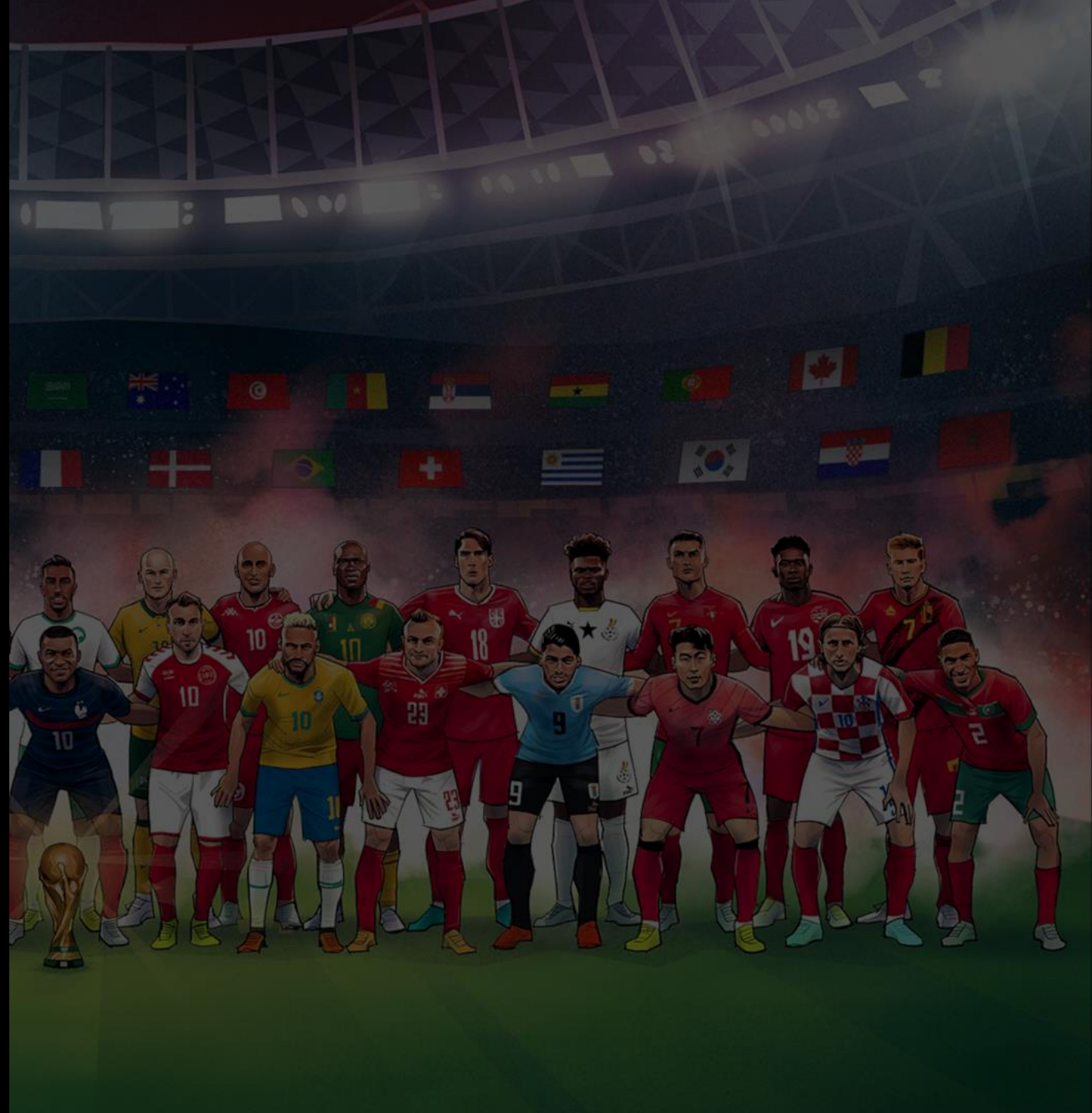
FIFA World Cup Chatbot

Aymane, Pascal, Jasmina, Lukas

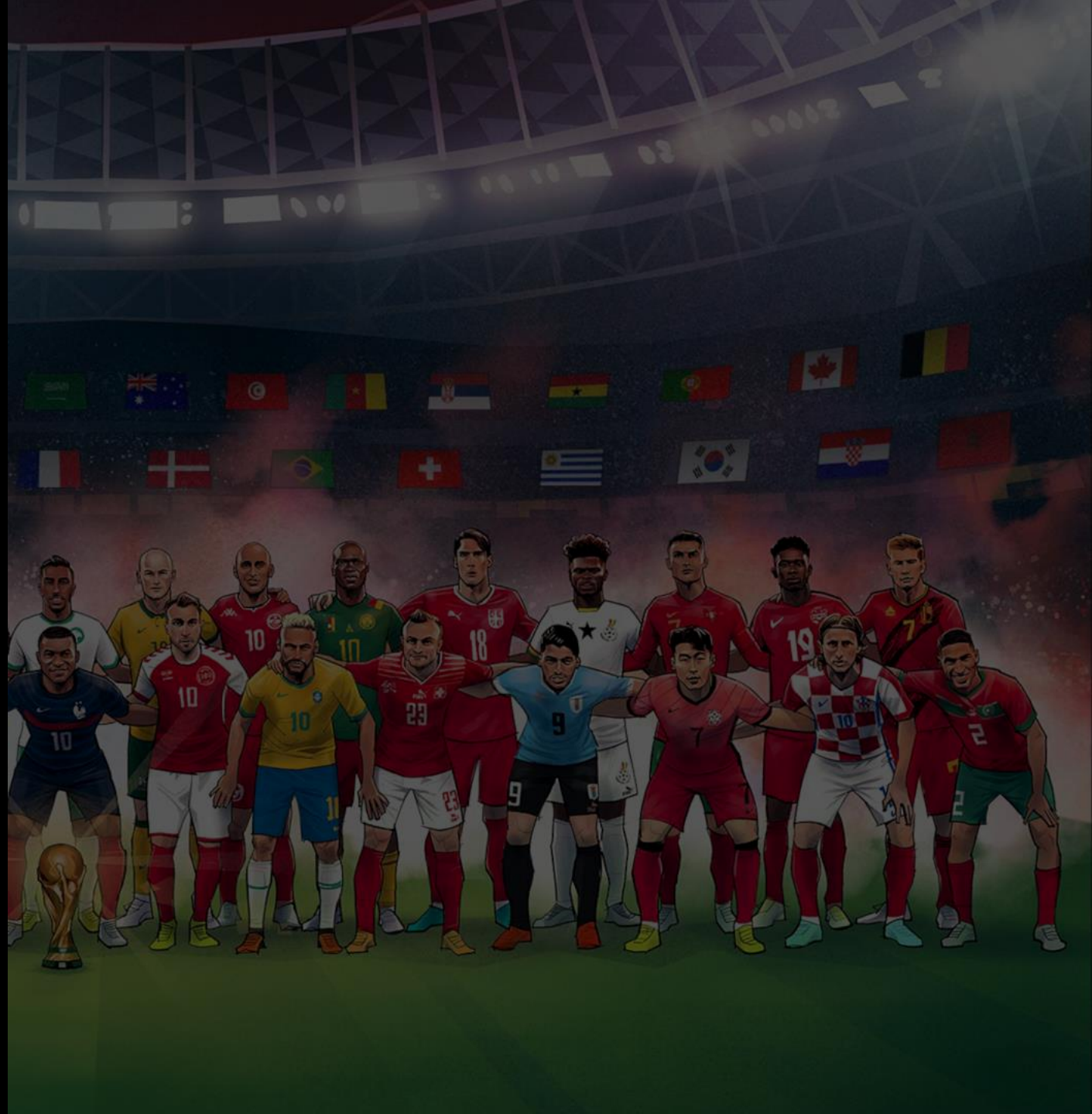
Agenda

1. Motivation
2. Recherche
3. Datengrundlage
4. Funktionsweise Chatbot
5. NLP – Pipeline
