Reverb

Metamorphosis of one-way audio into dynamic and interactive conversations through conversational AI

Shikha Shah

Reverb
by
Shikha Shah

Metamorphosis of oneway audio into dynamic and interactive conversations through conversational AI.

Current content like audiobooks and podcasts often leaves us passive, unable to engage or take notes. Not being able to search or research more about what we are listening to when consuming audio media while driving, cooking, or exercising leads to added situational disability. I am trying to make audio content more interactive and conversational using AI. For the scope of my thesis, I am focusing on creating interactive podcasts and how we could use AI conversations to increase knowledge retention. But it's not just about innovation; it's about ethics. The thesis explores how we could use the upcoming technologies to create new interactions for electracy (electronic literacy).

In the contemporary digital age, enhancing knowledge retention across all forms of literacy and audio media consumed from the internet is a key focus , especially considering the predominantly monologic nature of current audio content such as audiobooks and podcasts. The passive engagement in these activities, often paired with multitasking endeavors like driving or cooking, renders the user unable to actively seek information or jot down notes about discussed topics or unfamiliar terms. Despite the existence of prior art utilizing Artificial Intelligence (AI) tools for content generation and summarization by startups like Descript [28] and Speechify [16, 29], and AI character generation by groups like the Human AI Interaction team at MIT Media Labs, a gap persists in creating an interactive and conversational podcast experience that more deeply engages listeners, fostering a richer understanding and stronger connection with the content. The crux of the problem revolves around coming up with new ways to interact that make them easier to access, even when people are multitasking or in situations similar to having a disability. Employing methods such as Voice Cloning, Text-to-Speech and Speech-to-Text conversions through the Whisper API [46], utilizing the Chat GPT API for text summarization and note-taking [7, 13], and using Llama Index to retain the information context [37], this project aims to transform monologic audio content into a dialogic or conversational format, thereby enriching the user experience by making it more engaging and informative. However, the implementation of these methods is not without challenges, particularly in navigating data privacy and securing user consent, which are paramount to ethical practice in utilizing personal data for voice cloning and customization of experiences. The significance of this endeavor is underscored by the advancements in AI and Natural Language Processing, which have reached a zenith, enabling the conversion of monologues into seemingly realistic conversations. Looking ahead, future work will delve into crafting a personalized audio media consumption experience, wherein user queries could seamlessly connect to another podcast providing the answers, thereby amalgamating different content pieces. This opens up avenues where listeners can potentially engage in discussions with AI-generated podcast narrators, ushering in novel dimensions in conversation, storytelling, and information consumption, and thereby redefining the paradigms of user engagement in audio media platforms.

Reverb
by
Shikha Shah

A thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Design
at the
University of California, Berkeley

Fall, 2023

Faculty Director Signature and Date

Associate Director Signature and Date

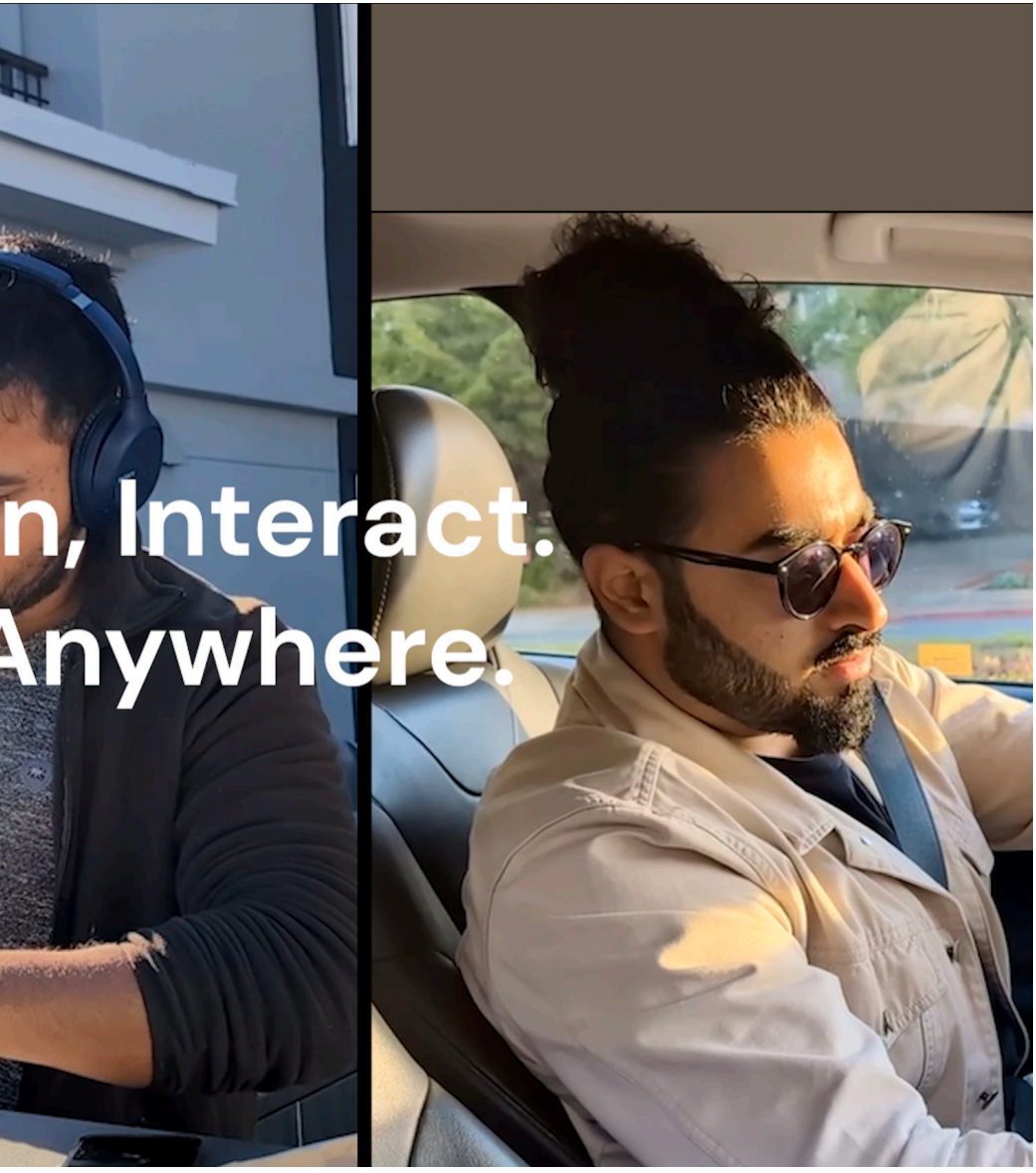
Acknowledge-ments

Conveying my deep thankfulness for everyone who supported me during my thesis is quite a challenge. Their support took many forms, from lending a sympathetic ear, offering crucial feedback, joining me for extended walks, partaking in late-night meals at the studio, driving me home, giving indispensable design suggestions, to just being there for me. My gratitude towards all of you is immense.

Rathin Shah
Heena Shah
Jitesh Shah
Kyle Steinfeld
Hugh Dubberly
Eric Paulos
Yoon Bahk
Hila mor
Govind Balakrishnan
Amanda Macgraw
Ananth Nayak
Yemoon Cho
David Duarte
Elena Hoshizaki
Grace Thompson
Gracy Kureel
Haesung Park
Helena Kent
Kabeer Andrabi
Jahnavi Jambolkar
Karthika Baiju
Neel Shah
Peiyao Wang
Samriddho Ghosh
Tomas Garcia
Wonjoon Oh
Zack Dive
Roshan Mohan
Vidit Nhargav
Qingzhu Zhang
Carmela Wilkins
John Berchbill
Justin Trainor
Zeping Fei
Shravani Nimbolkar
Ankit Bhawsar

Table of contents


Listen, Lear
Anytime, A


n, Interact.
Anywhere.

