Text2Playlist: Generating Personalized Playlists from Text on a Music Streaming Platform

Mathieu Delcluze
Antoine Khoury
Clémence Vast
Valerio Arnaudo
Léa Briand
Walid Bendada
Thomas Bouabça
Deezer
Paris, France
research@deezer.com

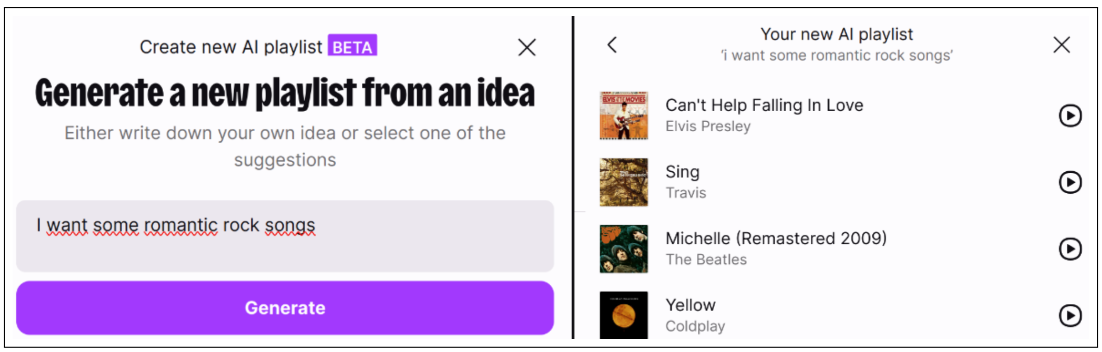

Figure 1: Interface of the "Text2Playlist" tool on the website version of Deezer, creating a personalized playlist generated from an idea given by a text from a Deezer user.

Abstract

The streaming service Deezer heavily relies on the search to help users navigate through its extensive music catalog. Nonetheless, it is primarily designed to find specific items and does not lead directly to a smooth listening experience. We present Text2Playlist, a stand-alone tool that addresses these limitations. Text2Playlist leverages generative AI, music information retrieval and recommendation systems to generate query-specific and personalized playlists, successfully deployed at scale.

CCS Concepts

• **Information systems** → *Retrieval tasks and goals*; *Recommender systems*; *Personalization.*

Keywords

Generative AI, Music Information Retrieval, Real-World Deployment

ACM Reference Format:
Mathieu Delcluze, Antoine Khoury, Clémence Vast, Valerio Arnaudo, Léa Briand, Walid Bendada, and Thomas Bouabça. 2025. Text2Playlist: Generating Personalized Playlists from Text on a Music Streaming Platform. In *Nineteenth ACM Conference on Recommender Systems (RecSys '25).* ACM, New York, NY, USA, 4 pages.

1 Introduction

Search engines of online content platforms are essential to explore large catalogs of items [23]. Traditionally, search systems have been optimized for *narrow* intent queries, where users have a *focus* mindset aimed at a *navigational* goal [16, 35], i.e., looking for specific entities such as music tracks, products and books [2, 12, 24, 33]. In contrast, users with an *exploratory* mindset, aimed at an *informational* goal, use *broad* intent queries (e.g., in music domain "Chill vibes on a rainy afternoon") [27, 35, 36]. In particular, the search engine of the French music streaming service Deezer helps 16 million users from 180 countries access more than 120 millions of music tracks. Despite supporting both intent queries, its design prioritizes navigation: the small tool bar does not encourage long queries and search results cannot be easily transformed into playlists.

