

# Generowanie muzyki przy pomocy dużych modeli językowych

Filip Ręka

Wydział Informatyki





# Modele LLM w kontekście tekstu i muzyki

- Tekst i muzyka mają strukturalne podobieństwa, opierające się na wzorcach i sekwencjach do przekazywania emocji. Oba mają hierarchiczną organizację: zdania tworzą akapity, a nuty i akordy tworzą melodie i harmonie.
- Tekst przestrzega zasad składni i gramatyki, a muzyka zasad rytmu i harmonii.
- Muzyka ma aspekt pionowy (akordy), którego nie ma w tekście
- NLP może być użyteczne w analizie muzyki.
- LLM zrewolucjonizowały przetwarzanie języka naturalnego i mogą być dostosowane do generowania muzyki

# Sposoby reprezentacji muzyki

- Partytura
- WAV
- MusicXML
- MIDI
- ABC

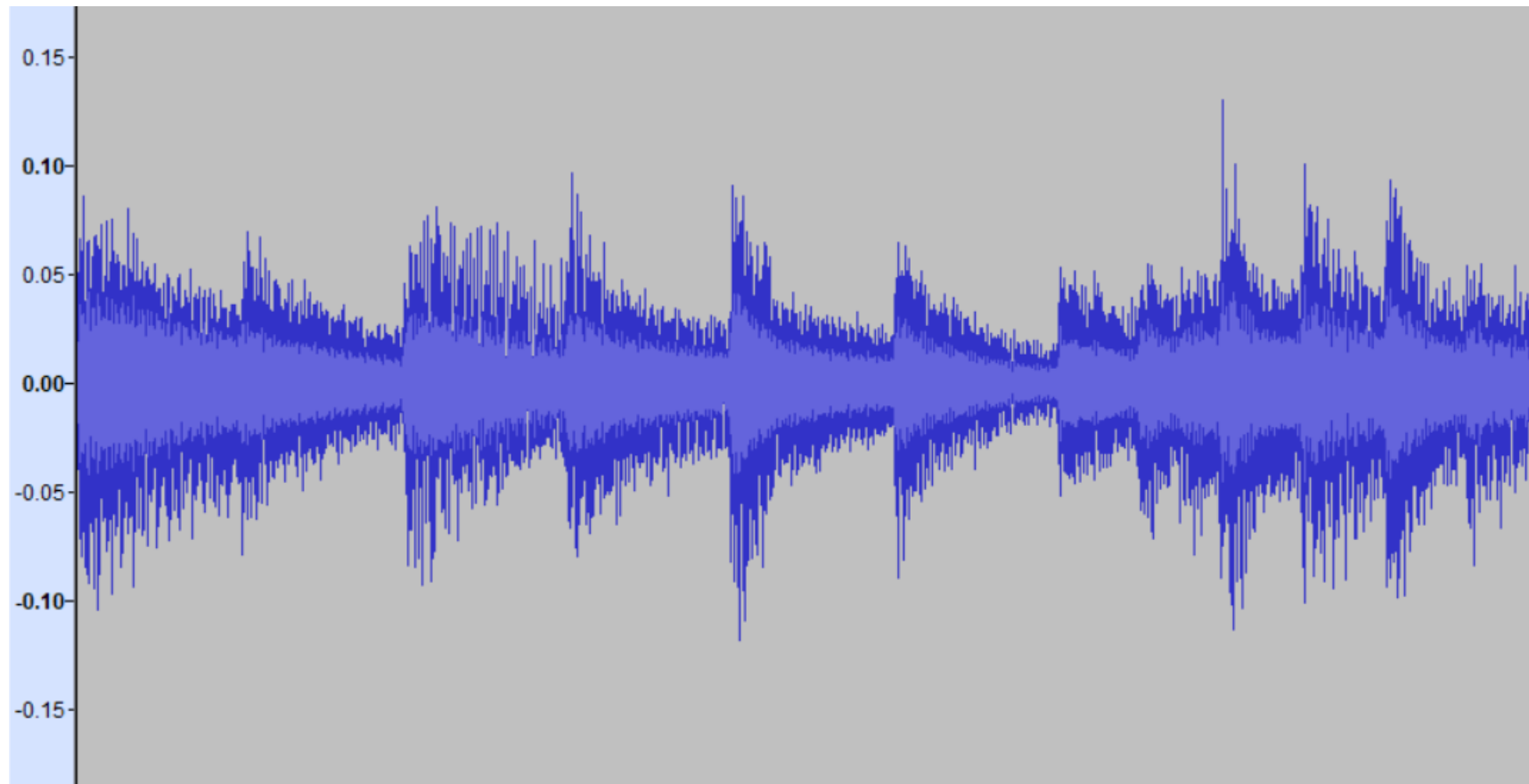


# Partytura

$\text{♩} = 120$

The musical score is written for four staves, organized into two systems of two staves each. The top system consists of two treble clef staves, and the bottom system consists of two bass clef staves. The key signature is one flat (B-flat), and the time signature is 4/4. The tempo is indicated as 120 beats per minute. The notation includes various musical symbols such as notes, rests, accidentals, and slurs, representing a piece of music.

# WAV



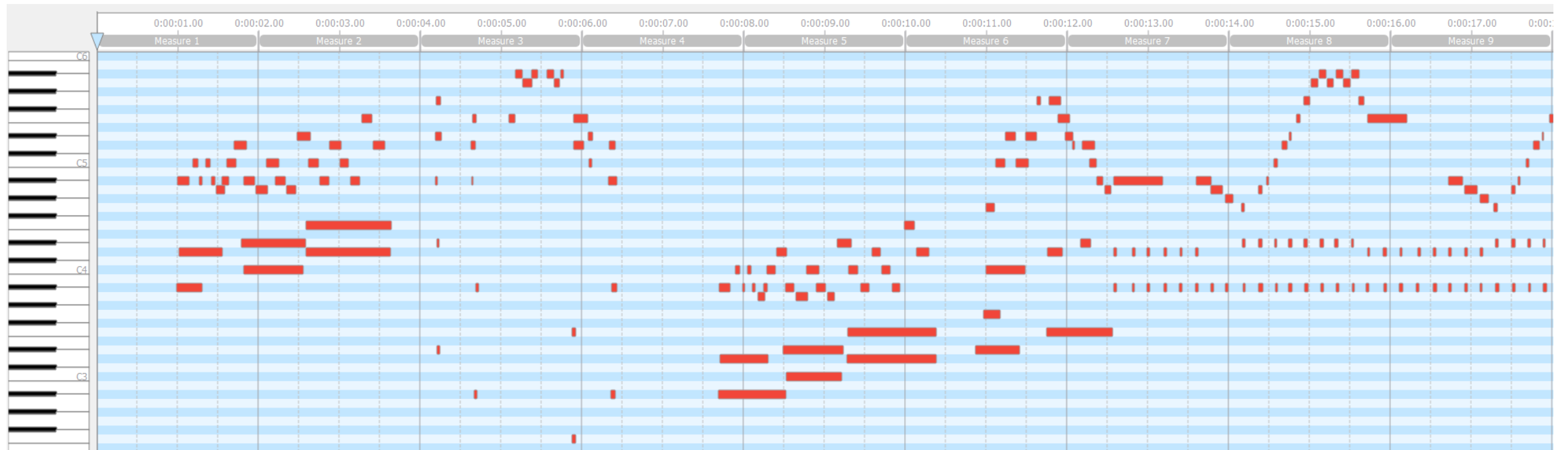


# MusicXML



```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE score-partwise PUBLIC
    "-//Recordare//DTD MusicXML 4.0 Partwise//EN"
    "http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise version="4.0">
  <part-list>
    <score-part id="P1">
      <part-name>Music</part-name>
    </score-part>
  </part-list>
  <part id="P1">
    <measure number="1">
      <attributes>
        <divisions>1</divisions>
        <key>
          <fifths>0</fifths>
        </key>
        <time>
          <beats>4</beats>
          <beat-type>4</beat-type>
        </time>
        <clef>
          <sign>G</sign>
          <line>2</line>
        </clef>
      </attributes>
      <note>
        <pitch>
          <step>C</step>
          <octave>4</octave>
        </pitch>
        <duration>4</duration>
        <type>whole</type>
      </note>
    </measure>
  </part>
</score-partwise>
```

