Big Data Scenarios
Task 1

Laurenz Brahner, Jan Landmann

Table of contents

Summarization of news articles

With this code we demonstrate the ability of the transformer models to summarize texts.

The user can select an article from a data set of different news articles, as well as the desired result length.
The desired article is then summarized ar

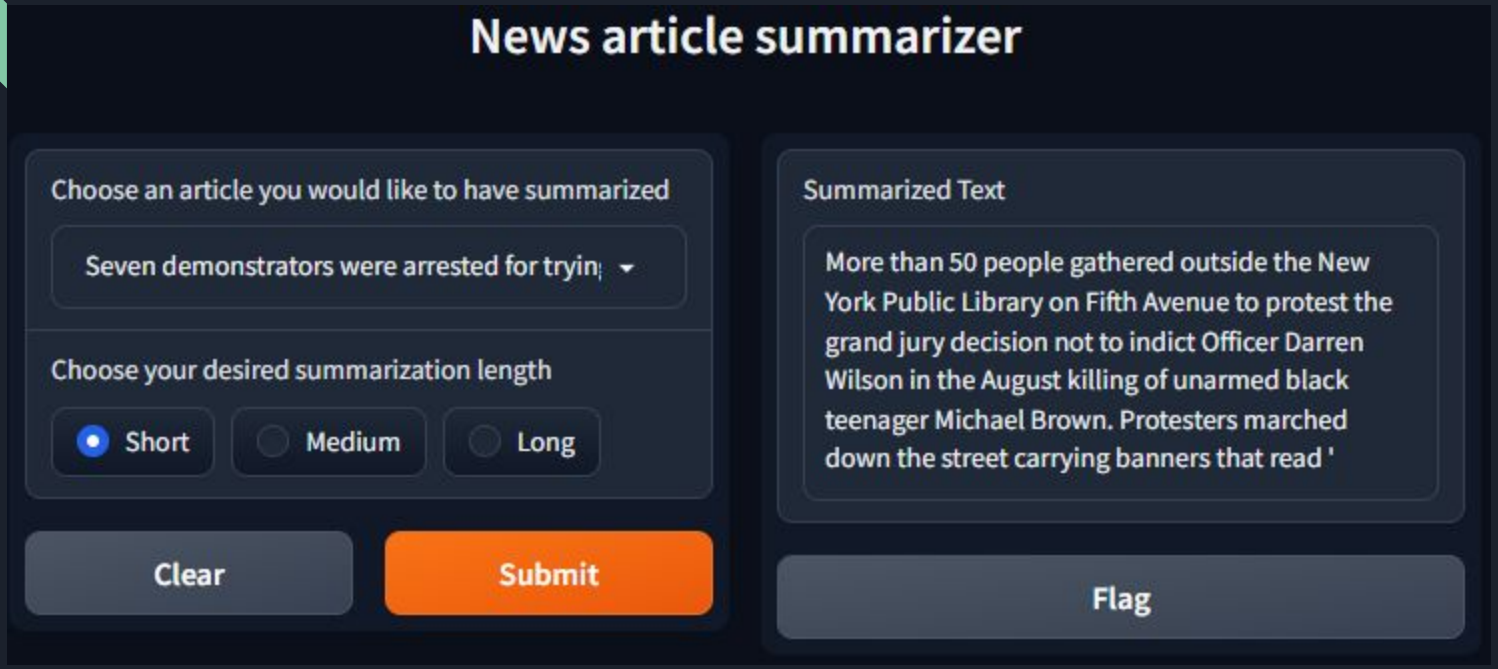
Summarization of news articles

Limitation:

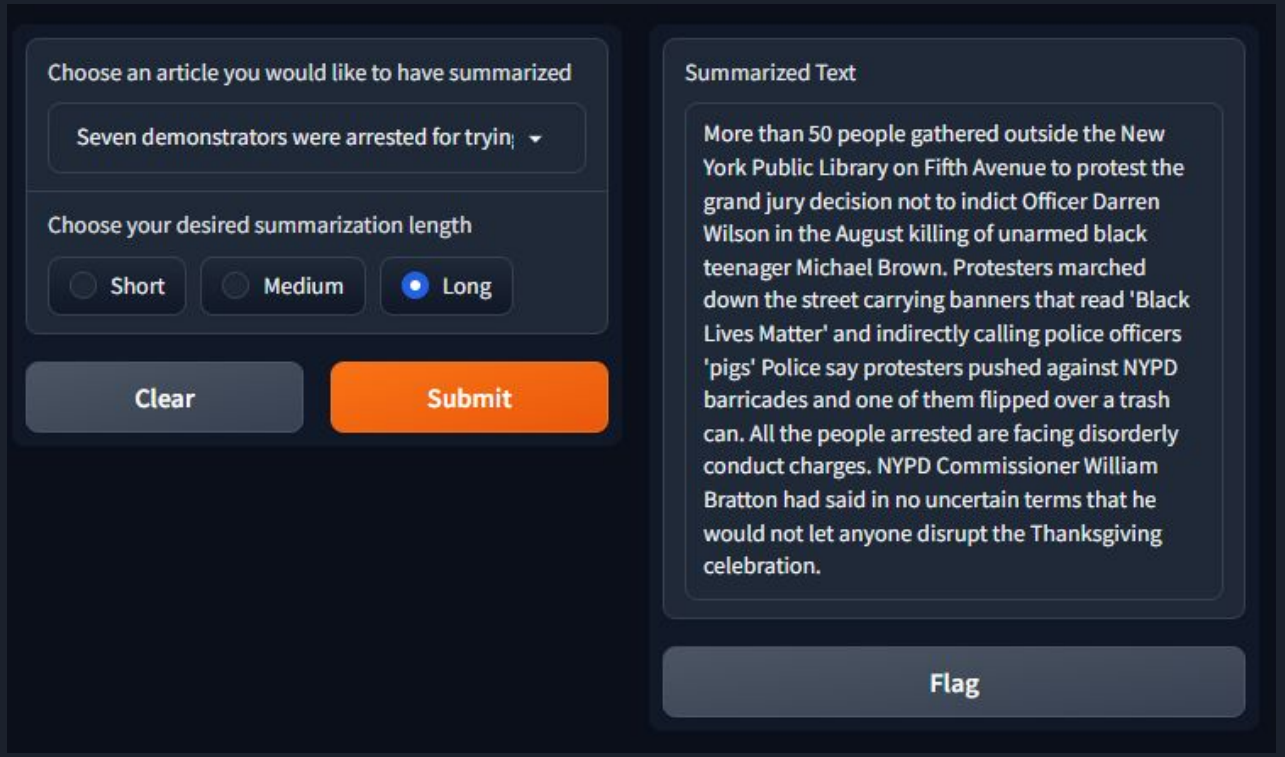
The maximum input length is limited. If only truncated, the missing context may influence the result negatively.

