

# Applications of AI and NLP to advance Music Recommendations on Voice Assistants

Ashlesha V Kadam

Amazon, Seattle WA, USA.

Received: 30 Apr 2023; Received in revised form: 26 May 2023; Accepted: 03 Jun 2023; Available online: 10 Jun 2023

©2023 The Author(s). Published by AI Publications. This is an open access article under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>)

**Abstract**— This paper builds on the popular use case of music requests on voice assistants like Siri, Google Assistant, Alexa, and others and explores the different AI and NLP techniques. The paper particularly focuses on how each of these techniques can be applied in the context of musical recommendations and experiences on voice assistants. It enumerates specific problems in the space of music recommendations and illustrates how specific techniques like multi-armed bandits can be applied.

**Keywords**— applications of AI, music, recommendations, voice assistants, voice technology.

## I. INTRODUCTION

Voice Assistants like Siri, Google Assistant, Alexa and others have become ubiquitous, and using them for tasks like listening to music, getting the latest news and performing simple tasks has become commonplace. A natural use case on voice assistants is to use them to play music, leading to the pursuit of personalized and immersive music experiences on voice assistants reaching new heights. Leveraging the power of Artificial Intelligence (AI) and Natural Language Processing (NLP), we are seeing groundbreaking advancements that are revolutionizing music recommendations on voice assistants. In this article, we explore the cutting-edge research and frameworks that underpin these advancements. We will refer to a hypothetical voice assistant – Nova – throughout this article for illustrative purposes.

## II. EVOLUTION OF RECOMMENDATIONS ON VOICE ASSISTANTS

What started as simple rule-based systems (e.g. when a user asks Nova to play pop music, Nova might every time start with playing a static ‘90s pop playlist that someone has curated, followed by a 2000s pop playlist and then back to the ‘90s one and so on), followed by collaborative filtering approaches (e.g. if Nova uses the logic of “users who like artist X’s music also like artist Y’s music” to

serve Y’s music to fans of X’s), personalized music recommendations have been continuously refined. But with the emergence of advanced techniques in AI and NLP, music recommendations on voice assistants have become more dynamic, hyper personalized and context-aware.

## III. APPLYING AI TECHNIQUES IN VOICE ASSISTANTS FOR MUSIC

Below are AI techniques, their brief overview and how each of them can be applied to the use case of music on voice assistants.

### 1.1 Deep Learning Architectures

Deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers have demonstrated success in music recommendation tasks.

1.1.1 CNNs can extract meaningful features from spectrograms, capturing time-frequency characteristics of music. Through a series of [1] convolutional layers, these networks learn hierarchical representations, enabling them to recognize complex patterns in music data. *Application:* These CNN-based models have been successful in tasks like music genre classification, mood analysis, and similarity-based recommendation systems. For example, Nova might be able to analyze Taylor Swift’s song, Evermore, to be more pensive but her song Me to be more

fun and upbeat. Similarly, Nova might be able to understand a user’s musical taste by knowing their liked tracks and recommend tracks that are similar to those liked tracks to help users discovery music that’s new to them. Or Nova might be able to leverage CNNs to ingest the music catalog with its meta data including genre information, where the CNN learns genre-specific features and maps those features to the genres, so it can classify tracks into different genres. See Fig. 1 for an illustrative example.

1.1.2 RNNs perform well at modeling sequential data [2]. The recurrent nature of these networks allows them to capture temporal dependencies and long-term dependencies within music, enabling them to understand musical context. RNN variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), can effectively capture patterns in rhythms, enabling building of more nuanced and context-aware recommendations. *Application:* These networks can learn from past sequences of music and generate predictions for the next item in a playlist or suggest music based on previous listening history. For example, Nova might sense that if you have skipped a few metal songs that you don’t like metal and not queue those up for listening. Or if you have been exploring a genre for the first time, like hip-hop, even if your go-to genres have been country and jazz, Nova might suggest a hip-hop album when you request for some music.

1.1.3 Transformer models, which were first introduced for NLP tasks, such as the widely known BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformers), have advanced the understanding of context and semantics in music data. Using self-attention mechanisms, transformers can capture global dependencies, allowing for more comprehensive music representations. These models excel at understanding user queries, deciphering user intent, and contextualizing music recommendations based on the broader context. *Applications:* Improved semantic understanding can allow for Nova to recommend meaningful music for requests of the shape “upbeat hip-hop music for workout” instead of literally searching for a song or artist with the name “upbeat hip-hop”. Similarly, if a user asks Nova for “some relaxing songs to unwind on a Friday evening”, these models can help leverage this contextual information to recommend the “right” music.

