Palabra Clave, (La Plata), abril - septiembre 2026, vol. 15, núm. 2, e282. ISSN 1853-9912
Universidad Nacional de La Plata
Facultad de Humanidades y Ciencias de la Educación
Departamento de Bibliotecología

Inteligência artificial em audiolivros: aplicações e perspectivas

Artificial intelligence in audiobooks: applications and perspectives

La inteligencia artificial en los audiolibros: aplicaciones y perspectivas

Suellen Souza Gonçalves
Universidade Federal de Minas Gerais, Brasil
https://ror.org/ 0176yjw32
Instituto Federal do Norte de Minas Gerais, Brasil
https://ror.org/03w6rv149
suesouzag@gmail.com
https://orcid.org/0000-0002-9330-2440

Patrícia Nascimento Silva
Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Brasil
https://ror.org/ 0176yjw32
patricians@ufmg.br
https://orcid.org/0000-0002-2405-8536

Resumo

O uso de técnicas de Inteligência artificial (IA) no contexto dos audiolivros tem ampliado as possibilidades de acessibilidade, personalização e imersão, permitindo desde o reconhecimento e a síntese de voz até experiências multimodais interativas e recomendações personalizadas, além de potencializar a recuperação de conteúdo e ampliar o acesso à informação. Este estudo teve como objetivo identificar, na literatura acadêmica, estudos sobre o uso da IA em audiolivros. Para tanto, foi realizada uma revisão de literatura nas bases *Scopus, Web of Science, ACM Digital Library, IEEE Xplore* e *Scielo,* entre maio e agosto de 2025, resultando na seleção e análise de 35 artigos. Os resultados revelam que os trabalhos concentram-se em quatro categorias: (i) reconhecimento de fala; (ii) síntese de voz e personalização; (iii) experiências baseadas em voz; e (iv) IA generativa e LLMs. Observou-se que predominam estudos técnicos voltados para o Reconhecimento Automático de Fala e Síntese de Voz, enquanto experiências baseadas em voz e aplicações de LLMs ainda aparecem de forma emergente, indicando tendências futuras. Os audiolivros também são frequentemente utilizados como *corpus* técnico para o desenvolvimento de modelos, com poucos estudos voltados à melhoria direta da experiência de uso, além de uma escassez de pesquisas na área da Ciência da Informação. Conclui-se que, apesar dos avanços recentes, há lacunas relativas à falta de estudos centrados no usuário e ao uso predominante dos audiolivros como *corpus* técnico, assim como poucos aspectos éticos e sociais. Este panorama oferece subsídios teóricos e práticos para pesquisas futuras na área.

Palavras-chave: Audiolivros, Inteligência artificial, Ciência da informação, Revisão de literatura.

Recepción: 30 Septiembre 2025 | Aceptación: 16 Febrero 2026 | Publicación: 01 Abril 2026

Cita sugerida: Gonçalves, S. S y Silva, P. N. (2026). Inteligência artificial em audiolivros: aplicações e perspectivas. *Palabra Clave (La Plata), 15*(2), e282. https://doi.org/10.24215/18539912e282

Abstract

The use of Artificial intelligence (AI) techniques in the context of audiobooks has expanded the possibilities for accessibility, personalization and immersion, covering aspects from voice recognition and synthesis to interactive multimodal experiences and personalized recommendations, in addition to enhancing content retrieval and expanding access to information. This study aimed to identify studies on the use of AI in audiobooks in the academic literature. To this end, a literature review was conducted in the Scopus, Web of Science, ACM Digital Library, IEEE Xplore and Scielo databases, between May and August 2025, resulting in the selection and analysis of 35 articles. The results reveal that the studies focus on four categories: (i) speech recognition; (ii) voice synthesis; and personalization; (iii) voice-based experiences; and (iv) generative AI and LLMs. It was observed that technical studies focused on Automatic Speech Recognition and Voice Synthesis predominate, while voice-based experiences and LLM applications are still emerging, indicating future trends. Audiobooks are also frequently used as technical *corpora* for model development, with few studies focused on directly improving the user experience, in addition to a scarcity of research in the field of Information Science. It can be concluded that, despite recent advances, there are gaps related to the lack of user-centered studies, the predominant use of audiobooks as a technical *corpus* as well as few ethical and social aspects. This overview provides theoretical and practical support for future research in the area.

Keywords: Audiobooks, Artificial intelligence, Information science, Literature review.

Resumen

El uso de técnicas de inteligencia artificial (IA) en el contexto de los audiolibros ha ampliado las posibilidades de accesibilidad, personalización e inmersión, abarcando aspectos que van desde el reconocimiento y la síntesis de voz hasta experiencias multimodales interactivas y recomendaciones personalizadas, además de mejorar la recuperación de contenidos y ampliar el acceso a la información.El objetivo de este estudio era identificar investigaciones sobre el uso de la IA en audiolibros en la literatura académica. Para ello, se llevó a cabo una revisión bibliográfica en las bases de datos *Scopus, Web of Science, ACM Digital Library, IEEE Xplore* y *Scielo*, entre mayo y agosto de 2025, lo que dio como resultado la selección y el análisis de 35 artículos. Los resultados revelan que los estudios se centran en cuatro categorías: (i) reconocimiento de voz; (ii) síntesis de voz y personalización; (iii) experiencias basadas en la voz; y (iv) IA generativa y LLM. Se observó que predominan los estudios técnicos centrados en el reconocimiento automático del habla y la síntesis de voz, mientras que las experiencias basadas en la voz y las aplicaciones LLM aún están emergiendo, lo que indica las tendencias futuras. Los audiolibros también se utilizan con frecuencia como *corpus* técnicos para el desarrollo de modelos, con pocos estudios centrados en mejorar directamente la experiencia del usuario, además de la escasez de investigaciones en el campo de la ciencia de la información. Se puede concluir que, a pesar de los avances recientes, existen lagunas relacionadas con la falta de estudios centrados en el usuario, el uso predominante de audiolibros como corpus técnico y la escasa atención prestada a los aspectos éticos y sociales. Esta visión general proporciona un apoyo teórico y práctico para futuras investigaciones en este ámbito.

Palabras clave: Audiolibros, Inteligencia artificial, Ciencia de la información, Revisión bibliográfica.

1. Introdução

Com o passar dos anos, os suportes destinados à informação passaram por transformações significativas para acompanhar os avanços tecnológicos e as novas formas de comunicação. A transição do papel para os dispositivos digitais marcou uma ampliação nos meios de acesso aos conteúdos informacionais. O surgimento de mídias eletrônicas, como *CD*s, *tablets* e *smartphones*, contribuiu para tornar os documentos mais dinâmicos, fáceis de recuperar e com maior circulação entre os usuários (Lourenço, 2005; Silva & Neves, 2013).

Ao acompanhar essa evolução dos suportes informacionais, os audiolivros passaram a ocupar um espaço importante como um formato alternativo de acesso ao conteúdo escrito. Sua origem está associada à invenção do fonógrafo por Thomas A. Edison, em 1887, tecnologia que possibilitou as primeiras gravações de trechos literários (Rubery, 2016).[1] O audiolivro consiste na reprodução sonora de um texto, permitindo que a informação cultural seja consumida por meio da escuta. Em muitos casos, esse formato é enriquecido com recursos adicionais, como efeitos sonoros e trilhas musicais, podendo ser narrado por autores, locutores profissionais, leitores amadores ou por sistemas de voz sintetizada (Have & Pedersen, 2019).

A incorporação de tecnologias digitais ampliou as possibilidades de produção e distribuição dos audiolivros, permitindo não apenas a gravação tradicional, mas também a aplicação de sistemas inteligentes em sua criação e personalização. Com o avanço da Inteligência Artificial (IA), surgem soluções capazes de sintetizar vozes naturais e ajustar entonações de acordo com o gênero textual, tornando o acesso mais dinâmico e inclusivo (Chen *et al*., 2023).

A IA tem se consolidado como uma aliada importante na promoção da inclusão digital, sobretudo por sua capacidade de imitar processos cognitivos humanos e apoiar o desenvolvimento de tecnologias assistivas avançadas. Essas soluções ampliam as possibilidades de acesso à informação, favorecendo práticas de interação mais acessíveis e inclusivas (Russell & Norvig, 2016). Entre as técnicas que sustentam a IA, destacam-se o aprendizado de máquina, o processamento de linguagem natural e outras formas de automação, que contribuem para a criação de sistemas inteligentes aplicáveis em diferentes áreas do conhecimento (Goodfellow, Bengio & Courville, 2016).

Apesar dos avanços das tecnologias de IA e da crescente popularização dos audiolivros, observa-se uma lacuna na literatura quanto à aplicação dessas técnicas em plataformas digitais voltadas à leitura e à escuta. Rubery (2016) e Have & Pedersen (2019) apontam que as discussões sobre audiolivros ainda se concentram em seus aspectos culturais e narrativos, havendo poucas investigações sobre sua estrutura informacional e tecnológica. Assim, o problema desta pesquisa consiste em compreender de que modo as técnicas de IA têm sido aplicadas aos audiolivros e em que medida podem contribuir para aprimorar a recuperação e a acessibilidade da informação nesses ambientes digitais.

Diante deste contexto, tem-se as seguintes questões norteadoras deste estudo: os audiolivros possuem interface com a IA? Quais técnicas de IA têm sido utilizadas nos audiolivros? Quais funcionalidades e recursos dos audiolivros utilizam a IA? Desta forma, esta pesquisa tem como objetivo identificar estudos sobre o uso da IA em audiolivros. O estudo integra uma pesquisa de doutorado. A tese em andamento está sendo desenvolvida no Programa de Pós-Graduação *stricto sensu* em Gestão e Organização do Conhecimento (PPGGOC), da Universidade Federal de Minas Gerais, na linha de pesquisa Gestão e Tecnologia da Informação e Comunicação (GETIC), com foco na recuperação de informação em plataformas de audiolivros.A relevância desta análise está na necessidade de compreender o potencial da IA como recurso de apoio aos audiolivros, contribuindo para ampliar a inclusão social em ambientes digitais. Espera-se que os resultados obtidos sirvam de base teórica e prática para a criação de soluções inovadoras, capazes de aprimorar a recuperação e o acesso a conteúdos nesse formato. Ao considerar o panorama atual, reconhece-se que existem diferenças entre os audiolivros voltados à inclusão, geralmente produzidos por iniciativas voluntárias e destinados ao uso de tecnologias assistivas, e aqueles desenvolvidos para o grande público em plataformas

comerciais. No entanto, esta pesquisa concentra-se no contexto das plataformas comerciais de audiolivros, que apresentam maior alcance, atualização tecnológica e potencial de integração de recursos inteligentes. O foco decorre dos resultados obtidos em uma pesquisa anterior das autoras (Gonçalves & Silva, 2025), na qual foram analisadas cinco plataformas comerciais, cujas limitações identificadas na representação e na recuperação de informação servem de base para o aprimoramento da proposta atual. Assim, o estudo busca propor recomendações que tornem essas plataformas mais acessíveis e eficazes, promovendo uma recuperação de informação mais inclusiva e alinhada às necessidades dos usuários. A revisão aqui apresentada corresponde a uma etapa inicial de um projeto de doutorado em andamento, que pretende identificar e analisar soluções baseadas em Inteligência Artificial aplicáveis à melhoria da acessibilidade e da experiência do usuário em ambientes de audiolivros.

