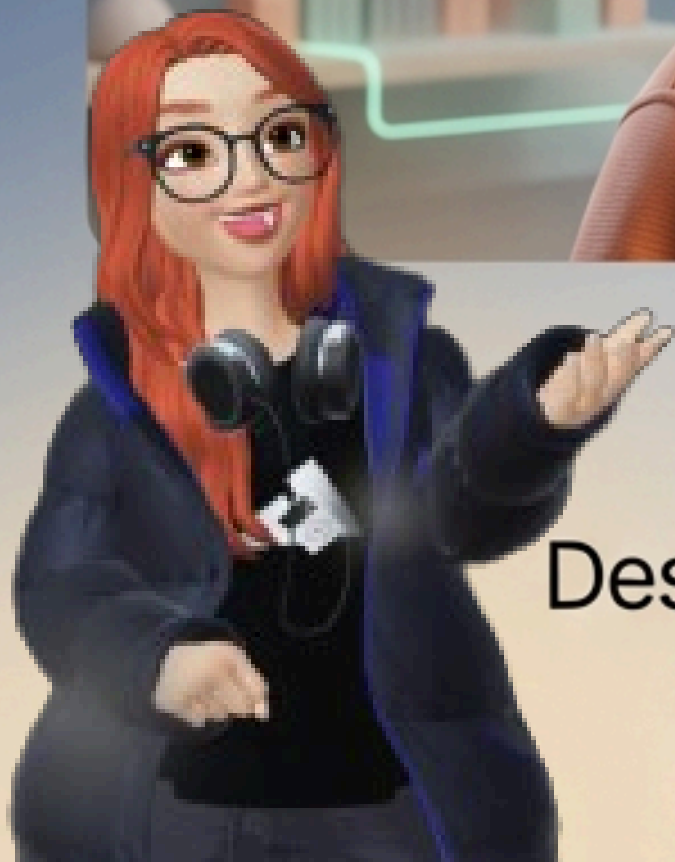


9 Arquitecturas RAG explicadas con Alpacas

Guía experta para pasar de "Basic RAG" a sistemas de IA avanzados.

