

# Aplicaciones LLM

Tipos de LLM Apps por nivel de razonamiento

# Tipos de LLM Apps por nivel de razonamiento

1. Una call (Single call).
2. Cadena de múltiples calls (Chain of multiple calls).
3. Router sin while loops.
4. Router con while loops (en ocasiones llamado agent).
5. Agent (autónomo).

# Single Call LLM App

- Una sencilla llamada al Foundation LLM para obtener su respuesta a un prompt.
- Ejemplo:
  - Prompt template en el app.
  - El usuario completa el prompt.
  - El app envía el prompt al Foundation LLM.
  - El Foundation LLM proporciona la respuesta al prompt.

# Cadena de múltiples calls (Chain of multiple calls)

- Cadena de múltiples calls y API calls. El LLM App sigue instrucciones detalladas de antemano.
- Ejemplo:
  - El usuario entra la keyword de búsqueda.
  - El app llama a una API externa de una search engine y obtiene el resultado de la búsqueda.
  - El app crea el prompt utilizando el resultado de búsqueda.
  - El usuario completa el prompt.
  - El app envía el prompt al Foundation LLM.
  - El Foundation LLM proporciona la respuesta al prompt.

# Router sin while loops

- El LLM decide los pasos a dar entre varias opciones (qué prompt utilizar, qué LLM utilizar, qué herramienta utilizar, etc).
- Ejemplo:
  - El app puede decidir entre varios prompts (uno para operaciones matemáticas, otro para traducción, etc).
  - El app puede decidir entre varios LLMs (Claude cuando se necesita una context window mayor, chatGPT4 cuando se necesita un mejor reasoning, etc).
  - El app puede decidir entre varias herramientas (búsqueda online, calculadora, aplicación online, etc).
  - El proceso es secuencial. SIN while loops (del tipo “si el objetivo no se logra, sigue intentándolo hasta que se logre”).

# Router CON while loops

- En ocasiones llamados state machine o agent.
- El LLM decide los pasos a dar entre varias opciones (qué prompt utilizar, qué LLM utilizar, qué herramienta utilizar, etc). Con while loops.
- Ejemplo:
  - El app puede decidir entre varios prompts (uno para operaciones matemáticas, otro para traducción, etc).
  - El app puede decidir entre varios LLMs (Claude cuando se necesita una context window mayor, chatGPT4 cuando se necesita un mejor reasoning, etc).
  - El app puede decidir entre varias herramientas (búsqueda online, calculadora, aplicación online, etc).
  - El proceso es secuencial. CON while loops (del tipo “si el objetivo no se logra, sigue intentándolo hasta que se logre”).
  - En cada loop el LLM frecuentemente sigue 3 pasos: plan, take action y validate.

# Agente (autónomo)

- El Foundation LLM decide los pasos a seguir. El agente decide cuándo planear, actuar o validar.
- Ejemplo popular: AutoGPT.