

Affective Computing - Empathic Artificial Intelligence

Dr. Marco Maier, Philipp Altmann

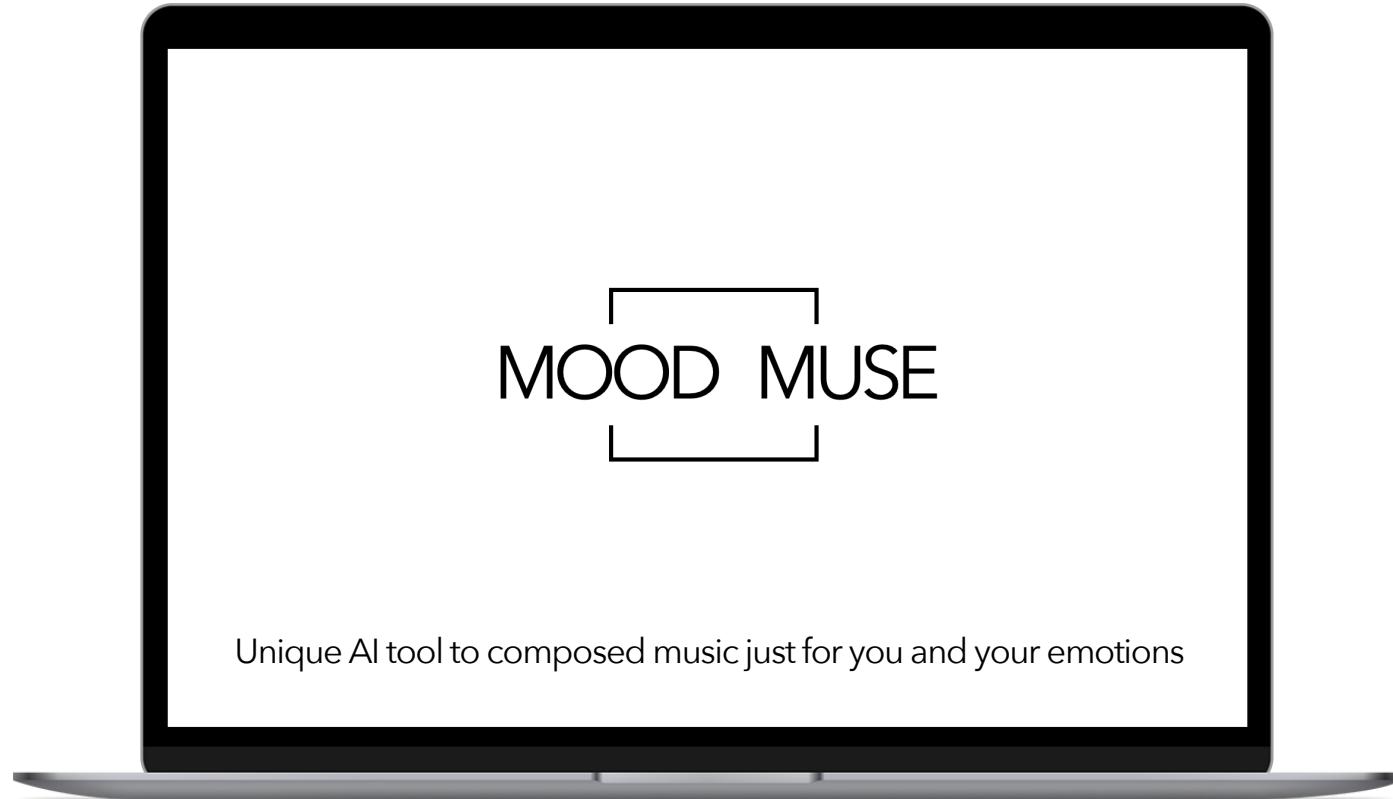
Final presentation, July 19th 2023

Ardit Mazreku, Franziska Wörle, Laura Schröder, Lea Goerl, Phi Linh Phan

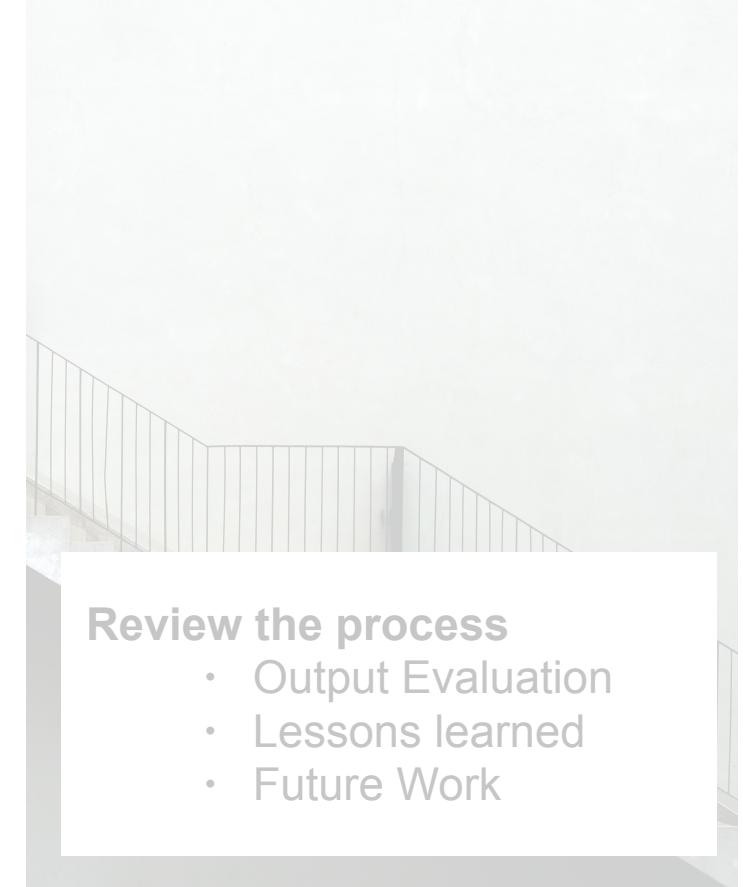