6. Fazit und Ausblick



Motivation

- Fußball = Volkssportart
- FIFA WM



Datengrundlage

Daten für Antwort

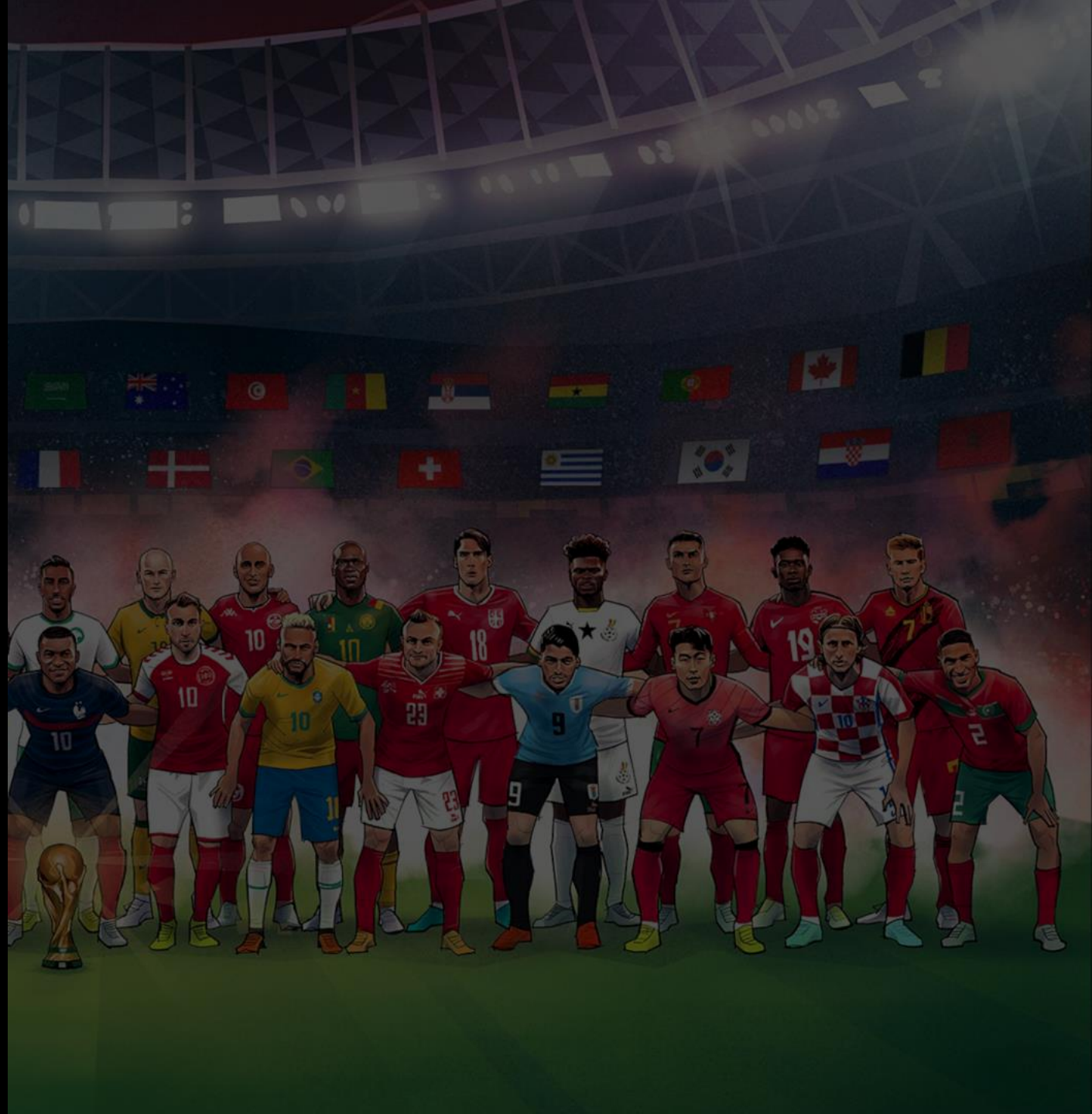
- API
- Kosten
- Datengrundlage schlecht
- Keine API für den Anwendungsfall
- Kaggle Dataset