2. Inteligência artificial aplicada em audiolibros

Desde a década de 1960, a Ciência da Informação (CI) vem dedicando parte de suas pesquisas à compreensão dos processos de organização, tratamento e uso da informação, buscando facilitar sua recuperação em diferentes contextos (Saracevic, 1996). Nesse cenário, a CI também tem se voltado à análise dos diferentes suportes informacionais e das tecnologias empregadas para facilitar o acesso e a mediação do conhecimento (Borko, 1968; Saracevic, 1996).

Os audiolivros se destacam por possibilitar o acesso à leitura por meio da escuta, favorecendo diferentes perfis de usuários, inclusive pessoas com deficiência visual ou com dificuldades de leitura (Have & Pedersen, 2019; Rubery, 2016). Sua evolução acompanha os avanços tecnológicos, passando do fonógrafo às plataformas digitais. O audiolivro é um formato editorial distinto dos demais, pois transforma o texto escrito em um produto sonoro. Nesse sentido, seu processo de produção aproxima-se das práticas comuns às indústrias da música e do rádio (Have & Pedersen, 2019).

O audiolivro pode ser acessado em diversos dispositivos, o que o torna uma das formas mais acessíveis de disponibilizar a leitura ao público. Ele pode ser reproduzido em aplicativos para celulares de diferentes sistemas operacionais, em arquivos nos formatos MP3 ou MP4, em tablets, computadores e até em sistemas de GPS veiculares (Schittine, 2022).

Diante da flexibilidade de acesso e do aumento do audiolivro como recurso de leitura e inclusão, torna-se importante discutir como avanços tecnológicos podem potencializar ainda mais esse formato. Nesse contexto, a IA surge oferecendo ferramentas capazes de transformar a experiência de escuta por meio de suas técnicas. O campo da IA começou a ganhar visibilidade a partir da década de 1950, impulsionado pelo avanço dos computadores durante a Segunda Guerra Mundial e pela publicação do artigo "*Computing Machinery and Intelligence*", na revista *Mind*, de autoria de Alan Turing. Nesse trabalho, o pesquisador levantou questionamentos pioneiros sobre a capacidade das máquinas de simular o pensamento humano, estabelecendo um marco fundamental para a evolução da IA (Pinheiro & Oliveira, 2022). A IA é compreendida como uma área da Ciência da Computação (CC) dedicada à criação de sistemas capazes de reproduzir processos cognitivos humanos, como a resolução de problemas, a aprendizagem, a interpretação de informações e o raciocínio lógico (Barr & Feigenbaum, 1981).

Dessa forma, ao compreender os fundamentos da IA, entendida como a área da CC dedicada a simular processos cognitivos humanos (Barr & Feigenbaum, 1981; Pinheiro & Oliveira, 2022), e as particularidades do audiolivro como formato editorial e de mediação da leitura (Have & Pedersen, 2019; Schittine, 2022), torna-se possível estabelecer a conexão entre esses dois campos. A combinação dessas áreas justifica a realização desta pesquisa, pois evidencia o potencial da IA para ampliar o acesso, a personalização e a inclusão em plataformas de audiolivros, contribuindo para novas práticas e reflexões no âmbito da CI.

3. Metodologia

O presente estudo realizou, inicialmente, uma pesquisa bibliográfica e documental entre maio e agosto de 2025, visando embasar os conceitos teóricos e normativos que orientam esta investigação, abordando temas como audiolivros, técnicas de IA, acessibilidade digital, interoperabilidade e recuperação de informação. Em seguida, foi conduzida uma revisão de literatura, com a finalidade de identificar e analisar publicações científicas que tratam da aplicação de técnicas de IA em plataformas de audiolivros, integrando este levantamento ao primeiro objetivo específico de um projeto de doutorado em andamento.

A revisão de literatura foi estruturada a partir de um protocolo previamente definido, que visa garantir a transparência e a reprodutibilidade da pesquisa, adaptado de referências da área, conforme apresentado no Quadro 1, que contempla as seguintes etapas: definição do objetivo geral e das questões de pesquisa; identificação das fontes de informação; estabelecimento dos critérios de elegibilidade e exclusão; elaboração dos campos e das expressões de busca; e descrição dos procedimentos de seleção, extração e análise dos estudos.

Quadro 1
Protocolo de revisão de literatura

Protocolo de revisão de literatura	
Critérios	**Descrição**
Objetivo geral	Identificar estudos sobre o uso da IA em audiolivros
Questões a serem resolvidas	1. Os audiolivros possuem interface com a IA?
	2. Quais técnicas de IA têm sido utilizadas nos audiolivros?
	3. Quais funcionalidades e recursos dos audiolivros utilizam a IA?
Fontes de informação pesquisadas	Bases de dados: *Scielo, Scopus, Web Of Science, ACM Digital Library* e *IEEE Xplore*.
Critérios de elegibilidade	Idioma: inglês, espanhol, português.
	Sem delimitação de data.
	Tipologia documental: artigos de periódicos, trabalhos de eventos com revisão por pares, livros e capítulos de livros.
Critérios de inclusão e de exclusão	**Inclusão**: documentos conforme os critérios de elegibilidade (idioma e tipologia), na área de CI e em áreas afins, conforme objetivo, pergunta e escopo da revisão.
	Exclusão: documentos duplicados, não disponíveis na íntegra (via Portal Capes), documentos cujos títulos e/ou resumos não apresentem relação explícita ou implícita com o uso de IA aplicada a audiolivros, incluindo plataformas digitais de audiolivros, mediação por voz, recuperação de informação ou organização da informação sonora.
Campos de busca	Título, resumo e palavras-chave.
Expressões de busca	Inteligência artificial, Processamento de linguagem natural, Reconhecimento de fala, Aprendizado de máquina, Busca semântica, IA conversacional, Lógica fuzzy, Modelo de linguagem grande, Audiolivros, Plataformas de audiolivros e Livros falados.
	Obs.: as expressões foram utilizadas em português, inglês e espanhol.

***String* geral**	("inteligência artificial" OR "artificial intelligence" OR "inteligencia artificial" OR "processamento de linguagem natural" OR "natural language processing" OR "procesamiento de lenguaje natural" OR "reconhecimento de fala" OR "speech recognition" OR "reconocimiento de voz" OR "aprendizado de máquina" OR "machine learning" OR "aprendizaje automático" OR "busca semântica" OR "semantic search" OR "búsqueda semántica" OR "IA conversacional" OR "conversational AI" OR "IA conversacional" OR "lógica fuzzy" OR "fuzzy logic" OR "lógica difusa" OR "modelo de linguagem grande" OR "large language model" OR "modelo de lenguaje grande") AND ("audiolivros" OR "audiobooks" OR "audiolibros" OR "plataformas de audiolivros" OR "audiobook platforms" OR "plataformas de audiolibros" OR "livros falados" OR "spoken books" OR "libros hablados").
Procedimentos de seleção dos documentos recuperados	Inicialmente, foi realizada a leitura dos títulos dos documentos recuperados com o objetivo de verificar a pertinência do conteúdo em relação ao objetivo geral da pesquisa. Em seguida, procedeu-se à leitura dos resumos, a fim de refinar a seleção dos estudos mais alinhados à temática proposta. Os documentos que atenderam os critérios de relevância seguiram para as etapas de leitura completa e análise qualitativa.
Procedimentos de análise	Foi realizada a leitura completa dos documentos selecionados, com o objetivo de identificar as técnicas de IA utilizadas em audiolivros.
Critério de exclusão após análise dos documentos	Foram excluídos os trabalhos que, após leitura na íntegra, não apresentaram abordagem conceitual, teórica ou metodológica sobre o uso de técnicas de IA em audiolivros.
Tratamento	O *software* Parsifal foi utilizado para eliminar duplicidades e facilitar a triagem dos documentos. As informações extraídas dos estudos selecionados foram sistematizadas em uma planilha eletrônica (Excel).

Fonte: adaptado de Oliveira & Nascimento Silva (2024).

A revisão de literatura considerou estudos publicados em português, inglês e espanhol e indexadas nas bases *Scopus, Web of Science, Scielo, ACM Digital Library* e *IEEE Xplore*. O inglês foi incluído por ser a principal língua de circulação científica internacional, sendo universal. O português foi utilizado para contemplar a literatura nacional e possibilitar a recuperação de estudos produzidos no contexto brasileiro, que dialogam diretamente com a trajetória acadêmica da pesquisadora e com o campo da CI no país. Já o espanhol foi incorporado por representar uma produção científica próxima da realidade latino-americana e por ampliar o acesso a pesquisas desenvolvidas em países do mesmo continente, priorizando o Sul Global, cujos contextos informacionais apresentam pontos de convergência com o cenário brasileiro. Reconhece-se que o uso desses três idiomas constitui uma limitação da pesquisa, uma vez que não contempla estudos publicados em outras línguas. Mas esta revisão foi ampliada dentro das possibilidades linguísticas adotadas, oferecendo um panorama representativo, embora não exaustivo, sobre o tema.