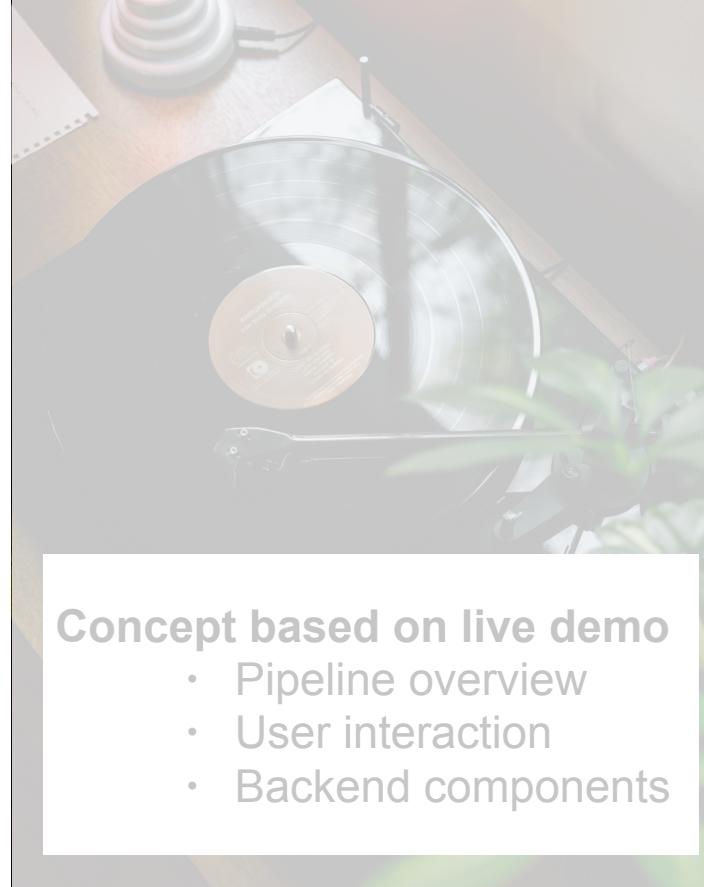
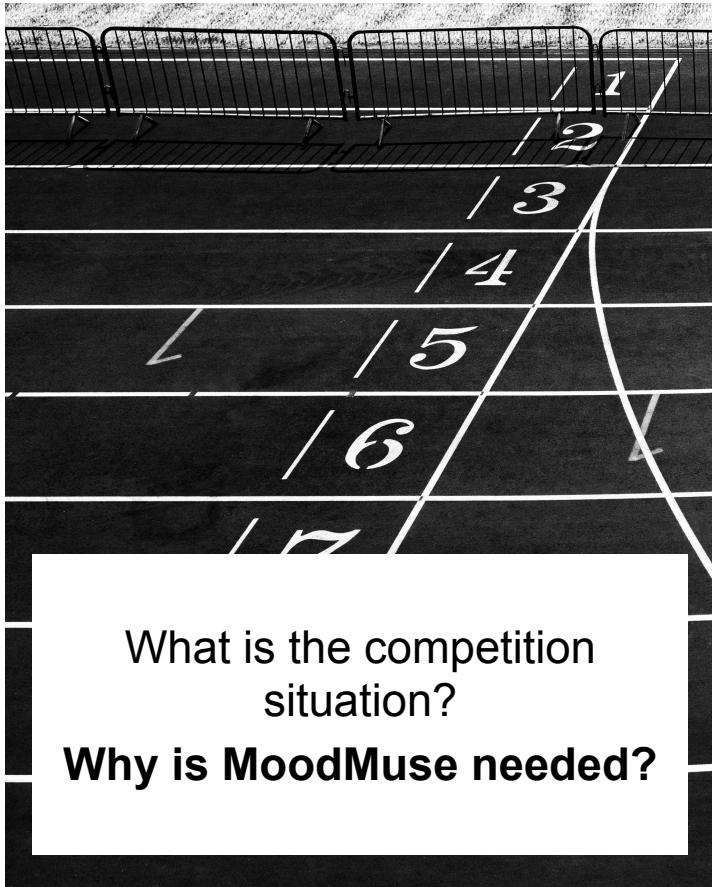


„I regularly post reels and photos on Instagram. I almost always add music to create the right atmosphere and feeling. Music brings it all to life.“ - **Franzi**