Whimsical Tech Editorial

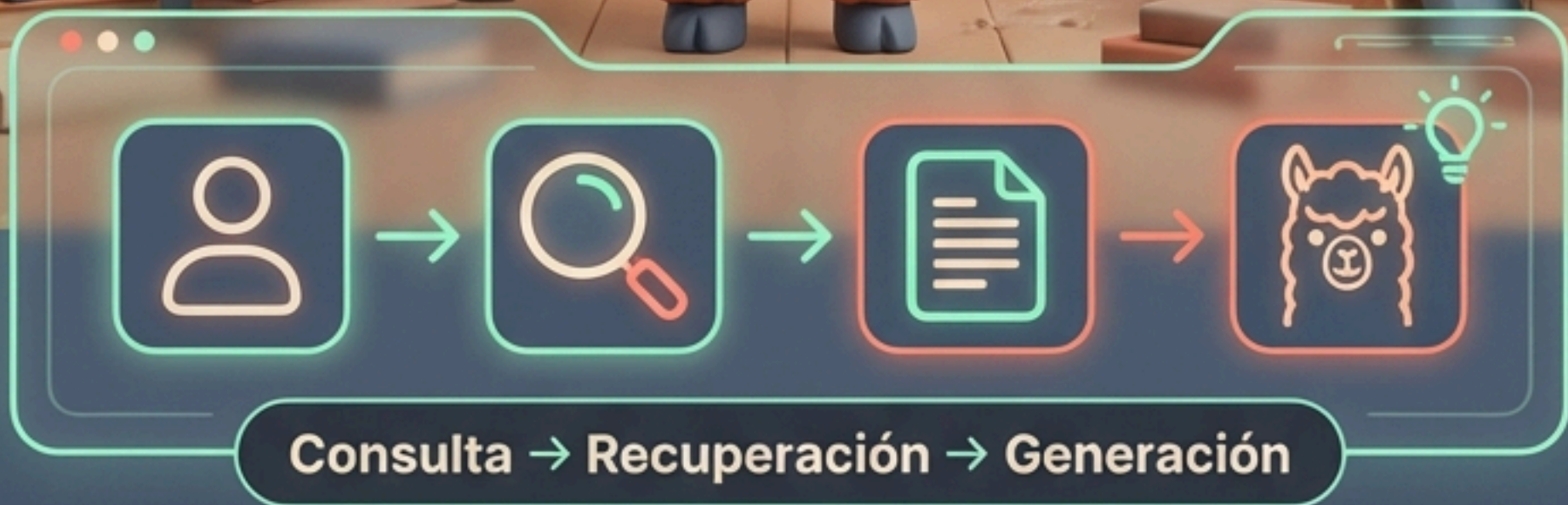
RAG LLM Vector DB Architecture



Desliza para evolucionar tu IA



1. La Alpaca Básica (RAG Ingenuo)



El Problema: La Alpaca es inteligente (LLM), pero su conocimiento es estático. Si no encuentra el libro exacto en la biblioteca (Base Vectorial), se inventa la respuesta.

⚠ Alucinaciones frecuentes

⚠ Contexto limitado y ruidoso

Nivel 1: La Alpaca Bibliotecaria (Búsqueda)



RAG Híbrido

Combina búsqueda semántica (vectores) con búsqueda léxica (palabras clave). Encuentra el dato exacto incluso con jerga técnica.



RAG Jerárquico

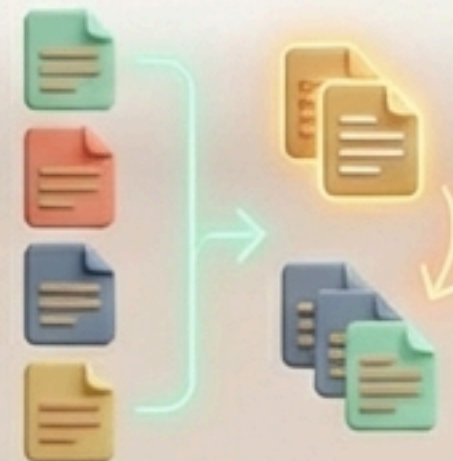
Primero lee el resumen. Si es útil, busca en los detalles. Máxima eficiencia para documentos largos.

Nivel 2: La Alpaca Chef (Procesamiento)



