OLA 01 PRIMERA OLA DE MEDICIÓN

Índice de visibilidad hospitalaria en IA

Costa Rica · Ola 1

Qué responden los principales modelos de inteligencia artificial al consultar sobre hospitales privados en Costa Rica.

Ventana de medición	**16 y 17 de agosto de 2026**
Panel de modelos	**9 configuraciones, 6 proveedores**
Preguntas formuladas	162 ítems; 1.568 respuestas por configuración
Respuestas analizadas	**14.112**
Hospitales evaluados	**6, más 1 control de sitio cerrado**

> Sextante mide la representación de los hospitales en inteligencia artificial. El índice combina presencia, posicionamiento y reputación en las respuestas. No evalúa resultados clínicos, seguridad del paciente ni calidad de atención.

RESUMEN LA VERSIÓN CORTA

Dos hospitales concentran las menciones de IA entre los seis estudiados

Todo lo que sigue sale de 14.112 respuestas dadas por nueve configuraciones de IA de seis proveedores, en español y en inglés.

64,4%

De las menciones a los seis hospitales corresponden a CIMA y Clínica Bíblica

93,5%

De las primeras menciones entre los seis recaen en esos mismos dos hospitales

9 de 9

Configuraciones describieron como abierto un hospital que cerró en 2018

01 **Dos hospitales encabezan la clasificación.**

CIMA obtiene 79,7 puntos y Clínica Bíblica 76,7. La Católica, en tercer lugar, alcanza 56,9. La diferencia entre el segundo y el tercero es de 19,8 puntos.

02 **Dos hospitales aparecen poco en preguntas abiertas.**

Hospital UNIBE aparece en el 5,4% y San Rafael Arcángel en el 1,9% de las preguntas abiertas. Cuando se consulta por ellos directamente, su nivel de detalle es 87,6 y 85,3 sobre 100. Son medidas distintas.

03 **Lo que los modelos equivocan varía más que lo que dicen.**

CIMA registra 99,2% de afirmaciones verificables correctas; Metropolitano, 74,7%. En este último, el 28,7% de las respuestas calificadas contiene algún error: casi tres de cada diez.

04 **Dejar que los modelos busquen cambia el orden.**

La Católica y Metropolitano intercambian posiciones entre las condiciones con y sin búsqueda. El cambio relativo es de 6,9 puntos. Esta comparación no mide una evolución en el tiempo.

PUBLICADO · 29 SEPT 2026 · COSTA RICA

Medición: 16 y 17 agosto 2026

Qué dicen los modelos cuando se consultan hospitales

Una consulta como «hospital en Costa Rica para un reemplazo de rodilla» produce una selección y una descripción. Sextante mide esa representación bajo un protocolo común.

Les hicimos las mismas preguntas a nueve configuraciones de IA, en español y en inglés, y registramos cada respuesta: cuáles hospitales nombraron, en qué orden, cómo los describieron y si los datos que afirmaron sobre ellos eran ciertos.

QUÉ NO ES ESTO

Una medida de calidad clínica

Nada de lo que hay aquí refleja resultados, tasas de complicación, mérito de acreditación ni calidad de atención. Un hospital puede ser excelente y estar mal representado por la IA, o al revés.

Lo que el paciente ve literalmente

Estas consultas llegan a los modelos de base. Las aplicaciones de consumo los envuelven en búsqueda adicional, personalización y reglas de seguridad específicas de salud. Esto es la disposición del modelo: un indicador reproducible, no una captura de pantalla.

Una medición de efectos comerciales

El estudio no observa decisiones de pacientes, reservas ni conversiones. Tampoco mide el efecto causal de cambios realizados por los hospitales en sus canales de información.