Content road

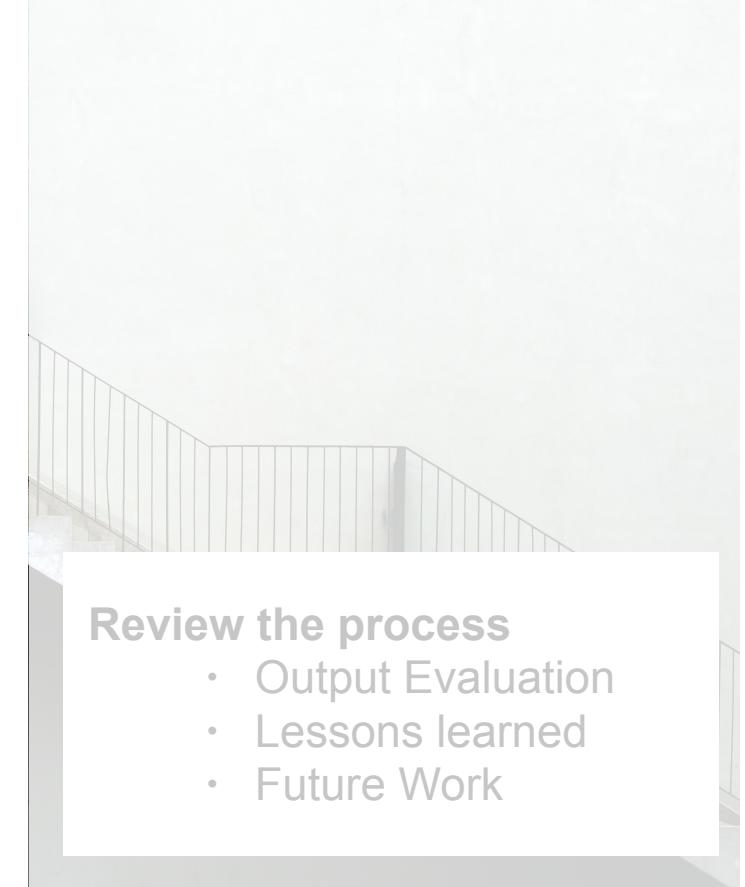
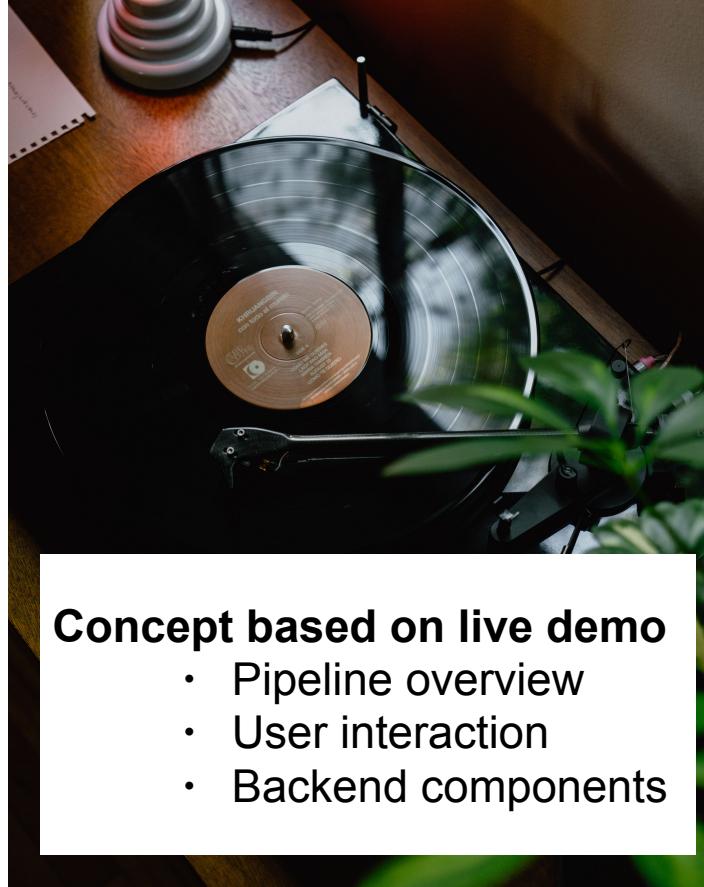
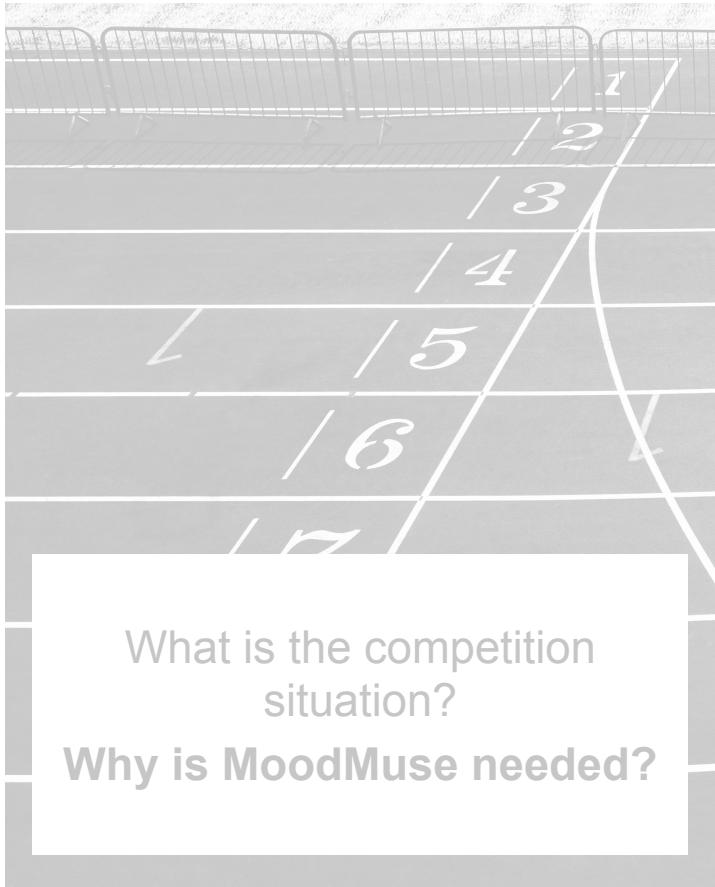


What is the competition situation and why is MoodMuse needed?

