Clicandseo.

EDICIÓN APRIL 2026 · V3.0

ESTUDIO EMPÍRICO · CLICANDSEO RESEARCH

Lo que los LLMs premian (y lo que no)

Estudio empírico sobre 21.614 citaciones reales de URL en ChatGPT, Claude, Gemini y Perplexity. 3 sectores · 1200 dominios analizados con métricas SEO · 189 URLs scrapeadas y clasificadas con Gemini 3 Flash.

Carlos González Alba

Fundador de Clicandseo · Senior SEO Consultant

CLICANDSEO
CLICANDSEO.COM
APRIL 2026

ÍNDICE

Lo que vas a leer

APERTURA

ANÁLISIS EN PROFUNDIDAD

APLICACIÓN PRÁCTICA

CIERRE

01 · RESUMEN EJECUTIVO

Lo que vas a encontrar en este informe

Este informe analiza empíricamente cómo los principales modelos de lenguaje (LLMs) — **ChatGPT (OpenAI)**, **Claude (Anthropic)**, **Gemini (Google)** y **Perplexity** — citan URLs cuando responden a preguntas reales en sectores variados.

La muestra: **21.614 citaciones** de URL recogidas durante **59 días**, sobre **173 prompts únicos** ejecutados en **3 sectores** distintos. Para cada URL se han recogido métricas SEO de Ahrefs (Domain Rating, posición Google, tráfico orgánico) y se ha clasificado el contenido con *Gemini 3 Flash*.

21.614 CITACIONES ANALIZADAS	**3.007** URLS ÚNICAS	**1.204** DOMINIOS DISTINTOS	**4** LLMS COMPARADOS
8.660 EJECUCIONES TOTALES	**1.200** DOMINIOS CON DR AHREFS	**189** URLS SCRAPEADAS	**454** KEYWORDS EXTRAÍDAS

HALLAZGOS CLAVE

Cinco hallazgos que cambiarán tu estrategia LLM

→ **Anthropic, Google y OpenAI citan casi siempre URLs ya en TOP-3 de Google.** De las URLs citadas por cada LLM que también rankean en Google: Anthropic = 99.1% en top-3, OpenAI = 97.1%, Google = 97.1%. Perplexity baja a 66.4% — y su mediana de posición Google es 2.0 frente a 1.0 de los otros LLMs.

→ **Cada LLM construye su propio universo de fuentes.** El 96.5% de las URLs citadas a lo largo del estudio son exclusivas a un único LLM. Sólo 6 URLs (el 0.2%) son citadas por los 4 LLMs simultáneamente.

→ **Anthropic exige más autoridad de dominio que Perplexity.** Domain Rating mediano de las fuentes citadas: Anthropic = 80.0, OpenAI = 73.0, Google = 64.0, Perplexity = 51.0. Anthropic cita dominios con DR>70 en el 71.4% de los casos; Perplexity sólo en el 25.6%.

→ **El sector dicta el perfil de fuentes más que el LLM.** Anthropic cita fuentes oficiales/institucionales en proporciones muy distintas según el sector: Sector B: 11.5%, Sector C: 7.0%, Sector A: 63.9%

→ **Perplexity cita contenido sustancialmente más largo.** Word count mediano del contenido citado: Perplexity = 1550, resto de LLMs = 798. Perplexity cita contenido aproximadamente 1.9x más largo en mediana.

02 · METODOLOGÍA

Cómo se construyó este estudio

Origen de los datos

Las citaciones LLM provienen de **Clicandseo**, plataforma SaaS de monitorización de visibilidad de marca en LLMs. Para cada cliente se han ejecutado prompts personalizados contra los modelos de **ANTHROPIC**, **GOOGLE**, **OPENAI**, **PERPLEXITY** durante el periodo **2026-02-26 → 2026-04-25** (59 días).

Las métricas SEO (Domain Rating, posición Google, tráfico orgánico, keywords) se han obtenido vía API de **Ahrefs**. La clasificación semántica del contenido (tipo, tono, intención, audiencia) se ha realizado con **Gemini 3 Flash** de Google DeepMind, usando structured output con JSON schema para garantizar consistencia. La detección de schema.org usa un extractor que combina JSON-LD, microdata y RDFa con aplanamiento de `@graph`.

Cobertura del estudio

MÉTRICA	VALOR
Sectores analizados (anonimizados)	**3**
Cuentas / clientes incluidos	**4**
LLMs observados	**4**
Prompts únicos ejecutados	**173**
Ejecuciones LLM totales (prompt × LLM × día)	**8.660**
Citaciones URL totales analizadas	**21.614**
URLs únicas citadas	**3.007**
Dominios únicos citados	**1.204**

MÉTRICA	VALOR
Dominios con Domain Rating de Ahrefs (cobertura 99.7%)	**1.200**
URLs scrapeadas (estructura HTML, schema)	**189**
URLs con métricas SEO completas (UR, tráfico, KW)	**200**
Keywords orgánicas extraídas (top-3 por URL)	**454**
URLs clasificadas semánticamente con Gemini 3 Flash	**189**

Sectores incluidos (anonimizados)

Sector A — Fintech

Sector B — Salud (Reproducción asistida)

Sector C — Salud (Depilación láser, ES y PT)

Los clientes se han anonimizado como 'Sector A/B/C'. Se reportan cifras agregadas, nunca específicas de un cliente que permitan re-identificarlo.

Tests estadísticos

Las correlaciones se calculan con **Spearman** (rango, no asume normalidad — apropiado para distribuciones de citaciones sesgadas). Las diferencias entre grupos se contrastan con **Kruskal-Wallis** (más de 2 grupos), **Mann-Whitney U** (2 grupos) o **chi-cuadrado** de independencia para variables categóricas. Para el test multidimensional con embeddings se usa **cosine similarity** sobre vectores de Gemini. Se considera significativo $p < 0.05$.

Cómo interpretar los p-values y tamaños de efecto. Reportamos p-value *y* coeficiente de Spearman ρ (tamaño del efecto). Para distinguir señal real de ruido: $|\rho| > 0.30$ con $p < 0.001$ son hallazgos sólidos; $|\rho|$ 0.10–0.30 con $p < 0.05$ son señales modestas que se replicarían con probabilidad razonable; resultados con $p \in [0.01, 0.05]$ deben tomarse con cautela porque, dado el número de tests realizados (~30), algún falso positivo es esperable sin corrección de múltiples comparaciones. Cuando un test reporta $p > 0.05$ (ej. FAQ schema, headings-pregunta), eso NO equivale a afirmar que el factor no funciona: significa que con esta muestra no encontramos evidencia suficiente, pero efectos pequeños o medianos podrían quedar indetectables. **Ausencia de evidencia ≠ evidencia de ausencia.**

Sobre el alcance de las conclusiones. Las URLs analizadas en bloques estructurales son las top-50 más citadas por cliente. Por tanto, los tests responden a *"qué diferencia las URLs muy citadas entre sí"*, no a *"qué*

hace que una URL sea citada vs no citada en absoluto". Cuando comparemos con estudios de mercado que analizan millones de páginas (citadas y no citadas), conviene tener presente que la pregunta de fondo puede ser distinta. Resultados como los del bloque 3.8 bis se interpretan dentro de este scope.

Hallazgos derivados exclusivamente de los datos. Las correlaciones se calculan con Spearman (no asume normalidad) y los tests de diferencia entre grupos con Kruskal-Wallis o Mann-Whitney. Las clasificaciones de tipo/tono/intent del contenido se obtienen con Gemini 3 Flash (Google DeepMind). Correlación ≠ causación: las relaciones halladas son asociativas.

EMPIEZA AHORA

¿Aparece tu marca cuando el LLM *no* te nombra?

Antes de leer 30 páginas de hallazgos, aplícalos a tu propia marca. Clicandseo monitoriza cada día tu visibilidad en ChatGPT, Claude, Gemini y Perplexity, y te dice qué URLs te citan, qué dominios te ganan y dónde estás perdiendo cuota frente a tus competidores.

→ **Prompts personalizados** contra los 4 LLMs principales, con periodicidad diaria.

→ **Share of voice** frente a competidores, en cada prompt y en agregado.

→ **URLs que te citan** y URLs que citan a tu competencia, listadas y descargables.

Empieza gratis →

7 días gratis · Plan freemium permanente · Sin tarjeta · app.clicandseo.com

PARTE I

La estructura del mercado LLM

¿Cómo se reparten las citaciones entre los modelos?

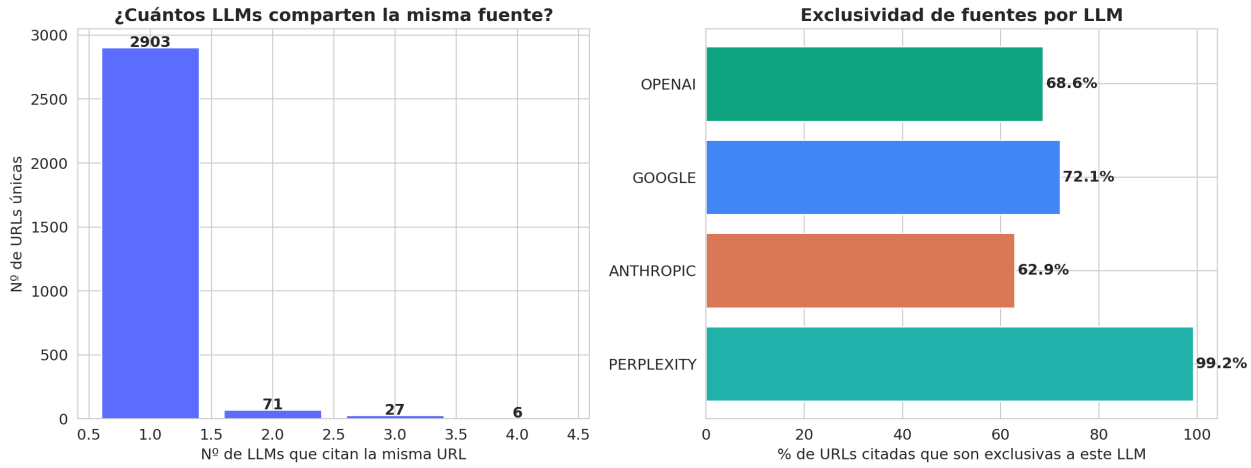
Antes de hablar de qué optimizar para aparecer en LLMs, hay que entender cómo se reparten las citaciones entre los cuatro grandes (Anthropic, Google, OpenAI y Perplexity).

Esta primera parte responde a cuatro preguntas fundamentales:

- ¿Comparten fuentes los LLMs?
- ¿Qué tipo de fuente prefiere cada uno?
- ¿Qué tan concentrado es el mercado de fuentes de cada LLM?
- ¿Son estables las citaciones día a día?

1.1 — Cada LLM construye su propio universo de fuentes

Sobre las **3.007 URLs únicas** que aparecen citadas a lo largo del estudio, sólo **6 URLs** son citadas por los cuatro LLMs simultáneamente. El **96.5%** aparecen en uno y sólo uno.



Izquierda: distribución del nº de LLMs que citan cada URL. La barra dominante es 'sólo 1 LLM'. Derecha: % de URLs que cita cada LLM y son exclusivas suyas.

Perplexity opera prácticamente en su propio universo: **99.2%** de las URLs que cita no las cita ningún otro LLM. En el resto: Anthropic 62.9%, OpenAI 68.6%, Google 72.1%.

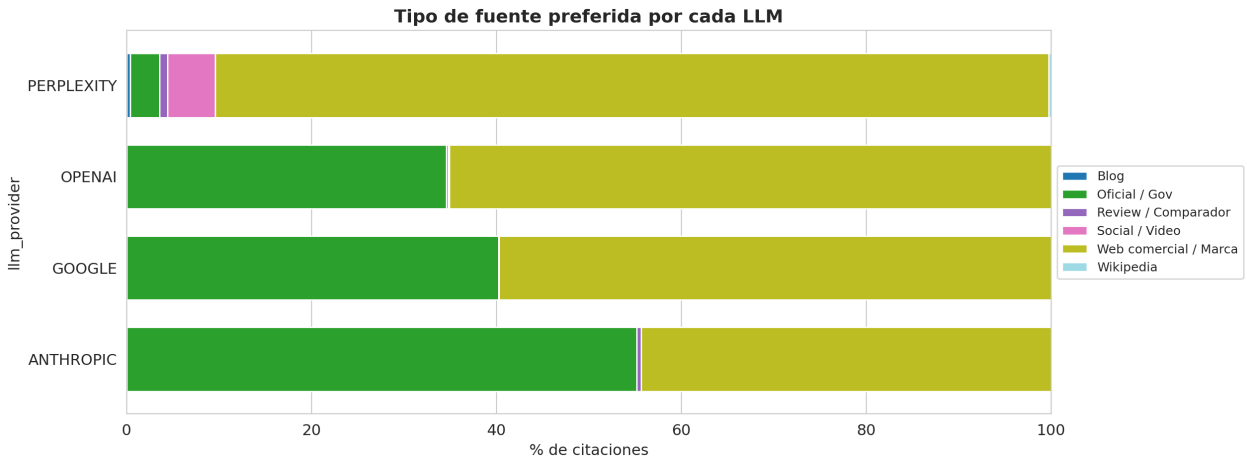
La idea de 'optimizar para LLMs' como una sola estrategia no se sostiene con los datos. Cada modelo aplica criterios distintos y, por tanto, exige un enfoque distinto.

