Natural Language Processing: a Hugging Face introduction

Diego M. Jiménez Bravo[1]

Assistant Professor of Computer Science and Artificial Intelligence

[1] Expert Systems and Applications Lab, Department of Computer Science and Automation, University of Salamanca

dmjimenez@usal.es

@dmjimenezbravo

University of the Itajaí Valley, Itajaí, Brazil

April 4th, 2023

VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

ESA LAB

Expert Systems and Applications Laboratory

Contents

UNIVALI

ESA LAB

Who am I?

1. Who am I?

UNIVALI

ESA LAB

Hi!

I'm Diego Manuel Jiménez Bravo.

🎓 Ph.D. in Computer Engineering.

👨‍🏫 Assistant Professor of Computer Science and Artificial Intelligence at University of Salamanca.

Where I come from?

2. Where I come from?

University of Salamanca

Over 800 years of history.

Expert Systems and Applications Lab

Research group with around 25 researchers from different areas.

Research focused on IoT, artificial intelligence, robotics, ...

Natural Language Processing

3. Natural Language Processing

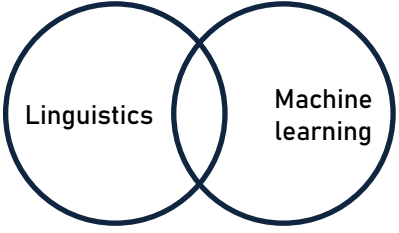
What is NLP?

3. Natural Language Processing

What is NLP?

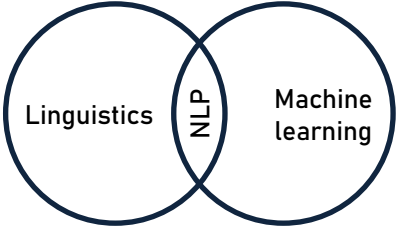
3. Natural Language Processing

What is NLP?



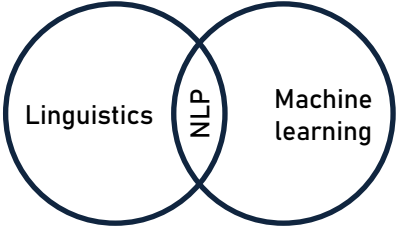
3. Natural Language Processing

What is NLP?



3. Natural Language Processing

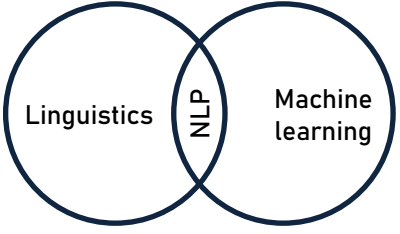
What is NLP?



Focused on learning and understanding everything related to human language.

3. Natural Language Processing

What is NLP?



Focused on learning and understanding everything related to human language.

Not only focused on understanding individual words but also on understanding the context of those words.

3. Natural Language Processing

UNIVALI

ESA LAB

NLP tasks

NLP tasks

- Text classification:

NLP tasks

- Text classification:

Input

```
1 pipe = pipeline("text-classification")
2 pipe("This talk is fantastic.")
```

3. Natural Language Processing

NLP tasks

- Text classification:

Input

```
1 pipe = pipeline("text-classification")
2 pipe("This talk is fantastic.")
```

Exit

```
1 [{'label': 'POSITIVE', 'score': 0.9998821020126343}]
```

3. Natural Language Processing

NLP tasks

3. Natural Language Processing

NLP tasks

- Fill mask:

NLP tasks

- Fill mask:

Input

```
1 pipe = pipeline("fill-mask")
2 pipe("The president of the USA is Mr. <mask>.",
  top_k=2)
```

NLP tasks

- Fill mask:

Input

```
1 pipe = pipeline("fill-mask")
2 pipe("The president of the USA is Mr. <mask>.",
  top_k=2)
```

Exit

```
1 [{'score': 0.20706287026405334, 'token': 1284, 'token_str': ' Obama',
2   'sequence': 'The president of the USA is Mr. Obama.'},
3  {'score': 0.18611690402030945, 'token': 140, 'token_str': ' Trump',
4   'sequence': 'The president of the USA is Mr. Trump.'}]
```

3. Natural Language Processing

NLP tasks

3. Natural Language Processing

NLP tasks

- Name Entity Recognition (NER):

NLP tasks

- Name Entity Recognition (NER):