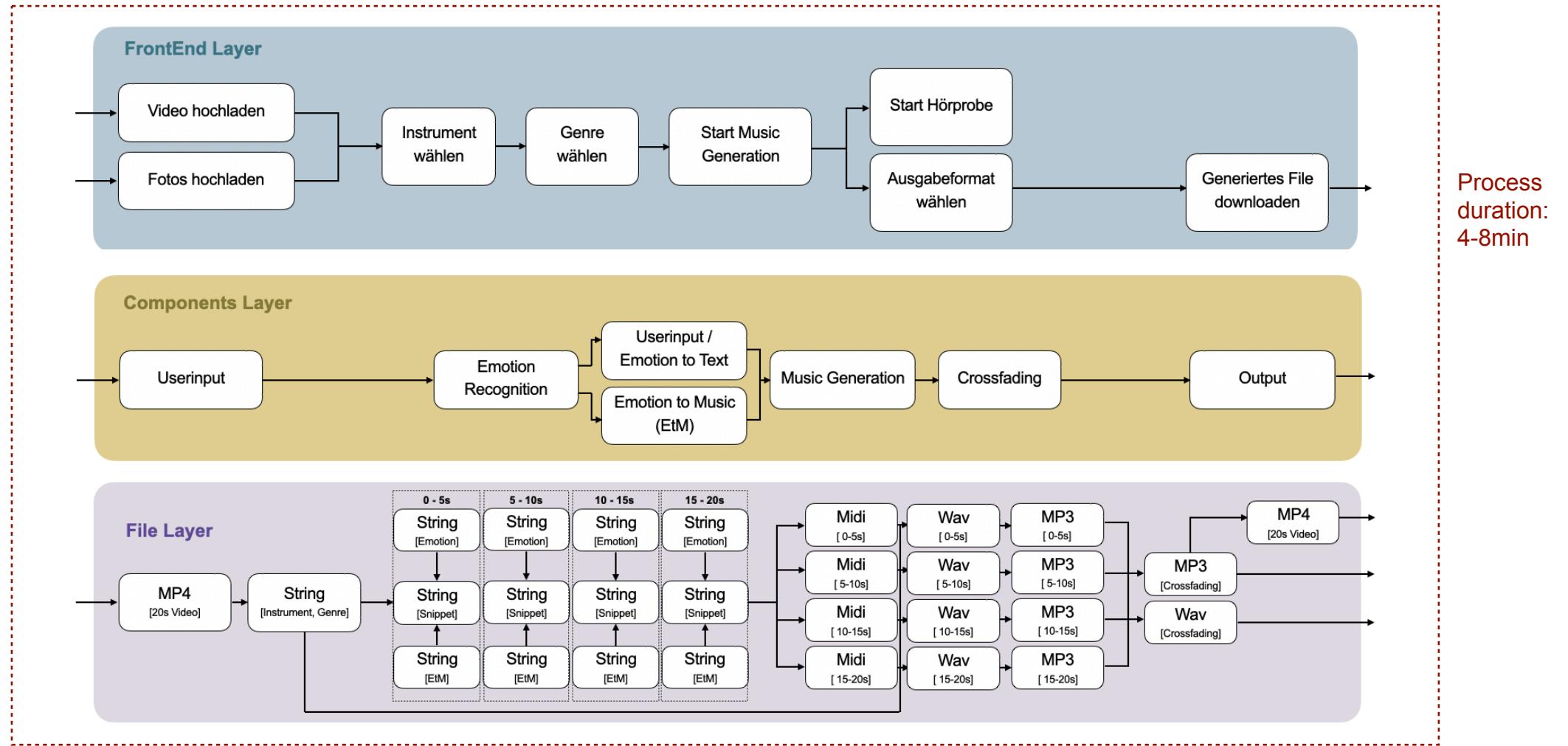
Landscape: AI tools for generating music

	Mubert	MusicLM	Jukebox	MusicGen	Magenta	Spotify
Input	Text	Text	Singing	Text	Text	Songs
Constraints	<ul style="list-style-type: none"> • Watermarks • No access • No API 	<ul style="list-style-type: none"> • Access with waiting list • No API 	<ul style="list-style-type: none"> • Input • Short samples • Small data set 			<ul style="list-style-type: none"> • Only song radio
Not included: Expressions of emotions, mood						

