust's expense. The companies that pull ahead over the next three years won't necessarily be the ones that move fastest with AI. They'll be the ones that pair it with openness: keeping humans in the loop, safeguarding creative rights, and never losing sight of the people whose talent made the publishing industry possible in the first place.

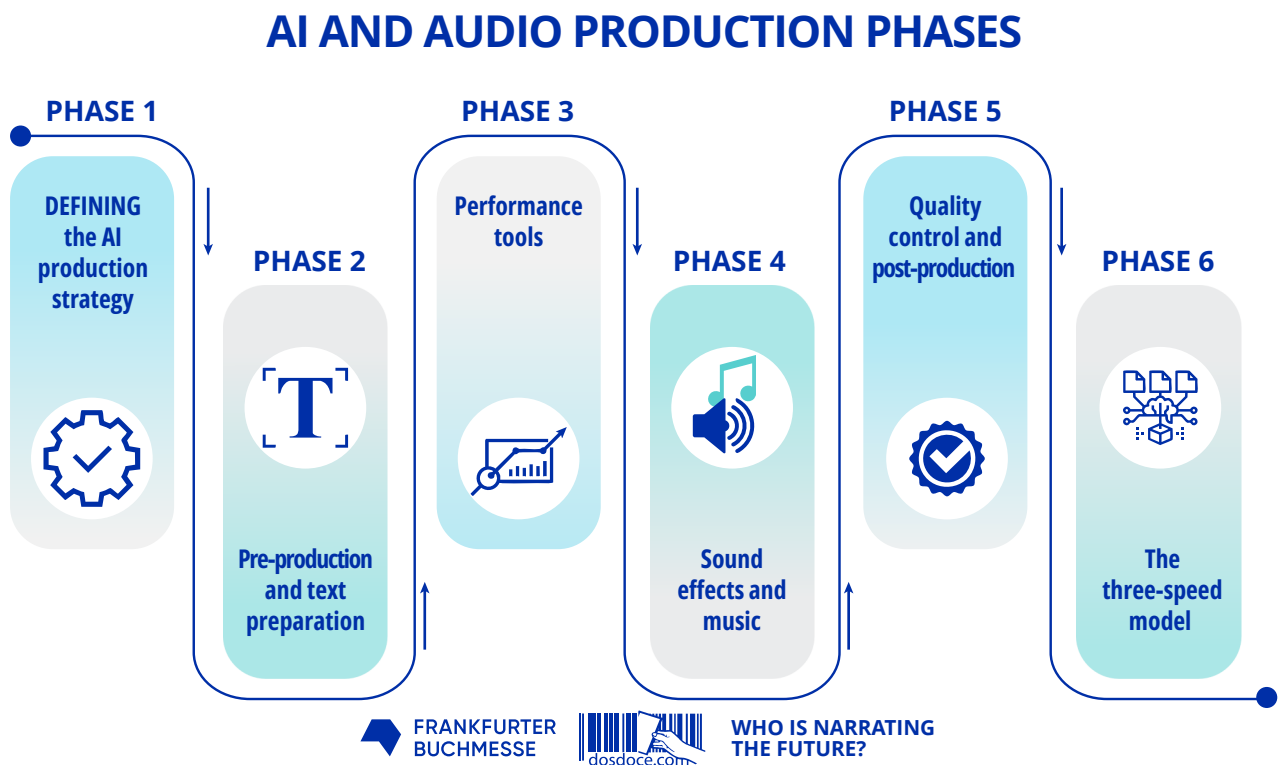
To help the publishing industry through this transition in a structured way, we propose the following four-lane action plan:

- Individual productivity: Embrace employees and authors using AI tools within existing day-to-day tasks and functions to gain efficiency.
- Workflow redesign: Redefine entire production processes by optimally coordinating collaborative work between humans and intelligent agents.
- Enhancing new value-added services: Identifying new go-to-market creation, discoverability and curation solutions based on higher reader / listener's personalization.
- Industry disruption: Actively search for new technologies that could transform the existing production and audiobook business model of the publishing industry.