Input

```
1 pipe = pipeline("ner")
2 pipe("The researcher of USAL, Diego, is coming to Brazil.")
```

3. Natural Language Processing

NLP tasks

- Name Entity Recognition (NER):

Input

```
1 pipe = pipeline("ner")
2 pipe("The researcher of USAL, Diego, is coming to Brazil.")
```

Exit

```
1 [{'entity': 'I-ORG', 'score': 0.99487865, 'index': 4, 'word': 'USA', 'start': 18,
  'end': 21},
2 {'entity': 'I-ORG', 'score': 0.9976356, 'index': 5, 'word': '##L', 'start': 21,
  'end': 22},
3 {'entity': 'I-PER', 'score': 0.9991905, 'index': 7, 'word': 'Diego', 'start': 24,
  'end': 29},
4 {'entity': 'I-LOC', 'score': 0.9995721, 'index': 12, 'word': 'Brazil', 'start': 44,
  'end': 50}]
```

3. Natural Language Processing

UNIVALI

ESA LAB

NLP tasks

NLP tasks

- Question answering:

3. Natural Language Processing

NLP tasks

- **Question answering:**

Input

```
1 pipe = pipeline("question-answering")
2 pipe(question="Who is the best professor at UNIVALI?", context="Cesar Albenes Zeferino
  is a teacher form UNIVALI, his research and teaching methods are out of this world.")
```

NLP tasks

- **Question answering:**

Input

```
pipe = pipeline("question-answering")
pipe(question="Who is the best professor at UNIVALI?", context="Cesar Albenes Zeferino
is a teacher form UNIVALI, his research and teaching methods are out of this world.")
```

Exit

```
{'score': 0.9811436533927917, 'start': 0, 'end': 22,
answer': 'Cesar Albenes Zeferino'}
```

3. Natural Language Processing

UNIVALI

ESA LAB

NLP tasks

UNIVALI

ESA LAB

NLP tasks

- Sumarization:

3. Natural Language Processing

NLP tasks

- Sumarization:

Input

```
pipe = pipeline("summarization")
pipe("Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.", min_length=5, max_length=30)
```

3. Natural Language Processing

NLP tasks

- Sumarization:

Input

```
pipe = pipeline("summarization")
pipe("Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.", min_length=5, max_length=30)
```

Exit

```
[{'summary_text': ' Natural language processing (NLP) refers to the branch of computer science concerned with giving computers the ability to understand text and spoken words in'}]
```

3. Natural Language Processing

UNIVALI

ESA LAB

NLP tasks

NLP tasks

- Text generation:

3. Natural Language Processing

NLP tasks

- Text generation:

Input

```
1 pipe = pipeline("text-generation")
2 pipe("In my last exam I was quite ")
```

3. Natural Language Processing

NLP tasks

- Text generation:

Input

```
1 pipe = pipeline("text-generation")
2 pipe("In my last exam I was quite ")
```

Exit

```
1 [{'generated_text': 'In my last exam I was quite \xa0confident about my
  decision of not getting the exam. I went with two options:- get the free
  exam and sign out the exam with a signed box or transfer from the other
  student to either a friend or'}]
```

3. Natural Language Processing

UNIVALI

ESA LAB

NLP tasks

3. Natural Language Processing

NLP tasks

- Translation:

3. Natural Language Processing

NLP tasks

- Translation:

Input

```
1 pipe = pipeline("translation_en_to_fr")
2 pipe("What is your name?")
```

3. Natural Language Processing

NLP tasks

- Translation:

Input

```
1 pipe = pipeline("translation_en_to_fr")
2 pipe("What is your name?")
```

Exit

```
1 [{'translation_text': 'Quel est votre nom?'}]
```

NLP techniques

4. NLP techniques

1950-1990

Heuristic methods: Offers short-term and immediate solutions in order to solve problems in a sub-optimal way.

- Regular expressions.
- WordNet
- Open mind questions.

4. NLP techniques

UNIVALI

ESA LAB

1950-1990

Heuristic methods: Offers short-term and immediate solutions in order to solve problems in a sub-optimal way.

- Regular expressions.
- WordNet
- Open mind questions.

1990-2010

Machine Learning: Trains a model with the data and provides a result.

- Naïve Bayes.
- Logistic Regression.
- SVMs.
- Linear Discriminant Analysis.
- HMM.