Summarization *Interface*

Summarization *Interface*



Summarization *Interface*



NER with sentences from wikipedia article

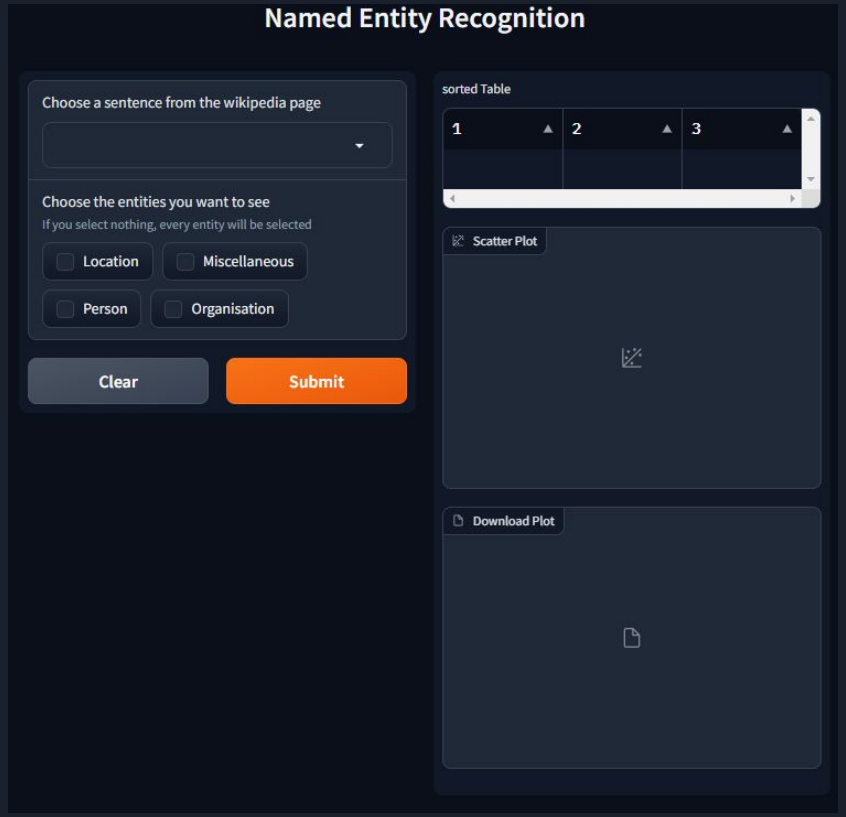
With this code we demonstrate the ability of the transformer models to perform named entity recognition.
For this purpose, a sentence can be selected from an already loaded Wikipedia article for which the NER is then performed.
You can either filter for specific entity types or output all entity types.
As a result, the program outputs a table showing the words sorted by frequency. The same information is also clearly displayed in a scatterplot.
The result is a table that is sorted in descending order by match. The results are also displayed graphically in a bar chart. The plot can be downloaded as HTML.

