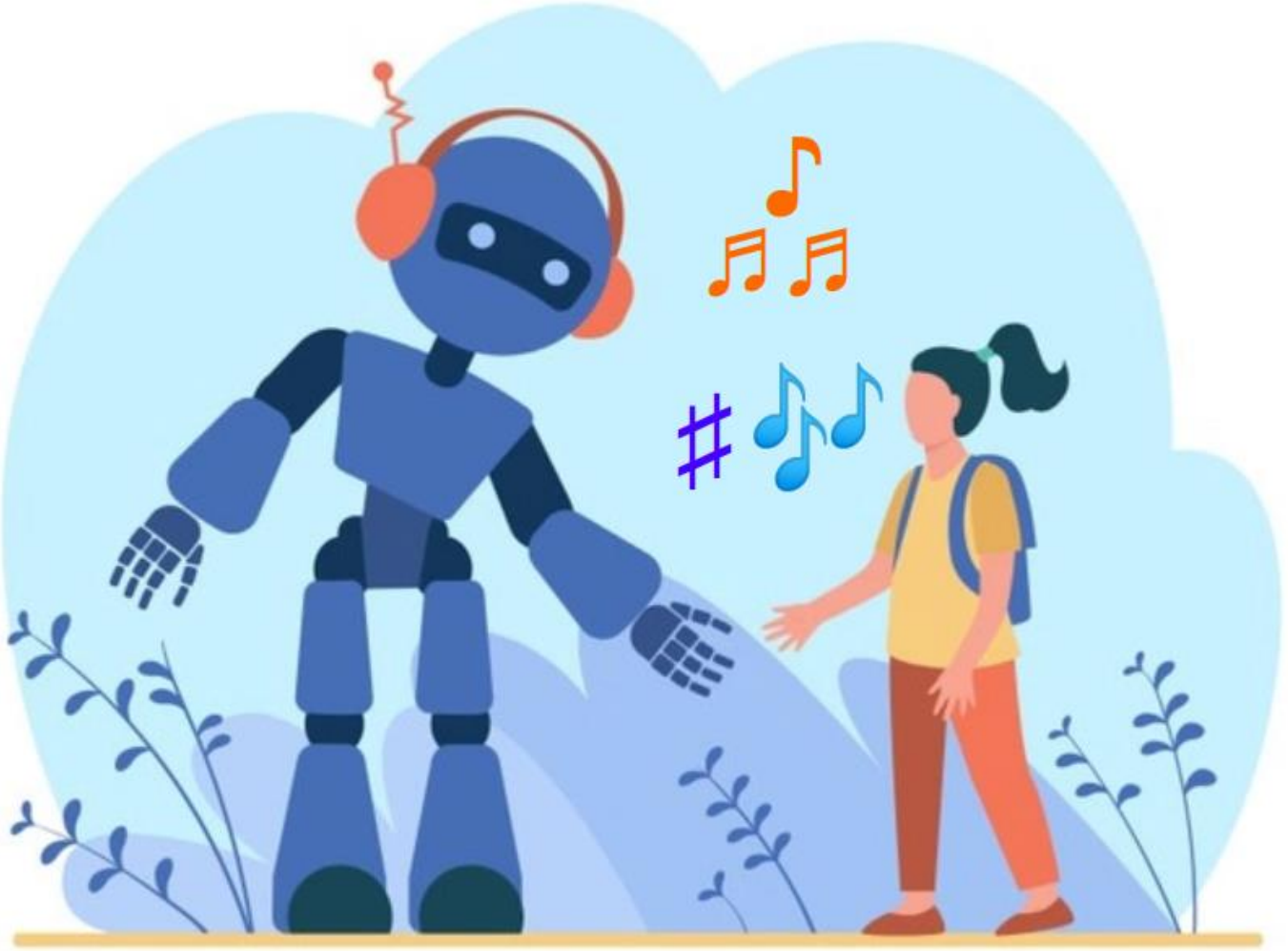
Let there be music: when AI learns to compose

