

# Text-to-Music Generation with Emotional Nuance: Multi- dimensional Emotion Space Modeling

---

By: Josephine, Malina, Wasif  
Team name: Resona



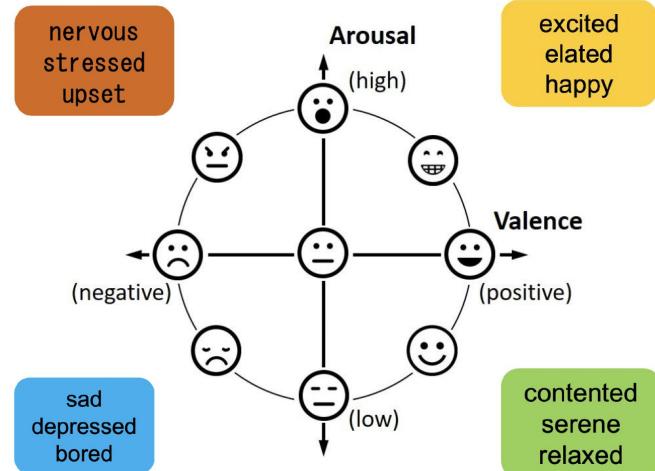
# Background & Research Hypothesis

**Problem:** "sad but hopeful" versus "purely tragic" ?

**Hypothesize:** VA-based emotion conditioning will enable music generation models to generate musically distinct, nuanced emotional outputs.

Examples	MusicGen
	Input Text
	An 80s driving pop song with heavy drums and synth pads in the background
	A cheerful country song with acoustic guitars
	90s rock song with electric guitar and heavy drums
	a light and cheerly EDM track, with syncopated drums, aery pads, and strong emotions
	lofi slow bpm electro chill with organic samples
	Punk rock with loud drum and power guitar

Russell's Circumplex Model of Affect (VA model)



Mahmood, Mahmood A., et al. "Segmentation-enhanced approach for emotion detection from EEG signals using the fuzzy C-mean and SVM." Scientific Reports 15.1 (2025): 31956.

# Methodology & Tools

## Step 1: Data Collection

## Step 2: Emotion Conditioning

## Step 3: Model Fine-tuning

## Step 4: Evaluation

## Step 5: Optimization

- Compile a standardized VA dataset
- Emotion Encoder
- Modified Cross-Attention
- Leveraging Low-Rank Decomposition (LoRA) method
- Interactive Demo
- Human-in-the-Loop Evaluation: A/B Testing & VA-Value Rating
- Refine the model based on the feedback gathered from the evaluations

 PyTorch



**Hugging Face**

Pre trained model: MusicGen



**gradio**

microsoft/**LoRA**

Code for loralib, an implementation of "LoRA: Low-Rank Adaptation of Large Language Models"



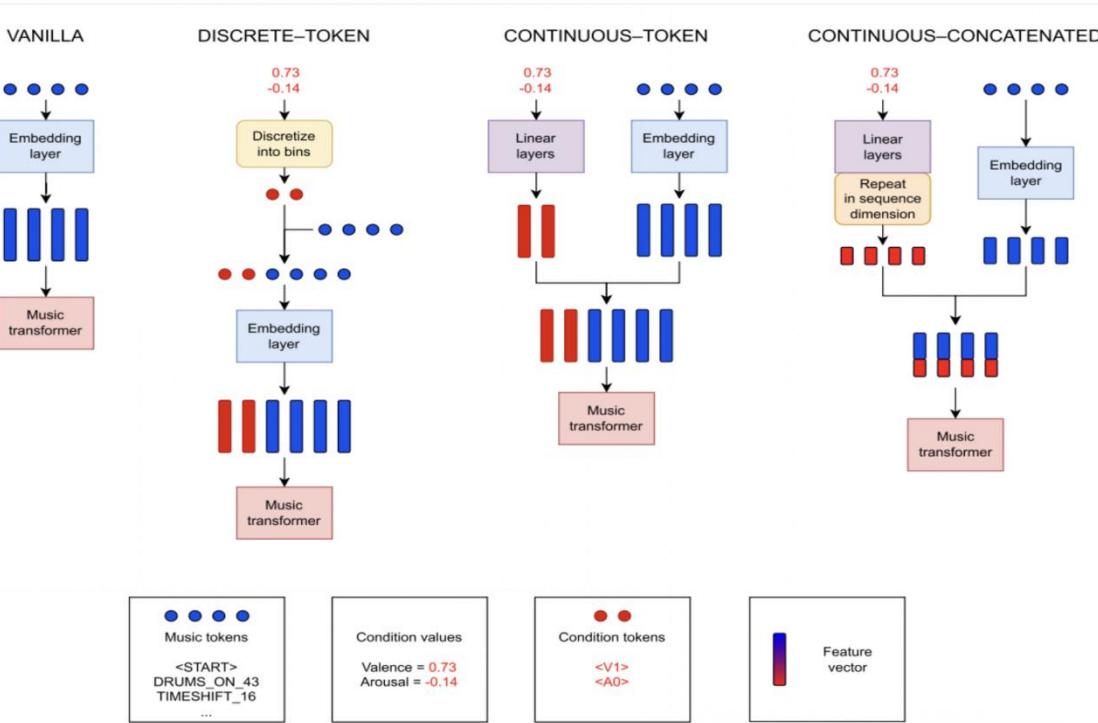
6 Contributors 81 Used by 1k Stars 85 Forks