SECTION 2

Outline of the AI and Audio Production Phases

The second section of this report describes an outline of the six phases that we consider essential when producing audio content with AI tools, as well as the fundamental role and weight of the human team in each of the phases:



Phase 1 — Defining the AI production strategy

Phase 2 — Pre-production and text preparation

Phase 3 — Performance tools

Phase 4 — Sound effects and music

Phase 5 — Quality control and post-production

Phase 6 — The three-speed model

The goal is not to list tools (which change every season) but to extract the editorial criteria, workflows, and lessons learned that allow AI to be adopted effectively, without sacrificing the editorial rigor that distinguishes a professional editorial-driven product from a mere technical demonstration.

■ 2.1 Defining the AI production strategy

The first phase of this report starts from an uncomfortable but necessary premise: AI in audiobook production is not, above all, a technological revolution, but a business decision. Before talking about tools and possible cost savings, it is necessary to analyze the industry's business opportunities in different markets and languages, the possibility of making a publisher's backlist profitable in the audio format, identifying which genres are underserved in audio, improving the efficiencies of human productions with AI touch-ups, pursuing the creation of categories that were previously economically unviable, etc. All key aspects of the audio business must be analyzed before diving into the redefinition of production processes.

Here is the uncomfortable truth about where the publishing industry's approach to AI has stalled: it focuses almost exclusively on cost reduction. Major publishers are implementing AI to cut spending on production processes and reduce staff —especially at entry levels— but these processes represent only the starting point of AI implementation, not its true potential.

The strategic questions that will define the next decade are barely being asked. Which processes can be fully automated? Which ones require human intervention at the start or end of the workflow, and which must remain entirely human? And, more importantly: how can AI help the publishing sector grow, rather than merely reducing its cost structure?

The opportunities are enormous and, to a large extent, untapped. AI-driven translation and production of e-books and audiobooks in multiple languages could bring key works and authors to entirely new audiences worldwide, at a scale and speed previously impossible. Redefining marketing and distribution —moving from a local scope to a truly global one— becomes feasible once language, narrative contextualization, and local marketing adaptation of these titles stop being a bottleneck. AI models themselves represent a potential new revenue stream, not merely a production tool. However, few players are analyzing these growth and business development possibilities with the seriousness they deserve.

To unblock this status quo, the first step begins with the CEO, senior executives, and the executive committee, embarking in a solid understanding of AI fundamentals to be able to define an AI & Audio strategy that takes into consideration the real technical capabilities of these technologies, as well as the human impact on narration, translation, casting, etc., in parallel to the catalogue growth potential and derived business opportunities.

Once the organization reaches a higher level of AI understanding, publishers and production studios will stop treating AI as a bolt-on tool for isolated tasks and instead redesign the full audio production workflow around it — from text preparation and editorial preproduction through voice synthesis, quality control, and technical postproduction — rather than simply layering new tools onto existing production pipelines.

The problem AI solves

As discussed during DPR's AudioNext Summit (May 2025)¹⁰, approximately 95% of eBooks have yet to be converted to audio; that's 5.1 million books that don't have an accompanying audiobook version.



The reason for this enormous gap is not lack of interest, but financial. For example, a professional narrator in Latin America charges between \$140 and \$160 per finished hour, and an audiobook based on a roughly 300-page print book, representing around 90,000 words, requires between 8 and 12 hours, putting traditional production between \$2,000 and \$3,000 per title. This can cost up to five times that in mature markets such as English-speaking or Nordic countries. For a publisher with 500 backlist titles, creating audiobooks of the entire catalogue with human voice can cost more than a million dollars and may not recoup the investment for more than 5 years. Synthetically voiced audiobooks do not present this same challenge.

This presents an opportunity for publishers to invest in. With growing demands for accessibility and the need to expand the audiobook catalogue in many languages and markets, there is no doubt that AI-narrated audiobooks have an important and expanding role to play.

A real cost comparison

One of the key aspects to consider in this first phase is the need to carry out a real analysis of costs and assumed savings. Comparing the cost of synthesis tokens with the total cost of a traditional production is a serious mistake, equivalent to comparing the price of paper with the cost of producing an entire book. A responsible AI production does not eliminate the stages of the traditional process such as text preparation, quality control, post-production, mastering, etc.; it merely enhances them.

When all these production stages are accounted for, including the actual quality-control listening time (at least 1.2 times the length of the audiobook, with iterations), the real cost reduction compared to human production stands between 25% and 50% in the second and third year after starting AI productions, not the impressive 85-95% savings in the first year that is often cited in AI voice providers' sales presentations.

