

## Topic 2: Open Source LLMs & Local Setup

### Assignment 1: Hugging Face Model Exploration

- **Objective:** Explore an open-source LLM on Hugging Face Hub.
- **Instructions:**
  1. Create a Python virtual environment and install `transformers` and `torch`.
  2. Pick a small open-source model (like `distilbert-base-uncased`).
  3. Load the model and tokenizer in Python.
  4. Use it to perform a simple task (e.g., text classification, summarization).
- **Deliverables:** Python script with the following:
  - Environment setup commands
  - Model loading code
  - Sample output from test text

#### Environmental Setup commands:

#### In Powershell:

#### Step 1: Open PowerShell as Administrator

- Click Start Menu → type "PowerShell" → Right click → Run as Administrator

#### Step 2: Change Execution Policy (Safe Way)

Run this command:

```
Set-ExecutionPolicy RemoteSigned -Scope CurrentUser
```

- RemoteSigned = lets you run **local scripts** (like `activate.ps1`) without restrictions, but still protects against untrusted scripts from the internet.
- -Scope CurrentUser = only applies to **your user account** (not system-wide, safe to use).

Press **Y** (Yes) when prompted.

#### Step 3: Verify the Change

Run:

```
Get-ExecutionPolicy -Scope CurrentUser
```

It should show:

RemoteSigned

## Step 4: Restart VS Code

Now, open VS Code normally. When you create or activate a virtual environment, just do:

```
.\hf_env\Scripts\activate
```

---

## Inside VS Code

### 1. Create Virtual Environment

In VS Code terminal (PowerShell or CMD), run:

```
python -m venv hf_env
```

### 2. Activate Environment

On Windows:

```
.\hf_env\Scripts\activate
```

### 3. Install Required Libraries

```
pip install torch transformers
```

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification,  
pipeline
```

### 4. Run Your Python Script

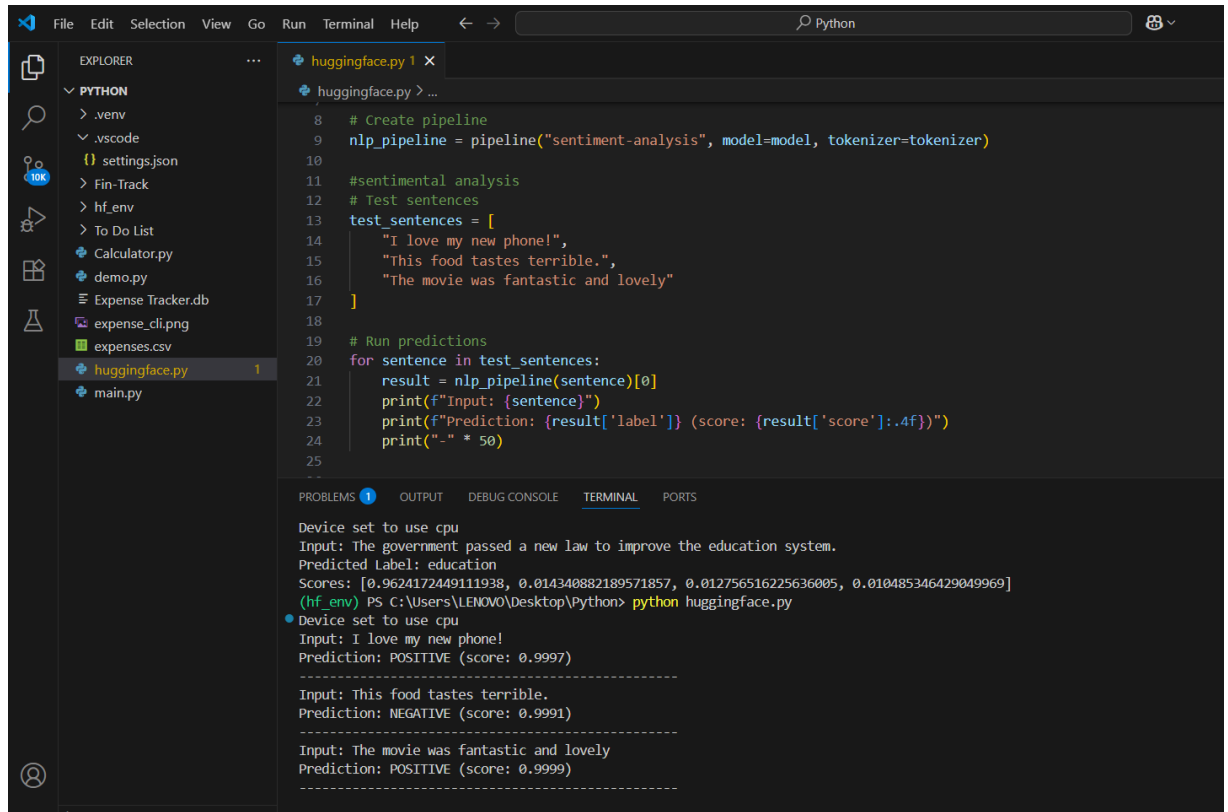
- Create a file, e.g., `huggingface.py`
- Add Hugging Face code (sentiment, summarization, classification).
- Run it in terminal:

```
python huggingface.py
```

## Output Screenshots:

### 1. Sentimental analysis

Model used : distilbert-base-uncased-finetuned-sst-2-english



```
File Edit Selection View Go Run Terminal Help
huggingface.py x
huggingface.py > ...
8 # Create pipeline
9 nlp_pipeline = pipeline("sentiment-analysis", model=model, tokenizer=tokenizer)
10
11 #sentimental analysis
12 # Test sentences
13 test_sentences = [
14     "I love my new phone!",
15     "This food tastes terrible.",
16     "The movie was fantastic and lovely"
17 ]
18
19 # Run predictions
20 for sentence in test_sentences:
21     result = nlp_pipeline(sentence)[0]
22     print(f"Input: {sentence}")
23     print(f"Prediction: {result['label']} (score: {result['score']:.4f})")
24     print("-" * 50)
25
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

Device set to use cpu  
Input: The government passed a new law to improve the education system.  
Predicted Label: education  
Scores: [0.9624172449111938, 0.014340882189571857, 0.012756516225636005, 0.010485346429049969]  
(hf\_env) PS C:\Users\LENOVO\Desktop\Python> python huggingface.py

Device set to use cpu  
Input: I love my new phone!  
Prediction: POSITIVE (score: 0.9997)  
-----  
Input: This food tastes terrible.  
Prediction: NEGATIVE (score: 0.9991)  
-----  
Input: The movie was fantastic and lovely  
Prediction: POSITIVE (score: 0.9999)  
-----

## 2. Text Summarization

Model used: sshleifer/distilbart-cnn-6-6

```
Run Terminal Help < -> Python
huggingface.py 1 x
huggingface.py > ...
26
27
28 #text summarization
29 print("\n=== Text Summarization ===")
30 summarizer = pipeline("summarization", model="sshleifer/distilbart-cnn-6-6")
31
32 long_text = """
33 Artificial Intelligence (AI) is a branch of computer science that aims to create machines that
34 can perform tasks that normally require human intelligence. These tasks include learning,
35 reasoning, problem-solving, perception, and language understanding. AI is widely used in
36 applications such as self-driving cars, virtual assistants, recommendation systems, and medical diagnosis.
37 """
38
39 summary = summarizer(long_text, max_length=50, min_length=25, do_sample=False)
40 print("Original Text:", long_text)
41 print("\nSummary:", summary[0]['summary_text'])
42
43
44
```

