

MuseChat: A Conversational Music Recommendation System for Videos

Supplementary Material

7. Summary of Supplementary Materials:

In this supplementary materials, we provide:

1. Details of Our Conversational Music Recommendation dataset, including data distribution of music videos, prompt template and examples of simulated conversations, see Section 8.
2. Details of the MVP model structure used for dataset construction (generate candidate music), see Section 9.
3. t-SNE visualization of music embeddings (from the recommendation module) associated with tags, see Section 10.
4. Additional Ablation Studies of the recommendation module, see Section 11.
5. Human Evaluation Details of the reasoning module, see Section 12.
6. Qualitative Results that show the demo of our MuseChat, see Section 13.

8. Dataset Details

In this section, we give details about our conversational music recommendation dataset including: (1) the distribution of music videos, and (2) the prompt template along with the examples of simulated conversations.

8.1. Data Distribution

To visualize the distribution of the music videos in the dataset, we plot top 20 music tags appeared in our dataset from MSD [2] dataset and MTT [19] dataset respectively. The distributions are shown in Figure 6.

To show the diversity of the applied tagging systems, we also study their correlation among top music tags, as shown in Figure 7. From this figure, we observe that similar tags have higher correlation coefficient. For example, Hip-Hop music is highly correlated with electronic music, but is less correlated with folk music. And the widely spread ranges of the coefficients show that our tagging systems are capable of capturing a diverse range of music elements for different music tracks.

8.2. Prompt Template and Simulated Conversations

Figure 8 illustrates the prompt template sent to GPT-3.5 for simulating conversations between a user and a recommendation system. Initially, we establish constrained rules to guide the generation of GPT-3.5. We then supply titles and top 5 music tags from each of two referenced datasets: MTT [19] and MSD [2]. These tags apply to both original music and candidate music examples. If metadata like

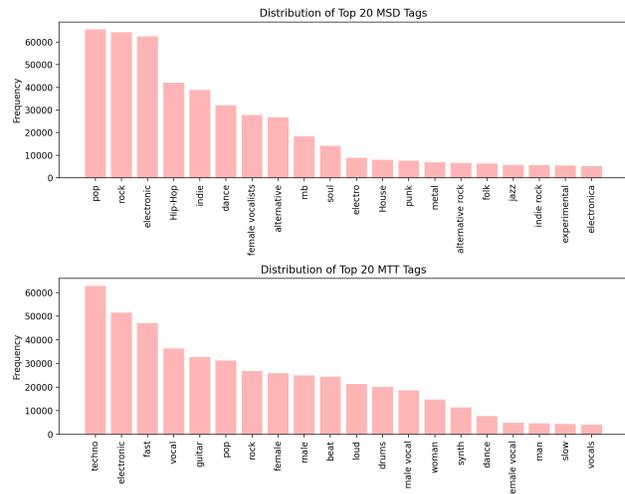


Figure 6. The distributions of the top 20 tags from the MSD [2] and MTT [19] systems used in our dataset.

official track names, album names, or artist names are available, they are also included in the prompt. Finally, we provide human-written conversation templates featuring original and candidate music examples. During generation, we input different pairs of original and candidate music, guiding GPT-3.5 to create new conversations based on the provided human-written examples. Figure 9 and Figure 10 show 10 examples of simulated conversations based on the prompt template shown in Figure 8.

9. MVP structure

The Music Video Pretrained (MVP) model is structured as a two-tower system, sharing a similar architecture with the models described in [23, 42]. This model was trained on a dataset comprising 3 million pairs of short internal music videos in our proprietary dataset. These high-quality music video pairs, created by professionals or popular influencers, cover various topics such as dance, food/cooking, travel, vlogs, and sports. The MVP model extracts a 10-second segment from the input video and captures 5 frames per second from this segment. Similarly, the MVP extracts a 10-second segment from the input music and samples the audio signal at a rate of 12,000 Hz. For video processing, the model employs the CLIP Image encoder [41]. In terms of audio processing, the MVP utilizes an Audio Spectrogram Transformer (AST) [17] to extract music embeddings. Both video and music embeddings are then projected into a shared 256-dimensional latent space using respective train-

