ASSIGNMENT

**__LLM APIs__

***Use Hugging Face APIs to integrate LLMs in applications.

1) Call Hugging face Inference APIs .

2) Build a proper question-answering interface.

3) Include exception handling

**__LLM and temperature tuning:__

1. Build a Python function `generate_story(prompt, temperature=0.7)` using:
    - `GPT2Tokenizer`
    - `GPT2LMHeadModel`
2. The function should:
    - Accept a text prompt and a temperature value
    - Generate a continuation using `generate()`
    - Include `do_sample=True`, `top_k`, `top_p`, and `repetition_penalty=1.2`
    - Return or print the generated story

3. Set pad token to avoid warnings:

```
pad_token_id=tokenizer.eos_token_id
```

4. Add exception handling in case of errors
5. Create test cases:

    - Use the same prompt with different temperature values: `0.3`, `0.7`, `1.0`, `1.3`

    - Observe and record how the generation changes

6. Mention comment lines properly and clearly