

Caracterización de personajes y generación de libretos para cortometrajes usando LLMs

William Salamanca, *Ingeniería de Sistemas y Computación, UPTC*

Yasser Cristancho, *Ingeniería de Sistemas y Computación, UPTC*

Abstract

The abstract goes here.

Index Terms

LLM, AI

I. INTRODUCCIÓN

LOS Modelos de Lenguaje Largo o Extenso, de su traducción en sus siglas en inglés LLMs (Language Large Models), han permitido evidenciar la capacidad potencial de generar texto o contenido para diferentes tareas. Contemplando lo anterior, en este proyecto se propone desarrollar un sistema que pueda analizar la narrativa de un libro con fin de explorar la aplicación de LLMs para la identificación de personajes con sus características detalladas y la generación de libretos dependiendo del contexto que se ingrese en términos de ubicación, época, entre otros.

II. PLANTEAMIENTO DEL PROBLEMA

III. METODOLOGÍA

A. Explorar LLMs

Se necesita tener una vista general de la variedad de LLMs que se encuentran actualmente implementados y desplegados.

B. Explorar documentación de los LLMs

Identificar aspectos relacionados a los requerimientos básicos y/o necesarios de los LLMs para poder hacer finetune de estos. Requerimientos de software y hardware, así como posibles comparativas que indiquen su desempeño en tareas de NPL, específicamente en las de identificar, caracterizar y clasificar personajes de libros. Además de la generación de libretos para cortometrajes a partir de personajes identificados, caracterizados y clasificados, y de un contexto histórico en el que se desarrolla el libreto.

C. Explorar fuentes de datos

Para el proceso de finetune de un LLM seleccionado se necesita tener varias fuentes de datos que en un principio nos provean de una descripción de las características de los personajes que se encuentran en este. Adicionalmente a lo anterior, se necesitan datos sobre contextos históricos con el fin de tener un contexto en el que se desarrolla el libreto para el cortometraje.

D. Comparativa inicial entre LLMs seleccionados

- Capacidad de los modelos para realizar las tareas de identificación, caracterización y clasificación de personajes, además de la generación de libretos con base a lo anterior y un contexto histórico dado.
- Tamaños y complejidades asociados a los modelos.
- Disponibilidad y acceso.
- Requerimientos de hardware.

E. Comparativa mediante ejercicios entre LLMs seleccionados

Partiendo de las mismas entradas y con las mismas condiciones de hardware, identificar el que tenga mejor rendimiento en términos de tiempo pero, teniendo en cuenta la calidad de la salida.

IV. DEFINICIÓN DE LA IDEA

V. LLMS IDENTIFICADOS

- GTP-1
- GTP-2
- GTP-3
- GTP-4
- Megatron-LM
- PaLM
- LaMDA
- GPT-Neo
- GPT-J
- BERT
- RoBERTa
- T5
- DistilBERT

VI. COMPARATIVA INICIAL ENTRE LLMS SELECCIONADOS

A. *GPT-Neo*

B. *T5*

VII. COMPARATIVA MEDIANTE EJERCICIOS ENTRE LLMS SELECCIONADOS

VIII. FUENTES DE DATOS SELECCIONADOS

IX. HALLAZGOS

X. CONCLUSIÓN

The conclusion goes here.