

PROYECTO LLMs : GENERADOR DE CONTENIDO

Planteamiento

La empresa Digital Content os ha contactado para crear un sistema de generación automático de contenido para diversos medios y audiencias utilizando inteligencia artificial generativa. Los medios que más interesan son: Blogs, Twitter/X, Instagram y linkedIn (se pueden incluir más si se desea). Quieren una prueba de concepto (poc) funcional, que les permita crear contenido de texto e imágenes, para automatizar sus publicaciones. Para ello os da completa libertad para utilizar los LLMS, APIs y modelos de imágenes que creáis oportunos. Un requisito que os han pedido es que quieren minimizar el gasto hasta que estén seguros de que quieren usar este sistema por lo que se pide que en la medida de lo posible se utilicen modelos en local, APIs gratuitas o con pruebas gratuitas, aunque ello implique tener limitaciones a la hora de generar contenido como límite de peticiones, o lentitud a la hora de generar el contenido.

Al sistema le debes pasar información básica de lo que quieres generar (tema, audiencia, plataforma, etc) y este debe generar contenido listo para publicar.

Han oído hablar de diferentes tecnologías (Modelos de Lenguaje, Generadores de imágenes, RAGs, Agentes, bases de datos de Vectores, etc) y quieren que exploréis sus posibilidades en el caso de uso de la generación de contenidos.

Quieren un sistema que sea fácilmente extensible por lo que se pide que se utilice frameworks para creación de aplicaciones basadas en LLMs (ej: LangChain) .

Plazos

Las propuestas de solución se deberán entregar el día **6 de Diciembre** (el 9 de Diciembre se presenta)

Condiciones de entrega

Para el día de la reunión, será necesario entregar:

- Presentación de 30 min (negocio + tecnica)
- Proyecto en github
- Artículo en Medium

Tecnologías que se pueden aplicar

- LangChain

- LangSmith
- LangGraph
- LlamaIndex
- CrewAI
- Ollama
- Groq
- Chroma
- Faiss
- Pinecone
- Huggingface
- Streamlit
- Gradio
- Chainlit

Niveles de Entrega

Nivel Esencial:

Crea contenido de texto sobre diferentes temas que proporciona el usuario, adaptado a diferentes plataformas y audiencias (blog posts, twitter/X, Instagram, SEO, divulgación, infantil etc) utilizando prompts. Para este nivel solo se requiere utilizar prompts.

Una interfaz web en la que interactuar.

Redactar y publicar un artículo en Medium explicando el proyecto que se ha creado.

Nivel Medio:

Dockerizar la aplicación.

Poder seleccionar entre al menos dos LLMs para la generación de contenido.

Posibilidad de añadir información de la empresa o persona que esté haciendo la publicación para que se añada como prompt a todo el contenido generado de forma que la generación de contenido esté personalizada.

Incluir imágenes relevantes en el contenido, de forma que quede integrado en el texto. Las imágenes pueden ser generadas con IA o obtenidas por otros medios (por ejemplo una API que te devuelva imágenes).

Nivel Avanzado:

Añadir trazabilidad de las peticiones y las respuestas realizadas a vuestro sistema. Podéis usar herramientas como LangSmith.

Incluir la funcionalidad de generar el contenido en Castellano, Inglés, Francés e Italiano.

Funcionalidad de noticias con información actualizada sobre los mercados financieros. Para ello se deberá proporcionar al modelo información actualizada por medio de APIs.

Desarrollar una funcionalidad de generación de contenido científico divulgativo que mejore la calidad del contenido generado en un área científica específica (por ejemplo, física cuántica, inteligencia artificial, biomedicina, astrofísica, etc.) para que sea comprensible para el público general. Para ello, se debe crear una arquitectura RAG (Retrieval-Augmented Generation) que extraiga y sintetice conocimiento relevante de documentos científicos provenientes de arXiv. Se tiene libertad para elegir tanto el tema específico dentro de las ciencias como los métodos empleados para buscar, seleccionar y procesar documentos científicos que alimenten la arquitectura RAG

Nivel Experto:

Aumentar el caso de uso del RAG científico utilizando grafos de conocimiento como fuente de contexto o información factual (Graph RAG).

Crear un sistema multiagente, en el se tengan agentes especializados en tareas concretas y se envíe al agente correcto cada caso de generación de contenido. Con acceso a diferentes LLMs específicos, prompts customizados, o acceso a herramientas dependiendo del caso de uso.

Añadir algún tipo de guardarraíles para evitar las alucinaciones y asegurar la calidad del contenido generado.