IMPLICACIÓN PRÁCTICA

Si tu marca depende de visibilidad LLM, identifica primero qué modelo usa más tu público objetivo y prioriza optimizar para sus criterios concretos. Repartir esfuerzos entre los cuatro por igual rinde poco en ninguno. **Dicho esto, si hay algo que parece funcionar para la mayoría de LLMs es seguir haciendo SEO clásico** — los datos lo confirman en los siguientes bloques.

1.2 — Tipo de fuente preferido por cada LLM

Clasificamos cada dominio citado en una de seis categorías heurísticas: *oficial / gov* (administraciones públicas, BOE, agencias tributarias, ministerios), *web comercial / marca* (sitios de empresa, blogs corporativos), *review / comparador* (Trustpilot, Capterra, comparadores), *social / video* (YouTube, redes sociales), *Wikipedia*, y *blog*. El % de citaciones por categoría revela la 'personalidad editorial' de cada LLM.



Stacked bar horizontal: % de citaciones de cada LLM partidas por tipo de fuente.

Dominios oficiales / gov por LLM:

- **ANTHROPIC**: oficial = 55.2%, comercial = 44.3%
- **GOOGLE**: oficial = 40.2%, comercial = 59.8%
- **OPENAI**: oficial = 34.6%, comercial = 65.1%
- **PERPLEXITY**: oficial = 3.2%, comercial = 90.2%

Anthropic muestra el sesgo más fuerte hacia fuentes oficiales/institucionales (>50% en el agregado), seguido por Google y OpenAI. Perplexity es la excepción radical: el 90%+ de sus citaciones van a contenido comercial/marca, y apenas un 3-4% a fuentes oficiales. Esto NO es un artefacto de la muestra: el patrón se mantiene en todos los sectores analizados (lo veremos en la Parte IV).

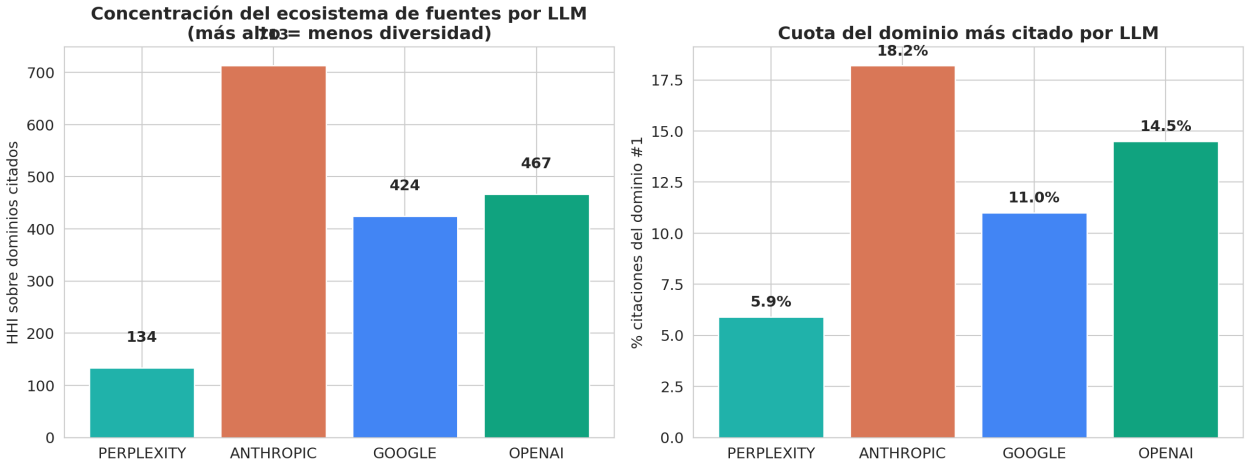
El motivo probable: Anthropic, Google y OpenAI emplean modelos con preferencia por fuentes con alto pedigrí editorial; Perplexity, al estar entrenado para responder con multitud de fuentes citables (típicamente 5-9 por respuesta), tiende a barrer el ecosistema comercial donde abundan listicles, comparadores y reviews.

IMPLICACIÓN PRÁCTICA

Si tu marca aspira a aparecer en Anthropic/OpenAI/Google AI, tu primera palanca es que el contenido sobre tu sector esté presente en (o referenciado desde) fuentes oficiales. Si tu sector NO tiene equivalente .gob potente, el techo en estos LLMs es bajo — y la apuesta natural es Perplexity, donde el contenido comercial bien estructurado tiene 20× más cuota.

1.3 — Perplexity diversifica fuentes 3-4× más que el resto

El índice Herfindahl-Hirschman (HHI) mide la concentración de un mercado: valores altos significan que pocos jugadores acaparan la mayoría. Aplicado al ecosistema de dominios citados por cada LLM, los resultados son extremos.



Izquierda: HHI sobre dominios citados por cada LLM. Cuanto más alto, menos diversificada es la fuente. Derecha: cuota de mercado del dominio más citado por LLM.

HHI sobre dominios: Perplexity = **133.6**, Anthropic = **713.0**, OpenAI = **466.9**, Google = **424.3**.

Perplexity cita 966 dominios distintos; el resto entre 99 y 175. En Perplexity, el dominio más citado tiene un **5.9%** de cuota; en Anthropic, OpenAI y Google ese top-1 ya supera el 17%.

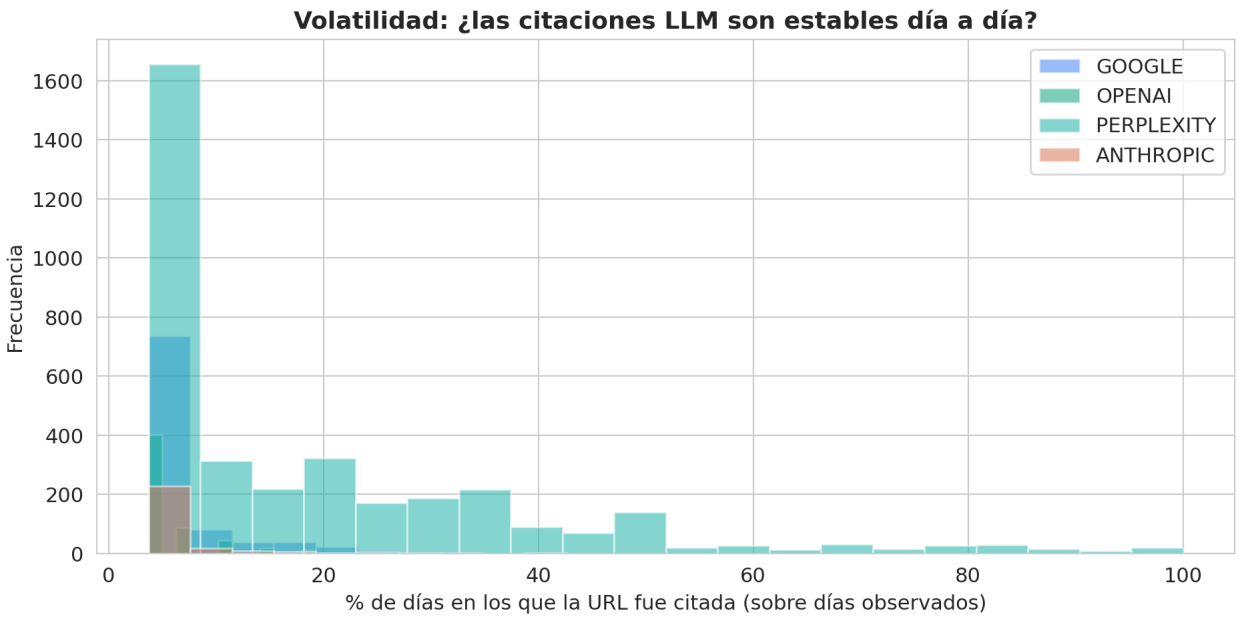
IMPLICACIÓN PRÁCTICA

Para una marca que aspira a entrar en el ranking de citaciones de un LLM, Perplexity es donde menos competencia tiene cada espacio: el universo de fuentes está más fragmentado y cabe más jugadores.

Importante: hay un posible sesgo metodológico a tener en cuenta. Perplexity, por diseño de producto, *siempre* enlaza URLs en sus respuestas (es su patrón de uso característico, citar fuentes explícitamente). El resto de LLMs no siempre adjuntan URLs; muchas veces responden con conocimiento interno sin citar fuente. Eso significa que Perplexity tiene una superficie de citación mucho mayor por construcción, lo que infla artificialmente la diversidad observada en el HHI. La diferencia real entre LLMs probablemente sea menor que la que muestran los datos brutos.

1.4 — La volatilidad de las citaciones es altísima

Sobre 27 días observados, miramos qué % de los días una misma URL aparece citada por una misma combinación (prompt, LLM). Una URL 'estable' la definimos como aquella que aparece en >70% de los días.



Histograma del % de días observados en los que cada URL aparece citada. La mayor parte de la masa está en valores bajos: las citaciones son volátiles.

Mediana de % de días en los que una URL aparece, por LLM:

- GOOGLE: **3.7%**
- OPENAI: **3.7%**
- PERPLEXITY: **11.1%**
- ANTHROPIC: **3.7%**

El % de citaciones 'estables' (>70% de los días) varía entre el 0.0% y 3.7% según el LLM. La mayoría de URLs aparecen sólo en una o dos respuestas y desaparecen.

IMPLICACIÓN PRÁCTICA

Un single-shot screenshot de 'qué cita el LLM' es insuficiente. La monitorización debe ser continua para detectar patrones reales por encima del ruido aleatorio diario. Las URLs verdaderamente "establecidas" en un LLM son una minoría.

Si quieres ver cómo se comportan las citaciones LLM en tu propio sector día a día, herramientas como **Clicandseo** automatizan esta monitorización continua sobre los 4 LLMs principales y te muestran qué URLs son recurrentes y cuáles aparecen y desaparecen por el ruido del modelo. Tienes 7 días gratis y plan freemium permanente en app.clicandseo.com.

PARTE II

La conexión SEO ↔ LLM

¿La autoridad del dominio y la posición Google predicen citaciones?

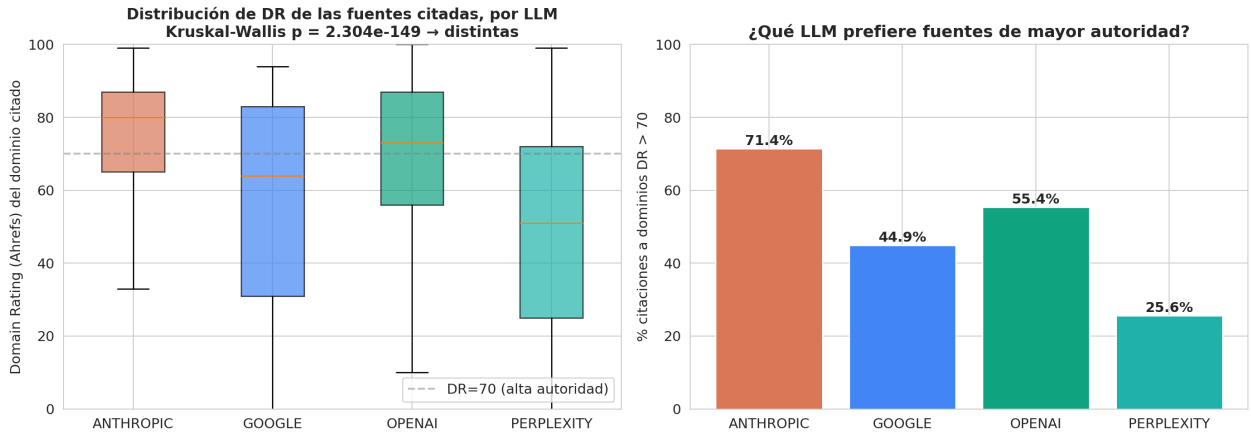
Para esta parte cruzamos las citaciones LLM con datos SEO de Ahrefs: Domain Rating (DR) de **1200 dominios** y métricas de posición Google + tráfico orgánico de las **200 URLs más citadas**. El SEO clásico ¿predice la visibilidad LLM?

LA RESPUESTA CORTA

Sí, es el factor que tienen en común con un altísimo porcentaje **3 de los 4 LLMs principales.**

2.1 — Anthropic exige más autoridad de dominio que Perplexity

Domain Rating (DR) es una métrica de Ahrefs (0-100) que mide la fortaleza del perfil de backlinks de un dominio. DR >70 = autoridad alta, <30 = baja.



