Beyond the Microphone: A Targeted Literature Review of Generative AI in Podcasting

Ilie-Alexandru Neamţu
MintViz Lab, Ştefan cel Mare University of Suceava
Suceava, Romania
alexandru.neamtu@usm.ro

Teresa Chambel
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisbon, Portugal
tc@di.fc.ul.pt

Radu-Daniel Vatavu
MintViz Lab, Ştefan cel Mare University of Suceava
Suceava, Romania
radu.vatavu@usm.ro

Abstract

The podcasting landscape is undergoing a significant transformation due to the integration of generative AI, which is reconfiguring creators' roles and how podcasts are experienced by listeners. In this paper, we present findings from a corpus of scientific articles identified in the ACM Digital Library and IEEE Xplore, which examine the intersection of podcasting and AI. Based on our analysis, two pillars emerge for AI integration into podcasting: (i) *AI as a productivity tool,* supporting human creators through smart editing and analytics, and (ii) *AI as an autonomous content generator,* enabled by agent-based processing, zero-shot audio generation, and high personalization in podcast delivery and consumption. These two pillars provide an account of how scientific contributions position AI within the podcasting landscape, and surface a transition from "podcasting then," as predominantly handcrafted and human-produced media, toward "podcasting now," as an increasingly AI-driven and personalizable audio experience.

CCS Concepts

• **Computing methodologies** → **Natural language processing**; • **Information systems** → **Multimedia information systems**; • **Human-centered computing** → *Human computer interaction (HCI).*

Keywords

Podcasting, Generative AI, Large Language Models, Text-to-Speech

ACM Reference Format:
Ilie-Alexandru Neamţu, Teresa Chambel, and Radu-Daniel Vatavu. 2026. Beyond the Microphone: A Targeted Literature Review of Generative AI in Podcasting. In *ACM International Conference on Interactive Media Experiences (IMX '26), June 09–11, 2026, Athlone, Ireland.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3788851.3815027

1 Introduction

Podcasting has rapidly evolved from a niche digital broadcasting format into a pillar of global media, education, and entertainment [18,29]. Historically, the production of high-quality podcast content has been a labor-intensive process, relying on human expertise in scripting, recording, and post-production editing [2,3]. However, recent advances in generative AI, driven by Large Language Models (LLMs) and text-to-speech (TTS) synthesis, have triggered a technological revolution in the audio domain [17,19,23], where AI-driven systems are capable of understanding complex contexts, generating synthetic voices with emotional depth, and automating content creation [10,32].

This revolution is transforming the very nature of the podcasting medium. Traditionally, podcasting has been a linear and passive broadcast experience. However, the integration of generative AI has introduced new adaptive audio formats, for the delivery of which creators collaborate with AI systems to synthesize content, while listeners leverage AI to personalize it. Yet, the scientific literature exploring the intersection of podcasting and AI tends to isolate specific topics, such as transcription accuracy [3] or multimodal segmentation [4,12,16], and a holistic understanding is missing about how AI innovations are redefining the podcasting ecosystem.

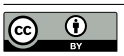
In this paper, we examine how generative AI is driving the transition from traditional audio broadcasting to a new ecosystem. Through a targeted literature review, we provide two main contributions: first, we present a descriptive analysis of the temporal evolution characterizing this emerging field in line with the mass availability of LLMs; second, we identify two distinct pillars, *AI as a productivity tool* for human creators and *AI as an autonomous content generator* for listeners. With these contributions, we offer an understanding of the current podcasting landscape, useful for future studies on production and audience engagement.

2 Scope and Method

To understand how AI is transforming podcasting, we conducted a targeted literature review coupled with thematic analysis, guided by the research question: *How does the scientific literature conceptualize the roles of AI in podcasting?* Answering this question maps the landscape of existing scientific work at the intersection of podcasting and AI, and helps characterize how AI supports creators, listeners, and learners with novel podcasting practices.