# MIDI



# Tokenizacja plików MIDI



Dur. 4.0	Vel. 80	Pitch A3	Tempo 116	Pos. 0	Bar	Dur. 0.4	Vel. 72	Pitch G3	Tempo 116	Pos. 27	Rest 0.4	Rest 1.0	Dur. 1.0	Vel. 84	Pitch A3	Tempo 116	Pos. 7	Dur. 1.0	Vel. 80	Pitch D3	Tempo 116	Pos. 0	Bar
----------	---------	----------	-----------	--------	-----	----------	---------	----------	-----------	---------	----------	----------	----------	---------	----------	-----------	--------	----------	---------	----------	-----------	--------	-----



# ABC

X:1

Q:1/4=120

V:1

L:1/16

M:4/4

K:C clef=G2

D4F4G4A4|d4c6B2A2B2|c4B8A4|d12^c4|

V:2

L:1/16

M:4/4

K:C clef=G2

A,4D6E2F4|A8^G4A2^G2|A6^G^F^G4E4|A6G2B4A4|

V:3

F,4A,4^A,4D4|F4E7DC2D2|E2F2B,2E4D2^C4|D6E2F4E4|

V:4

L:1/16

M:4/4

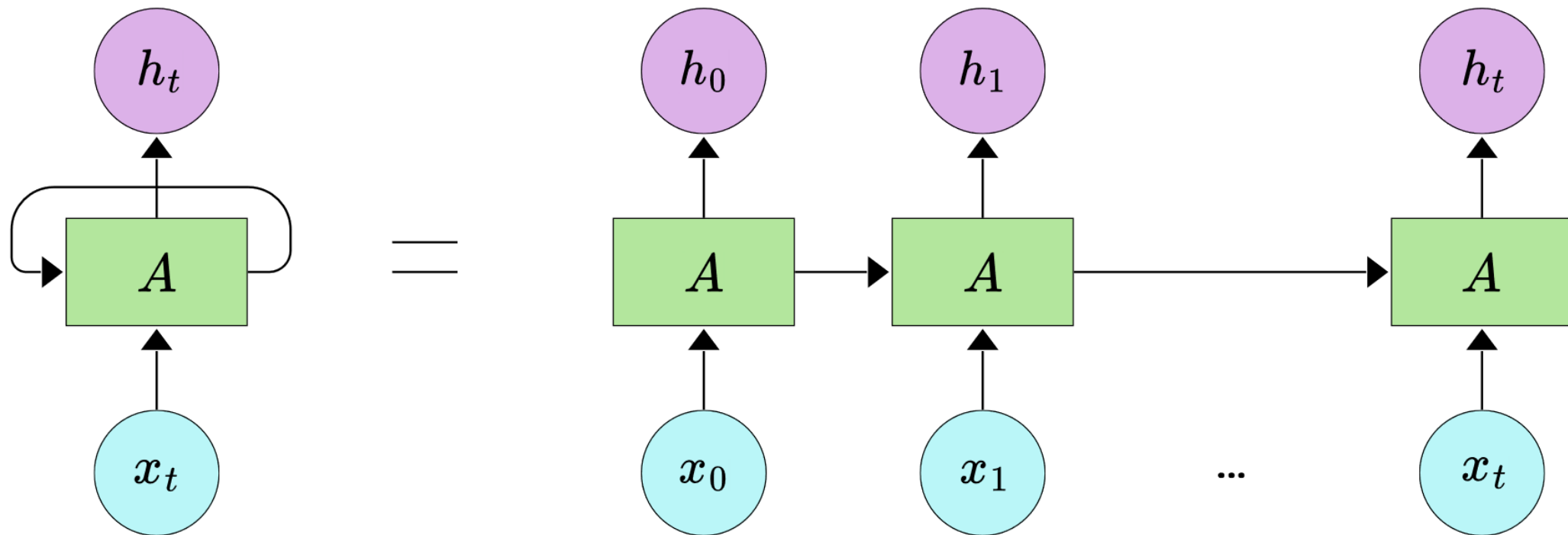
K:C clef=F4

D,8G,,4D,,4|D,4A,4E,4F,4|C,2D,2E,4E,,4A,,4|F,4^A,4A,2^G,2A,4|

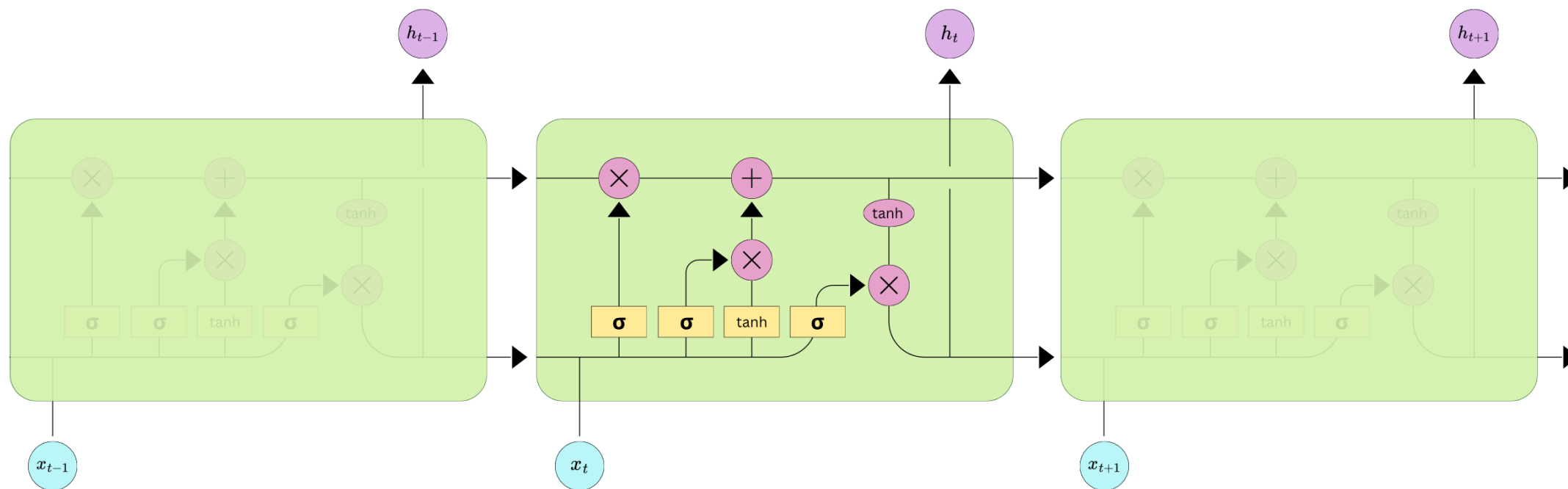
# Porównanie zapisów muzycznych

	rozmiar pliku na dysku	długość sekwencji
wav	3971116 B	992768
mp3	360951 B	992768
midi	1638 B	740
abc	953 B	562