As estratégias de busca foram definidas com base em combinações das expressões Inteligência artificial, Processamento de linguagem natural, Reconhecimento de fala, Aprendizado de máquina, Busca semântica, IA conversacional, Lógica fuzzy, Modelo de linguagem grande, Audiolivros, Plataformas de audiolivros e Livros falados, considerando-se as variações linguísticas em português, inglês e espanhol, o que resultou em uma *string* geral, que foi ajustada conforme a sintaxe de cada base, garantindo a uniformidade da busca em todas as fontes consultadas.[2] Todas as buscas foram realizadas por meio do Portal de Periódicos da CAPES, com acesso institucional via CAFe/UFMG, utilizando o campo de pesquisa avançada, *Advanced Search,* das bases para garantir a reprodutibilidade por demais pesquisadores. Adicionalmente, foi utilizada a ferramenta Parsifal, que permitiu organizar todos os documentos recuperados das bases e conduzir o processo de seleção.

O processo de seleção ocorreu em três etapas, seguindo o processo de análise de conteúdo conforme as etapas propostas por Bardin (2011), compostas por pré-análise, exploração do material e tratamento dos resultados. Na pré-análise, foi realizada a leitura flutuante dos 35 estudos selecionados, com o objetivo de familiarizar-se à temática dos estudos e identificar aspectos recorrentes relacionados ao uso da IA em audiolivros. Nessa etapa inicial, foram anotados elementos preliminares como objetivos centrais, técnicas de IA mencionadas, tipo de aplicação e contexto de uso.

Na etapa de exploração do material, procedeu-se à compilação sistemática dos estudos. Para isso, foi elaborada uma planilha de síntese no Excel, que reuniu informações sobre cada artigo, incluindo autores, ano de publicação, base de indexação, objetivos, métodos empregados, técnicas de IA utilizadas, tipo de dado analisado, finalidade da aplicação e principais achados. A planilha funcionou como instrumento de registro e comparação, permitindo identificar padrões, convergências, diferenças e recorrências entre os estudos. Cada coluna representou um elemento de análise, e a organização dos dados possibilitou observar como as técnicas de IA eram aplicadas e em que contextos se inseriam. Na planilha foi incluído também o DOI, área de conhecimento e observações que a pesquisadora utilizou para apontamentos que achou pertinentes sobre os estudos.

Na etapa de tratamento dos resultados e interpretação, os códigos registrados foram agrupados e analisados à luz do conceito de IA adotado neste estudo, conforme Russell e Norvig (2016), que compreendem a IA como sistemas capazes de executar tarefas associadas à percepção, aprendizado, raciocínio e geração de conteúdo. A partir dessa interpretação, as técnicas e aplicações identificadas foram organizadas em quatro categorias, definidas por sua aderência às dimensões funcionais da IA: ASR, associado à percepção auditiva e decodificação do sinal; TTS, relacionada à geração expressiva de fala; experiências baseadas em voz, que envolvem interação e adaptação comunicativa; e aplicações com IA generativa e modelos de linguagem, que representam formas mais complexas de raciocínio e produção automatizada de conteúdo.

Essas estratégias possibilitou agrupar trabalhos distintos em eixos de investigação convergentes, oferecendo uma visão estruturada sobre como a IA vem sendo aplicada aos audiolivros. Durante a codificação dos estudos, os elementos cognitivos definidos por Russell & Norvig (2016), como percepção, aprendizado, raciocínio e geração, foram utilizados como critérios de leitura e registro na planilha de síntese. Para cada artigo, foram analisados de forma articulada os objetivos, os métodos empregados e as contribuições indicadas pelos autores, identificando se a técnica utilizada buscava reconhecer padrões, gerar fala, produzir conteúdo, apoiar interações vocais ou resolver problemas específicos relacionados ao processamento de audiolivros. Essas informações foram registradas em colunas específicas da planilha, o que permitiu comparar os estudos de acordo com sua finalidade e com o tipo de capacidade cognitiva simulada pela técnica de IA.

Os estudos foram categorizados com base na finalidade das soluções propostas, considerando quatro grandes eixos temáticos: (1) ASR para audiolivros, que se refere ao conjunto de técnicas que convertem sinais de voz em texto. Nos estudos analisados, essa categoria agrupou trabalhos que utilizaram audiolivros como *corpus* para treinar, avaliar ou aprimorar modelos de reconhecimento de fala. Essa categoria evidencia tanto abordagens estatísticas clássicas quanto arquiteturas modernas, refletindo a trajetória evolutiva do ASR; (2) TTS e

personalização, categoria que contempla estudos que utilizam a síntese de voz como eixo central para compreender como a IA tem avançado na geração automática de fala em audiolivros, com foco na transição de métodos tradicionais para modelos neurais, capazes de produzir vozes mais naturais, expressivas e adaptáveis; (3) experiências baseadas em voz, que contempla estudos que exploram a voz como principal meio de interação em sistemas relacionados a audiolivros e conteúdos narrativos digitais, com foco no design de experiências auditivas que utilizam a voz como eixo central de engajamento, imersão e inclusão; e (4) IA generativa e *Large Language Models* (LLMs) em audiolivros. Esta categoria reúne estudos que aplicam modelos generativos de IA e LLMs para o enriquecimento semântico de metadados, a personalização de recomendações, a roteirização de conteúdos e a criação de experiências conversacionais com a capacidade dos modelos generativos de linguagem de transformar a experiência de leitura e escuta em algo mais interativo, imersivo e adaptado ao contexto do usuário. Essa categorização buscou oferecer uma visão panorâmica e organizada do estado da arte sobre o tema.

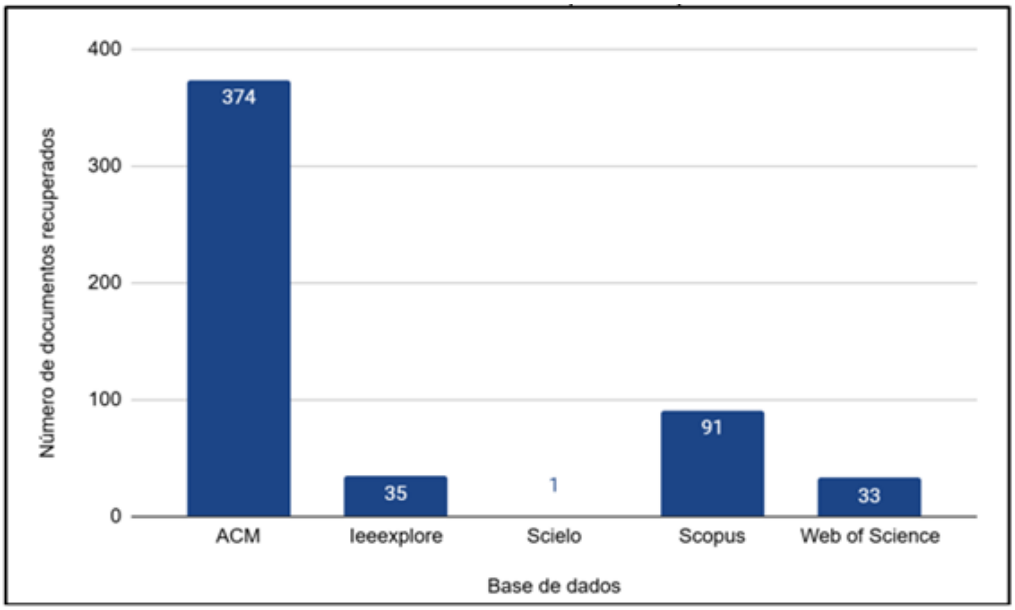
O *software* Parsifal foi utilizado para apoiar a gestão dos dados, eliminando duplicidades, organizando os resultados e facilitando o registro das decisões tomadas durante a seleção. Além disso, planilhas em Excel foram empregadas para registrar as quantidades de estudos recuperados, bem como para apoiar a sistematização e a organização do conteúdo dos artigos selecionados, contribuindo para uma maior clareza e rastreabilidade do processo. A execução da pesquisa teve início em maio e se estendeu até agosto de 2025, com a leitura integral dos artigos selecionados concentrada no mês de julho.

4. Resultados

A busca foi conduzida conforme os critérios previamente definidos no protocolo da revisão. Ao todo, foram recuperados 534 documentos relacionados à temática da pesquisa. A distribuição por base resultou em 33 registros na Web of Science, 91 na Scopus, 374 na ACM Digital Library, 35 na IEEE Xplore e 1 registro na Scielo (Gráfico 1).

Gráfico 1

Total de documentos recuperados por base de dados



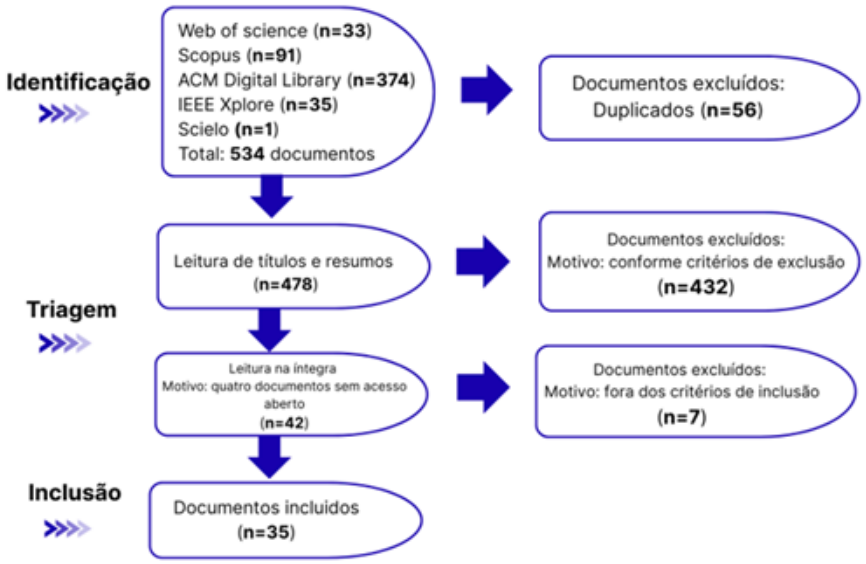
Fonte: dados da pesquisa (2025).

Após a importação dos arquivos no formato BibTeX para o *software* Parsifal, iniciou-se a etapa de identificação de duplicidades. Foram eliminados 56 registros duplicados, restando 478 documentos únicos para a análise inicial de título e resumo. Durante essa triagem, restaram 46 estudos, selecionados para leitura completa por apresentarem potencial relevância ao escopo da revisão, considerando-se sua relação com audiolivros e a aplicação de técnicas de IA.