Trainingsdaten

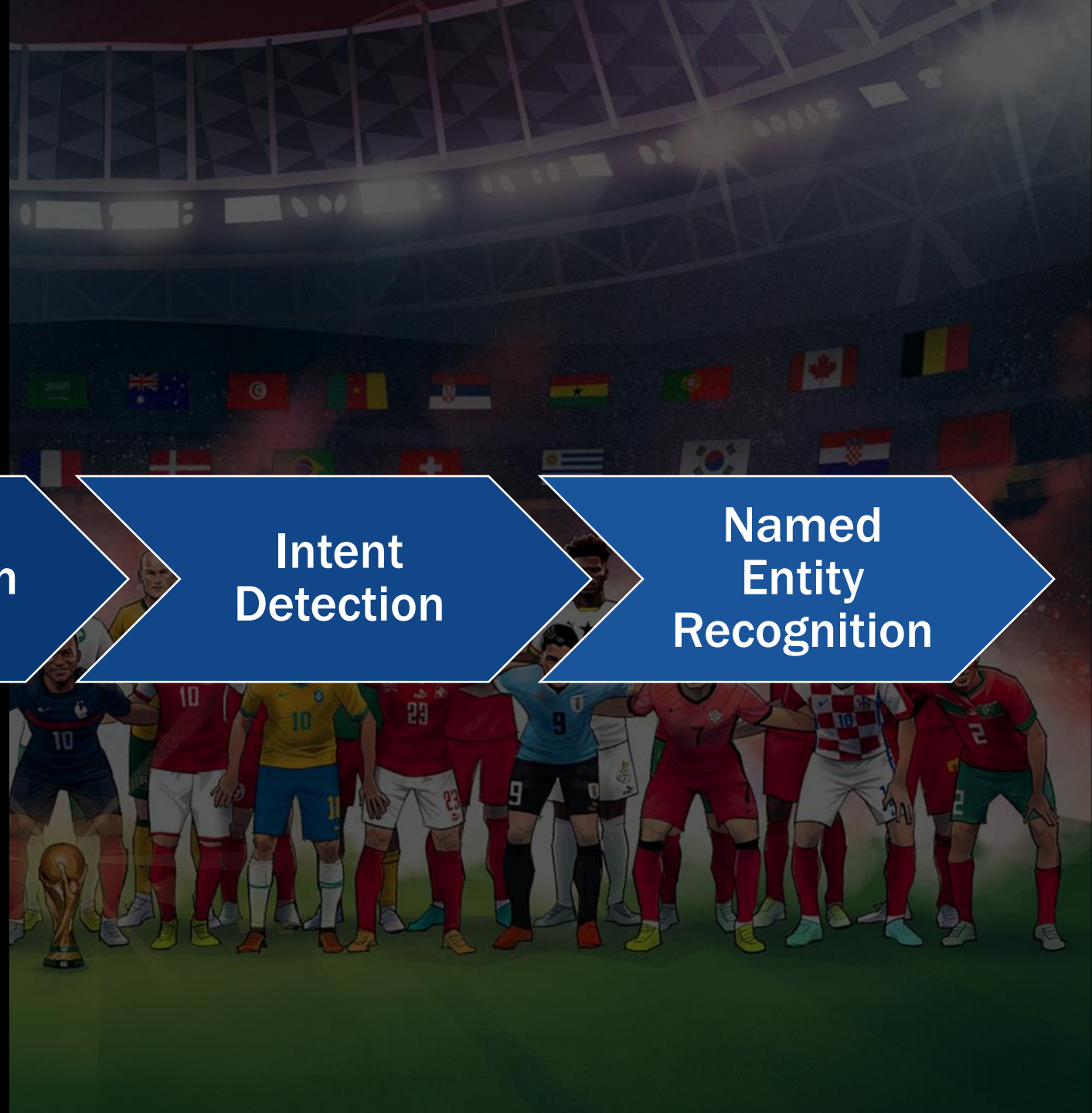
	query	category
0	Which team became 6th place in 2006?	PlacementTeam
1	Which team was world champion in the year 2010?	PlacementTeam
2	Who won in 1990?	PlacementTeam
3	Who was the winner in 1934?	PlacementTeam
4	Who became the world champion in 1938?	PlacementTeam
...	...	...
66	Who won the world cup in 1974?	firstPlace
67	Who was world-champion in 1950?	firstPlace
68	Who got the first place of the world cup in 2002?	firstPlace
69	Who was the winner of the world-championship i...	firstPlace
70	Who was the champion of the worldcup in 2010?	firstPlace

Funktionsweise

- Chatbot in JavaScript
  → Frage weiterleiten (JSON)
- Flask-App triggert NLP-Pipeline
  → Abfragen aus CSV-Files
- Antwort als JSON-File,
  Darstellung im Frontend
- Error-Handling



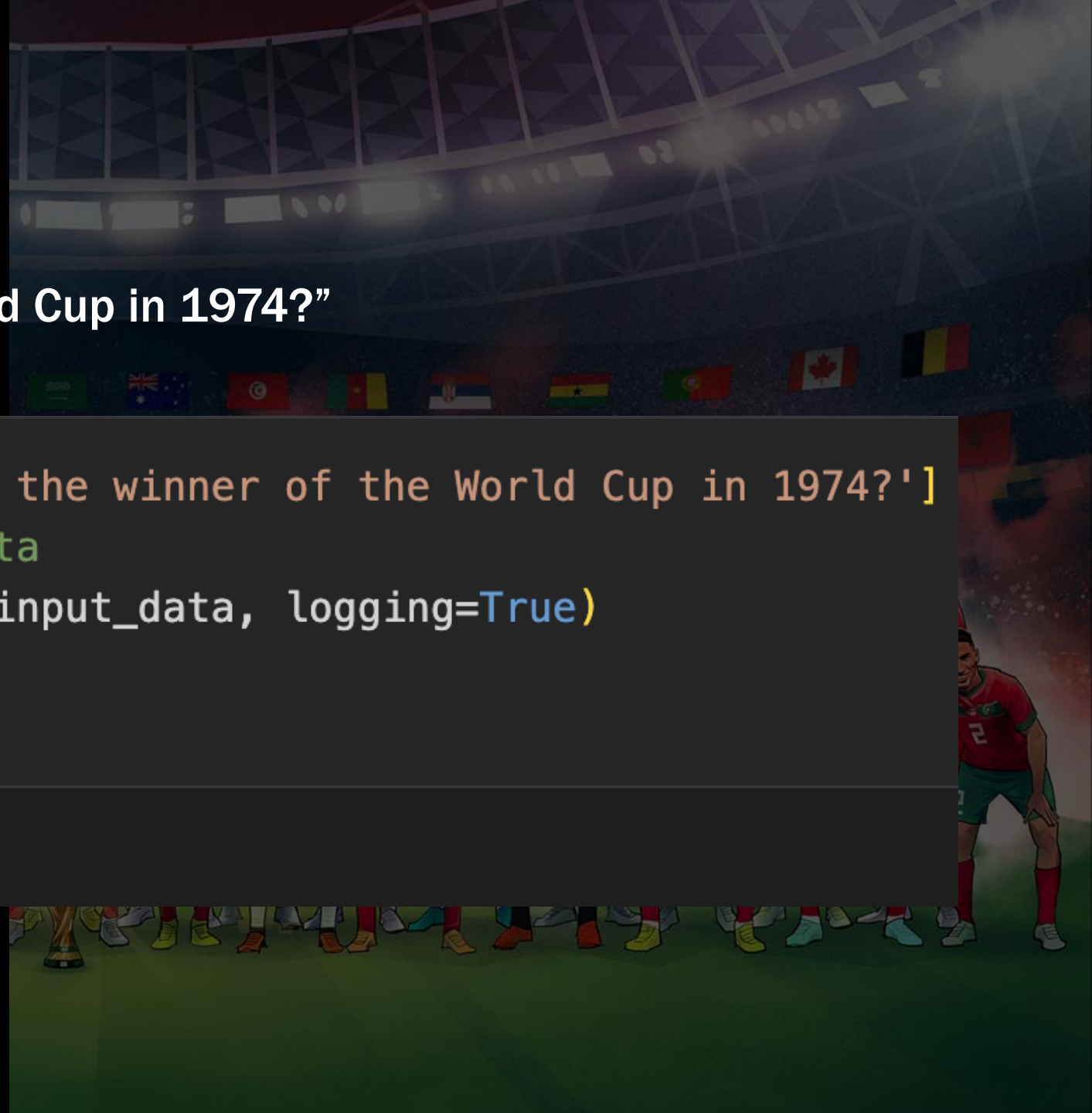
NLP Pipeline



Pre-Processing → Vectorization → Intent Detection → Named Entity Recognition

Pre-Processing Beispiel

“Who was the winner of the World Cup in 1974?”