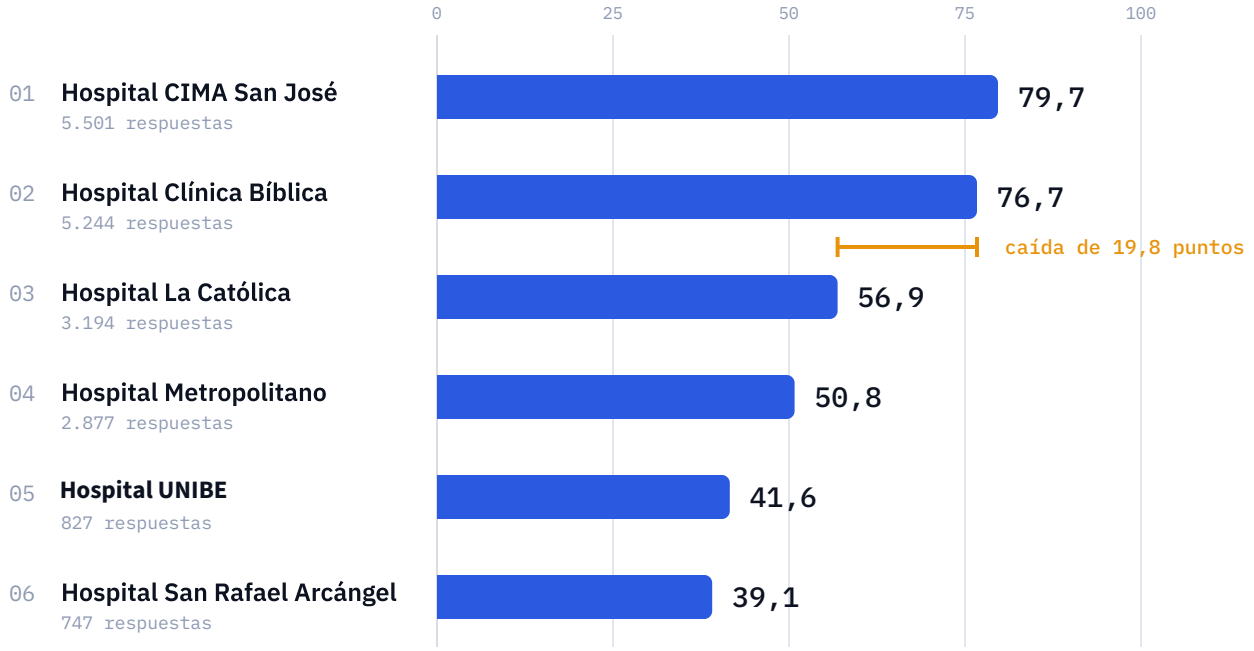
CÓMO SE DELIMITÓ EL PANEL

Se incluyeron seis hospitales privados con presencia relevante en las respuestas de IA según el criterio de selección del estudio. No es un censo de la oferta hospitalaria del país ni permite concluir que los hospitales fuera del panel tengan presencia nula.

Mencionar un hospital espontáneamente y describirlo cuando se pregunta por su nombre son capacidades distintas. Por eso el reporte combina frecuencia, detalle y exactitud, en lugar de interpretar una sola cifra como toda la visibilidad.

EXH 01 EL RANKING

Dos hospitales dominan la cima y después la tabla cae en picada



PUNTAJE SEXTANTE, DE 0 A 100

Más alto significa que los asistentes lo nombran y lo describen mejor

El índice combina presencia (40%), posicionamiento (30%) y reputación (30%). La clasificación principal corresponde a configuraciones con búsqueda habilitada. El posicionamiento incluye especificidad, exactitud factual e incidencia de errores.

INTERVALOS DE CONFIANZA AL 90%

CIMA: 79,4 a 79,9. Bíblica: 76,5 a 77,0. La Católica: 56,6 a 57,3.
Metropolitano: 50,4 a 51,2. UNIBE: 41,1 a 42,0. San Rafael: 38,7 a 39,5.

DE QUÉ ESTÁ HECHO EL PUNTAJE

¿Lo mencionan?
40% del puntaje

CIMA San José	69
Bíblica	62
La Católica	39
Metropolitano	32
UNIBE	20
San Rafael	18

¿Cómo lo describen?
30% del puntaje

CIMA San José	95
Bíblica	96
La Católica	77
Metropolitano	68
UNIBE	59
San Rafael	54

¿Hablan bien de él?
30% del puntaje

CIMA San José	79
Bíblica	77
La Católica	61
Metropolitano	58
UNIBE	53
San Rafael	52

EXH 02 CONCENTRACIÓN DE LA ATENCIÓN

CIMA y Clínica Bíblica reúnen el 64,4% de las menciones del panel

MENCIONES A LOS SEIS HOSPITALES: 64,4% PARA CIMA Y BÍBLICA

PRIMERAS MENCIONES ENTRE LOS SEIS: 93,5% PARA CIMA Y BÍBLICA

CIMA San José: 32,5% · Clínica Bíblica: 31,9% · La Católica: 18,2%

Metropolitano: 14,6% · Hospital UNIBE: 2,0% · San Rafael Arcángel: 0,7%

Ambas barras se normalizan entre los seis hospitales del panel. La primera distribuye todas sus menciones; la segunda, las primeras menciones. Si se toman todas las preguntas abiertas como base, CIMA y Clínica Bíblica suman 72,2% de primeras menciones. El estudio no midió acciones de pacientes.