Anna Aljanaki, 4. November 2021

Intelligent agents

- Alexa from Amazon Echo
- Siri (Apple)
- Google Assistant
- Alice (Yandex)
- Cortana (Microsoft)



Counterpoint

- A technique of writing music that demands creating several independent melodies (voices)
- The voices are simultaneous
- Was mostly used in European music in Renaissance and Baroque period

This presentation is based on:

- Anna Aljanaki, Stefano Kalonaris, Gianluca Micchi, Eric P. Nichols, **MCMA: A Symbolic Multitrack Contrapuntal Music Archive**, Empirical Musicology Review, 2021
- Stefano Kalonaris, Thomas McLachlan, Anna Aljanaki, **Computational Linguistics Metrics for the Evaluation of Two-Part Counterpoint Generated with Neural Machine Translation**, Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA), 2020
- Eric P. Nichols, Stefano Kalonaris, Gianluca Micchi, and Anna Aljanaki, 2021. **Modeling Baroque Two-Part Counterpoint with Neural Machine Translation**. In Proceedings of the International Computer Music Conference (ICMC). Santiago, Chile

Talking and singing simultaneously

- Simultaneous speech

Talking and singing simultaneously

- Simultaneous speech
- Simultaneous melodies

Talking and singing simultaneously

- Simultaneous speech
- Simultaneous melodies

Counterpoint

- Species counterpoint (strict rules)
- Free counterpoint
- Well separating the voices:
  - Prefer contrary motion, not parallel motion
  - Avoiding parallel fifths and octaves
- There are usually between 2 and 5 voices in counterpoint

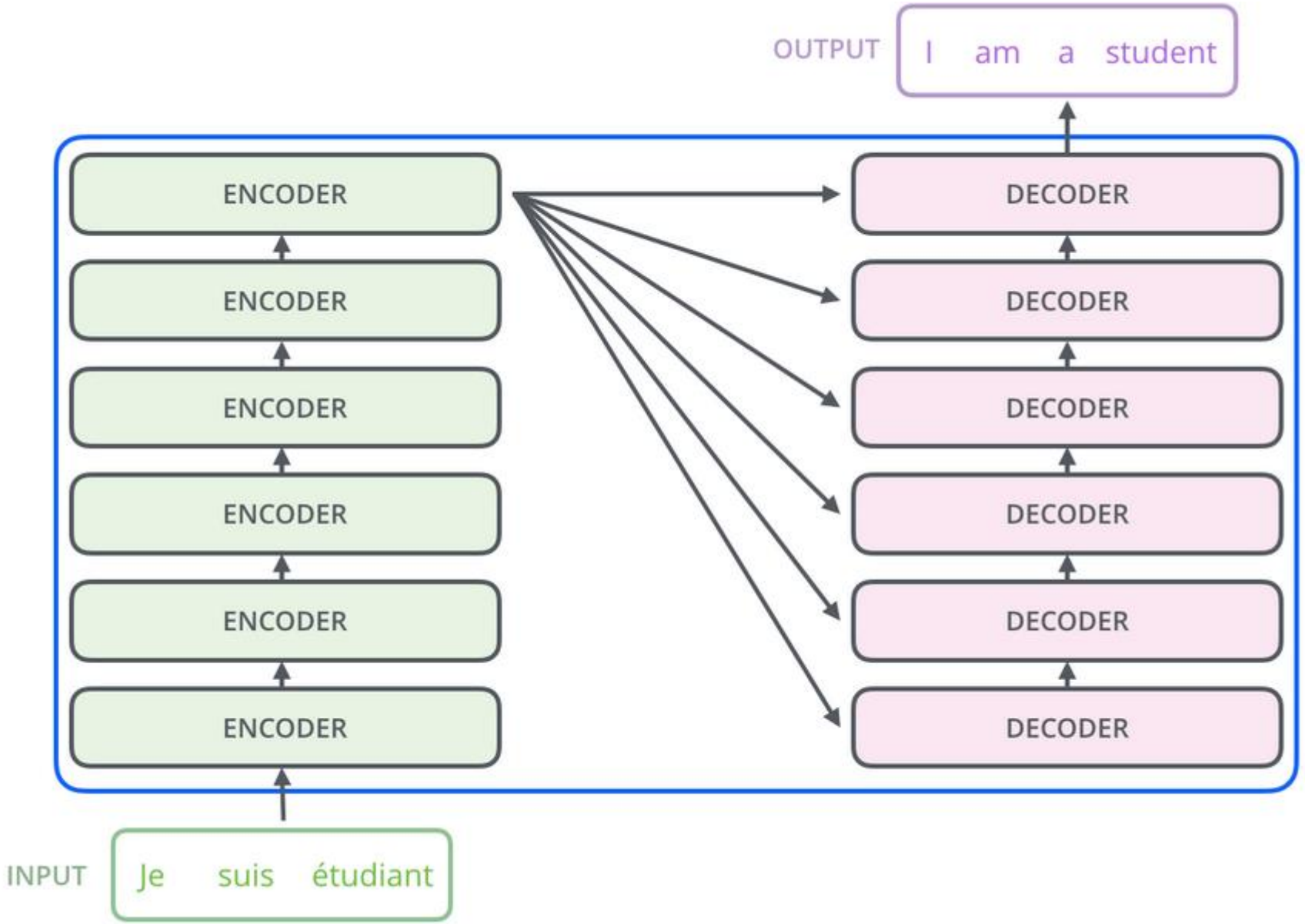
Short history of computer generated music