Chunking Semántico & Multimodal

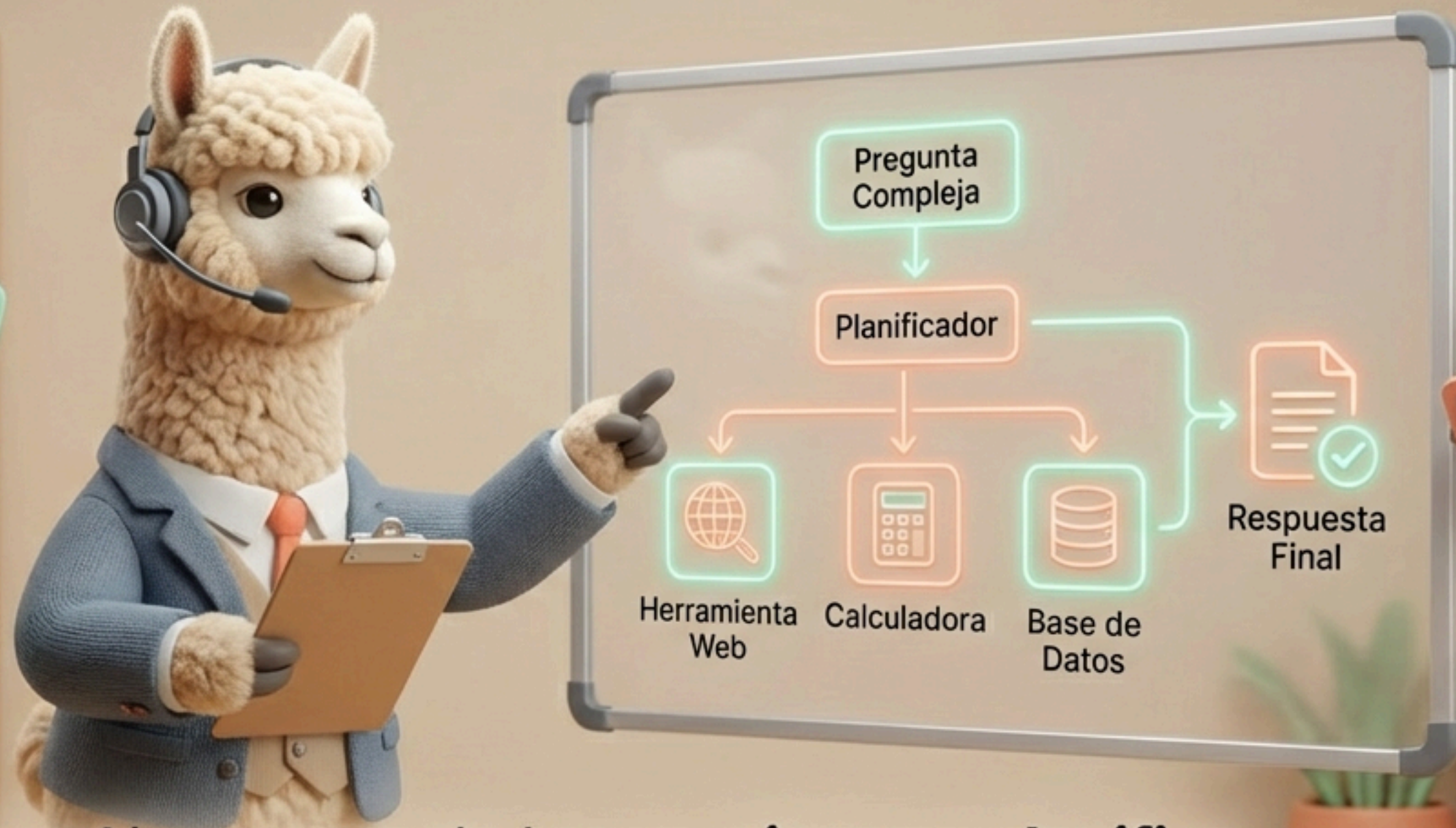
No cortes la información al azar. El Chef agrupa ideas relacionadas y procesa texto, imágenes y audio como ingredientes para una respuesta rica.



Reranking (Reordenamiento)

Una vez encontrados los ingredientes (docs), la Alpaca los prueba y reordena. Los más relevantes van al principio para que el LLM no se distraiga con ruido.

Nivel 3: La Alpaca Gerente (RAG con Agentes)



Esta Alpaca no solo busca, **piensa y planifica**.



Uso de Herramientas: Decide autónomamente si necesita buscar en Google, consultar una API o leer un PDF. Si falla, se corrige sola.

Nivel 4: La Alpaca con Memoria

La mayoría de las IAs tienen amnesia: olvidan lo que dijiste hace 5 minutos. Esta arquitectura cambia el juego.