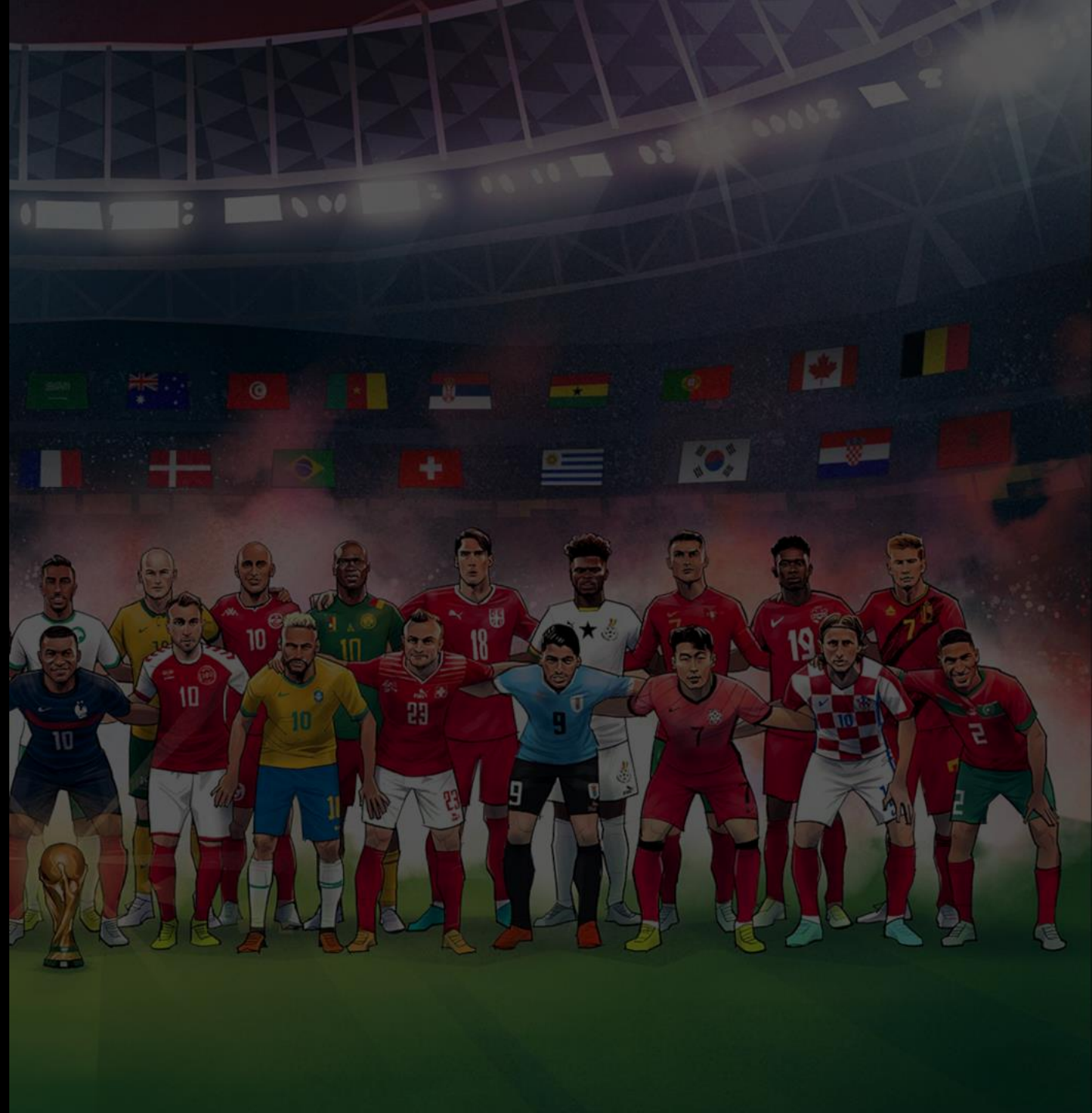
```
1  input_data = ['“Who was the winner of the World Cup in 1974?']
2  # cleaning the input data
3  cleanup = cleanup_text(input_data, logging=True)
4  cleanup[0]
```

✓ 0.2s

```
'winner world cup 1974'
```

Vectorization & Intent Detection

- grundlegende Funktionsweise
- Umsetzung
  - Word2Vec
  - tf-idf
  - Support Vector Classifier



Intent Detection - Step-by-Step Example

1. “Who was the winner of the World Cup in 1974?”

```
1  intent_categories = list(Encoder.classes_)
2  intent_categories
```

✓ 0.5s

```
['PlacementTeam',
 'TeamPlacement',
 'YearHost',
 'firstPlace',
 'year(avg)Goals',
 'yearGoals',
 'yearMatches']
```