Izquierda: distribución del DR de los dominios citados por cada LLM (boxplot). Derecha: % de citaciones que van a dominios DR>70.

DR mediano de las fuentes citadas: Anthropic = **80.0**, OpenAI = **73.0**, Google = **64.0**, Perplexity = **51.0**.

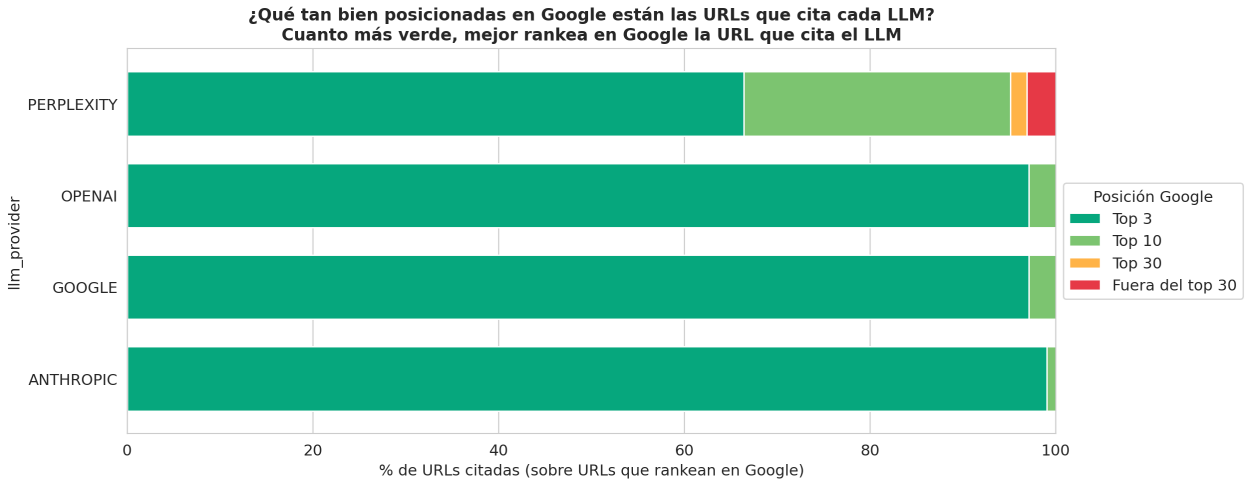
Anthropic cita dominios con DR>70 en el **71.4%** de los casos. Perplexity sólo en el **25.6%**. Test Kruskal-Wallis: H = 690.59, p = 2.304e-149 → diferencias estadísticamente significativas entre LLMs.

IMPLICACIÓN PRÁCTICA

Si tu dominio tiene DR <50, tu probabilidad de aparecer en Anthropic es mucho menor que en Perplexity. La autoridad de dominio tradicional sigue siendo una palanca relevante — sólo que más para unos LLMs que para otros.

2.2 — Anthropic, Google y OpenAI citan casi siempre URLs ya en TOP-3 de Google

Para cada URL citada que también rankea en Google, sacamos su mejor posición (la KW en la que mejor rankea). Es decir, si una URL tiene una KW en posición 1 y otra en posición 8, contamos posición 1.



% de URLs citadas (sobre las que rankean en Google) repartidas por bucket de posición Google. Verde = mejor posición.

% de URLs citadas que están en el top-3 de Google: Anthropic **99.1%**, OpenAI **97.1%**, Google **97.1%**. Perplexity baja a **66.4%** y su mediana de posición Google es **2.0** frente a **1.0** de los demás LLMs.

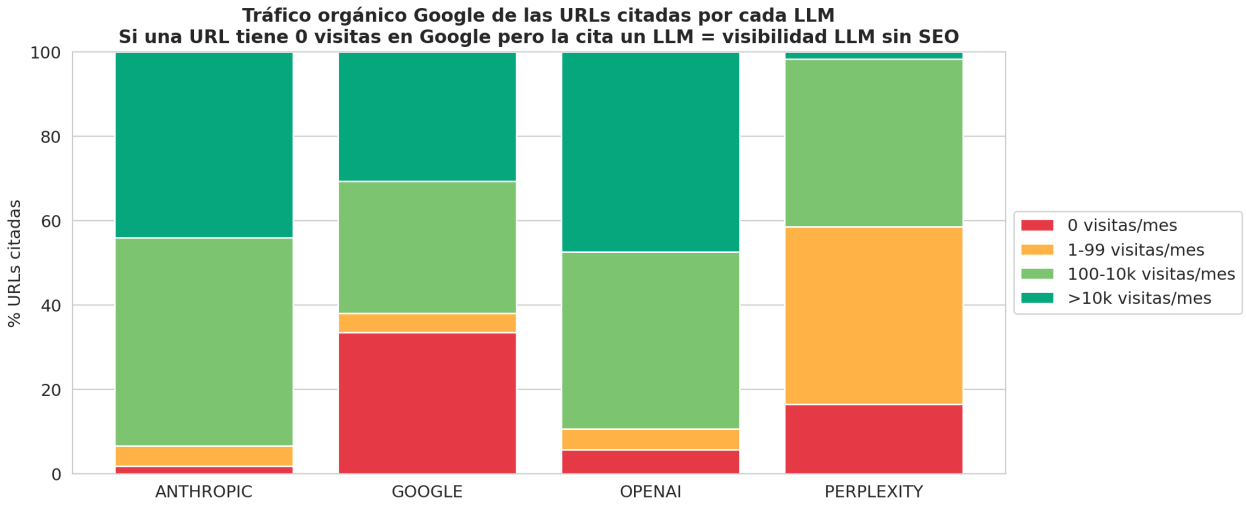
Esto es uno de los hallazgos más fuertes del estudio: **la visibilidad LLM en Anthropic/OpenAI/Google está casi totalmente determinada por estar en el top-3 de Google primero**.

IMPLICACIÓN PRÁCTICA

Para 3 de los 4 LLMs grandes, optimizar para LLMs ES optimizar para top-3 de Google. El SEO clásico no muere — se convierte en condición previa. Perplexity es la excepción: su tolerancia a posiciones medio-bajas es la única ventana de entrada para contenido nuevo o de marcas medianas.

2.3 — El 19% de las URLs citadas tiene cero tráfico orgánico

Cruce: ¿cuánto tráfico orgánico tiene la URL que cita el LLM? Una URL con 0 visitas en Google pero citada por un LLM = un caso interesante: visibilidad LLM sin SEO clásico.



Distribución del tráfico orgánico Google de las URLs citadas, por LLM. Rojo = 0 visitas/mes.

% de citaciones a URLs con 0 tráfico orgánico: **PERPLEXITY 16.4%**, **ANTHROPIC 1.8%**, **GOOGLE 33.5%**, **OPENAI 5.7%**

Mediana de tráfico orgánico de las URLs citadas: **PERPLEXITY 43**, **ANTHROPIC 7.020**, **GOOGLE 953**, **OPENAI 5.460**. Hay LLMs (especialmente Perplexity) que citan URLs casi sin tráfico Google. En el extremo opuesto, Anthropic cita URLs con tráfico mediano alto.

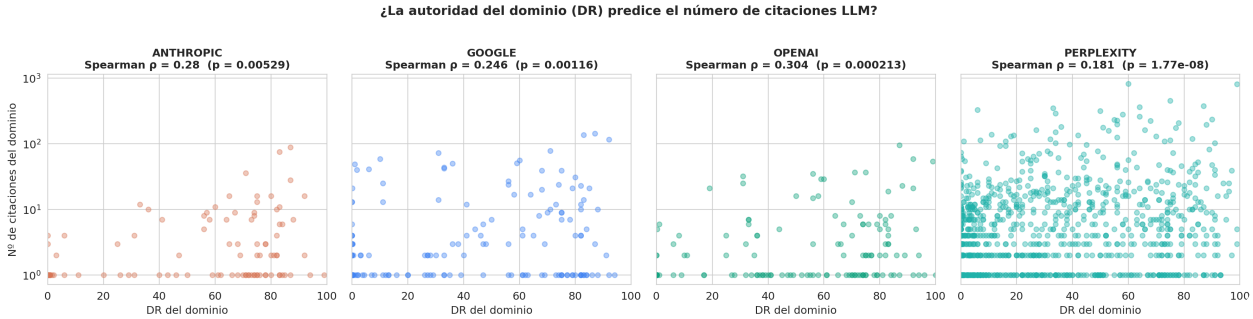
IMPLICACIÓN PRÁCTICA

La existencia de URLs con 0 tráfico orgánico citadas por LLMs sugiere que los LLMs NO son meramente cachés del SERP de Google. Probablemente tienen señales propias (estructura, freshness, schema, anchor text) que les permiten valorar URLs que el algoritmo de ranking de Google aún no ha posicionado.

Dos matices importantes: (1) este hallazgo depende de los datos de tráfico orgánico de un tercero — *Ahrefs* — que estima visitas a partir de su propio modelo, no son visitas reales medidas por Google Analytics; cabe la posibilidad de que esas URLs sí tengan algo de tráfico no detectado por Ahrefs. Y (2) **el porcentaje de URLs con 0 tráfico orgánico es minoritario** en todos los LLMs — la regla general sigue siendo que las URLs citadas tienen tráfico orgánico real. El hallazgo abre una hipótesis interesante, pero no contradice la regla general de la Parte II: el SEO clásico sigue siendo el mejor predictor.

2.4 — El DR predice citaciones, pero la correlación es moderada

Spearman entre DR del dominio y nº de citaciones que recibe ese dominio, por LLM. Spearman ρ varía de -1 a +1; ρ>0.3 = correlación moderada, ρ>0.5 = fuerte.



Scatter por LLM: cada punto es un dominio. Eje X = DR. Eje Y = nº de citaciones (log).

Correlación de Spearman entre DR y nº de citaciones, por LLM: **ANTHROPIC ρ = +0.28** (p = 0.00529, n = 98), **GOOGLE ρ = +0.246** (p = 0.00116, n = 171), **OPENAI ρ = +0.304** (p = 0.000213, n = 144), **PERPLEXITY ρ = +0.181** (p = 1.77e-08, n = 956). **Todas las correlaciones son positivas y estadísticamente significativas (p<0.05)**, pero ninguna llega a fuerte. Esto significa que el DR es uno de varios factores, no el determinante.

Otros factores que cabría explorar (fuera del alcance de este estudio por requerir más datos): edad del dominio, frecuencia de actualización, calidad de los anchor texts entrantes, presencia en Wikipedia, frescura del contenido.

IMPLICACIÓN PRÁCTICA

Construir DR sigue mereciendo la pena, pero no es la palanca milagrosa. Una marca con DR 50 y contenido excelente, fresco y bien estructurado puede competir con marcas DR 80 que tienen contenido obsoleto.

MIDE LO QUE IMPORTA

El SEO se mide ahora en los *LLMs*.

Acabas de ver que el 97-99% de las URLs citadas por Anthropic, OpenAI y Google ya están en top-3 de Google. La pregunta no es si el SEO sirve, es si lo estás midiendo bien. Los rankings de Google ya no cuentan toda la historia.

- → **Cruza posición Google con citaciones LLM** en una sola vista, para ver dónde rankeas pero no te están citando.
- → **Detecta gaps de visibilidad**: prompts donde tu competencia aparece y tú no.
- → **Auditoría AI Overviews**: monitoriza también tu presencia en los snippets de Google AI Mode.

Pruébalo 7 días gratis →

Sin tarjeta · Cancela cuando quieras · **Plan freemium permanente** · app.clicandseo.com

PARTE III

Anatomía del contenido citado

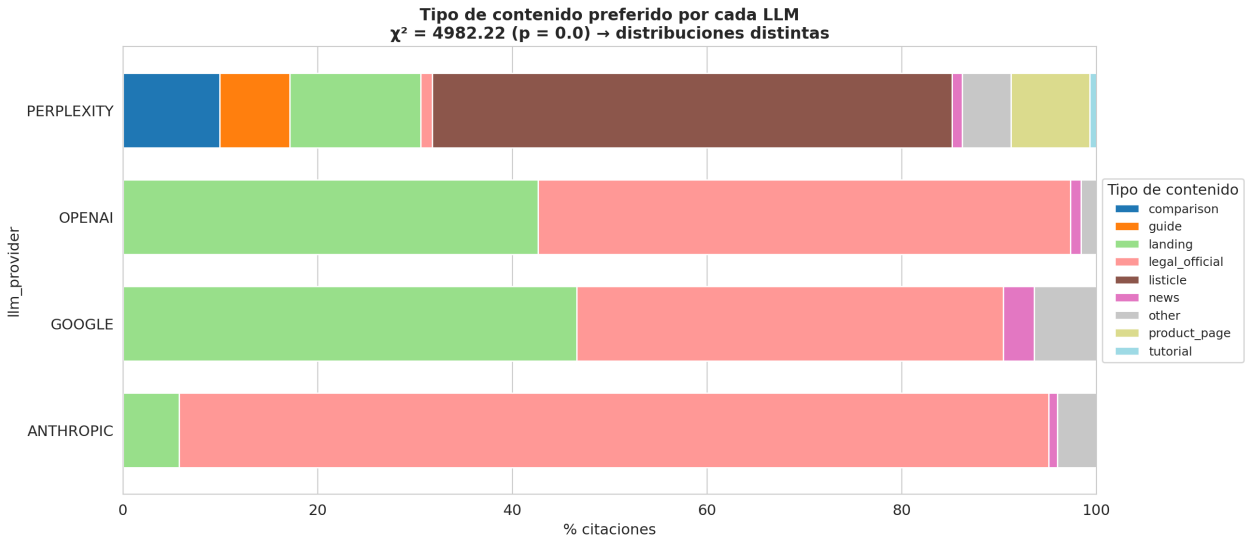
¿Qué tienen en común las páginas que ganan visibilidad LLM?