2.1 Scope

We start with an operational definition of AI in podcasting, which delimits the scope of our analysis in the broader context where AI is increasingly integrated into audio media in various ways. Specifically, we define *AI in podcasting* as the application of advanced machine learning (ML) models, such as LLMs, and synthetic audio technologies for automation, augmentation, and/or autonomous task execution across the podcasting lifecycle, including pre-production, core production, post-production, and consumption. In this context, our scope encompasses many possible AI systems and roles, from



IMX '26, Athlone, Ireland

ACM ISBN 979-8-4007-2448-0/2026/06
https://doi.org/10.1145/3788851.3815027

AI-driven tools that process audio data for multimodal segmentation [12] and trailer extraction [32] to autonomous systems that generate complex conversational formats, such as NewsPod's [19] synthesized interviews or PaperWave's [33] turning of academic articles into conversational podcasts. These generative capabilities show a profound technological shift from the medium's traditional roots, originally represented by simple episodic audio broadcasting distributed via RSS feeds [11].

2.2 Study Design

Literature reviews can take various forms, from exhaustive quantitative studies to qualitative thematic analyses [13]. Instead of conducting a formal systematic review, we adopted a targeted and synoptic [31] approach by focusing on a specific research question and targeting venues where contributions falling within our scope have been frequently published; subsequently, we implemented *identification, screening,* and *eligibility* stages, as described next.

Identification. We selected the ACM Digital Library and IEEE Xplore since they represent two major databases indexing peer-reviewed research in computer science. After conducting preliminary searches to familiarize ourselves with the existing literature, we decided to conduct searches targeting only document titles. The query formulated for the ACM Digital Library, "query": {Title:((podcast))} "filter": {ACM Content: DL}, returned a total of 442 results. For IEEE Xplore, the equivalent query ("Document Title":podcast) returned 237 results. Both queries were run on March 11, 2026. Overall, we identified a total of 679 raw results.

Screening. We filtered out results irrelevant from an academic point of view, as follows. For the ACM DL, we narrowed the results using the platform's filters to retain only full and short papers, extended abstracts, and demonstrations, which amounted to 81 results. This significant reduction occurred because the vast majority of the excluded entries were AI-generated audio podcasts automatically created to summarize research.[1] For IEEE Xplore, we applied the platform's advanced indexing filters on Speech Recognition, Critical Thinking, Language Model, and Learning Outcomes, which eliminated loosely related content, narrowing the IEEE Xplore corpus down to 19 potential papers.

Eligibility. We applied the following two eligibility criteria:

- EC_1: *Availability of the full text and peer-reviewed papers.* The papers must be peer-reviewed (e.g., journal articles, conference proceedings) and written in English, while their full text must be readily available.
- EC_2: *Thematic relevance.* The papers must explicitly investigate AI within podcasting. Specifically, papers were selected for further analysis if they provided frameworks, implementations, or empirical evidence for AI used in podcasting.

Following the eligibility stage, we retained a corpus of 15 scientific papers (marked with the ♦ symbol in the References section), which were subsequently included in the data extraction stage.

3 Results

Our analysis revealed a rapid evolution in both the volume of research and the roles of AI in podcasting. Next, we establish the descriptive metadata by detailing the temporal trajectory and primary publication venues; subsequently, we conduct an inductive thematic analysis that maps the integration of AI into podcasting.

[1]Artificial Intelligence (AI) Tools on the ACM Digital Library, https://dl.acm.org/generative-ai/summarizations.

Table 1: Thematic classification of the selected corpus (N=15) regarding AI integration in podcasting.

Main theme	%	AI applications in podcasting
AI as a productivity tool for podcasts	40%	Smart editing and content reuse [1,32]; Visual navigation driven by LLMs [10,26]; LLM-driven evaluation and analytics [3,24]
AI as an autonomous generator for podcasts	60%	Conversational formats and the democratization of science [27,33]; Automated pipelines and personalization [14,17]; Agentic architectures and complete autonomy [23]; The impact of generative audio on academic performance [9]; Reducing cognitive barriers [5,20]; Evaluation of creativity and pedagogical quality in synthetic content [6]

3.1 Temporal Evolution and Publication Venues

An analysis of the papers' distribution by year revealed an exponential increase in the past three years. Of the 15 papers in our corpus, only one [3] was published in 2023, followed by two papers [26,32] in 2024, whereas the absolute peak occurred in 2025, with 11 papers [1,5,6,9,10,14,17,20,24,27,33], and already one [23] available in 2026. The most frequent venue in our corpus was CHI (the CHI Conference on Human Factors in Computing Systems) [6,9,14,33], with 4 out of 15 papers (26.7%). Other SIGCHI venues, including DIS (Designing Interactive Systems) [32] and IMX (Interactive Media Experiences) [26], were each represented by a single paper. Among IEEE venues, ICIITCEE (Intelligent and Innovative Technologies in Computing, Electrical and Electronics) appeared twice [17,23], corresponding to 13.3% of the corpus.