History/ Prior Art

British philosopher Michael Oakeshott said “Conversation distinguishes the human being from the animal and the civilized man from the barbarian.” When did conversation begin? To raise this question is to ask: When did the faculty of language develop? Earlier, the medium of communication was predominantly oral [41]. Communities gathered, and stories were told, facilitating a direct and dynamic interaction between storytellers and their audience. Knowledge and narratives were not merely shared but were shaped where the audience was actively engaged in the process. The inception of the written word brought forth a paradigm shift in communication, pivoting towards a more asynchronous form of interaction like letters and books. But it still felt that we were a part of the conversation because it was an active activity [41, 67]. With technological advancements, telephony, and radio emerged as transformative mediums in the realm of communication. There was still a part of two-way conversations. Though the depth of the conversation was constrained, features like listener call-ins on the radio had an element of participation in the conversation. As we move towards the digital age, mediums such as audiobooks and podcasts have surfaced as prevalent platforms for information consumption and storytelling. These formats, while immensely accessible and diverse in content, predominantly entail passive consumption, where listeners absorb content without the capability for immediate, direct interaction with the creator.

Greg Ulmer introduced the term “electracy” to encapsulate the skills and competencies necessary to navigate and utilize digital media effectively [68]. In the contemporary digital epoch, individuals frequently resort to rapid online searches, utilizing platforms like Google, to learn unfamiliar terms or concepts manifested through practices such as summarization and strategic notetaking to revisit them and research more about them. Effective communication is fundamentally propelled by an intrinsic curiosity. Amidst an era characterized by diminishing attention spans, amplified by succinct media formats like Instagram Reels and YouTube Shorts, the expedited consumption of media has become increasingly pervasive. According to research, more than 57% of US adults consume news from audio media platforms [33]. Newspod, a research published by Berkeley researchers explored enhancing engagement on such platforms by autogenerating a customized news podcast based on a question-and-answer conversation structure by using natural language processing and text-to-speech technology. The research emphasizes methods to create audio content interactive and personalized [33].

Chinese podcasting apps like Ximalaya, Lizhi, Qingting FM, and Xiaoyuzhou FM were the first ones to introduce interactive podcasts [73]. Spotify was one of the first movers in the US. One of the earlier forms of interactivity in podcasts was including polls so that people who are listening can now participate too [43, 44, 73]. Traditionally





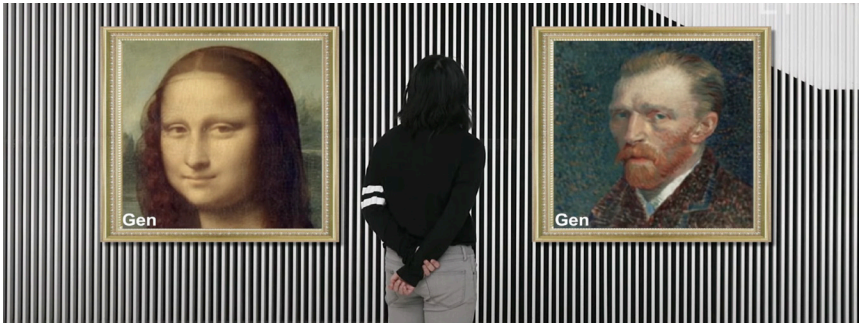

podcast apps also didn't provide enough granular information about your audience. To cope, many podcast creators in the U.S. set up Patreon pages, Substack newsletters, Discord channels, or Facebook groups to bypass this difficulty and reach their fans. Vince Major and Michael Tucker, hosts of the podcast Beyond the Screenplay, made podcasts interactive on Spotify to better understand their audience by gathering feedback from them. In the realm of children's audio entertainment, Pinna, a specialized audio streaming platform tailored for youngsters aged 3-12, introduced interactive storytelling [53]. This innovation, aptly named the "Yes No Audio" series, marked a significant departure from conventional narrative formats. It invited active participation from children, granting them the agency to influence the course of the stories by responding with a simple "yes" or "no" to prompts. This engaging approach immersed young listeners in the tales, transforming them from passive consumers into active co-creators of the narrative experience.

Since the emergence of Chat GPT, AI has made a notable imprint on the podcasting landscape. It has become instrumental in various aspects of the field, spanning from the generation of fresh topic concepts to the enhancement of content for better search engine performance [7, 13, 27]. AI is empowering podcast creators to simplify their processes and produce top-notch content that captivates and motivates listeners. In the realm of artificial intelligence, several key technologies have catalyzed transformative advancements in how we interact with and

process digital information. Speech-to-text capabilities are foundational in voice-activated assistants like Google Assistant and Amazon Alexa, revolutionizing user interactions through natural language understanding. Vector embeddings, essential in search algorithms and recommendation systems, play a pivotal role in Google's search engine and Netflix's content suggestions, enhancing user experiences through contextually relevant results. The GPT API, developed by OpenAI, has significantly advanced natural language processing, finding applications in sophisticated chatbots for customer service and in AI-driven content generation tools used in automated journalism. Similarly, technologies like the Llama Index, which retain contextual information in conversations, are integral in platforms such as IBM Watson for customer service interactions and in mental health applications where maintaining conversational flow is crucial. Many startups are based specifically out of this domain. For instance, Descript, which is funded by OpenAI and many other VCs, is a software that is highly regarded for its innovative approach to audio editing [28]. It transcribes audio, allowing users to edit the audio by editing the text, making the podcast editing process more intuitive and efficient. Speechify, another silicon-valley-based startup, uses AI to simplify podcast creation by automating tasks like content generation, transcription, and episode repurposing. It can clone voices for consistency, recommend content based on listener behavior, and even facilitate voice-overs in different languages while maintaining a natural sound and accent [16, 29].

The voice assistant technology, exemplified by Google assistant, Siri and Alexa, has seen widespread adoption in the United States with consumers at 85.4 million, 81.1 million and 73.7 million users respectively[74].The prevalence of Siri and Alexa illustrates the societal readiness and acceptance for voice-interactive technologies, setting a favorable backdrop for the introduction of more advanced systems like interactive podcasts and audiobooks. Recent breakthroughs in machine learning have paved the way for the creation of highly realistic AI-generated media, synthesizing prose, images, audio, and video data. Some podcasts, such as podcast.ai, are entirely crafted by AI, exemplified by a recent episode featuring a simulated interview between Joe Rogan and Steve Jobs [55]. While the concept of voice cloning and deep fakes might seem somewhat dystopian to some, they are being harnessed for positive uses as well.

The Human AI-Interaction group at MIT Media Labs, for instance, employs these technologies to enhance social connectivity and promote positive learning and well-being, creating AI characters of notable individuals or cartoons that can serve as virtual instructors [12, 51]. These remarkable prior works in the field lay the foundation for the thesis, which delves into the utilization of these AI tools and insights from related work to create an interactive, conversational podcast. This allows users to pose questions and actively participate in the audience. The thesis further investigates how this "intersecting monologue," described by novelist and essayist Rebecca West, who asserted that conversation is an illusion and merely overlapping monologues, can rekindle curiosity among individuals.



AI-generated characters for supporting personalized learning and well-being [51]

Motivation

The current advancements in AI make it an opportune time for the development of technologies like interactive podcasts and audiobooks [1, 4, 32]. AI has reached a point where it can understand and process natural language with remarkable accuracy, enabling real-time interactions that were previously not possible. Machine learning models can now contextualize and respond to user queries, making the experience of consuming audio content more personalized and engaging. Additionally, improvements in voice recognition and synthesis technologies allow these platforms to deliver responses in a natural, conversational manner [33]. This convergence of AI capabilities opens up new possibilities for how we interact with and consume media, making technologies such as interactive podcasts both feasible and potentially transformative. Although platforms may allow for subsequent discussions or comments, the intrinsic, dynamic interaction emblematic of oral traditions is substantially diluted. Michel de Montaigne said " studying books has a languid feeble motion, whereas conversation provides teaching and exercise all at once" [41]. One of the questions that is explored in this thesis is how do we bring an element of interactivity in listening to podcasts through conversational AI.