4. NLP techniques

UNIVALI

ESA LAB

1950-1990

Heuristic methods: Offers short-term and immediate solutions in order to solve problems in a sub-optimal way.

- Regular expressions.
- WordNet
- Open mind questions.

1990-2010

Machine Learning: Trains a model with the data and provides a result.

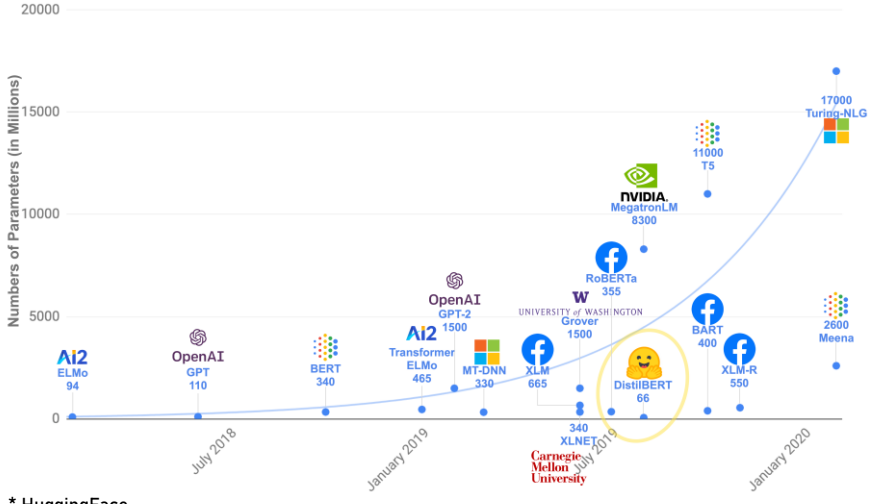
- Naïve Bayes.
- Logistic Regression.
- SVMs.
- Linear Discriminant Analysis.
- HMM.

2010-Today

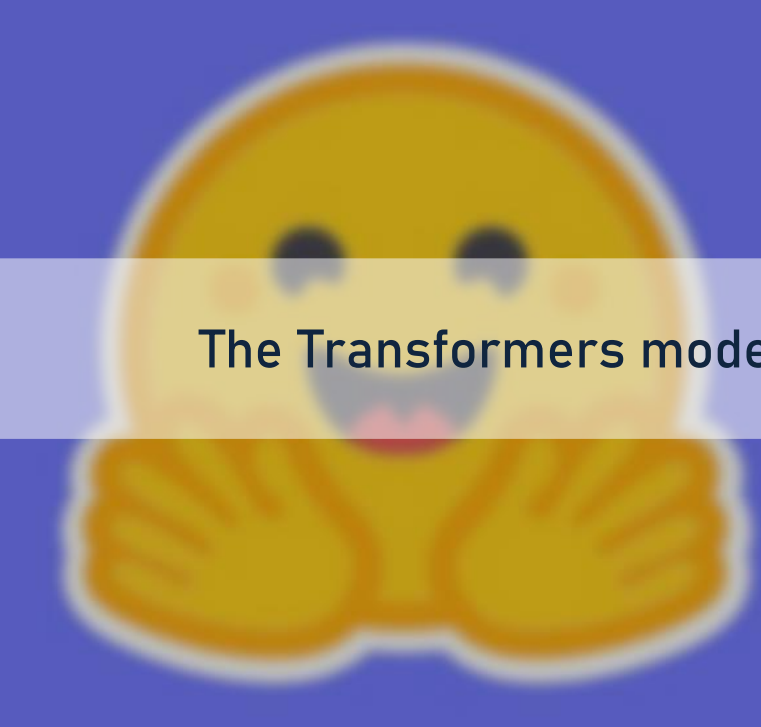
Deep Learning: Provides better results and its able to understand sequence data.

- LSTM.
- GRU.
- CNN
- Transformers.
- Autoencoder.
- …

4. NLP techniques



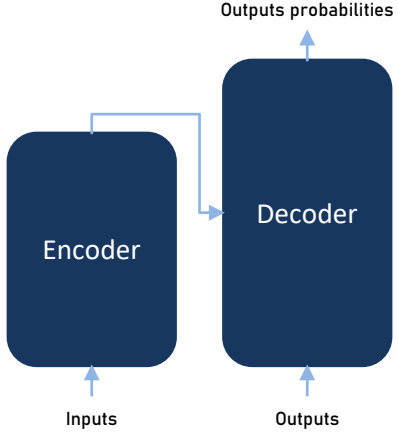
* HuggingFace.