64,4%

Se lo llevan los dos hospitales de arriba

2,7%

Corresponde a los dos últimos hospitales, sumados

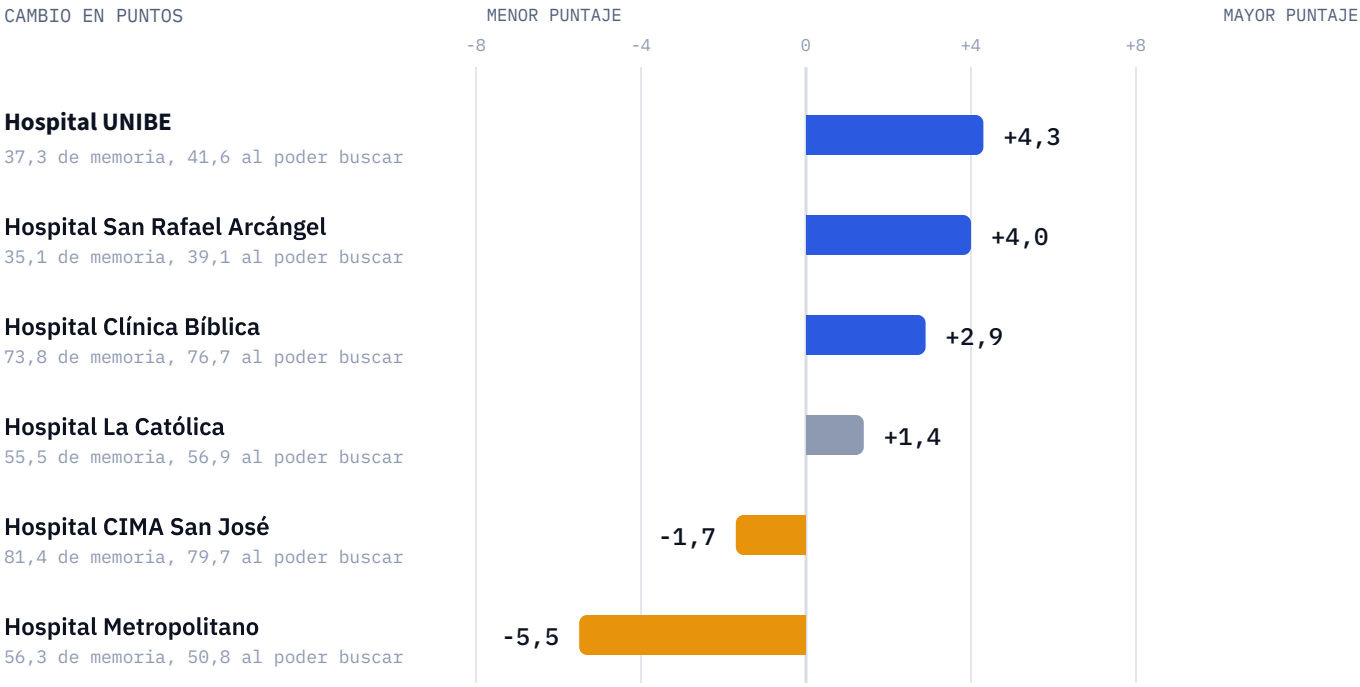
46x

Entre la mayor participación y la menor

La concentración muestra qué hospitales aparecen más en las respuestas de esta medición. No equivale a participación de mercado, preferencia de pacientes o calidad clínica. Tampoco permite afirmar que una respuesta del asistente entrene a la siguiente.

EXH 03 QUÉ CAMBIA CUANDO LOS MODELOS BUSCAN

Habilitar la búsqueda cambia los puntajes y algunas posiciones



CAMBIO EN EL PUNTAJE CUANDO EL MODELO PUEDE BUSCAR EN LA WEB

Se comparan los grupos de configuraciones con búsqueda habilitada y sin búsqueda, bajo el mismo banco de preguntas. Las barras muestran diferencias entre condiciones, no cambios en el tiempo ni efectos de una intervención hospitalaria.

