IOWA STATE UNIVERSITY

Translational AI Center

Natural Language Processing

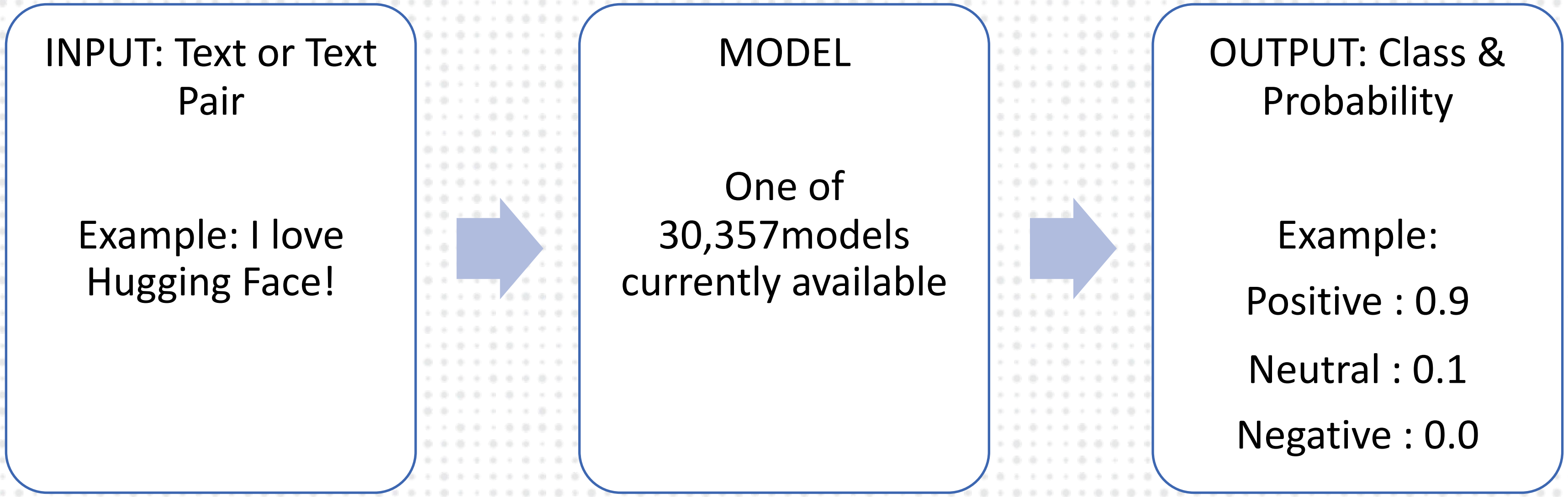
NLP TASKS

- There are a total of 11 NLP tasks per HuggingFace

1. Text classification
2. Zero shot classification
3. Token classification
4. Question & Answering
5. Table Question & Answering
6. Language translation
7. Summarization
8. Fill-Mask
9. Sentence similarity
10. Conversational
11. Text Generation

TEXT CLASSIFICATION

TEXT CLASSIFICATION



INPUT: Text or Text Pair

Example: I love Hugging Face!

MODEL

One of 30,357models currently available

OUTPUT: Class & Probability

Example:

Positive : 0.9

Neutral : 0.1

Negative : 0.0

TYPES OF TEXT CLASSIFICATION

- Sentiment Analysis
- Quora Question Pair
- Natural Language Inference (NLI)
- Multi-Genre NLI (MNLI)
- Question Natural Language Inference (QNLI)
- Grammatical Correctness