# Sieć RNN



# Sieć LSTM



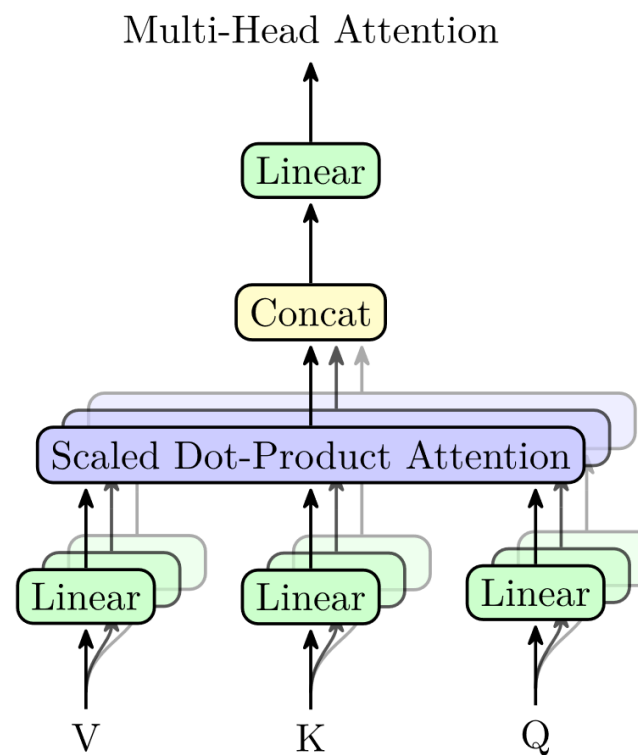
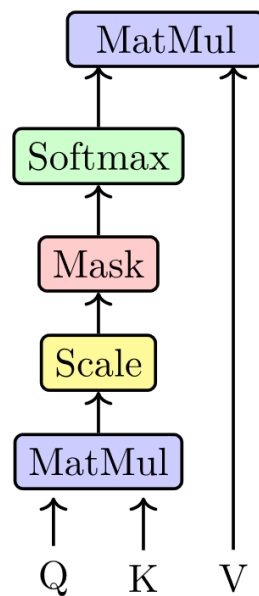
Dwa wektory wyjściowe:

- stan ukryty - *hidden state*
- stan komórki - *cell state*

Trzy „bramki”:

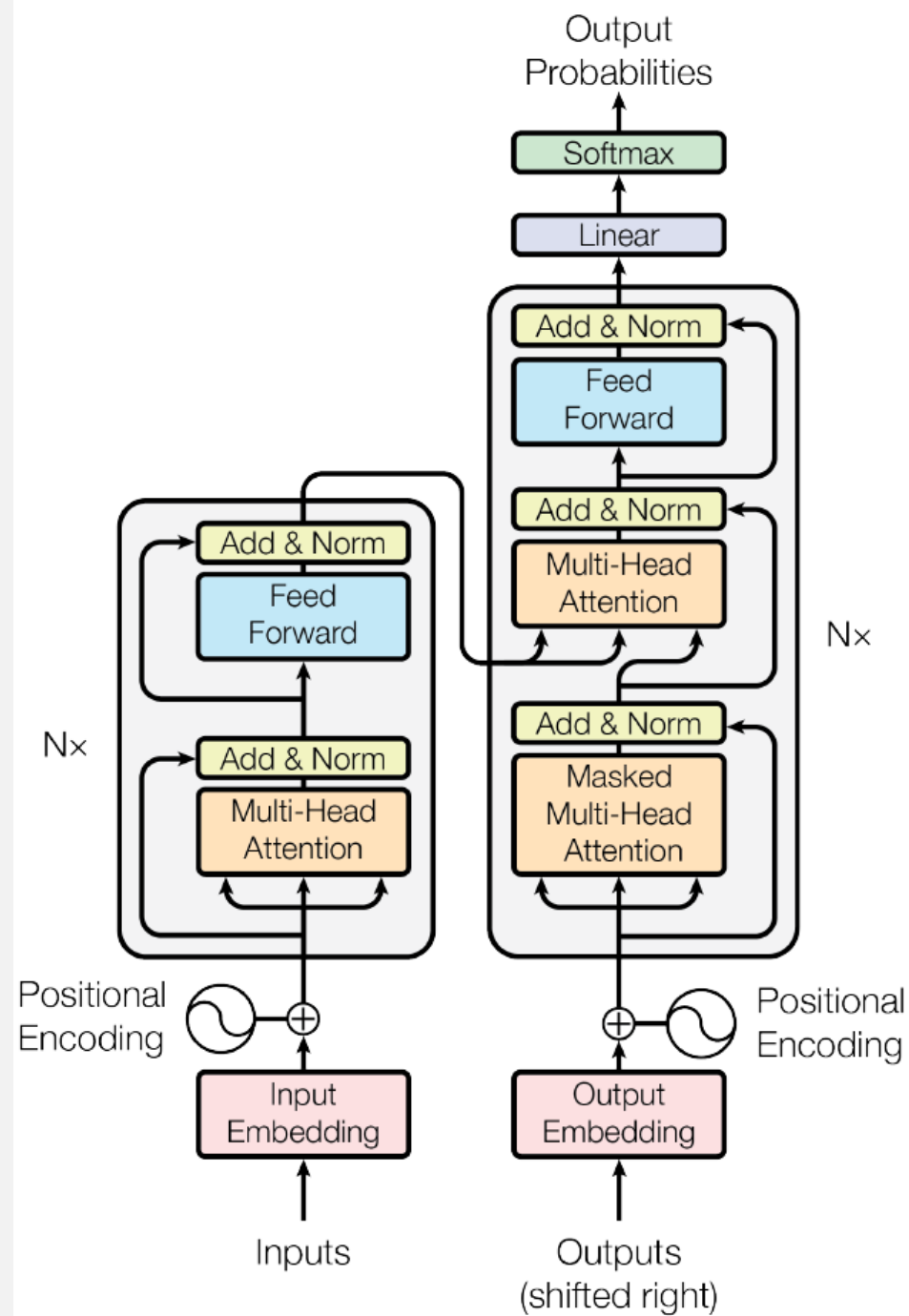
- wejścia
- zapomnienia
- wyjścia

# Attention



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

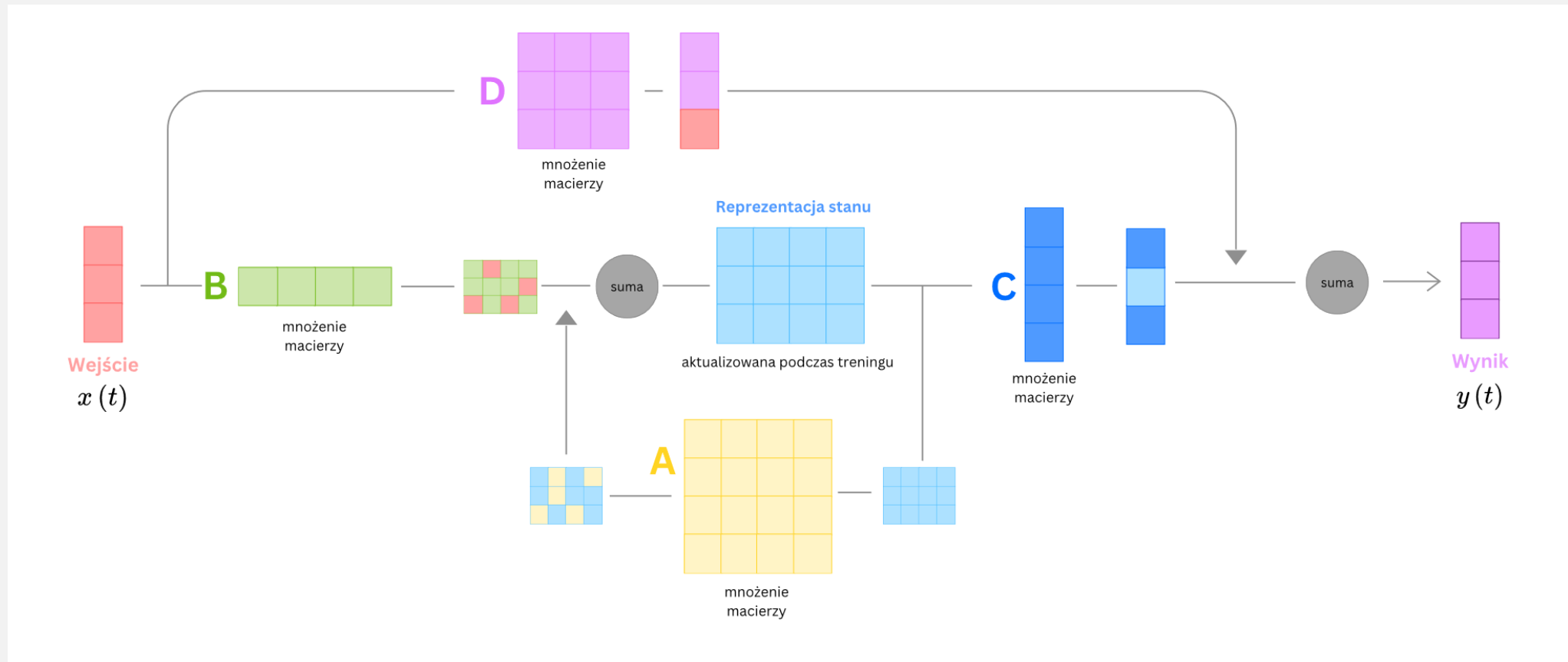
# Transformer



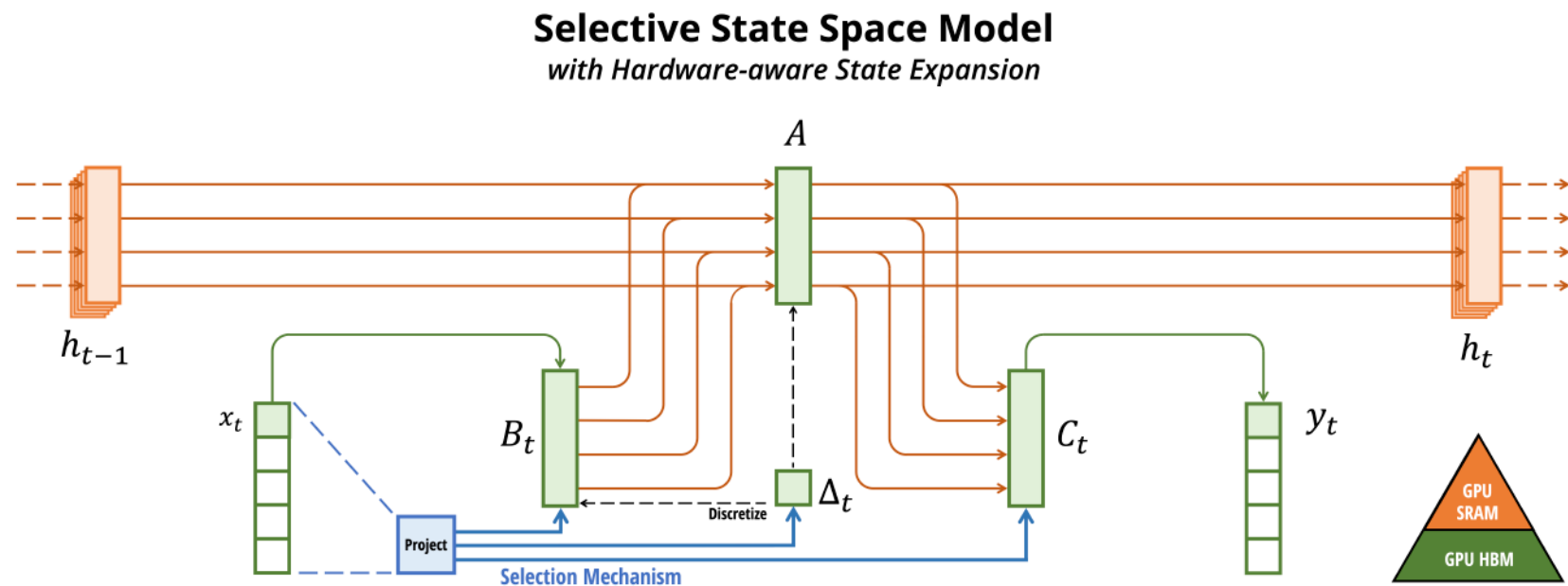
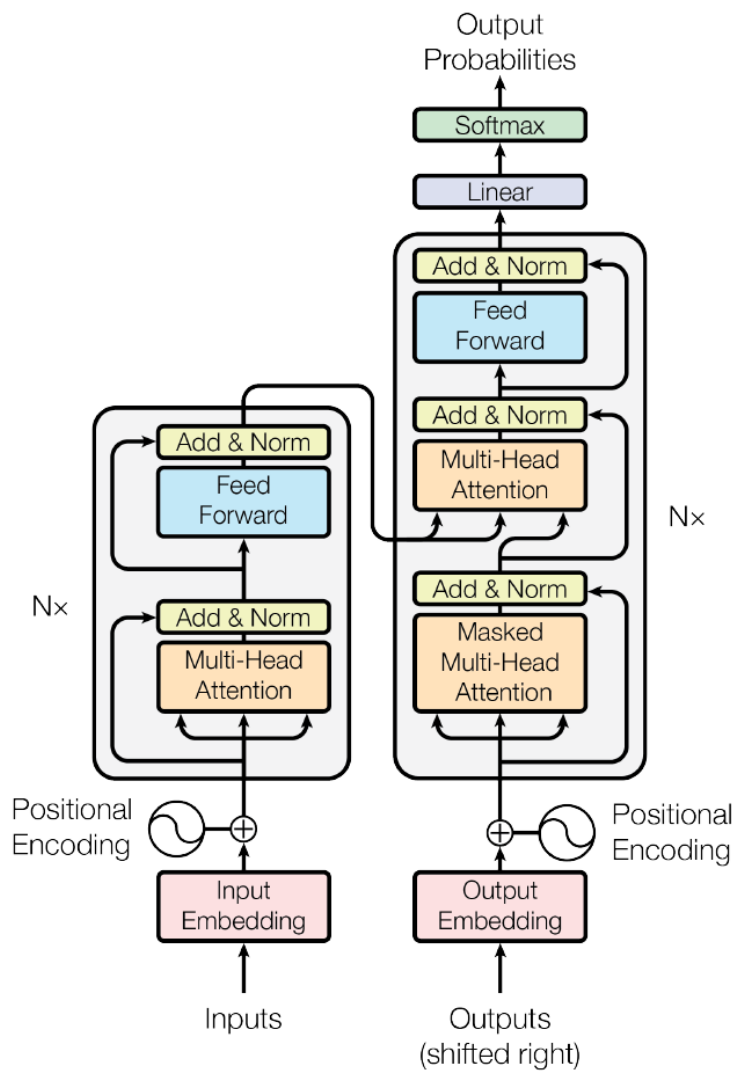
	Ala	ma	kota	i	psa		Ala	ma	kota	i	psa
Ala	0.13	0.18	0.16	0.15	0.18		Ala	0.13	$-\infty$	$-\infty$	$-\infty$
ma	0.68	0.02	0.08	0.14	0.02		ma	0.68	0.02	$-\infty$	$-\infty$
kota	0.06	0.25	0.14	0.11	0.23	→	kota	0.06	0.25	0.14	$-\infty$
i	0.21	0.14	0.16	0.17	0.14		i	0.21	0.14	0.16	0.17
psa	0.27	0.11	0.16	0.18	0.12		psa	0.27	0.11	0.16	0.18



