

AI Engineer Assignment

- Please submit your completed answer within **21 days** of receiving this assignment.
- Use **Gemini API** instead of OpenAI API

Source:

<https://aistudio.google.com/>

https://ai.google.dev/pricing#1_5flash

The assignment includes two questions:

Question 1: LLM integration (Score: 30%)

The task involves building an AI capable of language translation.

1.1 Single Text Translation: (Score: 15%)

You are asked to write a Python code using the OpenAI API to translate a given text into Vietnamese. (You should check the text if it's already the destination language).

For example, translating "Hello" into Vietnamese should return "Xin chào", but "Xin chào" should return the same.

1.2 Multiple Texts Translation: (Score: 15%)

Similar to 2.1, but the input is a list of texts. The Python code should accept a list of strings and return their translations in the specified language. For instance, translating ["Hello", "I am John", "Tôi là sinh viên"] into Vietnamese should return ["Xin chào", "Tôi tên là John", "Tôi là sinh viên"].

Question 2: Chatbot Development (Score: 70%)

Assignment Test: Chatbot Development from Website Data. The data is at

<https://www.presight.io/privacy-policy.html>

2.1 Data Access and Indexing (Score: 40%)

Tasked with creating a chatbot, begin by web crawling the specified website to gather relevant data, then preprocess and structure this data into a searchable index, ready for query retrieval.

(Short version: crawling then embedding data, you can use **selenium** or **requests**)

2.2 Chatbot Development (Score: 30%)

Develop a chatbot that employs natural language processing to comprehend user questions, searches the indexed data from 2.1 for the best match, and delivers the most accurate response drawn from the website's information.

(Use any distance/similarity metrics to get the best match paragraph then feed to LLM to get answer)

Deliverables

Please submit your response in PDF format. The solution should answer all questions.