Lesson learned: the argument is not just savings

Beyond budget and cost/time savings, publishers and streaming platforms producing content should analyze the creative and IP-related possibilities offered by AI tools, such as solving casting challenges without sacrificing artistic coherence, recovering the voice of deceased authors from sound archives (the recent case of Bolinda and Barbara Cartland), or personalizing the listener's experience by offering voice variants built on a human voice's master performance (Storytel's Voice Switcher). These arguments—harder to make in a publisher's internal meetings or industry forums than a simple 90% discount—are the ones that will build a financially sustainable and lasting audio industry.

Key takeaways from Phase 1:

- Analyze worldwide market and business opportunities (genre, platforms, competition).
- Identify creative and catalogue possibilities beyond short-term savings.
- Calculate the comparative cost: traditional production vs. AI voice synthesis, with the three cost layers (tokens, quality control, post-production).
- Structure the catalogue and the investment payback period.
- Resolve the contractual chain: recording rights, cloning consent, data policies, transparency with the listener.

2.2 Pre-production and text preparation

The second phase of production involves preparing the text for audio. It is crucial to determining the final quality of the audio content and involves two distinct levels.

First is text adaptation —linear and irreversible— to the nature of audio: what to do with tables, footnotes, referenced images, and structural breaks that are communicated visually. Second is preparation for the narration, which diverges depending on whether it is a human narrator (who intuitively resolves what is not explicit) or a synthesizer (which has no judgment of its own and needs every decision encoded in advance).

The workflow

In this phase we recommend that the human production team document a step-by-step sequential process: pre-production assessment (diagnosis of text quality prior to audio editing, density of special elements such as foreign words or names of places, linguistic

and text complexity such as number of characters or different tones within the text); extraction and cleanup of the original text; editorial treatment of each special element (tables, notes, images, lists, epigraphs); final human verification of the pre-production text; pronunciation guide; and final revision and standardization of the text —numbers, abbreviations, punctuation as a rhythm marker— before it enters the synthesizer.

Lesson learned: human verification cannot be delegated

Today no AI tool can fully operate without errors —an unclear rewritten sentence, a table converted to text with incorrect data. Detecting these discrepancies requires understanding the book's argument, and today only a human does that well.

Verification at each step, documented in a log with date, points reviewed, and corrections, is not just a quality control measure: it is also the record that demonstrates the rigor of the production process to an author, a literary agency, or a publisher still skeptical about the use of AI tools.

The pronunciation guide deserves special mention because of its importance: once built and verified, it serves as a compendium for future projects with those same references, places or characters, reducing the pre-production time for new titles.

Key takeaways from Phase 2:

- Distinguish the two levels: adapting the text for audio and preparing it for whoever performs it (narrator or synthesizer).
- Pre-production assessment: diagnosis of text quality, special elements, linguistic and register complexity.

- Extraction and cleanup of the original text (layout artifacts, encoding issues, syllable-division hyphens).
- Editorial treatment of tables, footnotes, images, lists, and epigraphs.
- Structured human verification of the prepared text (transitions, passages in other languages, reading aloud).
- Construction of the pronunciation guide by category (place names, technical terms, acronyms, foreign languages).
- Final normalization: numbers, abbreviations, punctuation, pace, and SSML tagging.
- Delivery of two final documents: script and pronunciation guide prepared for the human narrator, and normalized text with tagging for the chosen synthesizer.

■ 2.3 Performance tools

The third phase addresses the selection and direction of the voice synthesis engines themselves.

One of the main recommendations arising from our conversations with more than 50 industry professionals is that choosing an artificial voice without listening to it with real text from the work is one of the costliest mistakes a publisher or its production studio can make. Voice spec sheets say nothing about what truly matters —how that voice sounds reading real prose in the specific language of the project, or what happens when it encounters an acronym or a proper name it fails to pronounce correctly or that sounds strange.

The evaluation protocol

For these contexts, we propose a four-step protocol. First, build a test with real excerpts from the manuscript (narrative prose, a passage with problematic elements, a tone shift, and the book's opening paragraph). Second, generate each excerpt with neutral settings across several

potential voices from different vendors. Third, conduct a blind listening session evaluating naturalness, fluency, handling of problematic elements, consistency, and listening fatigue; and finally, record the decision along with its rationale as part of the project's documentation.