NER with sentences from wikipedia article

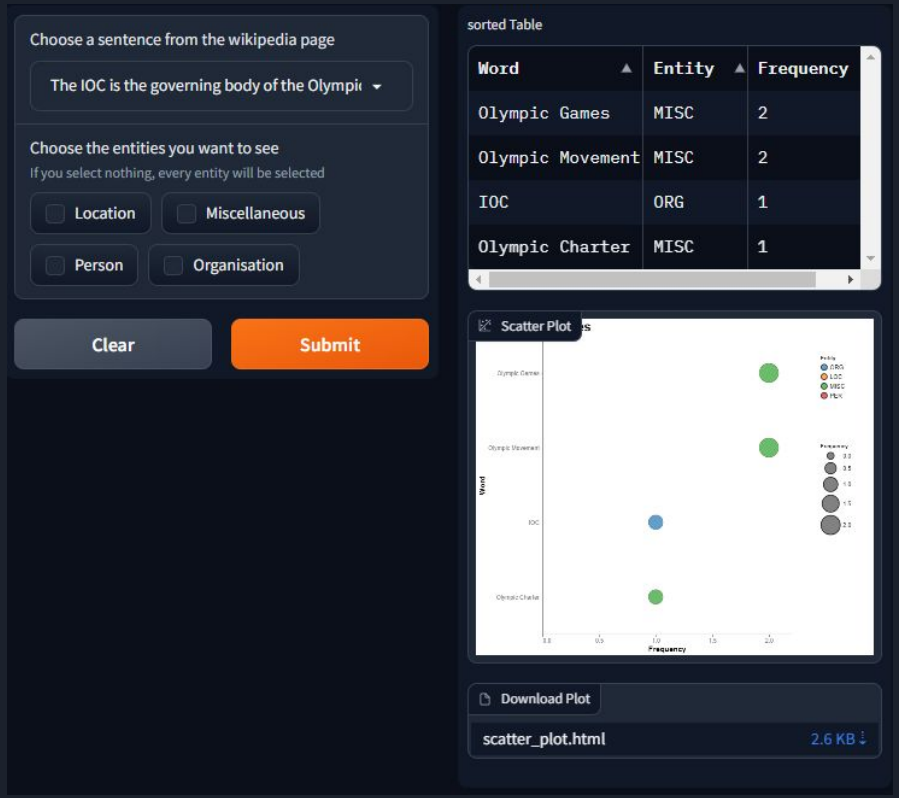
Limitations:

The model in its current state sometimes struggles to categorize entities correctly. The maximum input sequence length is also limited.

