# Using OpenAI Moderation API

## Exercise Overview:

In this exercise, we will use OpenAI's Moderation API to analyze text and retrieve category scores that indicate whether a given input violates predefined moderation categories.

## Instructions:

1. Create an OpenAI client instance using `OpenAI(api\_key="<OPENAI\_API\_TOKEN>")`.

2. Send a request to the Moderation endpoint using the `text-moderation-latest` model.

3. Analyze the moderation category scores from the response.

4. Print the moderation results.

## Corrected Python Code:

from openai import OpenAI  
  
# Initialize OpenAI Client  
client = OpenAI(api\_key="<OPENAI\_API\_TOKEN>")  
  
# Create a request to the Moderation endpoint  
response = client.moderations.create(  
 model="text-moderation-latest",  
 input="My favorite book is To Kill a Mockingbird."  
)  
  
# Print the category scores  
print(response.results[0].category\_scores)

## Explanation of the Fix:

1. \*\*Initializing OpenAI Client:\*\* The `OpenAI` client is initialized using a valid API key.  
2. \*\*Sending a Moderation Request:\*\* The input text is sent to OpenAI's moderation model (`text-moderation-latest`).  
3. \*\*Retrieving the Response:\*\* The API returns moderation results, including category scores that indicate potential violations.  
4. \*\*Printing the Scores:\*\* The relevant category scores are extracted and printed for analysis.