**kaggle**

**colab**

# Preliminary Results

## Four Architectures for Emotion Conditioning in Music Transformers



## Model performance comparison

Model	NLL	Top-1	Top-5
vanilla	0.7445	0.7784	0.9513
discrete-token	0.7375	0.7885	0.9536
continuous-token	0.7122	0.7895	0.9545
continuous-concatenated	<b>0.7075</b>	<b>0.7913</b>	<b>0.9548</b>

- **NLL (Negative Log-Likelihood):** Error metric **lower value** means better.
- **Top-1 and Top-5 Accuracy:** Accuracy metrics **higher value** means better.

S. Sulun et al. Symbolic music generation conditioned on continuous-valued emotions. IEEE Access, 10:44617–44626, 2022.

Spaces | patgpt4/MusicGen | like 12 | Running

## MusicGen

This is your private demo for [MusicGen](#), a simple and controllable model for music generation presented at:

**Input Text**  
a sad but hopeful piano piece

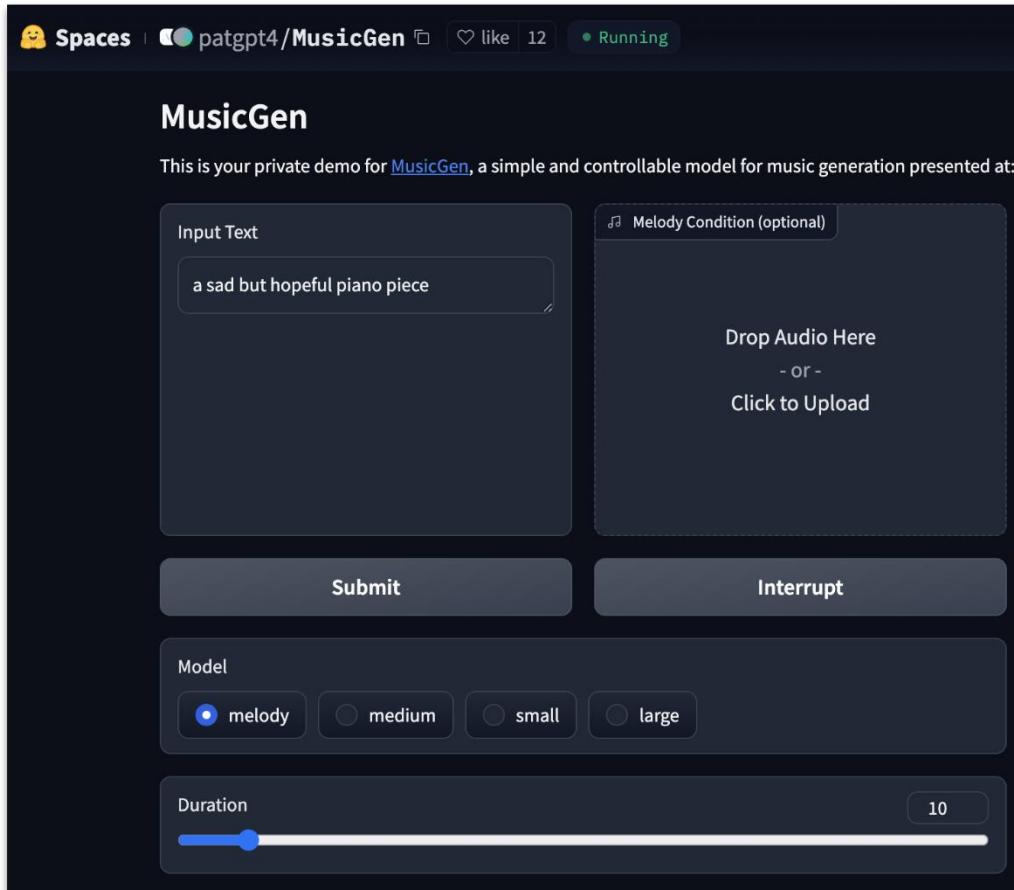
**Melody Condition (optional)**

Drop Audio Here  
- or -  
Click to Upload

**Submit**    **Interrupt**

**Model**  
 melody    medium    small    large

**Duration**



# Potential Impact & Outcomes

**Goals:** Introduces fine-grained, psychologically grounded emotion modeling into music generation.

**Contributions:** A VA-conditioned text-to-music framework enabling finer emotional control.

**Deliverables:** Micropublication, open-source code, interactive demo.

**Impact:** Advances emotion-aware AI music generation for potential applications of personalized media, therapy, and creative tools.

# Work Plan & Timeline

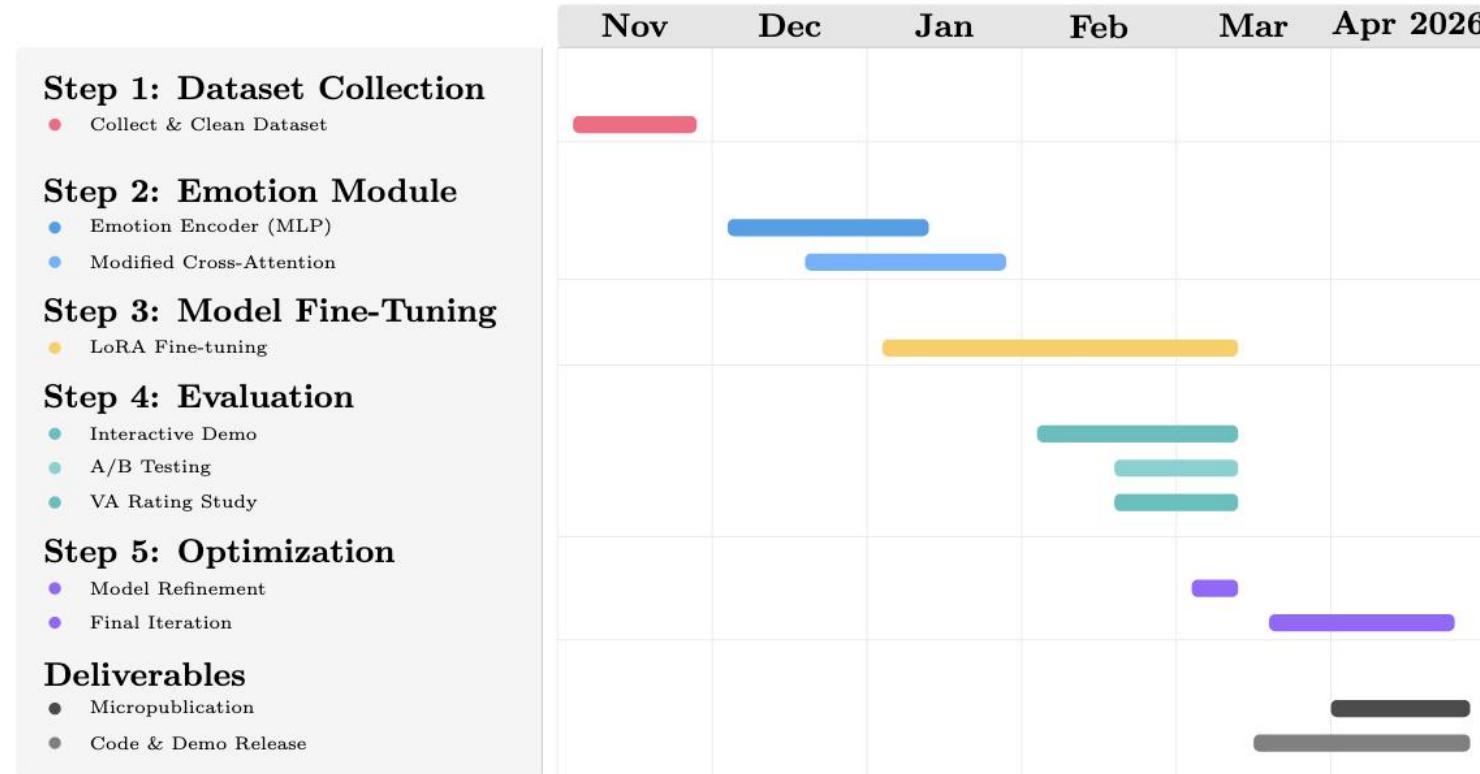


Figure 1: Project Gantt Chart: Nov 2025 - Apr 2026 Implementation Timeline