The Transformers model

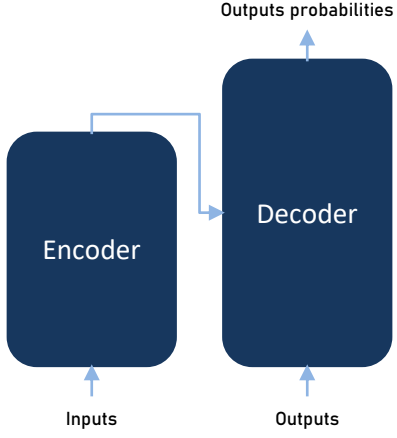
5. The Transformers model

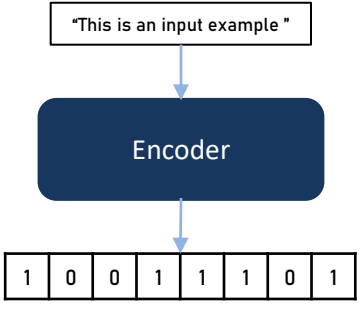
Architecture



5. The Transformers model

Architecture



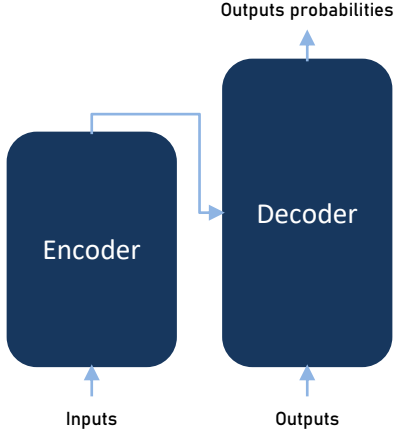


5. The Transformers model

UNIVALI

ESA LAB

Architecture

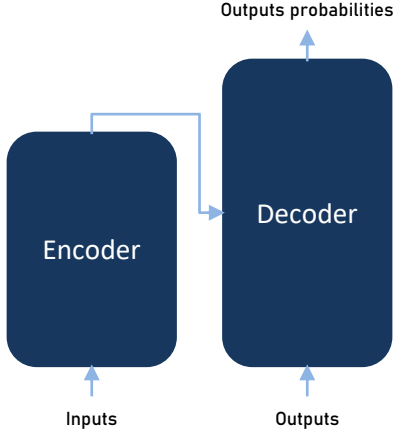


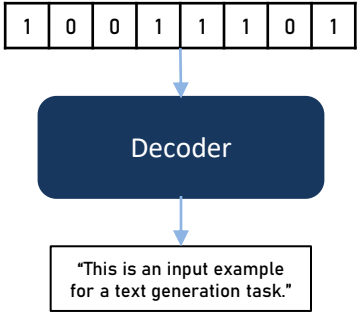
5. The Transformers model

UNIVALI

ESA LAB

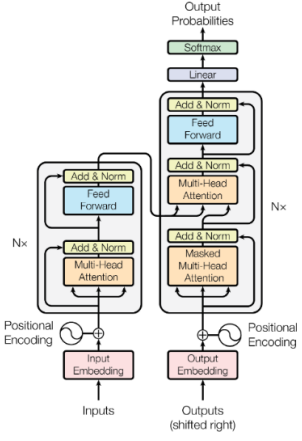
Architecture





5. The Transformers model

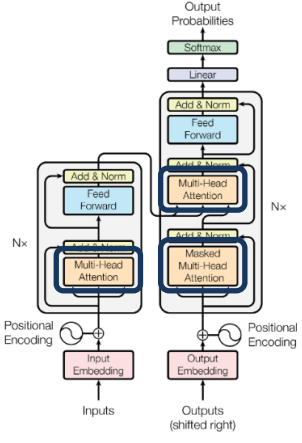
Architecture



* Vaswani et al.

5. The Transformers model

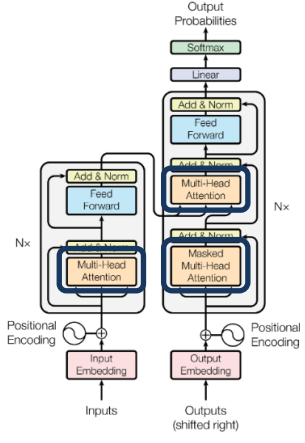
Architecture



* Vaswani et al.