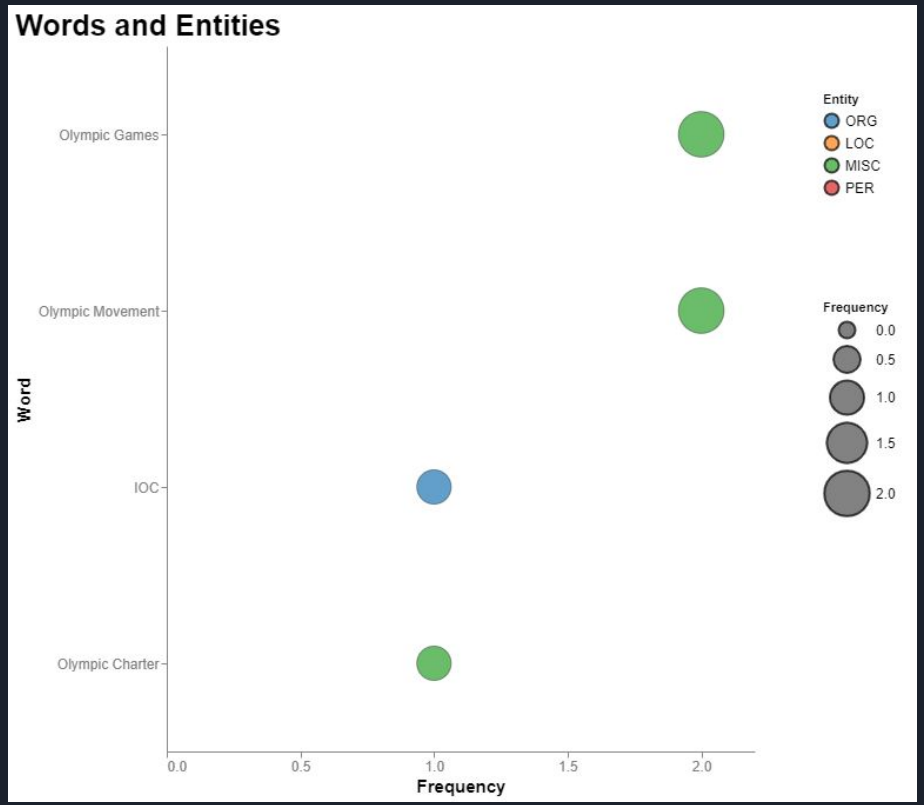
NER *Interface*



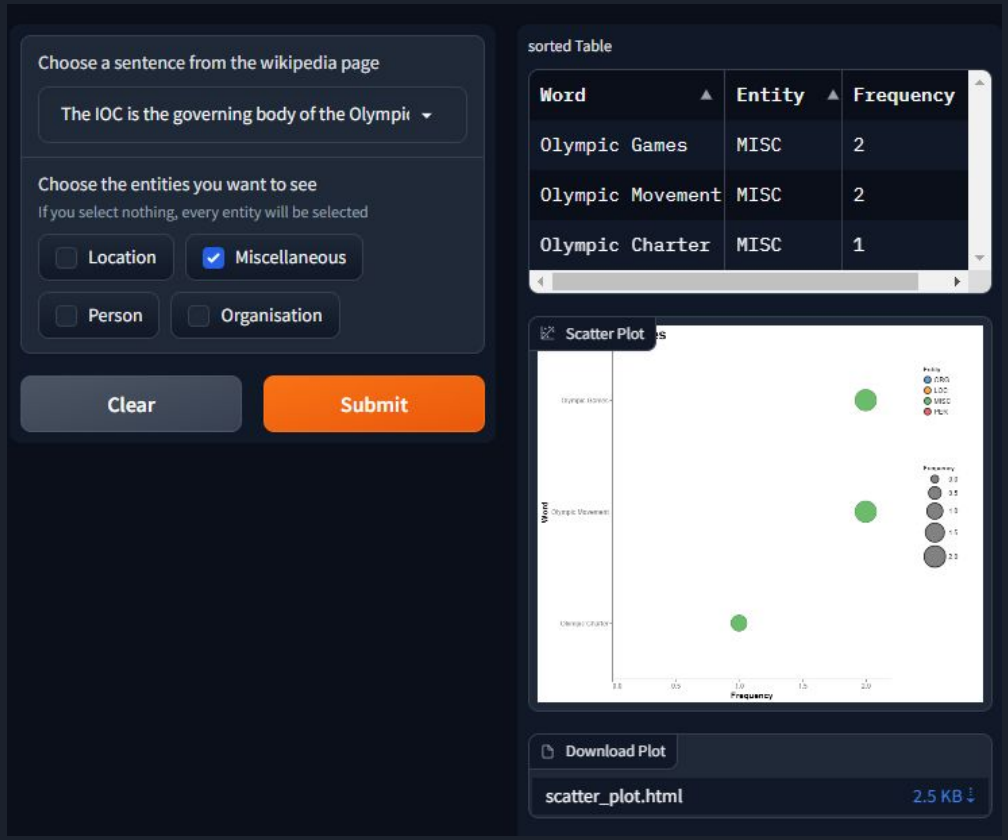
NER *Interface*



NER *Interface*



NER *Interface*



Sentiment analysis of yelp reviews

With this code we demonstrate the ability of the transformer models to perform sentiment analysis.

You can either enter a text yourself or select a review from a dataset from yelp reviews.

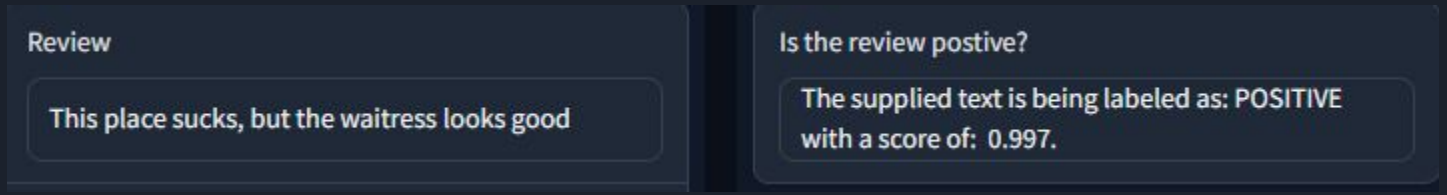
The program then determines whether the input is rated as positive or negative. The corresponding score is also displayed in numerical form.

Sentiment analysis of yelp reviews

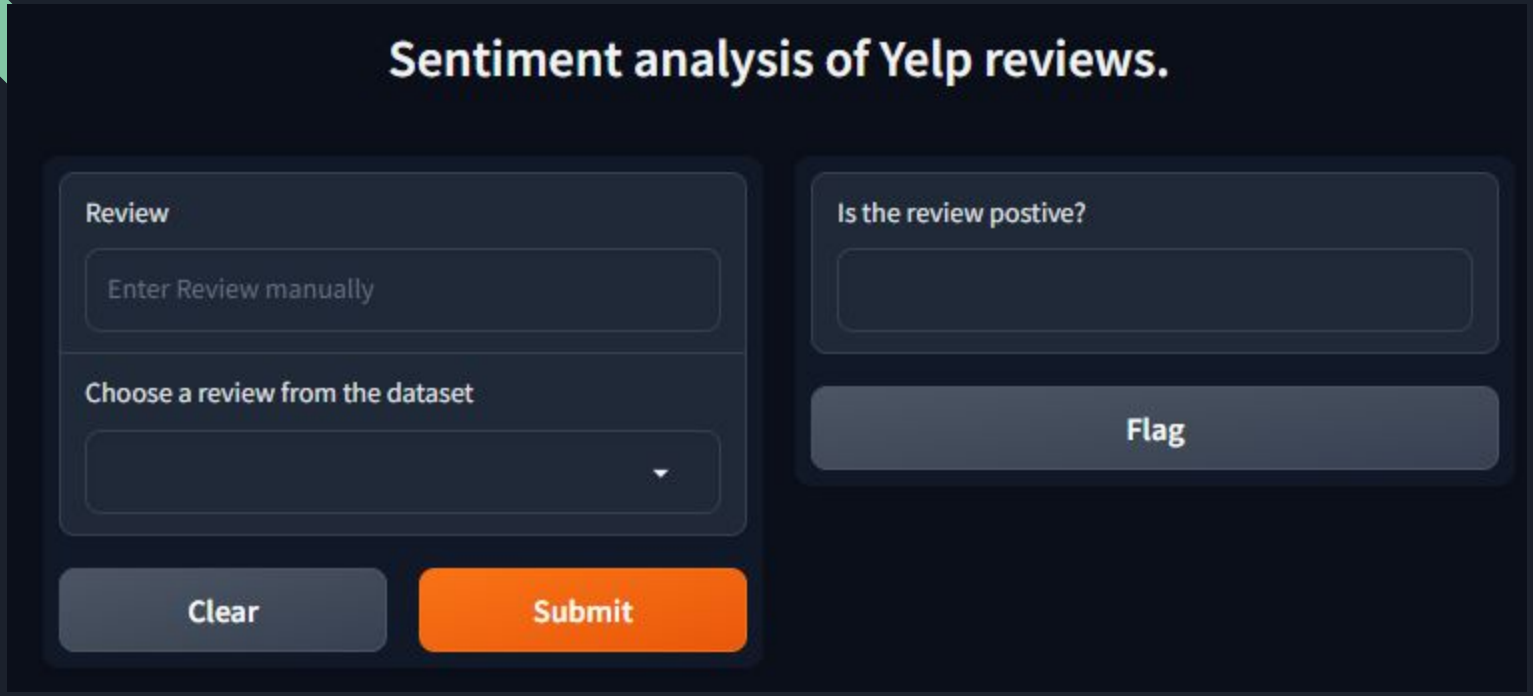
Limitations:

If the given input is neither positive nor negative the sentiment analysis is not able to determine whether the input is neutral.

