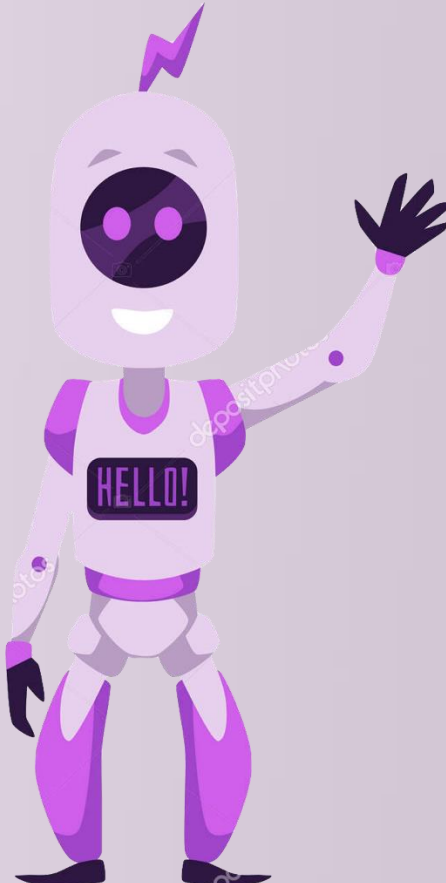
Music Generation from Textual Input

Giulia Saresini, Vanessa Maeder



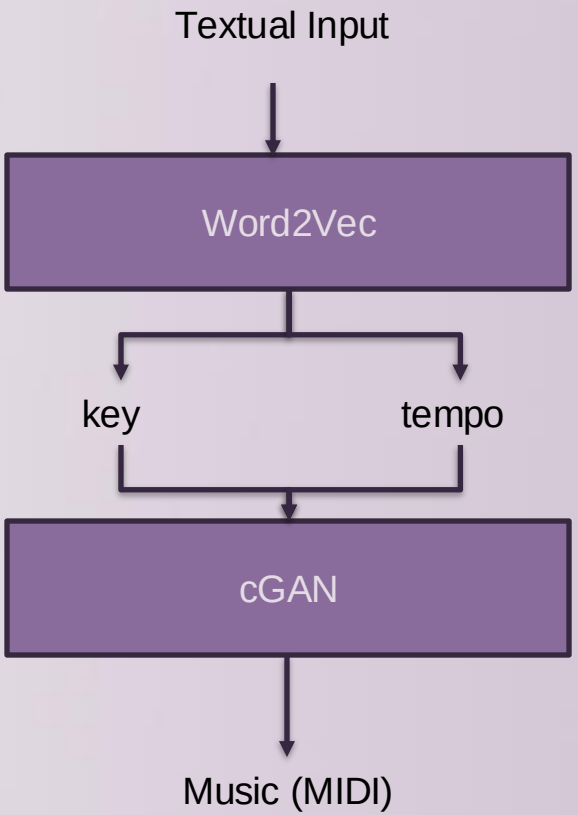
Motivation

Can a neural network truly understand and create music like human?

Music is an intricate form of expression, rich in **patterns**, **emotions**, and **creativity**. This project explores how neural networks can interpret and process musical structures and whether they can generate original compositions from scratch.

