BUILD IN PUBLIC

CASO REAL

Mi evaluador acertó las cuatro.

Lo que me importó fue dónde dudó.

Claude Code | Codex | Jev

Roberto J. Aguirre G. · Head of AI

EL STACK

Tres roles.
Tres herramientas.

ORQUESTADOR

Claude Code

Orquesta y verifica

Arma cada encargo y comprueba con evidencia antes de dar algo por cerrado.



CEREBRO Y MANOS

Codex

Diseña e implementa

Con esfuerzo alto planifica la solución. Con esfuerzo medio escribe el código.



EVALUADOR

Jev

Decide y evalúa el diff

No escribe una línea. Lee el cambio y los tests, responde preguntas cerradas y dice cuánto confía.

CASO REAL

Un cambio real. Cuatro preguntas.

¿Fallan los mismos tests que en la versión anterior?

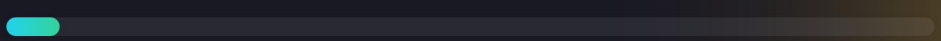


0.90

¿Qué causó los fallos?

el entorno **0.90** · ya existían **0.10** · el cambio **0**

¿Qué tan riesgoso es el cambio, de 0 a 3?



0.17

¿Alcanza para mergear sin revisar más?



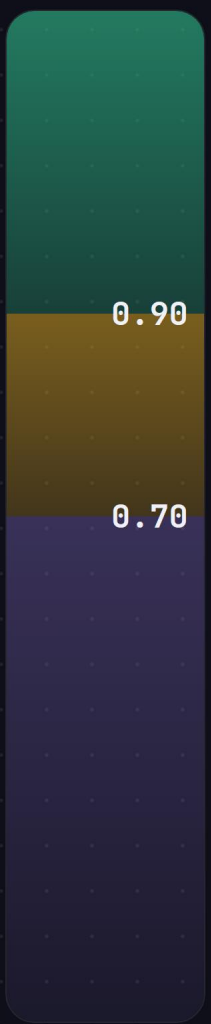
0.71

La única pregunta de juicio. Su menor confianza.

Las cuatro, correctas. Y dudó donde había que dudar.

LA REGLA

La confianza decide quién sigue.



0.90 o más

Sigue.

No molesto a ningún otro modelo.

Entre 0.70 y 0.90

Decide el orquestador.

Y deja por escrito por qué.

0.70 o menos

Escala al cerebro.

El modelo caro entra solo cuando hay ambigüedad real.

MEDIDO

Esa evaluación completa costó esto.

961
tokens de entrada

112
tokens de salida, que no se cobran

919 ms
las cuatro preguntas en una sola llamada

0.00004
dólares por evaluación

±0.02 de variación entre corridas idénticas. Ninguna decisión cambió.

LO QUE NO HACE

Saber qué no puede es parte de usarlo bien.

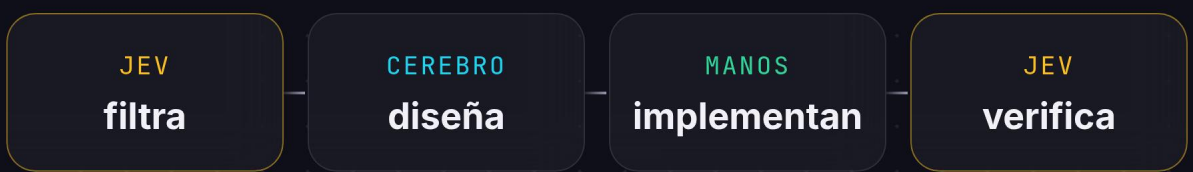
No diseña.

Solo acepta preguntas de `sí o no`, de `elección` o de `puntaje`. Elige entre opciones, nunca propone la que falta.

No busca la evidencia.

Razona muy bien sobre la que ya reuniste. Encontrarla sigue siendo trabajo tuyo.

POR ESO NO OCUPA LA SILLA DEL CEREBRO



LO QUE ME LLEVO

Un modelo que acierta es útil.

Uno que sabe cuándo dudar es confiable.

TE PREGUNTO

¿Qué decisión de un agente de IA sigues firmando a mano, y qué tendría que demostrarte para soltarla?