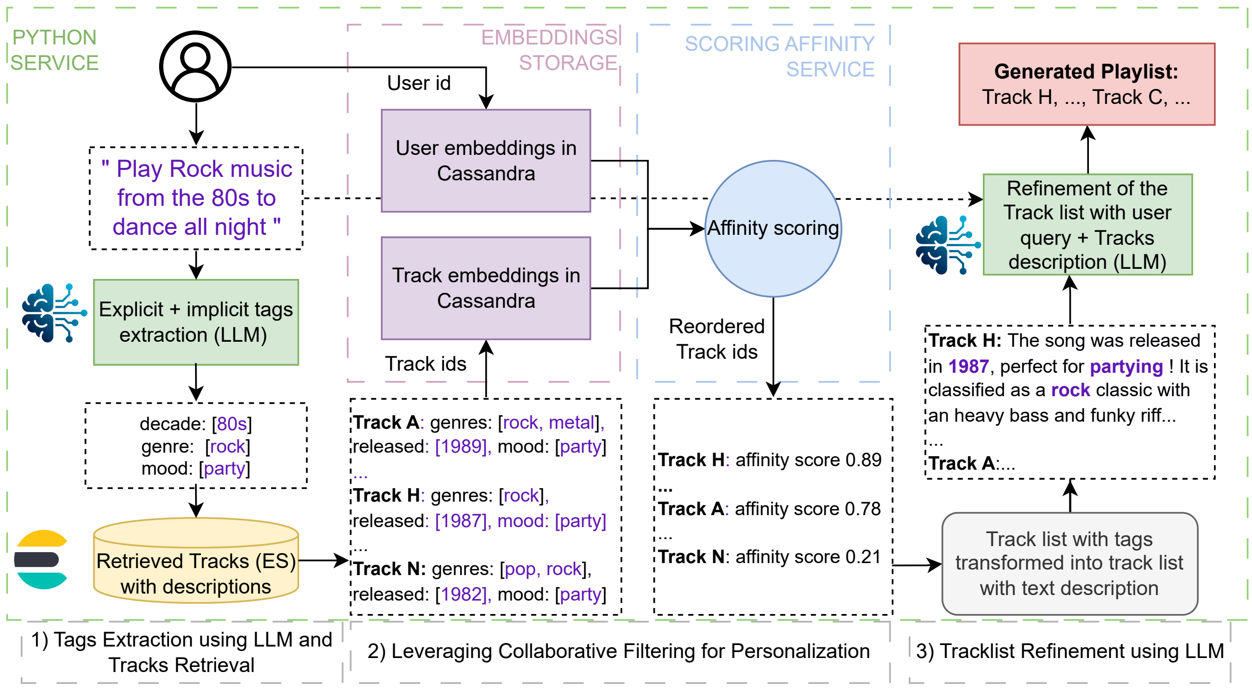

Figure 2: Overview of the Text2Playlist framework from Section 3, available for online requests in our production environment on Deezer.

In this paper, we present Text2Playlist, illustrated in Figure 1, a personalized playlist creation tool, explicitly dedicated for broad intent queries, distinct from the search feature. It takes advantage of the recent rise of Large-Language Models (LLMs) [34, 39, 40] and gets inspiration from Retrieval-Augmentation Generation (RAG) framework [26]. Text2Playlist has been deployed on Deezer mobile and web applications for a first test phase of 5% of premium users since July 2024 and 20% since October 2024. This paper is organized as follows. In Section 2, we further detail our motivations for such a system. In Section 3 we detail how we use together query extraction, various metadata sources, Deezer recommendation system and LLMs to build this solution. In Section 4, we explain how Text2Playlist is deployed on Deezer and present analysis based on the first data gathered. We conclude and discuss areas of improvement in Section 5.

2 Where to Explore and Save Music on Deezer

To explore the Deezer catalog, users can either rely on recommendation features or use the search function for more targeted music content. In this paper, we will focus on this latter way to explore and save music. Intent of the query - narrow vs broad - is identified by the search system thanks to a model combining search patterns and item clicks among historical queries [27, 35]. On Deezer, narrow queries comprise 80% of the total; they focus on providing the specific music entity specified in the query - it can be tracks, artists or albums. Although the remaining 20% of broad queries are less numerous and more challenging to satisfy, they are essential to address as they foster further catalog exploration and enrichment of personal libraries, both strong signals of engagement on the platform [16]. Despite these promising results, the current search bar is intentionally primarily designed for narrow queries and is not adapted to long, complex - broad queries. Besides, it is not straightforward to transform search results into a seamless music experience. Indeed, before listening to the search output, users must manually sort content into playlists, which can be time-intensive.

3 Text2Playlist, a Playlist Generation Tool from Text

Developed in 2024, the Text2Playlist engine, illustrated in Figure 1, aims to address these limitations. It encourages users to write their music needs and generates a personalized playlist, tailored to the query. Figure 2 provides a summary of Text2Playlist system in production, further described in this section.