Para esta parte se scrapearon las **189 URLs más citadas** y se clasificó cada una semánticamente con **Gemini 3 Flash** (tipo de contenido, tono, intención primaria, audiencia, branded vs editorial). Cruzando esa clasificación con qué LLM cita cada URL aparecen patrones muy claros.

3.1 — Cada LLM tiene un tipo de contenido favorito

Distribución del tipo de contenido (clasificación Gemini) que cita cada LLM. Los tipos detectados:

- **listicle** — listas numeradas o categorizadas (top 10, mejores X, X opciones).
- **guide** — guía exhaustiva sobre un tema, formato largo, tono educativo.
- **comparison** — comparativa A vs B (o varias opciones), con tablas o secciones por opción.
- **review** — opinión / análisis sobre un producto, servicio o marca.
- **definition** — qué es X, contenido enfocado a explicar un concepto puntual.
- **case_study** — caso real con resultados medibles (antes/después, métricas).
- **product_page** — ficha de producto o servicio (características, precio, CTA).
- **tutorial** — cómo hacer X paso a paso, instructivo.
- **landing** — página de captación con foco en conversión, CTA destacado.
- **legal_official** — regulación, ley, normativa o documento oficial / institucional.
- **news** — noticia con fecha reciente.



Composición del % de citaciones por tipo de contenido, por LLM. Test χ² para verificar si las distribuciones son distintas entre LLMs.

Tipo dominante por LLM:

- **ANTHROPIC**: legal_official (89.3%)
- **GOOGLE**: landing (46.6%)
- **OPENAI**: legal_official (54.7%)
- **PERPLEXITY**: listicle (53.4%)

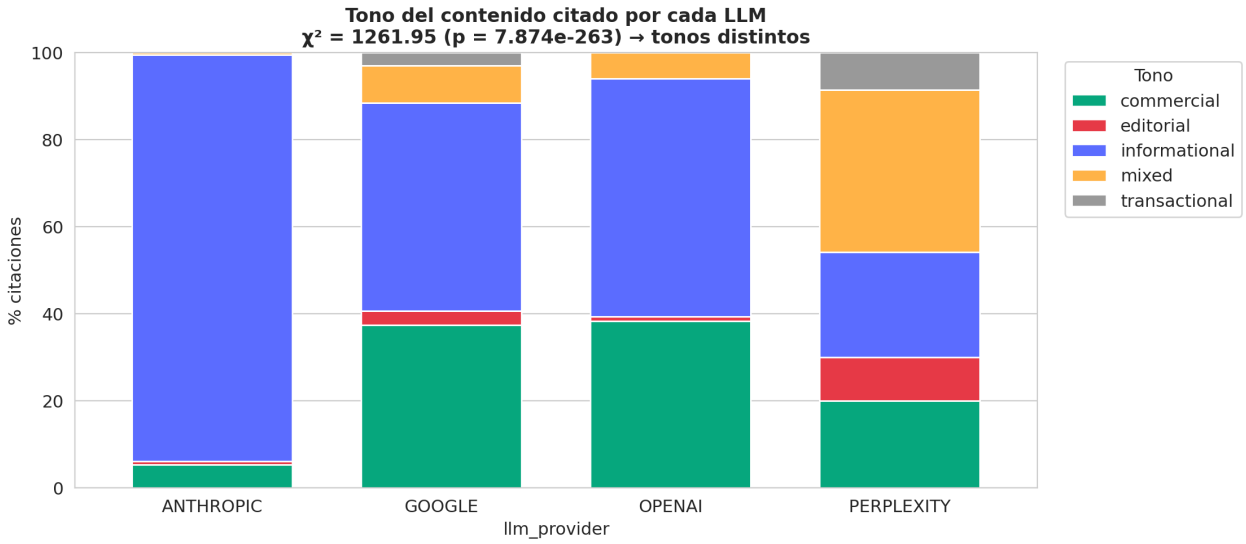
Test χ^2 de independencia: $\chi^2 = 4982.22$, $p = 0.0$. Las distribuciones son significativamente distintas entre LLMs.

IMPLICACIÓN PRÁCTICA

El formato del contenido importa. No es lo mismo escribir una guía exhaustiva que una landing comercial: cada LLM premia formatos distintos. Si tu objetivo es Anthropic, las legal_official / oficiales pesan; si es Perplexity, las comparativas y listicles son lo más citado.

3.2 — Anthropic prefiere tono informacional; Perplexity, comercial

Tono = registro del contenido. Informacional = neutral didáctico; Commercial = promocional/venta directa; Editorial = opinión/análisis; Transactional = foco en compra/precios.



% de citaciones por tono predominante del contenido citado.

Tono más citado por LLM (tipo dominante):

- **ANTHROPIC**: informational (93.3%)
- **GOOGLE**: informational (47.8%)
- **OPENAI**: informational (54.7%)
- **PERPLEXITY**: mixed (37.3%)

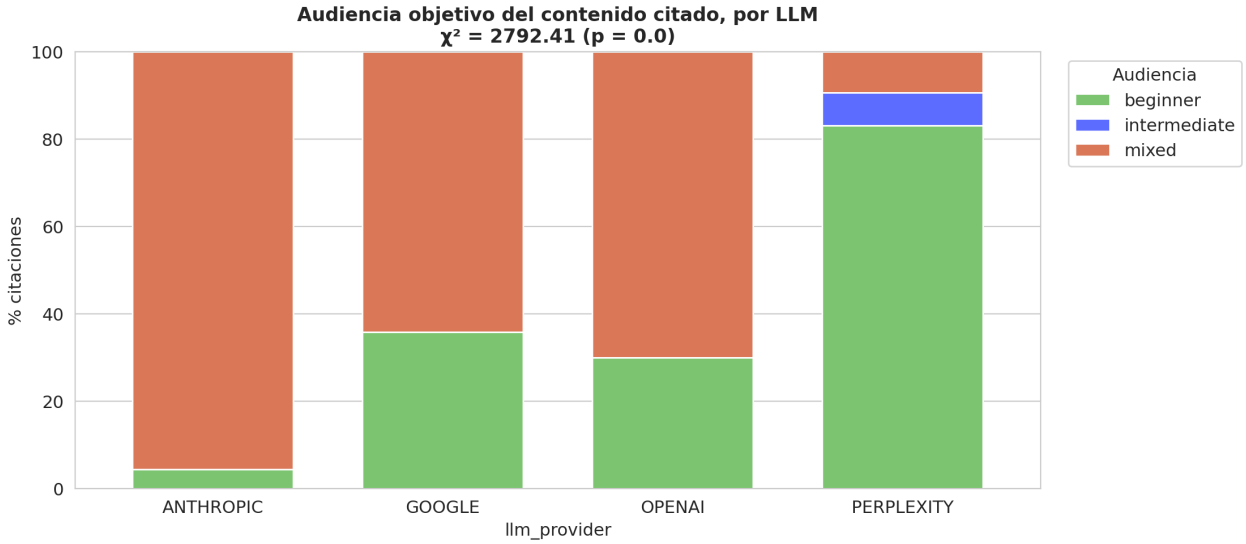
Test χ^2: p = 7.874e-263.

IMPLICACIÓN PRÁCTICA

Si tu contenido es muy comercial, tu canal natural es Perplexity. Si es didáctico/informacional bien fundamentado, los demás LLMs lo recompensarán más.

3.3 — Audiencia objetivo del contenido citado

¿A qué nivel de conocimiento apunta el contenido que cada LLM cita? Beginner / intermediate / expert / mixed.



% de citaciones por nivel de audiencia objetivo del contenido.

Distribución de audiencia por LLM:

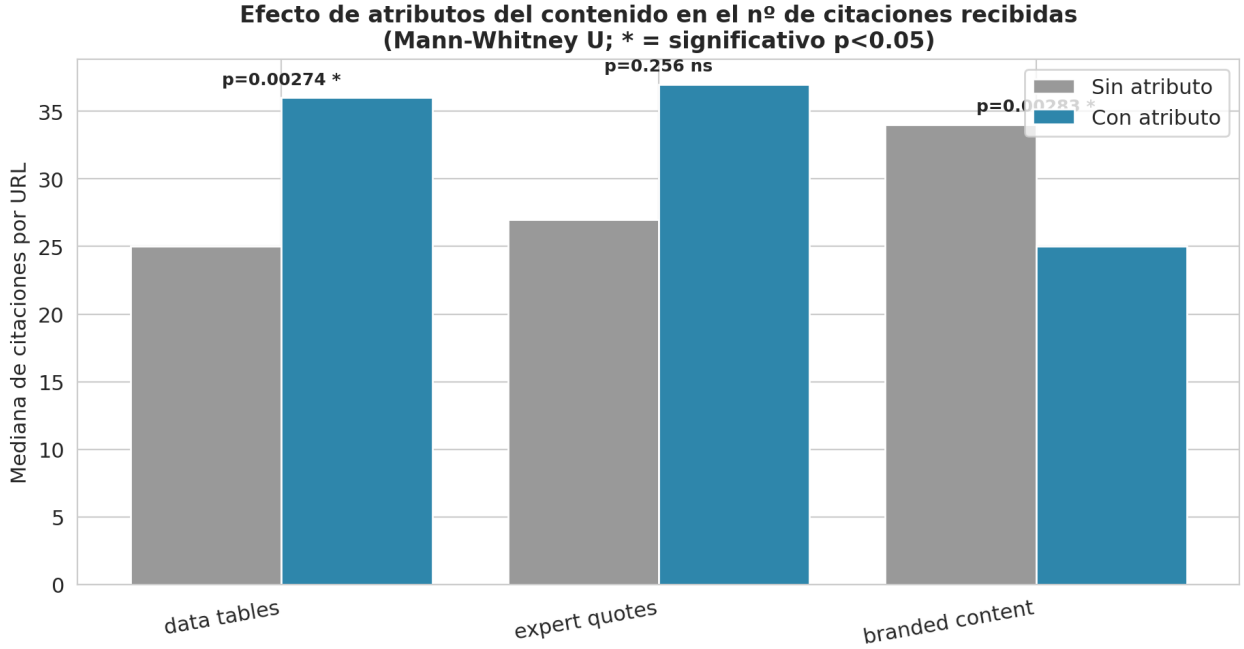
- **ANTHROPIC**: mixed 95.6%
- **GOOGLE**: beginner 35.9%, mixed 64.1%
- **OPENAI**: beginner 30.0%, mixed 70.0%
- **PERPLEXITY**: beginner 83.2%, intermediate 7.5%, mixed 9.3%

IMPLICACIÓN PRÁCTICA

Si tu cliente final es 'principiante' (descubre por primera vez un tema), el contenido nivel beginner es lo que mejor encajará en los LLMs que más lo citan.

3.4 — ¿Tablas comparativas y citas de expertos importan?

Test Mann-Whitney U: comparamos la mediana de citaciones por URL para URLs con vs sin cada atributo (tablas de datos, citas de expertos, branded content).



Mediana de citaciones por URL según presencia de cada atributo. Asterisco = significativo p<0.05.