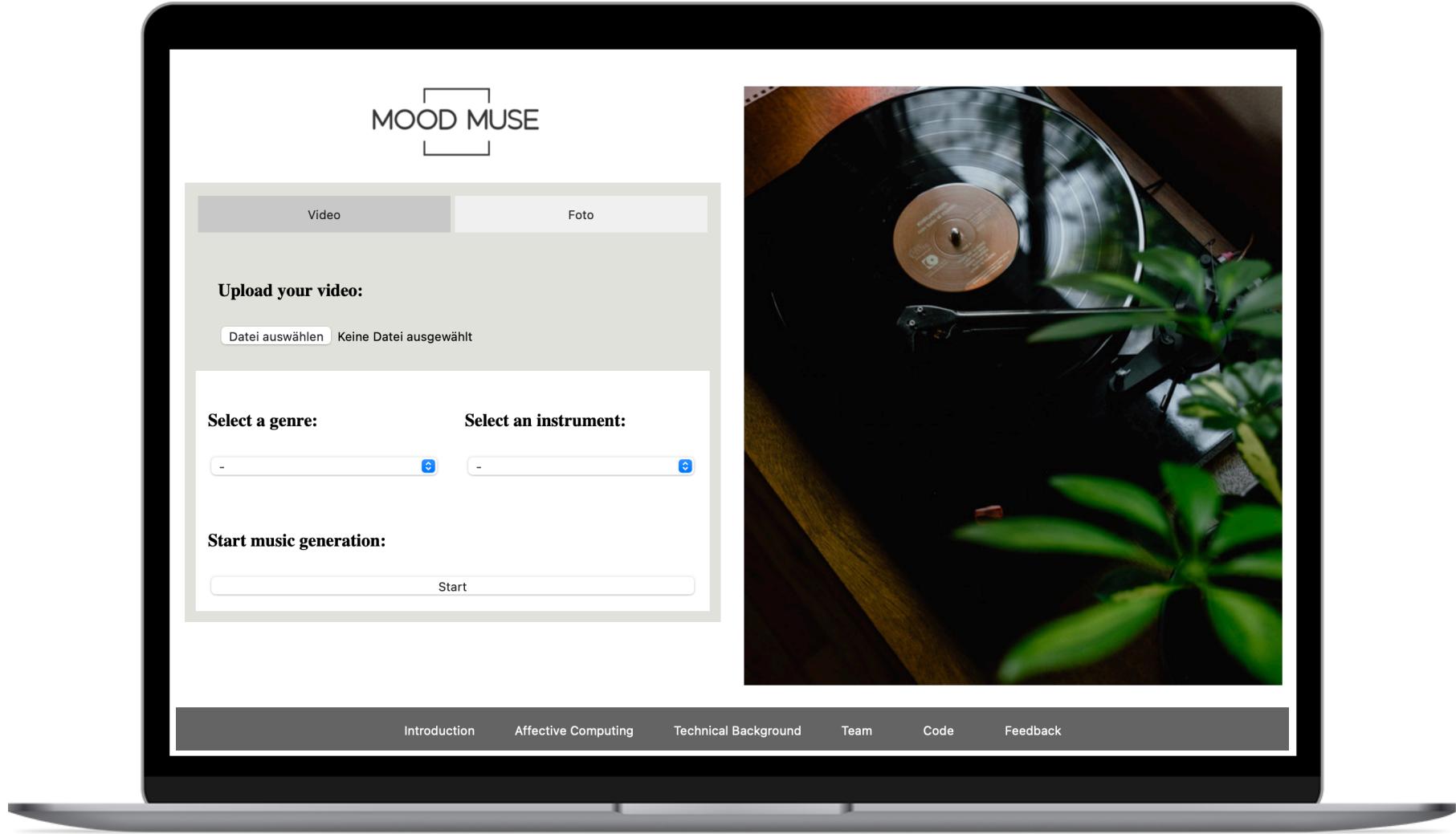
Content road



Pipeline overview

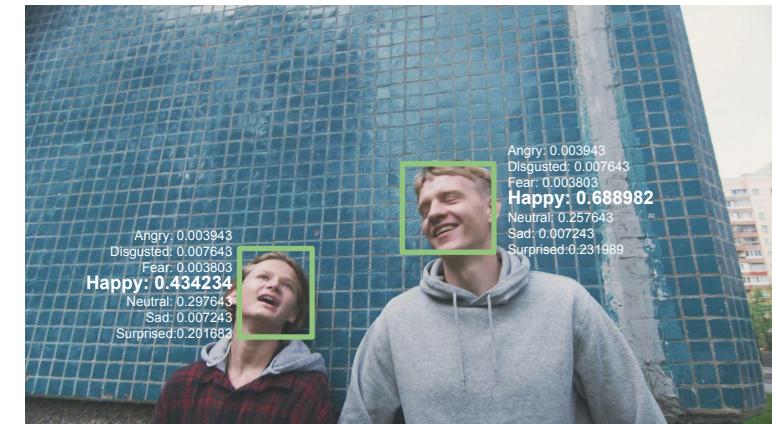
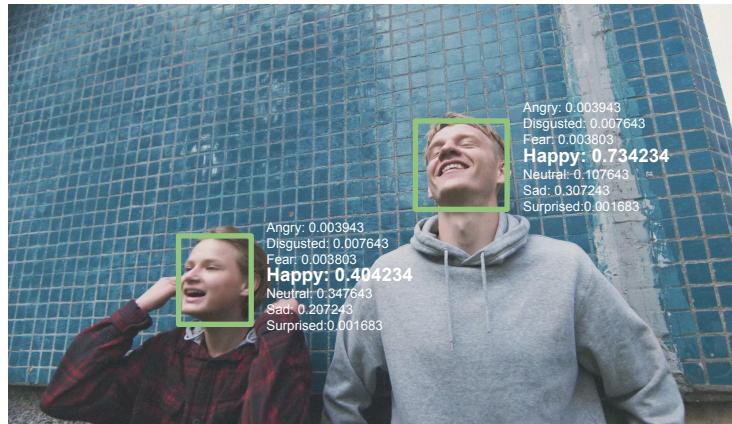
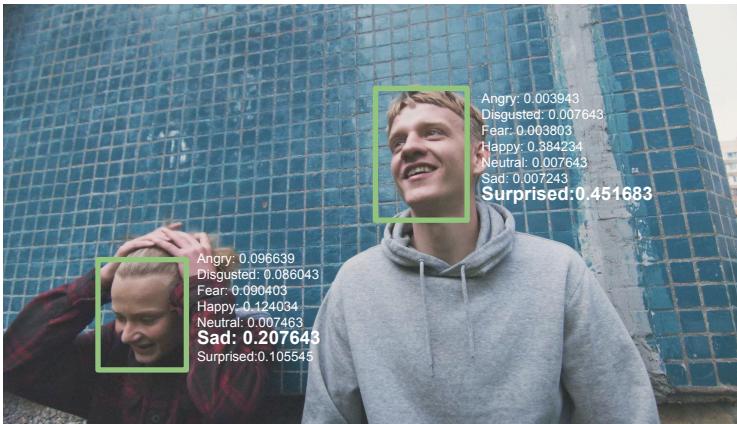