APPLICATIONS

- People Analytics
- Scientific Computing
- Marketing
- Sales

SENTIMENT ANALYSIS

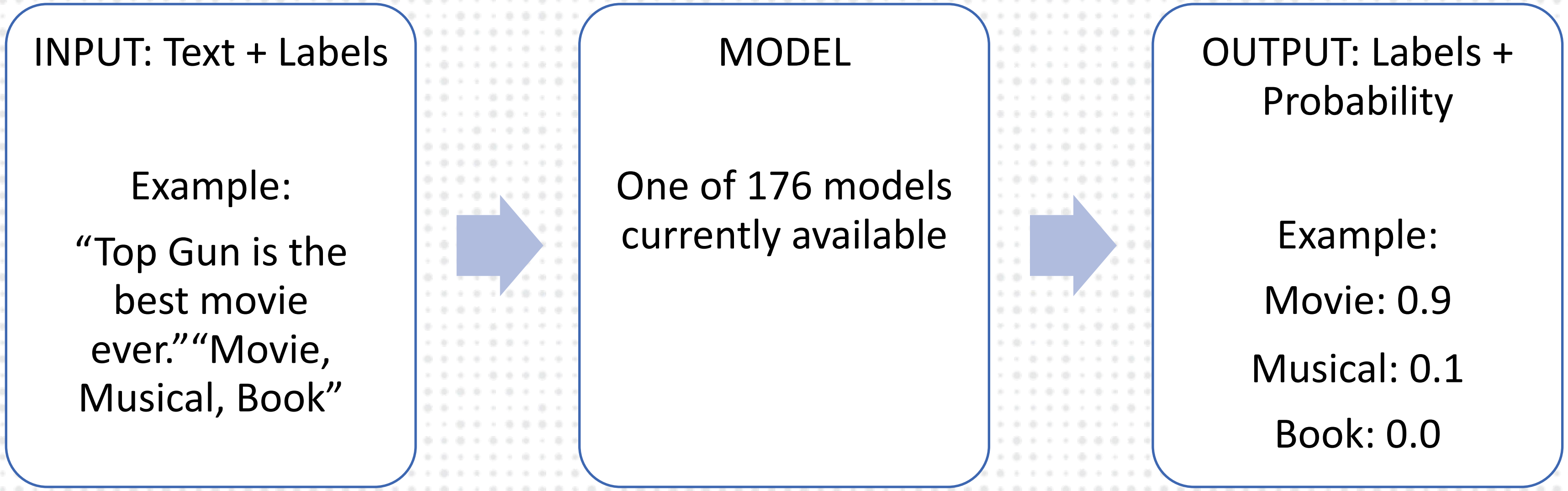
- Type of text classification task
- Analyzes text to determine associated emotion/sentiment
- Pipeline
  - Input: unstructured text data
  - Output: sentiment and probability
- Example:
  - Input: “I love coffee”
  - Output: {“positive” : 0.9, “neutral” : 0.1, “negative” : 0.0}

QUESTION NATURAL LANGUAGE INFERENCE (QNLI)

- Type of text classification task
- Designed to assess the ability of a model to determine if the answer to a question can befound in a given document.
- Pipeline
  - Input: text pair (question + context)
  - Output: entailment/not entailment
- Example:
  - Question: What percentage of marine life died during the extinction?
  - Sentence: It is also known as the "Great Dying" because it is considered the largestmass extinction in the Earth's history.
  - Label: not entailment

ZERO SHOT CLASSIFICATION

ZERO SHOT CLASSIFICATION



INPUT: Text + Labels

Example:

“Top Gun is the best movie ever.”“Movie, Musical, Book”

MODEL

One of 176 models currently available

OUTPUT: Labels + Probability

Example:

Movie: 0.9

Musical: 0.1

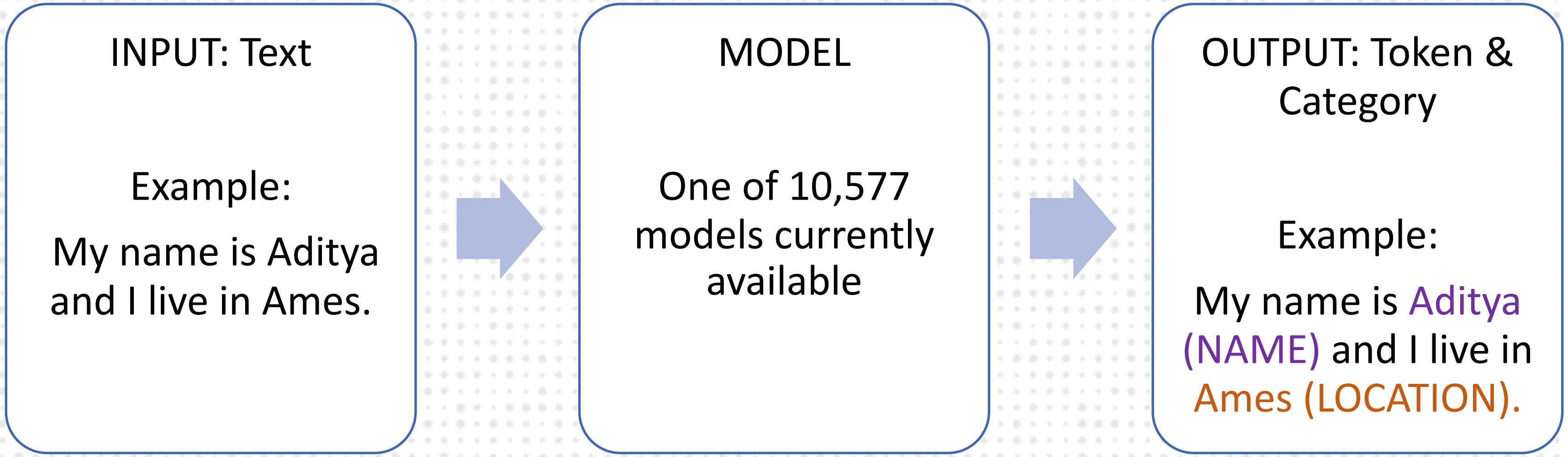
Book: 0.0

APPLICATIONS

- Multilingual sentiment analysis
- Topic classification
- Language identification

TOKEN CLASSIFICATION

TOKEN CLASSIFICATION



INPUT: Text

Example:

My name is Aditya and I live in Ames.

MODEL

One of 10,577 models currently available

OUTPUT: Token & Category

Example:

My name is Aditya (NAME) and I live in Ames (LOCATION).

APPLICATIONS

- Retail
- Oil & Gas
- Manufacturing
- Sales

TYPES OF TOKEN CLASSIFICATION

1. Named Entity Recognition (NER)
2. Part Of Speech tagging (POS)

NAMED ENTITY RECOGNITION

Named Entity Recognition (NER) - a natural language processing task that identifies namedentities in text and classifies them into predefined categories, such as person, organization,location, date, and time.

Example: The sentence "Barack Obama was born in Honolulu, Hawaii on August 4, 1961"contains the following named entities:

Barack Obama: Person
Honolulu: Location
Hawaii: Location
August 4, 1961: Date

PART OF SPEECH TAGGING

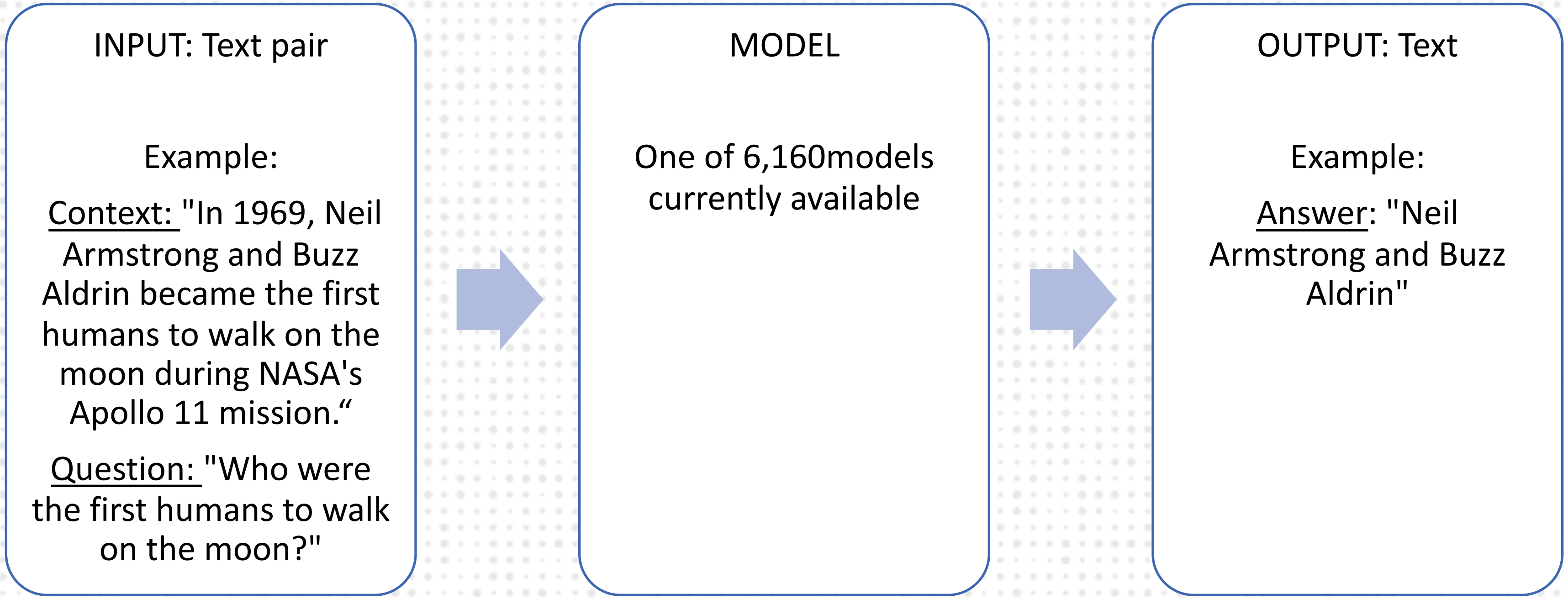
Part-of-speech (POS) tagging is a natural language processing task that assigns a part-of-speech tagto each word in a sentence. Part-of-speech tags indicate the grammatical category of a word, such asnoun, verb, adjective, adverb, preposition, conjunction, and determiner.