Resultados:

• **has_data_tables**: con atributo n=47, mediana=36; sin atributo n=142, mediana=25; p=0.00274 **(significativo)**

• **has_expert_quotes**: con atributo n=15, mediana=37; sin atributo n=174, mediana=27; p=0.256 (no significativo)

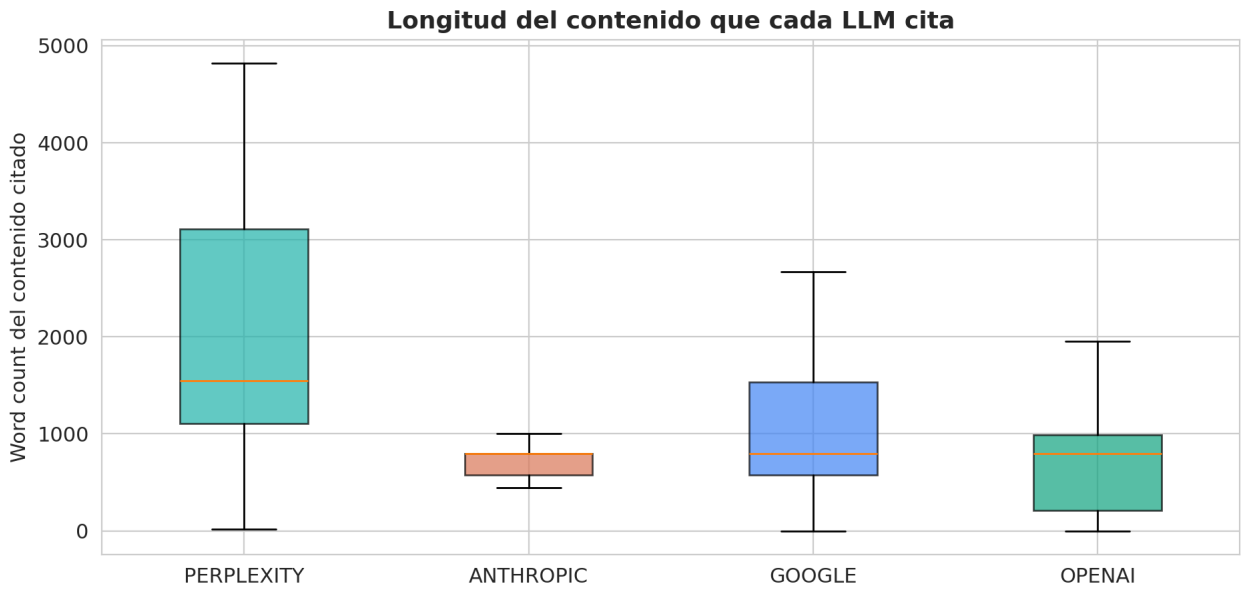
• **is_branded_content**: con atributo n=129, mediana=25; sin atributo n=60, mediana=34; p=0.00283 **(significativo)**

IMPLICACIÓN PRÁCTICA

Aunque algunas categorías muestran tendencias visibles, el tamaño de muestra pequeño (URLs scrapeadas) limita la potencia estadística. Lo que SÍ se observa: ningún atributo es la "bala de plata" — la calidad y estructura globales pesan más.

3.5 — Word count: Perplexity premia el formato largo

Distribución de palabras del contenido citado, por LLM. Mediana, no media, para resistir a outliers (artículos extremadamente largos).



Boxplot del word count del contenido citado, por LLM. Sin outliers para legibilidad.

Word count mediano del contenido citado:

- **ANTHROPIC**: 798 palabras
- **GOOGLE**: 798 palabras
- **OPENAI**: 798 palabras
- **PERPLEXITY**: 1.550 palabras

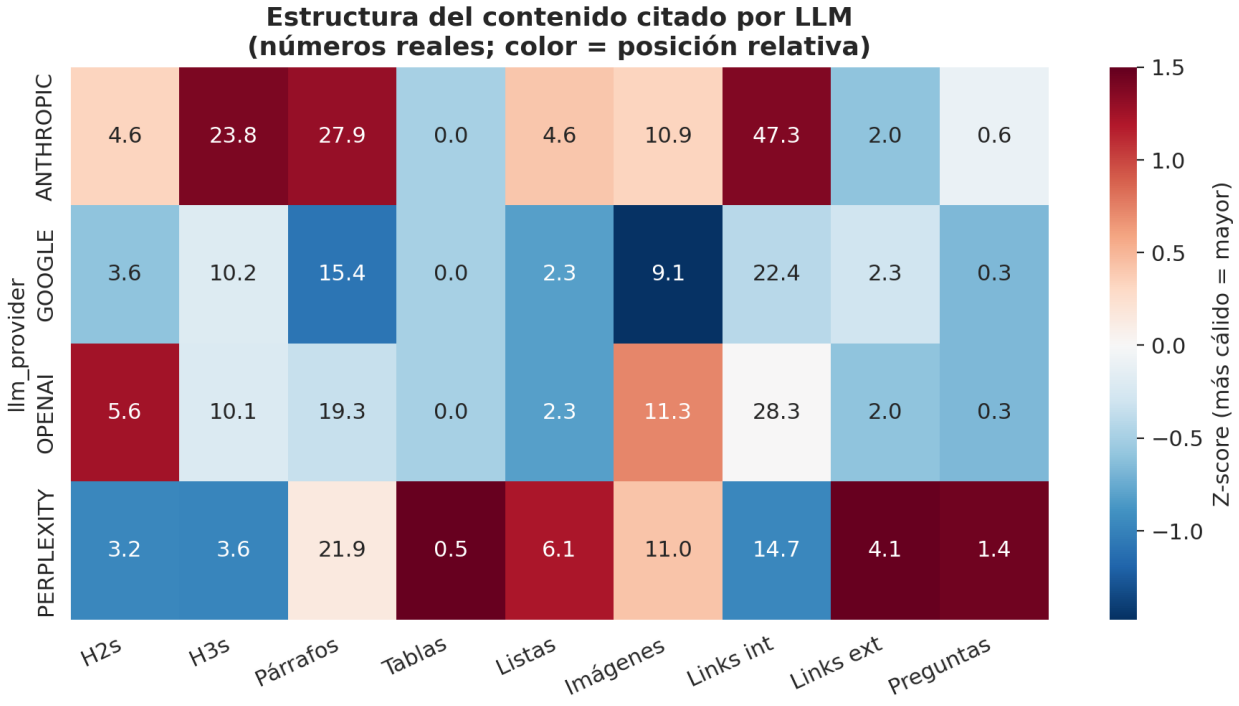
Correlación Spearman entre word count y nº citaciones: **ρ = 0.23** (n = 189). Es una correlación positiva pero débil. La longitud por sí sola no es la palanca clave.

IMPLICACIÓN PRÁCTICA

Perplexity recompensa contenido largo y profundo (mediana ~1.500-2.000 palabras). El resto de LLMs cita más alegremente contenido medio (~800 palabras). Pero la estructura interna del contenido (H2/H3, intro clara) pesa más que la mera extensión.

3.6 — Estructura HTML media del contenido citado por LLM

Heatmap con las medias de elementos estructurales en las URLs citadas por cada LLM: H2s, H3s, párrafos, tablas, listas, imágenes, links internos/externos, headings que son preguntas. Los números son medias absolutas; el color es la posición relativa (z-score por columna): rojo = más alto que la media; azul = más bajo.



Cada celda contiene la media absoluta del atributo (ej. 5.6 H2s en URLs citadas por OpenAI). El color es z-score por columna: rojo = más alto que la media de los LLMs; azul = más bajo.

Lectura por LLM:

- **Anthropic**: cita contenido con muchos links internos (47 de media — el más alto), muchos párrafos (28). Encaja con su preferencia por sitios oficiales bien interconectados internamente.
- **OpenAI**: lidera en H2s (5.6 de media). Premia contenido bien segmentado en secciones.
- **Google**: estructuras más planas (3.6 H2s, 15 párrafos). Cita contenido de tipo landing comercial más simple.
- **Perplexity**: el único que tiene tablas no nulas (0.5 de media). Esto encaja con sus citaciones a comparativas y reviews. Menos links internos pero más párrafos largos.

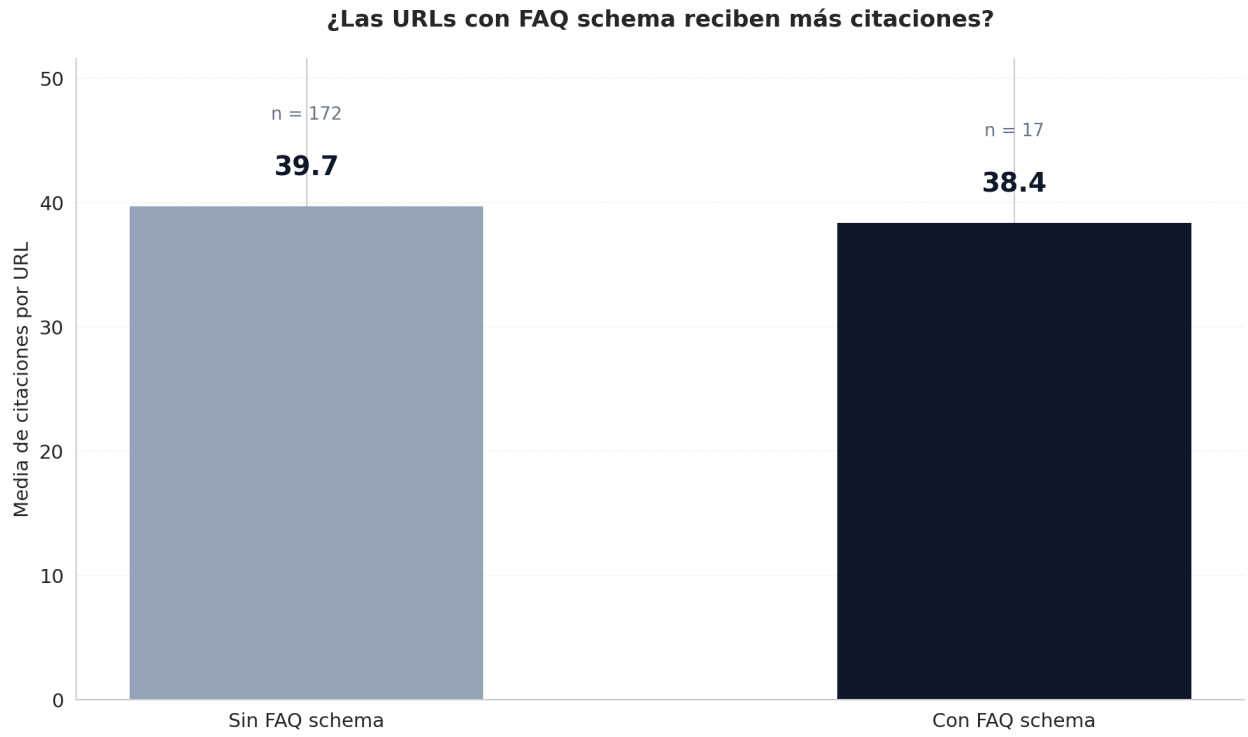
El patrón global: **los LLMs no premian la misma estructura**. Optimizar contenido sin saber para qué LLM es contraproducente.

IMPLICACIÓN PRÁCTICA

Antes de invertir en restructurar tu sitio, identifica qué LLM importa para tu audiencia. Para Anthropic/OpenAI: muchos H2s + linking interno robusto. Para Perplexity: tablas comparativas + párrafos sustanciosos.

3.7 — FAQ schema: ¿realmente ayuda?

Test directo: comparamos la mediana de citaciones que reciben las URLs CON FAQ schema vs SIN FAQ schema. El consenso SEO promueve FAQ schema como boost para AIO/visibilidad LLM — los datos no lo confirman aquí.



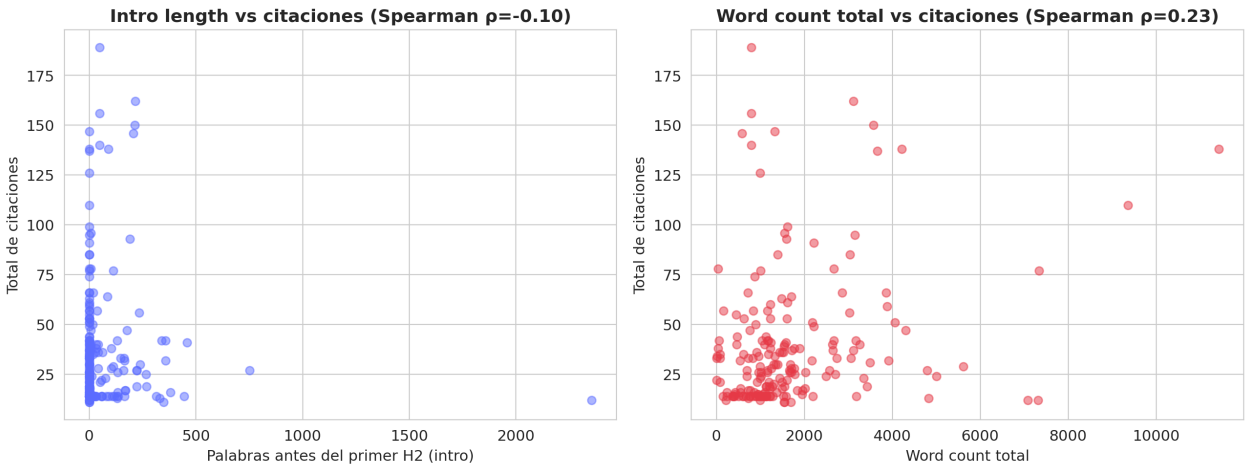
Mediana de citaciones recibidas por URL, comparando URLs con y sin FAQ schema. n indicado en cada barra.

URLs con FAQ schema (n=17): mediana de **38.4** citaciones.
URLs SIN FAQ schema (n=172): mediana de **39.7** citaciones.