User interaction



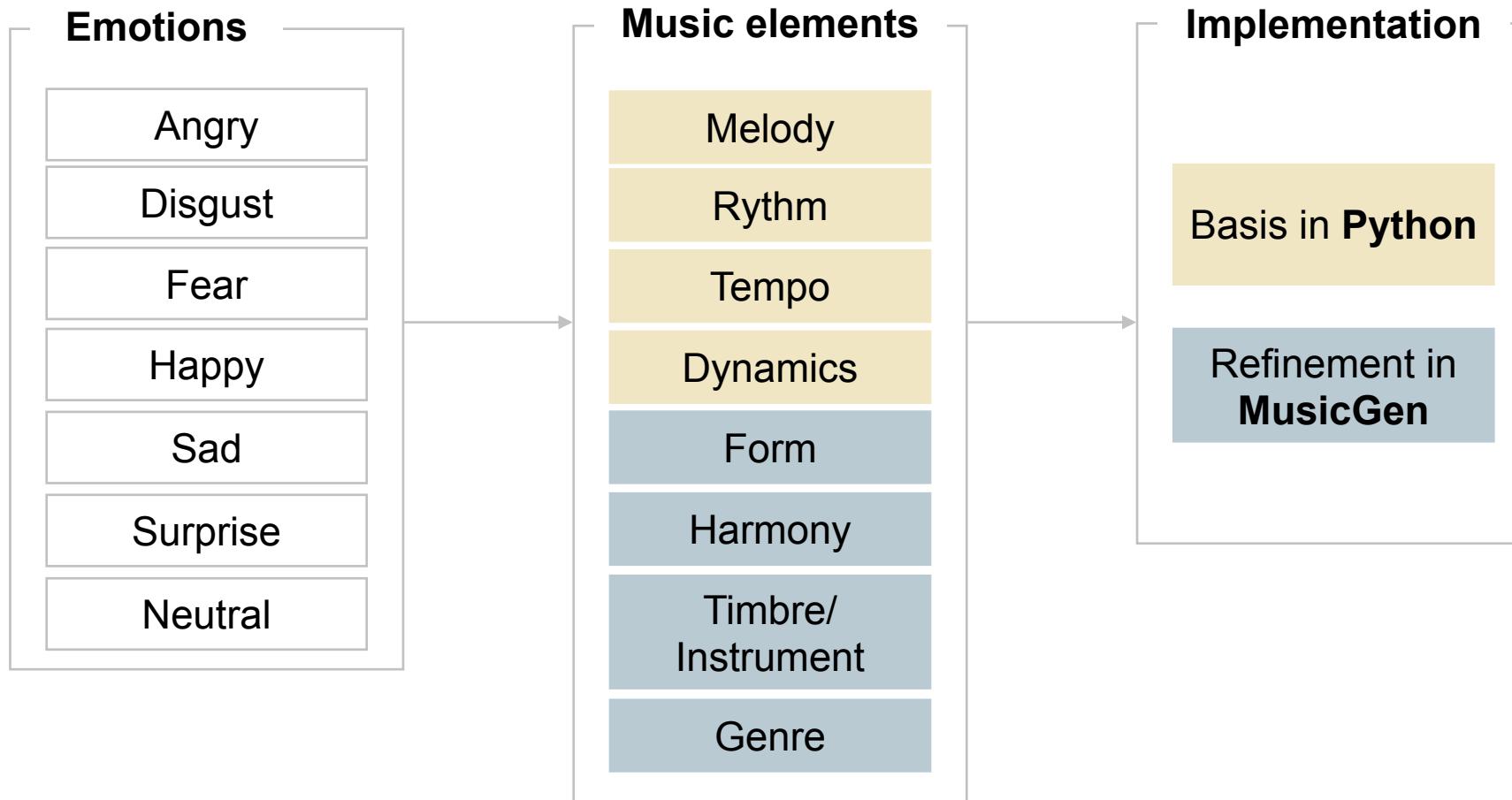
Backend Components - Face Emotion Recognition

- Dataset: FER2013
- Python package: FER
 - Face Detection MTCNN
 - Emotion Classification: mini-Xception
- Use: FER in Videos and Images



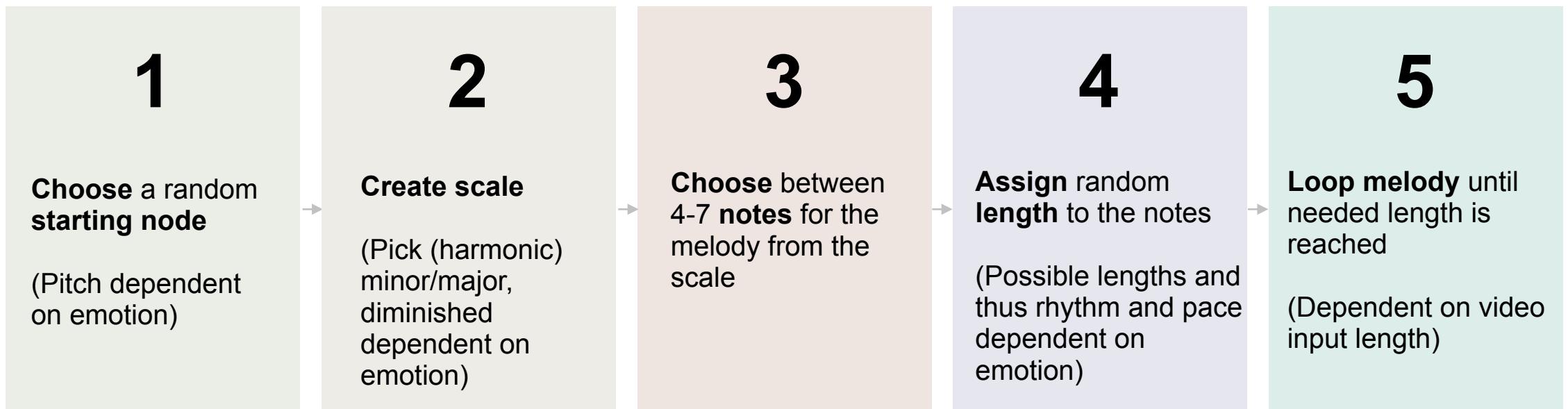
output = [[**Surprised**, Happy, Neutral], [**Happy**, Sad, Neutral], [**Happy**, Neutral, Surprised], [**Happy**, Neutral, Angry]]

Backend Components - *Mapping Emotion-to-Text / Emotion-to-Music*



Backend Components - *Melody creation: The musical logic behind it*

We use the Python library **midiutil** to generate melodies. For the **first** musical snippet, we...



To ensure a harmonic fade between the snippets, the **second** snippet has the **same or a parallel pitch** (e.g. when fading from major to minor, the parallel minor pitch is used).

Backend Components - *MusicGen as powerful tool*

We use the tool MusicGen by META to **refine our melodies** and as a **practical enhancement** to our composition. MusicGen...