Intent Detection - Step-by-Step Example

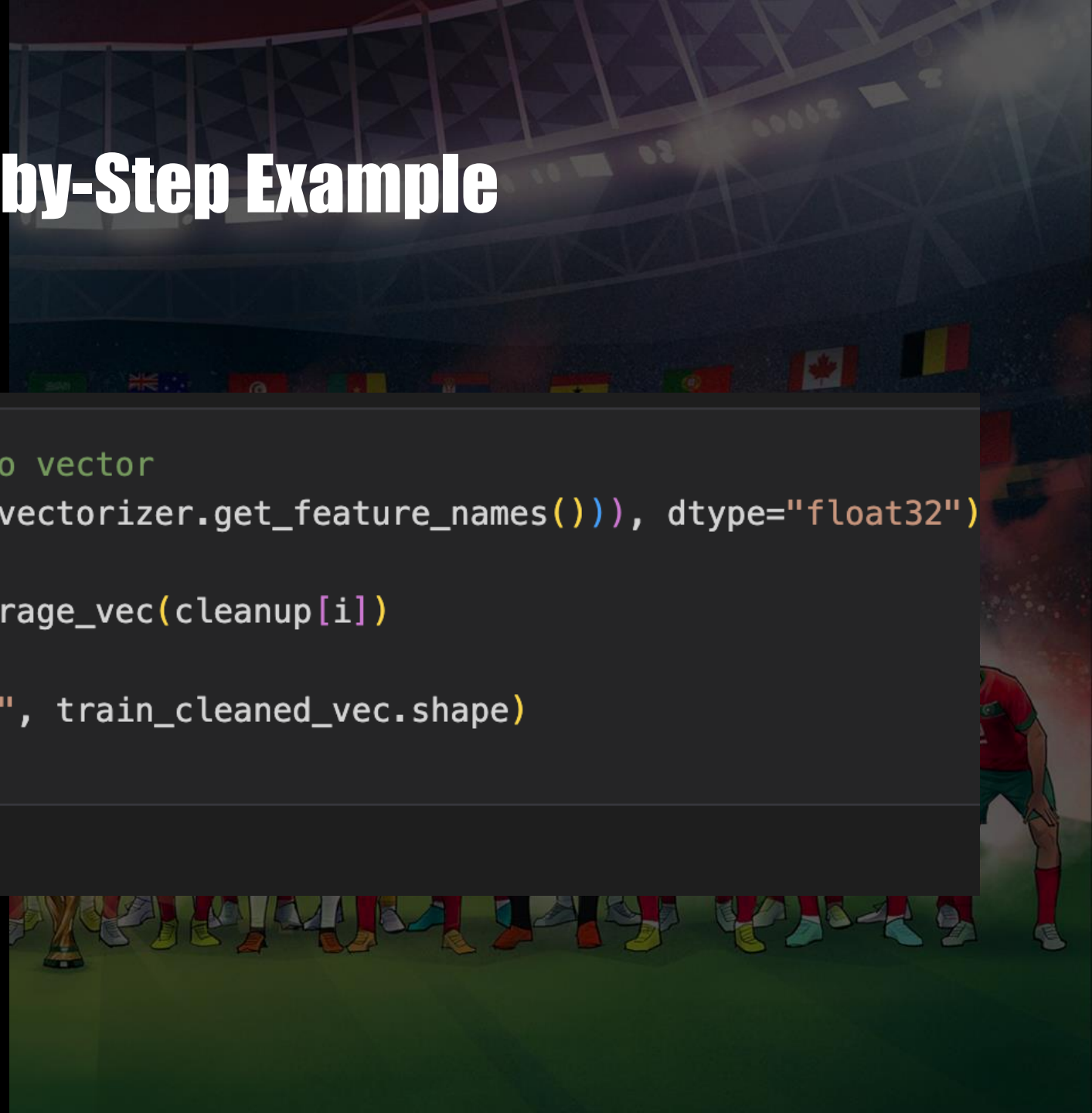
2.

```
# converting the cleaned data to vector
cleanup_vec = np.zeros((1, len(vectorizer.get_feature_names())), dtype="float32")
for i in range(len(cleanup)):
    cleanup_vec[i] = create_average_vec(cleanup[i])

print("Train word vector shape:", train_cleaned_vec.shape)
```

✓ 0.1s

```
Train word vector shape: (71, 88)
```

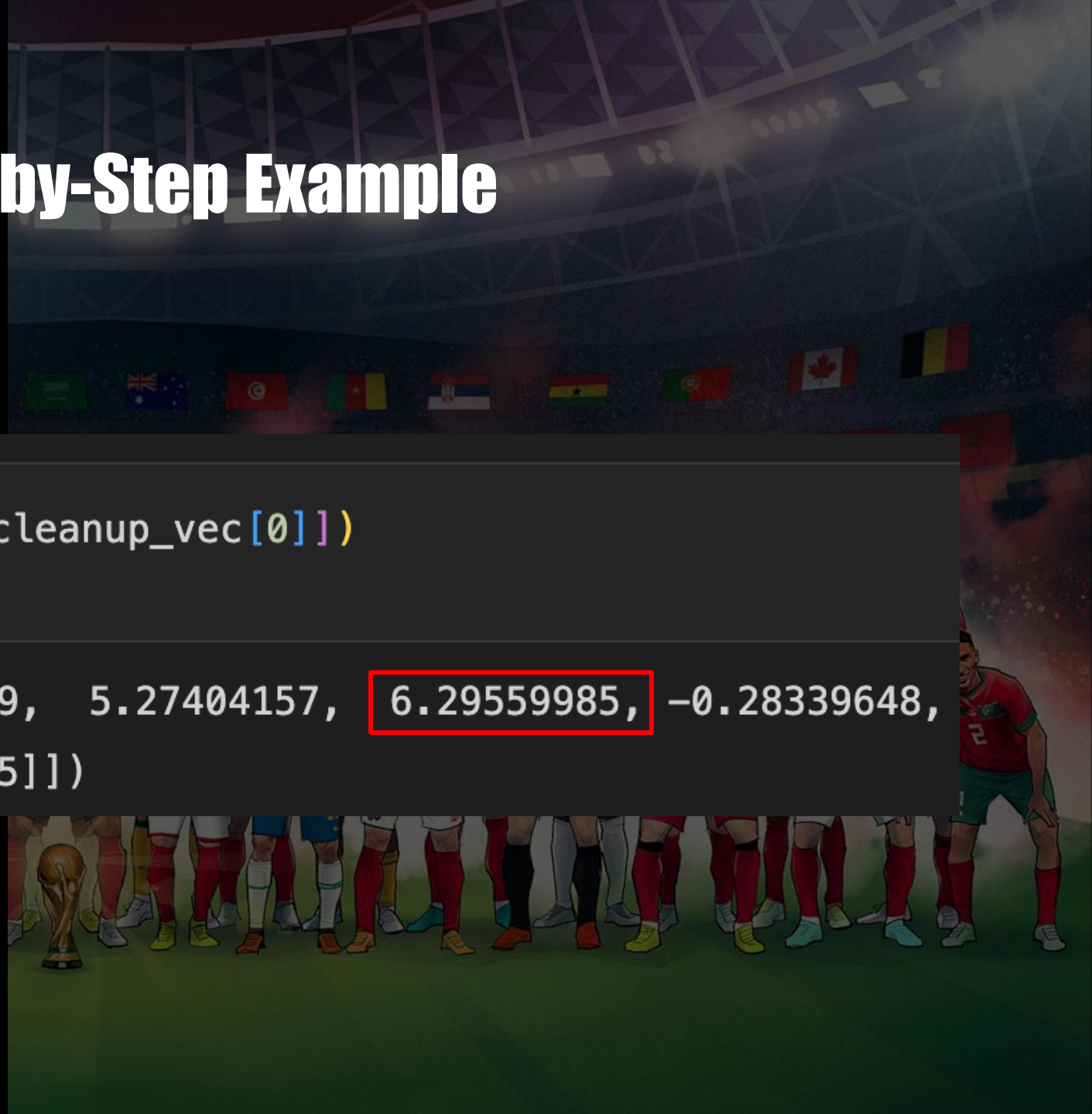


Intent Detection - Step-by-Step Example

3.

```
   1  grid.decision_function([cleanup_vec[0]])
 ✓  0.4s
```

```
array([[ 2.8382815 ,  0.75169329,  5.27404157,  6.29559985, -0.28339648,
         2.78789648,  2.82038755]])
```



