# Contraprueba

Una respuesta más votada también puede prometer algo que el manual no permite.

## Problema operativo

Antes de publicar una versión nueva de un asistente, el equipo de producto necesita saber qué respuestas siguen el procedimiento y cuáles introducen promesas sin respaldo. Cuando dos revisores discrepan, tiene que poder leer sus motivos y resolver el caso.

## Qué desarrollé

Diseñé un flujo de revisión por pares con paquetes ciegos, votos trazables y una rúbrica de fidelidad y utilidad. El coordinador reconstruye los resultados con una clave y puede reproducir las métricas y comparar desacuerdos en Python.

## Resultados y alcance

En el ejemplo guiado ficticio, la versión más votada corrige tres fallos e introduce dos. En un informe aparte, un juez mantuvo el mismo ganador al invertir las respuestas en 620 de 1.000 comparaciones de Arena-Hard y cambió de ganador en 140.




En el ejemplo gana la respuesta que ofrece consultar a la agencia. El motivo queda disponible junto a las notas de fidelidad y utilidad.

**Tecnologías** JavaScript · Python · Evaluación por pares

**Oferta investigada** Tecdata · AI Engineer Junior · IA Generativa y Agentes

Ver fallos que introduce la versión más votada  Proyecto · rcr-ia.eu/trabajo/contraprueba/  Ricardo Crespillo · LinkedIn