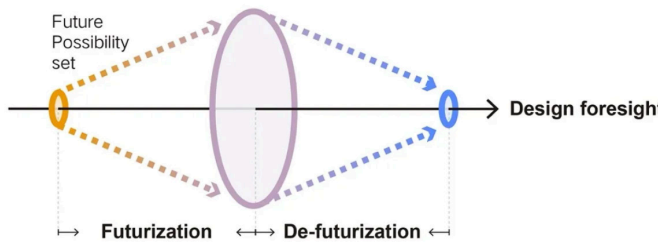
Electracy underscores the need to develop new literacies that align with the evolving digital landscape, ensuring effective communication, and interaction in contemporary mediums. diminishing attention spans might give rise to what Heidegger terms as 'pseudo-understanding' [69]. Metamorphoses in communication mediums, from oral to electronic, signify profound implications for our social interactions, cultural practices, and the construction of knowledge. While technological advancements have undeniably amplified our access to information and narratives, they also pose challenges pertinent to the depth and nature of our engagements with these stories and with each other. People spend most of their day in the virtual world– listening to a talk show while driving, and listening to an audiobook while cooking. podcasts and audio media have become prevalent, consumption often morphs into a passive activity. Concurrent engagement in other activities often relegates the act of further exploring–such as Googling information heard or researching intriguing topics–to a subsequent task. Given the pervasive decline in short-term memory amidst the populace, oftentimes, the impetus to delve deeper into captivating subjects heard about is forgotten. This gradual shift leans towards a trend of diminishing curiosity, potentially gesturing towards an eventual atrophy of inquisitive engagement in our digital society. In this context, the brevity of content consumption and information retrieval emerges not just as a trend, but as a nuanced, adaptive response to the digital age's communicative milieu. Hence another question explored in the thesis is using AI tools for notetaking and summarizing in an intuitive way when consuming audio media passively.

In a lecture at Berkeley, Michael Seibel highlighted a crucial aspect of contemporary information consumption: “Information is abundant but we need to put in efforts to get it. People become disinterested because of lack of speed of getting results.” This observation prompted a deeper reflection on the nature of information consumption today and the necessity for innovative forms of interaction. Seibel’s insights serve as a catalyst for rethinking how we access and engage with information in an era where immediacy often compliments depth, underlining the need for new methodologies in information dissemination and interaction. This raises several thought-provoking questions that particularly challenging designers to think about intuitive user interfaces and seamless integration of interactive features, content personalization and its reliability. For the scope of this, I would focus on exploring these through with Podcast as my primary medium.

Stop Multitasking on Your Phone: Media Multitasking May Lead to Attention Lapses and Poor Memory - Harvard University

Method/ Approach

The thesis employs three distinct frameworks: Futures Thinking, Design Thinking, and AI Thinking methods. Futures Thinking was bifurcated into two components: Defuturization aligns with the enhancement of existing realities, focusing on internal rationality and feasibility. Conversely, Futurization challenges the singular notion of reality, seeking alternative possibilities and embracing external societal and diversity factors. Thus, Futurization expands future possibilities, whereas Defuturization narrows down these options [11]. The Design thinking methodology began with an extensive phase of user and literature research, aimed at gaining a comprehensive understanding of the contextual background and user perspectives. The subsequent stage involved the generation of ideas and concepts through brainstorming sessions and developing interactive prototypes ensuring alignment with the research insights and user requirements [2, 18]. Considering the rapid advancements in technology and artificial intelligence, I propose the incorporation of AI Thinking methods alongside Futures Thinking and Design Thinking. The latter frameworks predominantly emphasize strategic and human-centric aspects. In contrast, AI Thinking centers on technological tools and infrastructures. Given the profound implications of AI technologies in areas like privacy, reliability, ethics, and their disruptive potential, there is a pressing need to integrate frameworks that critically assess the impact of these tools [9, 19]. Although AI has been extensively researched as a foresight tool in computer science, its pervasive influence across various fields necessitates its application in foresight design methodologies as well.



The Futurization and De-futurization with Design foresight [11]

1. Futures Thinking

In Futurization, I explored various future scenarios for the Conversational AI Podcast, considering advancements in AI and changing media consumption trends. This included potential impacts on education, entertainment, and user interaction. In Defuturization, I selected a realistic and achievable future for the project, grounded in current technological capabilities and market trends, while also considering user acceptance [11].

Within the Futures Thinking framework, the landscape of audio media consumption and conversational AI is characterized by several key trends and signals:

- Enhanced Personalization: AI advancements are enabling highly personalized audio experiences, adapting content to individual preferences and behaviors.
- Voice-Driven Interfaces: There is a noticeable shift towards using voice commands for media control, reflecting a trend towards voice as a primary interface in technology.
- Immersive Audio Experiences: Emerging technologies are being developed to create more immersive and interactive audio experiences, including 3D audio and virtual reality integrations.
- Seamless Conversational AI Integration: Conversational AI is increasingly being incorporated into everyday devices, facilitating smoother and more natural interactions.
- Ethical and Privacy Implications: The proliferation of conversational AI raises significant concerns around data privacy and the ethical use of AI.
- Accessibility Enhancements: Conversational AI is poised to substantially improve accessibility in audio media for individuals with disabilities.
- AI-Driven Content Production: The use of AI in content creation and distribution is streamlining and diversifying the production of audio media.
- Challenges of Deepfakes and Misinformation: The ability of AI to create realistic audio deepfakes presents challenges related to misinformation and media credibility.

These trends collectively indicate a future where audio media consumption is not only more interactive and personalized but also seamlessly integrated into daily life. However, this future also necessitates careful consideration of the ethical and privacy aspects associated with these technologies.


Interactive
Podcast
Hey Pod ...

2. Design Thinking

In this study, user interviews were conducted with 35 individuals across varied demographics, who listened to a diverse range of podcasts including comedy, historical discussions, political opinions, and financial topics.In the conducted research, it was discerned that a significant majority, approximately 90% of the participants, demonstrated pre-existing familiarity with digital AI assistants like Siri, Alexa, and Google Assistant. This prior exposure to such technologies had evidently equipped them with a comfort level in engaging in dialogues with a digital AI assistant. The insights gathered, as illustrated in the accompanying diagram, revealed a common challenge among users: while engaging with intellectually stimulating podcasts, they often forgot specific points they intended to research further. Their quick, hastily-written notes often lacked context, making them nonsensical upon later review. This issue was not due to a lack of curiosity but rather a loss of it. Common practices included jotting notes on readily available mediums, taking screenshots of relevant timestamps, scouring transcripts, or relying on memory. These observations informed the ideation process for "Reverb," with a focus on addressing these pain points through user-centered design, testing, and iterative prototyping. Subsequently, the development process involved various iterations to finalize the design of the podcast platform which I named as "Reverb".

3. AI-centered Thinking

We need to be conscious of how data is being generated in this application and how it could be used. I implemented the IDEA framework mentioned in HBR article on "How to change a GenAI future You can't predict" [72]. As shown in the figure, the framework is divided into 4 steps: identify, determine, extrapolate and anticipate. A potential framework can be explored where the podcaster actively participates in the content generation loop, maintaining user privacy and anonymity. This model would involve collating frequently asked or significant questions, especially those seeking the podcaster's perspective on various issues, and presenting them to the podcaster anonymously. Their responses and viewpoints could then be integrated back into the Reverb AI model. This approach would enable the AI to learn and generate responses that align more closely with the podcaster's opinions, especially on nuanced or sensitive topics. Such a system could also provide valuable insights to podcasters about their audience's interests, guiding content creation to enhance listener engagement and expand their follower base. This model, thus, presents a symbiotic framework for data collection and generation, ensuring content accuracy while fostering deeper engagement between

Final Design

The final design of the thesis is anchored in four primary pillars, each addressing key user pain points identified through research. These pillars are: 1) Conversations and Q&A, which enhance our capacity to articulate thoughts, resolve problems, and rapidly assimilate information through interactive dialogues; 2) Notetaking, critical for knowledge retention; 3) Summarization, facilitating the swift consumption of information; and 4) Interference or cross-referencing of various podcasts, a feature instrumental in stimulating user curiosity. These components collectively form the foundation upon which the entire design structure is built.