Na etapa de leitura do texto completo, observou-se que alguns documentos não estavam acessíveis integralmente, pois seu acesso dependia de pagamento, o que reduziu a amostra para 42 documentos com leitura efetivamente realizada. Ao final da análise completa, foram considerados 35 documentos elegíveis para compor a síntese qualitativa da revisão, atendendo aos critérios de inclusão e exclusão estabelecidos no protocolo, conforme apresenta a Figura 1.

Figura 1

Identificação dos estudos nas bases de dados

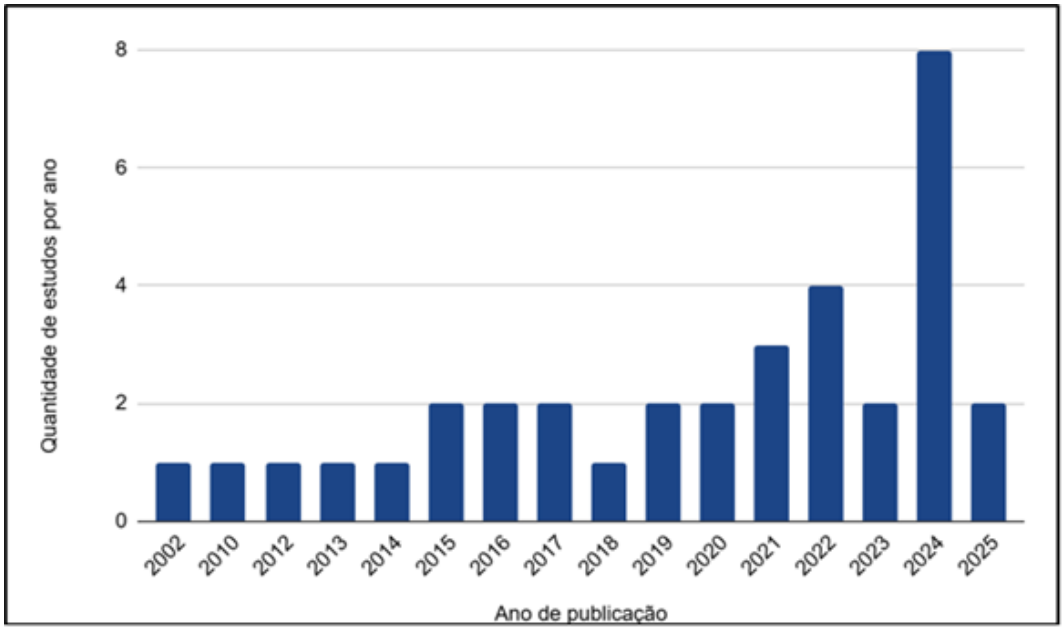


Fonte: adaptado de Page (2021).

A redução do número de estudos ocorreu por diferentes razões. Parte dos trabalhos tratava o audiolivro apenas como exemplo ou recurso pedagógico, sem aprofundar a aplicação de técnicas de IA. Outros artigos se limitavam a converter texto em áudio ou a realizar tarefas isoladas de ASR, sem propor ou analisar soluções específicas para plataformas de audiolivros. Além disso, alguns estudos apresentavam barreiras de acesso ao texto completo, restringindo sua avaliação integral. Os artigos selecionados, juntamente com as categorias e técnicas correspondentes, foram organizados e apresentados de forma sintética no Apéndice A.

A distribuição temporal das publicações evidencia um crescimento gradual pelo uso da IA. Entre 2002 e 2019, observa-se uma produção tímida e esparsa, limitada a uma ou duas publicações anuais. A partir de 2020, contudo, nota-se uma crescente, com destaque para 2021 (3 estudos), 2022 (4 estudos) e, sobretudo, 2024, que concentra oito publicações, configurando-se como o ano de maior produção sobre IA. Esse aumento pode ser associado ao avanço dos modelos de IA, em especial as arquiteturas de *deep learning* e, mais recentemente, os modelos de linguagem em larga escala, que ampliaram significativamente as aplicações em audiolivros. Assim, os dados sugerem que a temática alcançou maior maturidade nos últimos cinco anos, acompanhando o movimento global de incorporação da IA em produtos culturais e educacionais digitais. A distribuição das publicações por ano é apresentada no Gráfico 2.

Gráfico 2
Distribuição das publicações por ano



Fonte: dados da pesquisa (2025).

A distribuição das áreas de conhecimento, conforme classificação indicada nas bases de dados, evidencia a predominância da CC, presente em universidades e centros de pesquisa de diferentes continentes, com destaque para Estados Unidos, China e Índia, que concentram a maior parte das publicações, como apresentado na Figura 2, que mostra o mapa de distribuição geográfica, na qual indica os países de origem das instituições às quais os autores estão vinculados. O objetivo é evidenciar onde estão sendo conduzidas as pesquisas dos estudos recuperados.

Um aspecto relevante é a contribuição de laboratórios corporativos, como *Google Research, Microsoft* e *Spotify*, que reforçam o caráter aplicado das investigações. Em contraste, observa-se a presença pontual de pesquisas vinculadas à CI, à Comunicação e às Artes, indicando movimentos ainda iniciais de interdisciplinaridade. Essa heterogeneidade pode sugerir que, enquanto os avanços metodológicos e técnicos são liderados pela Computação, a interface com campos como Informação e Comunicação permanece emergente, abrindo espaço para a ampliação das discussões sobre a temática nessas áreas.

Figura 2
Distribuição dos estudos por local

Fonte: dados da pesquisa (2025).

Os documentos selecionados na revisão foram numerados, conforme indicado no Apéndice A, e organizados em quatro categorias temáticas, que consideraram os principais enfoques identificados na aplicação de IA em audiolivros, a fim de sintetizar e comparar as contribuições encontradas. As seções a seguir apresentam os resultados de cada categoria, com os respectivos estudos, indicando suas principais abordagens técnicas e metodologias, bem como as soluções propostas.

4.1 *Reconhecimento de fala para audiolivros*

Nesta categoria, foram identificados oito estudos relacionados ao reconhecimento de fala aplicado a audiolivros e tarefas conexas (*corpus* específico de audiolivros, pré-treino para ASR, correção pós-ASR, tradução fala-texto baseada em audiolivros e uso de ASR em interfaces por voz).

Os estudos D02, D15 e D25 concentram-se na construção de modelos e *corpus,*[3] para reconhecimento automático de fala. O estudo D02, de Gebreegziabher e Nürnberger (2020), propõe Redes Neurais Convolucionais (CNNs) leves para ASR, destacando sua eficiência em cenários com restrição computacional. Já o estudo D15, de Panayotov *et al*. (2015), apresenta o *corpus* LibriSpeech, baseado em audiolivros de domínio público, que se consolidou como referência para treinamento e avaliação de sistemas de fala. Em complemento, o estudo D25, de Tóth *et al*. (2010), investiga experimentos de ASR com audiolivros, reforçando a relevância desse domínio para avanços práticos em reconhecimento de voz.

No que diz respeito às interfaces, o estudo D12, de Oumard, Kreimeier & Götzelmann (2022), implementa e avalia uma interface de voz offline voltada para pessoas cegas ou com deficiência visual, evidenciando o potencial do ASR como tecnologia assistiva. O estudo D34, de Serralheiro *et al*. (2002), foca no alinhamento de palavras em livros falados digitais com Transdutores de Estado Finito Ponderados (WFSTs), destacando a integração entre fala e texto como recurso de navegação.

O estudo D18, de Sodhi *et al*. (2021), em relação a estratégias de pós-processamento e tradução, propõe o sistema Mondegreen para a correção de erros recorrentes no ASR, melhorando a legibilidade dos resultados. O estudo D24, de Gibadullin, Perukhin & Ilin (2021), explora redes neurais em larga escala para reconhecimento

de fala e tradução automática, ressaltando o papel do ASR como base para aplicações multilíngues. Por fim, o estudo D27, de Zhang *et al*. (2024), introduz o SpeechLM, um modelo de pré-treinamento que integra fala e texto em um espaço unificado, demonstrando ganhos expressivos em tarefas como ASR e tradução de fala.

É importante observar que os estudos desta categoria utilizam técnicas semelhantes. Os estudos D02, D24 e D25 exploram o uso de arquiteturas de Redes Neurais Recorrentes (RNNs) e suas variações, como as Redes Neurais de Memória de Longo Prazo (LSTM), em conjunto com técnicas de extração de atributos, como os Coeficientes Cepstrais em Frequência Mel (MFCCs), para aprimorar o desempenho do ASR. Já os estudos D15 e D34 reforçam a importância de combinar modelos estatísticos clássicos, como os Modelos Ocultos de Markov (HMMs) e os WFSTs, com Redes neurais profundas (DNNs) e rede neural Perceptron Multicamadas (MLP), configurando abordagens híbridas que marcaram a transição para o aprendizado profundo. O estudo D18, por sua vez, destaca a aplicação de redes neurais artificiais (ANNs) associadas à tradução automática neural (NMT) para o refinamento e a correção de transcrições, enquanto o D27 incorpora técnicas mais recentes, como a Classificação Temporal Conexionista (CTC), sua variação, embasada em Classificação Temporal Conexa Baseada em Unidades (UCTC), a Modelagem de Linguagem Mascarada Baseada em Unidades (UMLM) e os Transdutores de Rede Neural Recorrente (RNN-T), todas integradas em um modelo de pré-treinamento multimodal.

Os estudos dialogam com as reflexões de Jurafsky & Martin (2023), que descrevem a passagem do reconhecimento de fala fundamentado em modelos estatísticos, como os HMMs, para arquiteturas baseadas em aprendizado profundo, como redes neurais convolucionais e recorrentes, consolidando os audiolivros como um *corpus* estratégico para o avanço do campo. De modo convergente, Russell & Norvig (2016) já destacavam que a IA, ao combinar abordagens baseadas em regras e em redes neurais, tem ampliado sua capacidade de lidar com problemas complexos, como o reconhecimento automático de fala, favorecendo aplicações mais precisas e escaláveis. Além disso, os estudos apresentam a evolução das técnicas de ASR e também exemplificam como a IA sustenta inovações voltadas para ambientes digitais.