- RNN generation
  - FolkRNN
  - PerformanceRNN (2017)

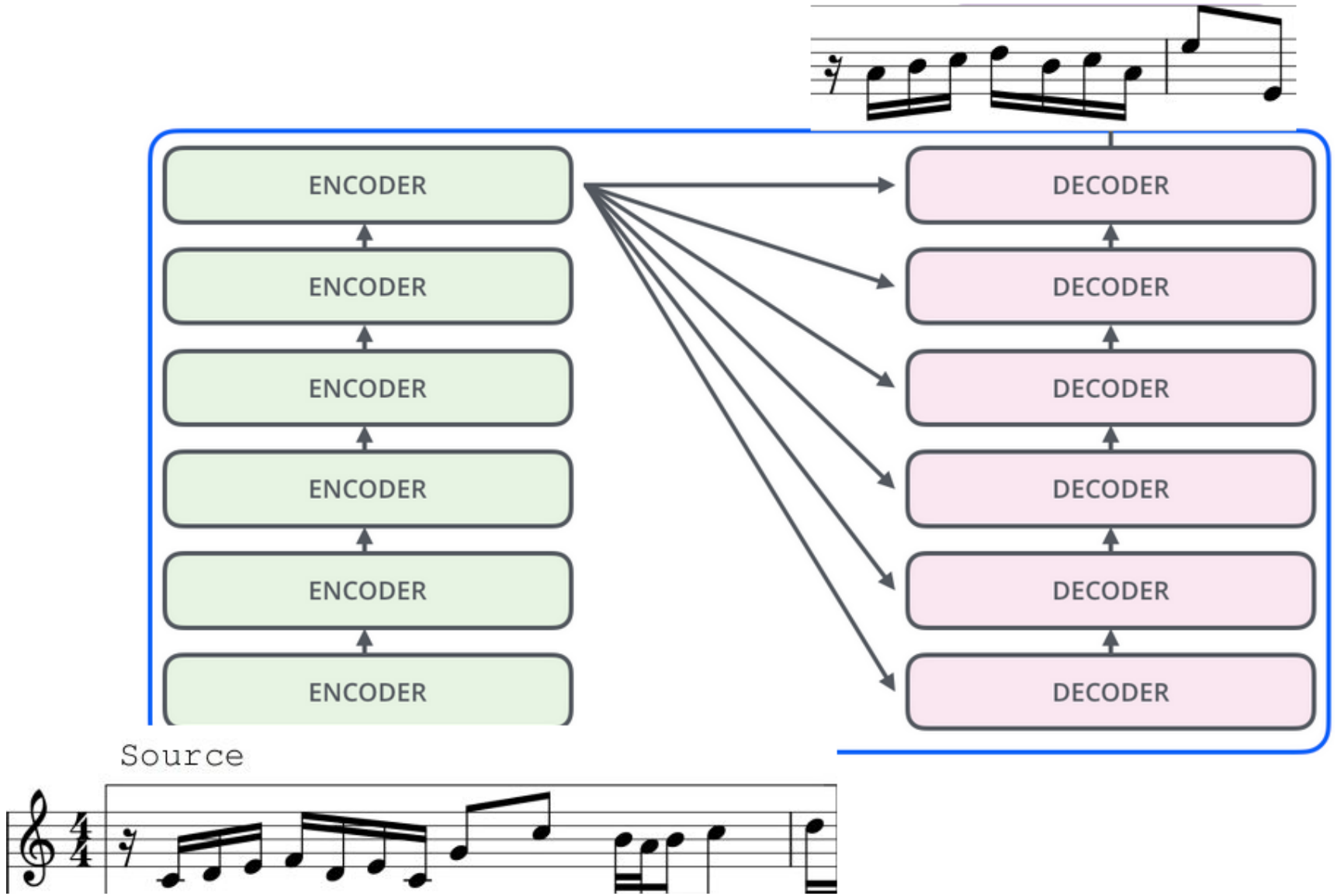
- WaveNET (digital signal sample level, 2016)