Reverb is segmented into three primary categories: Firstly, processing the podcast URL link to extract pertinent information about the podcast; secondly, processing user prompts to accurately extract relevant information in response; and thirdly, the design of the user interface (UI) for the webpage, which is pivotal for ensuring user-friendly interaction and accessibility. Each of these categories plays a crucial role in the seamless functionality and user experience of the thesis.

I chose Malcolm Gladwell's "Revisionist History" podcast for thesis prototyping due to its rich blend of historical and intellectual content [38]. The podcast frequently delves into complex topics, often filled with jargon and nuanced concepts, which naturally stimulates listeners to conduct further research for a deeper understanding. Additionally, the podcast presents information that is not only engaging but also valuable, often prompting listeners to take notes. This combination of informative, thought-provoking content and the necessity for additional research and note-taking makes it an ideal choice for testing and demonstrating the functionalities of the AI-driven podcast interaction system developed in this thesis.

From user interviews, it became evident that most podcast listeners are accustomed to accessing content through well-known platforms like Spotify, Apple Podcasts, and Google Podcasts. The interviews highlighted a reluctance to download new applications or transition to unfamiliar platforms. Additionally, the need for accessibility across various devices, such as phones and iPads, was emphasized. Consequently, these findings informed the decision to develop a web interface, enabling users to interact with podcasts by simply inputting their URLs. This solution was chosen to facilitate smoother design iterations and to cater to user preferences and habits as revealed in the interviews.

1. Webpage UI Design: Enhancing User Interaction and Accessibility

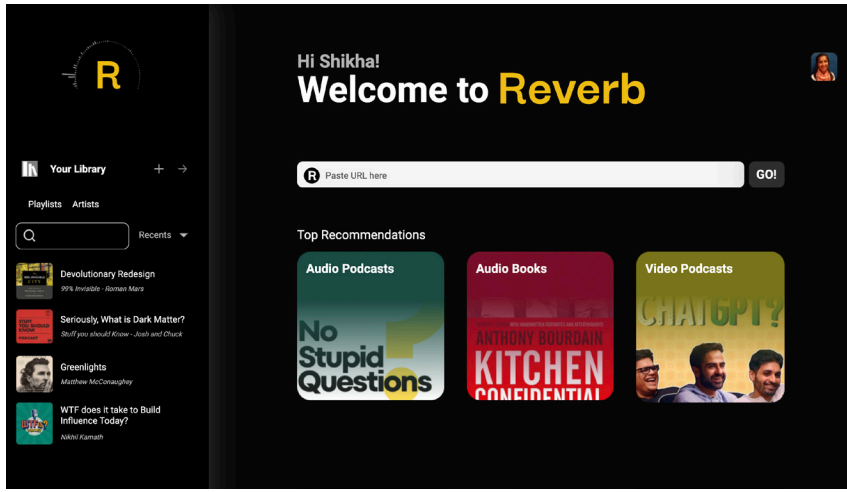
   In the initial stages of the project, I utilized Gradio for the frontend interface. Gradio is a tool that enables the integration of Python

code blocks (used for backend programming) to automatically generate a frontend webpage [62]. This proved particularly useful for quick iterations and functionality testing. For the scope of this thesis, I focused on easily available Podcast platforms like Youtube. To embed a video player on the webpage, I employed YouTube's iframe API (Application Programming Interface, is a set of rules and protocols for building and interacting with software applications), which facilitates the embedding of YouTube videos. This API allows access to podcast data such as audio, timestamps, descriptions, and captions via JavaScript. However, due to certain limitations with Gradio in handling these complex tasks, I eventually transitioned to using JavaScript and HTML for the webpage development.

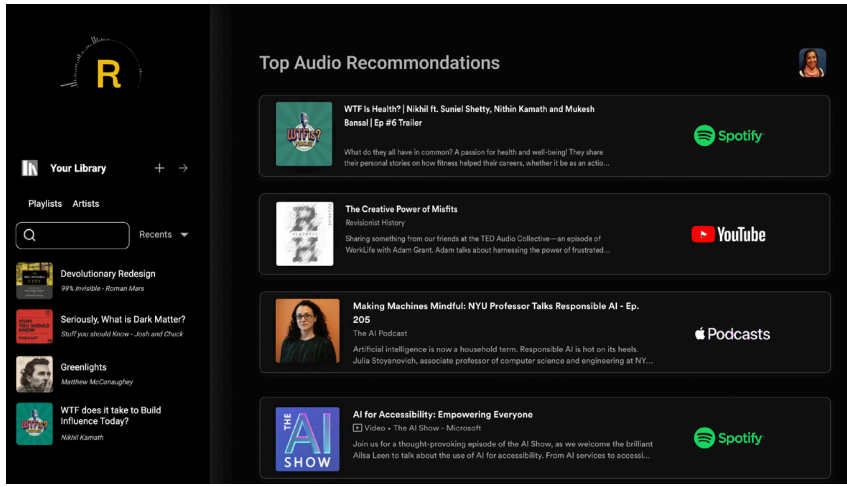
In this thesis, the user interaction flow with the web application is structured as follows: Upon opening the webpage, the user is prompted to grant microphone access, a step crucial for privacy. The user then inputs the podcast URL into a textbox and begins listening. Interaction with the podcast is initiated by the user saying a wake word, "Hey Pod." However, due to the advanced computational requirements of constant low-power detection and stringent privacy protocols, this thesis simulates the wake word through 'start recording' and 'stop recording' buttons. When the user starts recording to pose a question, the podcast automatically pauses, processes the prompt, and generates an audio response. This interactive process can be repeated until the user chooses to resume the podcast.

2. Extraction of Key Data from Podcast URLs

Upon retrieving podcast data from YouTube using an API, the audio file of the podcast is transcribed with OpenAI's Whisper API, which is a tool for converting spoken words into written text [46]. The transcription, along with the start and end timestamps of each sentence, is stored in a PDF format. For this thesis, a database was created containing Malcolm Gladwell's "Revisionist History" podcasts, including summaries of the podcast content. These summaries are converted into vector embeddings, a process that transforms text into a numerical format that a computer can understand. This transformation is crucial because it allows the AI smart assistant to search the database for other podcasts that might provide more detailed answers to user queries, thus enabling the smart assistant to reference relevant podcasts and supply the user with comprehensive information.