3.2 Themes in AI and Podcasting

To answer our research question, we applied inductive thematic analysis, consisting of initial open coding, pattern extraction, and theme definition. As a result, we classified the roles of AI in podcasting into two emerging themes (see Table 1 for details):

- **AI as a productivity tool**, supporting human-AI collaboration for smart editing, visual navigation, and analytics. In this category, AI does not replace human creators, but acts as a tool assisting them. This includes methods for editing content in collaboration with AI to accelerate production [1,3,32] and using LLMs for analysis and visual recommendations [10,26]; see Subsection 3.2.1 for details.
- **AI as an autonomous content generator** via automated processes, agentic architectures, zero-shot audio production with consumption impact through modality transformation and high content personalization; see Subsection 3.2.2.

3.2.1 AI as a productivity tool. Creating derived content from longer podcast episodes has always been a labor-intensive task. For example, Austin and Samuel [3] reported on transcription and

Figure 1: Word clouds generated from the titles of the papers included in our corpus (N=15), corresponding to the two themes: AI as a productivity tool (left) and AI as an autonomous content generator (right); see correspondences with Table 1. *Note:* the word clouds were created using https://www.wordclouds.com.

background noise removal tools. Wang et al. [32] developed PodReels, a human-AI co-creation system for extracting trailers; their system integrates active speaker diarization and filler word detection, using GPT-4 to suggest narrative segments. A major technical challenge in audio repurposing is represented by the loss of paralinguistic nuances during speech-to-text conversion. To address this issue, Adithya et al. [1] proposed the Emotion-Aware Multimodal Repurposing (EAMR) framework, where emotion is modeled as a computational feature and transcription unified with sentiment analysis. Unlike text or video, audio format is opaque, making navigation difficult for listeners. To address this aspect, Park et al. [26] explored the visualization of audio segments through generative models; by combining LLMs for theme extraction and image generation for visual cues, visual navigation of a podcast can eliminate the need for manual keyword searches. Lastly, AI can serve as an advanced analytical engine for podcast platforms. For example, Fabbri et al. [10] replaced traditional A/B testing with an LLM-as-a-Judge system for podcast recommendations. These contributions are reflected by specific keywords in Figure 1, left, such as "recommendations," "enhancing," "evaluating," and "co-creation."

3.2.2 AI as Autonomous Podcast Generator. The integration of LLMs and advanced speech synthesis has enabled new approaches to podcasting through modality transformation, which allows static texts to be converted into audio with minimal or no human intervention. For example, systems like PaperWave [33] can generate dialogue scripts from PDFs, while others integrate RAG architectures to create chatbots [17]. Modality transformation proves pedagogically competitive; for instance, Banihashemi et al. [5] showed that synthetic podcasts generated by NotebookLM can match or surpass human-authored explanations. Mysore et al. [23] proposed an agentic framework that coordinates specialized agents for ideation, script writing, review, and production to achieve zero-shot podcast creation. This autonomous capability is leveraged for extreme personalization, such as by adapting print media into news podcasts for users on the go [27] or through educational systems, such as PAIGE [9], which adapt the information from textbooks according to the student's academic major and instructional preferences.

AI-generated conversational formats reduce listening anxiety for non-native speakers, though they may not fully induce learning flow [20], while longitudinal studies indicate a risk of over-relying on synthetic audio summaries at the expense of critical reading [6]. To address these challenges, prior work has examined AI-generated podcasts for communicating research [14] and applied specific metrics [24] to assess the reliability of synthetic teachers. These contributions are reflected by specific keywords in Figure 1, right, such as "AI-generated," "personalized," "listening," and "comprehension."