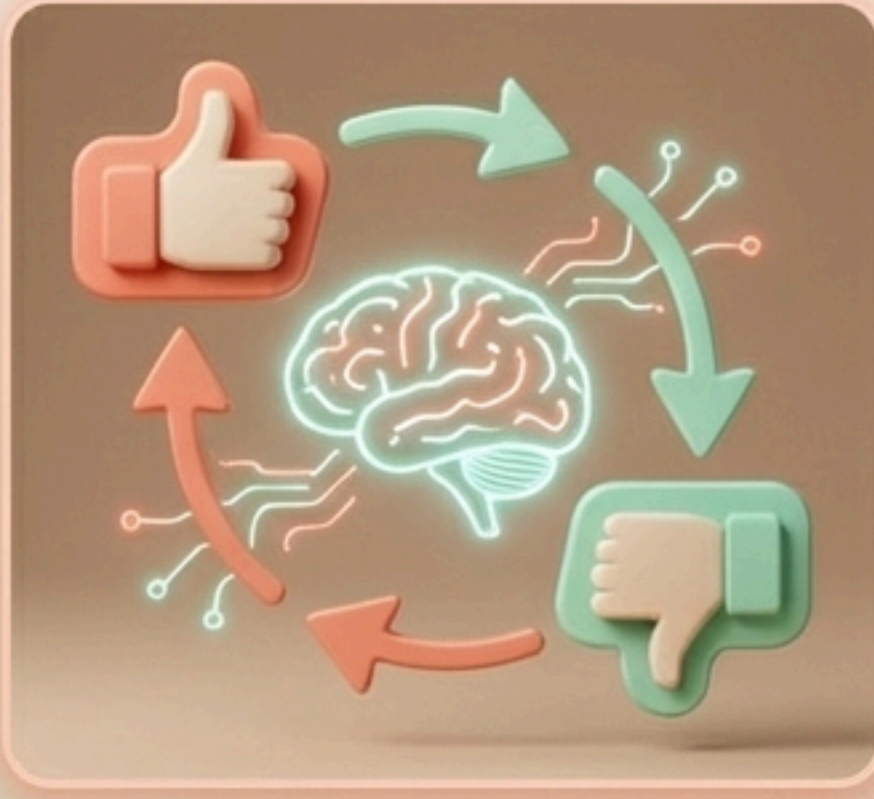
RAG de Memoria:

Incorpora un historial de largo plazo. Recuerda tus preferencias, proyectos anteriores y contexto de sesiones pasadas.

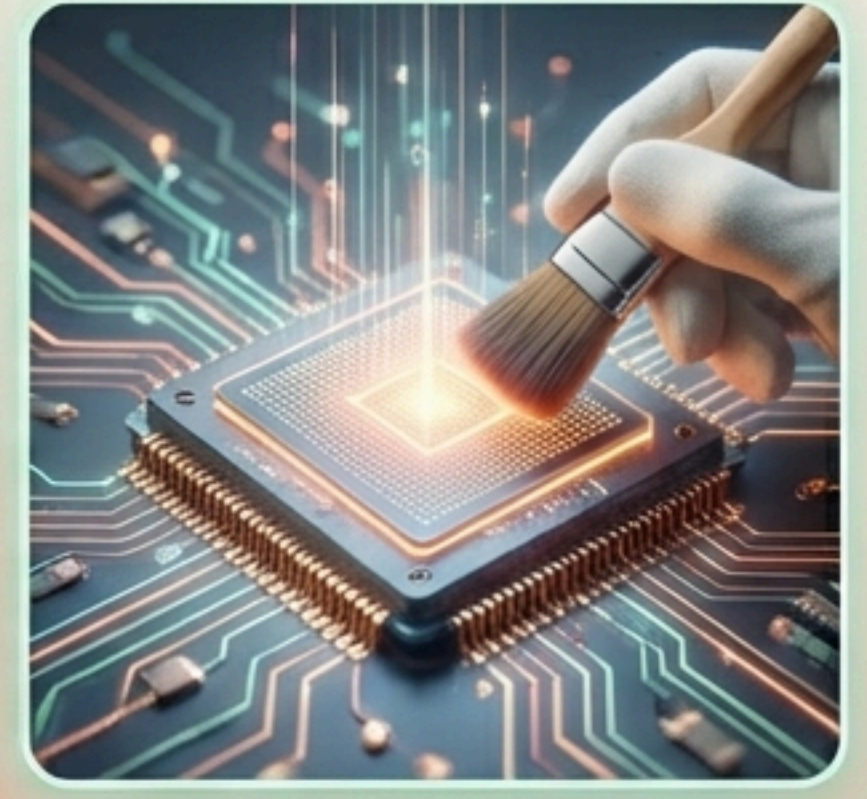


Nivel 5: La Alpaca Estudiosa (Mejora Continua)