Directing a synthetic voice is more like a translation than a conversation: the human production team must translate its years of production and directing experience, as well as its interpretive and artistic intuition, into a series of decisions that the model can process.

Architecture: cloud, local, and hybrid

The choice of where the synthesis model runs has consequences that go beyond cost per character: it affects manuscript privacy, dependence on an external provider, and the quality of regional accents.

Having listened to more than 50 industry professionals, we recommend deciding per project based on the following four variables: privacy, cost structure, tolerance for connectivity dependence, and the required accent quality.

Key takeaways from Phase 3:

- Build a test with real excerpts and evaluate candidate platforms through blind listening.
- Define the artistic direction brief: the narrator's voice profile, tone, and register.
- Calibrate SSML and empirically verify which tags the chosen engine honors.
- Build the voice profile sheet: locked parameters, reference excerpt, and clean-start protocol per chapter.

- Decide on the production architecture (cloud, local, or hybrid) based on privacy, cost, dependency, and accent quality.
- If applicable, run the voice cloning protocol with documented consent and a disclosure note.

■ 2.4 Sound effects and music

The fourth phase broadens the view beyond narration, beyond voices, to make key decisions about music, effects, and sound design (SFX).

These are younger, faster-changing tools than voice synthesizers, which is why the evaluation criteria matter most. Our lesson and recommendation is clear: prioritize clear legal provenance of the music library over whichever platform sounds best today.

In this context, we propose creating a shared internal table to manage key decisions on music, SFX, and podcasts, including the following fields: provenance of the training data, retention and privacy of uploaded material, real creative control (not just accepting whatever the model produces), output format compatible with the editing workflow, and provider stability.

Enriching this viewpoint, **André Calgaro, CEO Narratix (Brazil)**, indicates that “one part of the process that is still not so well covered is truly high-quality editing and post-production specifically for audiobooks. The tools that we have now in the market only help partially or offer limited time savings. With the growth of the industry, it’s clear that more sophisticated audio productions, with music and sound effects, such as audio dramas, are becoming an important trend. If AI can take care of more manual and repetitive (but still crucial) editing tasks, audio professionals can focus more on the creative side, including music composition, sound design, spatial audio, field recording and other activities that can enhance the storytelling.”

Key takeaways from Phase 4:

- Define which sound layers the project needs: musical identity, ambience/effects, its own podcast.
- Evaluate music and SFX platforms, verifying licenses at the official source.
- Generate and document each piece (sound logo, ambience, transitions) using layered, structured prompts: source, action, production.

■ 2.5 Quality control and post-production

The fifth phase starts from a well-known management principle: a good take is worth more than endless post-production.

Applying this concept to AI audio content production, strong pre-production is better than endless rounds of quality controls (QC). This concept ties in with the arguments and recommendations set out in Phase 2 of this report. QC checks do not manufacture quality; they only verify it. If errors turn up everywhere during review, the problem is not in QC but in the earlier phases, in pre-production or in the generation of the content itself.

In this context, we recommend separating the detection process from the correction tasks. Detecting means going through the material and logging findings without resolving them on the spot; correcting means clearing that log afterward.

Most of the professionals consulted during the research phase of this report recommend two complete passes / quality reviews —a reading pass, following the text to detect fidelity and interpretive consistency,

and a second technical pass to detect clicks, breaths, and artifacts, among other possible improvements.

These passes should be followed by a final sampling check, not three or four complete passes as was done years ago, which would be unworkable in terms of time and strain on the production team, given that each complete pass takes around 1.2 times the length of the audio.

AI tools do not decide what is wrong; rather, they help the production team by flagging where to look more closely. The golden rule is to provide the already verified and cleaned text as reference, as opposed to the original manuscript, so as not to contaminate the list with false positives.

Key takeaways from Phase 5:

- Draft the QC plan: planned passes, logging template, and checks carried over from earlier pre-production phases (see Phase 2).
- Run the material through a text-audio alignment tool and adjudicate each discrepancy (real or false positive).
- Run the reading pass and the technical pass by section, in a controlled listening environment, without correcting anything yet.
- Resolve the log: regenerate content/interpretation errors, clean up the signal for technical defects.
- Apply spatial and dynamics post-production only if the project requires it (audio drama or multi-voiced audio fiction).
- Consolidate the QC and post-production document, with lessons learned for the next project.