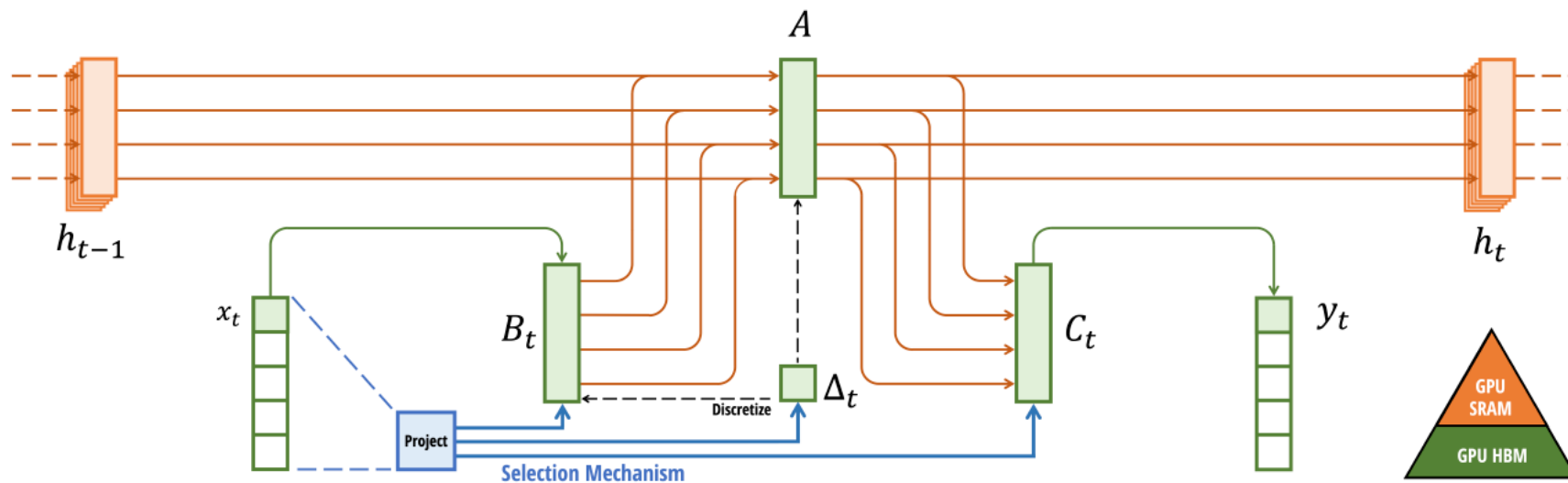
# Modele przestrzeni stanów



# Transformers vs Mamba



## Selective State Space Model with Hardware-aware State Expansion



## Mamba

Wprowadzone usprawnienia:

- Dyskretyzacja modelu (eksplorator rzędu zerowego)
- Reprezentacja konwolucyjna
- Macierz HIPPO
- Algorytm *selective scan*
- Algorytm *hardware-aware*