4.2 *Síntese de voz e personalização*

Nesta categoria, foram identificados 14 estudos que exploram desde "como" transformar audiolivros em vozes sintéticas até formas de controlar expressividade, emoção e contexto, elementos importantes para a personalização em plataformas.

Os estudos D07, D16 e D31 exploram a síntese por seleção de unidades em vozes treinadas com audiolivros. O estudo D07, de Godambe *et al*. (2016), propõe um método para desenvolver vozes a partir de seleção de unidades, mesmo quando não há alinhamento direto entre áudio e texto, ampliando o aproveitamento de gravações para a criação de fala sintética. O estudo D16, de Mamiya *et al*. (2013), propõe uma técnica de detecção de atividade de voz (VAD) levemente supervisionada, aplicada à síntese de voz a partir de audiolivros. O VAD integra uma etapa de pré-processamento no alinhamento de dados de áudio e texto, permitindo detectar automaticamente os limites de frases, em especial os silêncios finais, antes da aplicação da abordagem baseada em grafemas. Já o estudo D31, de Vít & Matoušek (2016), apresenta ajustes no algoritmo de seleção de unidades para lidar melhor com bancos de dados de fala não neutros e imprecisos utilizando vozes derivadas de audiolivros, buscando maior naturalidade.

Os estudos D03, D04 e D32 avançam no uso de audiolivros como *corpus* para o treinamento de TTS neurais. O estudo D03, de Kryvenchuk & Duda (2024), explorou como funcionam os sistemas automatizados de geração de audiolivros, os tipos de métodos de síntese de fala e seus desafios e limitações, apoiados por redes neurais, facilitando a produção em larga escala. O estudo D04, de Sri, Mounika & Yamini (2022), apresenta um protótipo capaz de converter arquivos de texto, imagem ou PDF em áudio, integrando reconhecimento e síntese de fala com foco em acessibilidade, uma vez que o objetivo dessa tecnologia é criar um sistema de reconhecimento de voz para pessoas com deficiência física. Já o estudo D32, de Chalamandaris *et al*. (2014),

investiga o uso de audiolivros como base para treinar modelos de TTS, destacando o potencial desse recurso para melhorar a naturalidade da fala sintética.

Os estudos D08, D13, D23 e D26 demonstram como técnicas de aprendizado profundo permitem capturar traços prosódicos e paralinguísticos. O estudo D08, de Pathak *et al*. (2024), combina a análise de sentimentos e TTS para gerar vozes mais expressivas e conscientes das emoções. O estudo D13, de Aldeneh, Perez & Mower Provost (2021), apresenta o Expressive Voice Conversion Autoencoder (EVoCA), que foca na extração de características paralinguísticas a partir de audiolivros, mostrando como redes neurais podem diferenciar emoções, entonações e estilos de fala. O estudo D23, de Chen, Braunschweiler & Gales (2015), aplica fatoração de voz e expressão em dados de audiolivros, propondo métodos para separar identidade vocal e características expressivas, o que favorece a adaptação e a personalização. Já o estudo D26, de Jin *et al.* (2024), apresenta o *SpeechCraft*, um conjunto de dados bilíngue voltado à síntese de voz expressiva e à personalização vocal, com objetivo de realizar anotação automática da fala para interpretação da expressividade que anota clipes de fala em ambiente real com descrições expressivas e vívidas da linguagem humana. Os áudios de fala são processados por uma série de classificadores especializados e modelos de legenda para capturar diversas características da fala, seguidos por um LLaMA ajustado para a geração de anotações personalizadas.

O estudo D28, de Székely *et al*. (2012), concentra-se na síntese de fala expressiva a partir de gravações amadoras de audiolivros, propondo técnicas de adaptação de voz que permitem um maior realismo, mesmo em *corpus* de menor qualidade acústica.

Por fim, os estudos D11, D33 e D06 ampliam a discussão sobre imersão e modulação adaptativa. O estudo D11, de Jennes, Blanckaert & Van den Broeck (2023), avalia a percepção de leitores quanto ao uso de áudio 3D em práticas digitais de leitura, ressaltando como a espacialização sonora contribui para uma maior imersão. O estudo D33, de Subramanian *et al*. (2024), aborda a modulação de voz em narrações de audiolivros, com o objetivo de fornecer *insights* práticos para otimizar a narração, levando a experiências de audiolivros dinâmicas, emocionalmente ressonantes e profundamente satisfatórias para os ouvintes. Já o estudo D06, de Xiao *et al*. (2024), apresenta uma estrutura de Pré-treinamento de Fala Contextual Contrastiva (CCSP) que busca integrar informações do texto contextual e das características da fala, permitindo aprender representações multimodais mais ricas.

Observa-se que os estudos transitam de métodos estatísticos para arquiteturas de *deep learning*, com destaque para CNNs, RNNs e Transformers aplicados à síntese de fala. Embora os estudos revelem avanços em personalização da voz por meio de técnicas de *deep learning*, observa-se que, nas plataformas de audiolivros, tais recursos ainda se manifestam de forma tímida. Gonçalves & Silva (2025) destacam que as plataformas de audiolivros possuem funcionalidades básicas, como retomar, pular ou retornar capítulos, ajustar a velocidade de reprodução, ativar o *sleep timer* e acessar o sumário do audiolivro, em geral restritas ao ambiente móvel, e que os recursos de personalização concentram-se em organizar e gerenciar audiolivros adquiridos ou emprestados, por meio de avaliações, resenhas, listas de desejados, favoritos, escuta *offline* e metas de leitura. Esses resultados sugerem que, enquanto a literatura acadêmica explora a expressividade e a adaptação das vozes sintéticas, as plataformas mantêm uma implementação mais limitada de personalização, centrada em aspectos de navegação e organização do conteúdo.

Pathak *et al*. (2024) explicam que a síntese de voz consiste na transformação de textos escritos em fala natural, viabilizando a criação de narrações sintéticas com qualidadepróxima à humana e com possibilidade de ajustes ao estilo desejado. Essa capacidade técnica amplia o potencial de uso da síntese de voz em audiolivros. Nesse sentido, Chalamandaris *et al*. (2014) complementam essa discussão ao destacar que tais recursos favorecem experiências de audição mais imersivas e personalizadas, capazes de ampliar a sensação de envolvimento do ouvinte e tornar os audiolivros mais acessíveis e atrativos para públicos diversos.

4.3 *Experiências baseadas em voz*

Nesta categoria, foram identificados nove estudos que exploram experiências baseadas em voz aplicando técnicas de IA para interação, engajamento e aprendizagem mediada por áudio.

Os estudos D01, D29 e D35 exploram a voz para a acessibilidade e a mediação social: o estudo D01, de Desai, Lundy & Chin (2023), propõe o uso de uma interface de Voz Interativa e Baseada em Narrativas (VUI) chamada "*Mystery Agent*" para engajar adultos mais velhos (60+ anos) no aprendizado informal de informações sobre saúde; o estudo D29, de Cruz *et al*. (2020), integra tecnologias como Conversão de Voz em Texto (*Speech-to-Text*) e reconhecimento de fala, que atendem especificamente pessoas com deficiência visual e outras deficiências físicas, utilizando dispositivos como smartphones, tablets e computadores capazes de emitir *feedbacks* de voz, ler textos por meio de um alto-falante ou fones de ouvido e receber a voz do usuário por meio de um microfone; e o estudo D35, de Brewer & Piper (2017), apresenta o *xPress*, uma comunidade de blog que destaca a importância da voz humana em plataformas sociais acessíveis para idosos com perda de visão.

Os estudos D19 e D30 avançam na disponibilização de conteúdo por voz: o estudo D19, de Laban *et al*. (2022), apresenta o sistema *NewsPod*, projetado para gerar podcasts de notícias automaticamente, utilizando técnicas de PLN e TTS, o que permite que o consumo de notícias seja mais acessível, dinâmico e interativo, possibilitando que os ouvintes façam perguntas durante a reprodução e recebam respostas automáticas; já o estudo D30, de Yang *et al*. (2018), compara recomendações via TTS com interfaces textuais, avaliando efeitos no engajamento. Em relação à leitura e à educação, o estudo D14, de Zhao & McEwen (2022), descreve um robô leitor que combina comandos de voz ASR, leitura em voz alta TTS e *chatbot* para sustentar rotinas de leitura compartilhada, e o estudo D09, de Bertulfo *et al*. (2017), propõe um audiolivro 3D com reconhecimento de voz (*Google Speech API*) para apoiar alunos cegos ou com baixa visão.

O estudo D10, de Park & Tsuruoka (2019), por sua vez, amplia a dimensão imersiva da leitura ao criar *bookscapes* em que som, música e paisagens sonoras são gerados dinamicamente a partir de aprendizado de máquina, ASR e música computadorizada, permitindo que o ambiente sonoro reaja em tempo real à voz do usuário. Por fim, o estudo D22, de Cotton, Vries & Tatar (2024), concentra-se em instigar questionamentos sobre o uso das tecnologias de clonagem e síntese neural de voz, evidenciando implicações éticas e de corporeidade nas experiências de escuta.

Bertulfo *et al*. (2017) destacam que as experiências baseadas em voz combinam recursos de reconhecimento de fala, síntese de voz e interatividade para criar ambientes imersivos e acessíveis, especialmente quando aplicadas a contextos educacionais e de entretenimento. Os estudos analisados nesta categoria mostra que a interação por voz pode ir muito além de comandos simples: é possível criar narrativas adaptativas, incorporar elementos multimodais, como áudio 3D, integrar assistentes virtuais e até utilizar recursos expressivos e emocionais para aumentar o engajamento. Essa evolução no design de experiências auditivas não apenas melhora a usabilidade e a imersão, mas também fortalece a inclusão digital, permitindo que diferentes públicos tenham acesso facilitado e mais apreciável ao conteúdo. Para Porcheron *et al*. (2018), a voz em interfaces digitais carrega dimensões sociais e expressivas que ampliam o engajamento, reforçando o potencial inclusivo e multimodal das experiências baseadas em voz.

4.4 *IA generativa e LLMs em audiolivros*

Nesta categoria, foram identificados quatro estudos que utilizam técnicas de IA generativa e LLMs aplicados a audiolivros.