Example: The sentence "The cat is on the mat" contains the following POS tags:

The: Determiner
cat: Noun
is: Verb
on: Preposition
the: Determiner
mat: Noun

QUESTION ANSWERING

QUESTION ANSWERING



INPUT: Text pair

Example:

Context: "In 1969, Neil Armstrong and Buzz Aldrin became the first humans to walk on the moon during NASA's Apollo 11 mission.“

Question: "Who were the first humans to walk on the moon?"

MODEL

One of 6,160models currently available

OUTPUT: Text

Example:

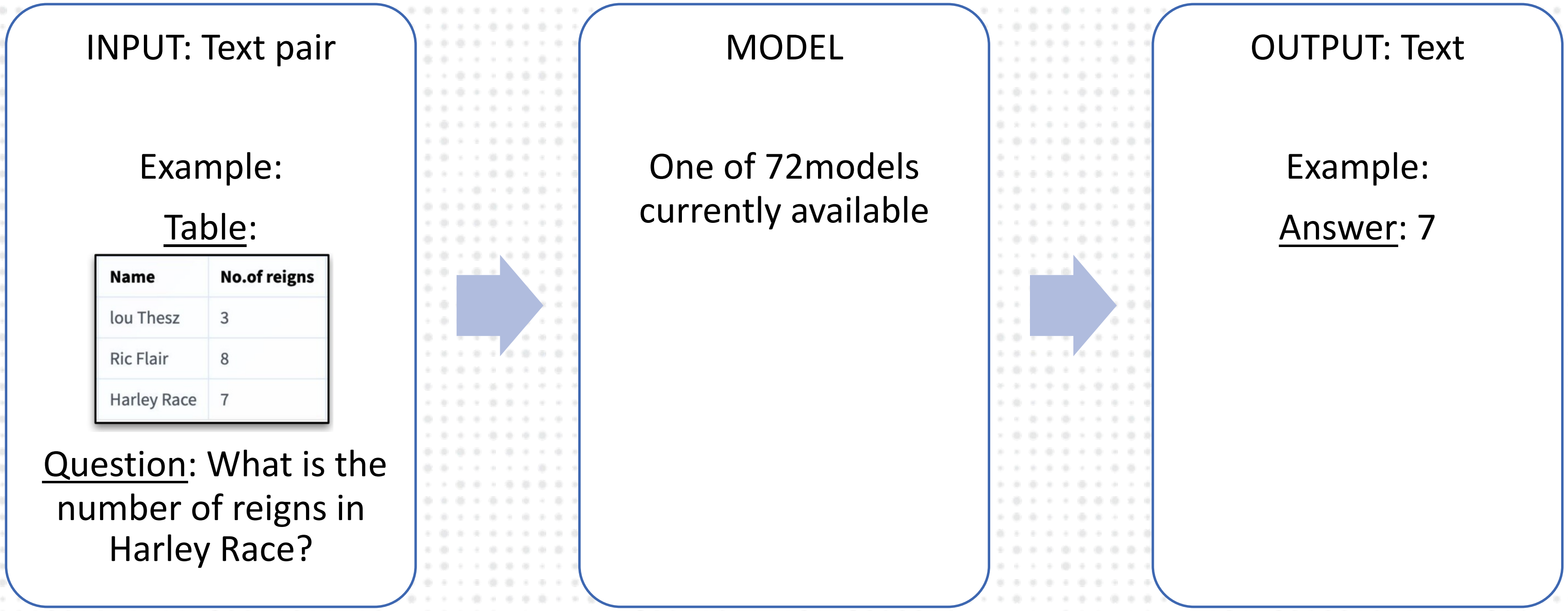
Answer: "Neil Armstrong and Buzz Aldrin"