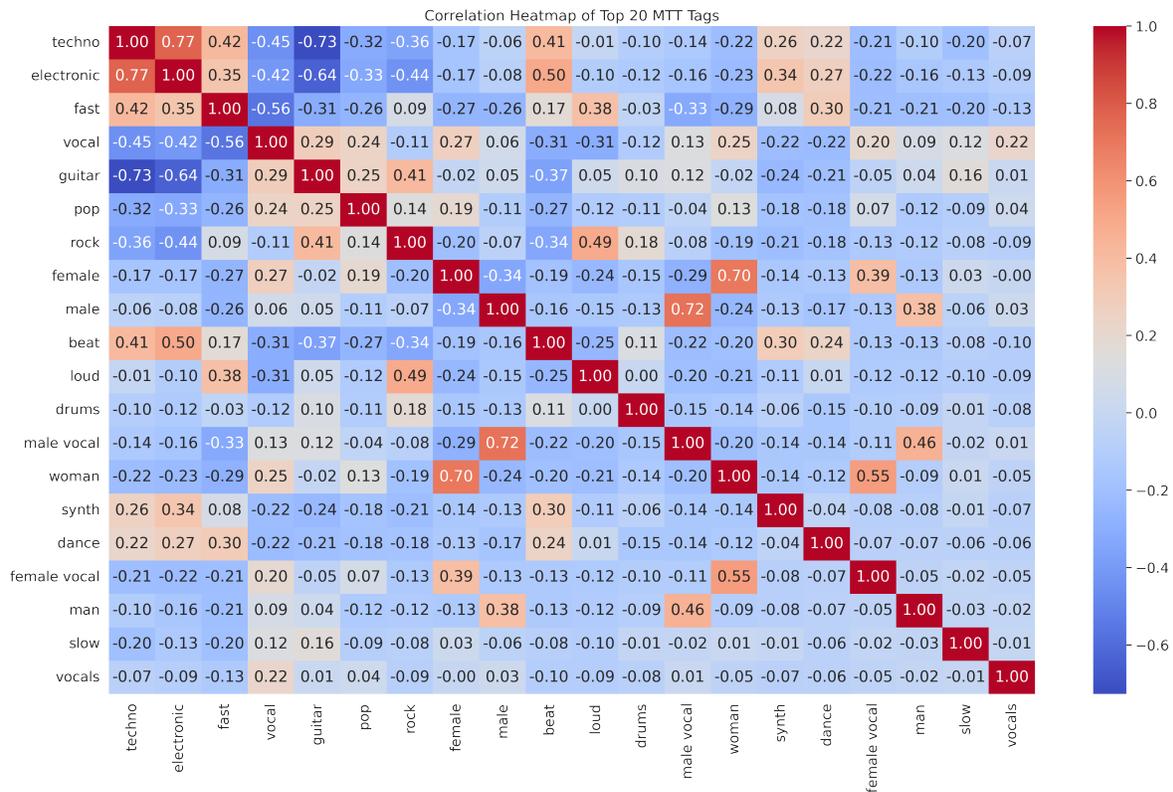
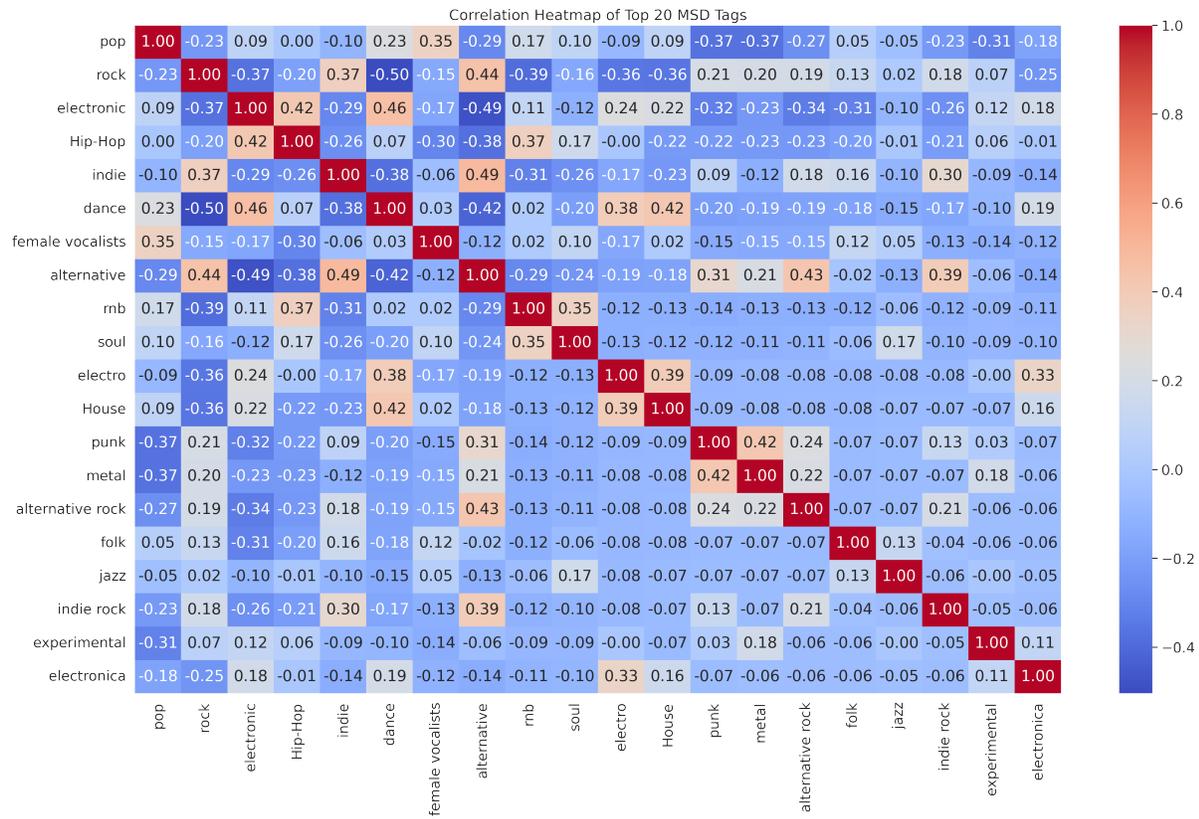


Figure 7. The correlation matrix of the top 20 tags from the MSD [2] and MTT [19] systems used in our dataset.

You are simulating a conversation between a user and a music recommendation system based on the original music and candidate music. You must follow the following rule:

- In the simulated conversation, user must not mention original music name or title.
- User can only mention metadata of the original music.
- Recommender must not use irrelevant descriptions in the recommendation.
- Recommender can use background knowledge to improve the quality of simulated conversation.
- Following is an example

Original music: Title: FREESTYLE "DOMINQUAIN" 2013. AVEC LA PHAZZ ET REENO. Track name: DOMINQUAIN. Album name: None. Artists: None. MSD tags: loud, fast, male, rock, male vocal. MTT tags: Hip-Hop, electronic, indie, dance, experimental.

Candidate music: Title: FIFA LB - Xequ-Mate (Feat: MagBoy Rsk) (VideoClip). Track name: None. Album name: None, Artists: None. MSD tags: Hip-Hop, electronic, chillout, ambient, electronica. MTT tags: techno, electronic, beat, synth, vocal.

User: Can you recommend a music for my dance video?

System: Based on your video, I recommend 'Xequ-Mate' by FIFA LB featuring MagBoy Rsk. It's a chillout and ambient track with electronic and techno elements. The song has a relaxing and atmospheric vibe, which could complement the visuals of your video. The vocals and synth in the track add a unique touch to the overall sound.

User: That sounds great, I like the electronic and chillout genre in your music recommendations. However, I would like to incorporate a bit more of a hip-hop style, especially with experimental and indie influences. I'm also interested in fast-paced beats and male vocals.

System: Allow me to recommend 'DOMINQUAIN' by La Phazz and Reeno. It's a freestyle track that combines elements of hip-hop, electronic, and indie genres. The song has experimental and dance influences, which could add a unique touch to your video. It was released in 2013, and the fast-paced beats and male vocals align with your preferences.

Given the following information, please simulate conversation.

Original music: Title:{title}. Track name:{track name}. Album name:{album name}. Artists:{artist}. MSD tags: {tags}. MTT tags: {tags}.

Candidate music: Title:{title}. Track name:{track name}. Album name:{album name}. Artists:{artist}. MSD tags: {tags}. MTT tags: {tags}.

Figure 8. Prompt Template given to GPT-3.5 for dialogue simulation

able linear layers. During training, the MVP model employs the Contrastive Multiview Coding Loss, aiming to increase the distance between mismatched video-music pairs (negative pairs) and decrease it for well-matched pairs (positive pairs). Upon completion of its training, the MVP model computes the similarity between video and music segments to select the most fitting music for a given video.