3.1 Tags Extraction using LLM and Tracks Retrieval

Tags are often characterized as keywords to describe key information (e.g., for music items we can talk about music genre, decade, mood, artist gender, language...). They are useful to boost relevance matching, help query reformulation and item recommendation [25, 28, 30]. A query may contain information that is explicitly or implicitly expressed. For example, in the query “I want music from the 90s for work”, we can extract the explicit decade tag “90s” named by the user, but we can also infer they may prefer "Focus" mood tracks to be able to work during their listening session. Thanks to a LLM, we deduce both explicit and implicit tags from the query [1, 20, 32]. Then we leverage tags already existing to describe the Deezer catalog. It comes from manual annotations from music experts, but also various internal models relying on audio content analysis and user-made playlists to expand the coverage [6, 9–11, 13, 15, 17–19, 31]. Tracks matching the extracted tags from the query using the LLM are retrieved and factored into JSON format.

3.2 Leveraging Collaborative Filtering for Personalization

Most of our personalized recommender systems [3, 4, 6–8] leverage latent models for Collaborative Filtering (CF) [5, 22]. By analyzing usage data on Deezer, they learn low-dimensional *embedding* vector representations of users and tracks, in a vector space where proximity reflects preferences. Then, they offer recommendations based on embedding similarity metrics [6, 8]. After computing the cosine similarities between the user and the tracks obtained in 3.1, the list is reordered in descending order of similarity, first tracks being the closest to the user profile.

3.3 Tracklist Refinement using LLM

Inspired by two-stage recommender systems [38] and RAG technique [26], a LLM is applied on the list of tracks and tags beforehand transformed into an unstructured text, to prioritize tracks that best match the original query. Additional rules such as artists diversity and overall quality of the playlist are also mentioned in the LLM prompt to optimize the final user experience.

4 Deploying Text2Playlist on Deezer

4.1 System Deployment

From a technical standpoint, this system was designed as a standalone framework, written in Python and running on a Kubernetes cluster. The LLM used for tags extraction (3.1) and playlist refinement (3.3) is Gemini Flash 1.5 [37]. This choice was mainly driven by cost considerations, as Gemini Flash costs are calculated based on the number of input and output tokens. Therefore, there is no costs incurred when there is no usage, which is beneficial when gradually rolling out such a feature that may not be in constant use. User and song-related data (3.1), including CF embedding vectors, mood scores and other catalog information, are exported daily in a Cassandra cluster. To retrieve the tracks matching the extracted tags we use Elasticsearch [14]. For user-song affinity scoring (3.2), we use a Golang application incorporating the Faiss library [21].

4.2 Empirical Analysis

After weeks of internal tests, Text2Playlist was released in July 2024 to 5% of premium users on mobile and web. Since October 2024, it has been expanded to 20% with success, which is promising for its wider roll-out. One key metric we analyze to gauge user satisfaction is the proportion of playlists generated by the feature that are

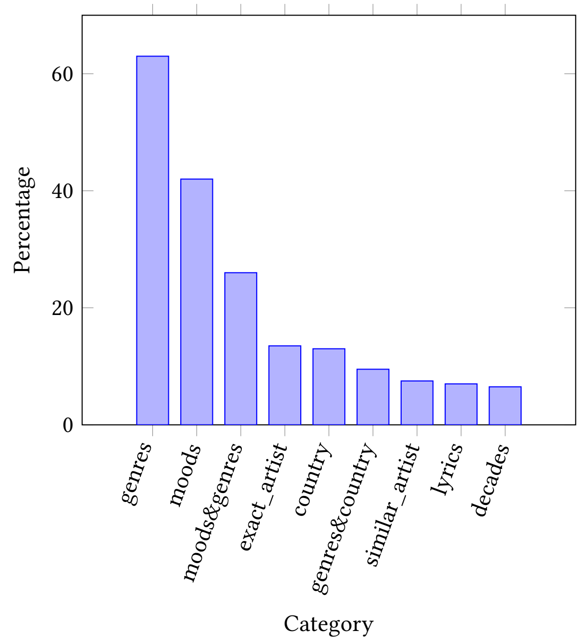

Figure 3: Percentage of user queries specifying each category

listened to in the following days: while it occurs for 27% of manual playlists, it is 45% for the generated playlists by Text2Playlist, indicating a positive engagement with the feature.

One of the main motivations behind this work was to analyse real data to better understand how users explore our music catalog via natural language queries. When we look at the usage of the Text2Playlist tool, it shows that over 62% of the queries include *genre* tags, making this the primary driver of the search behavior. Combined or not, genres and moods capture the majority of the user intent, while other tags, like decades or lyrics, are less frequent.