Home page of the UI Interface

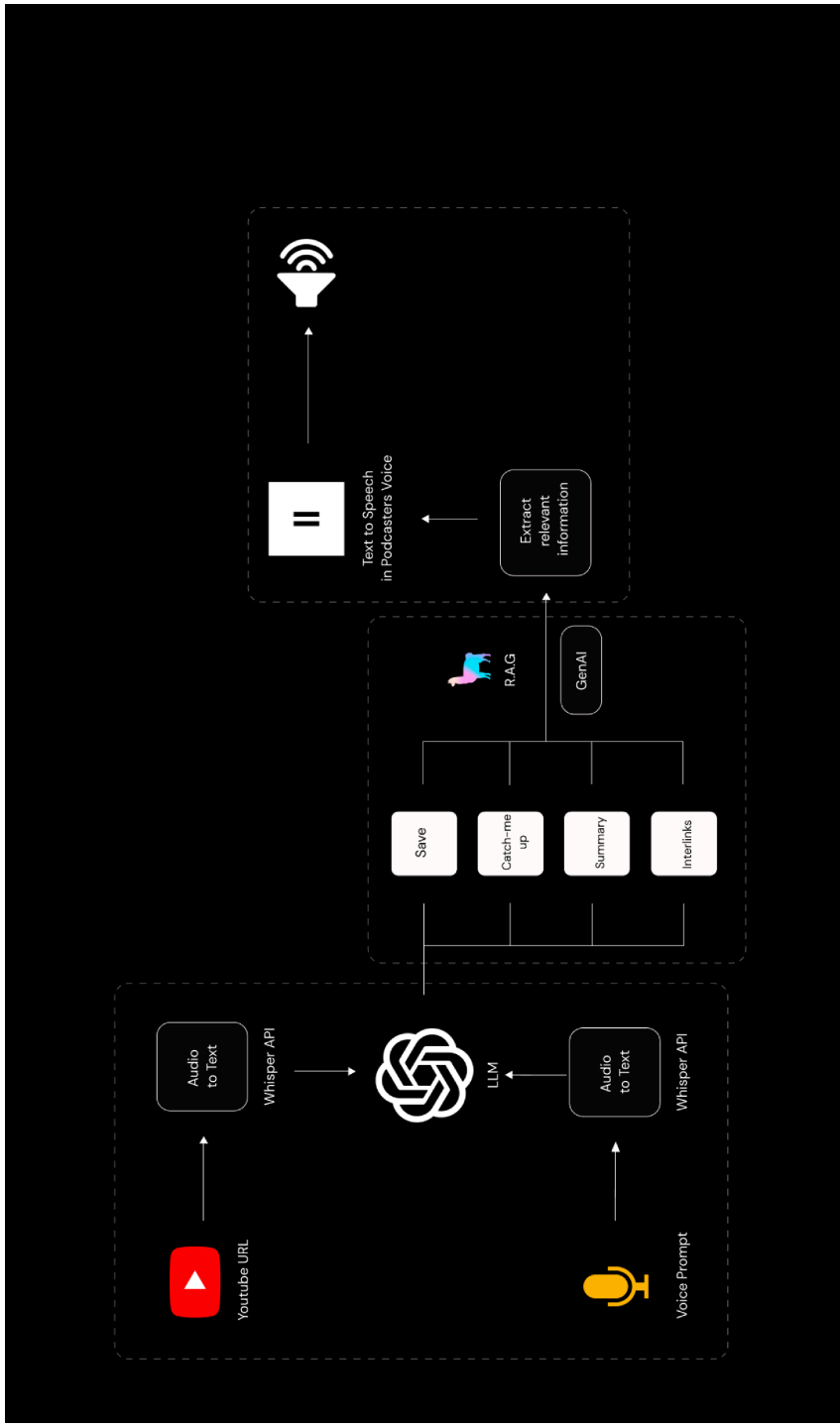


Customized recommendation of podcasts and the media plaforms.

3. User Prompt Analysis for Relevant Information Generation and Retrieval

When a user asks a question during the podcast, the system immediately pauses the podcast. The user's spoken question, along with the specific time it was asked, is recorded as an audio file. This audio is then converted into text using the Whisper API, a tool that accurately transcribes spoken words into written form. The transcribed text is then processed using the Llama Index. This tool helps in maintaining the context of the conversation by using the text from the podcast's transcript up to the point where the question was asked.This context, along with the user's question, is passed to the GPT API. The GPT API serves as a virtual podcaster, generating responses using two methods [21]. The first method, GenAI, searches for and creates relevant information to answer the question. The second method, RAG (Retrieval-Augmented Generation), specifically looks through the podcast's own content to find answers. This process ensures that the responses to the user's questions are not only contextually appropriate but also drawn from the most relevant sources, whether from within the podcast or from a broader range of information.

Once a response to the user's question is generated, I utilize Eleven Labs API to create a voice that sounds like the podcaster for the generated response [30]. Eleven Labs is a tool that can clone voices, making the AI assistant sound like a virtual version of the actual podcaster. This creates an immersive experience, almost as if the user is having a real conversation with the podcaster. Eleven Labs can modify voice parameters to make it sound very realistic. However, it's important to only clone voices with the podcaster's permission. Additionally, providing context to the GPT (Generative Pre-trained Transformer) during response generation ensures the information is relevant and within context, minimizing the risk of generating inaccurate or potentially harmful content. This approach aims to enrich the user experience while maintaining ethical standards in voice cloning.



System Diagram of the web interface techstack

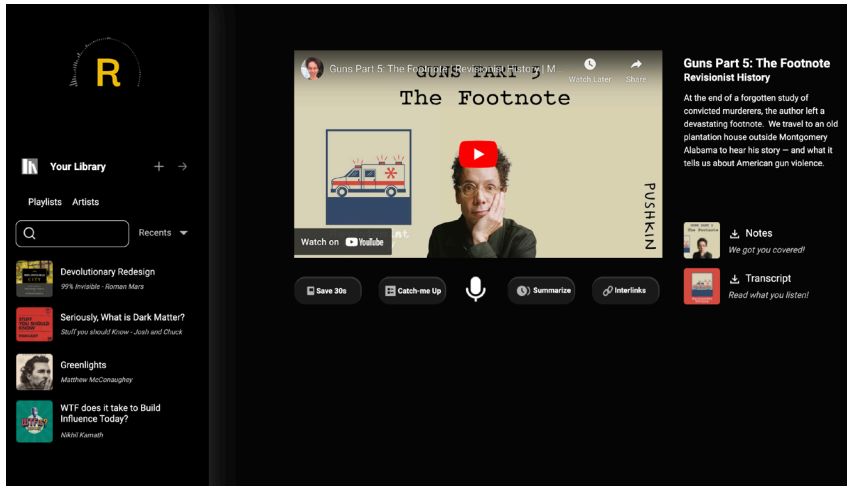
In this thesis, prompt designing, a crucial component of conversation AI, is meticulously addressed. Effective prompt design is integral for ensuring AI interactions are natural, relevant, and meaningful.

For this project, prompts were categorized into four distinct types to enhance user experience:

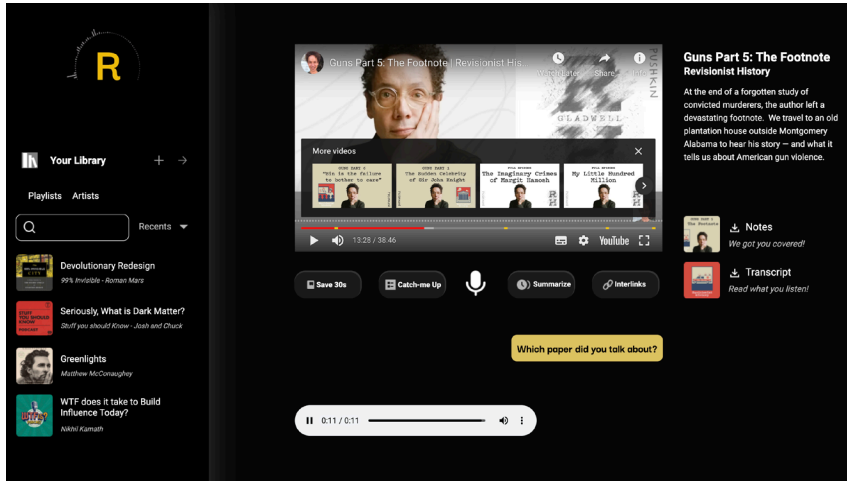
- General Queries: Allows users to engage in conversations and Q&A with the AI version of the podcaster.
- “Save” Command: When this word is included in the prompt, the system saves the last 15 seconds of the podcast into a file for future reference.
- “Summarize” Command: This prompt triggers the AI to intelligently understand and summarize the podcast content following the timestamp of the prompt.
- “Recommend ” Command: This command recommends podcasts based on unique conversations between the user and Reverb.

  These categorized prompts are designed to cater to different user needs, making the interaction with the podcast both efficient and user-friendly.