10. t-SNE visualizations

To further study the quality of music embeddings learned from MVT-Fusion module, we use t-SNE to project the original music features, extracted using the AST branch, from our conversational music recommendation dataset into a 2-dimensional space. We use the tag with highest proba-

bility from method [38] as the tag of the music track. For each music track in the test set, we extract the corresponding music feature using AST. As shown in the plots of both MSD and MTT tags in Figure 11, the music embedding is reasonably separated according to the tags.

11. More Ablation Studies

We include additional ablation studies to manifest the effectiveness of the design choices in our proposed MVT-Fusion module.

11.1. First Turn Retrieval Results

In the first dialogue turn, we assume that user uploads video only for music recommendation. To demonstrate the ca-

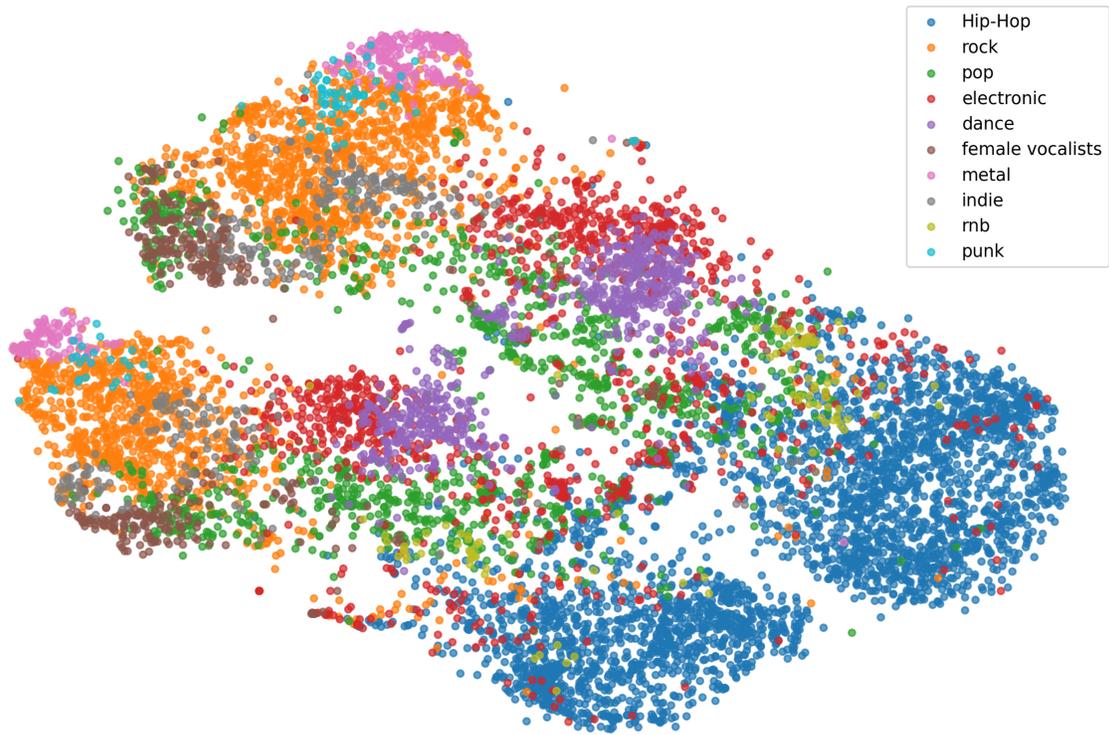
	<p>Original music: Title: New TV Wonders - Shine. Track name: Shine. Album name: None. Artists: New TV Wonders. MSD tags: rock, indie, alternative, indie rock, pop. MTT tags: rock, guitar, pop, male, male vocal.</p> <p>Candidate music: Title: Dengaz - Obrigado (feat. Dino) (Directed By DNG). Track name: Obrigado. Album name: AHYA. Artists: Dengaz. MSD tags: Hip-Hop, rock, pop, electronic, soul. MTT tags: techno, electronic, pop, vocal, fast.</p> <p>User: Can you recommend a music for my video?</p> <p>(1) System: Based on your video, I recommend 'Obrigado' by Dengaz. It's a track from the album 'AHYA' and features Dino. This song combines elements of hip-hop, rock, pop, and electronic music with soulful vocals. It has a techno and fast-paced sound too.</p> <p>User: That sounds great, but I also want to incorporate some rock and indie elements into my video. Do you have any recommendations that align with those genres?</p> <p>System: Certainly! How about 'Shine' by New TV Wonders? It's a rock and indie track from their album 'The Ostrich Maneuver'. The song has alternative and pop influences as well. It was released in spring 2012 and directed by Hans Frederik Munch.</p>
	<p>Original music: Title: Caroline & Damon "i knew you were trouble...". Track name: I Knew You Were Trouble. Album name: None. Artists: Madiilyn Bailey. MSD tags: female vocalists, pop, folk, rock, indie. MTT tags: female, woman, vocal, guitar, female vocal.</p> <p>Candidate music: Title: Crown The Empire / Voices (Official Music Video). Track name: None. Album name: None. Artists: None. MSD tags: metal, electronic, rock, experimental, alternative. MTT tags: loud, rock, fast, techno, electronic.</p> <p>User: Can you recommend a music for my video?</p> <p>(2) System: Based on your video, I recommend 'Voices' by Crown The Empire. It's a loud and fast-paced track with a mix of electronic and rock elements. The song has an experimental and alternative vibe, which can add intensity and energy to your video.</p> <p>User: That sounds great! I'm also looking for a song with a female vocal and a pop/rock sound. Do you have any recommendations?</p> <p>System: Certainly! I suggest 'I Knew You Were Trouble' by Madiilyn Bailey. It's a cover of the original song by Taylor Swift and has a pop/rock sound with a female vocal. This track can add a catchy and energetic element to your video while maintaining a similar vibe as the original song you mentioned.</p>
	<p>Original music: Title: Voodoo Terror Tribe "Wake of The White Devil" (Official Music Video) - Track name: Wake of the White Devil (Radio Edit). Album name: Symptoms Of Sin. Artists: VTT. MSD tags: metal, heavy metal, rock, experimental, punk. MTT tags: rock, loud, fast, metal, techno.</p> <p>Candidate music: Title: Leading Light - Culture Under Compulsion (OFFICIAL VIDEO) - Track name: None. Album name: None. Artists: None. MSD tags: metal, rock, electronic, alternative, indie. MTT tags: techno, loud, fast, rock, electronic.</p> <p>User: Hi, can you recommend a music for my video?</p> <p>(3) System: Of course! Based on your needs, I'd recommend 'Leading Light - Culture Under Compulsion'. This track offers a mix of metal, rock, electronic, alternative, and indie genres. The music is fast-paced and loud, which can add intensity to your video.</p> <p>User: Sounds good, but I'm looking for something with a heavier and more experimental sound.</p> <p>System: In that case, I would recommend 'Wake of the White Devil (Radio Edit)' by Voodoo Terror Tribe. This track falls under the metal genre and has heavy metal, rock, and experimental influences. The music has a punk vibe, which adds an edgy and intense feel to the sound.</p>
	<p>Original music: Title: "Rouv lo zyé" Dassby extrait "Ekri pou egzisté" - Track name: None. Album name: None. Artists: None. MSD tags: Hip-Hop, rock, soul, rnb, electronic. MTT tags: drums, beat, male, vocal, male vocal.</p> <p>Candidate music: Title: Encore Loin - Louis XIX (Dinozbar) - Track name: None. Album name: None. Artists: None. MSD tags: Hip-Hop, electronic, dance, soul, rnb. MTT tags: techno, male, male vocal, electronic, man.</p> <p>User: I'm looking for a music recommendation for my video.</p> <p>(4) System: Based on your video, I recommend 'Encore Loin' by Louis XIX. It's a dance track with electronic and soul influences, categorized as hip-hop. The song features male vocals and has a techno sound. The music video for 'Encore Loin' is available on their YouTube channel.</p> <p>User: That sounds great, but I would like to incorporate elements of rock and rnb into the music as well. I'm also interested in electronic beats and drums.</p> <p>System: Allow me to recommend 'Rouv lo zyé' by Dassby. It's a hip-hop track with rock and soul influences. The song also incorporates electronic elements and features drums. It was included in Dassby's 'Ekri pou egzisté' mixtape. The music video for 'Rouv lo zyé' is available on the Internet.</p>
	<p>Original music: Title: Young Dro - Oops My Bad(Official Video) - Track name: None. Album name: None. Artists: None. MSD tags: Hip-Hop, electronic, dance, electro, indie. MTT tags: loud, techno, fast, electronic, rock.</p> <p>Candidate music: Title: Face Da Beast "Tension" (OFFICIAL VIDEO) - Track name: None. Album name: None. Artists: None. MSD tags: Hip-Hop, rnb, soul, electronic, pop. MTT tags: techno, electronic, beat, synth, fast.</p> <p>User: Can you recommend a music for my video?</p> <p>(5) System: Based on your video, I recommend 'Tension' by Face Da Beast. It's a track with a fast-paced beat and electronic elements. The song falls under the genres of Hip-Hop, R&B, and Soul with electronic and pop influences. It was directed by Joshua Ramirez for RFE MEDIA/RAPFOREVER FILMS and offers quality HD visuals.</p> <p>User: That sounds great! I'm also interested in incorporating some elements of Young Dro's style. Can you recommend a track that combines Hip-Hop, electronic, and indie influences?</p> <p>System: Certainly! I suggest checking out 'Oops My Bad' by Young Dro. It's a track with a mix of Hip-Hop and electronic elements, and it falls under the genres of Hip-Hop, electronic, dance, electro, and indie. The song was released in 2012 and offers a unique blend of styles that could complement your video.</p>