4 Discussion

Based on our findings, we propose future research opportunities at the intersection of generative AI and podcasting practice, addressing creators, listeners, and learners.

4.1 Who's Podcasting Now?

In our analysis, we observed a shift from traditional human-led podcast production, including using AI as a tool, toward autonomous agent-based systems [23], where AI takes on ideation, script writing, voice synthesis, delivery, and may also act as a conversational companion. This transition makes podcast production faster for experienced creators and lowers barriers for novices and those without any production expertise. However, greater AI involvement in podcast generation raises new ethical and economic challenges, such as regarding authorship (who should be acknowledged for AI-generated content, scripts, or synthesis) and economic fairness (who is compensated when AI replicates specific human voices, content, or podcast formats). These challenges are application-dependent: when a synthetic podcast is a modality translation of pre-existing content, authorship and factual integrity are generally preserved; however, when an AI system mixes multiple sources to synthesize new content, there is an increased risk of flooding streaming platforms with synthetic audio that, despite sounding authoritative and polished, may contain fabricated information. Consequently, we recommend future research on developing automated mechanisms for ensuring factual accuracy and creator accountability in podcast production, specifically for zero-shot audio generation, before AI-generated episodes reach their listeners.

4.2 How Was That Podcast?

As the podcasting ecosystem integrates generative AI more deeply, traditional evaluation metrics, such as audio clarity, bit rate, or transcription accuracy, are no longer sufficient. For AI-generated podcasts, the experiential and communicative dimensions must be equally assessed, such as emotional resonance, conversational naturalness, narrative coherence, and the ability to produce creative and meaningful analogies. These are key to listener engagement, yet

the community currently lacks the means to evaluate the soft skills of synthetic podcast hosts. To address this aspect, we recommend future research that expands computational evaluation metrics for AI-generated podcasts [24] with new ones addressing affective expressiveness and conversational quality. These could be applied to determine whether AI-generated podcasts meet informational, entertainment, or educational criteria before deployment.

4.3 Can Listeners Tell, and Do They Care, When a Podcast Is Synthetic?

To understand the benefits and challenges of AI in podcasting, it is useful to situate our paper corpus in the broader context of AI development and application. As Haigh [15] argues, the field has shifted from 20th-century approaches, where humans explicitly encoded rules and logic, to 21st-century deep neural architectures and large-scale training. Our findings reveal that podcasting has mirrored this transition. For example, earlier work involved AI as an analytical or translation tool, with research focusing on semantic discovery and information retrieval, such as neural instant search [16], semantic metadata [4], multimodal topic segmentation [12], and predicting tasks [28]. When automation was first introduced to podcasting, it relied on simple text-to-speech for vocalizing existing content [7,22], while interactive podcasting [2,19] still had a human-authored logic. In contrast, more recent systems rely on LLMs to synthesize new content [23,33], making "podcasting now" of a very different kind from "podcasting then." As Haigh [15] cautions, LLMs can produce output that appears impressively human-like, but are prone to fabrication: they speak with confidence irrespective of their statements being true or not. This raises concerns for zero-shot and autonomous podcast generation, where highly polished podcasts may circulate despite containing weakly sourced or false information. Future research is recommended to investigate how listeners perceive, trust, and distinguish human-produced from AI-generated podcasts as well as to develop methods for detecting synthetic podcasts by taking the user's perspective into account.

4.4 How Are We Learning from Podcasts?

As synthetic audio technologies deliver personalized learning [9, 20], a balance must be sought between their pedagogical benefits and cognitive risks. On one hand, these technologies can make complex academic materials more accessible and personalized to individual learning paces; on the other hand, there is the risk of shallow understanding. Longitudinal observations suggest that learners may over-rely on AI-generated audio summaries to the detriment of critical reading of the original texts [6] and, as listeners become accustomed to passive consumption of synthetic content, they risk losing touch with the actual sources [14]. Future research is needed to investigate how podcasting can actively support critical reflection instead of simply facilitating access to content.