It also struggles to decode colloquial language or more complex inputs.

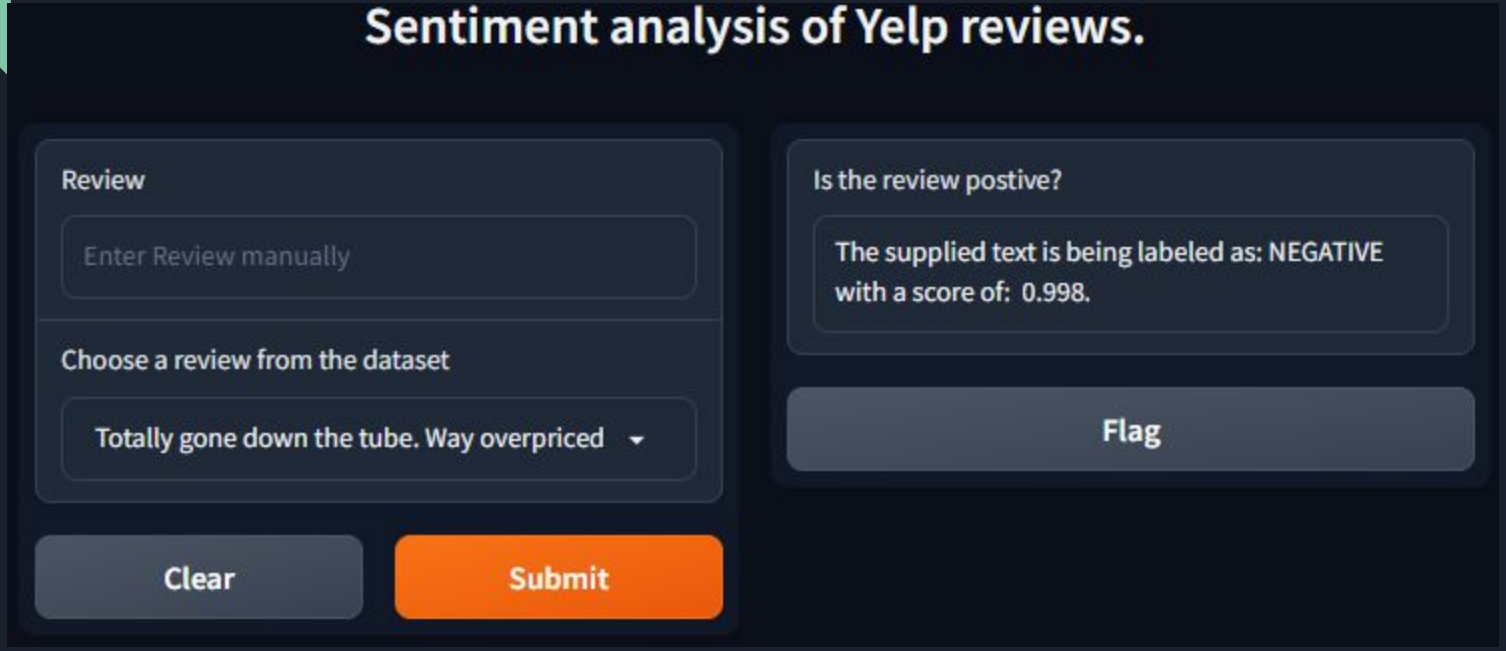


Sentiment analysis *Interface*



Sentiment analysis *Interface*

Sentiment analysis *Interface*



Translation of text with **language detection**

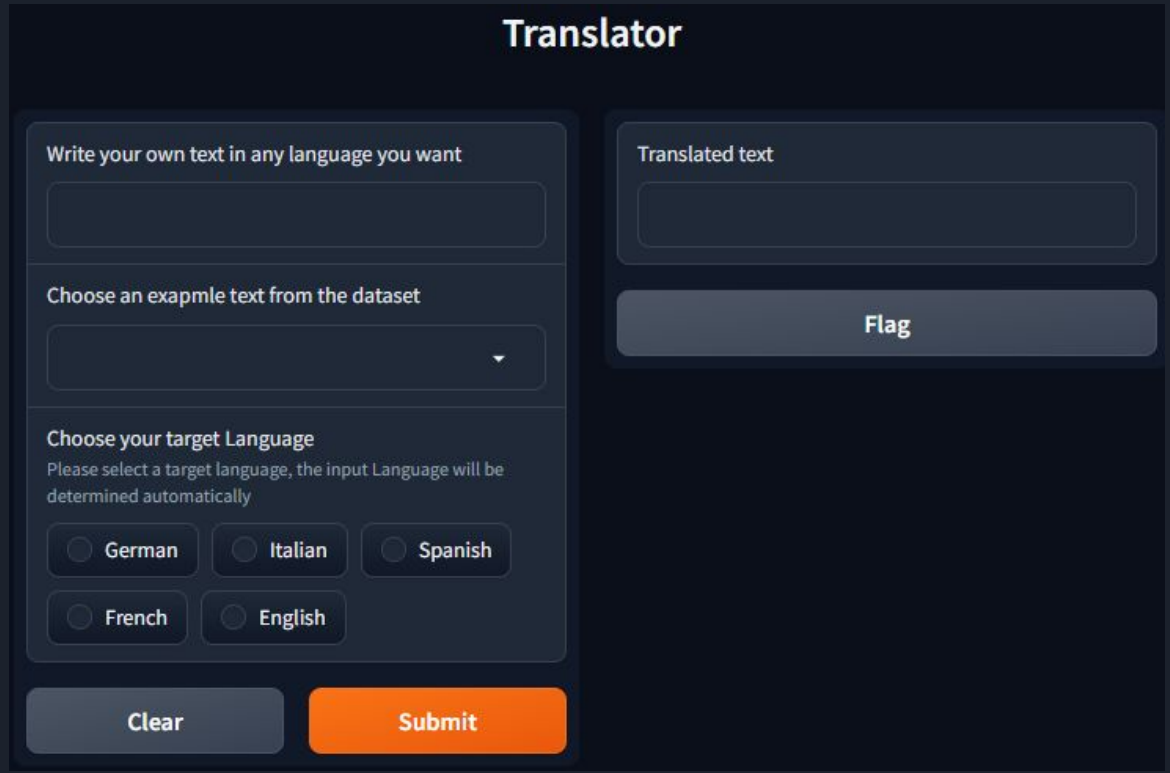
With this code we demonstrate the ability of the transformer