Figure 9. Examples of simulated conversations

	<p>Original music: Title: Respek - Pitch Black Official Music Video - Track name: None. Album name: None. Artists: None. MSD tags: Hip-Hop, rock, female vocalists, pop, indie. MTT tags: techno, vocal, female, vocals, woman.</p> <p>Candidate music: Title: ZIM'K x OUSSOS - MAFIA (CLIP) - Track name: None. Album name: None. Artists: None. MSD tags: Hip-Hop, electronic, dance, pop, rock. MTT tags: techno, fast, electronic, loud, male.</p> <p>User: Can you recommend a music for my video?</p> <p>(6) System: Based on your video, I recommend 'MAFIA' by ZIM'K x OUSSOS. It's an energetic track with elements of hip-hop, electronic, dance, and pop. The song has a fast tempo and loud electronic beats which can add excitement to your video.</p> <p>User: That sounds great! I like the electronic and dance elements in your recommendation. However, I would like to incorporate some female vocals and indie influences into the music as well.</p> <p>System: Allow me to recommend 'Pitch Black' by Respek. It's a hip-hop track with female vocals and indie influences. The song has a mix of rock and pop elements as well. It might be a fitting choice for your video.</p>
	<p>Original music: Title: In Her Eyes - Leroy Jönsson (Original) - Track name: In Her Eyes. Album name: You. Artists: Leroy Jönsson. MSD tags: rock, indie, alternative, acoustic, folk. MTT tags: guitar, male vocal, vocal, male, pop.</p> <p>Candidate music: Title: Into The Past by Nero: Mitchell Jackson Choreography - Track name: Into The Past. Album name: Into The Past. Artists: Nero. MSD tags: electronic, electro, electronica, dance, experimental. MTT tags: loud, techno, rock, fast, electronic.</p> <p>User: Can you recommend a music for my video?</p> <p>(7) System: Based on your video, I recommend 'Into The Past' by Nero. It's an electronic track from their album 'Into The Past'. The song has a loud and techno sound, which can add an energetic and experimental vibe to your video. It's also fast-paced, which can complement the visuals well.</p> <p>User: That sounds great! I also want to incorporate elements of rock and indie music, as well as acoustic and folk influences. Can you recommend something that combines these genres?</p> <p>System: Certainly! How about 'In Her Eyes' by Leroy Jönsson? It's a rock and indie track with acoustic and folk elements. The song features Leroy Jönsson's vocals and guitar, creating a melodic and heartfelt sound. It can provide a nice contrast to the electronic and experimental vibe of 'Into The Past'.</p>
	<p>Original music: Title: This is Me - Demi Lovato (Tradução) - Track name: This Is Me. Album name: None. Artists: Lauren Tenney, Colton Rudloff. MSD tags: pop, rock, female vocalists, alternative, indie. MTT tags: female, woman, pop, vocal, rock.</p> <p>Candidate music: Title: Dryante feat. Raku - Hollow (Godsmack full cover) - Track name: None. Album name: None. Artists: None. MSD tags: rock, folk, female vocalists, indie, acoustic. MTT tags: guitar, vocal, pop, female, singing.</p> <p>User: I'm looking for a music recommendation for my video.</p> <p>(8) System: Based on your video, I recommend 'Hollow' by Dryante feat. Raku. It's a cover of the song originally performed by Godsmack. This track falls under the rock genre, with folk and acoustic influences. It features female vocals and has a mellow and introspective sound. Although there isn't much description available, based on the tags and genre, it might be a good fit for your video.</p> <p>User: That sounds like a great recommendation. I like the rock and folk elements in your recommendation. However, I would like to incorporate more of Demi Lovato's style, especially her pop and alternative influences. I'm also interested in female vocals.</p> <p>System: Allow me to recommend 'This Is Me' by Lauren Tenney and Colton Rudloff. It's a track featuring Joe Jonas and is a cover of Demi Lovato's song. This track falls under pop, rock, and alternative genres. It includes female vocals and has a catchy and energetic sound that aligns with Demi Lovato's style.</p>
	<p>Original music: Title: Полина Гагарина - Кольбельная (HD) - Track name: Кольбельная. Album name: Попроси у облаков (2018 Version). Artists: Полина Гагарина. MSD tags: female vocalists, pop, soul, rnb, electronic. MTT tags: female, woman, female vocal, vocal, female voice.</p> <p>Candidate music: Title: Aneurisma - "Despierta" (Official Music Video) - Track name: None. Album name: None. Artists: None. MSD tags: metal, rock, electronic, experimental, alternative. MTT tags: loud, rock, techno, fast, electronic.</p> <p>User: Can you recommend a music for my video?</p> <p>(9) System: Based on your video, I recommend 'Despierta' by Aneurisma. It's a loud and fast-paced track that falls under the genres of metal, rock, and electronic. This song has experimental and alternative influences which can add an edgy and unique element to your video.</p> <p>User: That sounds great, but I would like to incorporate a bit of Полина Гагарина's distinctive style, especially her soul and R&B components. I'm also interested in female vocals and electronic elements.</p> <p>System: Allow me to recommend 'Кольбельная' by Полина Гагарина. It's a track from her album 'Попроси у облаков (2018 Version)'. This song combines elements of pop, soul, and R&B with electronic beats. The powerful female vocals and electronic sound could be a fitting choice for your video.</p>
	<p>Original music: Title: Black Swan - It's Just a Ride - Bill Hicks - Track name: It's Just a Ride. Album name: Revelations: Live in London. Artists: Bill Hicks. MSD tags: indie, rock, alternative, electronic, indie rock. MTT tags: rock, loud, guitar, fast, drums.</p> <p>Candidate music: Title: Quebre o Silêncio - Célebre Análise [Official Lyric Video] - Track name: None. Album name: None. Artists: None. MSD tags: electronic, metal, experimental, rock, ambient. MTT tags: techno, electronic, weird, synth, ambient.</p> <p>User: Can you recommend a music for my video?</p> <p>(10) System: Based on your video, I recommend 'Quebre o Silêncio' by Célebre Análise. It's an electronic and experimental track with metal and rock influences. The song has a techno and ambient sound, which aligns well with the alternative and electronic vibe you're looking for.</p> <p>User: That sounds interesting, but I would like to incorporate a bit more of Bill Hicks' style, especially his indie rock and alternative components. I'm also interested in guitar and fast-paced drums.</p> <p>System: Allow me to recommend 'It's Just a Ride' by Bill Hicks. It's a track from his album 'Revelations: Live in London'. It combines elements of indie rock, alternative, and electronic music with a focus on loud guitars and fast-paced drums that could be a fitting choice for your video.</p>