Os estudos D05, D17, D20 e D21 concentram-se no uso de IA generativa e LLMs para enriquecer experiências com audiolivros e conteúdos relacionados. O estudo A05, de Penha *et al*. (2025), propõe o uso de LLMs para enriquecer metadados e organizar listas contextuais de audiolivros em "prateleiras descritivas", ampliando a descoberta e a diversidade de recomendações na plataforma *Spotify*. Já o estudo D17, de Choi (2019), apresenta o protótipo LYRA, que combina modelos de conversação neural baseados em modelo de

sequência para sequência (seq2seq), ASR e TTS para criar narrativas interativas, em que crianças podem interagir com a história e alterar seus desfechos, evidenciando o potencial educativo da IA generativa.

Em complemento, o estudo D20, de Yahagi *et al.* (2025), desenvolve o sistema PaperWave, que transforma artigos científicos em *podcasts* conversacionais com roteiros gerados por LLMs e convertidos por TTS, ressaltando a viabilidade de adaptar conteúdos textuais complexos para formatos auditivos acessíveis e engajadores. Por fim, o estudo D21, de Nadai *et al.* (2024), aborda a recomendação personalizada de audiolivros por meio de redes neurais de grafos heterogêneos (HGNNs) integrados a um modelo Two-Tower (2T), combinando representações aprendidas de usuários e itens com LLM *embeddings* de descrições, resultando em ganhos expressivos em taxa de descoberta e consumo de novos audiolivros.

Observa-se que esses trabalhos exploram a capacidade dos LLMs e das técnicas de IA generativa de enriquecer metadados, criar experiências de narrativa interativa e aprimorar a personalização em sistemas de recomendação. Além de ampliarem a usabilidade e acessibilidade dos audiolivros, os estudos reforçam o papel estratégico da IA na transformação de conteúdos textuais em experiências multimodais, ao mesmo tempo em que indicam novos caminhos de pesquisa sobre a integração entre modelagem de linguagem, recomendação personalizada e design de interação no ecossistema digital de audiolivros (Brown *et al.*, 2020; Zhao & McEwen, 2022).

A contribuição dos estudos está em demonstrar os audiolivros como um campo estratégico de aplicação da IA, seja pela criação de *corpus* e ferramentas de base (ASR e TTS), seja pela proposição de novas experiências de leitura e escuta mediadas por voz, acessibilidade e personalização. A revisão evidencia não apenas avanços técnicos, mas também preocupações éticas, como no uso de síntese neural e clonagem de voz, além da valorização do papel social dos audiolivros em relação à inclusão digital. Além disso, é possível observar que, entre as categorias, há uma recorrência de técnicas, embora com ênfases diferentes, conforme o objetivo.

4.5 *Síntese da revisão*

Os estudos selecionados para a revisão foram categorizados conforme apresentado no Quadro 3.

Quadro 3

Resumo dos artigos por categoria

Categoria	Trabalhos (ID)
(1) Reconhecimento de Fala (ASR) para audiolivros	D02, D12, D15, D18, D24, D25, D27, D34
(2) Síntese de Voz (TTS) e personalização	D03, D04, D06, D07, D08, D11, D13, D16, D23, D26, D28, D31, D32, D33
(3) Experiências Baseadas em Voz	D01,D09, D10, D14, D19, D22, D29,D30, D35
(4) IA generativa e LLMs em audiolivros	D05, D17, D20, D21

Fonte: dados da pesquisa (2025).

Cabe destacar que, durante a análise, alguns estudos apresentaram características que poderiam enquadrá-los em mais de uma das categorias definidas. Nesses casos, optou-se por alocá-los na categoria cuja relação temática e metodológica se mostrou mais predominante. Outro aspecto importante é que determinadas técnicas de IA, como ASR, TTS, Redes Neurais Recorrentes (RNN e LSTM) e arquiteturas baseadas em Transformers, aparecem de forma recorrente em diferentes estudos e categorias, algumas delas constituindo, inclusive, a

principal abordagem da maioria dos trabalhos. Essa recorrência evidencia tanto a centralidade dessas técnicas no campo quanto sua versatilidade em múltiplas aplicações relacionadas aos audiolivros.

Conclui-se que os 35 estudos analisados responderam às perguntas da revisão. Em relação à primeira pergunta: Os audiolivros possuem interface com a IA?, os estudos analisados confirmam que, embora em diferentes estágios de maturidade, os audiolivros incorporam uma interface capaz de utilizar múltiplas técnicas de IA em seus sistemas.

Em relação à segunda pergunta: Quais técnicas de IA têm sido utilizadas nos audiolivros?, foram identificados desde métodos estatísticos, como HMMs, GMMs e WFSTs, até arquiteturas avançadas de aprendizado profundo, incluindo CNNs, RNNs, LSTMs, Transformers, além de modelos de LLMs e técnicas de pré-treinamento auto-supervisionado.

Já em relação à terceira pergunta: Quais funcionalidades e recursos dos audiolivros utilizam a IA?, os artigos apontaram funcionalidades como ASR, TTS, personalização por modulação expressiva, experiências multimodais baseadas em voz e recomendação personalizada mediada por LLMs e grafos.

Verificou-se que, embora ainda em estágios variados de maturidade, os audiolivros são utilizados como *corpus* para múltiplas técnicas de IA, desde o ASR e o TTS tradicionais até modelos generativos mais recentes, ampliando funcionalidades que vão da acessibilidade à recomendação personalizada. Essa constatação confirma a relevância do tema e aponta para a necessidade de investigações futuras que aprofundem tanto a eficiência técnica quanto as implicações sociais e éticas do uso da IA em audiolivros.

Considerações finais

Este estudo teve como objetivo identificar estudos sobre o uso da IA em audiolivros. Para isso, foi conduzida uma revisão de literatura com base em um protocolo de seleção, que resultou na análise de 35 artigos, recuperados nas bases *Scopus, Web of Science, ACM Digital Library, IEEE Xplore* e *Scielo*, entre maio e agosto de 2025.

Os estudos foram organizados em quatro categorias: (1) ASR em audiolivros; (2) TTS e personalização; (3) Experiências baseadas em voz; e (4) IA generativa e LLMs em audiolivros. Essas categorias evidenciaram como a IA tem sido aplicada de forma transversal, apoiando tanto o desenvolvimento técnico de modelos de fala e síntese quanto a criação de experiências de leitura mais acessíveis, expressivas e imersivas. As perguntas que nortearam esta revisão foram contempladas pelos 35 estudos avaliados, e os artigos analisados revelam um campo em expansão, que conjuga preocupações técnicas (como qualidade da fala sintética e robustez dos modelos) com aspectos sociais e éticos, como acessibilidade para pessoas com deficiência visual, impacto da clonagem de voz e valorização da experiência do usuário em ambientes digitais.

Apesar dos avanços da IA, o estudo identificou a limitação de que, embora os audiolivros sejam amplamente usados no campo técnico (ASR, TTS e LLMs), voltado a melhorar modelos de reconhecimento ou síntese de fala, o audiolivro funciona quase como um "sub objeto" ou "meio" para treinar IA, e não como fim em si. Poucos estudos refletem sobre a experiência real do usuário: como a escuta pode se tornar mais fluida, inclusiva, envolvente, personalizada, e até mesmo utilizar formas de busca e recuperação acessível.

Como contribuição, este estudo oferece um panorama da literatura, destacando o papel estratégico da IA no fortalecimento dos audiolivros como recurso inclusivo, personalizado e imersivo, identificando avanços e lacunas importantes, como o predomínio de estudos técnicos, poucas análises centradas no usuário, o uso dos audiolivros, sobretudo como *corpus*, e não como objeto de melhoria, a escassez de aspectos éticos e sociais e a presença de poucos trabalhos na área da CI. Tais temáticas sinalizam oportunidades para investigações futuras.

Além disso, a revisão permitiu identificar que os audiolivros, ao incorporarem a IA, não apenas ampliam o acesso à leitura, mas também se configuram como um campo de inovação tecnológica com grande contribuição social, por meio da ampliação do acesso e da própria acessibilidade.

Agradecimento

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio à pesquisa, processo 303721/2025-1.

Roles de colaboración

Suellen Souza Gonçalves: Conceitualização – Curadoria dos dados - Redação - Revisão e edição.
Patrícia Nascimento Silva: Administração do Projeto - Conceitualização - Redação - Revisão e edição.

Apêndice A Documentos selecionados e suas categorias

D01
Título: "A Painless Way to Learn": Designing an Interactive Storytelling Voice User Interface to Engage Older Adults in Informal Health Information Learning
Autores: Desai, Lundy & Chin (2023)
Categoria: Experiências baseadas em voz
Técnicas: Processamento de Linguagem Natural (PLN); Reconhecimento Automático de Fala (ASR); Síntese de Voz (TTS)
D02
Título: A Light-weight Convolutional Neural Network based Speech Recognition for Spoken Content Retrieval Task
Autores: Gebreegziabher & Nürnberg (2020)
Categoria: Reconhecimento de Fala para audiolivros
Técnicas: Redes Neurais Convolucionais (CNNs); Redes Neurais Recorrentes (RNNs); Redes Neurais Profundas (DNNs); Classificação Temporal Conexa (CTC); Coeficientes Cepstrais em Frequência Mel (MFCCs)
D03
Título: Audio Book Creation System Using Artificial Intelligence
Autores: Kryvenchuk & Duda (2024)
Categoria: Síntese de Voz e personalização
Técnicas: Rede Neural Recorrente com Memória de Longo e Curto Prazo (LSTM); TTS; Grandes Modelos de Linguagem (LLMs)
D04
Título: Audiobooks that converts Text, Image, PDF-Audio & Speech-Text: for physically challenged & improving fluency
Autores: Sri, Mounika & Yamini (2022)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; ASR; Compreensão da Linguagem Natural (NLU); PLN; Biometria de voz (Voice Biometrics); Gerenciamento de Diálogo (Dialog Management); Reconhecimento de Entidades Nomeadas (NER)
D05
Título: Contextualizing Spotify's Audiobook List Recommendations with Descriptive Shelves
Autores: Penha *et al*. (2025)
Categoria: IA Generativa e LLMs em audiolivros
Técnicas: LLMs
D06