Feedback Loop



Fine-Tuning



Feedback Loop

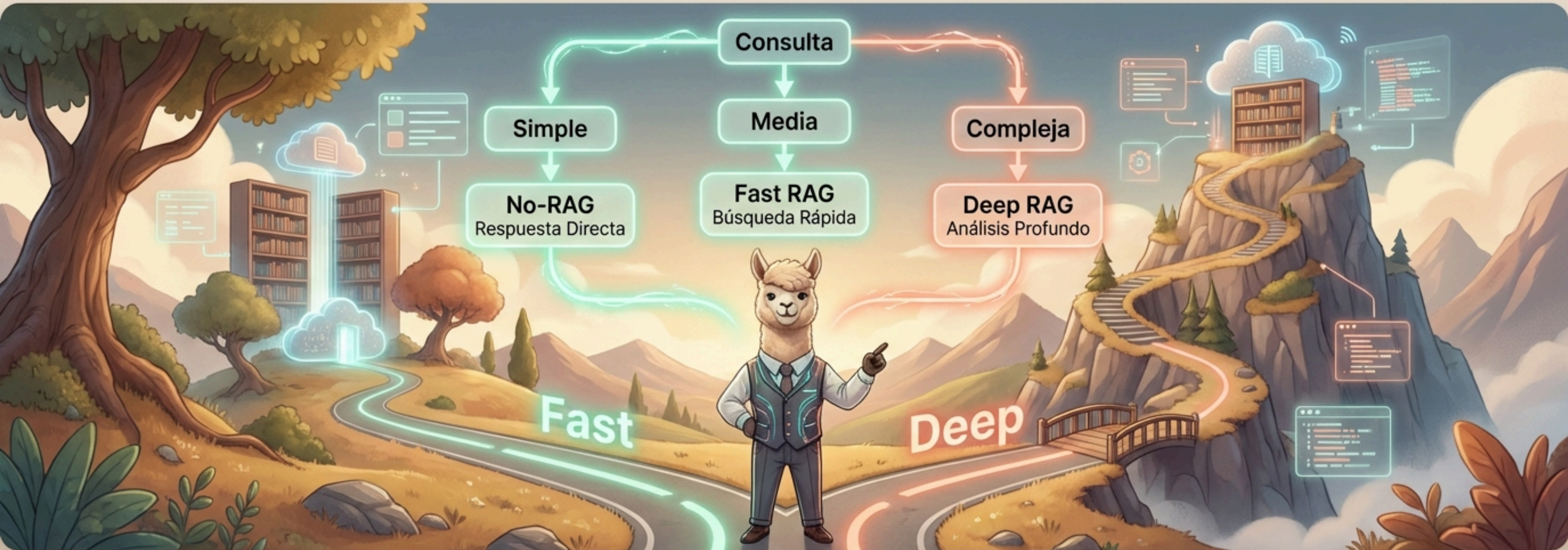
Aprende de sus errores. Si el usuario corrige una respuesta, el sistema ajusta el ranking para la próxima vez.

Fine-Tuning

Entrenamos el cerebro de la Alpaca con datos específicos de tu industria (Legal, Médico, Código) para que hable tu idioma experto.

Nivel 6: La Alpaca Estratega (RAG Adaptativo)

Eficiencia y Ahorro: No siempre hay que ir a la biblioteca.



El sistema evalúa la complejidad antes de actuar.
Ahorra tokens, dinero y reduce la latencia.

