

# Raport projektu, Paweł Dopierała

06-S4IN01-P00991

## Cel projektu

Celem projektu było stworzenie modelu, który może generować tekst o sztucznej inteligencji stylem powieści „Pana Tadeusza”.

## Dane

Powieść „Pan Tadeusz” autorstwa Adama Mickiewicza pochodzi z linku

<https://wolnelektury.pl/media/book/txt/pan-tadeusz.txt>

Wykorzystano niecałą połowę powieści ze względu na ograniczone zasoby.

Teksty o sztucznej inteligencji wygenerowano samodzielnie, wpisując odpowiednie zapytania do Chatu GPT, a następnie łącząc odpowiedzi w jeden tekst.

## Modele

Wykonano model w oparciu o <https://www.geeksforgeeks.org/lstm-based-poetry-generation-using-nlp-in-python/>

Architektura modelu:

```
Model: "sequential_2"
```

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 75, 100)	1072800
bidirectional_2 (Bidirectional)	(None, 75, 300)	301200
dropout_2 (Dropout)	(None, 75, 300)	0
lstm_5 (LSTM)	(None, 100)	160400
dense_4 (Dense)	(None, 10727)	1083427
dense_5 (Dense)	(None, 10728)	115089984

```
=====  
Total params: 117707811 (449.02 MB)  
Trainable params: 117707811 (449.02 MB)  
Non-trainable params: 0 (0.00 Byte)
```

Dużym problemem w zadaniu było odpowiednie uczenie modelu. Najpierw wyuczono go wieloma epokami na samej powieści „Pan Tadeusz”, następnie mniejszą ilością epok na zbiorze danych zawierających teksty o sztucznej inteligencji. Następnie dostrajano model, ucząc go na obu zbiorach danych oddzielnie, a także razem, aż do uzyskania satysfakcjonującego rezultatu.

## Ewaluacja

Nie wykonano tradycyjnej ewaluacji modelu ze względu na charakterystykę projektu. Przykładowe działanie sieci widoczne jest na poniższym obrazie:

```
predict("CNI", 24)
'CNI «wielmożni nieruchomości głowę lecz weźmiem na świat ich umiała się wtłoczyć na końcu które w w chleba gałeczki sieci neuronowych i zdolność do generowania'
```

```
predict("GANy", 17)
'GANy i w dawnej surowości prawidłach wychował zakazy żołnierszczyzny na sklepieniu sieci neuronowych w w przetwarzaniu języka naturalnego'
```

## Wnioski

Uzyskane wygenerowane teksty są satysfakcjonujące. Widoczny jest styl pisania zgodny ze stylem powieści „Pan Tadeusz”, a także zdolność do generowania informacji w tematyce sztucznej inteligencji. Teksty zawierają błędy, które prawdopodobnie wynikały ze zbyt małej ilości danych i liczby epok w czasie uczenia modelu. Niestety, ze względu na ograniczone zasoby, było to konieczne.