Antes de migrar, se mide con una carga real.

01

Elige

Una operación con volumen: soporte, RAG, documentos o código.

02

Línea base

Coste, calidad y latencia de hoy, con casos reales.

03

Compara

Los mismos casos con modelos abiertos, en las mismas condiciones.

04

Decide

Si el coste por tarea aceptada no baja, no se migra.

COSTE POR TAREA ACEPTADA

`coste completo ÷ tareas aceptadas`

QUÉ APORTAS

- Volumen de uso y factura actual de esa carga.
- Qué cuenta como tarea aceptada.
- Una muestra de casos, anonimizada.

QUÉ RECIBES

- Calidad, latencia y coste, con sus fuentes.
- Una recomendación: mantener, combinar o PoC.
- También por escrito si no compensa.

METODOLOGÍA COMPLETA

helmcode.com/es/guias/optimizacion-costes-ia