models to translate texts with speech recognition. The user can enter any text in the input field and select the desired output language from a list of five languages.

During input, the entered language is automatically recognized and translated into the selected language.

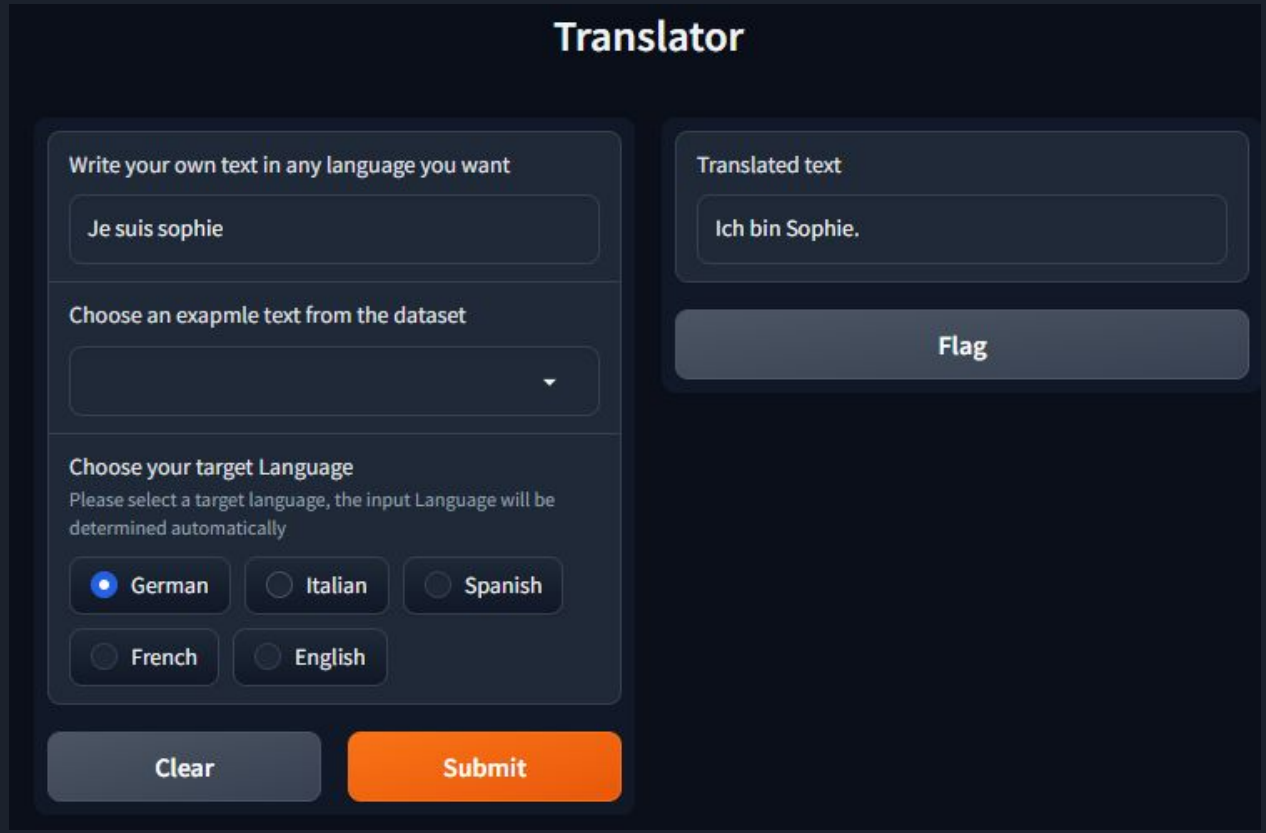
Translation of text with **language detection**