2.6 The three-speed model

The final phase invites publishers and streaming platforms to change scale, that is, to move from producing a single title or a small collection of audiobooks to producing an entire catalogue.

The central idea is to dismantle the false dilemma that arises at many book-world conferences and debates —human voice or synthetic voice?

THE THREE-SPEED MODEL

First

PREMIUM HUMAN PRODUCTION

Reserved for the year's flagship titles with the greatest literary, editorial, or potential commercial value, as well as titles with the greatest interpretive demands or technical complexity.



Second

HYBRID PRODUCTION (HUMAN + AI TOOLS)

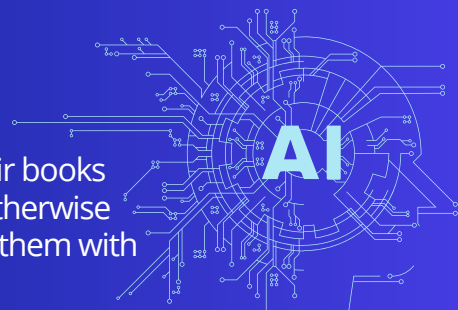
which will be the most common production model for most publishers and streaming platforms



Third

FULLY AI PRODUCTION

Reserved for bringing backlist authors and their books to life, as well as niche categories that would otherwise never get an audio format because producing them with human voices would be unviable.



As illustrated in the seventh graph of this report, the publishing sector AI dilemma is resolved by establishing three production models:

- First: Premium human production reserved for the year's flagship titles with the greatest literary, editorial, or potential commercial value, as well as titles with the greatest interpretive demands or technical complexity.
- Second: Hybrid production (human production assisted by AI tools) which will be the most common production model for most publishers and streaming platforms.
- Third: Fully AI production reserved for bringing backlist authors and their books to life, as well as niche categories that would otherwise never get an audio format because producing them with human voices would be unviable.

In the third model, human intervention will be complementary primarily reviewing processes, workflows and outputs. Revenue from these productions will help publishers and platforms have more budget to produce more titles in the two models described above.

In short, assigning a title to one of these models is determined by crossing two factors: the title's commercial value and the text's interpretive demand, as well as applying three determining filters: rights and consent, language and accent, and realistic budget and timeline.

To ensure orderly management of that coordination, we recommend three steps:

- Standardization: developing checklists that determine the criteria for each speed.
- Creating a library of reusable assets (voice profile sheets, pronunciation guides, templates) that makes the second title cost less than the first,

- and finally, having a stage-by-stage tracking board that makes the status of each title visible.

Most important is that the person responsible for the AI audiobook production team should not be perceived merely as a tool manager. Our experience shows that this role is more like an orchestra conductor who knows which instrument comes in at each moment and who reserves their attention —the most expensive and valuable human resource in the digital age we live in— for the editorial decisions that truly cannot be delegated.

Key takeaways from Phase 6:

- Map the entire catalogue across the three production models described, avoiding the bias of producing out of inertia.
- Build the title-by-title allocation table: commercial value × interpretive demand × editorial value, plus the three filters of rights, language, and budget.
- Design the task allocation for the hybrid workflow (before/during/after) for each title that requires it.
- Set up the scale-management system: stage-by-stage board, library of reusable assets, per-model checklists, and real capacity calculations.
- Update the full production record (Phases 1 to 6) as a living template to manage the publisher's / platform's future catalogue (new releases, acquisitions, etc.)

■ 2.7 Takeaways to adopt AI effectively

Below, we summarize the main takeaways, recommendations and key points to keep in mind when designing an AI adoption strategy for audio production:

- The economic benefits are real, but medium/long term. Cost reduction with AI tools compared to traditional human production stands between 20% and 50% in the second and third year, not the 85-95% figures promoted by technology providers in the audio industry. Comparing complete workflows, not isolated cost lines, is essential for realistic budgeting and determining annual returns on investment.
- Text preparation determines the quality of the final audio more than any other step. Investing time and editorial judgment in extraction, cleanup, treatment of special elements, and building the pronunciation guide pays off many times over in every subsequent phase.
- The legal and ethical framework —explicit consent for voice cloning, data retention policies, transparency with the listener— must be resolved before production begins, not after. Publishers, platforms, and production companies that build that solid foundation will be better positioned once regulation within the audio industry becomes established.
- Endless QC is a symptom of poor pre-production or generation, not a virtue of the process. Separating detection from correction and sizing the effort realistically (approximately 1.2 times the length of the audio per pass) makes the process sustainable.