... performs **music generation** based on **text description** and **melody conditioning**

... is trained on a dataset of 20,000 hours of licensed music

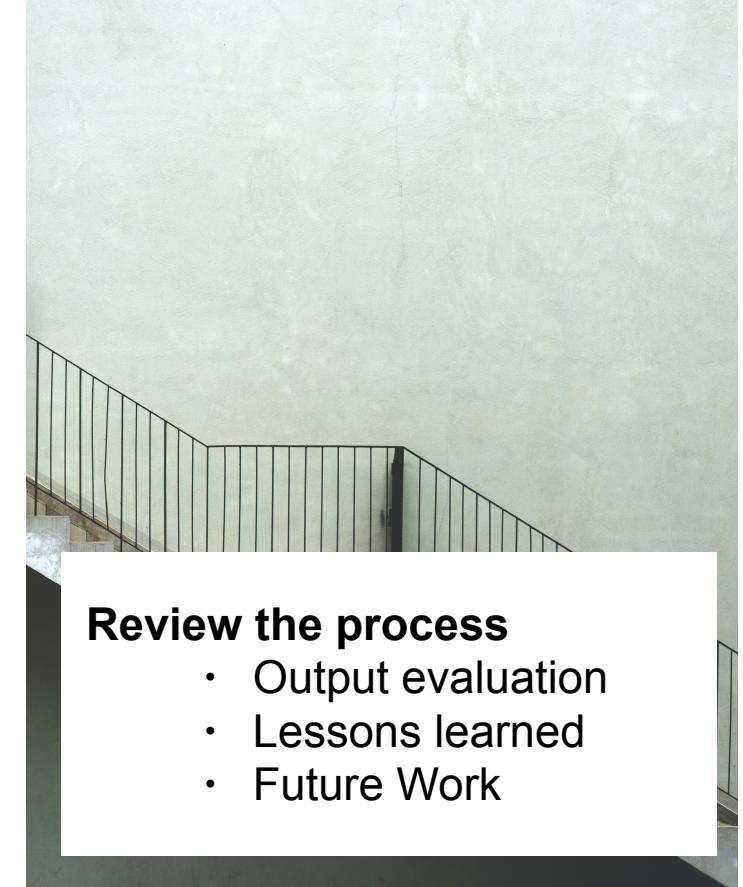
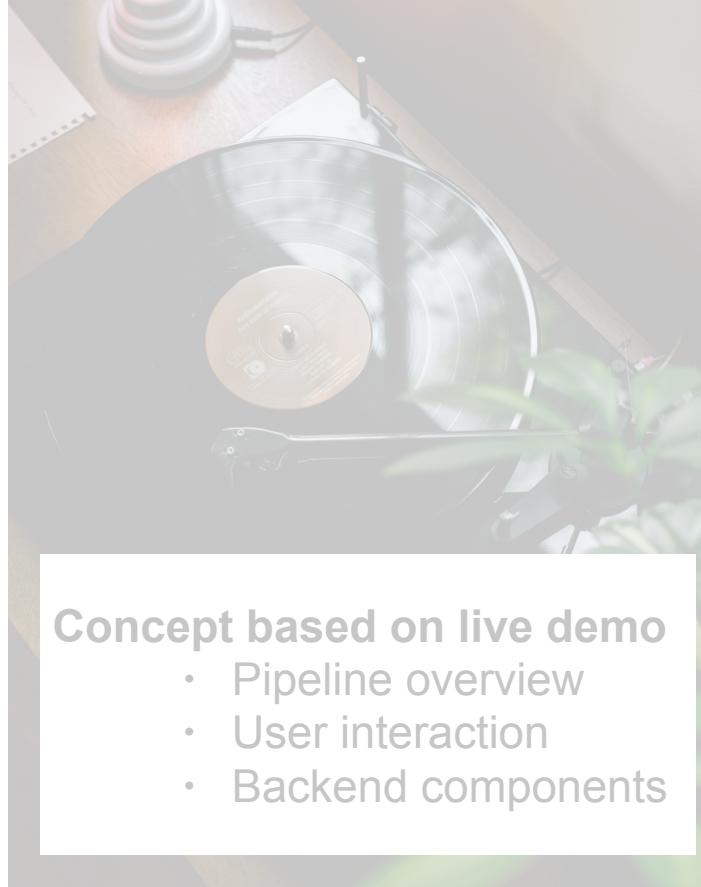
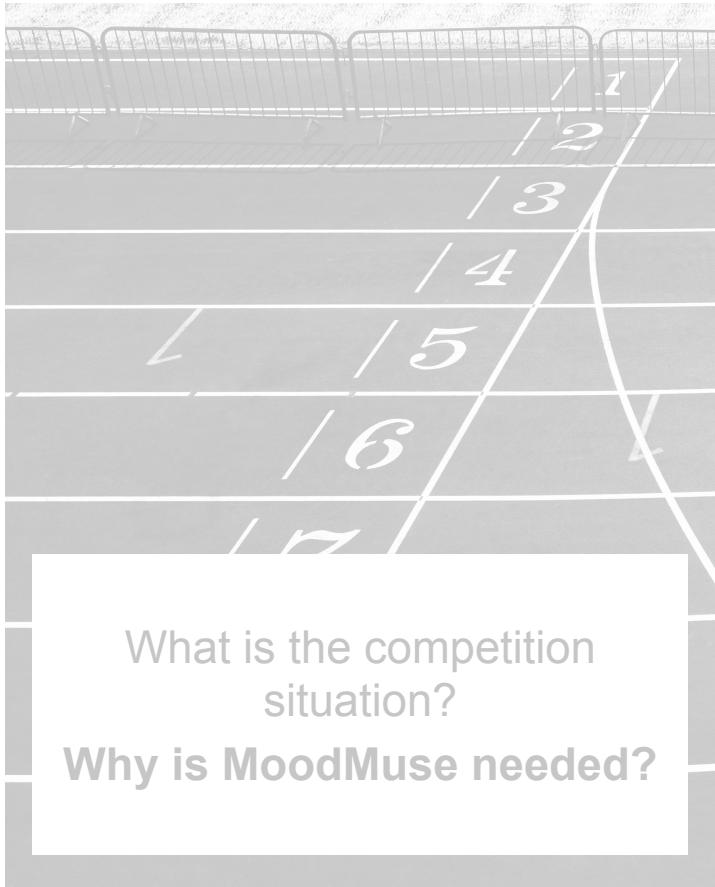
... incorporates a language model which utilizes quantized units from the EnCodec audio tokenizer (provides high-fidelity neural audio compression and tokenization)



♫ Our melody ensures that there is some musicality to it

➡️ MusicGen allow us to easier incorporate genres, instruments, accompaniment, dynamic volumes

Content road



Output evaluation - *Melodie refinement: MusicGen Do's*

The text prompt, which accompanies our melody, need to be precise yet concise.