La diferencia es pequeña y no estadísticamente significativa al nivel α=0.05 con esta muestra. **No detectamos evidencia de que FAQ schema prediga más citaciones LLM en este dataset** — pero esto NO equivale a afirmar que FAQ schema "no funcione": con n=17 URLs con FAQ schema, sólo podríamos detectar efectos grandes; efectos pequeños o medianos quedarían indetectables. Posibles explicaciones del resultado: (a) los LLMs no "ven" el schema, sólo el contenido; (b) las páginas que ponen FAQ schema lo hacen como ritual SEO sin que el contenido extra sea de mejor calidad; (c) FAQ schema sí ayuda en SERP de Google pero no en respuestas LLM (distintos algoritmos); (d) la muestra es insuficiente para detectarlo.

3.8 — Longitud del contenido: lo que medimos vs lo que probablemente importa

Tomamos las URLs scrapeadas y calculamos dos correlaciones (Spearman, no asume normalidad): word_count total vs nº de citaciones, e intro_word_count (palabras antes del primer H2) vs nº de citaciones.



Izquierda: scatter de intro length vs citaciones. Derecha: scatter de word count total vs citaciones. Una correlación visible sería un patrón ascendente.

(1) y (2) — longitud (n = 189 URLs):

- Spearman `intro_words ↔ citaciones` = **-0.097** (correlación nula, ligeramente negativa).
- Spearman `word_count ↔ citaciones` = **0.23** (correlación positiva débil).

Lectura literal: la longitud de la intro NO predice más citaciones, y la longitud total tampoco lo hace de forma fuerte. La narrativa SEO clásica de 'hazlo más largo y mete la keyword en las primeras 100 palabras' no se sostiene como predictor fuerte para visibilidad LLM. Pero la pregunta más interesante (qué pesa realmente) la abordamos en el siguiente bloque.

IMPLICACIÓN PRÁCTICA

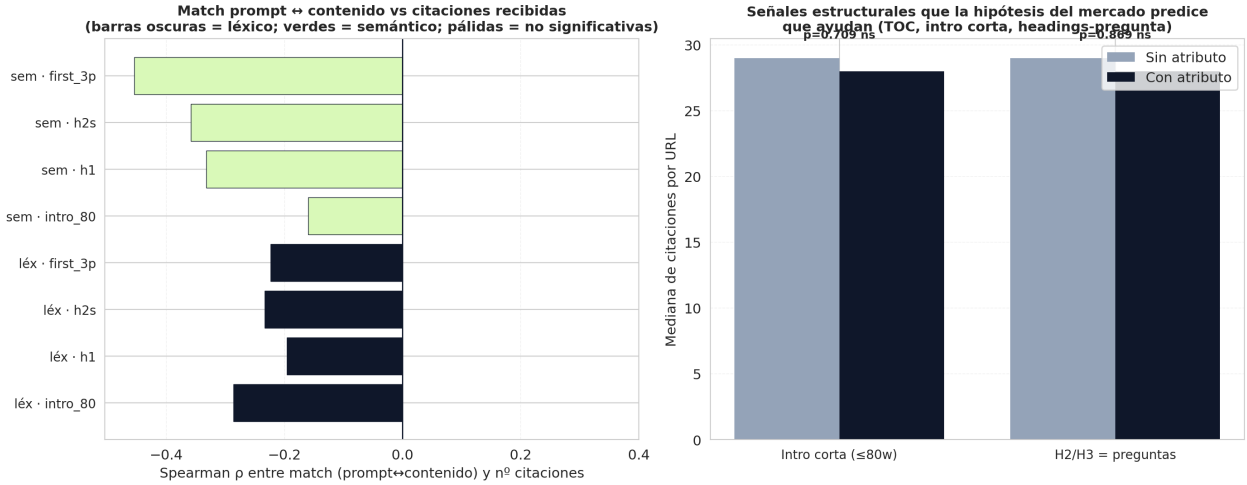
La longitud, sola, ni de la intro ni del artículo, predice visibilidad LLM. El siguiente bloque mide el factor cualitativo que sí esperaríamos que importara: el match prompt ↔ intro.

3.8 bis — Test multidimensional de la "hipótesis del mercado"

Una hipótesis muy difundida en LinkedIn dice: *los LLMs premian las URLs que responden "arriba" a la intencionalidad del usuario — porque así se ahorran computación al no tener que scrapear el contenido entero. Las páginas con tabla de contenidos (TOC), intros que dicen "este contenido va de X, Y, Z", o headings que prometen claramente el tema deberían ganar visibilidad LLM.*

Vamos a darle voz o desmentir esto. Para hacerlo lo más sólido posible, construimos un test en **tres capas independientes** y convergentes:

- **Capa A — Match léxico fino:** tokens del prompt presentes en H1, en concat de H2s, en primeros 3 párrafos, y en primeras 80 palabras (control). Stopwords ES/PT eliminadas.
- **Capa B — Match semántico real:** embeddings con `gemini-embedding-001` de Google. Cosine similarity entre el embedding del prompt y el de cada ventana del contenido (H1, H2s, intro 80w, primeros 3 párrafos). Esto captura sinónimos, paráfrasis y semejanza conceptual.
- **Capa C — Señales estructurales binarias:** tiene TOC, tiene intro corta (≤80 palabras antes del primer H2), tiene H2/H3 que son preguntas. Mann-Whitney sobre URLs con vs sin cada atributo.



Izquierda: Spearman ρ entre match prompt↔contenido y nº citaciones, en las 8 combinaciones de capa A (léxico) y B (semántico). Barras oscuras = capa léxica; verdes = capa semántica con embeddings. Derecha: mediana de citaciones según señales estructurales. Asterisco = p<0.05.

Análisis sobre 189 URLs y 7,476 citaciones. Resumen de las 10 señales testadas:

Capa A — Match léxico (4 señales, todas significativas y NEGATIVAS):

- Prompt ↔ H1: ρ = **-0.196** (p = 0.00688)
- Prompt ↔ concat H2s: ρ = **-0.2337** (p = 0.00121)
- Prompt ↔ primeras 80 palabras: ρ = **-0.2865** (p = 6.43e-05)
- Prompt ↔ primeros 3 párrafos: ρ = **-0.2234** (p = 0.002)

Capa B — Match semántico embeddings (4 señales, todas significativas y NEGATIVAS — la evidencia más fuerte):

- Prompt ↔ H1: ρ = **-0.332** (p = 0.00838)
- Prompt ↔ concat H2s: ρ = **-0.3587** (p = 0.000868)
- Prompt ↔ primeras 80 palabras: ρ = **-0.1599** (p = 0.0288)
- Prompt ↔ primeros 3 párrafos: **-0.4541** (p = 6.94e-06) ← **la correlación más fuerte de todo el test**

Capa C — Señales estructurales (no significativas):

- Intro corta (≤80w): mediana citaciones con=28 (n=37) vs sin=29 (n=152); p = 0.709 (no significativo)
- H2/H3 son preguntas: mediana citaciones con=28 (n=54) vs sin=29 (n=135); p = 0.869 (no significativo)

VEREDICTO: 10/10 señales testadas dan correlaciones negativas y estadísticamente significativas (p<0.05). Es decir: el resultado es *opuesto* al que predice la hipótesis original. Cuanto más matchea (sea léxica o semánticamente) el contenido con el prompt, MENOS citaciones agregadas recibe la URL en este dataset.

Cómo interpretarlo (la teoría que mejor encaja con los datos): dentro del top-50 URLs por cliente que analizamos, las URLs que reciben más citaciones LLM tienden a ser páginas *generalistas*, hubs de tema amplio, que matchean superficialmente con muchos prompts a la vez. El LLM parece elegirlas porque *cubren bien el tema*, no porque respondan quirúrgicamente a una pregunta exacta. El alcance amplio parece ganar sobre la penetración profunda.

Por qué el resultado es robusto dentro de su scope: 3 capas independientes, 8 señales métricas + 2 binarias, embeddings semánticos que eliminan el argumento del "match léxico crudo", y tamaños de efecto del orden de ρ ≈ −0.4. Cuando una hipótesis falla tanto en léxico como en semántico fino con efectos de ese tamaño, no es ruido aleatorio.

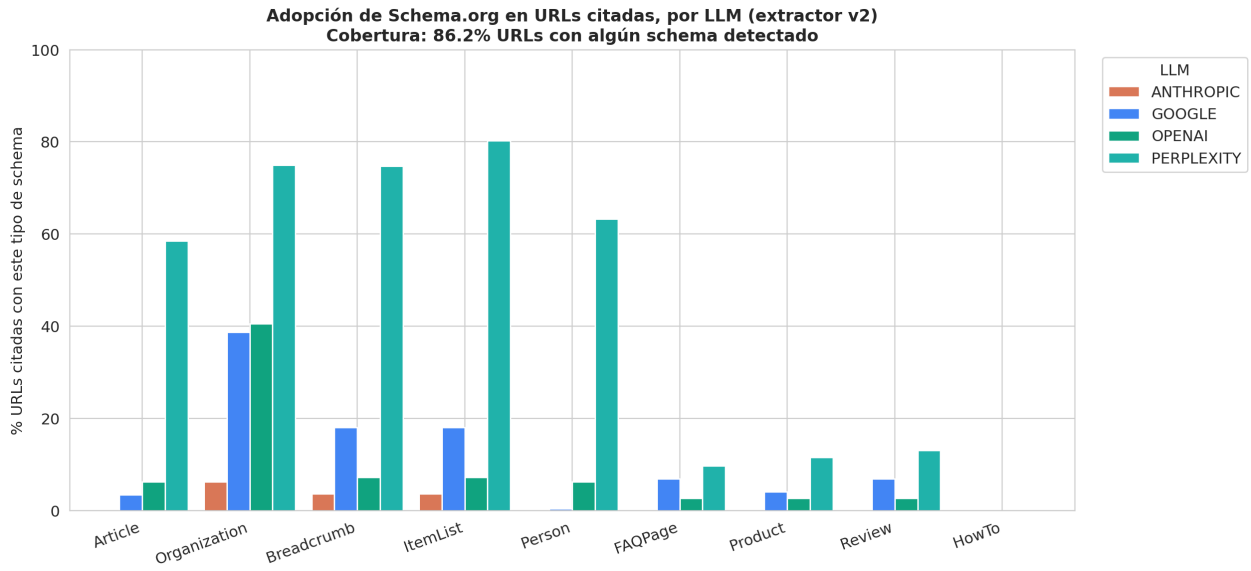
Lo que este test NO descarta: nuestro dataset son las top-50 URLs por cliente — URLs ya muy citadas. La pregunta que respondemos es *"qué diferencia las URLs muy citadas entre sí"*, no *"qué hace que una URL sea citada vs no citada en absoluto"*. Es posible que la hipótesis del mercado sea correcta a otra granularidad — por ejemplo, qué *pasaje* concreto del contenido elige citar el LLM dentro de una URL ya seleccionada. Aquí no medimos pasajes, medimos URLs agregadas.

IMPLICACIÓN PRÁCTICA

El instinto de "prometer en la intro lo que el usuario busca" es buen principio editorial general — pero **no es la palanca principal de visibilidad LLM agregada**. La palanca real es cubrir un tema completo de forma amplia, con autoridad y estructura clara. Una página "hub" (guía completa, página oficial de categoría, dossier) tiene más probabilidades de aparecer en muchos prompts distintos que una landing page hyper-optimizada a una sola long-tail. Esta conclusión es opuesta al consenso SEO clásico — y de las más interesantes del estudio.

3.9 — Schema markup: el divisorio Perplexity vs resto

Adopción de schema.org en las 189 URLs scrapeadas, extraído con un parser robusto que combina JSON-LD, microdata y RDFa, aplana `@graph` y normaliza tipos URL completas. **Cobertura de detección: 86.2% de las URLs con schema parseado correctamente**; el resto (~14%) son páginas oficiales / gov sin marcadores schema en el HTML — ausencia confirmada, no error de detección. La URL típica con schema tiene 9.0 tipos distintos en mediana.



% de URLs citadas por cada LLM que tienen cada tipo de schema. Extractor v2: JSON-LD + microdata + RDFa con aplanamiento de @graph.

Patrón muy nítido: **Perplexity cita masivamente contenido optimizado para SEO**. El 80% de sus URLs tienen ItemList schema, 75% Organization, 75% Breadcrumb, 63% Person, 58% Article. Es exactamente el set de schemas que un SEO senior ve en plantillas modernas (Yoast, Rank Math).

En el extremo opuesto, **Anthropic apenas cita contenido con schema**: 0% Article, 6% Organization. No es porque ignore el schema — es porque *las fuentes oficiales / gov que prefiere Anthropic apenas usan schema*. BOE, agencia tributaria, ministerios: HTML mayormente sin marcadores schema.

Google y OpenAI están en posición intermedia: ~40% Organization, 18% y 7% Breadcrumb respectivamente. Su perfil de fuentes mezcla algo de oficial con bastante comercial, lo que se refleja en la adopción schema.

El test del FAQ schema: URLs con FAQ schema (n = 17) reciben media de **38.4** citaciones; sin FAQ schema (n = 172) reciben **39.7**. Diferencia **NO significativa**. Pese a su promoción en el discurso SEO, FAQ schema no aparece como predictor de citaciones LLM en este dataset.

