

# 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

**L**OS 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

El objetivo principal de este proyecto es desarrollar una solución innovadora que aproveche la potencia de los modelos de lenguaje grandes (LLM) para facilitar y agilizar dos tareas clave en la producción de cortometrajes: la caracterización de personajes y la generación de libretos para cortometrajes.

En la etapa de preproducción de una obra, guion o cortometraje, la creación de personajes detallados y bien definidos es fundamental para construir una historia cautivadora y convincente. Sin embargo, este proceso suele ser laborioso y requerir una amplia investigación y esfuerzo creativo por parte de los guionistas y directores. Mediante la implementación de un modelo LLM especializado en caracterización de personajes, se busca automatizar y agilizar esta tarea, permitiendo la generación de descripciones detalladas y coherentes de personajes a partir de una breve entrada o premisa.

Por otra parte, la escritura del libreto es una etapa crucial y desafiante en la producción cinematográfica. Requiere habilidades narrativas sólidas, conocimiento del formato específico de guiones y una gran inversión de tiempo. Al utilizar un modelo LLM entrenado específicamente para la generación de libretos, se pretende agilizar este proceso, brindando a los guionistas y directores una herramienta poderosa que les permita obtener borradores iniciales de libretos a partir de una idea o premisa básica.

El proyecto contempla el desarrollo de dos modelos LLM separados, uno para cada tarea, con el fin de maximizar su eficacia y rendimiento. Estos modelos serán entrenados con conjuntos de datos específicos y relevantes para cada tarea, utilizando técnicas de aprendizaje supervisado.

La implementación de estos modelos LLM en el flujo de trabajo de producción de cortometrajes tiene el potencial de ahorrar tiempo y recursos, al tiempo que fomenta la creatividad y la exploración de ideas. Los guionistas y directores podrán enfocarse en refinar y perfeccionar los resultados generados por los modelos, en lugar de dedicar largas horas a tareas repetitivas o de baja complejidad.

## 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

Modelo	Capacidad de Tokenización (Límite de Contexto)	Tipo	Número de Parámetros
GPT-1	512 tokens	Pago	117M
GPT-2	1024 tokens	Código libre	1.5B
GPT-3	2048 tokens	Pago	175B
GPT-4	Hasta 8192 tokens o más, dependiendo de la configuración	Pago	175B+
Megatron-LM	2048 tokens	Código libre	8.3B
PaLM	2048 tokens	Pago	540B
LaMDA	Hasta 2048 tokens	Pago	137B
GPT-Neo	2048 tokens	Código libre	1.3B y 2.7B
GPT-J	2048 tokens	Código libre	6B
BERT	512 tokens	Código libre	110M
RoBERTa	512 tokens	Código libre	355M
T5	512 tokens	Código libre	11B
DistilBERT	512 tokens	Código libre	66M

TABLE I  
CAPACIDAD DE TOKENIZACIÓN, TIPO Y NÚMERO DE PARÁMETROS DE LLMs IDENTIFICADOS

## VI. COMPARATIVA INICIAL ENTRE LLMs SELECCIONADOS

### A. GPT-Neo

”Hace 6 años, estaba volando sobre el desierto del Sahara y se estropeó el motor de mi avión. Estaba completamente solo.”

### B. T5

## VII. COMPARATIVA MEDIANTE EJERCICIOS ENTRE LLMs SELECCIONADOS

### VIII. FUENTES DE DATOS SELECCIONADOS

- Libros de cuentos y relatos cortos: Los libros de cuentos y relatos cortos son una fuente excelente de fragmentos narrativos concisos y autosuficientes, que podrían servir como base para desarrollar cortometrajes. Algunos ejemplos son los cuentos de Edgar Allan Poe, Anton Chéjov, Jorge Luis Borges, entre otros.
- Antologías de historias cortas: Las antologías recopilan historias cortas de diversos autores y géneros, lo que te brindaría una amplia variedad de estilos y temáticas para trabajar.
- Novelas cortas: Las novelas cortas, también conocidas como ”novellas”, son obras narrativas más extensas que un cuento, pero más breves que una novela completa. Podrías extraer fragmentos o capítulos completos de estas obras.
- Obras de teatro de un acto: Las obras de teatro de un acto, al ser más cortas y condensadas, son una excelente fuente de diálogos y tramas compactas que podrían adaptarse fácilmente a cortometrajes.
- Repositorios de guiones y libretos: Existen repositorios en línea que compilan guiones y libretos de películas, series de televisión y obras de teatro. Aunque no son fragmentos literarios, podrían ser útiles para aprender sobre la estructura y el formato de los libretos.
- Sitios web y plataformas de escritura creativa: Algunas plataformas en línea, como Wattpad, Inkitt o WritingPrompts, albergan historias cortas, relatos y escritos creativos de autores emergentes. Estos podrían ser una fuente interesante de material.

