

Language-based representations for music recommendation

Gustave Cortal

Deezer research proposal

Context and objective

- ▶ The research proposal aims to develop language-based representations for musical items and users

Context and objective

- ▶ The research proposal aims to develop language-based representations for musical items and users
- ▶ Large Language Models (LLMs) can be used to generate textual descriptions of items and user profiles

Context and objective

- ▶ The research proposal aims to develop language-based representations for musical items and users
- ▶ Large Language Models (LLMs) can be used to generate textual descriptions of items and user profiles
- ▶ This enable explainable and flexible music recommendation systems

Context and objective

- ▶ The research proposal aims to develop language-based representations for musical items and users
- ▶ Large Language Models (LLMs) can be used to generate textual descriptions of items and user profiles
- ▶ This enable explainable and flexible music recommendation systems
- ▶ The research strategy is articulated around three interconnected axes: user modeling, item modeling, and music recommendation

User modeling

User modeling

How can we leverage LLMs to obtain scrutable natural language profiles?

User modeling

How can we leverage LLMs to obtain scrutable natural language profiles?

- ▶ Preference modeling has relied on interaction data to produce embeddings

User modeling

How can we leverage LLMs to obtain scrutable natural language profiles?

- ▶ Preference modeling has relied on interaction data to produce embeddings
- ▶ Vector-based methods lack interpretability and raise scrutability issues

User modeling

How can we leverage LLMs to obtain scrutable natural language profiles?

- ▶ Preference modeling has relied on interaction data to produce embeddings
- ▶ Vector-based methods lack interpretability and raise scrutability issues
- ▶ Scientific literature establishes a strong link between musical preferences and psychological traits

User modeling

How can we leverage LLMs to obtain scrutable natural language profiles?

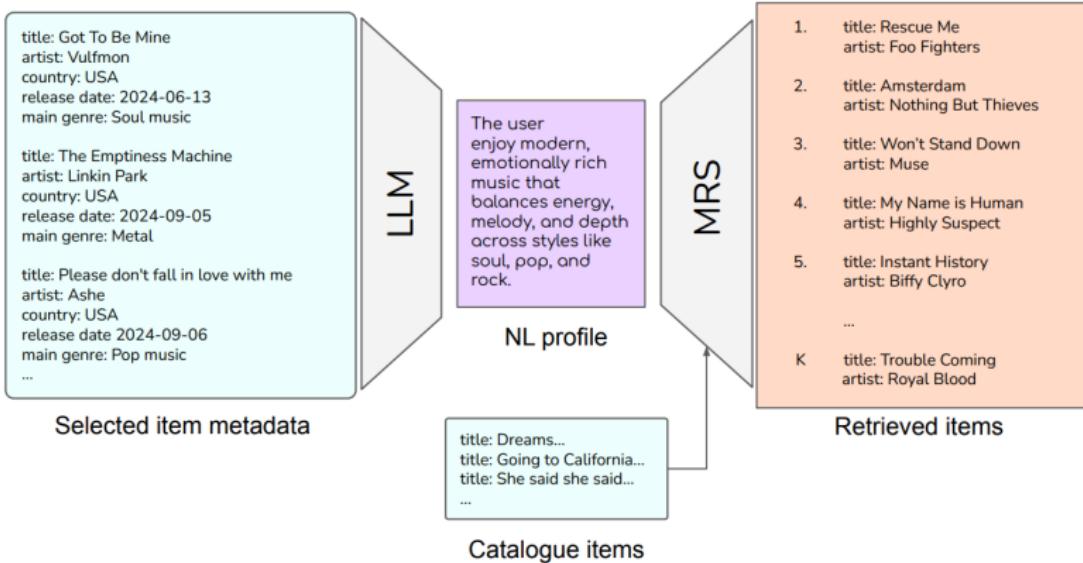
- ▶ Preference modeling has relied on interaction data to produce embeddings
- ▶ Vector-based methods lack interpretability and raise scrutability issues
- ▶ Scientific literature establishes a strong link between musical preferences and psychological traits
- ▶ We could use LLMs to infer these states and preferences, transforming them into natural language profiles