- The smartest decision is not choosing between human voice and synthetic voice but assigning the right model to each title in a catalogue, with criteria defined in writing before production begins, crossing editorial and commercial value with interpretive demand.
- At every stage —from text cleanup to mastering— the machine proposes and the human with judgment accepts, corrects, or rejects. That principle, more than any specific tool, is what distinguishes a professional editorial production from automatic generation.
- Tools will change over time. What does not change is the evaluation criteria: legal provenance of data, real creative control, privacy, and human verification at every point where editorial judgment cannot be delegated.

In summary, effectively integrating AI into editorial audio production is not about finding the right tool, but about building a management process —documented, verifiable, and repeatable— into which any tool can be inserted with full confidence into the production at any phase of audio content production. When moving from a few production titles to many, the bottleneck stops being production overloads and becomes management and team coordination.

■ 3. Sources

1. The Whitepapers of Frankfurter Buchmesse
<https://www.buchmesse.de/en/whitepaper>
2. Audiobook Global Growth Report (2024)
<https://www.dosdoce.com/2024/11/28/the-audiobook-industry-will-grow-more-than-25-annuallyaround-the-world-in-the-next-7-years/>
3. Netflix Earning Results
https://s22.q4cdn.com/959853165/files/doc_financials/2026/q2/Netflix-Inc_Earnings-Call_2026-07-16T00_00_00_English-1.pdf
4. The Working Narrator
<https://www.dosdoce.com/2026/07/01/the-working-narrator-el-trabajo-detras-del-trabajo-de-los-narradores-de-audiolibros/>
5. Global Audio AI-Driven Tools & Services Map
<https://www.dosdoce.com/2025/09/15/ai-and-audio-report/>
6. PARIX Audio Day 2026
<https://www.dosdoce.com/2026/03/17/parix-audio-day-2026-one-of-the-most-international-gatherings-in-the-audio-industry/>
7. Audio Publishers Association Sales Survey
<https://www.audiopub.org/surveys>
8. Edison Research for Spoken
https://www.prnewswire.com/news-releases/in-largest-study-of-its-kind-us-fiction-audiobook-consumers-rate-spoken-multi-cast-higher-than-human-narration-302824457.html?tc=eml_cleartime
9. Artificial Intelligence and Publishing (OCR Torino)
<https://www.dosdoce.com/2026/05/20/artificial-intelligence-and-publishing-torino-book-fair-conference-2026/>
10. All About Digital Publishing
<https://all-about-digital-publishing.com/en/1/start>

■ 4. Credits



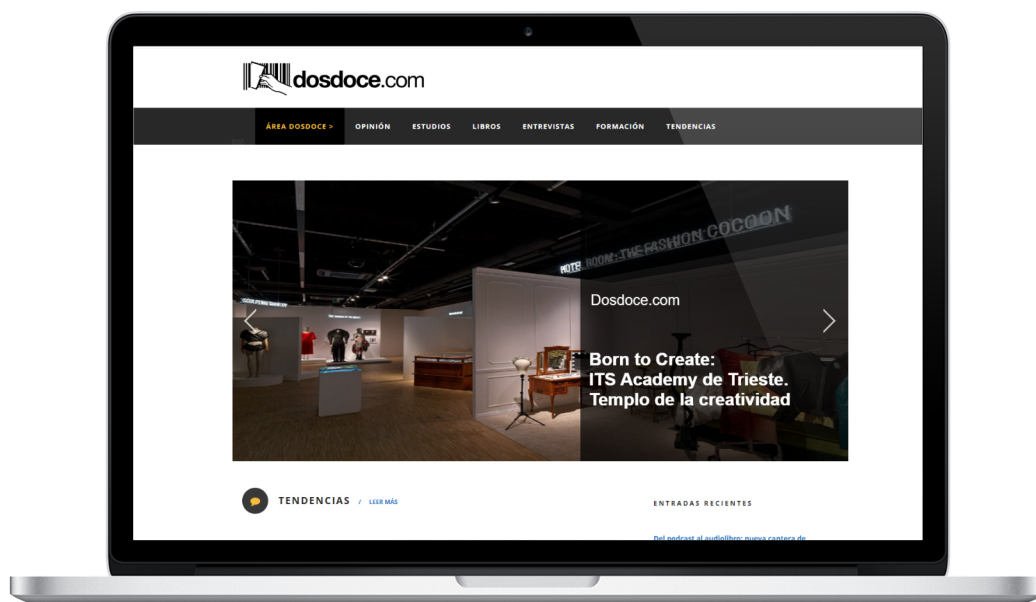
Dosdoce.com was created in March 2004 with the purpose of trying to better understand the digital era in which we live. Dosdoce's goal has been to encourage professionals in the creative industries to digitally transform their organizations to grow their business activities through the benefits derived from the Internet.

