

5C34 : ADVANCED AI EMOTION DRIVEN MUSIC GENERATOR

Alba Carrilo Martinez
Haoran Zhang
Ishan Samir Thoke

Trinity College Dublin
MSc. Electronic Information Engineering
Project: Plan Presentation

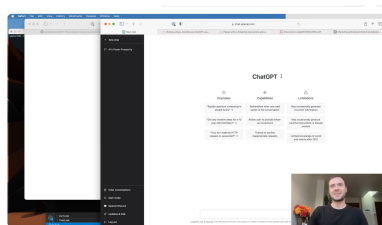
2026.2.5



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Motivation and Problem Statement

- In the first project meeting we realised our interests lie in the signal processing + ML domain and were looking to building a multi-modal, interactive AI.
- Easy, expressive text → structured music (ABC notation)
- We found a dataset and went through the youtube video by Andrej Karpathy on building GPT from scratch.
- We also went hunting on Kaggle and GitHub to look for codes and jupyter notebooks but as they as
- OpenAI is not Open Anymore :(



Let's build GPT: from scratch, in code, spelled out.

Why not train it from scratch?

And we decided against it

- **Data requirements:** large, curated corpora required to match SOTA; dataset collection/cleaning is expensive.
- **Compute & time:** full pre-training needs weeks on many GPUs impractical for project timeframe and resource limits, also I doubt we have enough money to pay for GPUs
- **Diminishing returns:** smaller fine-tuning + prompt engineering often equals or exceeds custom small-scale training for demo tasks.
- **Preferred approach (after a series of emails with Rishabh):** Integrate pre-trained models from hugging face and optimize the pipeline, change parameters accordingly rather than full pre-training.

Reality catching up to us was little harsh but all we wanna do is BUILD SOMETHING COOL!

What we decided to finally build

Title: Solution: Pre-Trained Model Pipeline

Emotion Classification Model

Model:
bhadresh-savani/distilbert-base-uncased-emotion

Architecture: DistilBERT (distilled version of BERT)

The screenshot shows the Hugging Face model card for 'musicgen-small-900h'. The card includes a title, a description of the model's capabilities, and a table of download statistics. The description states that MusicGen is a text-to-music model capable of generating high-quality music samples conditioned on text descriptions or audio prompts. It is a single-stage auto-regressive Transformer model trained over a 324k Emotion Tokenizer with 4 codebooks sampled at 30 Hz. The card also lists the authors (Meta, Facebook, and the Hugging Face team) and the license (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International).

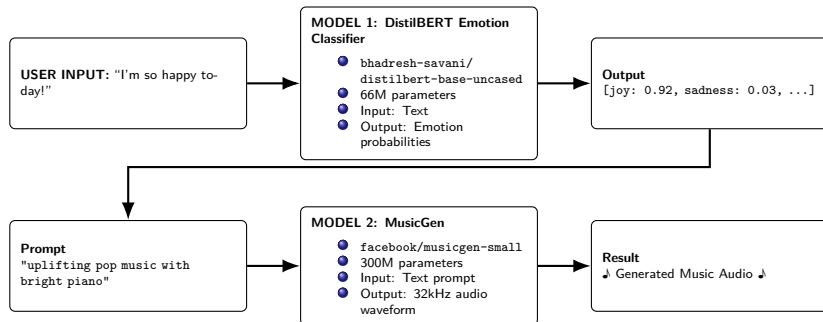
Music Generation Model

Model:
facebook/musicgen-small (or musicgen-medium)

Architecture: Transformer-based audio generation model

The screenshot shows the Hugging Face model card for 'distilbert-base-uncased-emotion'. The card includes a title, a description of the model's capabilities, and a table of download statistics. The description states that this is a distilled version of the BERT model, trained on the English Wikipedia dataset. The card also lists the authors (Hugging Face) and the license (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International).

Pipeline Overview



Thank you!

Alba Carrilo Martinez
Haoran Zhang
Ishan Samir Thoke