



Evaluación Completa del Chatbot por Pregunta

Selecciona un experimento para visualizar

eval_v1_asistente_rrhh



Resultados individuales por pregunta

| | pregunta | prompt_version | chunk_size | chunk_overlap | run_id | correctness | relevancia |
|----|--|-------------------|------------|---------------|-----------------------------------|-------------|------------|
| 1 | ¿Cuántos gramos de azúcar y de carbohidratos necesito para un maratón de 511? | v1_asistente_rrhh | 512 | 50 | 10a0b74030a1410301301720870c00a44 | 0 | |
| 2 | ¿Debo estirar antes o después del entrenamiento para mejorar la flexibilidad? | v1_asistente_rrhh | 512 | 50 | ee239f78f551445e8257df414afd76d2 | 1 | |
| 3 | ¿Debo estirar antes o después del entrenamiento para mejorar la flexibilidad? | v1_asistente_rrhh | 512 | 50 | 0b542b1940bc4fd58f42e3662092ed0c | 1 | |
| 4 | ¿Debo estirar antes o después del entrenamiento para mejorar la flexibilidad? | v1_asistente_rrhh | 512 | 50 | dbde8068cc4d420fb6908444ac0f2717 | 1 | |
| 5 | ¿Es suficiente entrenar fuerza 3 veces por semana para prevenir lesiones en triatlón? | v1_asistente_rrhh | 512 | 50 | 32b651b207874a6aa7b472e5f8b23a6b | 1 | |
| 6 | ¿Es suficiente entrenar fuerza 2 veces por semana para prevenir lesiones en triatlón? | v1_asistente_rrhh | 512 | 50 | 1e7bfe3613cc479abe0b13b0051d9f9d | 0 | |
| 7 | ¿Es suficiente entrenar fuerza 1 vez por semana para prevenir lesiones en triatlón? | v1_asistente_rrhh | 512 | 50 | ea4a7d3629f240b18df786b190408102 | 0 | |
| 8 | ¿Cómo debería ser mi taper de 3 semana antes de un maratón? | v1_asistente_rrhh | 512 | 50 | 315590e42c894f2191420d0ece65018f | 1 | |
| 9 | ¿Cómo debería ser mi taper de 2 semana antes de un maratón? | v1_asistente_rrhh | 512 | 50 | 4567b902998a46ceaa85740197ccaee | 0 | |
| 10 | ¿Cómo debería ser mi taper de 1 semana antes de un maratón? | v1_asistente_rrhh | 512 | 50 | 9c8de41f79d44c70823409744f8e518d | 1 | |
| 11 | ¿Cuántos días necesito para aclimatarme a 3000 m de altitud antes de una carrera de trail? | v1_asistente_rrhh | 512 | 50 | a6741dc8e5624b5292caba8eab21d950 | 0 | |



Selecciona criterio(s) para comparar

Criterios:

correctness coherence toxicity harmfulness relevance

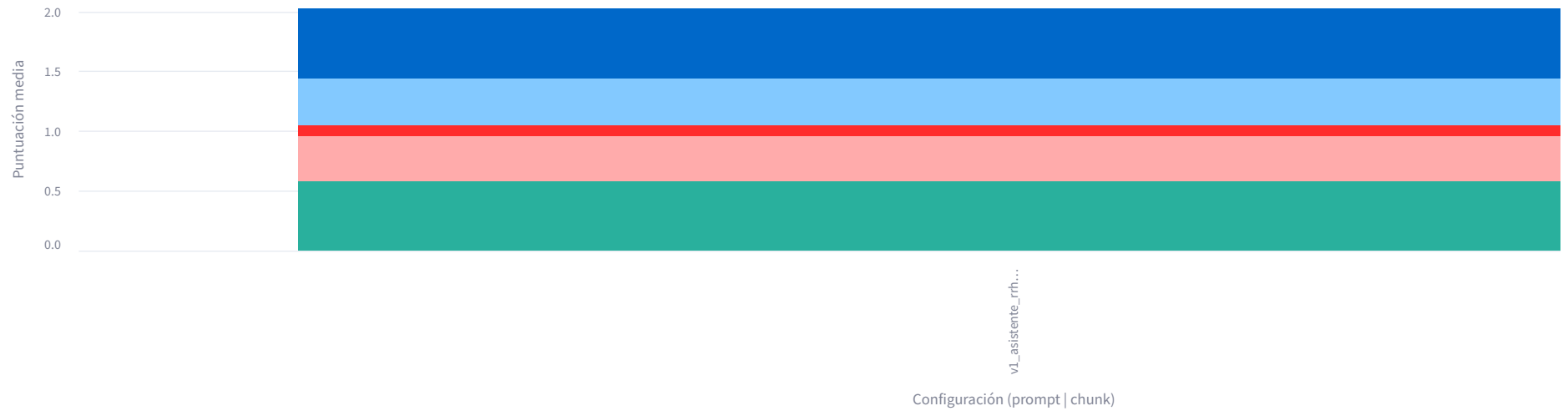


Desempeño agrupado por configuración

No results

```
at http://localhost:8501/static/js/7673.0828a05e.chunk.js:2:334183
at http://localhost:8501/static/js/7673.0828a05e.chunk.js:2:349353
at http://localhost:8501/static/js/7673.0828a05e.chunk.js:2:293066
at http://localhost:8501/static/js/7673.0828a05e.chunk.js:2:293268
at http://localhost:8501/static/js/7673.0828a05e.chunk.js:2:289446
at http://localhost:8501/static/js/7673.0828a05e.chunk.js:2:289688
```

```
at DL (http://localhost:8501/static/js/main.cc5b8325.js:2:2268181)
at http://localhost:8501/static/js/main.cc5b8325.js:2:3046033
```



☐ Mostrar razonamientos del modelo 🧠

Notas

- Todas las métricas se calculan con la clase `LabeledCriteriaEvalChain`.
- Asegúrate de **registrar** cada métrica como `<criterio>_score` en MLflow para que se cargue correctamente.
- Razonamientos* se esperan en el tag/param `eval_reasoning`; ajusta el nombre según tu pipeline.
- Si quieres comparar estadísticas adicionales (p. ej. desviación estándar) puedes extender la sección de agrupación.