El Kit de Herramientas de la Alpaca



Tu Rebaño de Arquitecturas RAG

✓ 1. Básico (Linear)

✓ 2. Híbrido & Jerárquico

✓ 3. Agentes (Autónomo)

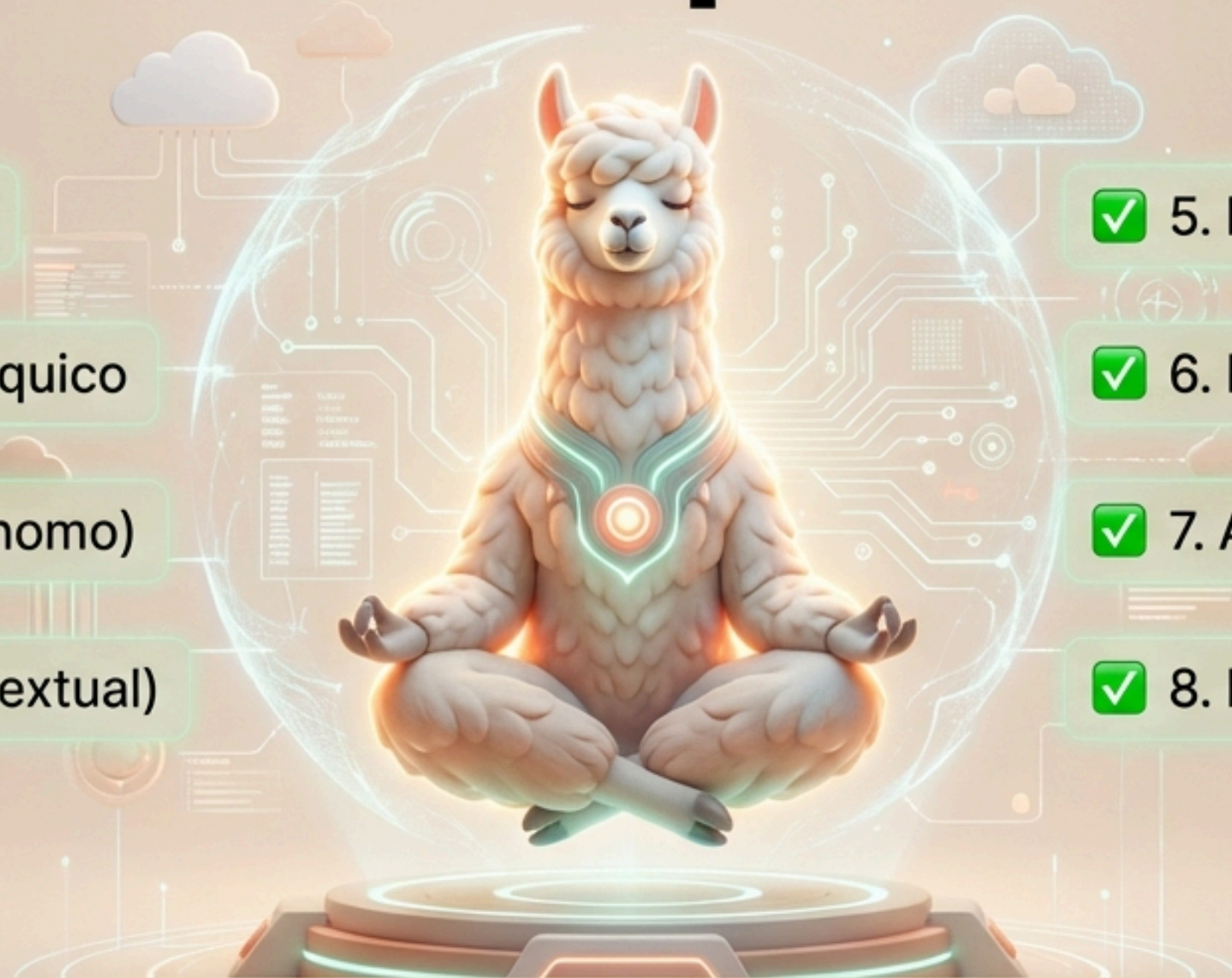
✓ 4. Memoria (Contextual)

✓ 5. Multimodal (Media)