4 de 6

Suben cuando los modelos pueden buscar

2 de 6

Bajan cuando pueden buscar

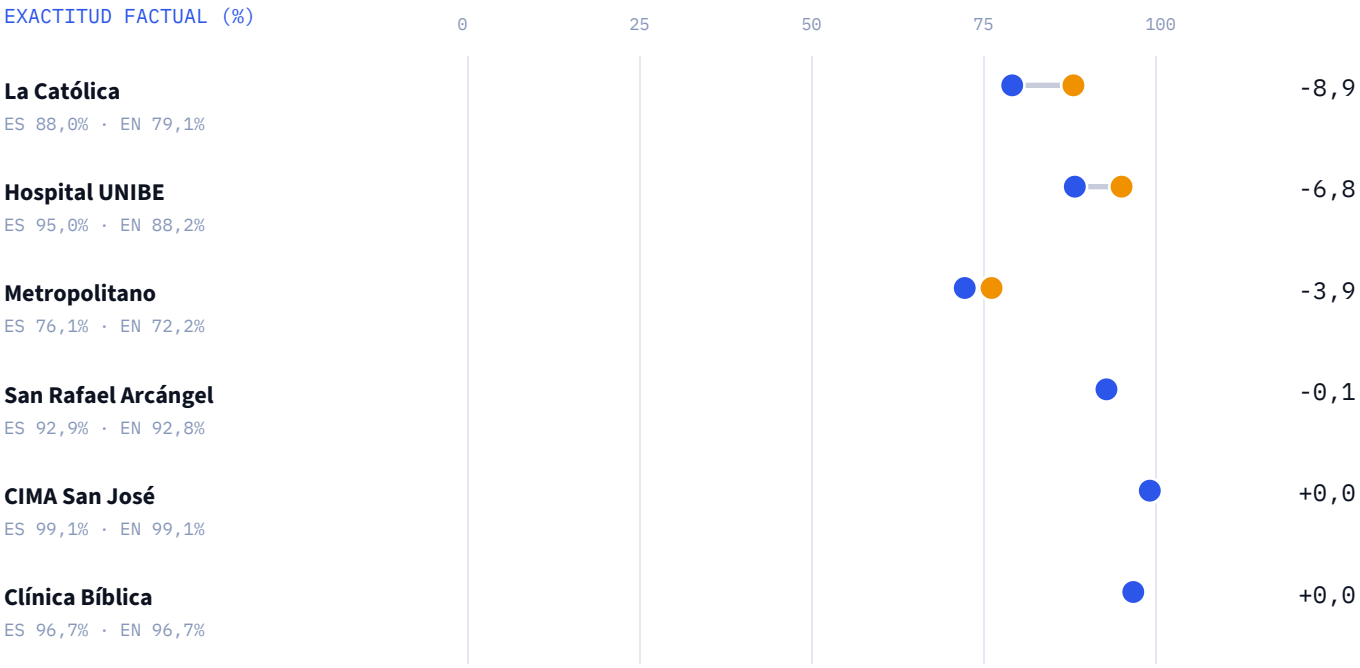
5,5

El movimiento más grande, en puntos

La dirección del cambio varía por hospital. Tener búsqueda habilitada tampoco implica usarla en cada respuesta. Estos resultados no identifican qué página o dato produjo la diferencia, ni demuestran que una modificación específica mejoraría el puntaje.

EXH 04 ESPAÑOL E INGLÉS

Los datos correctos también varían según el idioma



Base: afirmaciones verificables calificadas en el corte por idioma G5. La diferencia se expresa en puntos porcentuales. No es una medición de pacientes, turismo médico ni decisiones de atención.

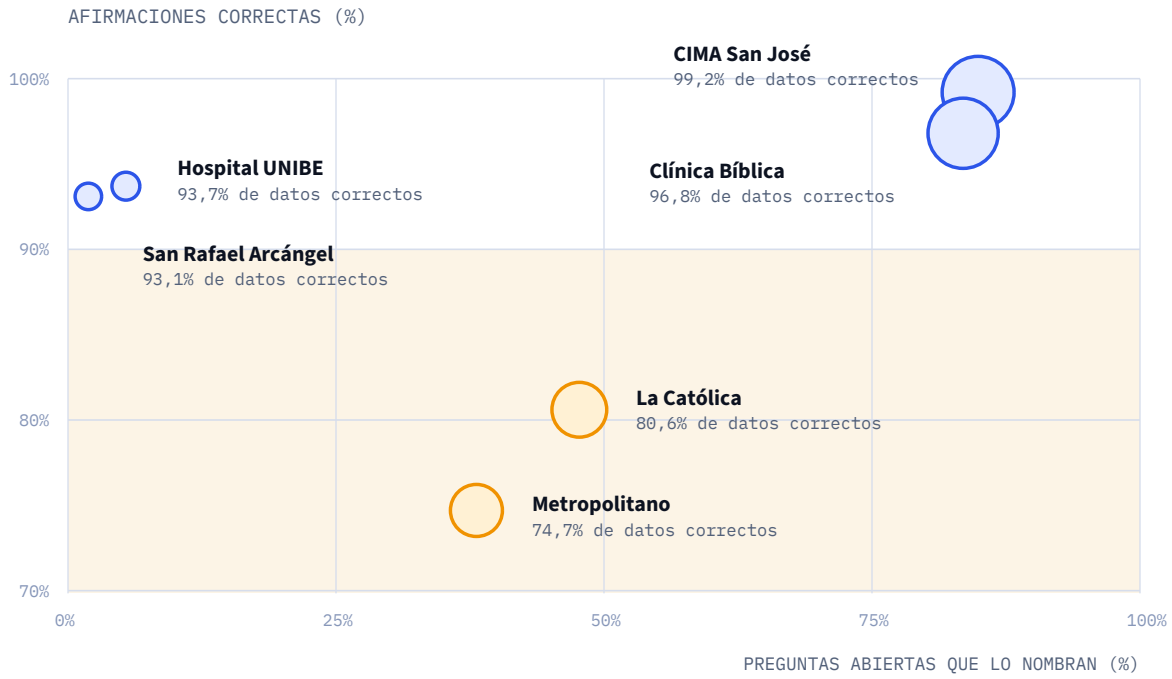
AFIRMACIONES CALIFICADAS	ESPAÑOL	INGLÉS
La Católica	1.494	1.917
Hospital UNIBE	337	398
Metropolitano	1.341	1.508
San Rafael Arcángel	551	649
CIMA San José	4.493	5.902
Clínica Bíblica	3.950	5.247