- **Transformers**
  - MuseNET (72 layer GPT-2 based net with 24 attention heads)

Transformer



Transformer



Multitrack Contrapuntal Music Archive

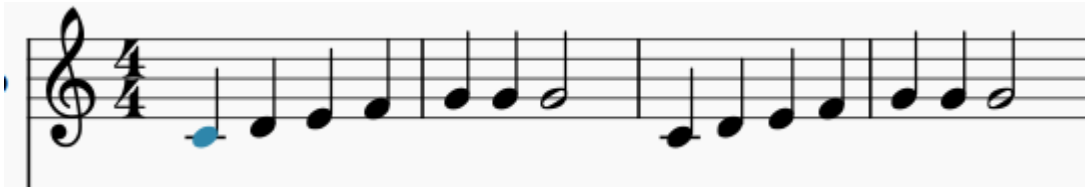
- mcma.readthedocs.io
- Separated each voice into separate track
- Each part is monophonic as much as possible
- Removed common MIDI problems:
  - Timing inaccuracies during performance
  - Embellishments added by performers

Multitrack Contrapuntal Music Archive

Composer	Number of Pieces	Publication Years	Total duration (crotchets, hours)
Albinoni, Tomaso	72	1708 – 1712	9,406.75 ♩ ≈ 1h30’
Bach, Johann Sebastian	178	1722 – 1750	33,171.25 ♩ ≈ 5h10’
Becker, Dietrich	45	1674	7,505.00 ♩ ≈ 1h10’
Buxtehude, Dietrich	88	1694 – 1696	11,019.00 ♩ ≈ 1h40’
Lully, Jean-Baptiste	92	1682	15,597.00 ♩ ≈ 2h20’

Representation

- C4:1.0 D4:1.0 E4:1.0 F4:1.0 G4:1.0 G4:1.0 G4:2.0



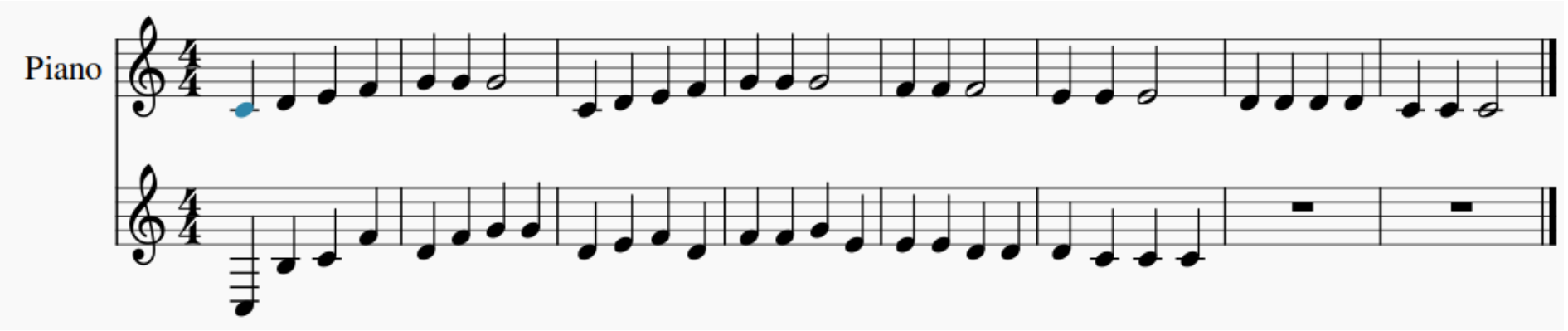
- Or with beat token (relative positional encoding)
- b1 C4:1.0 b2 D4:1.0 b3 E4:1.0 b4 F4:1.0 b5 G4:1.0 b6 G4:1.0 b7 G4:2.0