4.5 Toward Podcasting as a Multimedia Learning Experience

AI-generated and mediated podcasts have the capacity to extend beyond audio broadcasting and function as explanatory media that structure information for comprehension [27,33]. Just as hypermedia brought structure and interaction to multimedia, helping to deal with cognitive load for reflective learning [8,25], conversational AI can provide support in the conceptualization and construction learning phases, and then especially in the dialogue phase, allowing learners to express, discuss and strengthen their emerging understanding [8]. In this context, AI-mediated podcasting may help transform listening into an experiential learning process in which knowledge emerges through interaction. While risks, such as those described in the previous subsections, also exist in multimedia scenarios, there are many opportunities to foster critical thinking through these dialogues and enrich learning with multiple media. Since audio by itself can be quite opaque, generative AI can help with structuring, visualization, and navigation [26]. Other media types, such as video, can bring realism and authenticity through visual information [8]. Future research should investigate how AI-mediated interactive multimedia podcasting can effectively support learning by considering foundational principles of media and cognition [8,21,25,30] within AI-driven conversational scenarios.

5 Conclusion

We examined in this work the integration of AI into podcasting by analyzing relevant scientific literature from the two most representative databases for computer science research. Notably, the volume of this research has seen exponential growth over the past three years, aligning with the rapid rise of LLMs. In our analysis, we identified a clear shift from AI as a passive support tool toward an autonomous creator of podcast content. Future work will expand our analysis by considering new publications and extending queries to other electronic databases to identify further scientific contributions. What is clear is that podcasting, once a niche digital broadcasting format, is undergoing a broad transformation in how content is created and mediated through AI-based systems as well as how it is encountered and experienced by listeners, with new risks and challenges, but also new opportunities to explore.

References

[1] S. Adithya, G.A. Rohan, A.R. Muthudhanush, and Devi Naveen. 2025. Emotion-Aware Multimodal Repurposing for Podcasts: The Conceptual Design of PodTo-Post.ai. In *Proceedings of the IEEE Pune Section International Conference (PuneCon '25)*. IEEE, USA, 1–6. doi:10.1109/PuneCon67554.2025.11379248 ♦.

[2] Pedro Almeida, Pedro Beça, Telmo Silva, Marcelo Afonso, Iulia Covalenco, and Carolina Duarte Nicolau. 2022. A Podcast Creation Platform to Support News Corporations: Results from UX Evaluation. In *Proceedings of the ACM International Conference on Interactive Media Experiences (IMX '22)*. ACM, New York, NY, USA, 343–348. doi:10.1145/3505284.3532978

[3] Andrea Austin and Anita Samuel. 2023. Enhancing Podcasting by Leveraging AI. *eLearn* 2023, 10, Article 2 (2023). doi:10.1145/3626767.3625304 ♦.

[4] Maryam Aziz, Alice Wang, Aasish Pappu, Hugues Bouchard, Yu Zhao, Benjamin Carterette, and Mounia Lalmas. 2021. Leveraging Semantic Information to Facilitate the Discovery of Underserved Podcasts. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM '21)*. ACM, New York, NY, USA, 3707–3716. doi:10.1145/3459637.3481934

[5] Ali Banihashemi, Amirali Shahriary, and Yadollah Yaghoobzadeh. 2025. Human vs NotebookLM for Educational Podcasts: A Controlled Experiment on Two General Topics. In *Proceedings of the 15th Int. Conference on Computer and Knowledge Engineering*. IEEE, USA, 1–6. doi:10.1109/ICCKE68588.2025.11273844 ♦.

[6] Dennis Benner, Jannik Rauch, Andreas Janson, and Jan Marco Leimeister. 2025. An Explorative Diary Study of AI-Generated Podcasts in University Education: Benefits, Challenges, and Future Directions. In *Proceedings of the Extended Abstracts of the CHI Conf. on Human Factors in Computing Systems (CHI EA '25)*. ACM, New York, NY, USA, Article 70, 8 pages. doi:10.1145/3706599.3719957 ♦.