Out solution:

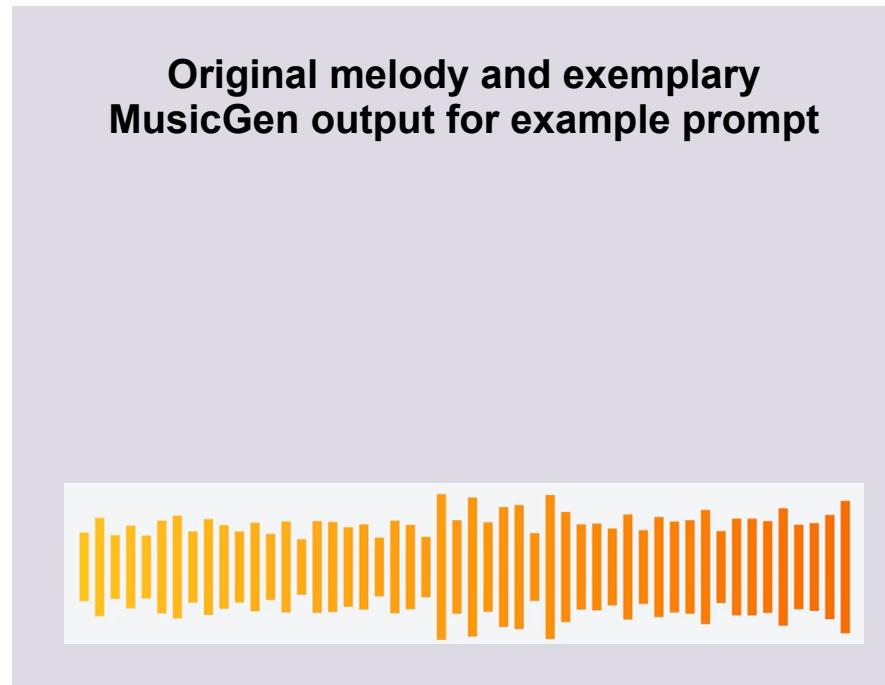
Template

```
„A“ + <adjective 1> +  
      <adjective 2> +  
      <genre> + „song with“  
      <instrument>
```

Example

```
„A spooky and  
suspenseful folk song  
with violin.“
```

Directly definition:	genre, instrument (chosen by user)
Indirectly definition:	Volume
Music Gen adds:	Accompaniment, further instruments



Output evaluation - *Melody refinement: MusicGen Dont's*

Poorly chosen melody length:

- Too short: unpredictable composition, might contain 5 seconds of silence or just noisiness
- Too long: prompt instructions not well met, processing time quite long

More words/more precise descriptions do not improve outcome quality

Not all prompts yield the desired outcome

- Some don't work reliably: choir singling scarcely
- Some don't work with melody conditioning: soft, fast, ..

95% desired outcome, 5% time unpredictability



Lessons learned - *Technical Surrounding*

- **Colab**
 - Version control is very difficult
 - Parallel work on several features is difficult
 - Multiple notebooks required for testing
- **Frontend** hosting is not very developer friendly via Google Drive and Colab
- **Transitioning** from Google Drive became too complex at the moment of realisation

No attractive alternative providing cloud storage, programming environments and collaboration tools

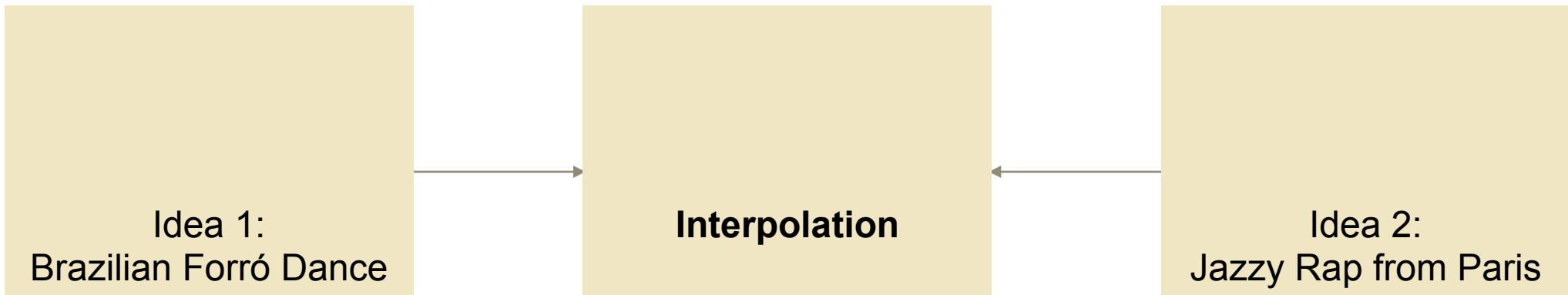
Lessons learned - Scope of the Tool



Faster focus would have been beneficial

Future Work

- Live Video Recording
- Better runtime performance
- More attributes for photo and video recognition
 - Natur and landscape
 - Objects
- Interpolation of the musical sections instead of crossfade for smoother and more interesting transitions



User interaction