La Católica presenta la mayor diferencia de este corte: 88,0% en español y 79,1% en inglés.

G5 corresponde a un corte distinto de la tabla general G3: no debe agregarse ni compararse directamente con ella. La composición exacta de configuraciones de G5 no está especificada en los agregados disponibles.

EXH 05 EXACTITUD FRENTE A ALCANCE

Que hablen mucho de usted y que lo describan bien son dos problemas distintos



El tamaño del círculo representa respuestas analizadas. El eje de exactitud comienza en 70%.

El estudio calificó cuatro familias de afirmaciones contra fichas de referencia: acreditación, estado operativo, ubicación y propiedad. Precio, servicios y resultados quedaron fuera de estas tasas por el alcance de la verificación.

99%

La mejor exactitud del panel

75%

La peor exactitud del panel

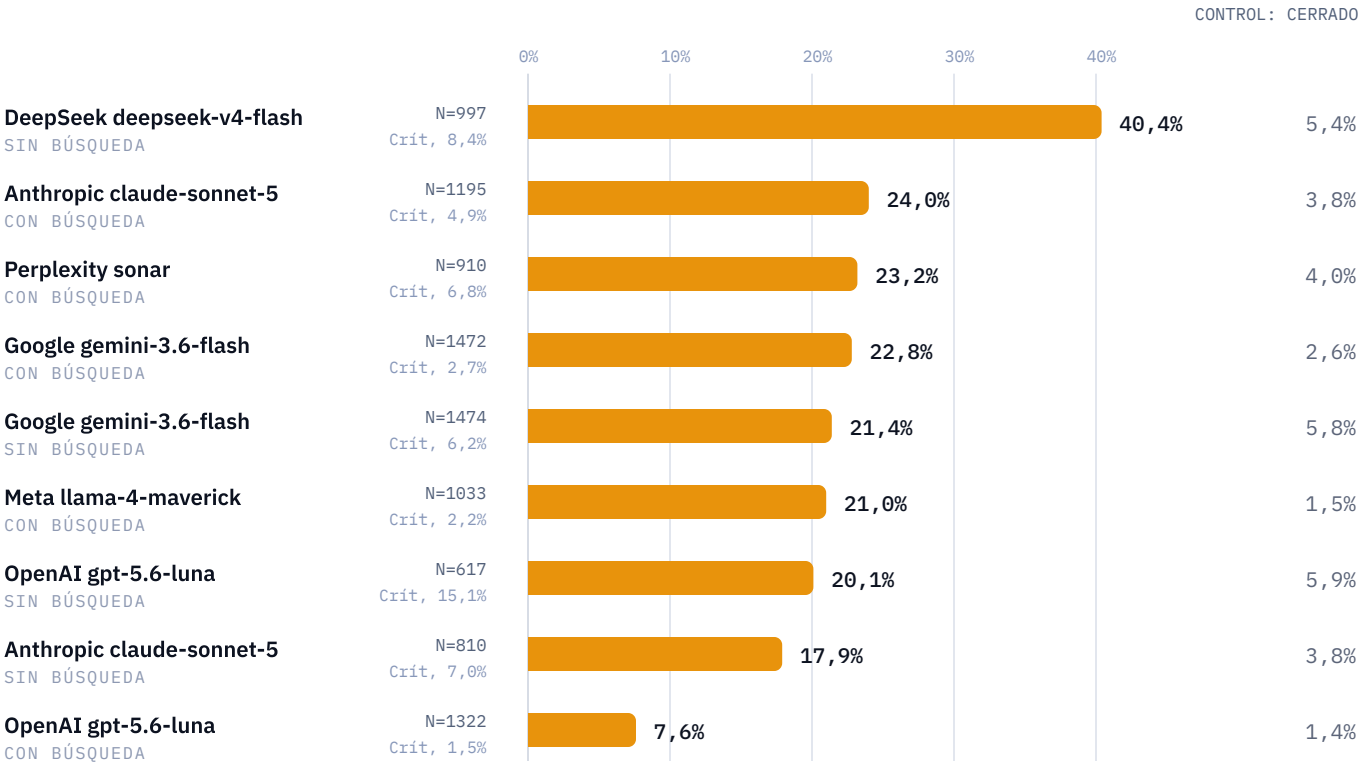
4

Familias de afirmaciones calificadas en el estudio

Una institución puede aparecer frecuentemente y ser descrita con errores; otra puede aparecer poco y tener alta exactitud. Las tasas usan bases diferentes: preguntas abiertas para menciones y afirmaciones verificables calificadas para exactitud.

EXH 06 LA PRUEBA DEL HOSPITAL CERRADO

Las nueve configuraciones presentaron como abierto un hospital cerrado



RESPUESTAS CON AL MENOS UN ERROR DE HECHO

Un hospital que dejó de operar en 2018 se incluyó como control. La columna derecha usa 1.568 respuestas por configuración. Las barras de error usan solo respuestas calificadas (N indicado junto al modelo); «Crít.» muestra la tasa de errores críticos sobre esa misma base.