APPLICATIONS

- Healthcare
- e-Commerce
- Manufacturing
- Finance
- Legal

TABLE QUESTION ANSWERING

TABLE QUESTION ANSWERING



INPUT: Text pair

Example:

Table:

Name	No.of reigns
lou Thesz	3
Ric Flair	8
Harley Race	7

Question: What is the number of reigns in Harley Race?

MODEL

One of 72models currently available

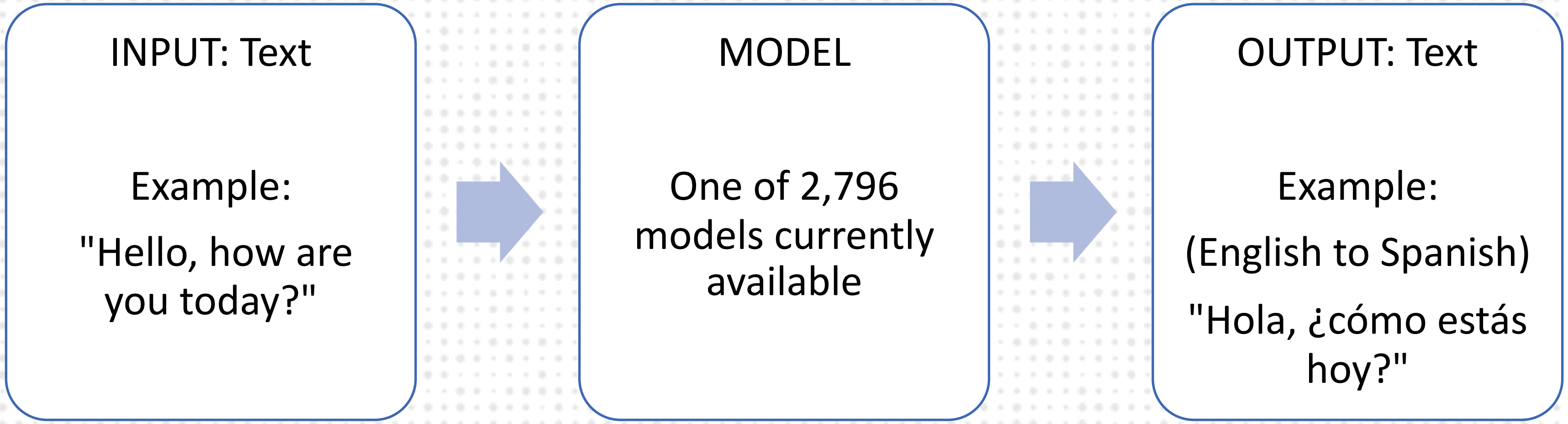
OUTPUT: Text

Example:

Answer: 7

LANGUAGE TRANSLATION

TOKEN CLASSIFICATION



INPUT: Text

Example:

"Hello, how are you today?"