Título: Contrastive Context-Speech Pretraining for Expressive Text-to-Speech Synthesis
Autores: Xiao *et al*. (2024)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; ASR
D07
Título: Developing a unit selection voice given audio without corresponding text
Autores: Godambe *et al*. (2016)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; DNN; ASR; MFCCs; Modelos Ocultos de Markov (HMM); Modelo de Mistura Gaussiana (GMM)
D08
Título: Emotion-Aware Text to Speech: Bridging Sentiment Analysis and Voice Synthesis
Autores: Pathak et al. (2024)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; ASR; HMM; LSTM; Rede Neural de Memória de Longo e Curto Prazo Bidirecional (BLSTM); PLN
D09
Título: Gabay tinig: a 3D interactive audiobook with voice recognition for visually-impaired and blind preschool students using mobile technologies
Autores: Bertulfo et al. (2017)
Categoria: Experiências baseadas em voz
Técnicas: ASR; TTS; PLN
D10
Título: Generative bookscapes: Towards immersive and interactive book reading
Autores: Park & Tsuruoka (2019)
Categoria: Experiências baseadas em voz
Técnicas: CNNs
D11
Título: Immersion or Disruption? Readers' Evaluation of and Requirements for (3D-) audio as a Tool to Support Immersion in Digital Reading Practices
Autores: Jennes, Blanckaert & Van den Broeck (2023)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; MFCCs
D12
Título: Implementation and Evaluation of a Voice User Interface with Offline Speech Processing for People who are Blind or Visually Impaired
Autores: Oumard, Kreimeier & Götzelmann (2022)
Categoria: Reconhecimento de Fala (ASR) para audiolivros
Técnicas: ASR; NLU; Gerenciamento de Diálogo (DM)
D13
Título: Learning Paralinguistic Features from Audiobooks through Style Voice Conversion
Autores: Aldeneh, Perez & Provost (2021)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; MFCsC; LSTM; BLSTM; ASR
D14
Título: "Let's read a book together": A Long-term Study on the Usage of Pre-school Children with Their Home Companion Robot

Autores: Zhao & McEwen (2022)
Categoria: Experiências baseadas em voz
Técnicas: ASR; PLN; TTS
D15
Título: Librispeech: An ASR Corpus Based on Public Domain Audio Books
Autores: Panayotov et al. (2015)
Categoria: Reconhecimento de Fala para audiolivros
Técnicas: DNN; GMM; HMM
D16
Título: Lightly supervised GMM VAD to use audiobook for speech synthesiser
Autores: Mamiya *et al*. (2013)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; HMM; Detecção de atividade de voz (VAD); GMM
D17
Título: LYRA: An Interactive Storyteller
Autores: Choi (2019)
Categoria: IA Generativa e LLMs em audiolivros
Técnicas: Modelo Neural de Conversação (Seq2Seq); TTS
D18
Título: Mondegreen: A Post-Processing Solution to Speech Recognition Error Correction for Voice Search Queries
Autores: Sodhi et al. (2021)
Categoria: Reconhecimento de Fala para audiolivros
Técnicas: Tradução Automática Neural (NMT); Redes Neurais Artificiais (ANNs)
D19
Título: NewsPod: Automatic and Interactive News Podcasts
Autores: Laban *et al*. (2022)
Categoria: Experiências baseadas em voz
Técnicas: PLN; TTS
D20
Título: PaperWave: Listening to Research Papers as Conversational Podcasts Scripted by LLM
Autores: Yahagi *et al*. (2025)
Categoria: IA Generativa e LLMs em audiolivros
Técnicas: LLMs; TTS
D21
Título: Personalized Audiobook Recommendations at Spotify Through Graph Neural Networks
Autores: Nadai *et al.* (2024)
Categoria: IA Generativa e LLMs em audiolivros
Técnicas: LLMs; Redes Neurais em Grafos Heterogêneos (HGNNs)
D22
Título: Singing for the Missing: Bringing the Body Back to AI Voice and Speech Technologies
Autores: Cotton, Vries & Tatar (2024)
Categoria: Experiências baseadas em voz
Técnicas: DL; ASR; TTS; CTC; RNN; DNN
D23
Título: Speaker and Expression Factorization for Audiobook Data: Expressiveness and Transplantation
Autores: Chen, Braunschweiler & Gales (2015)

Categoria: Síntese de Voz e personalização
Técnicas: TTS; LSM; ASR; Rede Neural Perceptron Multicamadas (MLP)
D24
Título: Speech Recognition and Machine Translation Using Neural Networks
Autores: Gibadullin, Perukhin & Ilin (2021)
Categoria: Reconhecimento de Fala para audiolivros
Técnicas: LSTM; RNN; MFCCs
D25
Título: Speech Recognition Experiments with Audiobooks
Autores: Tóth *et al*. (2010)
Categoria: Reconhecimento de Fala para audiolivros
Técnicas: HMM; MFCC; GMM; ANN; Transdutores de Estado Finito Ponderados (WFST)
D26
Título: SpeechCraft: A Fine-Grained Expressive Speech Dataset with Natural Language Description
Autores: Jin *et al*. (2024)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; LLM; Grande Porte da Meta (LLaMA)
D27
Título: SpeechLM: Enhanced Speech Pre-Training With Unpaired Textual Data
Autores: Zhang *et a*l. (2024)
Categoria: Reconhecimento de Fala para audiolivros
Técnicas: CTC; Classificação Temporal Conexa Baseada em Unidades (UCTC); Modelagem de Linguagem Mascarada Baseada em Unidades (UMLM); Transdutor de Rede Neural Recorrente (RNN-T)
D28
Título: Synthesizing expressive speech from amateur audiobook recordings
Autores: Székely *et al.* (2012)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; HMM; WFST; ASR
D29
Título: Talkie: An Assistive Web-based Educational Application Using Audio Files and Speech Technology for the Visually Impaired
Autores: Cruz *et al*. (2020)
Categoria: Experiências baseadas em voz
Técnicas: ASR; TTS
D30
Título: Understanding user interactions with podcast recommendations delivered via voice
Autores: Yang *et al.* (2018)
Categoria: Experiências baseadas em voz
Técnicas: TTS
D31
Título: Unit-selection speech synthesis adjustments for audiobook-based voices
Autores: Vít & Matoušek (2016)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; HMM; MFCC; Deep Learning (DL)
D32
Título: Using audio books for training a text-to-speech system
Autores: Chalamandaris *et al.* (2014)

Categoria: Síntese de Voz e personalização
Técnicas: TTS; HMM
D33
Título: Voice Modulation in Audiobook Narration
Autores: Subramanian *et al.* (2024)
Categoria: Síntese de Voz e personalização
Técnicas: TTS; LLM; LSTM; LLaMA; Máquina de vetor de suporte (SVM); CNN; RNN
D34
Título: Word alignment in digital talking books using WFSTs
Autores: Serralheiro *et al*. (2002)
Categoria: Reconhecimento de Fala para audiolivros
Técnicas: MLP; HMM; WFST
D35
Título: xPress: Rethinking Design for Aging and Accessibility through an IVR Blogging System
Autores: Brewer & Piper (2017)
Categoria: Experiências baseadas em voz
Técnicas: TTS; Sistema de Resposta Audível Interativa (IVR)

Referências

Aldeneh, Z., Perez, M. & Mower Provost, E. (2021). Learning paralinguistic features from audiobooks through style voice conversion. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. https://aclanthology.org/2021.naacl-main.377/

Bardin, L. (2011). *Análise de conteúdo*. Edições 70.

Barr, A. & Feigenbaum, E. A. (1981). *The handbook of artificial intelligence*. William Kaufmann Inc.

Bertulfo, L. C., Razon, J. M., San Juan, J. L., Sambrano, J. C. & Medina, R. P. (2017). Gabay Tinig: A 3D interactive audiobook with voice recognition for visually-impaired and blind preschool students using mobile technologies. *Proceedings of the 3rd International Conference on Communication and Information Processing (ICCIP '17)*. https://dl.acm.org/doi/10.1145/3162957.3162979

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge University Press.

Borko, H. (1968). Information science: What is it? *American documentation, 19*(1). https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090190103

Brewer, R. N. & Piper, A. M. (2017). XPress: Rethinking design for aging and accessibility through an IVR blogging system. *Proceedings of the ACM on Human-Computer Interaction, 1*(CSCW). https://dl.acm.org/doi/10.1145/3139354

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. L*Advances in neural information processing systems (NeurIPS 2020)*. https://arxiv.org/abs/2005.14165

Chalamandaris, A., Raptis, S., Karabetsos, S. & Tsiakoulis, P. (2014). Using audio books for training a text-to-speech system. En N. Calzolari et al. (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA). https://aclanthology.org/L14-1645/

Chen, L., Braunschweiler, N. & Gales, M. J. F. (2015). Speaker and expression factorization for audiobook data: Expressiveness and transplantation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23*(4). https://ieeexplore.ieee.org/document/6995936

Chen, X. *et al.* (2023). StyleSpeech: self-supervised style enhancing with VQ-VAE-based pre-training for expressive audiobook speech synthesis. *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.48550/arXiv.2312.12181

Choi, D. H. (2019). LYRA: an interactive and interactive storyteller. *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE. https://ieeexplore.ieee.org/document/8919562

Cotton, K., de Vries, K. & Tatar, K. (2024). Singing for the missing: bringing the body back to AI voice and speech technologies. *Proceedings of the 9th International Conference on Movement and Computing*. ACM. https://dl.acm.org/doi/10.1145/3658852.3659065

Cruz, J. R. D., Villanueva, C. M., Abarro, J. A. & Villanueva, J. L. (2020). Talkie: an assistive web-based educational application using audio files and speech technology for the visually impaired. *Proceedings of the 2020 The 6th International Conference on Frontiers of Educational Technologies*. ACM. https://dl.acm.org/doi/10.1145/3404709.3404748

Desai, S., Lundy, M. & Chin, J. (2023). "A painless way to learn": designing an interactive storytelling voice user interface to engage older adults in informal health information learning. *CUI '23: Proceedings of the 5th International Conference on Conversational User Interfaces* (No. 5). ACM. https://doi.org/10.1145/3571884.3597141

Gebreegziabher, N. H. & Nürnberger, A. (2020). A light-weight convolutional neural network based speech recognition for spoken content retrieval task. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, *Toronto, ON, Canada*. IEEE. https://doi.org/10.1109/SMC42975.2020.9282956

Gibadullin, R. F., Perukhin, M. Y. & Llin, A. V. (2021). Speech recognition and machine translation using neural networks. *International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, *Sochi, Russia*. IEEE. https://doi.org/10.1109/ICIEAM51226.2021.9446474

Godambe, T., Singh, R., Sitaram, S. & Choudhury, M. (2016). Developing a unit selection voice given audio without corresponding text. *EURASIP Journal on audio, speech, and music processing,* (6). https://doi.org/10.1186/s13636-016-0084-y

Gonçalves, S. S. & Silva, P. N. (2025). Requisitos funcionais para recuperação de informação em audiolivros: uma análise nas plataformas. *Informação & informação, 30*(1), 354-372. https://doi.org/10.5433/1981-8920.2025v30n1p354

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org/

Have, I. & Pedersen, B. (2019). The audiobook circuit in digital publishing: voicing the silent revolution. *New media & society, 22*(3), 409-428. https://doi.org/10.1177/1461444819863407

Jennes, I., Blanckaert, E. & Van den Broeck, W. (2023). Immersion or disruption: Readers' evaluation of and requirements for (3D-) audio as a tool to support immersion in digital reading practices. *IMX '23: Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*. ACM. https://dl.acm.org/doi/10.1145/3573381.3596151

Jin, Z. *et al*. (2024). SpeechCraft: a fine-grained expressive speech dataset with natural language description. *Proceedings of the 32nd ACM International Conference on Multimedia*. ACM. https://dl.acm.org/doi/10.1145/3664647.3681674

Jurafsky, D. & Martin, J. H. (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall.

