UNIVERSIDADE NOVE DE JULHO - UNINOVE
PROJETO EM SISTEMAS INTELIGENTES

EXCLUÍDOS OS DADOS SOBRE OS AUTORES EM ATENDIMENTO A LGPD - LEI GERAL DE PROTEÇÃO DE DADOS

GERANDO PERGUNTAS COM LARGE LANGUAGE MODEL

São Paulo
2024

EXCLUÍDOS OS DADOS SOBRE OS AUTORES EM ATENDIMENTO A LGPD - LEI GERAL DE PROTEÇÃO DE DADOS

GERANDO PERGUNTAS COM LARGE LANGUAGE MODEL

Projeto apresentado a Universidade Nove de Julho - UNINOVE, como parte dos requisitos obrigatórios para obtenção do título de BACHAREL CIÊNCIA DA COMPUTAÇãO.

Prof. Orientador: Edson Melo de Souza, Dr.

São Paulo
2024

RESUMO

Contexto:Muitas vezes as intelgências artificiais baseadas em linguagem são tratadas como programas que funcionam como assistentes pessoais que respondem coisas, porém, essa é uma noção errônea e superficial, apenas a ponta de um iceberg formado por uma arquitetura de redes neurais aplicadas em toneladas de texto. Primordialmente seu maior triunfo é a capacidade de analisar um texto, consumir informacão nao estruturada. Pensando nisso Como uma pessoa que consome muito conteudo didático pelo youtube, um método que uso para absorver o conteudo é sempre me fazer perguntas sobre o que eu assisti (ou ouvi).Entao como produto desse projeto busquei desenvolver uma aplicação que gere perguntas de um determinado video, ou texto. **Objetivo**: Gerar uma API web e um front-end que aplicasse uma LLM para gerar perguntas dado um texto ou link para um video do yotube. **Resultados**:Como resultado tivemos uma API web que gera perguntas dada um determinado texto ou link de um video do youtube. Sendo aplicado um simples front end apenas com intuito de demostrar um exemplo de aplicao planejada. **Conclusão**: Como comun de um modelo baseado em redes neurais, sua acertividade é muito intensa (ou muito boa, ou muito ruim),tendo as qualidades das perguntas uma hetereogenidedade pelo idioma, script ou fatores do próprio youtube.

Palavras-chave: Inteligência Artificial, Conteúdo Educacional,Ferramentas Educacionais, Redes Neurais.

ABSTRACT

Contextualization:Artificial intelligence systems based on language are often treated as programs functioning as personal assistants that respond to queries. However, this is an erroneous and superficial notion, merely the tip of an iceberg formed by a neural network architecture applied to tons of text. Primarily, their greatest triumph is the ability to analyze a text and consume unstructured information. As someone who consumes a lot of educational content on YouTube, a method I use to absorb the content is to always ask myself questions about what I watched (or heard). Therefore, as the product of this project, I sought to develop an application that generates questions from a given video or text. **objective**: To create a web API and a front-end application that utilizes a large language model (LLM) to generate questions given a text or a link to a YouTube video. **Results**:As a result, we developed a web API that generates questions from a given text or a YouTube video link. A simple front-end was applied with the sole purpose of demonstrating an example of the planned application. **Conclusion**: As is common with a neural network-based model, its accuracy can be quite variable (either very good or very bad), with the quality of the questions displaying heterogeneity due to factors such as language, script, or the specific characteristics of YouTube content.

Keywords: Artificial Intelligence, Educational Content, Large Language Model, Educational Tools.

SUMÁRIO

1 INTRODUÇÃO

1.1 O que é são LLMs ?

1.1.1 Definição

LLMs é a singla em inglês para Large Language Model(grandes modelos de linguagem),que se tratam de um modelo de inteligência artificial treinado com aprendizado profundo(deep learning) supervisionado com uma grande quantidade de textos.

1.1.2 Aplicações

esses modelos são aplicados para analisar e gerar texto de forma convicente em aplicações como: atendimento ao cliente, mecanismo de busca, redação e geração de texto numa forma geral; e no ambito da educação tem um potencial imenso como para a sitesse de textos, Gerar resumos e no nosso caso criar pergunta a partir de analises prévias.

1.2 Objetivo

O objetivo desse trabalho é criar uma aplicação que utiliza um modelo de LLM para gerar perguntas de textos e videos do youtube, que tem entre seus conteúdos uma grande quantidade de videos informativos tais como: aulas, documentarios, podcasts e audiolivros, aos quais pode ser aplicado para gerar perguntas que ajudam a fixar o aprendizado e praticar o conteudo consumido.

1.3 O que é youtube ?

youtube é uma plataforma de compartilhamento de video fundada em fevereiro de 2005,onde o usuario, tanto pessoa fisica ou juridica pode hospedar videos de caracter pessoal, musical, educacional, televisivo e etc...

2 FUNDAMENTAÇÃO TEÓRICA

Resumo do capítulo

O trabalho aprensentado é para aplicação prática de conhecimentos já consolidados, com o fim no aprendizado e familiarização dos mesmos, sendo sua base teórica focada na aplicação, e não no desenvolvimento de novas abordagens.