MODEL

One of 2,796 models currently available

OUTPUT: Text

Example:

(English to Spanish)

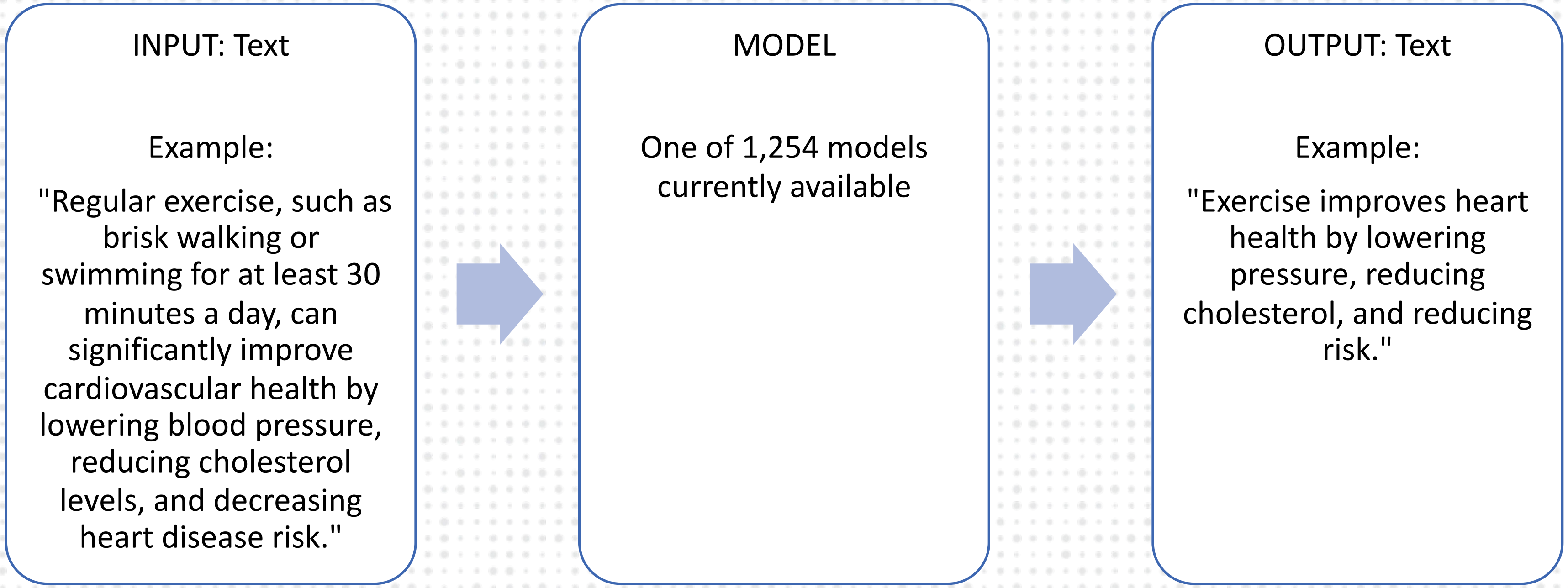
"Hola, ¿cómo estás hoy?"

APPLICATIONS

- Cross-language information retrieval
- Product and service localization
- Document analysis
- Speech translation
- Media translation

SUMMARIZATION

SUMMARIZATION



INPUT: Text

Example:

"Regular exercise, such as brisk walking or swimming for at least 30 minutes a day, can significantly improve cardiovascular health by lowering blood pressure, reducing cholesterol levels, and decreasing heart disease risk."

MODEL

One of 1,254 models currently available

OUTPUT: Text

Example:

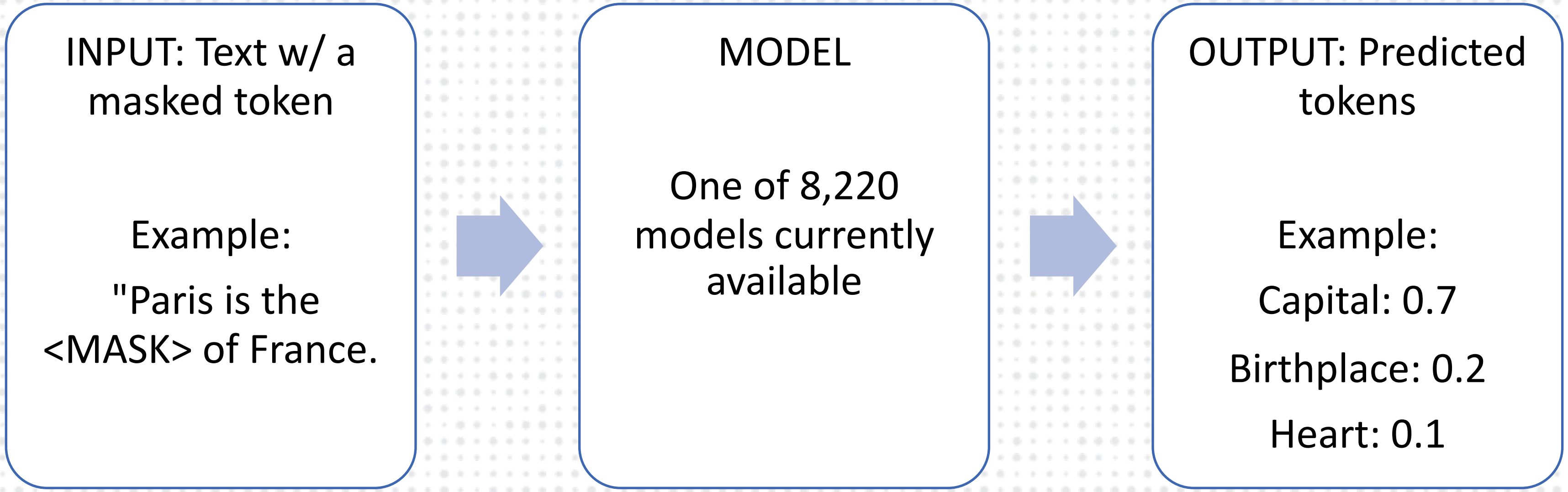
"Exercise improves heart health by lowering pressure, reducing cholesterol, and reducing risk."

APPLICATIONS

- Article summarization
- Meeting minutes
- Chatbot conversations
- Recommendation summary
- Text to reel conversion

FILL MASK

FILL MASK



INPUT: Text w/ a masked token

Example:

"Paris is the <MASK> of France.

MODEL

One of 8,220 models currently available

OUTPUT: Predicted tokens

Example:

Capital: 0.7

Birthplace: 0.2

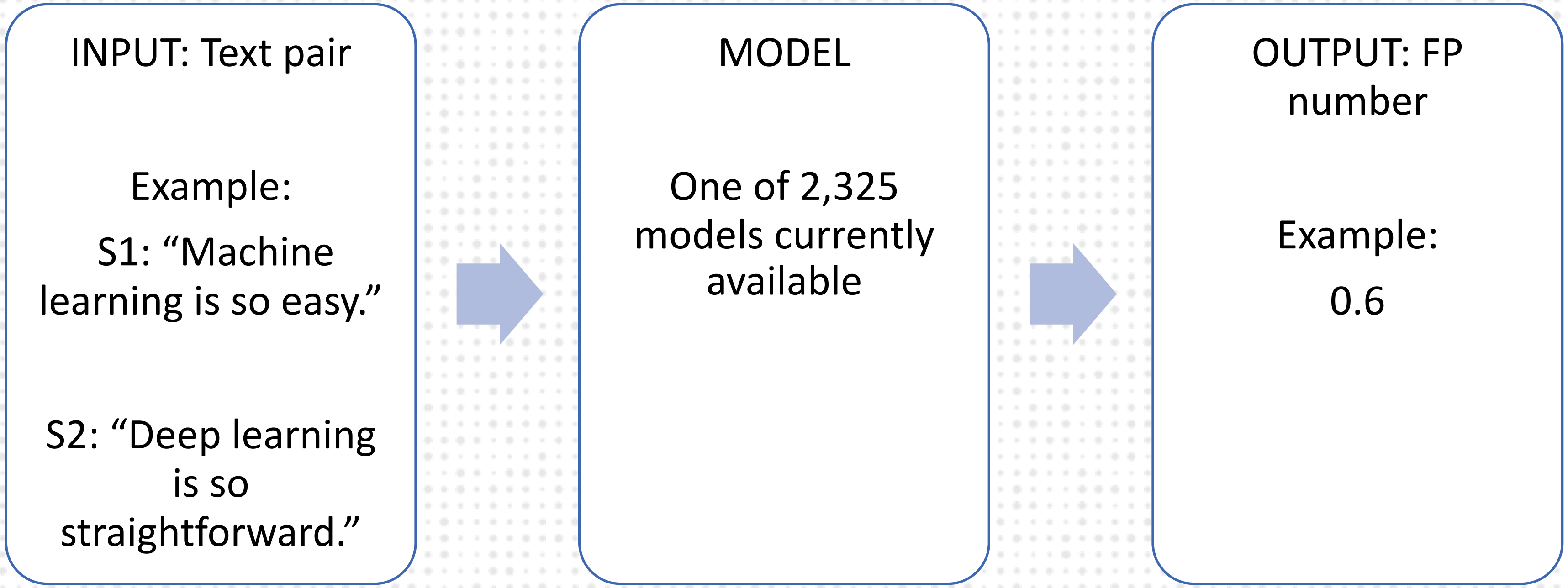
Heart: 0.1

APPLICATIONS

- Code completion in programming
- Legal document analysis
- Historical text analysis
- Creative writing

SENTENCE SIMILARITY

SENTENCE SIMILARITY



INPUT: Text pair

Example:

S1: “Machine learning is so easy.”

S2: “Deep learning is so straightforward.”

MODEL

One of 2,325 models currently available

OUTPUT: FP number

Example:

0.6

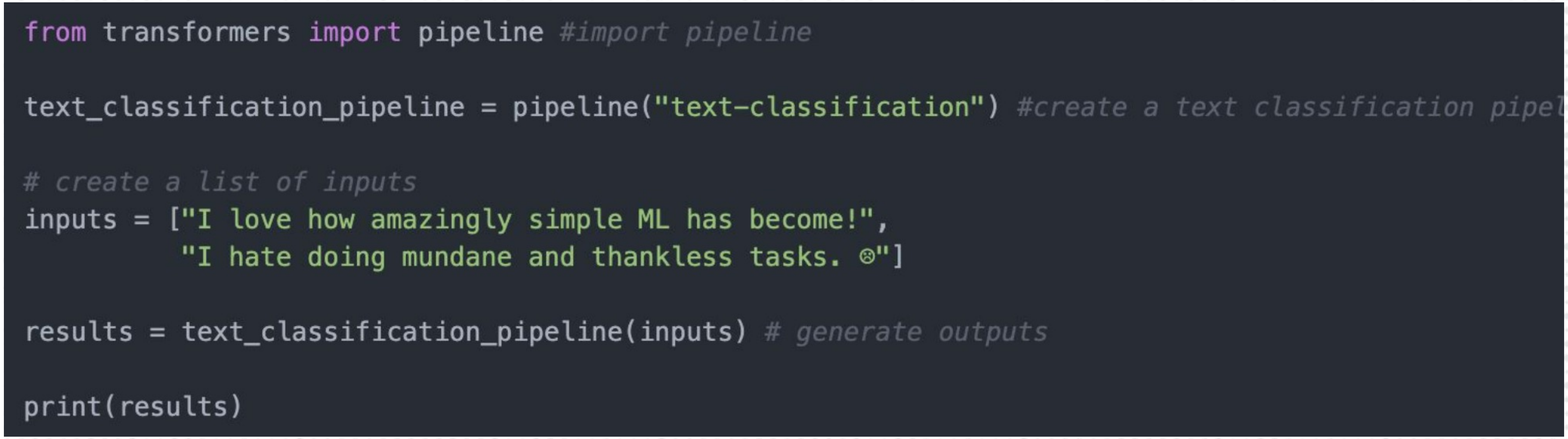
APPLICATIONS

- De-duplication
- Plagiarism and copyright violations
- Improving search engine results
- Chatbots and virtual assistants
- Recommender systems
- Automated customer support

TYPES OF INFERENCE

- Real-time inference
  - For an end-user application
  - Low-latency
  - Online processing
- Batch inference
  - For data analytics or business intelligence
  - High throughput
  - Offline processing

HUGGING FACE PIPELINE



```
from transformers import pipeline #import pipeline

text_classification_pipeline = pipeline("text-classification") #create a text classification pipel

# create a list of inputs
inputs = ["I love how amazingly simple ML has become!",
          "I hate doing mundane and thankless tasks. ☹"]

results = text_classification_pipeline(inputs) # generate outputs

print(results)
```

HUGGING FACE RESOURCES

- HUGGING FACE RESOURCES
- https://huggingface.co
- Tasks - Hugging Face
- Models - Hugging Face
- Code snippets

THANK YOU!