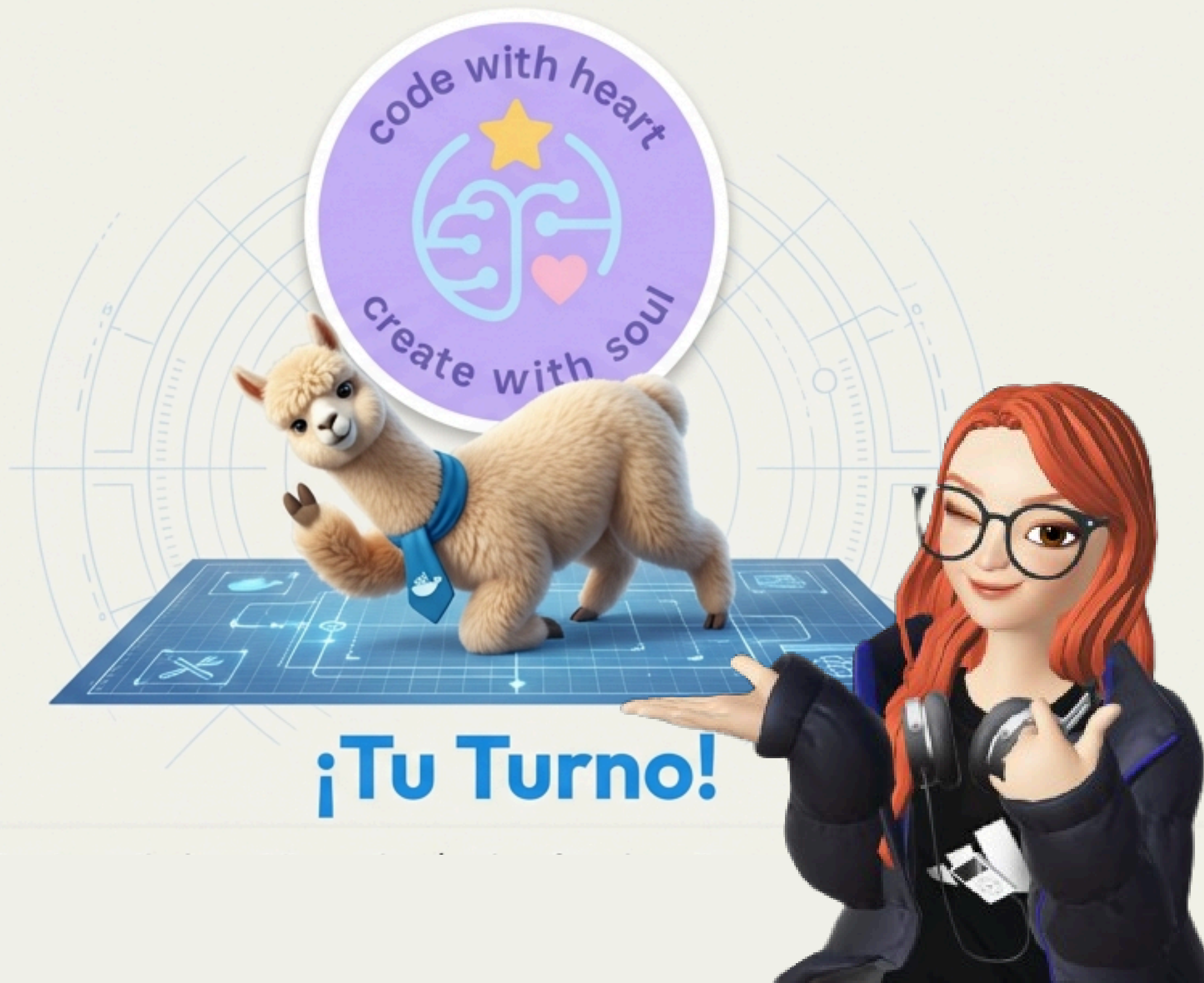
✓ 6. Feedback (Learning)

✓ 7. Adaptativo (Smart Routing)

✓ 8. Distribuido (Scale)



¿Cuál necesita tu proyecto? 🙌 Guárdalo para diseñar tu próximo sistema.



¡Tu Turno!

Sígueme para más consejos de ingeniería, creatividad y un poco de magia. ✨