User modeling

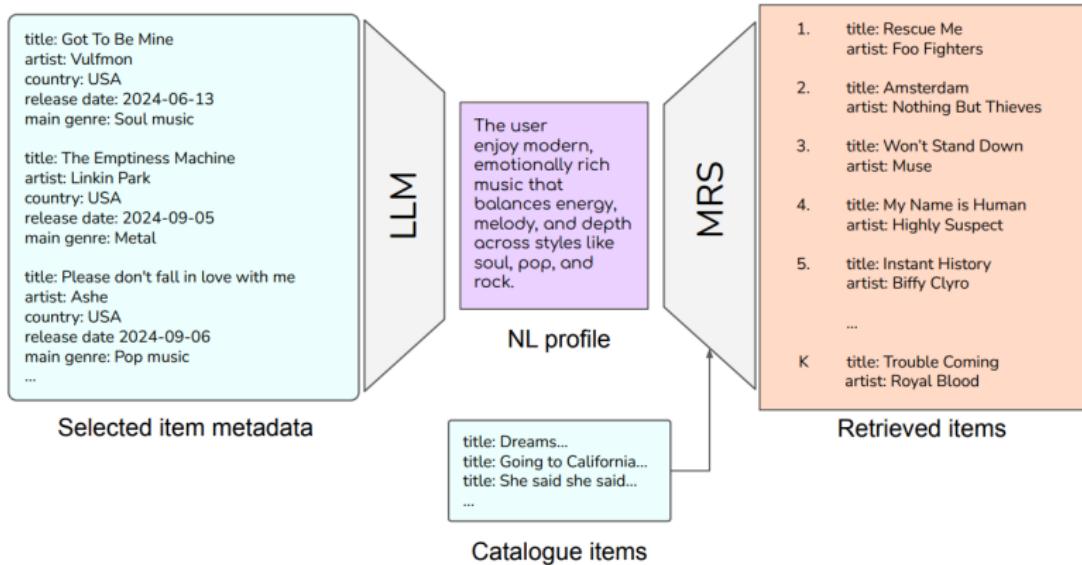
How can we leverage LLMs to obtain scrutable natural language profiles?

- ▶ Preference modeling has relied on interaction data to produce embeddings
- ▶ Vector-based methods lack interpretability and raise scrutability issues
- ▶ Scientific literature establishes a strong link between musical preferences and psychological traits
- ▶ We could use LLMs to infer these states and preferences, transforming them into natural language profiles
- ▶ Natural language descriptions serve as a transparent base that the user can verify and refine, and provide a context for the recommendation engine

User modeling



User modeling



→ Natural language profiles involve preferences, listening intents, personality, mood and emotion

User modeling

How can we better modelize user preferences, especially regarding their affective states?

User modeling

How can we better modelize user preferences, especially regarding their affective states?

- ▶ Existing research has explored mood and emotion detection from music listening data

User modeling

How can we better modelize user preferences, especially regarding their affective states?

- ▶ Existing research has explored mood and emotion detection from music listening data
- ▶ However, current models often rely on simplistic emotion categories (e.g., discrete emotions or valence-arousal dimensions)

User modeling

How can we better modelize user preferences, especially regarding their affective states?

- ▶ Existing research has explored mood and emotion detection from music listening data
- ▶ However, current models often rely on simplistic emotion categories (e.g., discrete emotions or valence-arousal dimensions)
- ▶ There is a need for more nuanced representations that capture the complexity of human emotions

User modeling

How can we better modelize user preferences, especially regarding their affective states?

- ▶ Existing research has explored mood and emotion detection from music listening data
- ▶ However, current models often rely on simplistic emotion categories (e.g., discrete emotions or valence-arousal dimensions)
- ▶ There is a need for more nuanced representations that capture the complexity of human emotions
- ▶ Incorporating cognitive science theories of emotion can enhance the understanding of user affective states

User modeling

How can we better modelize user preferences, especially regarding their affective states?

- ▶ Existing research has explored mood and emotion detection from music listening data
- ▶ However, current models often rely on simplistic emotion categories (e.g., discrete emotions or valence-arousal dimensions)
- ▶ There is a need for more nuanced representations that capture the complexity of human emotions
- ▶ Incorporating cognitive science theories of emotion can enhance the understanding of user affective states

User modeling

How can we better modelize user preferences, especially regarding their affective states?

- ▶ Existing research has explored mood and emotion detection from music listening data
- ▶ However, current models often rely on simplistic emotion categories (e.g., discrete emotions or valence-arousal dimensions)
- ▶ There is a need for more nuanced representations that capture the complexity of human emotions
- ▶ Incorporating cognitive science theories of emotion can enhance the understanding of user affective states

We could reuse some ideas to better modelize items' affective dimensions

Linguistic and cognitive science theories

Psychological theories and emotion annotation schemes

Psychological theories and emotion annotation schemes

What are current limitations and interesting research directions?