Figure 10. Examples of simulated conversations. Continued from Figure 9

t-SNE of music features - MSD tags



t-SNE of music features - MTT tags

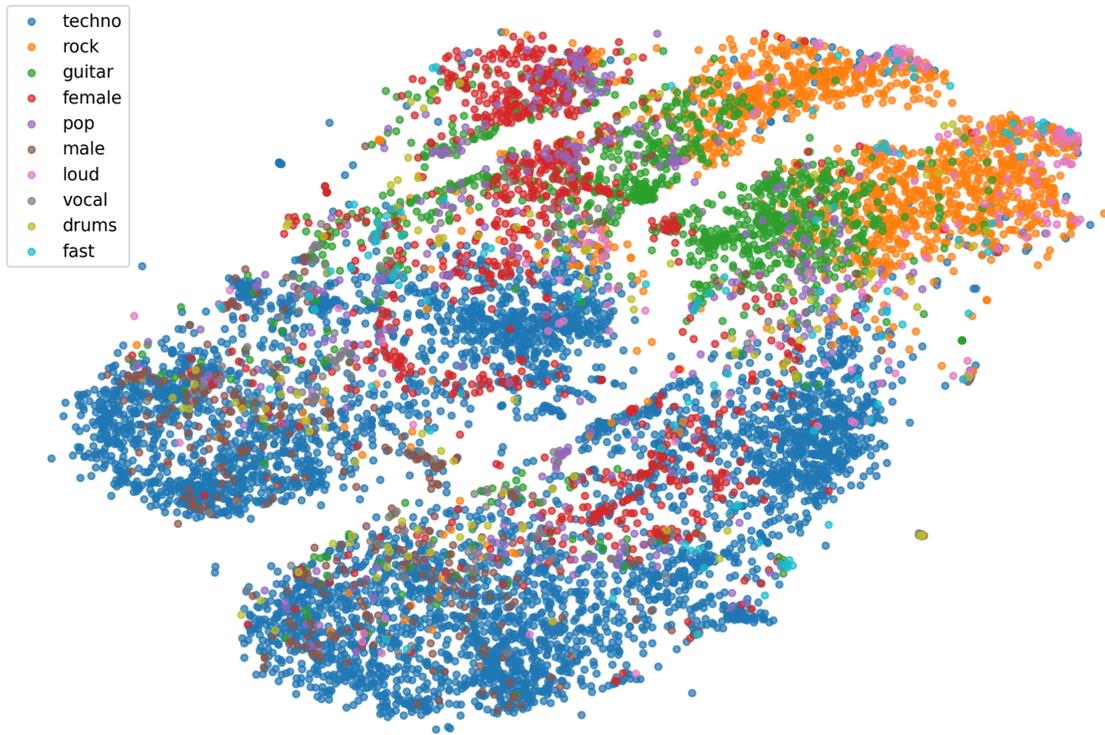


Figure 11. t-SNE visualizations of music features in MSD [2] and MTT [19] tags.

pability of our proposed MVT-Fusion module in retrieving the music initially without user prompt as inputs, we compare different fusion baseline methods against our method. The results are shown in Table 5. While all of these models are trained on three modalities – video, candidate music and text, they achieve similar good results in music retrieval solely based on video as input. It is note-worthy that such tri-modalities models (MR 5) even outperform the MVP model (MR 7) with only video and music (Table 1 in main text).