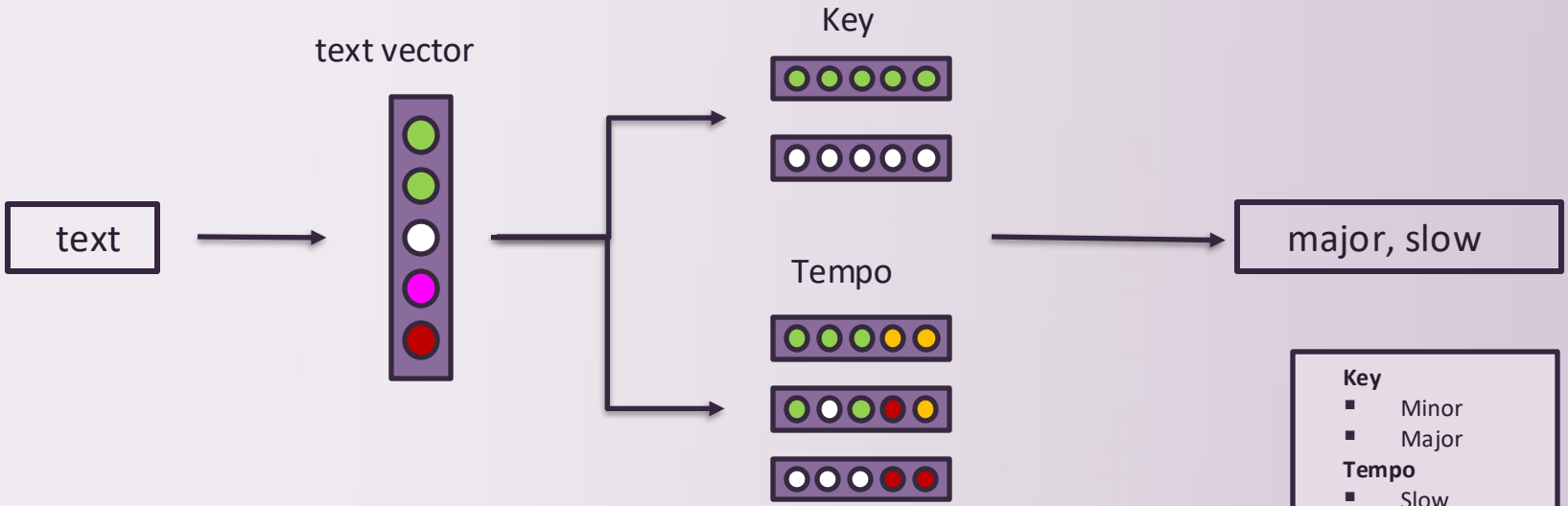
Approach to the Problem

To address the question, we adopted a pipeline that begins with a **Word2Vec** NLP model to process textual input. This output is then passed through a **conditional GAN** trained to generate piano compositions, using the MAESTRO dataset as the foundation for learning musical patterns.



Word2Vec

- en_core_web_lg (spacy)



Conditional GAN - MAESTRO

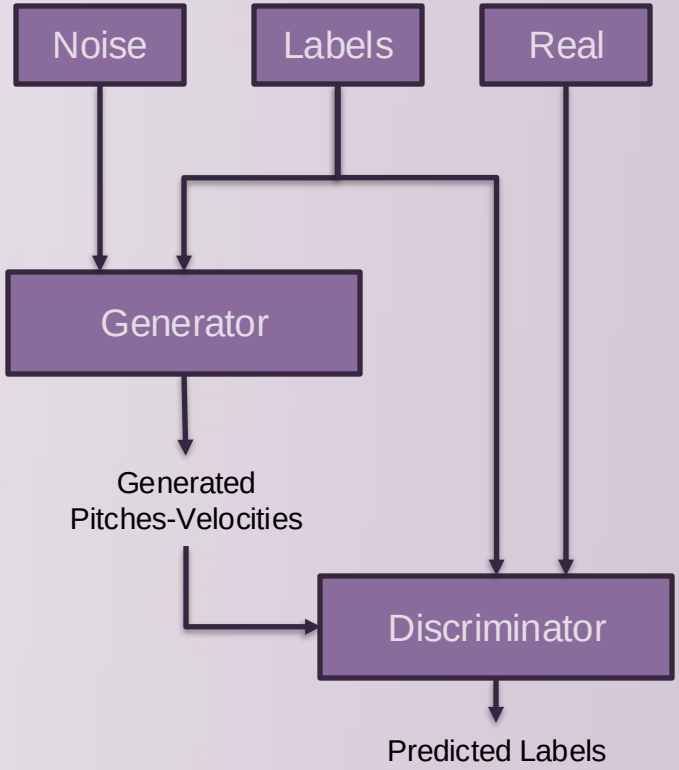
To train the conditional GAN, we decided to work with the MAESTRO dataset, a large collection of MIDI and audio recordings of **classical piano performances**.

This dataset provides rich temporal and dynamic information, making it ideal for training models to generate realistic and expressive music.

Conditional GAN - STRUCTURE

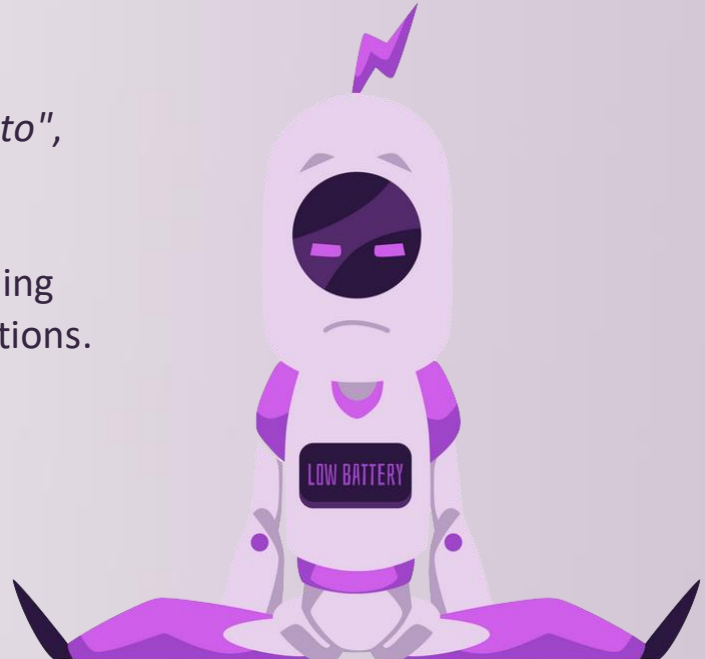
To generate music, we provided our model with sequences containing the following elements:

- **Pitches**: the specific notes in a musical scale, defining the melody.
- **Velocity**: the intensity or force of a note, influencing its loudness and expression.
- **Key**: indicates whether the composition is in a Major or Minor scale, defining its mood and tonality.



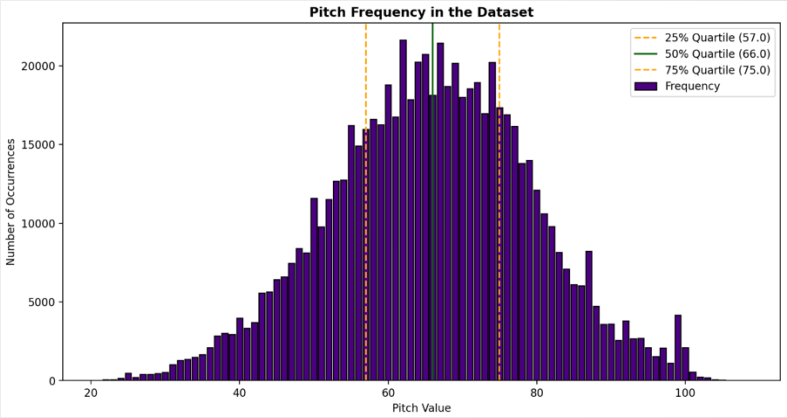
Word2Vec - CHALLENGES

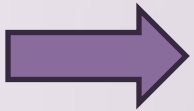
- **Ambiguity in musical terms:** the "moderate" tempo was not being predicted unless it was explicitly mentioned in the textual input.
- **Influence of common words:** frequent words like *"and"*, *"to"*, etc., can affect the quality of the learned representations.
- **Handling negations:** Word2Vec struggles with understanding negations, which can alter the meaning of musical descriptions.

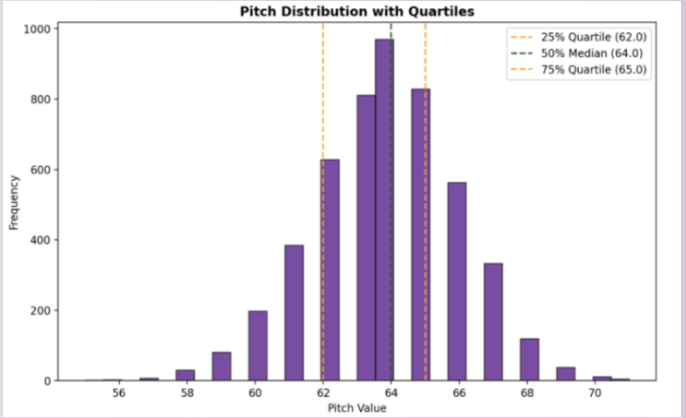


cGAN – RESULTS and CHALLENGES

What was our model able to do?

