# Questão 4: Com base na implementação da Questão 2 (Parte 2), explique as principais limitações de utilizar LangChain para integrar a API da OpenAI Gemini.

- *Latência de Resposta:* Devido aos grandes servidores, as requisições não demoram para acontecer. O problema seria se o usuário possuir uma rede fraca.
- *Limites de Uso da Gemini:* O limite de tokens é 1000000 na versão grátis.
- *Desafios de Escalabilidade e Custo:* Aplicações com um grande número de usuários simultâneos enfrentam desafios devido a limites de requisição da API. Além disso, o custo da API aumenta proporcionalmente ao número de tokens processados, tanto no prompt quanto na resposta.
- *Qualidade das traduções geradas em comparação com outros modelos:* O modelo é extremamente rápido e versátil, conseguindo entender e trazer uma resposta condizente.

# Questão 5: Com base na aplicação desenvolvida na 3 (Parte 2), explique as limitações de usar LangChain para integrar o modelo HuggingFace de tradução.

- *Desempenho e tempo de resposta:* O tempo de resposta é ok. A única demora é para baixar o modelo.
- *Consumo de recursos computacionais:* Utilizando a biblioteca pytorch ou tensorflow, voce consegue regular os recursos computacionais. Mas, no geral, se a máquina for boa não haverá muitos problemas.
- *Possíveis limitações no ajuste fino do modelo:* O modelo já vem especializado em tradução, não tendo a necessitade de fine tuning.
- *Comparação com o uso direto da API HuggingFace:* A API HuggingFace oferece um método rápido e eficiente para acessar modelos prontos sem a necessidade de configurar.