Module	MR ↓	R@1 ↑	R@5 ↑	R@10 ↑
Sum-Fusion	5	21.70	49.49	63.27
Self-Attn Fusion	5	21.44	50.30	64.09
Cross-Attn Fusion	5	20.74	48.83	63.10
MuseChat (ours)	5	20.74	48.83	63.10
Chance	250	0.20	1.00	2.00

Table 5. First Turn Retrieval Performance using different fusion methods.

11.2. Impact of Input Modality at the Test Stage

We also examine how each modality contributes to the performance of MuseChat recommendation module. At the test time, we test two variants: (1) Use video as the only input, (2) Use candidate music along with user prompt text as inputs without addition of the video feature. As shown in Table 6, the model without the video feature perform worst, suggesting visual information as the most significant role to the retrieval performance. Moreover, when video information is supplemented with candidate music along with user prompt, the retrieval performance gets boosted further, matching both the context of the video and the user’s preference.

Model	MR ↓	R@1 ↑	R@5 ↑	R@10 ↑
MuseChat (music+text)	22	7.37	23.16	34.04
MuseChat (video)	5	20.74	48.83	63.10
MuseChat (ours)	2	32.79	63.92	76.53
Chance	250	0.20	1.00	2.00

Table 6. Impact of modality inputs on MuseChat at the test stage.

11.3. Modality Fusion Order Ablation

In our proposed MVT-Fusion module, candidate music and user prompt text are fused before combined with the video feature. To demonstrate the efficacy of this strategy in terms of the fusion order, we explore two alternative fusion orders: (1) Integrating candidate music with video with cross-modal attention followed by addition to text, (2) Merging

text with video with cross-modal attention followed by addition to candidate music. Both alternatives keep the same internal architecture as the MVT-Fusion module, with the order of fusion being the only difference. As shown in Table 7, MuseChat’s module outperforms the other two methods on both MR and recall metrics. We believe that the success of our strategy is attributed to the effective separation of visual information as an independent branch, reflecting its importance of contribution in the retrieval.

Model	MR ↓	R@1 ↑	R@5 ↑	R@10 ↑	SR ↑
Music-Video Fusion	3	27.97	59.14	72.55	30.21
Text-Video Fusion	4	25.96	57.16	71.05	26.05
MuseChat (ours)	2	32.79	63.92	76.53	40.49
Chance	250	0.20	1.00	2.00	0.40

Table 7. Ablation studies on fusion strategies in terms of fusion order

12. Human Evaluation Details of the Reasoning Module

We use 5-point MOS (Mean-Opinion-Score) scale to measure correctness, musicality and clarity of the reasoning results. We include the following questions:

- Correctness: How accurately does the system identify the music information?
 - 1: Provides incorrect or irrelevant information about the music.
 - 2: Identifies basic information but with significant inaccuracies.
 - 3: Generally correct in identifying basic information, though some inaccuracies are present.
 - 4: Accurately identifies most music information with minor errors.
 - 5: Accurately and consistently identifies all relevant music information.
- Musicality: How well does the system describe the music’s characteristics in terms of depth and insight?
 - 1: Incorrect explanation of the music’s characteristics.
 - 2: Basic explanation, lacking depth.
 - 3: Adequate, covering essential characteristics with some depth.
 - 4: Comprehensive explanation with considerable depth.
 - 5: Exceptionally insightful and detailed explanation, capturing the essence of the music’s characteristics.
- Clarity: How well does the system convey its reasoning in terms of clarity, coherence, and completeness?
 - 1: Outputs are generally incomplete or in the wrong format.
 - 2: Outputs are mostly incomplete with some correct format elements.

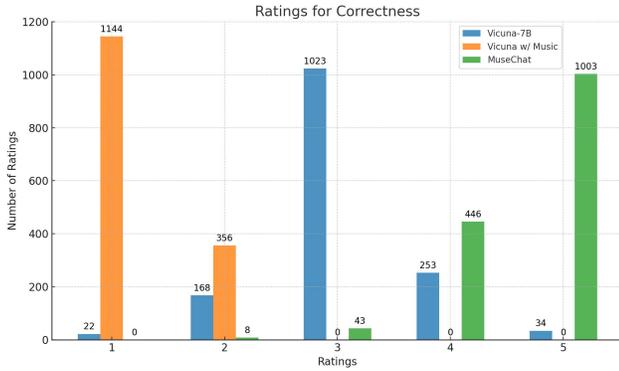


Figure 12. Human Evaluation Results for Correctness Aspect of Reasoning Module in MuseChat.

- 3: Outputs convey main information but are not fully complete.
- 4: Outputs are mostly complete with minor errors or format issues.
- 5: Outputs are consistently complete, clear, and correctly formatted.

Participants use the information from the webpage of corresponding YouTube video as the ground truth for correctness evaluation, listen to the retrieved music and compare with the reasoning text response to evaluate musicality, and assess the clarity of the response text as a whole. To ensure robustness in our evaluation, we select 30 participants with different levels of music backgrounds from music lovers to music performers with years of experience. In the survey, participants are presented with a multiple-choice question to assess their musical skill level. They are asked to select from one of three options: (A) Music lover and show familiarity with the task; (B) Music practitioner with basic knowledge of music along with practical experience; and (C) Music expert who is experienced in at least one type of musical performance. The reported distribution of skill level among the participants is: (A) 12, (B) 11 and (C) 7. Such diversity of skill levels among the participants enhances the overall robustness of the final evaluation. To reduce the variance of the survey thus ensuring the quality, each participant is asked to review 50 reasoning text responses from each model: Vicuna-7B, Vicuna w/ Music, and MuseChat, focusing on assessing model performance in terms of correctness, musicality, and clarity. We show the distribution of ratings respectively in Figures 12 to 14.

13. Qualitative Results

We provide additional qualitative results in Figures 15 to 17. We also include 4 video demos to show the capability of MuseChat in music recommendation under the dialogue setting. Please see the attached files in the supplementary bundle.

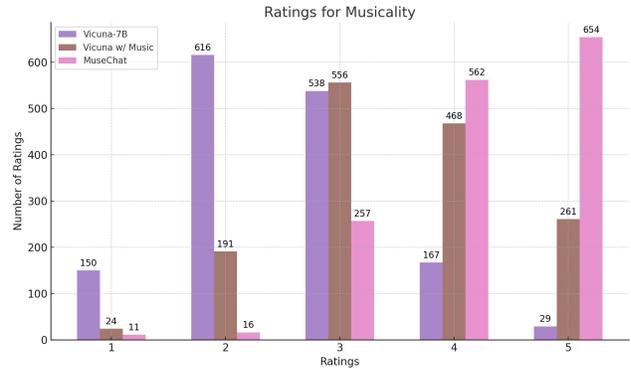


Figure 13. Human Evaluation Results for Musicality Aspect of Reasoning Module in MuseChat.