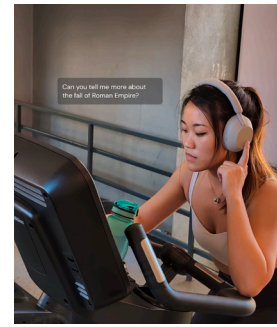
Incorporating the methodologies and analyzing future trends, drivers, and signals, while integrating insights from user research and considering technological and ethical constraints, the thesis culminates in a design predicated on a future scenario. This scenario envisions a world where the speed of information exchange and depth of knowledge acquisition are paramount, even amidst multitasking. It foresees an era where interaction with digital AI assistants is akin to engaging with an alter ego, transcending in-person conversations. It portrays a future where life is navigated in a unique harmony of asynchronous and synchronous experiences.

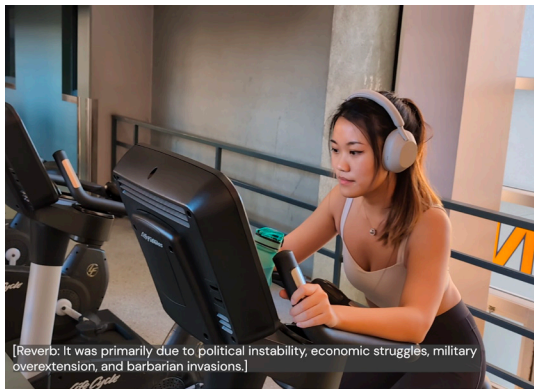


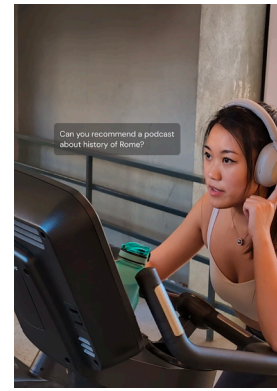
Podcast of the URL serached with all the touch and voice command features

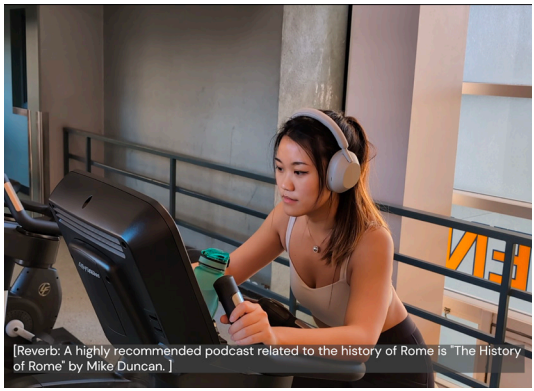


Feedback of the voice command provided as text and provoded with audio response









User flow showing contextual, summarize and interlink features

Discussion

In the scope of this thesis, it was observed that the prevalent familiarity with digital AI assistants such as Siri, Alexa, and Google Assistant played a pivotal role in facilitating users' acclimatization to conversational AI podcast technology. This ease of adaptation was evident as users demonstrated a readiness to interact with the AI for information retrieval, paralleling their usage patterns with existing digital assistants. Notably, the Reverb system's customization capabilities, encompassing content summarization, note-taking, and the simulation of interaction with a podcast host, garnered positive reception. A significant majority, approximately 90% of users, indicated a strong preference for these personalized features, underscoring the growing demand for tailored interactive experiences in digital media consumption.

The implementation of cross-referencing in podcast episodes remarkably ignited the listeners' curiosity. Suppose a listener is deeply engrossed in Malcolm Gladwell's "Revisionist History" podcast, specifically the episode "Guns Part 5: The Footnote". If they inquire about a topic that's covered in a different episode, say "Guns Part 3: A Shooting Lesson", the digital AI assistant could cleverly bring up this relevant episode in its response. This innovative feature not only enriched the listener's journey through interconnected narratives but also effortlessly bridged the gap in discovering similar content. By weaving a web of references across episodes, it transformed the podcast experience into an engaging exploration, organically guiding listeners through a labyrinth of stories and insights. This approach not only heightened engagement but also subtly served as a personalized guide to related content, mirroring the intuitive recommendations of a knowledgeable friend.

The integration of conversational AI in podcast platforms, as exemplified by the "Reverb" technology, while offering significant personalization and engagement benefits, also brings forth substantial ethical and privacy challenges. Issues such as constant microphone monitoring present significant privacy intrusions, while the risks associated with voice cloning for fraudulent activities or deepfake creation highlight the potential for misuse. A major legal and ethical quandary is the absence of legal recognition for voice ownership, which complicates the protection of individual vocal identities [42]. Moreover, the accuracy and potential biases in AI-generated responses pose further ethical concerns. In instances where a podcaster's voice is cloned to enhance user engagement, with their consent, it becomes crucial to ensure that the AI does not generate misleading or factually incorrect information. This necessitates transparent communication with users, clarifying that

responses generated by AI, though in the podcaster's voice, are not reflective of the podcaster's actual thoughts or opinions.

This leads to the critical question of the extent to which AI should be trained. Should the AI encompass all publicly available information related to the podcaster to mimic their thoughts, or should it be restricted to the content within the podcast to prevent biases? The broader implications of such decisions on content authenticity and user trust must be carefully considered. Given these complexities, "Reverb" and similar technologies necessitate the development of robust frameworks that address these multifaceted ethical, privacy, and accuracy concerns. This includes establishing clear guidelines on data usage, ensuring transparency in AI operations, and creating mechanisms to protect against the misuse of synthetic voice technologies. In essence, while these technologies herald a new era of interactive digital media, they also underscore the need for vigilant and comprehensive approaches to safeguard ethical standards and user trust in this rapidly evolving domain.

Future works and Envisionments

This thesis presents a vision for expanding into various future applications, notably in the realm of distance learning education. Imagine an environment where students, while listening to prerecorded online lectures, have the ability to interact, asking questions as though they were physically present in the classroom, or directly taking notes from the video. This technology could also revolutionize engagement in audiobooks and enhance children's educational experiences through interactive dialogues with their favorite animated characters. Additionally, its application in museum audio tours could invigorate historical learning, transforming it into dynamic, conversational experiences.

From a technical standpoint, the thesis would delve into optimizing wake word detection for voice-based technologies like Reverb, designed to facilitate interaction during passive activities such as cooking or driving. The current technology, while robust, relies on a tech stack that includes Whisper API for user prompt understanding, vector embeddings for cross-podcast referencing, Llama Index for context retention, GPT API for information generation, and Eleven Labs API for voice synthesis. This setup, however, results in a 15-20 seconds delay in generating responses. Future research would focus on refining these technologies to achieve near real-time conversational flow.

A significant area of future exploration would be the integration of this technology into established podcast platforms like Spotify and Apple Podcasts, as opposed to a standalone web interface. This integration would utilize existing databases of podcasters, subject to their consent for voice cloning, thereby creating digital AI assistants for an enhanced and engaging podcast experience. Furthermore, the thesis suggests a potential framework where podcasters actively participate in the content generation process, thereby ensuring the accuracy and relevance of AI-generated content.

Conclusion

The culmination of this thesis demonstrates a significant leap in enhancing the interactivity of audio media consumption, particularly podcasts, through the innovative application of AI technologies. “Reverb,” the designed system, successfully transforms traditional, passive listening experiences into dynamic, interactive dialogues, where users can engage directly with the content. “Reverb,” was subjected to user testing with 20 individuals, yielding a high satisfaction rate of 86%. The primary feedback from users emphasized the need for integration with existing podcast platforms. This suggests a strong preference for a seamless and unified listening experience, where users can leverage the interactive features of “Reverb” within the familiar environments of platforms they already use. This feedback is invaluable for future enhancements, underscoring the importance of compatibility and ease of use in the broader adoption of such innovative technologies.By integrating advanced AI tools for voice cloning, context retention, and prompt processing, the project not only addresses the prevailing gap in conversational media but also pioneers a new paradigm in digital audio interaction. The ethical considerations and technical challenges encountered underscore the need for continual development and refinement in this domain. The future potential of this technology is vast, with implications for education, entertainment, and beyond. This work lays a foundational framework for future explorations in making digital media consumption more engaging, interactive, and personalized, steering towards a future where technology and human curiosity intersect more seamlessly.