Limitations

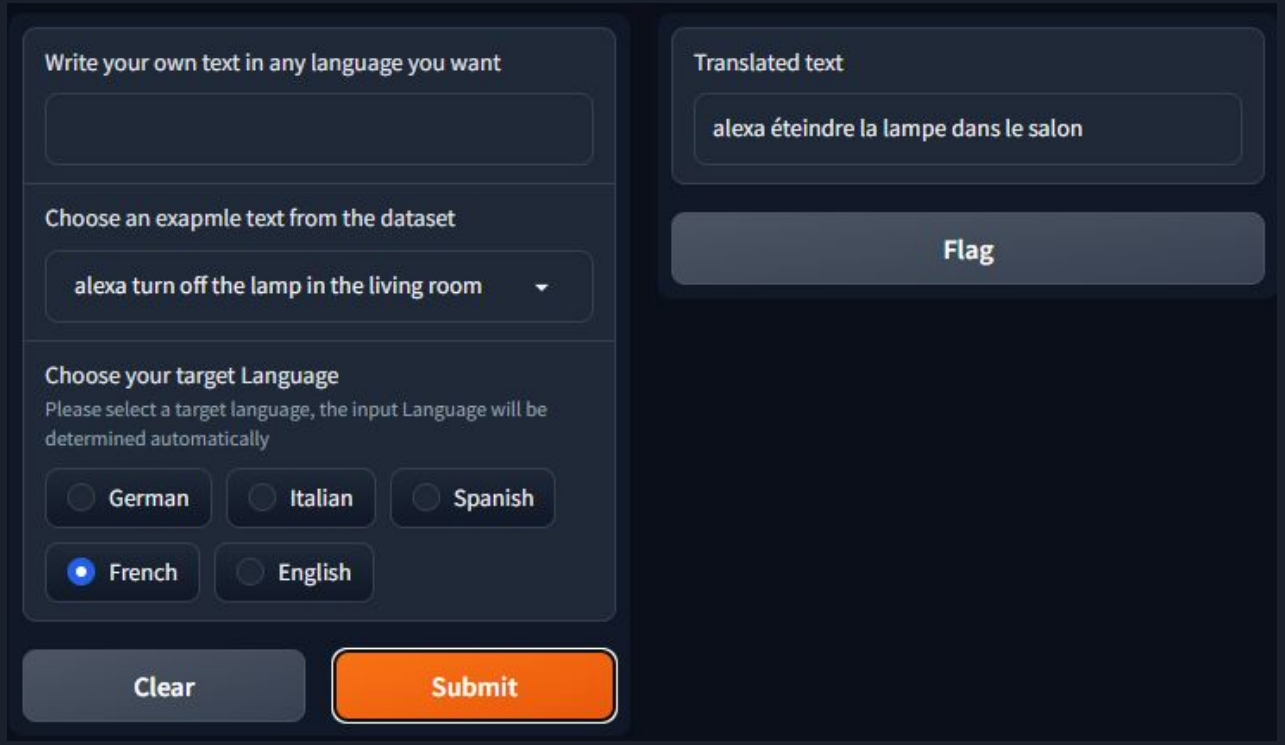
Translation *Interface*



Translation *Interface*



Translation *Interface*



Zero-shot classification of recipes and corresponding cuisines

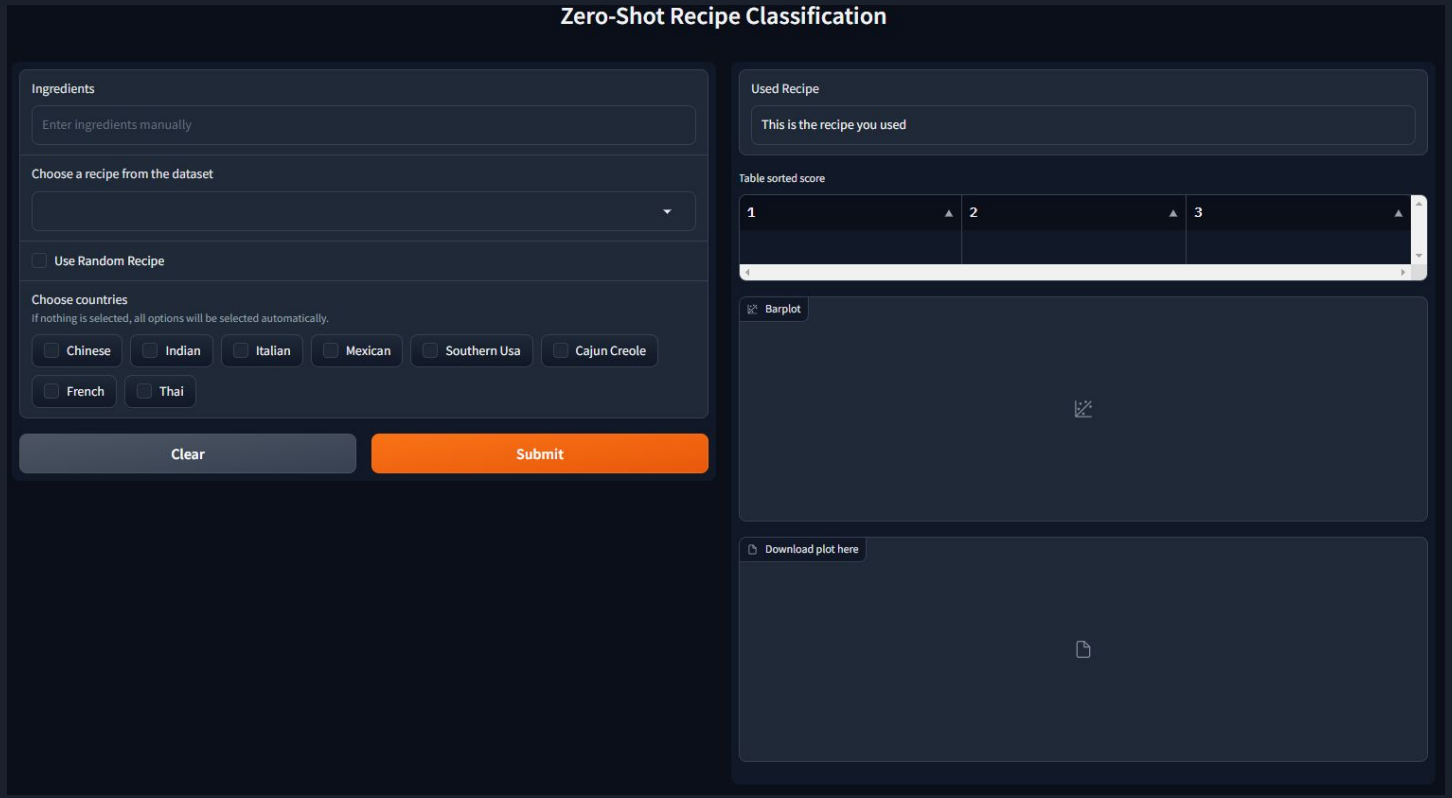
With this code we demonstrate the ability of the transformer models to perform a zero shot classification. You have the option of making an entry yourself, selecting from a pool of existing ingredient lists or having an entry selected automatically from the pool. you can then select one or more national cuisines from predefined values. If none are selected, all are selected automatically. In this example, the national cuisines serve as labels for the zero shot classification.
The result is a table that is sorted in descending order by match. The results are also displayed graphically in a bar chart. The plot can be downloaded as HTML.