Psychological theories and emotion annotation schemes

What are current limitations and interesting research directions?

Psychological theories	In text, emotion is...	Example
Basic emotions theory	a category	"I love philosophy." → joy
Constructivist theories	a continuous value with an affective meaning	"His voice soothes me." → valence (4/5), arousal (1/5)
Appraisal theory	a continuous value with a cognitive meaning composed of semantic roles	"I received a surprise gift." → sudden (4/5), control (0/5) "Louise (experiencer) was angry (cue) towards Paul (target), because he didn't inform her (cause)."

Which verbal signs are used to infer expressed emotions?

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")
- ▶ *suggested* with the description of a situation which generally, in a given sociocultural context, leads to an emotion ("She gave me a gift")

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")
- ▶ *suggested* with the description of a situation which generally, in a given sociocultural context, leads to an emotion ("She gave me a gift")

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")
- ▶ *suggested* with the description of a situation which generally, in a given sociocultural context, leads to an emotion ("She gave me a gift")

→ Different emotion expression modes are more or less difficult to interpret

How to integrate psychological theories of emotion?

How to integrate psychological theories of emotion?

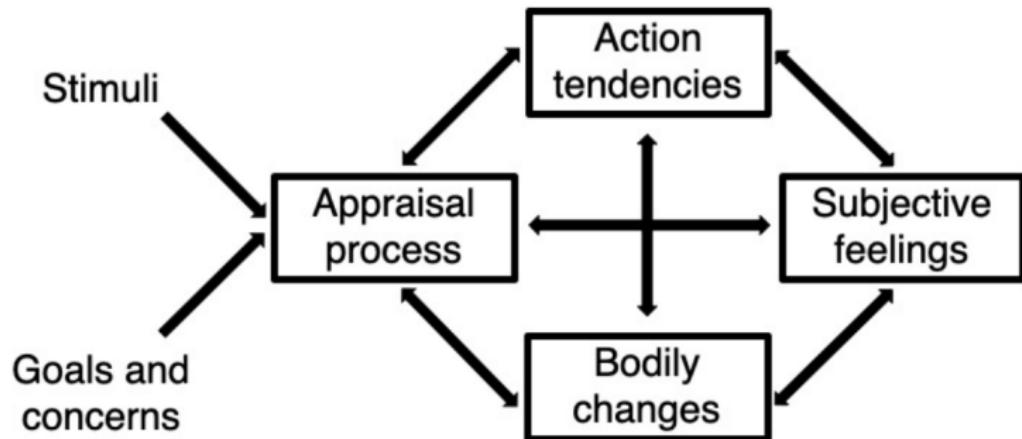


Figure: Integrated framework for emotion theories. Emotional episodes are synchronized changes in four components (Scherer, 2022).

Item modeling

Item modeling

Music item models leverage collaborative and content data

Item modeling

Music item models leverage collaborative and content data

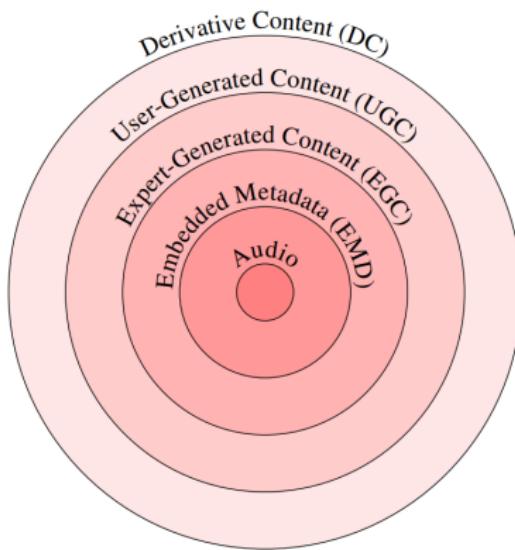


Figure: The *onion model* of music content. A visualization of content layers accumulating over time, ranging from objective descriptors (inner) to high-level, subjective cultural data (outer).

Item modeling

How can we construct language-based item representations that capture both intrinsic content and individual/cultural context?

Item modeling

How can we construct language-based item representations that capture both intrinsic content and individual/cultural context?

We could use LLMs for several roles:

Item modeling

How can we construct language-based item representations that capture both intrinsic content and individual/cultural context?

