Developing Curricula for Artificial Intelligence and Robotics (DeCAIR)
618535-EPP-1-2020-1-JO-EPPKA2-CBHE-JP

Co-funded by the
Erasmus+ Programme
of the European Union

DeCAIR Lab Experiment Form

Author(s)	Gheith Abandah
Author Organization Name(s)	The University of Jordan
Work Package Number & Title	Work Package 8: Establishing and developing AIR labs
Activity Number & Title	Activity 8.4: Design of the lab manuals that includes the experiment that will be conducted by the students enrolled in the labs.
Work Package Leader	Sobhi Abou Shahin, Beirut Arab University

Due Date of Delivery	30/9/2023	**Project Month**	M33
Submission Date	7/8/2023	**Project Month**	M31

Revision History

Version	Date	Author	Description	Action *	Page(s)
1	7/8/2023	Gheith Abandah	Original (base) document	C	1-3
2	19/5/2024	Gheith Abandah	Update for Spring 2024	U	1-4
3					
4					

(*) Action: C = Creation, I = Insert, U = Update, R = Replace, D = Delete

Disclaimer

Email: DeCAIR@ju.edu.jo

Project Website: http://DeCAIR.ju.edu.jo/

Developing Curricula for Artificial Intelligence and Robotics (DeCAIR)
618535-EPP-1-2020-1-JO-EPPKA2-CBHE-JP

Co-funded by the
Erasmus+ Programme
of the European Union

Course Title	Natural Languages Processing
Course Number	0907753
Experiment Number	2
Experiment Name	Hugging Face's Transformers Library
Objectives	The students experiment with Hugging Face's Transformers Library.
Introduction	This is an advanced experiment in NLP. The student solves five exercises to practice some advanced skills in NLP.
Materials	Computer with Python integrated development environment (IDE) software installed (PyCharm is recommended), or Jupyter Notebook (Google Colab is recommended). Dataset files: None
Procedure	## Exercise 1: Question Answering The following code specifies a context and a question. Complete this code to answer the question using the given context using the *DistilBERT base uncased distilled SQuAD* model.

```python
from transformers import pipeline

# Initialize the question answering pipeline with
#   the distilbert model
qa_pipeline = pipeline("question-answering",
                model="distilbert-base-uncased-distilled-squad")

# Define the context and the question
context = """
Hugging Face is a technology company based in New York and Paris.
They are known for their Transformers library, which is widely
used in natural language processing.
"""
question = "Where is Hugging Face based?"
```

Exercise 2: Text Summarization
The following code specifies a text. Complete this code to summarize this text using the T5 Small model.

```python
from transformers import pipeline

# Initialize the summarization pipeline with the t5-small model
summarizer = pipeline("summarization", model="t5-small")
```

Developing Curricula for Artificial Intelligence and Robotics (DeCAIR)
618535-EPP-1-2020-1-JO-EPPKA2-CBHE-JP

Co-funded by the
Erasmus+ Programme
of the European Union

```
# Define the text to summarize
text = """
Hugging Face is a technology company based in New York and Paris.
They are known for their Transformers library, which is widely
used in natural language processing.
The Transformers library provides thousands of pre-trained models
to perform tasks on texts such as classification, information
extraction, question answering, summarization, translation, and
text generation in over 100 languages.
Hugging Face has democratized the use of state-of-the-art machine
learning models by making them accessible to researchers and
practitioners around the world.
"""
```

Exercise 3: Text Classification

The following code specifies text samples. Complete this code to classify these samples using the `distilbert-base-uncased-finetuned-sst-2-english` model.

```python
from transformers import pipeline

# Initialize the text classification pipeline with the bert-base-
uncased model
classifier = pipeline("text-classification")

# Example texts for classification
texts = [
    "I love this product! It's amazing.",
    "This is the worst movie I have ever seen.",
    "The restaurant was okay, not great but not terrible either."
]
```

Exercise 4: Text Classification/Manual

Solve the above exercise using the `TFAutoModelForSequenceClassification` and `AutoTokenizer` classes imported below.

```python
import tensorflow as tf
from transformers import AutoTokenizer
from transformers import TFAutoModelForSequenceClassification
```

Developing Curricula for Artificial Intelligence and Robotics (DeCAIR)
618535-EPP-1-2020-1-JO-EPPKA2-CBHE-JP

Co-funded by the
Erasmus+ Programme
of the European Union

	Exercise 5: Generating Shakespearean Text Use the Hugging Face Transformers library to download a pretrained language model capable of generating text (e.g., GPT), and try generating more convincing Shakespearean text. You will need to use the model's `generate()` method—see Hugging Face's documentation for more details.
Data Collection	Capture the output of your code for the above five exercises.
Data Analysis	None
Required Reporting	Submit your code and the captured output.
Safety Considerations	Standard safety precautions related to using computer.
References	1. Course slides available on https://www.abandah.com/gheith/ 2. H. Lane, C. Howard, and H. Hapke, Natural Language Processing in Action: Understanding, analyzing, and generating text with Python, Manning, 2019. 3. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow, O'Reilly, 3rd Edition, 2022.