Kryvenchuk, Y. & Duda, O. (2024). Audio book creation system using artificial intelligence. *IEEE 19th International Conference on Computer Science and Information Technologies (CSIT)*, *Lviv, Ukraine*. IEEE. https://doi.org/10.1109/CSIT65290.2024.10982688

Laban, P., Dusek, O., Sharma, R. & Rieser, V. (2022). NewsPod: automatic and interactive news podcasts. *27th International Conference on Intelligent User Interfaces*. ACM. https://dl.acm.org/doi/10.1145/3490099.3511147

Lourenço, C. de A. (2005). *Modelagem de dados como ferramenta de análise de padrões de metadados em bibliotecas digitais: O padrão de metadados brasileiro para teses e dissertações segundo o modelo entidade-relacionamento* (Tese de doutorado). Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte. https://repositorio.ufmg.br/handle/1843/EARM-6ZGNZC

Mamiya, Y., *et al.* (2013). Lightly supervised GMM VAD to use audiobook for speech synthesiser. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. https://ieeexplore.ieee.org/document/6639220

McEnery, T. & Hardie, A. (2012). *Corpus linguistics*. Cambridge University Press.

Nadai, M. de, Silva, T., Faria, R. & Ribeiro, L. (2024). Personalized audiobook recommendations at Spotify through graph neural networks. *WWW '24: Companion Proceedings of the ACM Web Conference*. ACM. https://dl.acm.org/doi/10.1145/3589335.3648339

Oliveira, D. T. De & Nascimento Silva, P. (2024). Representação e recuperação de dados governamentais abertos: uma revisão de literatura. *RDBCI: Revista digital de biblioteconomia e ciência da informação,* (22)*,* e024029. https://doi.org/10.20396/rdbci.v22i00.8675828

Oumard, C., Kreimeier, J. & Götzelmann, T. (2022). Implementation and evaluation of a voice user interface with offline speech processing for people who are blind or visually impaired. *PETRA '22: Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments*. ACM. https://dl.acm.org/doi/10.1145/3529190.3529197

Page, M. J. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMG, 372*(71). https://doi.org/10.1136/bmj.n71

Panayotov, V., Chen, G., Povey, D. & Khudanpur, S. (2015). Librispeech: An ASR corpus based on public domain audio books. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia. IEEE. https://doi.org/10.1109/ICASSP.2015.7178964

Park, T. H. & Tsuruoka, T. (2019). Generative bookscapes: towards immersive and interactive book reading. *International Computer Music Conference, New York City Electroacoustic Music Festival*. Fulcrum. https://www.fulcrum.org/epubs/9880vt18d?locale=en#page=3

Pathak, A., Sharma, V., Singh, R. & Choudhury, M. (2024). Emotion-aware text to speech: bridging sentiment analysis and voice synthesis. *3rd International Conference for Innovation in Technology (INOCON), Bangalore, India*. IEEE. https://doi.org/10.1109/INOCON60754.2024.10512224

Penha, G., Santos, L., Almeida, F. & Hauff, C. (2025). Contextualizing Spotify's audiobook list recommendations with descriptive shelves. En C. Hauff *et al.* (Eds.), *Advances in information retrieval* (Lecture notes in computer science, 15576). Springer. https://doi.org/10.1007/978-3-031-88720-8_26

Pinheiro, M. & Oliveira, H. (2022). Inteligência artificial: estudos e usos na ciência da informação no Brasil. *Revista ibero-americana de ciência da informação, 15*(3), 950-968. https://doi.org/10.26512/rici.v15.n3.2022.42767

Porcheron, M., Fischer, J. E., Reeves, S. & Sharples, S. (2018). Voice interfaces in everyday life. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Paper 640). ACM. https://dl.acm.org/doi/10.1145/3173574.3174214

Rubery, M. (2016). *The untold story of the talking book*. Harvard University Press.

Russell, S. & Norvig, P. (2016). *Inteligência artificial: uma abordagem moderna*. GEN LTC.

Saracevic, T. (1996). Ciência da informação: origem, evolução e relações. *Perspectivas em ciência da informação, 1*(1). http://hdl.handle.net/20.500.11959/brapci/37415

Serralheiro, A., Ferreira, C. & Costa, P. (2002). Word alignment in digital talking books using WFSTs. En M. Agosti & C. Thanos (Eds.), *Research and advanced technology for digital libraries*. Springer. https://link.springer.com/chapter/10.1007/3-540-45747-X_37

Schittine, D. (2022). Audiolivros: Desafios de produção, voz do narrador e público-leitor. *Scripta, 26*(56), 256-269. https://doi.org/10.5752/P.2358-3428.2022v26n56p256-269

Silva, M. B. da & Neves, D. A. de B. (2013). A aplicação da teoria facetada em banco de dados, através da modelagem conceitual. En M. E. B. C. de Albuquerque, L. S. M. A. da Silva, & R. C. C. de Araújo (Eds.), *Representação da informação: um universo multifacetado*. Editora da UFPB. https://doi.org/10.22477/vii.widat.206

Sodhi, S. S., Singh, A., Ghosh, S. & Shrivastava, M. (2021). Mondegreen: A post-processing solution to speech recognition error correction for voice search queries. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. ACM. https://dl.acm.org/doi/10.1145/3447548.3467156

Souza Gonçalves, S. & Nascimento Silva, P. (2026). Dados da pesquisa: inteligência artificial em audiolivros: aplicações e perspectivas. Mendeley data, V2, https://doi.org/10.17632/vtb3ygt62k.3

Sri, K. S., Mounika, C. & Yamini, K. (2022). Audiobooks that converts text, image, PDF-audio & speech-text: For physically challenged & improving fluency. *International Conference on Inventive Computation Technologies (ICICT), Nepal*. IEEE. https://doi.org/10.1109/ICICT54344.2022.9850872

Subramanian, V., Patel, A., Rao, P. & Kumar, S. (2024). Voice modulation in audiobook narration. *2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. IEEE. https://ieeexplore.ieee.org/document/10851662

Székely, É., O'Connor, N. & Gobl, C. (2012). Synthesizing expressive speech from amateur audiobook recordings. *2012 IEEE Spoken Language Technology Workshop (SLT)*. IEEE. https://ieeexplore.ieee.org/document/6424239

Tóth, L., Grósz, T., Gosztolya, G. & Hoffmann, I. (2010). Speech recognition experiments with audiobooks. *Acta cybernetica, 19*(4), 669-682. https://cyber.bibl.u-szeged.hu/index.php/actcybern/article/view/3792

Vít, J. & Matoušek, J. (2016). Unit-selection speech synthesis adjustments for audiobook-based voices. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, speech, and dialogue*. Springer. https://link.springer.com/chapter/10.1007/978-3-319-45510-5_38

Xiao, Y., Li, H., Zhou, K., Zhang, J. & Liu, Y. (2024). Contrastive context-speech pretraining for expressive text-to-speech synthesis. *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24), October 28-November 1, 2024, Melbourne, VIC, Australia*. ACM. https://dl.acm.org/doi/10.1145/3664647.3681348

Yahagi, Y., Tanaka, M., Saito, T. & Nakamura, S. (2025). PaperWave: listening to research papers as conversational podcasts scripted by LLM. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM. https://dl.acm.org/doi/10.1145/3706599.3706664

Yang, L., Krause, M., Seipp, K. & Ricci, F. (2018). Understanding user interactions with podcast recommendations delivered via voice. *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM. https://dl.acm.org/doi/10.1145/3240323.3240389

Zhang, Z., Wu, Y., Li, X. & Chen, S. (2024). SpeechLM: enhanced speech pre-training with unpaired textual data. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 32*. IEEE. https://ieeexplore.ieee.org/document/10476749

Zhao, Z. & McEwen, R. (2022). Let's read a book together: a long-term study on the usage of pre-school children with their home companion robot. *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Sapporo, Japan*. IEEE. https://doi.org/10.1109/HRI53351.2022.9889672

Notas

1 Segundo Rubery (2016), o fonógrafo criado por Thomas Edison em 1877, mesmo apresentando limitações nas gravações iniciais, permitia registrar pequenas rimas infantis e trechos de versos. A partir da tecnologia idealizada por Thomas A. Edison, tornou-se possível conceber o audiolivro, que tinha como objetivo incluir os soldados que retornaram da guerra com deficiência visual.

2 O quadro com as *strings* está disponível em Souza Gonçalves & Nascimento Silva (2026)

3 Um *corpus* pode ser definido como uma coleção organizada de textos ou outros dados linguísticos utilizada para fins de pesquisa e desenvolvimento. Segundo Biber, Conrad & Reppen (1998), a linguística de *corpus* parte do estudo de

grandes coleções de textos (*corpora*) armazenadas e analisadas eletronicamente. De forma complementar, McEnery e Hardie (2012) ressaltam que um *corpus* pode reunir tanto textos escritos quanto transcrições de fala, constituindo-se como uma base fundamental para análise e aplicação em tecnologias de linguagem.