[7] Marcelo Carvalho Afonso, Pedro Almeida, Pedro Beça, Telmo Silva, and Iulia Covalenco. 2022. Usability of Text-To-Speech Technology in Creating News Podcasts using Portuguese of Portugal. In *Proceedings of the ACM International*

Conference on Interactive Media Experiences (IMX '22). ACM, New York, NY, USA, 363–368. doi:10.1145/3505284.3532968

[8] Teresa Chambel, Carmen Zahn, and Matthias Finke. 2006. Hypervideo and Cognition: Designing Video-Based Hypermedia for Individual Learning and Collaborative Knowledge Building. In *Cognitively Informed Systems: Utilizing Practical Approaches to Enrich Information Presentation and Transfer*. IGI Global, Hershey, PA, USA, 26–49. doi:10.4018/978-1-59140-842-0.ch002

[9] Tiffany D. Do, Usama Bin Shafqat, Elsie Ling, and Nikhil Sarda. 2025. PAIGE: Examining Learning Outcomes and Experiences with Personalized AI-Generated Educational Podcasts. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, New York, NY, USA, Article 896, 12 pages. doi:10.1145/3706598.3713460 ♦.

[10] Francesco Fabbri, Gustavo Penha, Edoardo D'Amico, Alice Wang, Marco De Nadai, Jackie Doremus, Paul Gigioli, Andreas Damianou, Oskar Stål, and Mounia Lalmas. 2025. Evaluating Podcast Recommendations with Profile-Aware LLM-as-a-Judge. In *Proceedings of the 19th ACM Conference on Recommender Systems (RecSys '25)*. ACM, New York, NY, USA, 1181–1186. doi:10.1145/3705328.3759305 ♦.

[11] Michael W. Geoghegan and Dan Klass. 2007. *Podcast Solutions: The Complete Guide to Audio and Video Podcasting* (2nd ed.). Apress, Berkeley, CA, USA. doi:10.1007/978-1-4302-0473-2

[12] Iacopo Ghinassi, Lin Wang, Chris Newell, and Matthew Purver. 2023. Multimodal Topic Segmentation of Podcast Shows with Pre-trained Neural Encoders. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR '23)*. ACM, New York, NY, USA, 602–606. doi:10.1145/3591106.3592270

[13] Maria J. Grant and Andrew Booth. 2009. A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies. *Health Information & Libraries Journal* 26, 2 (2009), 91–108. doi:10.1111/j.1471-1842.2009.00848.x

[14] Quan Connie Gu, Daniel Hickey, and Kimiko Ryokai. 2025. When AI Tells Their Story: Researchers' Reactions to AI-Generated Podcasts as a Tool for Communicating Research. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. ACM, New York, NY, USA, Article 597, 6 pages. doi:10.1145/3706599.3719971 ♦.

[15] Thomas Haigh. 2025. Artificial Intelligence Then and Now. *Commun. ACM* 68, 2 (2025), 24–29. doi:10.1145/3708554

[16] Helia Hashemi, Aasish Pappu, Mi Tian, Praveen Chandar, Mounia Lalmas, and Benjamin Carterette. 2021. Neural Instant Search for Music and Podcast. In *Proc. of the 27th ACM Conference on Knowledge Discovery & Data Mining (KDD '21)*. ACM, New York, NY, USA, 2984–2992. doi:10.1145/3447548.3467188

[17] Arya Jalindar Kadam, Prasad Rajaram Kute, Chinmay Ashok Kale, Kushal Bhoraji Pathave, Chaitali Shewale, and Pawan Wawage. 2025. Enhancing Comprehension with LLM, TTS, and RAG: Transcription of Text into Podcasts and Chatbots. In *Proceedings of the International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics*. IEEE, USA, 1–6. doi:10.1109/IITCEE64140.2025.10915367 ♦.

[18] Daniel Köhler, Sebastian Serth, and Christoph Meinel. 2023. On Air: Benefits of Weekly Podcasts Accompanying Online Courses. In *Proceedings of the 10th ACM Conference on Learning @ Scale (L@S '23)*. ACM, New York, NY, USA, 311–315. doi:10.1145/3573051.3596178

[19] Philippe Laban, Elicia Ye, Srujay Korlakunta, John Canny, and Marti Hearst. 2022. NewsPod: Automatic and Interactive News Podcasts. In *Proceedings of the 27th International Conference on Intelligent User Interfaces (IUI '22)*. ACM, New York, NY, USA, 691–706. doi:10.1145/3490099.3511147

[20] Qianqian Liu and Yan Zhao. 2025. The Impact of NotebookLM Generated Podcasts on EFL News Listening Comprehension, Flow Experience, and Anxiety: A Quasi-Experimental Study. In *Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area Education Digitalization and Computer Science Int. Conference (EDCS '25)*. ACM, New York, NY, USA, 665–670. doi:10.1145/3746469.3746573 ♦.

