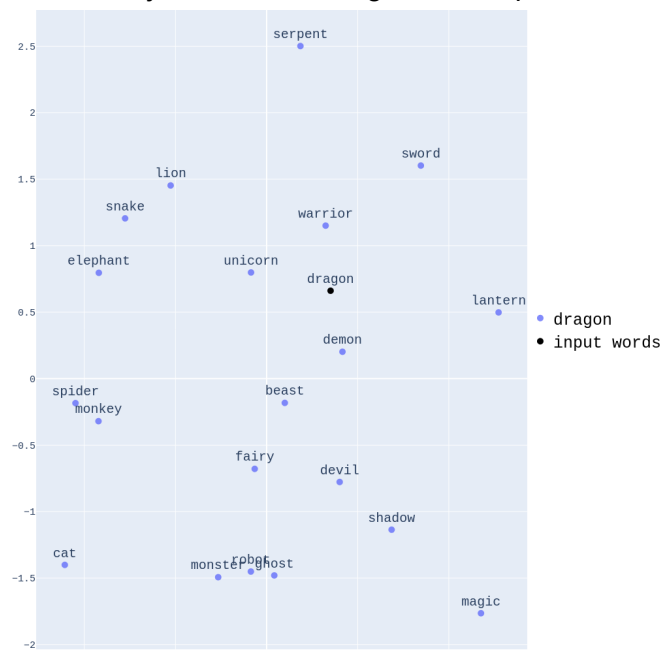


- Tytuł: Twitter bot
- [Kod projektu](#)
- Działanie projektu:
 - dwa konta na Twitterze: [Input](#), [Output](#)
 - wyszukanie w Tweepie konta Input słowa kluczowego poprzedzonego znakiem # (możliwy też parametr poprzedzony znakiem \$) i wygenerowanie przez Output odpowiedzi na tego tweeta
 - typy komend: #hello, #embedding, #emotion_detection oraz przypadek, gdy nie wykryto komendy
- Nauczanie maszynowe
 - Word embedding
 - Komenda:: #embedding
 - Dane: [glove.6B.100d.txt](#)
 - Redukcja wymiarowości 100D -> 2D z wykorzystaniem PCA
 - Wizualizacja w 2D dla danego słowa, np.



- Wykrywanie emocji w tekście
 - Komenda:: #emotion_detection
 - Dane: ponad 40k rekordów z dwoma kolumnami:
 - text - krótki tekst bez żadnych znaków interpunkcyjnych
Wszystko z małej litery.
 - jedna z następujących emocji: 'surprise', 'joy', 'love', 'anger', 'fear', 'sadness'
 - Podział na zbiór treningowy, testowy i walidacyjny.
 - Zamiana tych krótkich tekstów na liczby, gdzie każde słowo jest unikalnie przypisane do pewnej liczby z użyciem Tokenizera z tensorflow; dodanie paddingu do utworzonych sekwencji, żeby wyrównać ich długość.

■ Następująca architektura:

```
model = tf.keras.models.Sequential([
    tf.keras.layers.Embedding(10000, 16, input_length=50),
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(20, return_sequences=True)),
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(20)),
    tf.keras.layers.Dense(6, activation='softmax')
])

model.compile(
    loss='sparse_categorical_crossentropy',
    optimizer='adam',
    metrics=['accuracy']
)
```

■ Uczenie:

```
Epoch 1/5
7816/7816 [=====] - 109s 13ms/step - loss: 0.3180 - accuracy: 0.8736 -
val_loss: 0.1327 - val_accuracy: 0.9288
Epoch 2/5
7816/7816 [=====] - 103s 13ms/step - loss: 0.1155 - accuracy: 0.9343 -
val_loss: 0.1086 - val_accuracy: 0.9346
Epoch 3/5
7816/7816 [=====] - 102s 13ms/step - loss: 0.0978 - accuracy: 0.9398 -
val_loss: 0.1015 - val_accuracy: 0.9375
Epoch 4/5
7816/7816 [=====] - 103s 13ms/step - loss: 0.0924 - accuracy: 0.9412 -
val_loss: 0.0990 - val_accuracy: 0.9359
Epoch 5/5
7816/7816 [=====] - 103s 13ms/step - loss: 0.0890 - accuracy: 0.9424 -
val_loss: 0.0979 - val_accuracy: 0.9377
```

■ Wynik na zbiorze testowym:

```
2606/2606 [=====] - 14s 5ms/step - loss: 0.0955 - accuracy: 0.9397
```

- Pretrenowany transformer użyty do wygenerowania dalszej części tekstu na podstawie pierwotnego tekstu:

```
generator = pipeline('text-generation', model = 'gpt2')
```