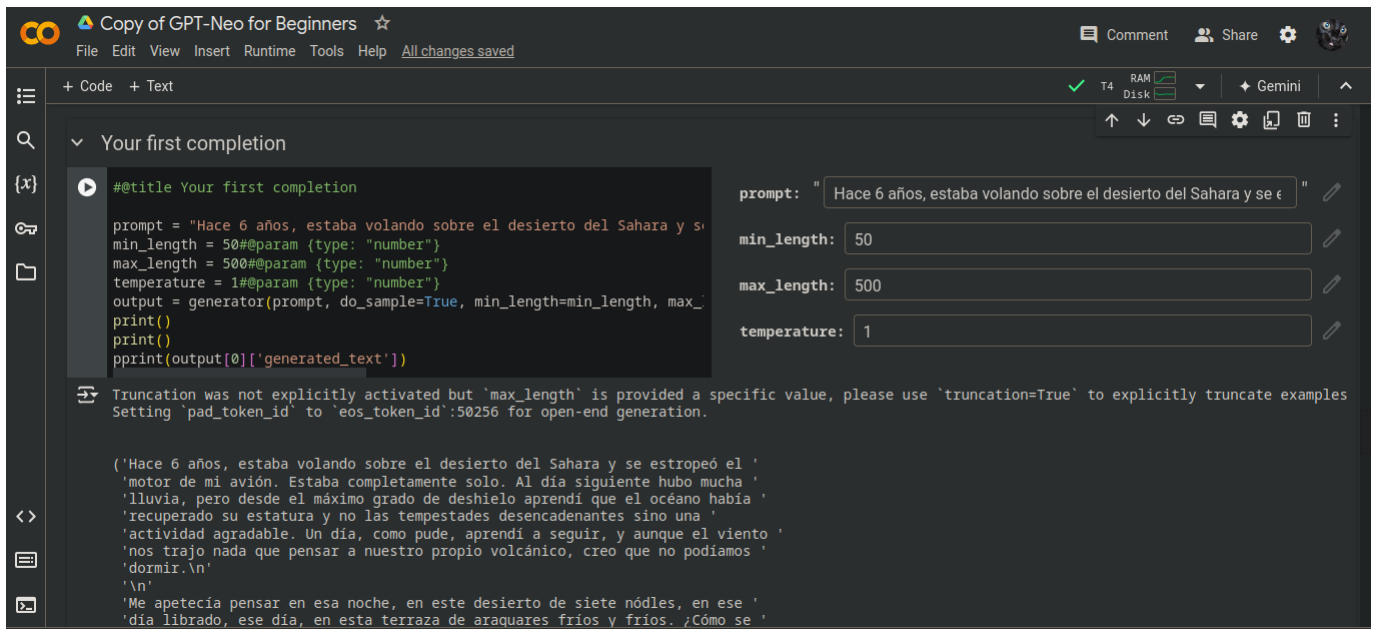


Fig. 1. Ejemplo GPT-Neo

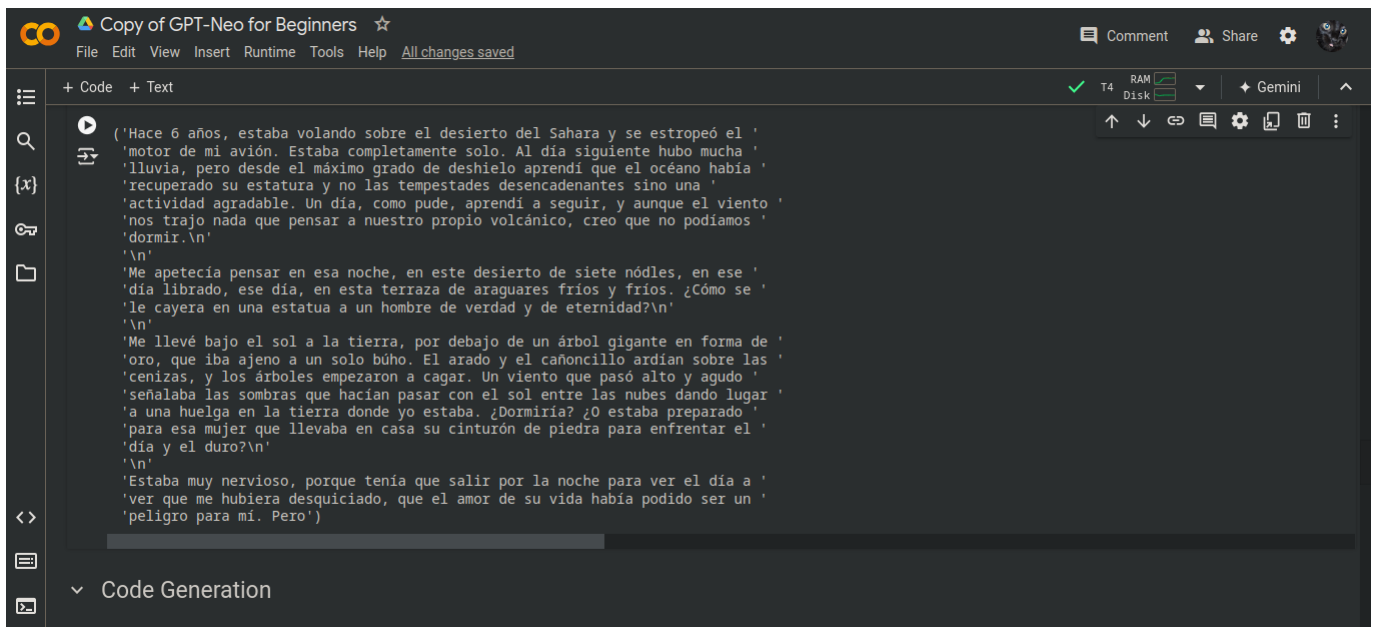


Fig. 2. Ejemplo GPT-Neo

## IX. HALLAZGOS

## X. CONCLUSIÓN

The conclusion goes here.