Lastly, we report most frequent tags of moods asked by the users are "Chill" and "Party", representing even nearly half of the requested moods. These observations provide valuable insights illustrating what kind of music users want through broad queries on Deezer over time.

Note: our experiments will be further illustrated and discussed in the RecSys Workshop EARL associated with this article. Resources related to this talk will be available on: https://github.com/deezer/text2playlist-recsys2025.

5 Conclusion

In this paper, we presented Text2Playlist, a stand-alone tool designed for generating personalized playlists from text at scale. Text2Playlist feature was successfully deployed on the music streaming service Deezer in 2024. Beyond demonstrating promising performance, this system provides valuable insights about users' music needs and how they articulate them (e.g., what are the most popular moods recognized by the LLM in the queries 4.2?). Our team also plans to increase the coverage and diversity of the extracted tags (e.g., could we use LLM or lyrics to enrich even more music representation [10]?). Besides, as shown in 4.2, users often need to reformulate queries: with the surge of assistant-driven interactions, we could refactor Text2Playlist into a conversational tool [29].

References

[1] Xiao Bai, Xue Wu, Ivan Stojkovic, and Kostas Tsioutsiouliklis. 2024. Leveraging Large Language Models for Improving Keyphrase Generation for Contextual Targeting. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management.* 4349–4357.

[2] David Bainbridge, Sally Jo Cunningham, and J Stephen Downie. 2003. How people describe their music information needs: A grounded theory analysis of music queries. (2003).

[3] Walid Bendada, Théo Bontempelli, Mathieu Morlon, Benjamin Chapus, Thibault Cador, Thomas Bouabça, and Guillaume Salha-Galvan. 2023. Track Mix Generation on Music Streaming Services using Transformers. In *Proceedings of the 17th ACM Conference on Recommender Systems.* 112–115.

[4] Walid Bendada, Guillaume Salha, and Théo Bontempelli. 2020. Carousel Personalization in Music Streaming Apps with Contextual Bandits. In *Proceedings of the 14th ACM Conference on Recommender Systems.* 420–425.

[5] Dheeraj Bokde, Sheetal Girase, and Debajyoti Mukhopadhyay. 2015. Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey. *Procedia Computer Science* 49 (2015), 136–146.

[6] Théo Bontempelli, Benjamin Chapus, François Rigaud, Mathieu Morlon, Marin Lorant, and Guillaume Salha-Galvan. 2022. Flow Moods: Recommending Music by Moods on Deezer. In *Proceedings of the 16th ACM Conference on Recommender Systems.* 452–455.

[7] Léa Briand, Théo Bontempelli, Walid Bendada, Mathieu Morlon, François Rigaud, Benjamin Chapus, Thomas Bouabça, and Guillaume Salha-Galvan. 2024. Let's Get It Started: Fostering the Discoverability of New Releases on Deezer. In *European Conference on Information Retrieval.* Springer, 286–291.

[8] Léa Briand, Guillaume Salha-Galvan, Walid Bendada, Mathieu Morlon, and Viet-Anh Tran. 2021. A Semi-Personalized System for User Cold Start Recommendation on Music Streaming Apps. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.* 2601–2609.

[9] Jeong Choi, Anis Khlif, and Elena Epure. 2020. Prediction of user listening contexts for music playlists. In *Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA).* 23–27.

[10] Rémi Delbouys, Romain Hennequin, Francesco Piccoli, Jimena Royo-Letelier, and Manuel Moussallam. 2018. Music mood detection based on audio and lyrics with deep neural net. *arXiv preprint arXiv:1809.07276* (2018).

[11] Elena V Epure, Guillaume Salha, and Romain Hennequin. 2020. Multilingual music genre embeddings for effective cross-lingual music item annotation. *arXiv preprint arXiv:2009.07755* (2020).

[12] Jean Garcia-Gathright, Brian St. Thomas, Christine Hosey, Zahra Nazari, and Fernando Diaz. 2018. Understanding and evaluating user satisfaction with music discovery. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval.* 55–64.

[13] Yuan Gong, Yu-An Chung, and James Glass. 2021. Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778* (2021).

[14] Clinton Gormley and Zachary Tong. 2015. *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine.* " O'Reilly Media, Inc.".

[15] Romain Hennequin, Jimena Royo-Letelier, and Manuel Moussallam. 2018. Audio based disambiguation of music genre tags. *arXiv preprint arXiv:1809.07256* (2018).