REVERB
Information at your fingertips

Bibliography

1. How to Make Your Podcast Global and Interactive with AI | LinkedIn. https://www.linkedin.com/pulse/how-make-your-podcast-global-interactive-ai-junaid-awan/. Accessed 8 Oct. 2023.
2. Aliotta, Chiara. “Design the Future through the Power of Storytelling.” Medium, 10 Nov. 2023, https://chiara-aliotta.medium.com/design-the-future-through-the-power-of-storytelling-5cc7a1a329c9.
3. Berry, Richard. “A Golden Age of Podcasting? Evaluating Serial in the Context of Podcast Histories.” Journal of Radio & Audio Media, vol. 22, no. 2, 2015, pp. 170–78. search.library.berkeley.edu, https://doi.org/10.1080/19376529.2015.1083363.
4. Best Ai Podcasting Tools to Streamline Your Workflow (2023). https://riverside.fm/blog/ai-podcasting-tools. Accessed 8 Oct. 2023.
5. Carman, Ashley. “The next Big Thing in Podcasts Is Talking Back.” The Verge, 12 Oct. 2021, https://www.theverge.com/2021/10/12/22722468/spotify-amazon-facebook-audio-podcast-polls-interact.
6. Carr, Nicholas G. The Shallows: What the Internet Is Doing to Our Brains. Second edition., W.W. Norton & Company, 2020.
7. “ChatGPT and the Future of AI-Generated Content for Podcasts.” AIContentfy, 27 Jan. 2023, https://aicontentfy.com/en/blog/chatgpt-and-future-of-ai-generated-content-for-podcasts.
8. Chiong, Cynthia, and Judy S. DeLoache. “Learning the ABCs: What Kinds of Picture Books Facilitate Young Children’s Learning?” Journal of Early Childhood Literacy, vol. 13, no. 2, June 2013, pp. 225–41. SAGE Journals, https://doi.org/10.1177/1468798411430091.
9. Daniel, Silveira. An Empathetic Design Framework for Humanity-Centered AI: A Preventative Approach to Developing More Holistic, Reliable, and Ethical ML Products. OCAD University, 2 May 2023, https://openresearch.ocadu.ca/id/eprint/4044/.
10. Dennis. “Engaging Podcast Content: 13 Tips to Create Better Content.” Castos, 1 Jan. 2019, https://castos.com/engaging-podcast-content/.
11. “Design Foresight: A Design Approach That Marries the Futurization and De-Futurization * Journal of Futures Studies.” Journal of Futures Studies, 1 Nov. 2023, https://jfsdigital.org/design-foresight-a-design-approach-that-marries-the-futurization-and-de-futurization/.
12. Donath, Judith S. “Anthropomorphic Visualization: Depicting Participants in Online Spaces Using the Human Form.” MIT Media Lab, https://www.media.mit.edu/publications/anthropomorphic-visualization-depicting-participants-in-online-spaces-using-the-human-form/. Accessed 11 Oct. 2023.
13. Dykstra, Maria. “Supercharge Your Podcast Success with AI: How to Use Chat GPT.” TreDigital, 20 Apr. 2023, https://tredigital.com/supercharge-your-podcast-success-how-to-use-chatgpt-to-boost-your-content-strategy-examples/.

14. “Electronic Monuments.” University of Minnesota Press, https://www.upress.umn.edu/book-division/books/electronic-monuments. Accessed 20 Nov. 2023.
15. “---.” University of Minnesota Press, https://www.upress.umn.edu/book-division/books/electronic-monuments. Accessed 8 Oct. 2023.
16. Free AI Voice Cloning: Clone Your Voice In 30 Seconds! 4 Oct. 2023, https://speechify.com/voice-cloning/.
17. Gabon, Alain. “Review of Heuretics: The Logic of Invention.” SubStance, vol. 25, no. 1, 1996, pp. 146–51. JSTOR, https://doi.org/10.2307/3685243.
18. Gandhi, Santhosh. “Futures Thinking and Design Thinking Simply Explained!” Medium, 22 Mar. 2022, https://bootcamp.uxdesign.cc/future-thinking-and-design-thinking-simply-explained-d65716d67651.
19. Geurts, Amber, et al. “New Perspectives for Data-Supported Foresight: The Hybrid AI-Expert Approach.” FUTURES & FORESIGHT SCIENCE, vol. 4, no. 1, 2022, p. e99. Wiley Online Library, https://doi.org/10.1002/ffo2.99.
20. Goldberg, Cait. “THE NEW BRAIN: How the Modern Age Is Rewiring Your Mind.” Science News, vol. 167, no. 1, 1 Jan. 2005, p. 15.
21. “GPT4ALL-Voice-Assistant/Main.Py at Main · Ai-Austin/GPT4ALL-Voice-Assistant.” GitHub, https://github.com/Ai-Austin/GPT4ALL-Voice-Assistant/blob/main/main.py. Accessed 20 Nov. 2023.
22. Graham, Amara. “Capturing Audio in the Browser For ‘Wake Words.’” Voice Tech Podcast, 29 Apr. 2019, https://medium.com/voice-tech-podcast/capturing-audio-in-the-browser-for-wake-words-cc4972263773.
23. Gregoire, Henri, et al. An Enquiry Concerning the Intellectual and Moral Faculties and Literature of Negroes. Taylor & Francis Group, 1996.
24. Gregory L. Ulmer. https://english.ufl.edu/gregory-l-ulmer/. Accessed 8 Oct. 2023.
25. Hobart, Michael E. “Review of The Information: A History, a Theory, a Flood.” Technology and Culture, vol. 55, no. 2, 2014, pp. 489–90.
26. ---. “The Information: A History, a Theory, a Flood by James Gleick (Review).” Technology and Culture, vol. 55, no. 2, 2014, pp. 489–90. search.library.berkeley.edu, https://doi.org/10.1353/tech.2014.0045.
27. How Can Podcasters Use AI & ChatGPT to Their Advantage? https://www.thepodcasthost.com/business-of-podcasting/ai-podcasting/. Accessed 8 Oct. 2023.
28. How to Make a Podcast | Descript. https://www.descript.com/podcasting. Accessed 8 Oct. 2023.
29. How To Use AI To Create Podcasts Speechify. 11 Jan. 2023, https://speechify.com/blog/use-ai-create-podcasts/.
30. “Introduction.” ElevenLabs, https://elevenlabs.io/docs/api-reference/introduction. Accessed 20 Nov. 2023.

31. Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement - Yiyi Li, Ying Xie, 2020. https://journals.sagepub.com/doi/10.1177/0022243719881113. Accessed 11 Oct. 2023.
32. Knibbs, Kate. “Generative AI Podcasts Are Here. Prepare to Be Bored.” Wired. www.wired.com, https://www.wired.com/story/generative-ai-podcasts-boring/. Accessed 8 Oct. 2023.
33. Laban, Philippe, et al. “NewsPod: Automatic and Interactive News Podcasts.” 27th International Conference on Intelligent User Interfaces, Association for Computing Machinery, 2022, pp. 691–706. ACM Digital Library, https://doi.org/10.1145/3490099.3511147.
34. Landa, Jose Angel Garcia. “Linkterature: From Word to Web Or: Literature in the Internet - Internet as Literature - Literature as Internet - Internet in Literature.” SSRN Electronic Journal, 2006. www.academia.edu, https://www.academia.edu/104883490/Linkterature_From_Word_to_Web_Or_Literature_in_the_Internet_Internet_as_Literature_Literature_as_Internet_Internet_in_Literature.
35. Larsen, Nicole E., et al. “Do Storybooks with Anthropomorphized Animal Characters Promote Prosocial Behaviors in Young Children?” Developmental Science, vol. 21, no. 3, 2018, p. e12590. Wiley Online Library, https://doi.org/10.1111/desc.12590.
36. Levkowitz, H. “The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think.” Choice, vol. 50, no. 2, 2012, pp. 317-.
37. LlamaIndex 0.9.4. https://docs.llamaindex.ai/en/stable/. Accessed 20 Nov. 2023.
38. Malcolm Gladwell Is Lord Of All Things Overlooked and Misunderstood | Rich Roll Podcast - YouTube. https://www.youtube.com/watch?v=agyic7NFtRQ&ab_channel=RichRoll. Accessed 20 Nov. 2023.
39. McCormick, Samuel. “The Chattering Mind: A Conceptual History of Everyday Talk.” The Chattering Mind, University of Chicago Press, 2020. www-degruyter-com.libproxy.berkeley.edu, https://www.degruyter.com/document/doi/10.7208/9780226677804/html.
40. Metz, Cade. “Google Made a Chatbot That Debates the Meaning of Life.” Wired. www.wired.com, https://www.wired.com/2015/06/google-made-chatbot-debates-meaning-life/. Accessed 8 Oct. 2023.
41. Miller, Stephen, and Carol V. Wright. Conversation: A History of a Declining Art. Yale University Press, 2006. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/berkeley-ebooks/detail.action?docID=3420089.
42. Moore, Schuyler. “Who Owns Voice And Image Artificial Intelligence Rights?” Forbes, https://www.forbes.com/sites/schuylermoore/2022/10/28/who-owns-voice-and-image-artificial-intelligence-rights/. Accessed 20 Nov. 2023.
43. Naomi. “Podcasts Get Interactive With New Q&A and Polls

Features." Spotify, 30 Sept. 2021, https://newsroom.spotify.com/2021-09-30/podcasts-get-interactive-with-new-qa-and-polls-features/.
44. ---. "Spotify's AI Voice Translation Pilot Means Your Favorite Podcasters Might Be Heard in Your Native Language." Spotify, 25 Sept. 2023, https://newsroom.spotify.com/2023-09-25/ai-voice-translation-pilot-lex-fridman-dax-shepard-steven-bartlett/.
45. Nickolaisen, Michelle. "The Ultimate Guide to Becoming a Better Podcast Listener." Medium, 19 Mar. 2018, https://medium.com/@_chelleshock/the-ultimate-guide-to-becoming-a-better-podcast-listener-9fcdef3c8c05.
46. OpenAI Platform. https://platform.openai.com. Accessed 20 Nov. 2023.
47. ---. https://platform.openai.com. Accessed 20 Nov. 2023.
48. ---. https://platform.openai.com. Accessed 20 Nov. 2023.
49. ---. https://platform.openai.com. Accessed 20 Nov. 2023.
50. ---. https://platform.openai.com. Accessed 20 Nov. 2023.
51. Overview ‹ Theme | Human-AI Interaction — MIT Media Lab. https://www.media.mit.edu/projects/theme-virtual-humans/overview/. Accessed 8 Oct. 2023.
52. Ph.D, Riza C. Berkan. "Interactive AI Podcasting Debut." Medium, 9 June 2020, https://medium.com/@rizaberkan/interactive-ai-podcasting-debut-a4aa88d7ca07.
53. Pinna. Pinna Launches the First Voice Activated Interactive Podcasts. https://www.prnewswire.com/news-releases/pinna-launches-the-first-voice-activated-interactive-podcasts-301567317.html. Accessed 8 Oct. 2023.
54. Podcast Statistics and Trends (& Why They Matter). https://riverside.fm/blog/podcast-statistics. Accessed 8 Oct. 2023.
55. Podcast.Ai. https://podcast.ai/. Accessed 8 Oct. 2023.
56. Randall, David. The Concept of Conversation: From Cicero's Sermo to the Grand Siècle's Conversation. Edinburgh University Press, 2018. JSTOR, https://www.jstor.org/stable/10.3366/j.ctt1tqxvh0.
57. Rhetoric & Apparatus Theory | Writing Commons. 7 Jan. 2023, https://writingcommons.org/section/rhetoric/rhetoric-apparatus-theory/.
58. Science Gallery Atlanta and WABE Unveil New Interactive Podcast, 'Calls for Justice' | Emory University | Atlanta GA. https://news.emory.edu/stories/2023/08/er_science_gallery_atlanta_podcast_28-08-2023/story.html. Accessed 8 Oct. 2023.
59. Sokołowski, Filip. "Generative Genius: I Brought Steve Jobs Back to Life (and You Can Too)." Medium, 22 Feb. 2023, https://medium.com/@fillsoko/generative-genius-i-brought-steve-jobs-back-to-life-and-you-can-too-f94330911a31.
60. "Soulless Podcast Advice." Soulless Podcast Advice, https://soullesspodcastadvice.com/. Accessed 8 Oct. 2023.
61. Speech Synthesis: Generate AI Audio & Voiceovers. https://

elevenlabs.io/speech-synthesis. Accessed 20 Nov. 2023.
62. Team, Gradio. Gradio. https://gradio.app. Accessed 20 Nov. 2023.
63. “The Future of Podcasting with Artificial Intelligence.” Podcastle Blog, 24 Jan. 2023, https://podcastle.ai/blog/ai-in-podcasting/.
64. The next Big Thing in Podcasts Is Talking Back - The Verge. https://www.theverge.com/2021/10/12/22722468/spotify-amazon-facebook-audio-podcast-polls-interact. Accessed 8 Oct. 2023.
65. The Psychology of Conversation - Research Summary - Faculty & Research - Harvard Business School. https://www.hbs.edu/faculty/Pages/item.aspx?research=7741. Accessed 20 Nov. 2023.
66. Three Reasons Podcast Creatives Should Embrace, Not Fear, AI – Sounds Profitable. https://soundsprofitable.com/article/three-reasons-podcast-creatives-should-embrace-not-fear-ai/. Accessed 8 Oct. 2023.
67. Turkle, Sherry. Reclaiming Conversation: The Power of Talk in a Digital Age. Penguin Press, 2015.
68. Ulmer, Gregory. The Learning Screen From Networked Book. www.academia.edu, https://www.academia.edu/37590082/The_Learning_Screen_From_Networked_Book. Accessed 8 Oct. 2023.
69. Ulmer, Gregory L. Applied Grammatology : Post(e)-Pedagogy from Jacques Derrida to Joseph Beuys. Johns Hopkins University Press, 2019. directory.doabooks.org, https://doi.org/10.1353/book.67868.
70. ---. Heuretics: The Logic of Invention. Johns Hopkins University Press, 1994.
71. Urquiza-Haas, Esmeralda G., and Kurt Kotrschal. “The Mind behind Anthropomorphic Thinking: Attribution of Mental States to Other Species.” Animal Behaviour, vol. 109, Nov. 2015, pp. 167–76. ScienceDirect, https://doi.org/10.1016/j.anbehav.2015.08.011.
72. Webb, Amy. “How to Prepare for a GenAI Future You Can’t Predict.” Harvard Business Review, 31 Aug. 2023. hbr.org, https://hbr.org/2023/08/how-to-prepare-for-a-genai-future-you-cant-predict.
73. “What Spotify and Apple Can Learn from Chinese Podcasting Apps.” Rest of World, 8 Sept. 2021, https://restofworld.org/2021/what-spotify-and-apple-can-learn-from-chinese-podcasting-apps/.
74. Wohr, James. “Voice Assistants: What They Are and What They Mean for Marketing and Commerce.” Insider Intelligence, https://www.insiderintelligence.com/insights/voice-assistants/. Accessed 20 Nov. 2023.
75. Zeldin, Theodore. Conversation. HiddenSpring, 2000.

Disclosure
I used ChatGPT for copy-editing the first draft of this section and to brainstorm topic titles given my set of references.