5. The Transformers model

Architecture

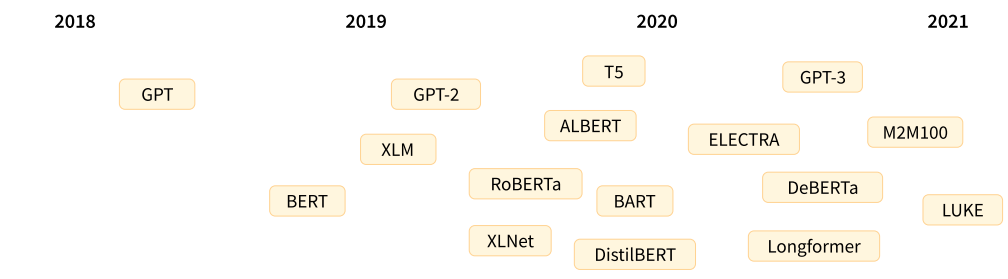


* Vaswani et al.

- Attention layers:
  - It focuses attention on certain fragments of the input, ignoring the rest, when other information in the input is being processed.
  - Decoder's attention layer use all the input information.
  - Encoder's attention layers works sequentially; so, they can only use the already processed data for prediction.

5. The Transformers model

History



* HuggingFace.

About Hugging Face

6. About Hugging Face

UNIVALI

ESA LAB

What is HF?

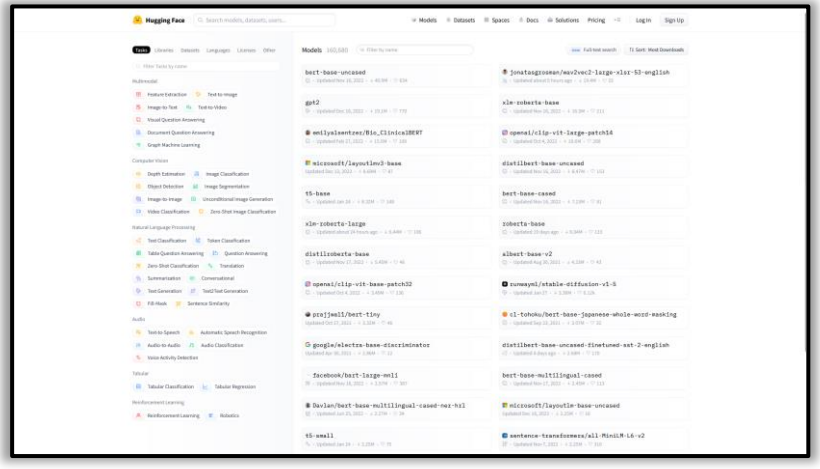
- A tech company focused on artificial intelligence and deep learning.
- A platform for deep learning.
- An open source library for deep learning.