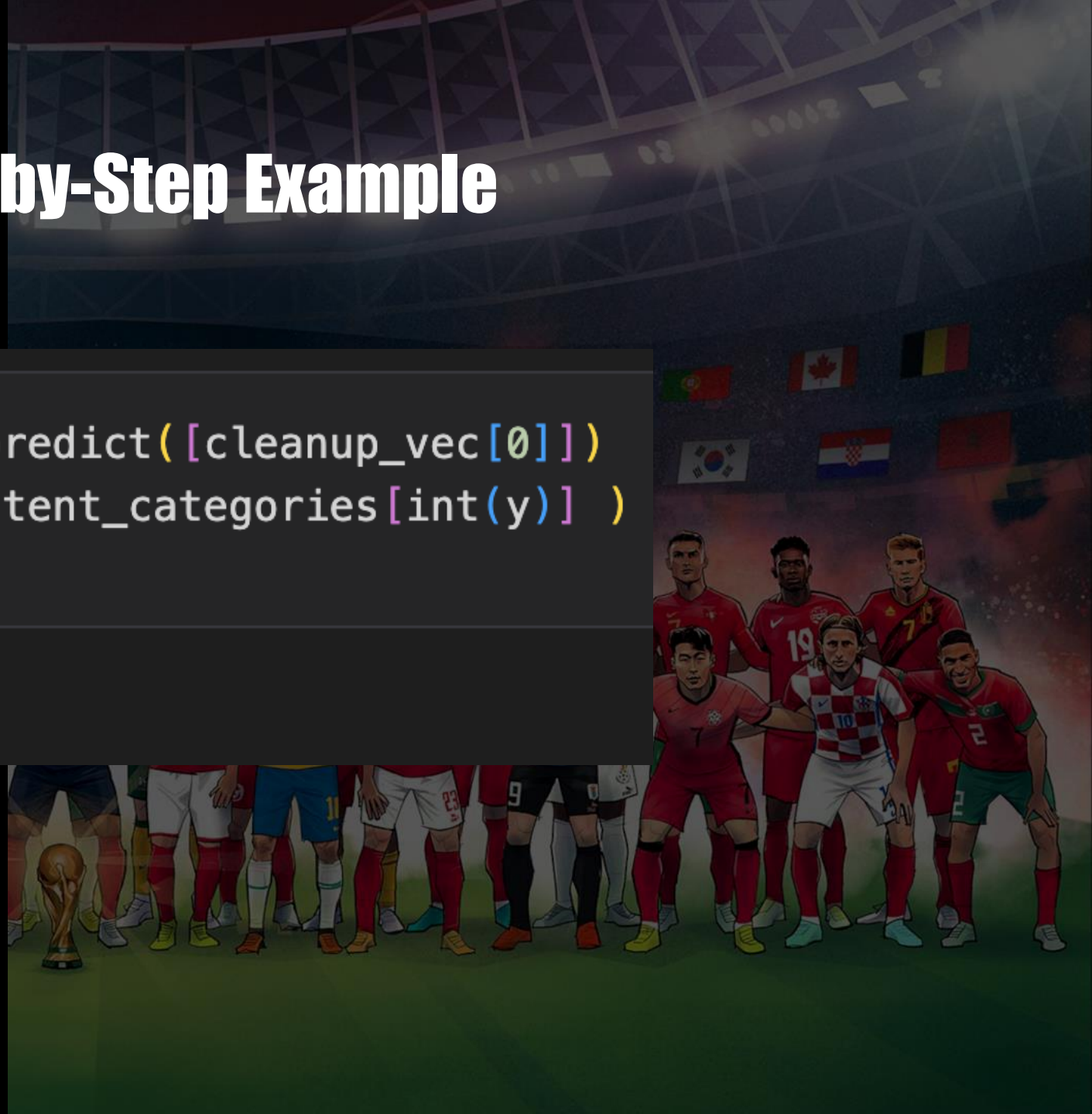
Intent Detection - Step-by-Step Example

4.

```
1  y = grid.predict([cleanup_vec[0]])
2  print(y,intent_categories[int(y)] )
```