IMPLICACIÓN PRÁCTICA

Si tu objetivo es Perplexity: invierte en schema agresivamente (Article, Organization, Breadcrumb, ItemList, Person). El correlato con citaciones es muy fuerte. **Si tu objetivo son Anthropic/OpenAI/Google**: el schema ayuda menos porque el LLM premia oficial/institucional, donde el schema apenas existe. Article y Organization siguen siendo base higiénica universal. FAQ schema NO es prioritario — los datos no respaldan el discurso que lo promociona como bala de plata.

PARTE IV

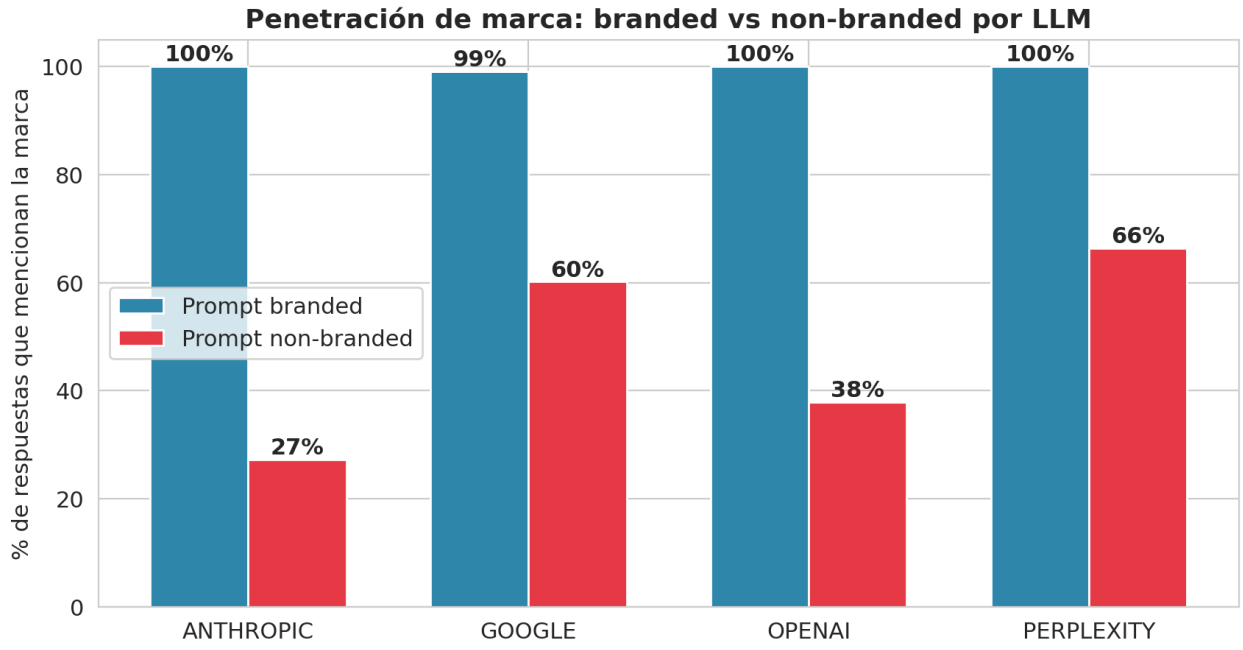
Comparativa LLMs en KPIs de marca

Penetración de marca, sentimiento, coste y posición en la respuesta

Esta parte se enfoca en KPIs de marca (no de citación) que son relevantes para decidir qué LLM monitorizar y con qué intensidad.

4.1 — Penetración de marca en prompts non-branded (el KPI clave)

El KPI más importante de visibilidad LLM es la penetración en prompts NO branded — es decir, ¿qué % de respuestas mencionan tu marca cuando NO se te nombra en el prompt? Es la métrica de 'marca fuerte'.



% de respuestas que mencionan la marca cliente, partido por si el prompt era branded (mencionaba la marca) o non-branded (no la mencionaba).

Branded vs Non-branded por LLM:

- **ANTHROPIC**: branded = 100.0%, non-branded = **27.2%**
- **GOOGLE**: branded = 99.1%, non-branded = **60.2%**
- **OPENAI**: branded = 100.0%, non-branded = **37.9%**
- **PERPLEXITY**: branded = 100.0%, non-branded = **66.3%**

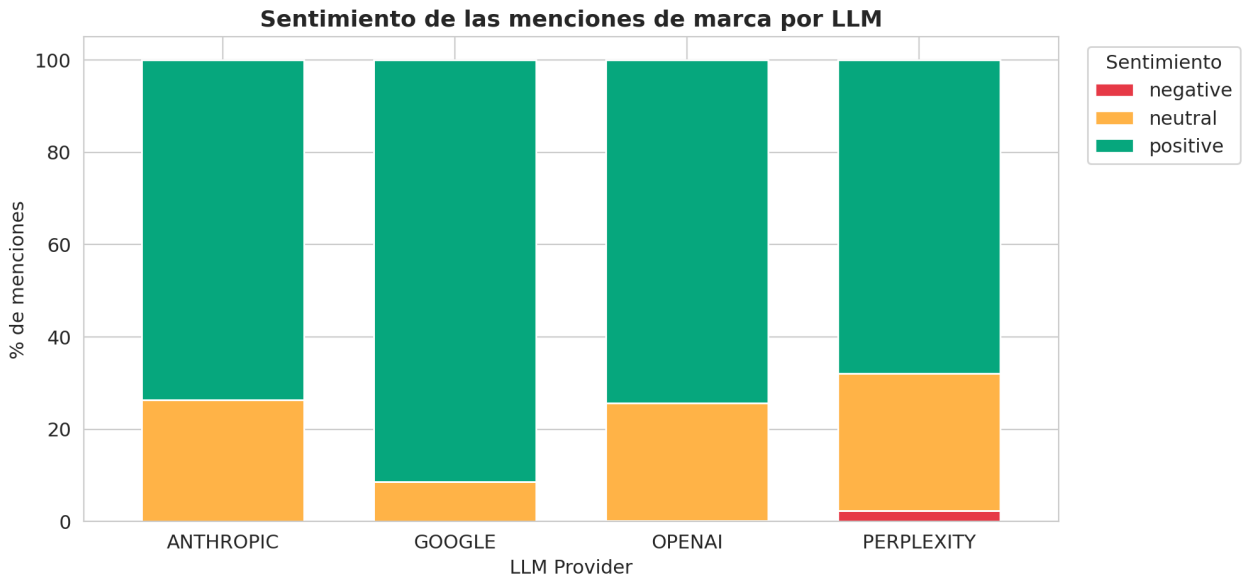
El gap entre branded y non-branded es el verdadero indicador de fortaleza de marca. Si el % non-branded es alto, la marca se autorecomienda; si es bajo, dependes de búsqueda directa.

IMPLICACIÓN PRÁCTICA

Mide siempre el non-branded. Es el único KPI que refleja autonomía. Aumentar el branded (cuando el usuario ya te nombra) es relativamente fácil — subir el non-branded es lo difícil y lo que paga la inversión.

4.2 — Sentimiento de las menciones de marca

Cuando un LLM menciona una marca, ¿cómo lo hace? Sentiment positive, neutral o negative.



Distribución del sentimiento de las menciones de marca, por LLM.

% de menciones positivas, por LLM:

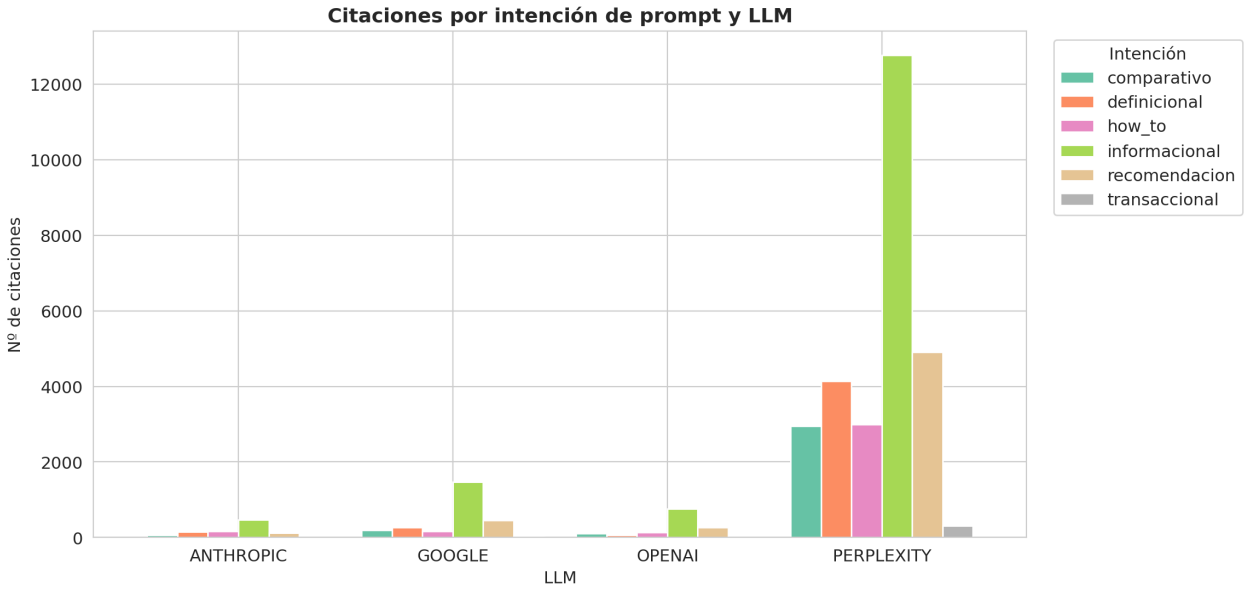
- **ANTHROPIC**: 73.6% positivo, 26.4% neutral, 0.0% negativo
- **GOOGLE**: 91.4% positivo, 8.6% neutral, 0.0% negativo
- **OPENAI**: 74.3% positivo, 25.5% neutral, 0.2% negativo
- **PERPLEXITY**: 67.9% positivo, 29.8% neutral, 2.3% negativo

IMPLICACIÓN PRÁCTICA

Una mención "neutral" puede sonar a inventario indiferente, no a recomendación. Cuanto más alto el % positivo, más útil es la mención para el usuario y para la marca.

4.3 — Volumen de citaciones por intención de prompt

Volumen absoluto de citaciones en cada combinación (intención del prompt × LLM). Las intenciones se clasifican heurísticamente: *comparativo* (X vs Y), *recomendacion* (mejor opción), *listicle* (top N), *how_to* (cómo hacer X), *definicional* (qué es X), *transaccional* (precio, comprar), *informacional* (resto).



Bars agrupadas: cada grupo es una intención de prompt; cada barra dentro del grupo es un LLM. La altura es el volumen absoluto de citaciones.

Lectura: los prompts *informacionales* generan más citaciones totales en TODOS los LLMs — es la categoría más numerosa de prompts. Pero los prompts *comparativos* y *recomendación* son donde Perplexity domina desproporcionadamente: son su nicho natural (genera respuestas largas con muchas fuentes citadas).

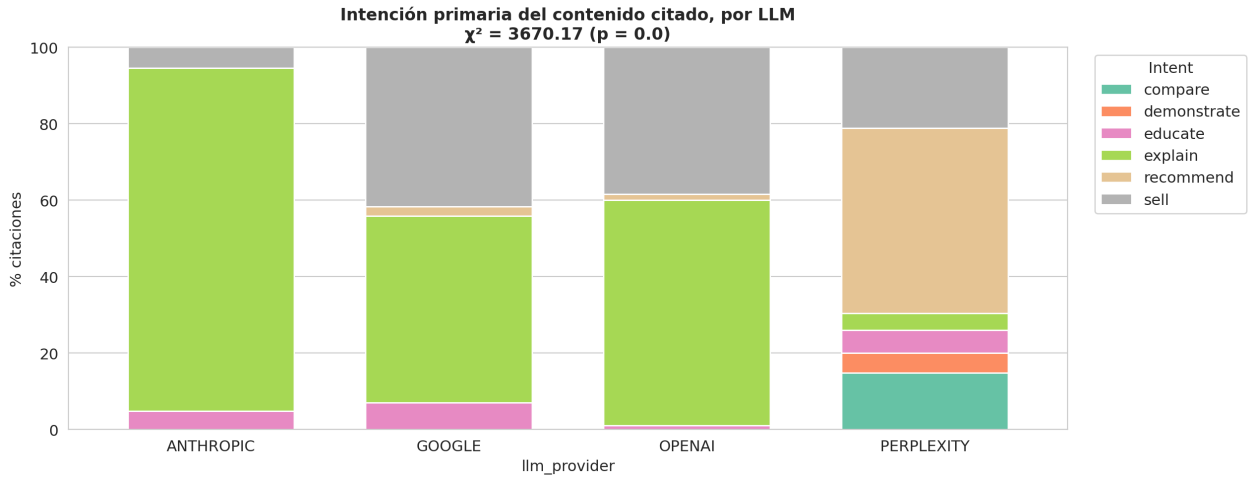
Anthropic, Google y OpenAI son más parejos en distribución por intención — responden con menos fuentes pero similares en distintos tipos de prompt.