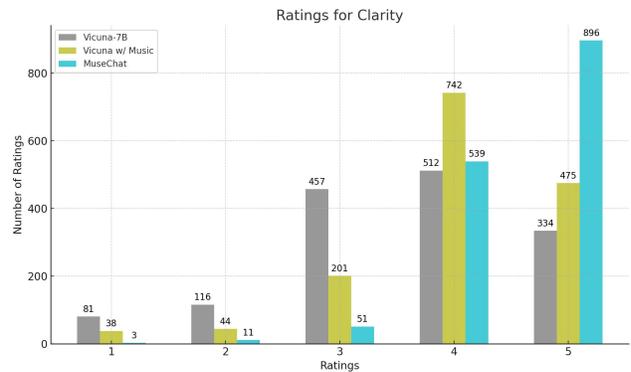


Figure 14. Human Evaluation Results for Clarity Aspect of Reasoning Module in MuseChat.

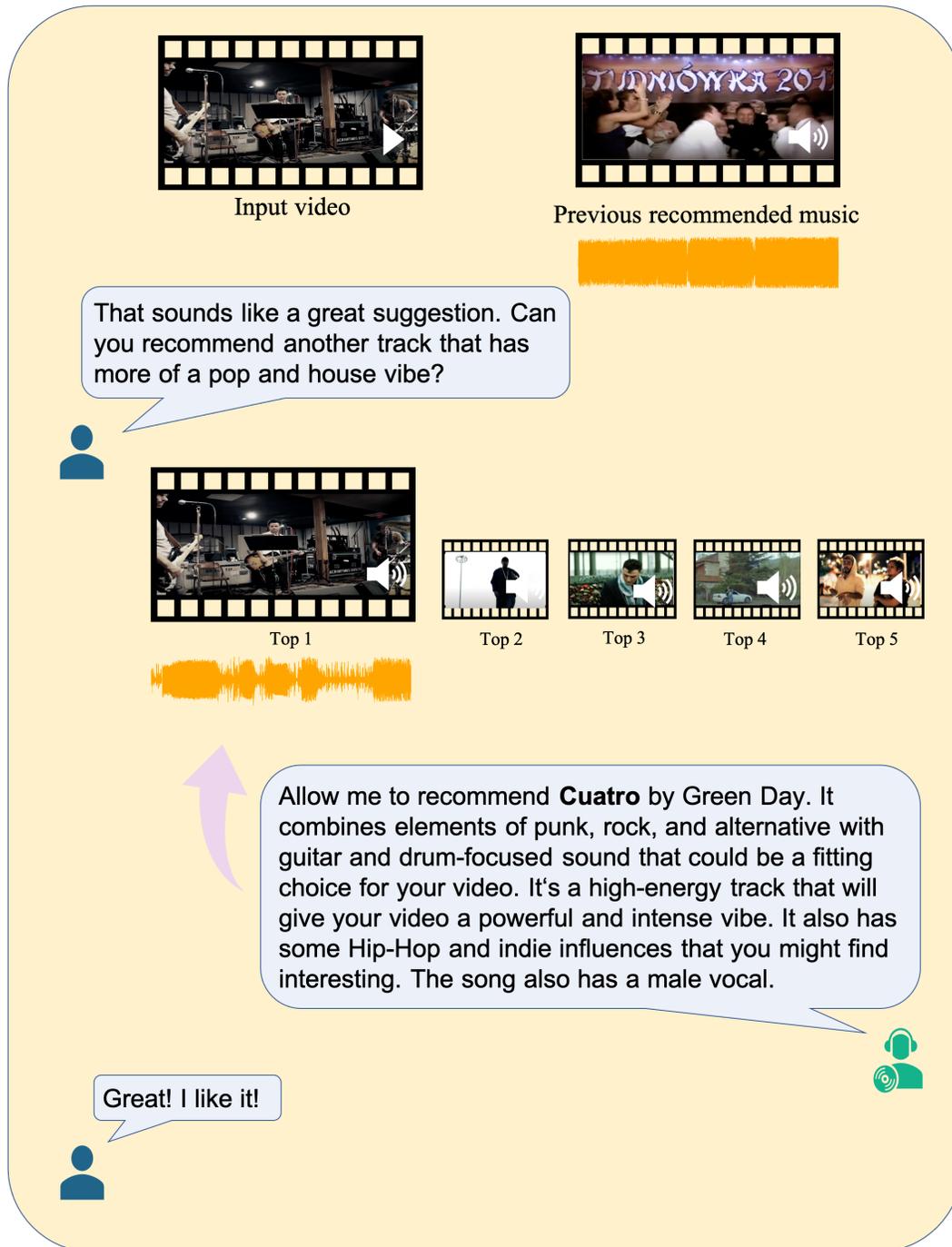


Figure 15. Example showing MuseChat’s capability in improving recommendation results by learning user’s preferences, contextual music and video content.