Las configuraciones conservaron información desactualizada sobre el estado operativo del hospital de control. El hallazgo corresponde a las respuestas de los modelos. Revisar la vigencia de los datos institucionales es pertinente, pero este estudio no probó qué cambios en las fuentes corregirían las respuestas.

EXH 07 LÍNEAS DE SERVICIO

CIMA y Clínica Bíblica encabezan las ocho líneas evaluadas

	Bariátrica S1	Estética S2	Dental S3	Ortopedia S4	Oftalmo S5	Cardio S6	Fertilidad S7	Urología S8
CIMA San José	72	77	70	75	66	75	63	72
Bíblica	72	66	59	68	63	73	61	72
La Católica	52	43	33	55	48	55	34	55
Metropolitano	52	40	32	42	43	37	34	48
UNIBE	37	42	33	37	36	29	27	40
San Rafael Arcángel	36	33	26	32	38	29	25	39

PUNTAJE POR LÍNEA, DE 0 A 100

Fuerte · Sólido · Discreto · Débil

Los dos hospitales encabezan las ocho líneas de servicio, aunque el orden del resto varía por especialidad. Estos puntajes describen representación en IA; no evalúan resultados clínicos ni calidad de cada servicio.

CÓDIGO	LÍNEA DE SERVICIO	PROCEDIMIENTO INSIGNIA
S1	Bariátrica	manga gástrica
S2	Estética y plástica	aumento de senos
S3	Implantes dentales	implantes dentales All-on-4
S4	Ortopedia (cadera y rodilla)	reemplazo total de rodilla
S5	Oftalmología	cirugía de cataratas
S6	Cardiovascular electiva	angioplastia programada
S7	Salud femenina y fertilidad	tratamiento de FIV
S8	Urología y cirugía general	operación de hernia por laparoscopía

ACCIONES LECTURA INSTITUCIONAL

Tres áreas que los equipos hospitalarios pueden revisar

01 Revisar la información institucional verificable

Acreditación, estado operativo, ubicación y propiedad son las cuatro familias evaluadas. Mantener información oficial clara y vigente permite contrastar lo que publican las instituciones con lo que responden los modelos.