Model

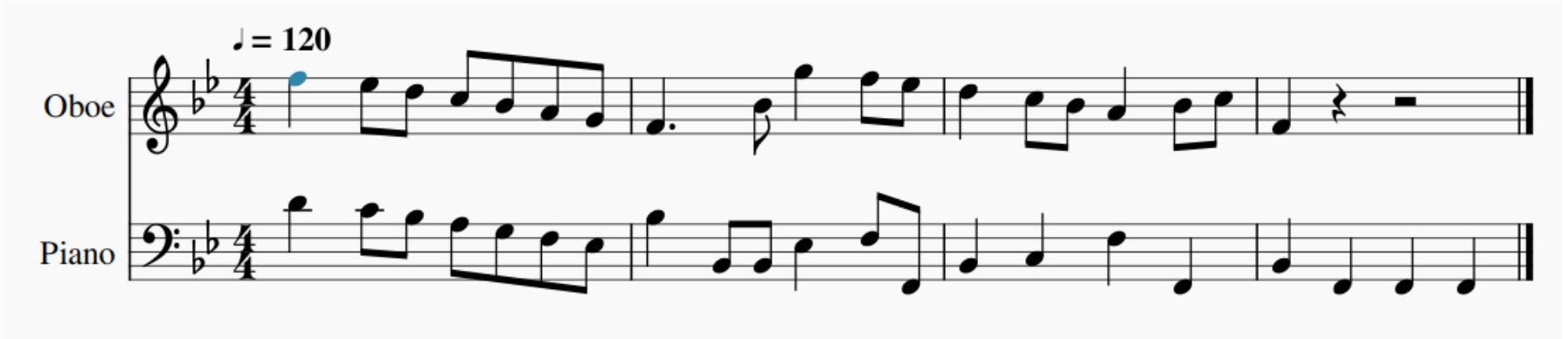
- A Transformer model with 6 layers and 8 attention heads
- Query, key and value matrices
- Each row of the query matrix represents a token from the target music phrase, while each key-value pair is taken from the source music phrase

