

Primer MVP

Introducción

El objetivo de este proyecto es crear un sistema de IA generativa que pueda ayudar a las personas a crear anuncios de Coca-Cola. El sistema tendrá tres componentes principales:

- Un chatbot que podrá generar ideas generales para los anuncios.
- Un modelo de imágenes generativo que podrá crear imágenes de Coca-Cola, personas, paisajes, etc.
- Un modelo de texto generativo que podrá crear texto para los anuncios.

Este reporte se centrará en la explicación de los notebooks de los modelos que se probaron para el proyecto.

Notebook Llama 2

En el notebook de Llama-2, utilizaremos el modelo de Llama-2, un modelo de lenguaje de código abierto desarrollado por Meta AI. Este modelo de transformador está pre-entrenado en un vasto conjunto de datos que abarca tanto texto como código. Llama-2 es capaz de generar texto, traducir idiomas, crear diversos tipos de contenido creativo y responder preguntas con información precisa. Además, en este contexto, Llama-2 actuará como un chatbot, proporcionando ideas para especificaciones, generación de texto para modelos generativos de imágenes y mejoras para anuncios de Coca-Cola.

Un aspecto destacado de este notebook es la creación y despliegue de una interfaz web amigable para el usuario. Esta interfaz tiene la capacidad de realizar tareas de generación de texto, traducción de idiomas y responder preguntas. Todo esto se logra gracias al sólido modelo de lenguaje Llama-2-7b-chat-hf.

El proceso comienza con la adquisición e instalación del software y archivos necesarios, asegurando que todos los requisitos previos estén en su lugar para garantizar una ejecución sin contratiempos. Posteriormente, se inicia la interfaz web.

Una vez que la interfaz web está en funcionamiento, los usuarios pueden acceder a ella fácilmente simplemente navegando a <http://localhost:7860> en sus navegadores web. En esta interfaz de usuario, los usuarios pueden ingresar sus solicitudes, y la plataforma responderá de inmediato generando texto, facilitando la traducción de idiomas o proporcionando respuestas a sus preguntas.

Notebook SDXL TUNED

Este cuaderno Colab tiene un objetivo específico: entrenar y desplegar un modelo de generación de imágenes basado en indicaciones textuales. Se ofrece a los usuarios un proceso de configuración simplificado, que les permite configurar varios aspectos del proyecto. Estos parámetros incluyen la elección del modelo, el nombre del proyecto, el prompt y los hiperparámetros esenciales para el entrenamiento. En particular, el código permite enviar el modelo entrenado al centro de modelos Hugging Face, una plataforma para compartir modelos pre-entrenados. Tras la configuración, se inicia el proceso de entrenamiento, aprovechando el modelo seleccionado y los parámetros definidos por el usuario. Además, se incluye una sección de inferencia posterior al entrenamiento en la que se demuestra la capacidad del modelo para generar imágenes a partir de mensajes de texto.

Se hizo un intento de utilizar AutoTrain DreamBooth para realizar fine-tuning de un modelo con imágenes de latas de Coca-Cola. Sin embargo, se encontró que las bibliotecas en el entorno de Colab no estaban actualizadas y no coincidían con las requeridas por la biblioteca xTransformers. Como resultado, en este momento, solo se pudo realizar un primer intento de cómo se vería una lata de Coca-Cola utilizando el modelo sin aplicar el proceso de fine-tuning.

Notebook Fontjoy

En este cuaderno se encuentra incrustada la página de Fontjoy, una página de generación de combinaciones de estilos de textos por medio del uso del emparejamiento de estilos, el cual tiene como base el usar vectores de estilos, el cual percibe los estilos en base a ciertos factores y los coloca en base a sus valores en cada vector. La selección de estilos de la que se pueden elegir son todos parte de google font, por lo que son gratuitos y de libre uso comercial.

Una de las particularidades de la página bajo la cual fue creada consiste en que puedes ajustar la generación de los estilos a utilizar en sus combinaciones en base a los contrastes que se deseen en relación con otros estilos, siendo capaz de bloquear unos estilos para que estos no cambien mientras que se continúa buscando por otro estilo en caso de encontrar uno que no resulte ser el deseado.

Oportunidades de mejora

Llama 2:

En particular, la velocidad y la capacidad de respuesta de la interfaz de usuario (UI) podrían optimizarse para una experiencia más fluida. En algunos casos, la UI podría haber mostrado un rendimiento lento, lo que podría desanimar a los usuarios. Considerar la optimización de la UI, quizás mediante la implementación de técnicas de carga diferida o mejora del procesamiento en segundo plano, podría ser un valioso avance.

SDXL:

Una oportunidad de mejora se presenta en la posibilidad de realizar un proceso de fine-tuning en futuras iteraciones de este cuaderno. Esto permitiría abordar los desafíos experimentados en la generación de imágenes de latas de Coca-Cola, las cuales mostraron ciertas deficiencias en su calidad y en la coherencia del texto asociado. La ejecución de un proceso de fine-tuning podría optimizar y ajustar el modelo para obtener resultados de mayor precisión y cohesión en la generación de contenido visual y textual.

Fontjoy:

En cuanto a Fontjoy si bien es de uso libre y puede ayudar mucho a elegir a seleccionar un estilo siendo ya una herramienta utilizada ya por diseñadores, esta tiene varios fallos, en primer lugar no dispone de una API, dado eso fue la necesidad de incrustar la página en el notebook, los demás problemas son en relación a las medidas de distancias vectoriales de los vectores de estilos, ya que pueden tener magnitudes similares entre sí que conlleven a un resultado no tan deseado. Sin embargo la mayoría de las alternativas requieren de un coste monetario o no disponen de un API o alguna otra forma de utilizarlo como la incrustación en este caso.