2.1 Arquitetura e Funcionamento dos LLMs

Os LLMs, baseados em arquiteturas de Transformer, revolucionaram o campo de processamento de linguagem natural (NLP). A arquitetura Transformer, introduzida por Vaswani et al. (2017), utiliza mecanismos de atenção para melhorar a capacidade do modelo de entender e gerar texto coerente e contextualmente relevante . Este modelo processa o texto em paralelo, ao contrário de métodos sequenciais anteriores, permitindo uma análise mais rápida e eficiente de grandes volumes de dados

2.2 Aplicações de LLMs em Geração de Conteúdo

Os LLMs têm sido aplicados em diversas áreas, incluindo a geração de conteúdo automatizado. No contexto de aplicativos web, a capacidade de um LLM para gerar perguntas pode ser particularmente útil em cenários como educação, atendimento ao cliente, e chatbots interativos. Ao gerar perguntas relevantes e contextualmente apropriadas, os LLMs podem melhorar significativamente a interatividade e a eficácia de tais sistemas .

2.3 Desafios e Limitações

Apesar dos avanços, a utilização de LLMs para a geração de perguntas enfrenta alguns desafios. Um dos principais desafios é garantir que as perguntas geradas sejam sempre relevantes e precisas. Problemas como a geração de perguntas sem sentido ou inadequadas para o contexto podem ocorrer devido a limitações no treinamento e nos dados utilizados pelo modelo. Além disso, questões éticas e de privacidade também devem ser consideradas, especialmente quando o modelo acessa dados sensíveis dos usuários .

Lacuna de Pesquisa 1. Descrever aqui a lacuna de pesquisa. Se tiver mais que uma, criar outro bloco.

***Pergunta* 1.1.** Aqui vai a pergunta de pesquisa 1.

***Pergunta* 1.2.** Aqui vai a pergunta de pesquisa 2.

3 METODOLOGIA

3.1 Visão Geral

Os processo desenvolvimento do software contemplam o desenvolvimento das funções que aplicam o modelo, criar o back-end que ira manipular as requisicoes http,criar o fron-end no qual o usuario ira interagir e os teste.

3.2 Back end

Desenvolvido com python e Flask como um servidor que tem três end-points base

3.2.1 A home

que é caracterizado pelo end-point "/"que recebe uma requicisão tipo get onde o servidor retorna um a paginal inicial que permite usaurio entrar com o link do video ou texto desejado de se gerar as perguntas. A partir dela que o usuario é direcionado as outras duas.

3.2.2 Text_post

que é caracterizado pelo end-point "/text_post "onde ele chama a funcao que aplica o modelo ao texto passado no parametro no body da requicicao e retorna uma pagina com as perguntas geradas.

3.2.3 Yb_post

que é caracterizado pelo end-point "/yb_post "que no body da requicicao recebe o link do video em questao, chama uma funcao que transcreve o video para se ter o texto dele, e aplica o modelo sobre esse texto retornando uma pagina com o texto do video, e as questoes geradas.

3.3 Front end

foi usado html e css, e as paginas sao retornada pelo servidor atraves da funcao "render_template()"presente no flask que além de nos permitir enviar html e css para o client, nos permite inserir codigos python caracterizados pela abertura e fechamento de chaves duplas "{{codigo em questao}}", e nos permite passar variaves do servidor como argumentos para a pagina, no caso, um dicionario com as informacoes necesarias incluido

um array com as perguntas geradas.

3.4 Obtendo a transcrição do video do youtube

para se obter a transcrição do texto do video foi usado uma api em forma de biblioteca chamada "youtube_transcript_api", porém, a qualidade da transcrição depende dos recursos do video em si,videos com transcricoes geradas automaticamentes sao menos precisas que as geradas manualmente.

3.5 Gerandos as Questões

para a gerção usamos o modelo lmqg, que possui codigo aberto e esta disponível tanto no github, pypi, e hugginface, no nosso caso, obtemos eles pela hugginface.o Modelo recebe um texto de contexto com o limite de até 512 caracteres, entao criammos uma funcao que pega o texto, divide o numero de caracteres pelo inteiro da divisao 512 + 1. por exemplo, um texto de 1200 caracteres seria divido por 3, onde aplicamos a geracao para cada uma dessas fraçoes, no caso, do caractere 1 ate o 400, do 401 ate o 800 e do 801 até o 1200.

4 RESULTADOS E CONCLUSÕES

Com o aprendizado, desse e semestres anteriores e pesquisas implementei uma LLM num protótipo de aplição web, que gerasse questões a partir pricipalmente de um video do youtube. A qualidade do resultado depende muito do: idioma do video,tema do texto ou video de contexto, se o autor do video acrescentou uma transcricao oficial ou se ela foi gerado pelo algoritmo do youtube entre outras coisas.

Um problema desse tipo de inteligencia artificial é que ele sempre responde, mesmo quando nao tem embasamento suficiente, por exemplo, ele sempre ira gerar questões, porém, as vezes elas são muito boas, ou nao tem muito sentido. Outro desafio encontrado foi a demora para se obter as respostas, a funcao de analise e geração ela é consideravelmente custosa, e a sua limitação de caracteres pede q ela seja rechamada algumas vezes.

Essa aplicação é só um exemplo de como a programação e a inteligência artificial pode ser usada para fins ditáticos, onde tive muito saprendizados tanto sobre inteligência artificial baseada em modelos de linguagens e redes neurais, tanto nas suas aplicações, por exemplo, para se lidar com o tempo de uma requição enquanto servidor esta rodando o modelo.

REFERÊNCIAS BIBLIOGRÁFICAS

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. Not cited.

REDDY, S.; CHEN, D.; MANNING, C. D. Coqa: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 7, p. 249–266, 2019. Not cited.