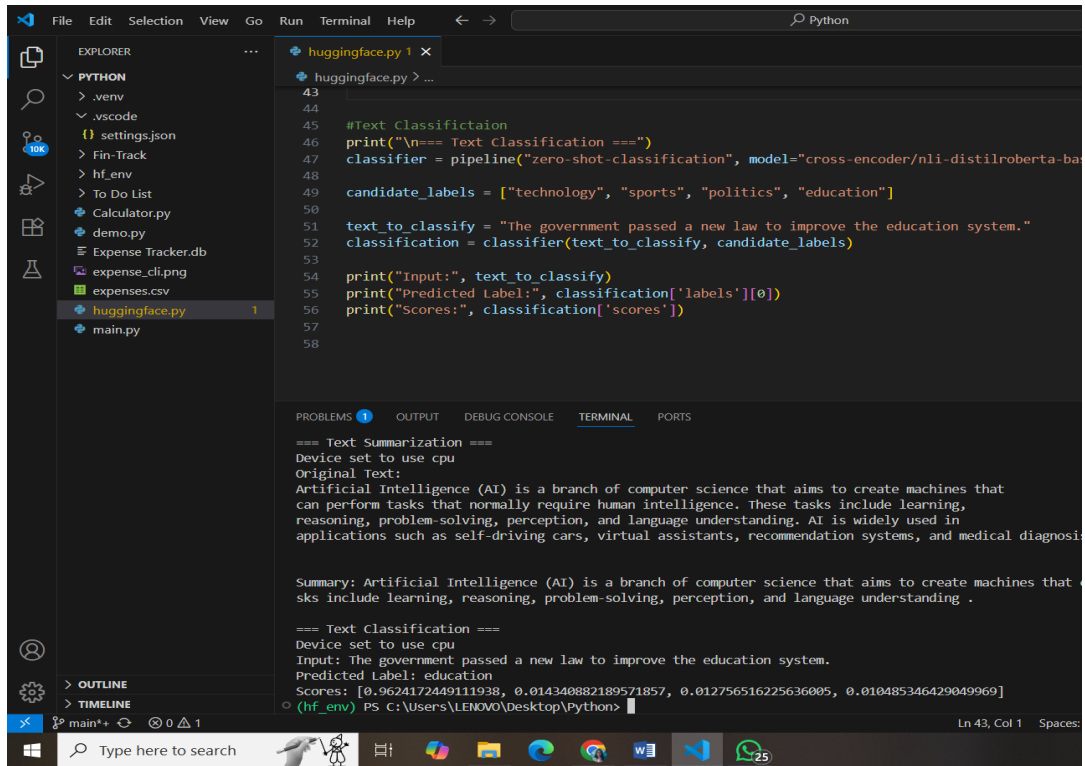
PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS powershell +v

```
==== Text Summarization ====
Device set to use cpu
Original Text:
Artificial Intelligence (AI) is a branch of computer science that aims to create machines that
can perform tasks that normally require human intelligence. These tasks include learning,
reasoning, problem-solving, perception, and language understanding. AI is widely used in
applications such as self-driving cars, virtual assistants, recommendation systems, and medical diagnosis.

Summary: Artificial Intelligence (AI) is a branch of computer science that aims to create machines that can perform tasks that normally require human intelligence . These ta
sks include learning, reasoning, problem-solving, perception, and language understanding .
```

### 3. Text Classification

Model used: cross-encoder/nli-distilroberta-base



The screenshot shows a Visual Studio Code editor with a Python file named `huggingface.py` open. The file contains a script for text classification using the `cross-encoder/nli-distilroberta-base` model. The script defines a list of candidate labels, a text to classify, and a classification function. The output of the script is displayed in the terminal window at the bottom.

```
43
44
45 #Text Classificaion
46 print("\n=== Text Classification ===")
47 classifier = pipeline("zero-shot-classification", model="cross-encoder/nli-distilroberta-base")
48
49 candidate_labels = ["technology", "sports", "politics", "education"]
50
51 text_to_classify = "The government passed a new law to improve the education system."
52 classification = classifier(text_to_classify, candidate_labels)
53
54 print("Input:", text_to_classify)
55 print("Predicted Label:", classification['labels'][0])
56 print("Scores:", classification['scores'])
57
58
```

The terminal output shows the results of the text classification:

```
=== Text Summarization ===
Device set to use cpu
Original Text:
Artificial Intelligence (AI) is a branch of computer science that aims to create machines that
can perform tasks that normally require human intelligence. These tasks include learning,
reasoning, problem-solving, perception, and language understanding. AI is widely used in
applications such as self-driving cars, virtual assistants, recommendation systems, and medical diagnosis.

Summary: Artificial Intelligence (AI) is a branch of computer science that aims to create machines that
sks include learning, reasoning, problem-solving, perception, and language understanding .

=== Text Classification ===
Device set to use cpu
Input: The government passed a new law to improve the education system.
Predicted Label: education
Scores: [0.9624172449111938, 0.014340882189571857, 0.012756516225636005, 0.010485346429049969]
```