# Przeprowadzenie eksperymentów

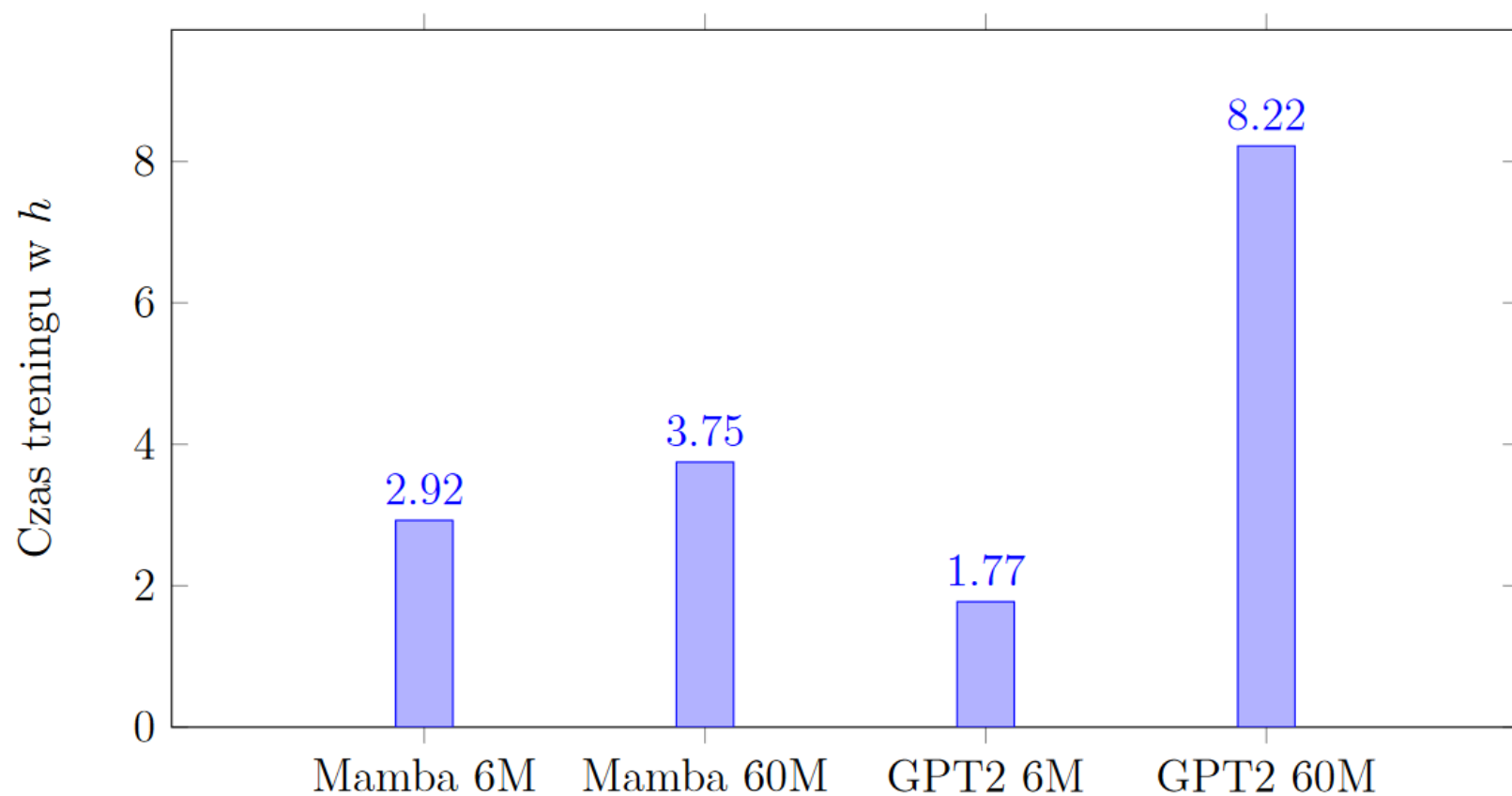
—

# Trenowane modele

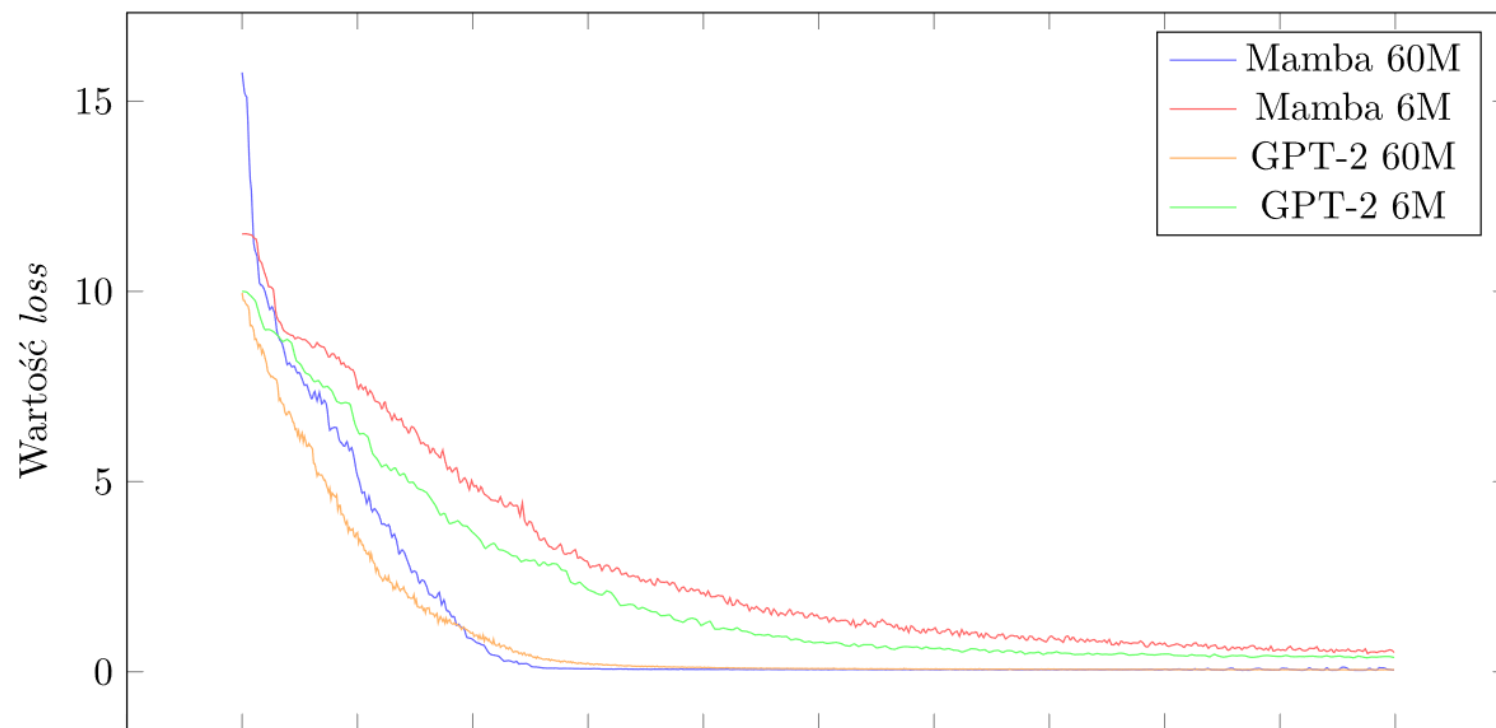


- GPT-2 6M
- GPT-2 60M
- Mamba 6M
- Mamba 60M

# Czas treningu 100k kroków

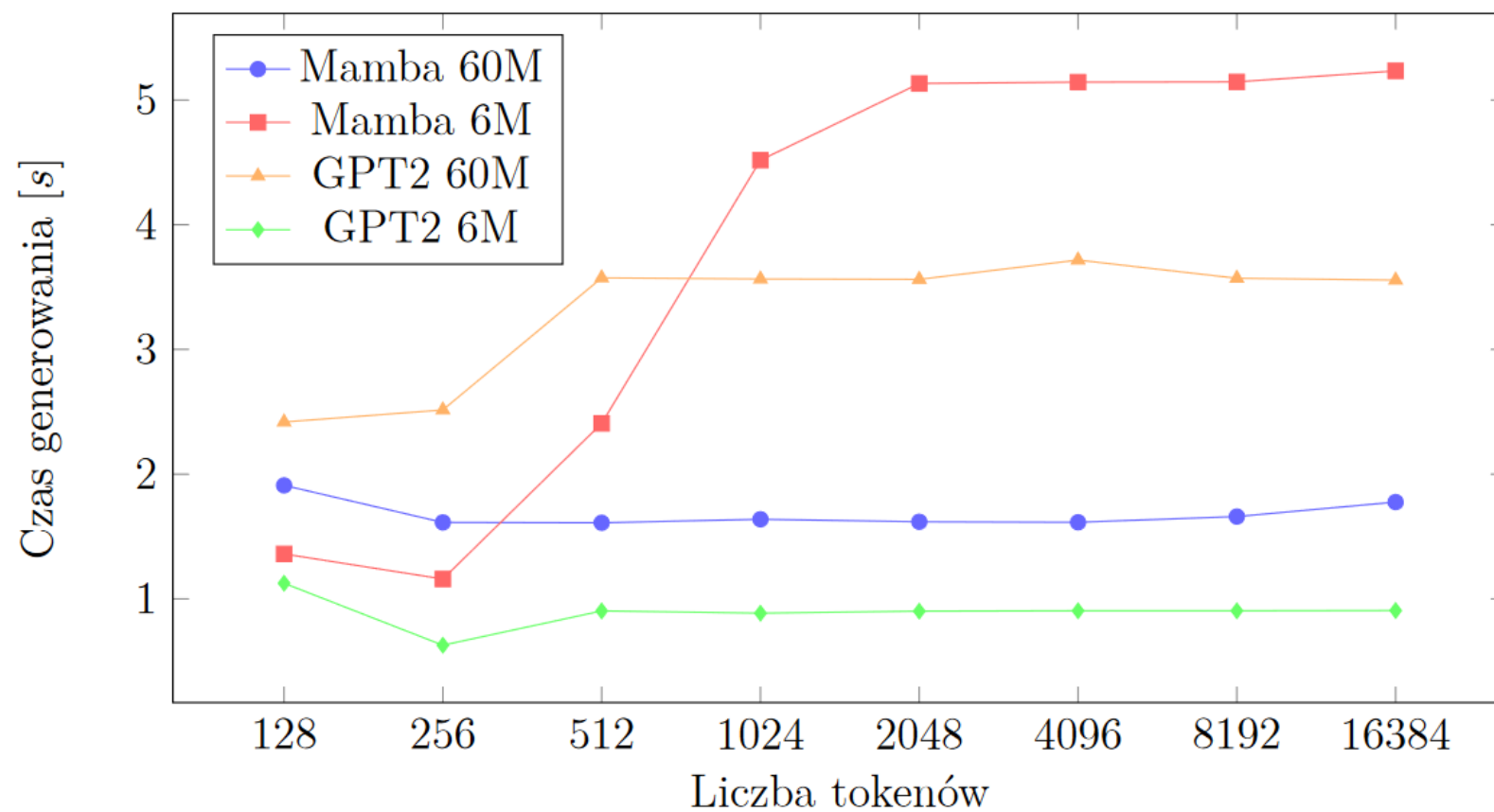


# Wartość funkcji straty podczas treningu





# Czas inferencji



# Przykład generacji modelu

X:1

S:2

B:9

E:4

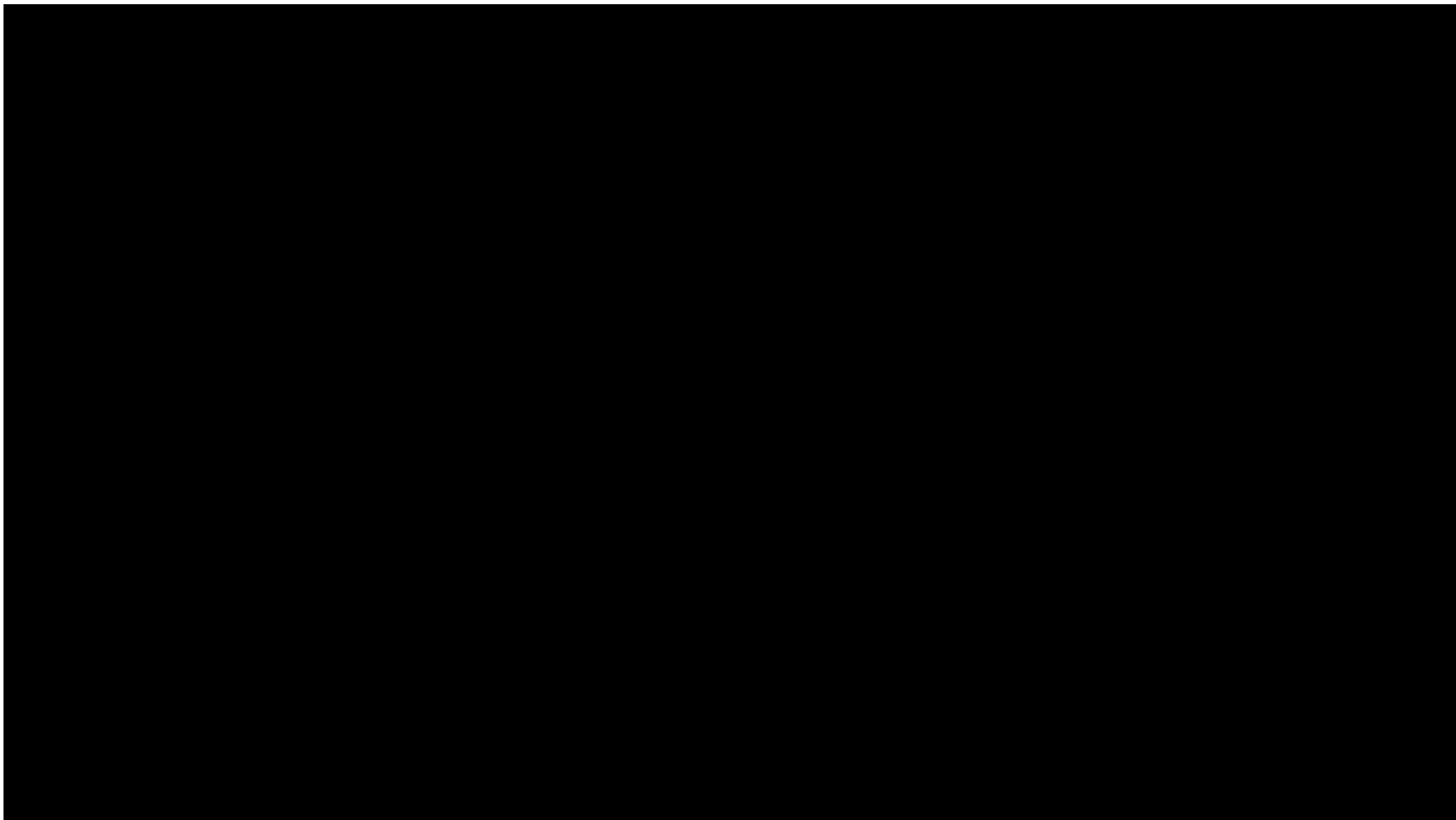
L:1/8

M:3/4

K:C

```
D/2d/2d Ad fa | a/g/e/f/ g/f/e/d/ ^cA | d2 dA df | ag/f/ e^c A2 | D2 d^c
de | f2 ef/g/ a/g/f/e/ | d2 df e^c |1 d2 D4 :|2 d2 D2 fg |: a2 g2 f2 | ed
^cd e/f/g | a2 g2 f2 | ed ^cd e/f/g | a2 g2 f2 | ed ^cd ef/g/ | a2 g2 fe
|1 d2 de fg :|2 d2 D4 ||
```





Weryfikacja wyników innymi modelami LLM

# ChatMusician

---

ChatMusician to otwarty model językowy, który integruje zdolności muzyczne. Model został zaprojektowany na bazie modelu LLaMA2 z 7 miliardów parametrów, jest trenowany i dostrajany z wykorzystaniem notacji muzycznej ABC, traktując muzykę jako drugi język. Dzięki temu ChatMusician jest w stanie