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

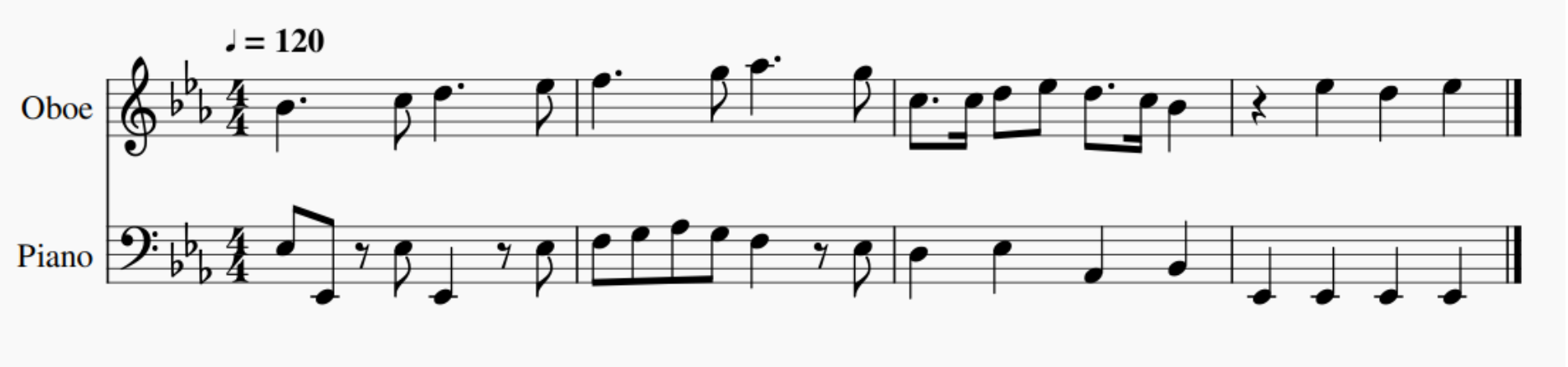
and the model answers



with actual baroque music



with actual baroque music



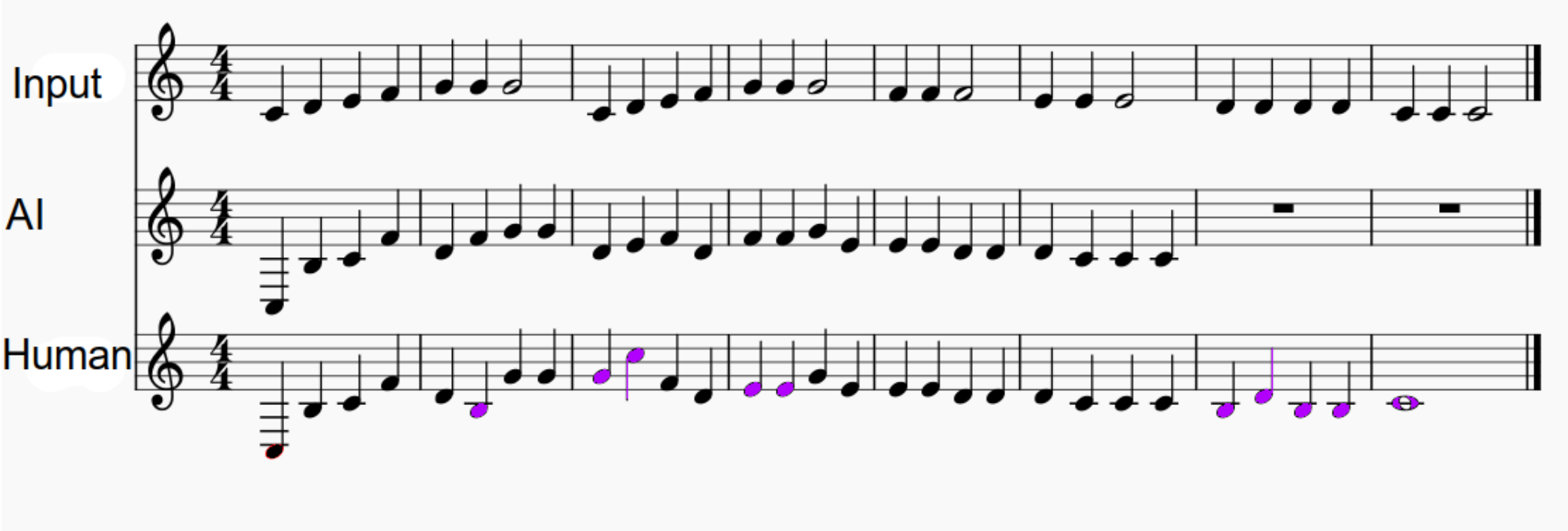
Evaluation

- Bilingual Evaluation Understudy (BLEU) (Papineni et al., 2002)
- Word Error Rate (WER)
- Recall-Oriented Understudy for Gisting Evaluation (ROUGE) (Lin, 2004)
- Perplexity (Brown et al., 1992)

Music evaluation

- Subjective evaluation
- Musical turing tests
- Human edit rate

Human edit rate



Transformers for business

- Machine translation
- Text generation
- Text summarization
- Chatbots