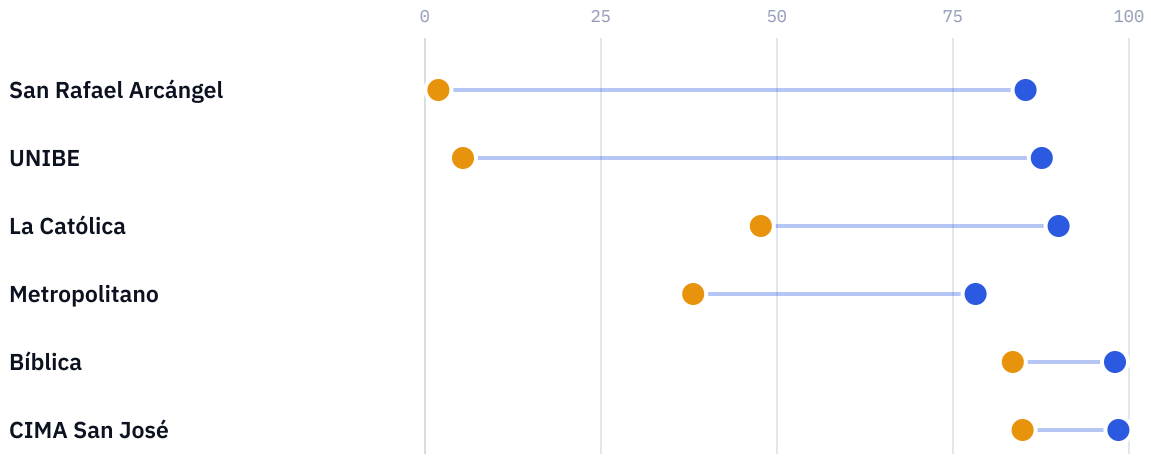
02 Distinguir reconocimiento de mención espontánea

UNIBE y San Rafael tienen puntajes de detalle de 87,6 y 85,3 sobre 100, pero aparecen en menos del 6% de las preguntas abiertas. Esta diferencia orienta qué revisar; no cuantifica un resultado comercial recuperable.

03 Revisar cada idioma con su propia base

El corte G5 muestra diferencias de exactitud entre español e inglés. Conviene contrastar ambos idiomas y sus denominadores. El estudio no identifica a los usuarios reales ni atribuye un idioma a un tipo de paciente.

DOS DIMENSIONES DE VISIBILIDAD



NARANJA: MENCIÓN ESPONTÁNEA (%)

AZUL: DETALLE (0 A 100)

Son indicadores con unidades distintas. Su resta no representa puntos porcentuales recuperables. Los hospitales pueden actualizar sus fuentes; el efecto en los modelos debe medirse de nuevo y no está garantizado por este estudio.

Nueve configuraciones, un banco de preguntas congelado, 14.112 respuestas codificadas

El banco contiene 162 ítems activos en español e inglés. Cada configuración produjo 1.568 respuestas, para un total de 14.112. Se incluyeron consultas abiertas, escenarios de paciente, preguntas por nombre, comparaciones directas y verificación de datos.

PROVEEDOR	MODELO	ACCESO A LA WEB
OpenAI	`openai/gpt-5.6-luna`	Sí
Google	`google/gemini-3.6-flash`	Sí
Anthropic	`claude-sonnet-5`	Sí
Perplexity	`perplexity/sonar`	Sí
Meta	`meta-llama/llama-4-maverick`	Sí
OpenAI	`openai/gpt-5.6-luna`	No
Google	`google/gemini-3.6-flash`	No
Anthropic	`anthropic/claude-sonnet-5`	No
DeepSeek	`deepseek/deepseek-v4-flash`	No

Las configuraciones con acceso a la web deciden cuándo buscar. La clasificación principal usa ese grupo. Las consultas se realizaron a los modelos bajo el protocolo del estudio y no reproducen literalmente las aplicaciones de consumo.

Un modelo juez, Claude Sonnet 5, codificó las respuestas con un esquema fijo. Las afirmaciones se contrastaron contra fichas de referencia hospitalarias. El índice combina tres pilares.

PILAR	PESO	LA PREGUNTA QUE RESPONDE
Presencia	**40%**	¿Lo mencionan del todo, y qué tan arriba?
Posicionamiento	**30%**	¿Lo que dicen es específico, y es exacto?
Reputación	**30%**	¿Hablan bien de él, y lo escogen sobre sus rivales?

Los intervalos al 90% provienen de 1.000 remuestreos bootstrap. Reflejan variación de muestreo, no toda la incertidumbre del modelo juez. La estabilidad de una posición no implica exactitud clínica ni certeza absoluta.

Ocho medidas para distinguir presencia, detalle y exactitud

HOSPITAL	UMR	FMR	SOV	ARD	FAR	HALL	NSS	H2H
Hospital CIMA San José	84,9	49,4	32,5	98,5	99,2	1,1	82,0	76,8
Hospital Clínica Bíblica	83,5	22,8	31,9	98,0	96,8	4,5	81,5	70,9
Hospital La Católica	47,7	2,8	18,2	90,0	80,6	23,5	70,6	41,0
Hospital Metropolitano	38,1	1,3	14,6	78,2	74,7	28,7	68,0	38,0
Hospital UNIBE	5,4	0,3	2,0	87,6	93,7	7,7	65,4	26,7
Hospital San Rafael Arcángel	1,9	0,6	0,7	85,3	93,1	10,0	63,2	30,7