Zero-shot classification of recipes and corresponding cuisines

Limitations:

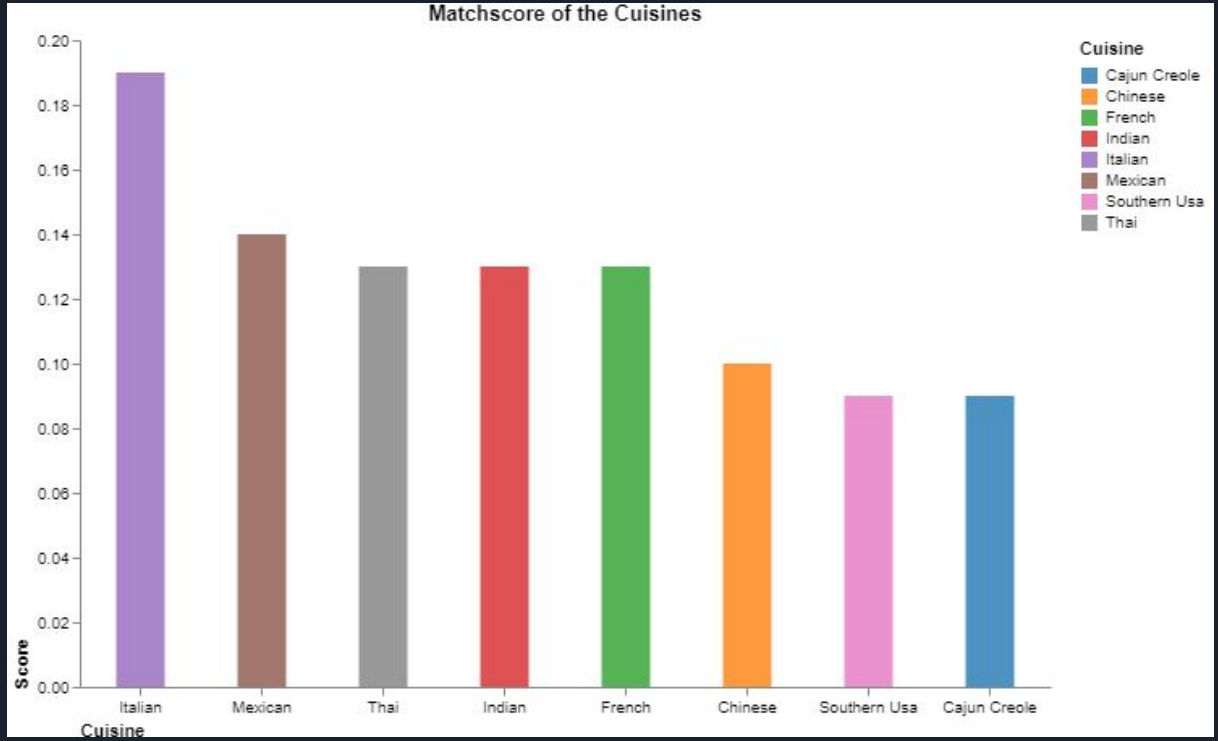
The model sometimes struggles to choose between different labels if the input is associated with multiple labels.

Zero-shot classification *Interface*



Zero-shot classification *Interface*

Zero-shot classification *Interface*



Conclusion

- It works already quite well
- but especially the NER had problems with our data

→ So we have to train our models with our data to get better results