[16] Christine Hosey, Lara Vujović, Brian St. Thomas, Jean Garcia-Gathright, and Jennifer Thom. 2019. Just give me what I want: How people use and evaluate music search. In *Proceedings of the 2019 chi conference on human factors in computing systems.* 1–12.

[17] Karim M Ibrahim, Elena V Epure, Geoffroy Peeters, and Gael Richard. 2020. Should we consider the users in contextual music auto-tagging models?. In *21st International Society for Music Information Retrieval Conference.*

[18] Karim M Ibrahim, Elena V Epure, Geoffroy Peeters, and Gael Richard. 2022. Exploiting device and audio data to tag music with User-Aware listening contexts. *arXiv preprint arXiv:2211.07250* (2022).

[19] Karim M Ibrahim, Jimena Royo-Letelier, Elena V Epure, Geoffroy Peeters, and Gael Richard. 2020. Audio-based auto-tagging with contextual tags for music. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 16–20.

[20] Bayan Issa, Muhammed Basheer Jasser, Hui Na Chua, and Muzaffar Hamzah. 2023. A comparative study on embedding models for keyword extraction using KeyBERT method. In *2023 IEEE 13th International Conference on System Engineering and Technology (ICSET).* IEEE, 40–45.

[21] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-Scale Similarity Search with GPUs. *IEEE Transactions on Big Data* 7, 3 (2019), 535–547.

[22] Yehuda Koren and Robert Bell. 2015. Advances in Collaborative Filtering. *Recommender Systems Handbook* (2015), 77–118.

[23] Sudarshan Lamkhede and Sudeep Das. 2019. Challenges in search on streaming services: netflix case study. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 1371–1374.

[24] Audrey Laplante. 2008. Everyday life music information-seeking behaviour of young adults: an exploratory study. (2008).

[25] Sanghoon Lee, Mohamed Masoud, Janani Balaji, Saeid Belkasim, Rajshekhar Sunderraman, and Seung-Jin Moon. 2017. A survey of tag-based information retrieval. *International Journal of Multimedia Information Retrieval* 6 (2017), 99–113.

[26] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[27] Ang Li, Jennifer Thom, Praveen Chandar, Christine Hosey, Brian St Thomas, and Jean Garcia-Gathright. 2019. Search mindsets: Understanding focused and non-focused information seeking in music search. In *The World Wide Web Conference.* 2971–2977.

[28] Jianguo Li, Yong Tang, and Jiemin Chen. 2017. Leveraging tagging and rating for recommendation: RMF meets weighted diffusion on tripartite graphs. *Physica A: Statistical Mechanics and its Applications* 483 (2017), 398–411.

[29] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive conversational agents in the post-chatgpt world. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval.* 3452–3455.

[30] Kelong Mao, Xi Xiao, Jieming Zhu, Biao Lu, Ruiming Tang, and Xiuqiang He. 2020. Item tagging for information retrieval: A tripartite graph neural network based approach. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 2327–2336.

[31] Gabriel Meseguer-Brocal, Dorian Desblancs, and Romain Hennequin. 2024. An Experimental Comparison of Multi-View Self-Supervised Methods for Music Tagging. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 1141–1145.

[32] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. 2023. A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435* (2023).

[33] Gustavo Penha, Enrico Palumbo, Maryam Aziz, Alice Wang, and Hugues Bouchard. 2023. Improving content retrievability in search with controllable query generation. In *Proceedings of the ACM Web Conference 2023.* 3182–3192.

[34] Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066* (2019).

[35] Bruno Sguerra, Marion Baranes, Romain Hennequin, and Manuel Moussallam. 2022. Navigational, informational or punk-rock? an exploration of search intent in the musical domain. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization.* 202–211.

[36] Ning Su, Jiyin He, Yiqun Liu, Min Zhang, and Shaoping Ma. 2018. User intent, behaviour, and perceived satisfaction in product search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.* 547–555.

[37] Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530* (2024).

[38] Xinyang Yi, Ji Yang, Lichan Hong, Derek Zhiyuan Cheng, Lukasz Heldt, Aditee Kumthekar, Zhe Zhao, Li Wei, and Ed Chi. 2019. Sampling-bias-corrected neural modeling for large corpus item recommendations. In *Proceedings of the 13th ACM conference on recommender systems.* 269–277.

[39] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).

[40] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, et al. 2023. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. *arXiv preprint arXiv:2302.09419* (2023).