rozumieć i generować muzykę za pomocą tekstowego tokenizatora, bez konieczności stosowania zewnętrznych struktur neuronowych czy tokenizatorów stworzonych specjalnie dla notacji muzycznej.

**PROMPT:** I have some music in ABC notation that I generated. Could you analyze it and tell me if it sounds natural? Specifically, I would like your feedback on the overall musicality, the sense of harmony, and whether the melody and chord progressions are coherent and pleasing. Here is the ABC notation of the music:

X:1

L:1/8

M:3/4

K:C

D/2d/2d Ad fa | a/g/e/f/ g/f/e/d/ ^cA | d2 dA df | ag/f/ e^c A2 |  
D2 d^c de | f2 ef/g/ a/g/f/e/ | d2 df e^c |1 d2 D4 :|2 d2 D2 fg |:  
a2 g2 f2 | ed ^cd e/f/g | a2 g2 f2 | ed ^cd e/f g | a2 g2 f2 |  
ed ^cd ef/g/ | a2 g2 fe |1 d2 de fg :|2 d2 D4 ||



**ANSWER:** Overall Musicality: 1. The piece has a pleasant and flowing rhythm with a steady tempo. It features a mix of different instruments and voices, creating an interesting texture. The melodies are catchy and memorable, making the piece enjoyable to listen to.