We could use LLMs for several roles:

- ▶ LLMs as *knowledge bases* to generate music tags, genres, or styles

Item modeling

How can we construct language-based item representations that capture both intrinsic content and individual/cultural context?

We could use LLMs for several roles:

- ▶ LLMs as *knowledge bases* to generate music tags, genres, or styles
- ▶ LLMs as *information extractors* to correct or augment metadata

Item modeling

How can we construct language-based item representations that capture both intrinsic content and individual/cultural context?

We could use LLMs for several roles:

- ▶ LLMs as *knowledge bases* to generate music tags, genres, or styles
- ▶ LLMs as *information extractors* to correct or augment metadata
- ▶ LLMs as *summarizers* of heterogeneous textual data to create a concise summary of a song

Music recommendation

Music recommendation

How can we leverage natural language queries for context-aware music recommendation using LLMs?

Music recommendation

How can we leverage natural language queries for context-aware music recommendation using LLMs?

- ▶ Users often employ broad concepts, such as “music for sports during a rainy day”, which implies a need to align music with specific contexts

Music recommendation

How can we leverage natural language queries for context-aware music recommendation using LLMs?

- ▶ Users often employ broad concepts, such as “music for sports during a rainy day”, which implies a need to align music with specific contexts
- ▶ Given an user profile and a natural language query, how to recommend the best music items?

Music recommendation

How can we leverage natural language queries for context-aware music recommendation using LLMs?

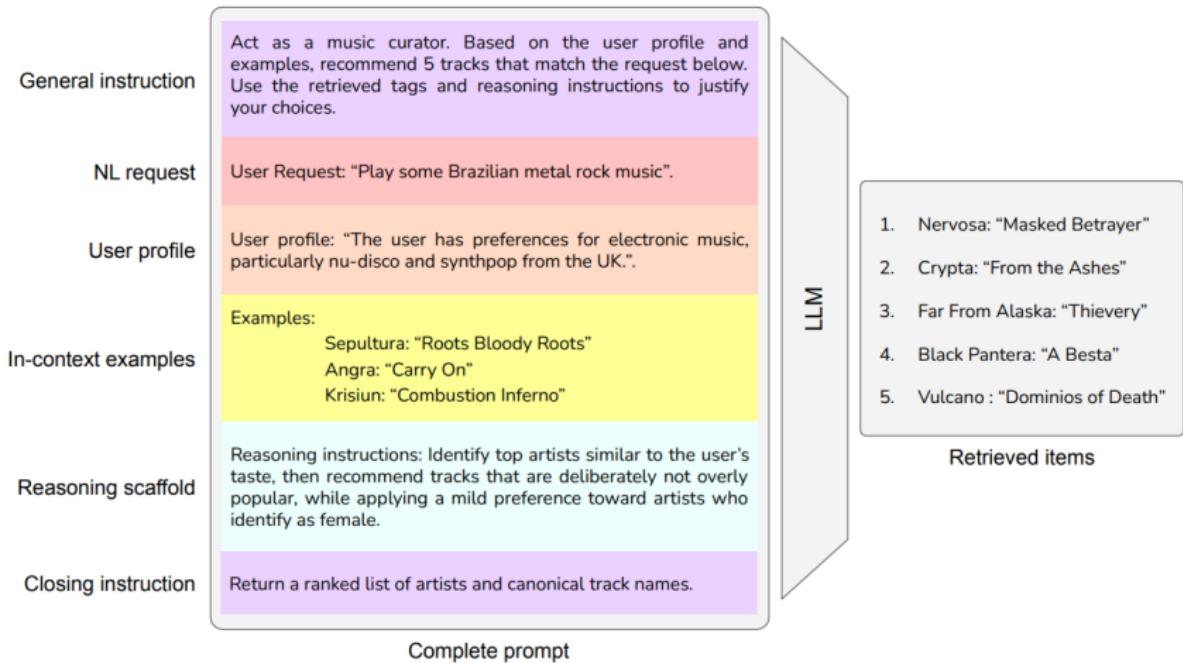
- ▶ Users often employ broad concepts, such as “music for sports during a rainy day”, which implies a need to align music with specific contexts
- ▶ Given an user profile and a natural language query, how to recommend the best music items?
- ▶ Previous methods often rely on keyword matching, which limits the expressiveness of user intents

Music recommendation

How can we leverage natural language queries for context-aware music recommendation using LLMs?