CÓDIGO	COMO LO LLAMA ESTE INFORME	QUÉ CUENTA
UMR	**Lo nombran sin que se lo pregunten**	De cada pregunta abierta, con qué frecuencia aparece del todo
FMR	**Lo nombran de primero**	Con qué frecuencia es el primer hospital que aparece en la respuesta
SoV	**Participación en las menciones**	Porción de las menciones a los seis hospitales del panel
ARD	**Detalle que conocen si se les pregunta**	Cuánto saben de verdad cuando se les pregunta directamente por él
FAR	**Datos correctos**	De las afirmaciones que se pueden verificar, cuántas son ciertas
HALL	**Respuestas con error**	Proporción de respuestas calificadas con algún error
NSS	**Tono de la descripción**	Qué tan favorablemente se escribe sobre él
H2H	**Gana comparaciones directas**	Frente a un rival nombrado, a cuál de los dos escoge el modelo

Presencia se construye con las primeras cuatro. Posicionamiento, con la especificidad, la exactitud de los datos y el inverso de la tasa de error. Reputación, con el tono, el récord cara a cara y con cuánta fuerza se recomienda al hospital. Los pesos completos están en la nota metodológica que acompaña este informe.

NOTAS QUÉ NO PUEDE AFIRMAR ESTE ESTUDIO

Estas salvedades son parte del resultado

Cualquier cifra que se cite de este informe debería llevarlas consigo.

01 Modelos, no las apps que usa el paciente

Las consultas llegan a los modelos de base. Los productos de consumo agregan búsqueda, personalización y barandas específicas de salud que los modelos crudos no aplican.

02 Buscar es decisión del modelo

Las configuraciones con acceso a la web pueden buscar pero deciden cuándo. Son uniformemente capaces de buscar, no uniformemente buscadoras.

03 Los intervalos son más angostos que la verdad

Reflejan solo la variación de muestreo. Volver a leer respuestas idénticas coincide en el 72% a 76% de las afirmaciones individuales, así que la exactitud y las tasas de error cargan más incertidumbre de la que sugieren sus intervalos. Las diferencias pequeñas entre hospitales en esas dos medidas no deberían tratarse como reales.

04 Cuatro familias de afirmaciones dentro de este protocolo

Se contrastaron acreditación, estado operativo, ubicación y propiedad. Cerca del 42% de las afirmaciones, sobre precio, servicios o resultados, quedó fuera de la verificación por el alcance de las fichas de referencia.

05 Las parejas cara a cara las escogió un analista

Se escogieron diez parejas para enfrentar a cada hospital contra rivales cercanos. Ese récord carga el 30% del pilar de Reputación sobre una cantidad comparativamente pequeña de comparaciones.

06 Una configuración corrió sobre una herramienta de desarrollo

La configuración de Anthropic con acceso a la web corrió a través de una herramienta de línea de comandos para desarrolladores y no de una API de tipo consumo, y parte de sus respuestas cargan el encuadre de esa herramienta. Los registros afectados están marcados en los datos de base.

07 Esto es una fotografía, no una tendencia

Las versiones de los modelos, los índices de búsqueda y la presencia web de los hospitales cambian. Estos resultados describen una sola ventana de medición y hay que volver a medirlos para que se conviertan en tendencia.

08 No es una medida de calidad clínica

Un puntaje alto significa que un asistente habla bien de un hospital. No significa que ese hospital sea clínicamente mejor que otro.

NOTAS REPRODUCIBILIDAD

Cada cifra de aquí se puede rastrear hasta la respuesta que la produjo

Las respuestas crudas, las codificaciones estructuradas y las afirmaciones calificadas se conservan completas. Los hashes de abajo fijan los insumos que generaron esta ola, así que cualquier número de este informe se puede recalcular desde el corpus del que salió.

Hash del banco de preguntas	4e13a5c3d380e906
Hash del universo de hospitales	6b3222290b0cc3c5
Versión del pipeline	2.1.0
Respuestas analizadas	14.112
Esquema de codificación	Un solo modelo juez, una versión de prompt
Ventana de medición	16 y 17 de agosto de 2026

Sextante. Costa Rica, edición 1

Sebastián Urbina Cañas y Roberto Echeverria Cardenas
Autores del estudio
Roberto Echeverria Cardenas, coautor y consultor en inteligencia artificial aplicada a la salud.

Informe fuente emitido el 21 de agosto de 2026.
Edición en español para prensa: 28 de setiembre de 2026.
Mide representación en IA; no evalúa calidad clínica.