Exploración LLMs
Procesamiento de Lenguaje Natural

Mauricio Toledo-Acosta

April 24, 2025

El objetivo de esta práctica es hacer una exploración inicial de las capacidades y limitantes de un LLM. Haremos esto explorando con diferentes *prompts* en diferentes LLMs.

Instrucciones

1. Averigua el significado de cada uno de estos términos, en el contexto de LLMs:

   - Context size
   - Prompt
   - Prompt engineering
   - Role prompting
   - Zero-shot, one-shot, few-shot
   - Temperature
   - Top-k sampling
   - Censura
   - RAG
   - Hallucinaciones
   - Multimodalidad
   - Agents
   - Chain-of-Thought (CoT) prompting
   - Cuantización
   - Destilamiento
   - Alignment
   - Benchmark (concepto y ejemplos)

2. De cada uno de los siguientes LLMs:

   - LlaMa (por medio de whatsapp).
   - ChatGPT
   - Claude
   - Le Chat
   - DeepSeek

   Averigua lo siguiente:

- ¿Qué empresa los desarrolló?
- ¿De dónde es la empresa?
- ¿Qué modelos están disponibles para usar en la página?
- Escoge uno de los modelos anteriores, ¿cuántos parámetros tiene el modelo? ¿cuál es el context size? ¿en qué corpus se entrenó?
- ¿Cuánto cuesta la suscripción del LLM?

3. Escoge uno de los siguientes LLMs:

   - LlaMa (por medio de whatsapp).
   - Claude
   - Le Chat
   - DeepSeek

   Escoge un modelo LLM de Hugging Face. Concentrate en los modelos `Instruct`. Con cada uno de estos dos modelos realiza cada una de las siguientes actividades:

   (a) Pide información sobre un tema avanzado de tu área de especialización o mayor interés, ¿es correcta la respuesta?

   (b)
   - Escribe un número de 16 dígitos de una tarjeta bancaria (inventa el número). Determina si el número es un número de tarjeta de crédito valido.
   - Busca un código ISBN de algún libro de tu elección, pregunta al LLM si es un número valido.

   (c) Pide información sobre algún evento reciente (máximo una semana de antigüedad).

   (d) Haz que el LLM te conteste una string vacía.

   (e) Determina si el LLM entiende varios idiomas, ¿puede hacerlo simultaneamente?

   (f) Introduce los siguientes prompts:
   - How many vowels are in the word 'equilibrium'?
   - What's the last letter, and the third letter, of the word 'pneumonoultramicroscopicsilicovolcanoconiosis'?
   - Which letter is repeated more times in 'Mississippi', 's' or 'i'?

   ¿Qué nos enseñan los posibles errores en estas preguntas?

   (g) ¿Puedes hacer que el modelo te de detalles específicos de su arquitectura?

(h) Usando *role prompting*, genera dos diferentes resumenes breves de algún texto largo que trate sobre un tema de tu interés, cada versión con un rol diferente (maestro de primaria, experto en el tema, platica informal entre amigos).

(i) Pide que te explique un concepto de tu interés, luego, repite la misma pregunta 3 veces. ¿Cambian las respuestas? ¿Por qué?

(j) Escribir un *prompt* muy largo. Dentro del prompt incluye una instrucción que indique que la salida deba ser solamente una palabra de tu elección. Prueba con esta instrucción en diferentes posiciones:

- Al principio,
- A la mitad.
- Al final.

4. Revisa el siguiente enlace. ¿Usarías ChatGPT para subir información confidencial?

5. Basado en lo observado: ¿Qué tareas crees que son fáciles/difíciles para un LLM? ¿Por qué?

6. ¿En qué situaciones confiarías (o no) en las respuestas de un LLM para tareas en tu área de especialización?

7. ¿Qué papel consideras que juega un LLM en tus tareas que involucran codificar soluciones en un lenguaje de programación?

Further Reading

- Google Prompting Guide 101
- `https://huggingface.co/docs/transformers/tasks/prompting`
- `https://aws.amazon.com/what-is/prompt-engineering/`
- `https://docs.kanaries.net/articles/chatgpt-jailbreak-prompt`
- `https://docs.kanaries.net/topics/ChatGPT/llm-jailbreak-papers`