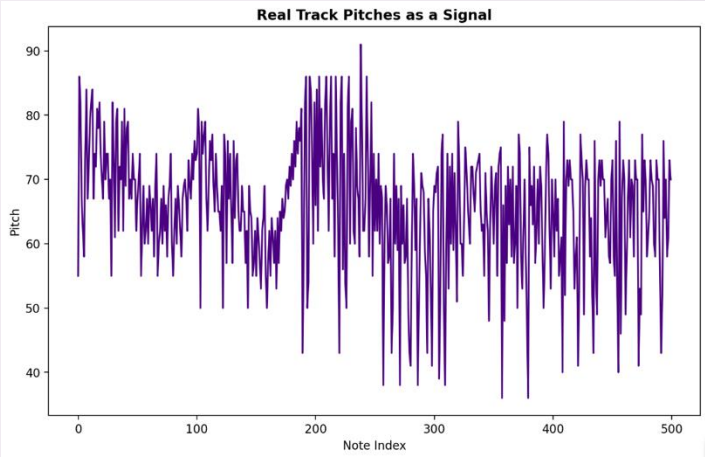




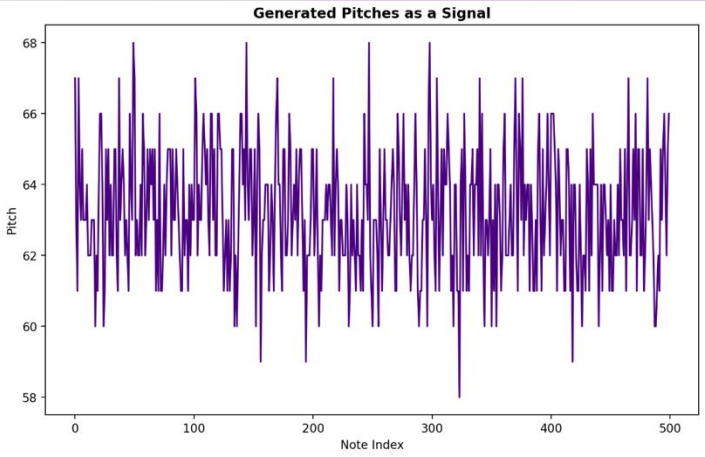


cGAN – RESULTS and CHALLENGES

What weren't we able to accomplish?



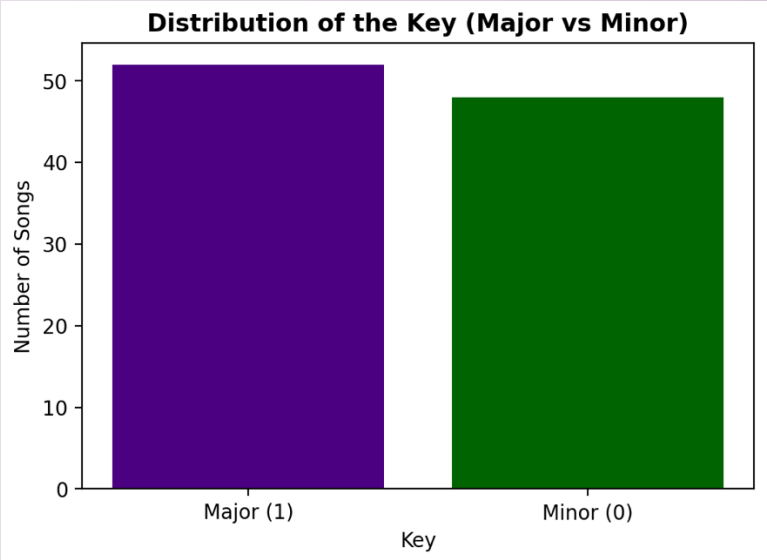
≠



cGAN – RESULTS and CHALLENGES

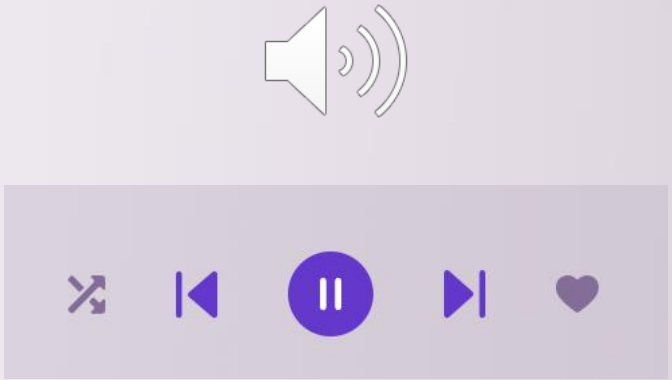
Unfortunatly…There is more

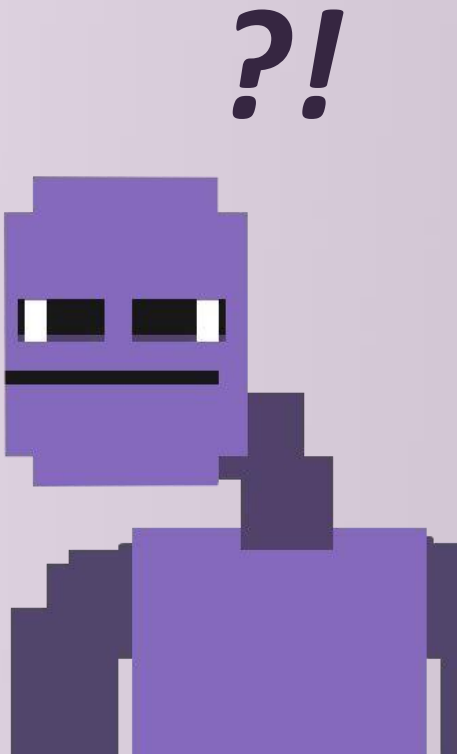
It is worth noting that the experiment involving key conditioning also did not yield successful results.



cGAN – RESULTS and CHALLENGES

Just to give you a concrete example:





Thank you for your Attention!

Any Question?