## 1.2 Knowledge Graphs

1.2.1 A knowledge graph [3] is a graph-based data structure that organizes information. In the context of music recommendations, knowledge graphs might capture the relationships between artists, albums, genres, user

preferences, and other music-related entities. Knowledge graphs can be used to encode semantic relationships, allowing deeper and nuanced understanding of music entities to provide recommendations. *Application:* Knowledge graphs can identify that artists X and Y are similar and hence fans of X might enjoy music by Y too. For example, Nova can help users discover new music that is relevant to them by deriving relationships like “artist X is influenced by artist Y” or “album A is similar to album B”.

## 1.3 Reinforcement Learning

1.3.1 These are some of the most powerful models, using feedback from the users to fine-tune recommendations. These models can also strike a balance between explore and exploit to keep trying to identify the best recommendation for a user in the given context while also not trapping them in their musical taste bubble. Within RL, multi-armed bandits (MABs) are an application that can be effective for music recommendations [4]. Finally, RLs can also help with optimization of multiple objectives while choosing recommendations. *Application:* These models can help continuously identify the best recommendation for a given customer intent and given context. For example, if a user simply asks Nova to play music, knowing when to play upbeat hip-hop vs. when to play focus music vs. when to play mellow folk music. Or if a user pulls up the screen of Nova, they might get different recommendations on the screen that help satisfy different stakeholder objectives, like showcasing a mix of both personalized and promoted music, ads and organic content, and more. Fig. 2 shows how the MAB chooses a recommendation and keeps improving its recommendations based on user feedback to served recommendations.

## IV. FIGURES AND TABLES

To ensure a high-quality product, diagrams and lettering MUST be either computer-drafted or drawn using India ink.

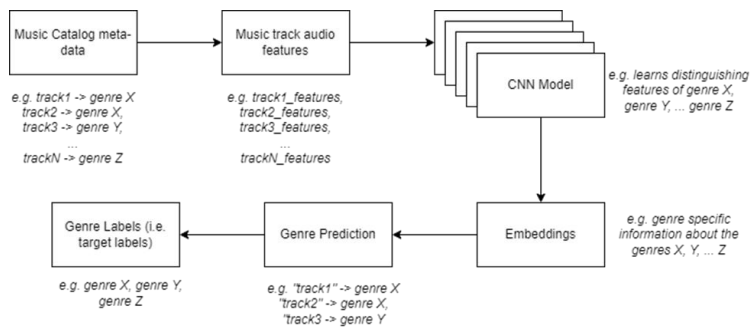


Fig. 1: Genre classification process using CNNs

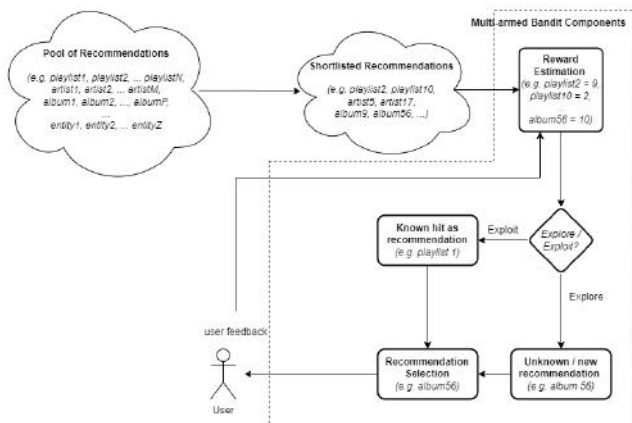


Fig. 2: Use of multi-armed bandit to serve a user request on Nova

## V. CONCLUSION

The sections above cover the most prominent AI and NLP techniques along with their specific applications in the context of music recommendations on voice assistants. As AI advances, our understanding of the users' preferences, along with better contextual and catalog metadata understanding, will allow us to further revolutionize this field.

## REFERENCES

- [1] Yandre M.G. Costa, Luiz S. Oliveira, Carlos N. Silla, "An evaluation of Convolutional Neural Networks for music classification using spectrograms" Applied Soft Computing, Volume 52, 2017, Pages 28-38
- [2] Priyank Jaini, Zhitang Chen, Pablo Carbajal, Edith Law, Laura Middleton, Kayla Regan, Mike Schaekermann, George Trimponias, James Tung, Pascal Poupart "Online Bayesian Transfer Learning for Sequential Data Modeling", Published as a conference paper at ICLR 2017
- [3] Mayank Kejriwal, "Knowledge Graphs: A Practical Review of the Research Landscape", Special Issue Knowledge Graph Technology and Its Applications
- [4] Chunqiu Zeng, Qing Wang, Shekoofeh Mokhtari, Tao Li, "Online Context-Aware Recommendation with Time Varying Multi-Armed Bandit"