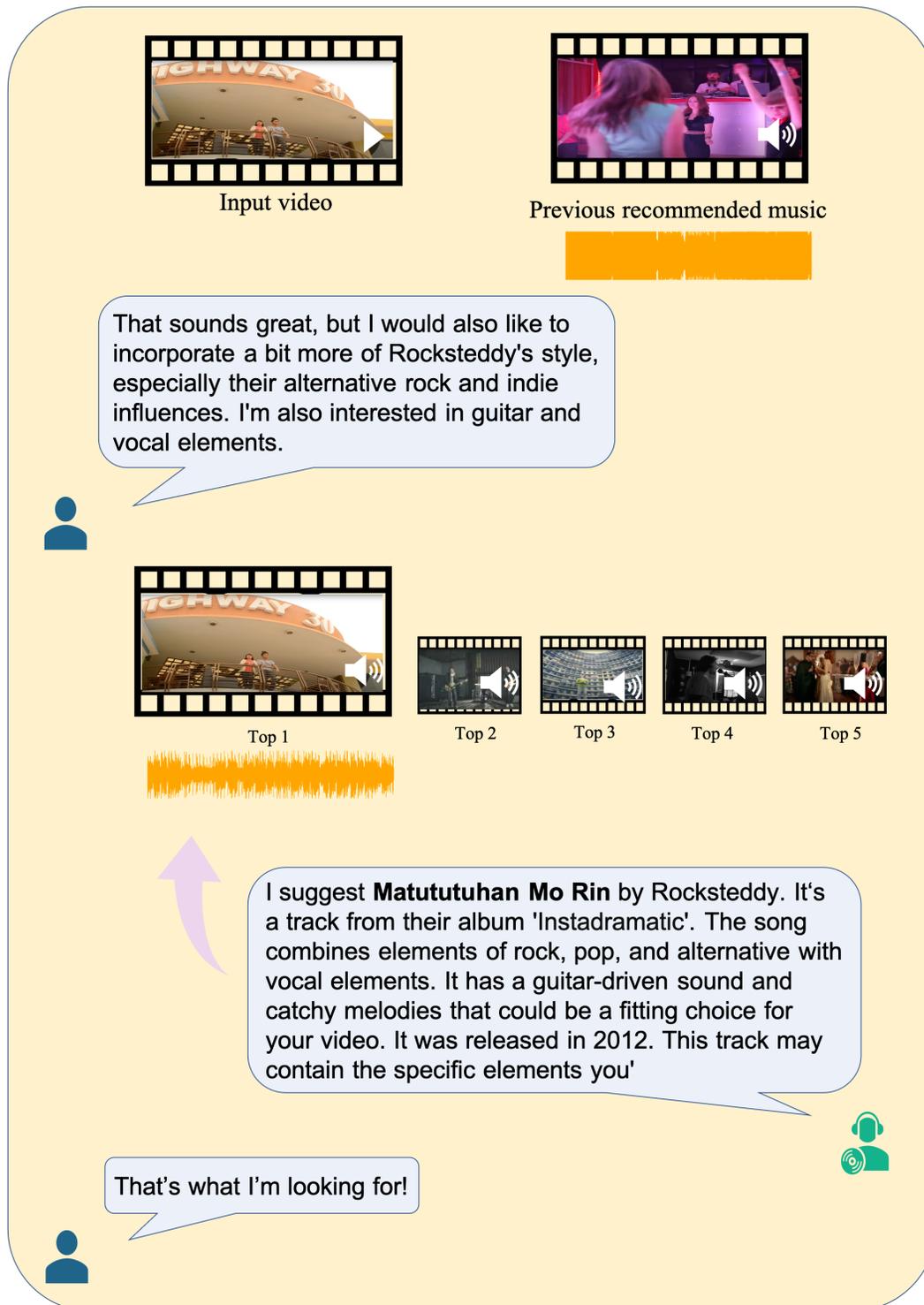


Figure 16. Example showing MuseChat's capability in improving recommendation results by learning user's preferences, contextual music and video content.

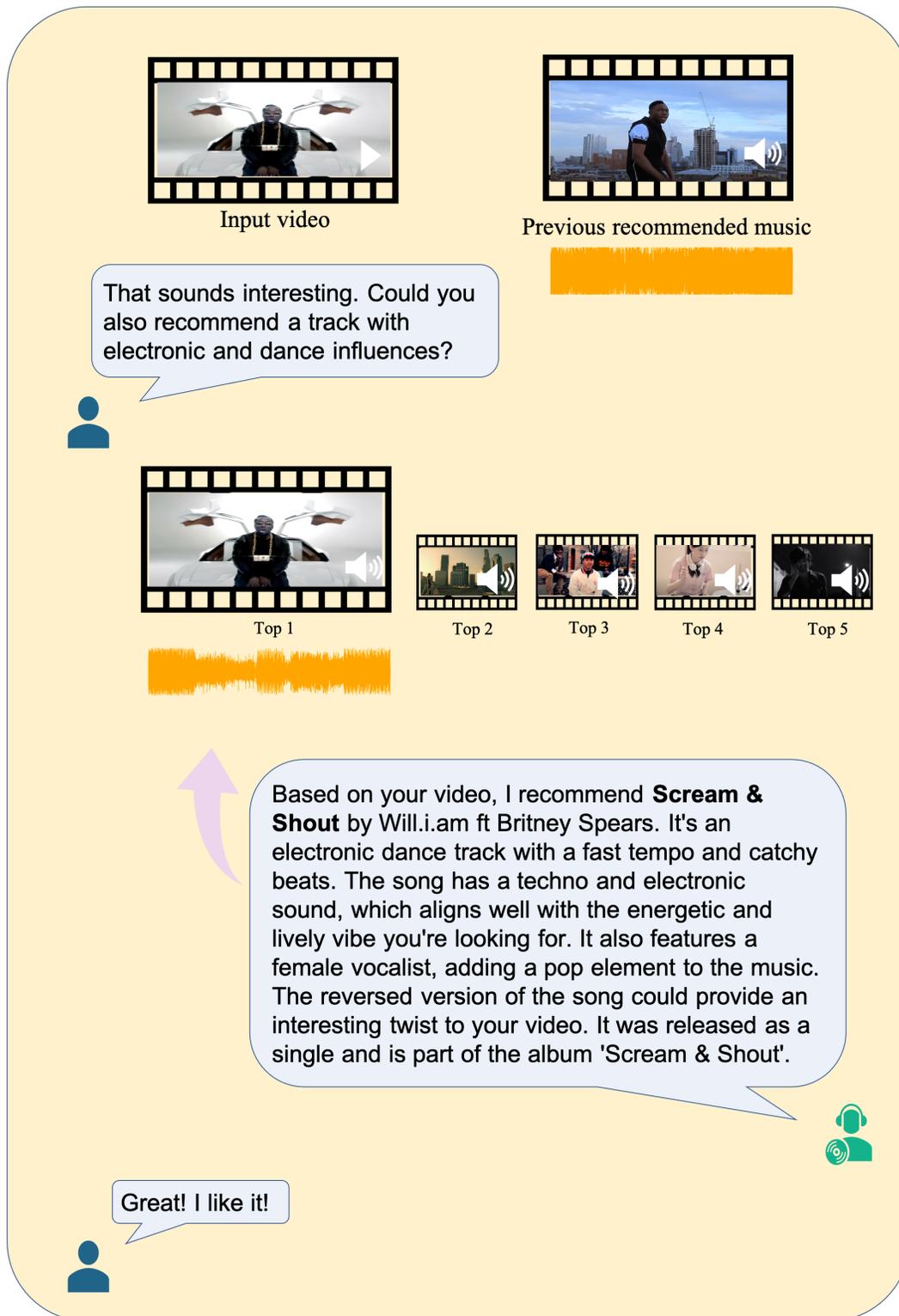


Figure 17. Example showing MuseChat’s capability in improving recommendation results by learning user’s preferences, contextual music and video content.