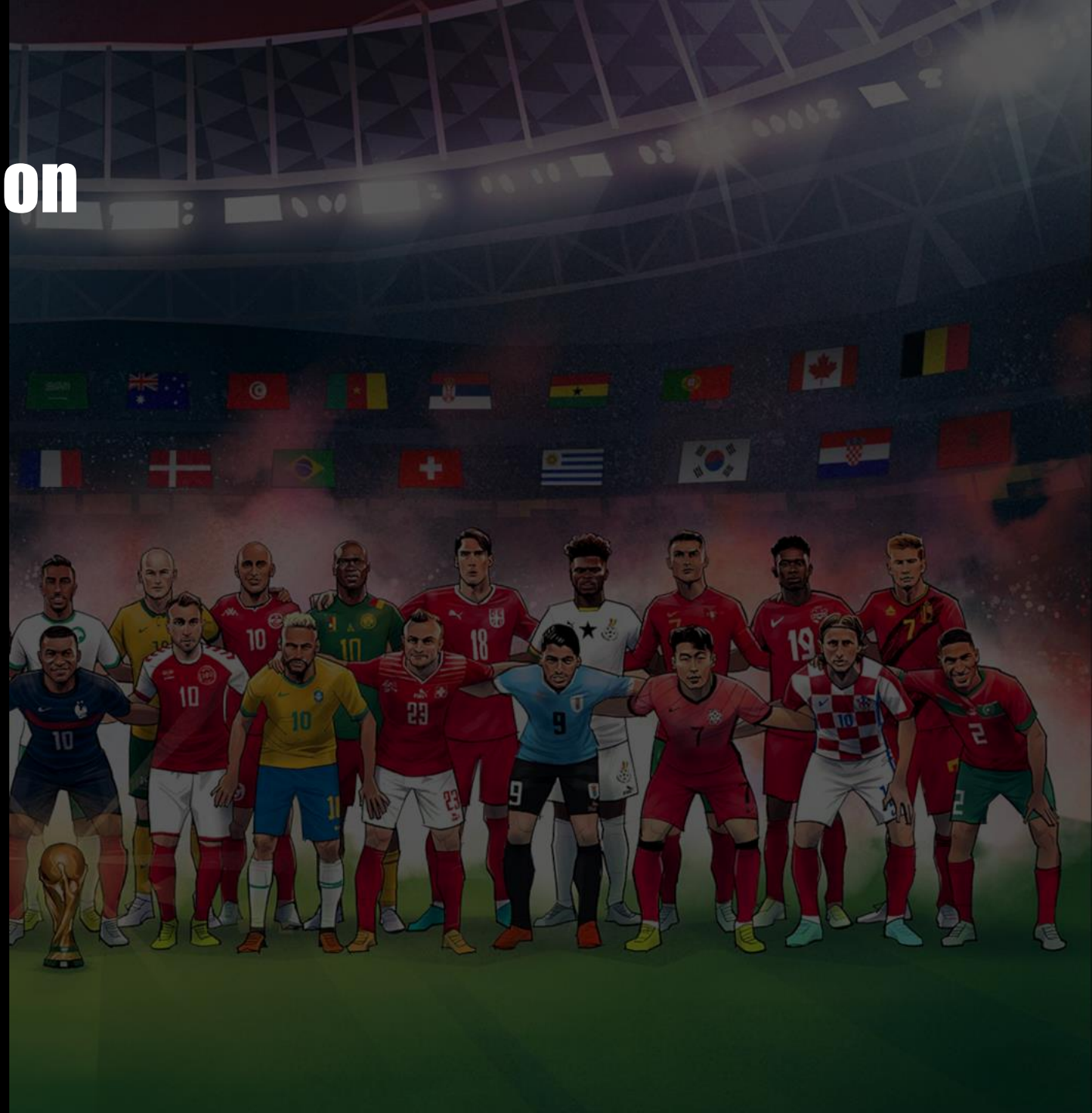
✓ 0.3s

```
[3] firstPlace
```



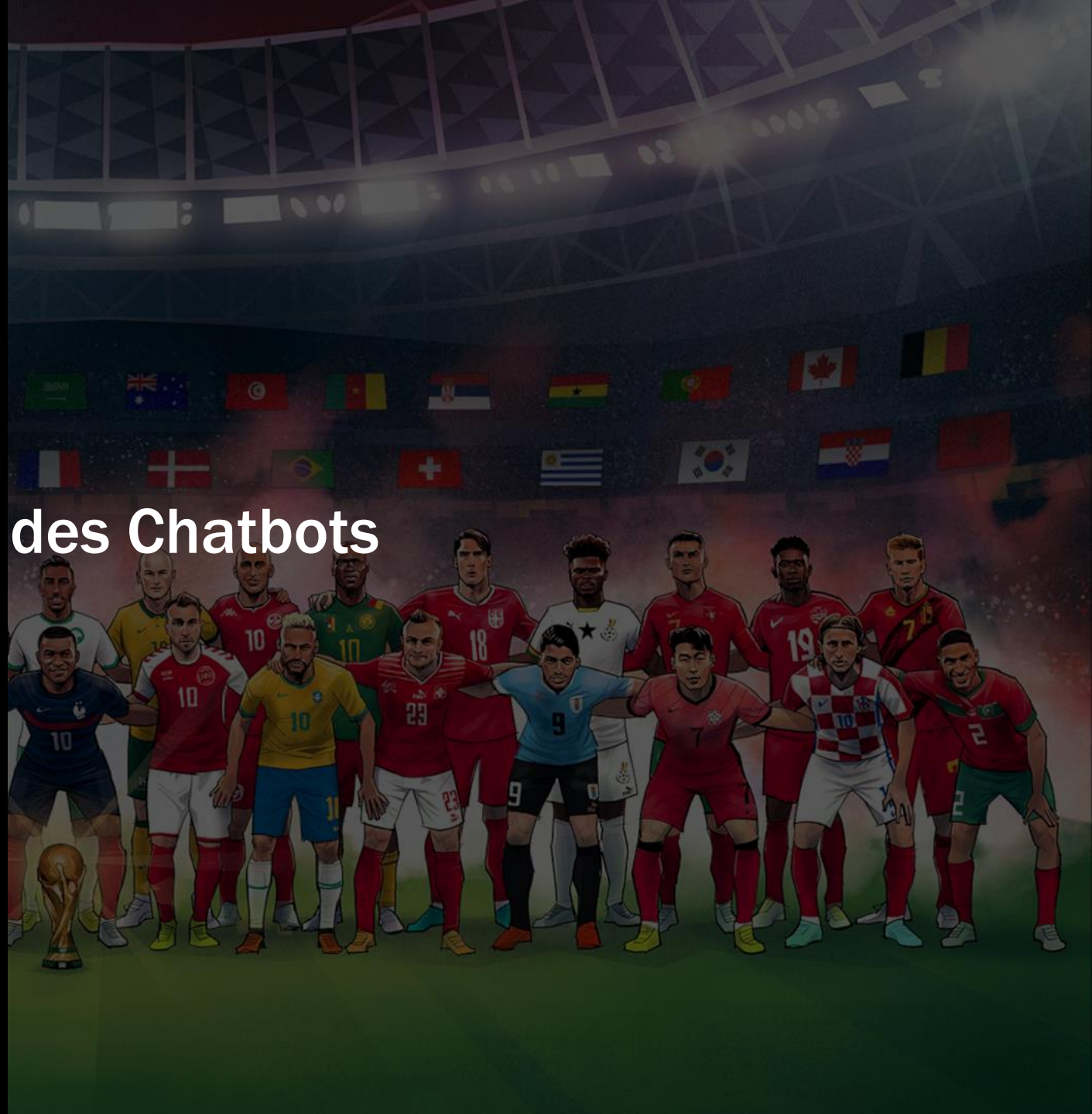
Named Entity Recognition

- SpaCy
- Entities: GPE, Date, Ordinal
- Numerizer
- Generierung von Antworten



Hands On

Demo des Chatbots



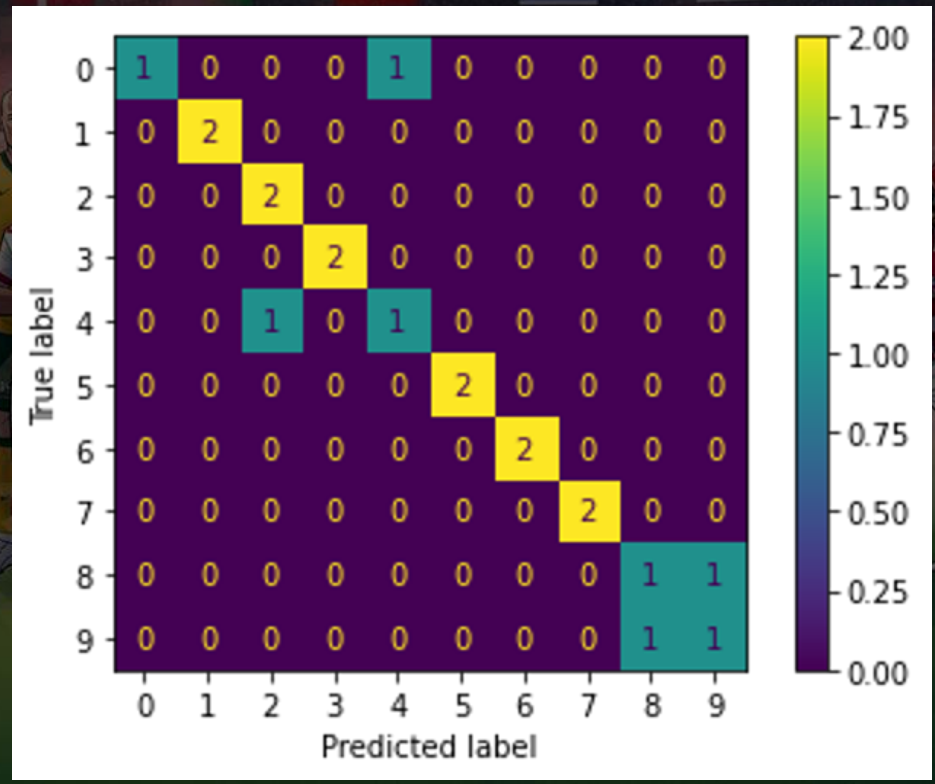
Fazit

- Antworten auf verschiedene Fragevarianten
- Erkennen allgemeiner - und Sportfragen-Intents

Anzahl Trainingsdaten: 90

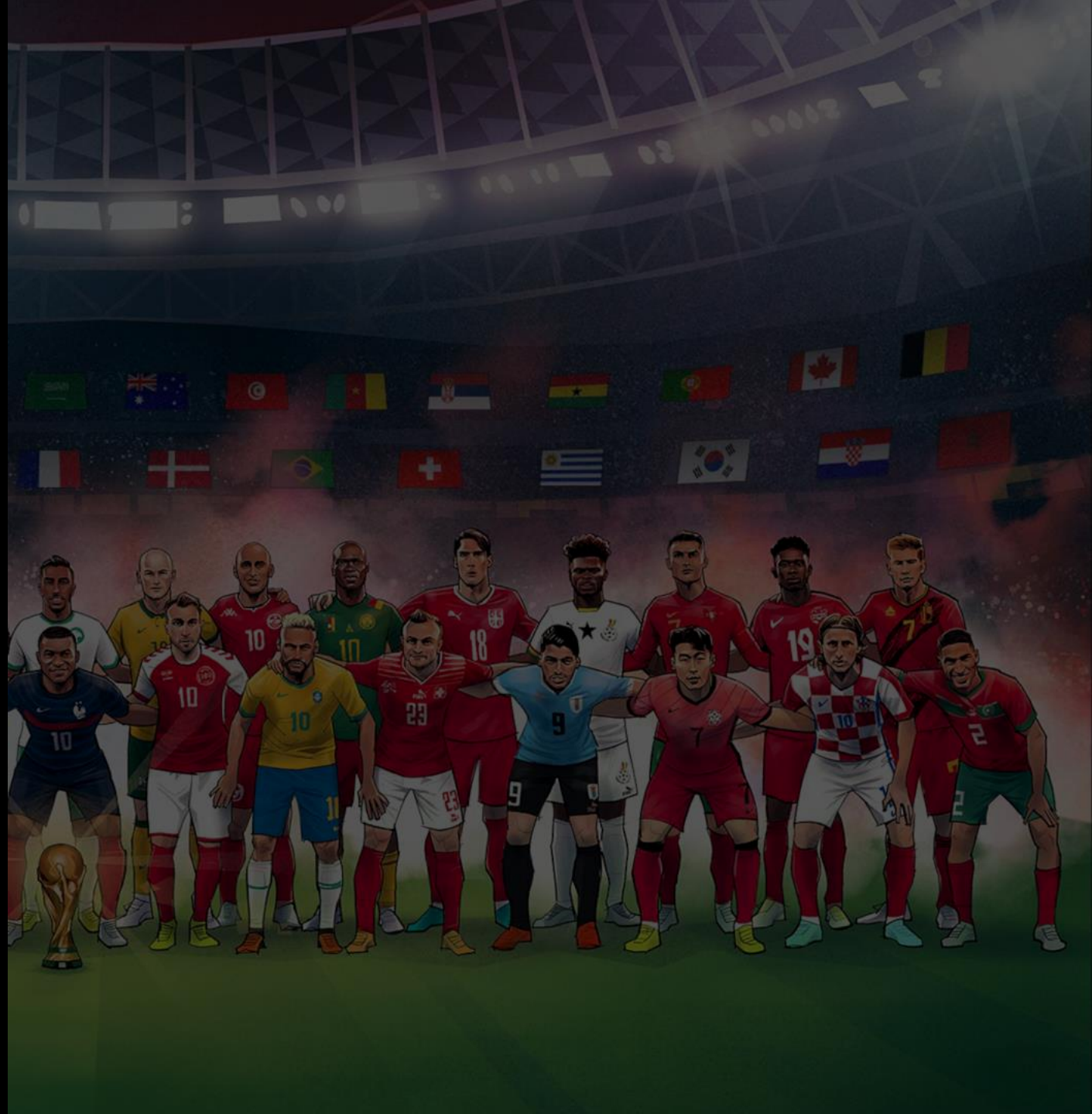
Anzahl Testdaten: 20

Accuracy: 80%



Ausblick

- Modellgenauigkeit
- Training weiterer Modelle
- Nachfragen an User
- API Anbindung
- Speech-to-Text
- Mobile Application



FIFA World Cup Chatbot

Aymane, Pascal, Jasmina, Lukas

Quellen

Intentanalyse

Liu, Bing, and Ian Lane. "Attention-based recurrent neural network models for joint intent detection and slot filling." arXiv preprint arXiv:1609.01454 (2016).

https://www.atlantis-press.com/journals/hcis/125963694

Cahn, Jack. "CHATBOT: Architecture, design, & development." University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science (2017).

https://www.kaggle.com/code/taranjeet03/intent-detection-svc-using-word2vec/notebook#)

Spacy

Numerizer: https://github.com/jaidevd/numerizer

https://spacy.io/usage/linguistic-features

Datengrundlage

https://www.kaggle.com/datasets/iamsouravbanerjee/fifa-football-world-cup-dataset?select=FIFA+-+2022.csv

Websiteentwicklung

https://github.com/patrickloeber/chatbot-deployment

Bilder

https://wallpapersden.com/2022-fifa-world-cup-hd-wallpaper/