6. About Hugging Face

UNIVALI

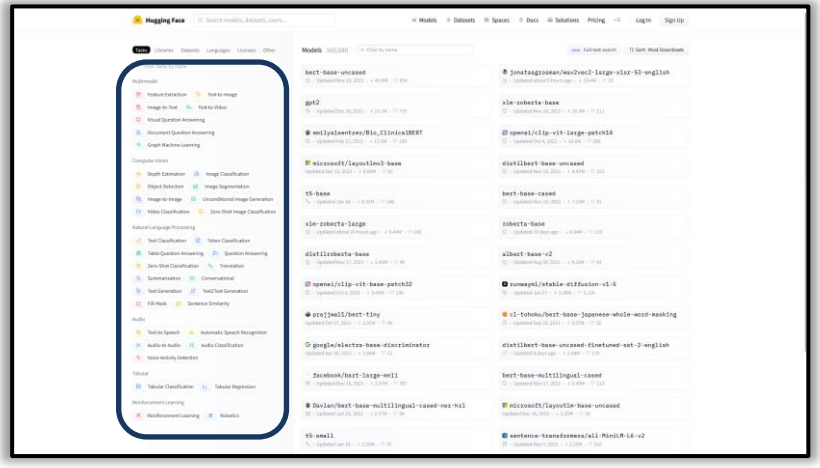
ESA LAB

HF platform



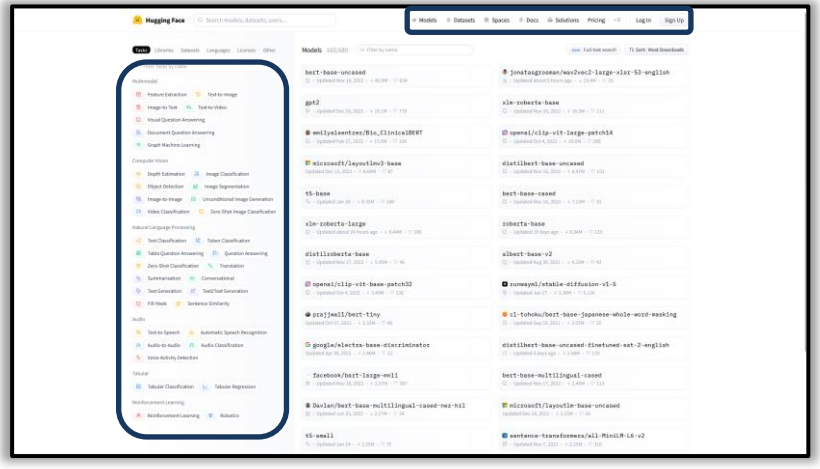
6. About Hugging Face

HF platform



6. About Hugging Face

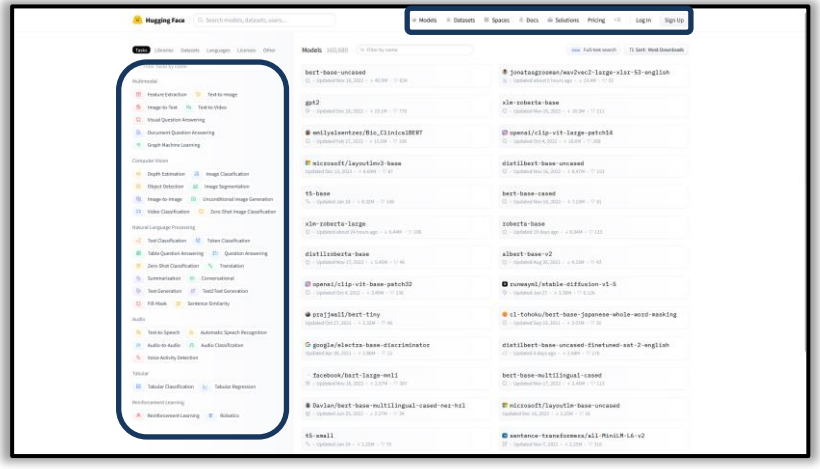
HF platform



6. About Hugging Face

HF platform

https://huggingface.co



6. About Hugging Face

HF libraries

Transformes

6. About Hugging Face

HF libraries

Transformes

Datasets

6. About Hugging Face

HF libraries

- Transformes
- Datasets
- Tokenizers

6. About Hugging Face

HF libraries

- Transformes
- Datasets
- Tokenizers
- Evaluate

6. About Hugging Face

HF libraries

- Transformes
- Datasets
- Tokenizers
- Evaluate
- Hub

6. About Hugging Face

HF libraries

- Transformes
- Datasets
- Tokenizers
- Evaluate
- Hub
- Diffusers

6. About Hugging Face

HF libraries

- Transformes
- Datasets
- Tokenizers
- Evaluate
- Hub
- Diffusers
- Gradio

6. About Hugging Face

UNIVALI

ESA LAB

HF libraries

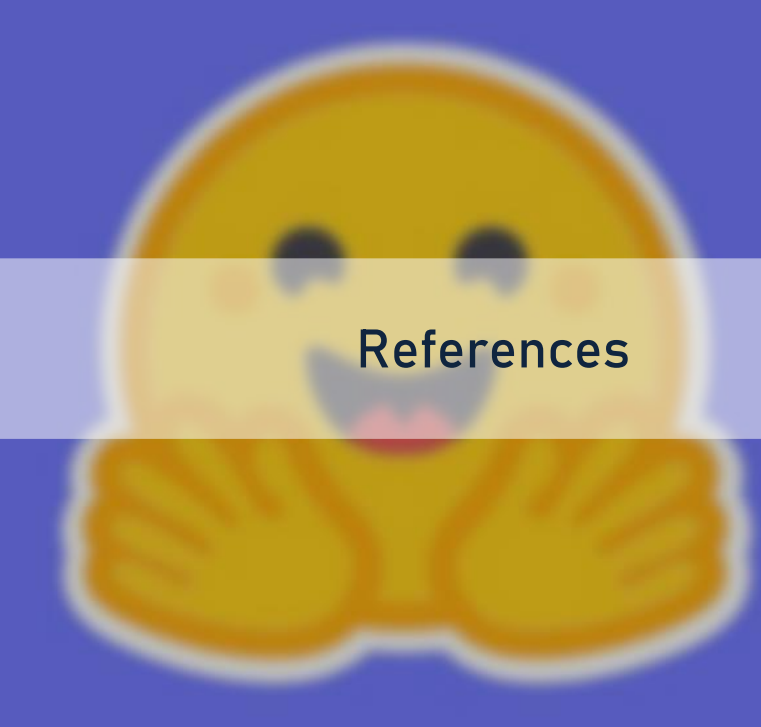
https://huggingface.co/docs

- Transformes
- Datasets
- Tokenizers
- Evaluate
- Hub
- Diffusers
- Gradio

Quick HF Tutorial

7. Quick HF tutorial

Google Colab notebook



References

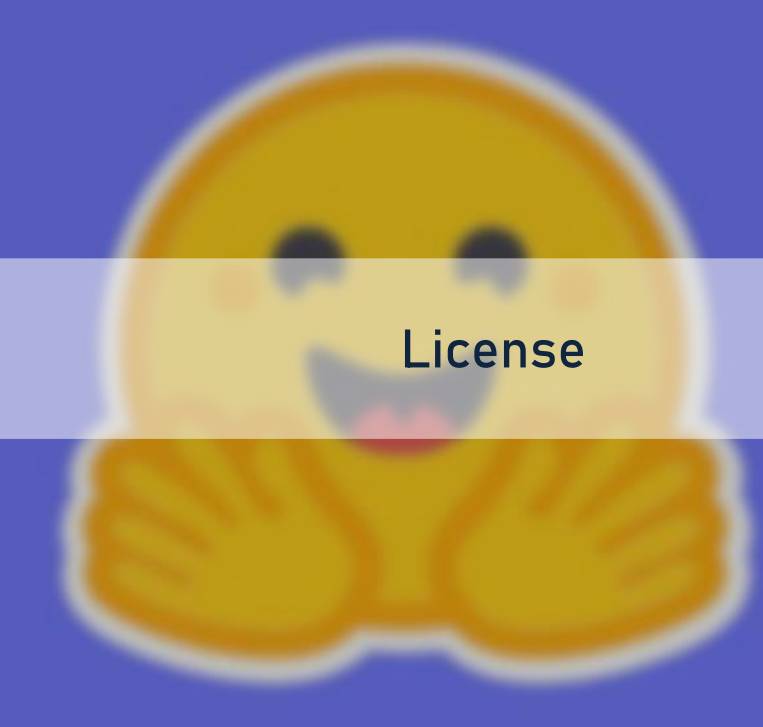
8. References

GitHub repository

NLPHuggingFaceIntroduction

8. References

- Introduction - Hugging Face Course. (s. f.). https://huggingface.co/course/chapter1/1
- Raaijmakers, S. (2022). Deep Learning for Natural Language Processing. Manning Publications.
- Rothman, D. (2022). Transformers for Natural Language Processing - Second Edition: Build, train, and fine-tune deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, and GPT-3 (2nd ed.). Packt Publishing.
- Tunstall, L., Werra, V. L., Wolf, T., & Geron, A. (2022). Natural Language Processing With Transformers: Building Language Applications With Hugging Face. Oreilly & Associates Inc.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Neural Information Processing Systems, 30, 5998-6008. https://arxiv.org/pdf/1706.03762v5



License

9. License

UNIVALI

You are free to:

copy, distribute and publicly communicate the work.

make derivative works.

Under the following conditions:

More information at https://creativecommons.org/licenses/by-nc-sa/4.0/.

Natural Language Processing: a Hugging Face introduction

Diego M. Jiménez Bravo[1]

Assistant Professor of Computer Science and Artificial Intelligence

[1] Expert Systems and Applications Lab, Department of Computer Science and Automation, University of Salamanca

dmjimenez@usal.es

@dmjimenezbravo

University of the Itajaí Valley, Itajaí, Brazil

April 4th, 2023

VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

ESA LAB

Expert Systems and Applications Laboratory