[21] Richard E. Mayer. 2020. *Multimedia Learning* (3rd ed.). Cambridge University Press, Cambridge. doi:10.1017/9781316941355

[22] Giulio Mori, Maria Claudia Buzzi, Marina Buzzi, and Barbara Leporini. 2010. Structured Audio Podcasts via Web Text-to-Speech System. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. ACM, New York, NY, USA, 1281–1284. doi:10.1145/1772690.1772902

[23] Ashwin N. Mysore, Shreesha M. Rao, M. Pramath, and V. Laxmi. 2026. A Framework for Autonomous Podcast Generation. In *Proceedings of the International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics*. IEEE, USA, 1–5. doi:10.1109/IITCEE67948.2026.11394302 ♦.

[24] Harini Narayanan. 2025. Quantifying Creativity in AI-generated Podcasts. In *Proceedings of the IEEE 49th Annual Computers, Software, and Applications Conference*. IEEE, USA, 1613–1618. doi:10.1109/COMPSAC65507.2025.00217 ♦.

[25] Donald A. Norman. 1993. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Addison-Wesley, Reading, MA, USA.

[26] Jimin Park, Chaerin Lee, Eunbin Cho, and Uran Oh. 2024. Enhancing the Podcast Browsing Experience through Topic Segmentation and Visualization with Generative AI. In *Proceedings of the ACM International Conference on Interactive Media Experiences (IMX '24)*. ACM, New York, NY, USA, 117–128. doi:10.1145/3639701.3656324 ♦.

[27] Daniel Phillips and Patrick Hosein. 2025. Generating Personalized News Podcasts from Print Media for Those on the Go and the Blind. In *Proceedings of the IEEE International Conference on Technology Management, Operations and Decisions*. IEEE, USA, 1–6. doi:10.1109/ICTMOD66732.2025.11371959 ♦.

[28] Naquee Rizwan, Nayandeep Deb, Sarthak Roy, Vishwajeet Singh Solanki, Kiran Garimella, and Animesh Mukherjee. 2025. Toxicity Begets Toxicity: Unraveling Conversational Chains in Political Podcasts. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25)*. ACM, New York, NY, USA, 11776–11784. doi:10.1145/3746027.3754553

[29] Hossein Shafiei, Ahmad Khonsari, Sina Kashipazha, and Aresh Dadlani. 2020. The Tongue Can Paint: Understanding Popularity in Podcasting Societies. In *Companion Proceedings of the Web Conference (WWW '20)*. ACM, New York, NY, USA, 90–91. doi:10.1145/3366424.3382712

[30] John Sweller, Paul Ayres, and Slava Kalyuga. 2011. *Cognitive Load Theory*. Springer, New York, NY, USA. doi:10.1007/978-1-4419-8126-4

[31] Mihail Terenti and Radu-Daniel Vatavu. 2021. How Do HCI Researchers Describe Their Software Tools? Insights From a Synopsis Survey of Tools for Multimodal Interaction. In *Companion Publication of the International Conference on Multimodal Interaction (ICMI '21 Companion)*. ACM, New York, NY, USA, 7–12. doi:10.1145/3461615.3485431

[32] Sitong Wang, Zheng Ning, Anh Truong, Mira Dontcheva, Dingzeyu Li, and Lydia B Chilton. 2024. PodReels: Human-AI Co-Creation of Video Podcast Teasers. In *Proceedings of the ACM Designing Interactive Systems Conference (DIS '24)*. ACM, New York, NY, USA, 958–974. doi:10.1145/3643834.3661591 ♦.

[33] Yuchi Yahagi, Rintaro Chujo, Yuga Harada, Changyo Han, Kohei Sugiyama, and Takeshi Naemura. 2025. PaperWave: Listening to Research Papers as Conversational Podcasts Scripted by LLM. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. ACM, New York, NY, USA, Article 689, 10 pages. doi:10.1145/3706599.3706664 ♦.