2. Sense of Harmony: The harmonies in this piece are well-crafted and create a rich and full sound. The use of various chords and progressions adds depth and complexity to the music. The harmonic choices are appropriate for the style and genre of the piece.

3. Melody and Chord Progressions: The melody and chord progressions in this piece are coherent and pleasing. They follow a logical progression and fit together well. The melodic lines are well-written and expressive, adding to the emotional impact of the piece. The chord progressions provide a solid foundation for the melodies and add interest and variety to the composition.

In conclusion, this piece of music has a strong sense of musicality, harmony, and coherence. The melodies and chord progressions are well-constructed and contribute to the overall appeal of the piece. It is a successful example of its genre and style.

**ANSWER:** Overall Musicality: The piece has a pleasant and flowing quality to it, with a mix of different rhythms and tempos. The use of dynamics and articulations adds depth and interest to the composition.

2. Sense of Harmony: The harmonies used in this piece are mostly consonant and create a sense of stability and balance. There are occasional dissonances that add tension and color to the music.

3. Melody and Chord Progressions: The melodies in this piece are well-crafted and follow a logical progression. They are easy to sing or play and have a clear sense of direction. The chord progressions are also coherent and support the melodies effectively.

In conclusion, this piece has a good overall musicality, sense of harmony, and coherence between its melody and chord progressions. It is an enjoyable and engaging composition.



# GPT-4

To evaluate the musicality and harmony of your ABC notation, I will provide an analysis based on the key, melody, rhythm, and chord progressions. Here is a detailed breakdown of your piece:

## ### Key and Meter

- \*\*Key:\*\* C major
- \*\*Meter:\*\* 3/4 (Waltz time)

## ### Melody and Harmony

### 1. \*\*Opening Phrase:\*\*

- `D/2d/2d Ad fa | a/g/e/f/ g/f/e/d/ ^cA`
  - This phrase introduces a mix of stepwise motion and leaps, giving a dynamic start. The alternation between the upper and lower notes creates a call-and-response effect. The use of the raised 4th degree (^c) adds a touch of chromaticism.

### 2. \*\*Middle Phrase:\*\*

- `d2 dA df | ag/f/ e^c A2`
  - The repetition of 'd' with variations in intervals suggests a thematic consistency. The leap to 'f' and back to 'd' maintains interest. The movement towards 'A' anchors the phrase back to the tonic's dominant, providing harmonic stability.

### 3. \*\*Transition:\*\*

- `D2 d^c de | f2 ef/g/ a/g/f/e/`
  - The use of the D minor chord (D-F-A) transitioning to E minor (E-G-B) and then back to fragments of D minor adds tension and resolution.

### 4. \*\*Closing Phrase:\*\*

- `d2 df e^c |1 d2 D4 :|2 d2 D2 fg`
  - This phrase returns to the tonic D and introduces a cadential figure (D-F-E-^C). The first ending resolves to the tonic D major chord, while the second ending moves into a continuation with 'fg', creating a sense of forward motion.

## ### Second Part

### 1. \*\*Opening Phrase:\*\*

- `a2 g2 f2 | ed ^cd e/f/g`
  - This section starts with a descending sequence, providing contrast to the first part. The melodic minor descent with a chromatic touch (e-^c-d) maintains interest.

### 2. \*\*Middle Phrase:\*\*

- `a2 g2 f2 | ed ^cd e/f g`
  - The repetition reinforces the theme while subtle variations in articulation (slurs) add expressiveness.

### 3. \*\*Transition and Closing:\*\*

- `a2 g2 f2 | ed ^cd ef/g/ | a2 g2 fe`
  - The alternation between stepwise motion and leaps ensures coherence and variety. The repeated 'a-g-f' figure creates familiarity, while 'e-^c-d ef/g/' keeps it fresh.

### 4. \*\*Final Cadence:\*\*

- `1 d2 de fg :|2 d2 D4 ||`
  - The final cadence resolves nicely with the repeating 'de fg' figure leading to the tonic 'D' again, ensuring a satisfying conclusion.

## ### Overall Musicality

- \*\*Melodic Coherence:\*\* The melody has a clear direction with a good mix of stepwise motion and leaps, maintaining interest.
- \*\*Harmony:\*\* The harmony is mostly implied through the melodic line with occasional chromatic touches adding tension and resolution. The frequent returns to tonic and dominant notes (C and G) provide a sense of harmonic stability.
- \*\*Rhythmic Flow:\*\* The piece maintains a consistent waltz rhythm (3/4 time), which gives it a dance-like quality, adding to its musicality.

## ### Conclusion

The piece sounds natural and coherent, with a pleasing melody and harmonic progression. The use of chromatic notes and variations in articulation adds interest and depth. It effectively balances repetition with variation, ensuring both familiarity and novelty for the listener.

Based on the ABC notation provided, here is an analysis of the music in terms of musicality, harmony, and melody coherence:

### ### Overall Musicality

The notation seems to follow a structure that could be pleasing. The use of different note values and ornamentations like the grace note `{g}` add interest to the melody. The notation implies a mix of staccato and legato sections, which can contribute to a dynamic musical flow.

### ### Harmony

The harmony is somewhat implied rather than explicitly stated. The melody suggests movement between different tonal centers, but the use of accidentals (like `^g`, `\_A`, `^c`, etc.) makes the harmony more complex. It moves away from a strictly diatonic C major scale, introducing chromaticism which can add color but might sound unusual without proper context or accompanying chords.

### ### Melody Coherence

The melody uses a variety of intervals and note values, which gives it a sense of motion. However, the frequent use of accidentals (`^g`, `\_A`, `^c`, `\_A`, `=B`) can make it sound less coherent if not handled carefully. These accidentals can create tension and release, which is good for musical interest but must be resolved correctly to sound pleasing.

### ### Specific Sections

- \*\*Measure 1\*\*:
- \*\*Measure 2\*\*:
- \*\*Measure 3\*\*:
- \*\*Measure 4\*\*:
- \*\*Measure 5\*\*:
- \*\*Measure 6\*\*:
- \*\*Measure 7\*\*:
- \*\*Measure 8\*\*:

### ### Conclusion

The melody and chord progressions can be coherent and pleasing if the chromaticism and accidentals are used intentionally to create tension and resolution. It may sound natural in a modern or experimental context but could be challenging in a more traditional tonal framework. To improve musicality, ensure that the chromatic notes resolve in a way that feels intentional, and consider adding harmonic support to clarify the underlying chord progressions.

# Wnioski

---

- Modele przeznaczone do generowania tekstu wykazują zdolności w analogicznych zastosowaniach muzycznych.
- Transformery oraz model Mamba radzą sobie z zadaniem generowania muzyki.
- Zwiększanie liczby parametrów modeli może prowadzić do bardziej skomplikowanych melodii.
- Innowacyjne metody oceny muzyki przy pomocy LLM mogą prowadzić do lepszego dotrenowania modeli.
- Możliwe jest stworzenie plug-inów do programów jak *FL Studio* czy *MuseScore* wspierających kompozytorów.
- Badania te stanowią początek rozwoju zaawansowanych modeli generatywnych dla muzyki.

# Dziękuję za uwagę

Filip Ręka