- ▶ Users often employ broad concepts, such as “music for sports during a rainy day”, which implies a need to align music with specific contexts
- ▶ Given an user profile and a natural language query, how to recommend the best music items?
- ▶ Previous methods often rely on keyword matching, which limits the expressiveness of user intents
- ▶ LLMs can interpret complex natural language queries, enabling more nuanced recommendations

Music recommendation



Music recommendation

How can we optimize prompts to improve recommendation quality?

Music recommendation

How can we optimize prompts to improve recommendation quality?

- ▶ Prompt engineering is crucial for eliciting desired responses from LLMs

Music recommendation

How can we optimize prompts to improve recommendation quality?

- ▶ Prompt engineering is crucial for eliciting desired responses from LLMs
- ▶ Manual prompt design can be time-consuming and may not yield optimal results

Music recommendation

How can we optimize prompts to improve recommendation quality?

- ▶ Prompt engineering is crucial for eliciting desired responses from LLMs
- ▶ Manual prompt design can be time-consuming and may not yield optimal results
- ▶ Automated prompt optimization techniques can explore the prompt space to find effective formulations

Music recommendation

How can we optimize prompts to improve recommendation quality?

- ▶ Prompt engineering is crucial for eliciting desired responses from LLMs
- ▶ Manual prompt design can be time-consuming and may not yield optimal results
- ▶ Automated prompt optimization techniques can explore the prompt space to find effective formulations
- ▶ We could improve recommender prompts by optimizing natural language instructions and few-shot examples

Music recommendation

How can we optimize prompts to improve recommendation quality?

- ▶ Prompt engineering is crucial for eliciting desired responses from LLMs
- ▶ Manual prompt design can be time-consuming and may not yield optimal results
- ▶ Automated prompt optimization techniques can explore the prompt space to find effective formulations
- ▶ We could improve recommender prompts by optimizing natural language instructions and few-shot examples
- ▶ We could improve LLM-as-a-judge prompts to better measure subjective dimensions of recommendation quality (e.g., groundedness, discovery quality, personalization gain, profile fidelity, and cultural/linguistic coverage)

DSPy prompt optimization

```
1 class ExtractInfo(dspy.Signature):
2     """Extract structured information from text."""
3
4     text: str = dspy.InputField()
5     title: str = dspy.OutputField()
6     headings: list[str] = dspy.OutputField()
7     entities: list[dict[str, str]] = dspy.OutputField(desc="a list of entities and their metadata")
8
9     module = dspy.Predict(ExtractInfo)
10
11    text = "Apple Inc. announced its latest iPhone 14 today." \
12        "The CEO, Tim Cook, highlighted its new features in a press release."
13    response = module(text=text)
14
15    print(response.title)
16    print(response.headings)
17    print(response.entities)
```

Conclusion

Conclusion

- ▶ The research proposal aims to develop language-based representations for music recommendation using LLMs

Conclusion

- ▶ The research proposal aims to develop language-based representations for music recommendation using LLMs
- ▶ Three interconnected axes: user modeling, item modeling, and music recommendation

Conclusion

- ▶ The research proposal aims to develop language-based representations for music recommendation using LLMs
- ▶ Three interconnected axes: user modeling, item modeling, and music recommendation
- ▶ User modeling: *How can we leverage LLMs to obtain scrutable natural language profiles? How can we better modelize user preferences, especially regarding their affective states?*

Conclusion

- ▶ The research proposal aims to develop language-based representations for music recommendation using LLMs
- ▶ Three interconnected axes: user modeling, item modeling, and music recommendation
- ▶ User modeling: *How can we leverage LLMs to obtain scrutable natural language profiles? How can we better modelize user preferences, especially regarding their affective states?*
- ▶ Item modeling: *How can we construct language-based item representations that capture both intrinsic content and individual/cultural context?*

Conclusion

- ▶ The research proposal aims to develop language-based representations for music recommendation using LLMs
- ▶ Three interconnected axes: user modeling, item modeling, and music recommendation
- ▶ User modeling: *How can we leverage LLMs to obtain scrutable natural language profiles? How can we better modelize user preferences, especially regarding their affective states?*
- ▶ Item modeling: *How can we construct language-based item representations that capture both intrinsic content and individual/cultural context?*
- ▶ Music recommendation: *How can we leverage natural language queries for context-aware music recommendation using LLMs? How can we optimize prompts to improve recommendation quality?*