Throughout these 20+ years, Javier Celaya, founding partner of Dosdoce.com, has advised leading companies in the digital economy such as Bookwire, Storytel, Podimo, De Marque, Skeelo, Benetech, TheCopia, Bayard, Aniara.One, Eeditis, among others.

Dosdoce has also produced more than 300 studies on the use of new technologies in different areas of the cultural sector (business models, distribution strategies, digital content creation, etc.) and published thousands of articles to help professionals in the sector understand the opportunities offered by the internet.

Throughout his career, Javier has shared his knowledge in numerous master's and training courses, teaching at the Advanced Publishing Institute at New York University (NYU), the Master's in Publishing at the Autonomous University of Madrid (UAM), the online Master's in Criticism and Cultural Communication at the University of Alcalá and El Cultural magazine, the Master's in Cultural Management at Carlos III University, as well as the Master's in Publishing at the Complutense University of Madrid. For over five years, he was co-director of the

Master's in Corporate and Institutional Communication 2.0 and the master's in digital Publishing at the University of Alcalá, as well as Academic Co-Director of the first two editions of the Master's in Publishing Management and Digital Marketing at the University of Alcalá in collaboration with the Unidad Editorial School (ESUE). He is also part of the faculty at PARIX, a training school for book professionals managed by the Germán Sánchez Ruipérez Foundation. On this online school platform, Javier coordinates four courses related to the audio industry: one on AI & Audio Production Processes; another one on key aspects to consider on podcast, audio series, and audio drama production strategies; another on the creation and distribution of audio content for the publishing sector; and the latest one produced analyzes the Trends on Latin America Storytelling & Audio Content.



About Frankfurter Buchmesse

Frankfurter Buchmesse is the international publishing industry's leading trade fair and the most important international marketplace for content – from novels and audiobooks to academic databases and stories for films, games and virtual reality experiences. Around 120,000 publishing professionals from more than 130 countries meet here with partners from the technology sector and from related creative and cultural industries, sparking new partnerships and business models. Since 1976, the book fair has featured an annual guest country, which showcases its book market, literature and culture to attendees in a variety of ways. Frankfurter Buchmesse also organizes the participation of German publishers at international book fairs and hosts trade events throughout the year in major international markets. The 2026 Frankfurter Buchmesse will take place from October 7th-11th. For more information and tickets, please visit www.buchmesse.de/en.



About Frankfurt Audio

From audiobooks to streaming platforms and podcasts: Frankfurter Buchmesse's Frankfurt Audio is all about the spoken word. Latest innovations, trends and services from the audio sector are presented by audio players and service providers in Halls 4.0 and 4.1 and on the fair's stages in Frankfurt. A programme of panel discussions, talks and get-togethers provides the opportunity for audio and publishing professionals to connect and gain insights into different audio areas and markets. The book fair's Audio Ambassador 2026, Ama Dadson, helps to connect interested members of the publishing community with audio exhibitors at the book fair. For more detailed information about the audio area and events related to Frankfurt Audio, please visit www.buchmesse.de/en/audio.

Imprint

Who is narrating the future? AI adoption, use and outlook in audio production

© Dosdoce.com September 2026

Research Team: [Valle Dessaint](#), [Michele Cobb](#), [David Roa](#) and [Javier Celaya](#)

All rights reserved. No part of this publication may be reproduced, stored in any retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of Dosdoce.com.

Cover photo: [Roberto Carreras](#) ©

Graphic Designer: [Génesis Arriechi](#)

Contact details:

Javier Celaya

 jcelaya@dosdoce.com

+34.606.367.708

 www.dosdoce.com

 www.linkedin.com/in/javiercelaya