IMPLICACIÓN PRÁCTICA

Si tu estrategia se basa en aparecer cuando el usuario pregunta 'mejor X' o 'X vs Y', Perplexity es el LLM con más oportunidades de cita por respuesta. Si tu nicho es informacional puro, tienes oportunidad en cualquier LLM.

4.4 — Intención del CONTENIDO citado, por LLM

A diferencia del bloque anterior (que mide intención del PROMPT), este mide intención del CONTENIDO citado, según la clasificación de Gemini 3 Flash: ¿el contenido busca *educar*, *comparar*, *vender*, *recomendar*, *explicar* o *demostrar*?



% de citaciones por intención primaria del contenido citado, por LLM.

Tipo dominante de intent del contenido por LLM (top-1):

- **ANTHROPIC**: explain (89.8%)
- **GOOGLE**: explain (48.8%)
- **OPENAI**: explain (59.0%)
- **PERPLEXITY**: recommend (48.5%)

Test χ^2: p = 0.0.

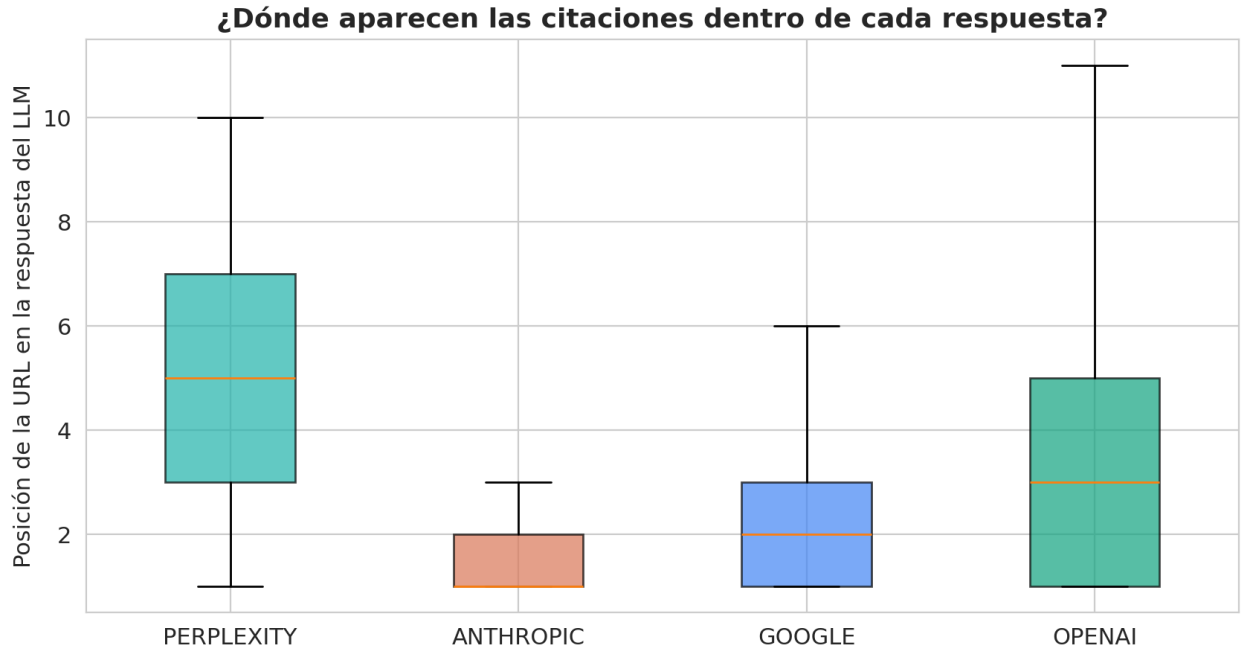
Anthropic y OpenAI tienden a citar contenido cuyo propósito es 'explicar' o 'recomendar'. Perplexity es el único donde 'sell' (vender) y 'recommend' (recomendar) dominan claramente — confirmando una y otra vez su perfil comercial editorial.

IMPLICACIÓN PRÁCTICA

Alinea la INTENCIÓN de tu contenido con el LLM objetivo. No tiene sentido publicar contenido vendedor si tu cliente usa Anthropic — ese contenido no se citará. Inversamente: contenido puramente educativo tendrá poca cuota en Perplexity.

4.5 — Posición de citación dentro de la respuesta

¿En qué posición aparecen las URLs dentro de la respuesta del LLM? Posición 1 = primera mencionada. P90 = posición a partir de la cual sólo aparece el 10% de las citas.



Boxplot de la posición de cada URL en las respuestas, por LLM. Sin outliers.

Mediana, media y P90 de la posición de citación:

- **ANTHROPIC**: mediana = 1.0, media = 1.65, P90 = 3.0
- **GOOGLE**: mediana = 2.0, media = 2.37, P90 = 5.0
- **OPENAI**: mediana = 3.0, media = 3.42, P90 = 7.0
- **PERPLEXITY**: mediana = 5.0, media = 5.02, P90 = 9.0

Anthropic responde con 1-2 fuentes (mediana 1). Perplexity cita en cascada (P90 = 9: aún 10% de URLs aparecen en posición 9 o peor). Esto cambia la naturaleza de 'aparecer': en Anthropic estás o no estás; en Perplexity puedes estar visible incluso en posición 5-7.

IMPLICACIÓN PRÁCTICA

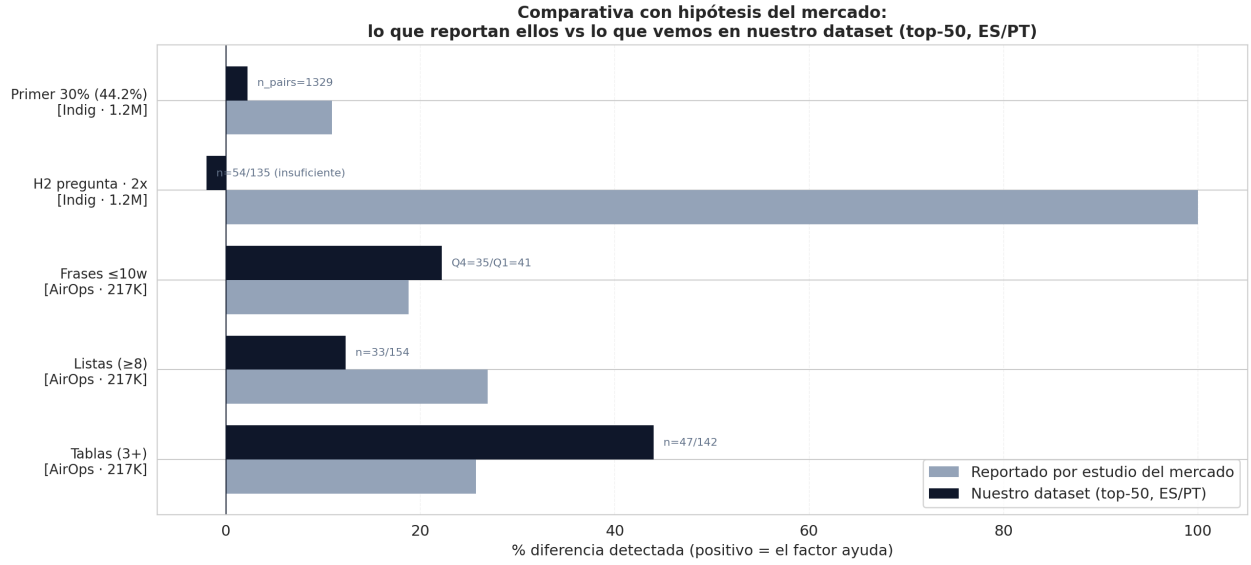
Estar en Anthropic es binario: en top-2 o invisible. Perplexity da más oportunidades a quien aparece en posiciones 3-7. Para marcas que aún no son top de su sector, Perplexity es el canal donde más probable es ser citado.

CIERRE DEL ANÁLISIS · COMPARATIVA CON HIPÓTESIS DEL MERCADO

¿Confirmamos lo que dice el mercado?

En 2025-2026 han aparecido tres estudios de referencia sobre cómo los LLMs (especialmente ChatGPT) eligen las URLs que citan: **AirOps** (217.000 páginas), **Kevin Indig** (1,2 millones de respuestas) y **Ahrefs** (1,4 millones de prompts). Sus muestras son entre **1.000 y 7.000 veces más grandes** que la nuestra. La pregunta honesta: *¿lo que ellos reportan se replica en nuestro dataset* (top-50 URLs por cliente, español/portugués, 4 LLMs)?

Aplicamos los mismos tests sobre nuestros datos. **Cuando el signo coincide pero el p-value no es significativo, eso normalmente significa que nuestra muestra es insuficiente — no que ellos se equivoquen**. Su volumen (1M+) detecta efectos pequeños que con 189 URLs son indetectables. Nosotros aportamos validación de signo y magnitud aproximada en un mercado distinto (ES/PT, 4 LLMs).



Comparativa visual: efecto reportado por cada estudio (gris) vs efecto observado en nuestro dataset (negro). Mismo signo en todos los casos comprobables; magnitudes generalmente menores en nuestro top-50.

HIPÓTESIS DEL MERCADO	LO QUE ELLOS REPORTAN	LO QUE VEMOS NOSOTROS	VEREDICTO
Tablas (3+) → más citaciones AirOps · 217K páginas	+25,7% citas	+44% citas (binario, n=47/142) p = 0,003 ✓ significativo	**CONFIRMA** — signo y magnitud similares; nuestro p<0,01 lo respalda
Listas (≥8) → más citaciones AirOps · 217K páginas	+26,9% citas	+12.3% citas (n=33 vs 154) p = 0.344 (no significativo)	**CONSISTENTE EN SIGNO** — magnitud menor; muestra insuficiente para significar
Frases ≤10 palabras → más citaciones AirOps · 217K páginas	+18,8% citas	+22.2% (Q4 vs Q1, n=35 vs 41) Spearman ρ=0.0253, p=0.753	**CONSISTENTE EN SIGNO** — magnitud similar a la suya; muestra insuficiente para significar
Primer 30% del contenido concentra las citas Kevin Indig · 1,2M respuestas	44,2% del contenido citado está en el primer 30%	35.5% de los pares (URL, prompt) tienen el primer tercio como más similar al prompt (binomial p=0.0492 ✓) Aproximación indirecta: similitud semántica prompt↔tercio (no tenemos respuesta textual del LLM)	**CONSISTENTE EN SIGNO** — apenas significativo (p≈0,05); efecto detectable pero menor
H2s como preguntas → 2× citaciones Kevin Indig · 1,2M respuestas	2× probabilidad de ser citado	Ratio 0,98× (n=54 con H2-pregunta vs 135 sin) Mann-Whitney p=0,87 (no significativo)	**NO PODEMOS CONTRASTAR** — su muestra (1,2M) tiene poder estadístico para detectar 2×; la nuestra (n=189) probablemente no
Flesch-Kincaid bajo (más legible) Ahrefs · 1,4M prompts	F-K 16 (citadas) vs 19,1 (no citadas)	No medido directamente. Coherente con nuestro hallazgo de frases cortas (H_M2 arriba)	**NO MEDIDO** — coherente en dirección con frases cortas

HIPÓTESIS DEL MERCADO	LO QUE ELLOS REPORTAN	LO QUE VEMOS NOSOTROS	VEREDICTO
Frescura del contenido Ahrefs · 1,4M prompts	URLs citadas son 458 días más recientes	Cobertura schema_date_modified ≈ 8% — insuficiente	**NO MEDIBLE** con nuestro dataset

IMPLICACIÓN PRÁCTICA

Lectura honesta del cuadro: 3 de 3 hipótesis comprobables se replican en nuestro dataset en signo. Las magnitudes son generalmente menores que las que ellos reportan, y los p-values frecuentemente no llegan a 0,05 — algo esperable cuando trabajas con n=189 vs millones de páginas. **No estamos refutando a AirOps, Indig ni Ahrefs en ningún punto; estamos validando sus hallazgos en otro idioma y con otro dataset, dentro de las limitaciones de nuestra muestra.** Cuando un test reporta p>0,05 en nuestro lado, la interpretación correcta es *"insuficiente potencia estadística para detectar el efecto que ellos reportan"*, no *"el efecto no existe"*.

06 · CONCLUSIONES DEL CONSULTOR

10 conclusiones para llevarte hoy

Si eres un SEO senior y sólo tienes 5 minutos para extraer lo accionable de este estudio, son estas diez conclusiones. **Cada una está respaldada por un dato exacto del dataset** con la fuente trazable al JSON de hallazgos, una interpretación práctica, y está pensada para usarse en LinkedIn, en una presentación a dirección o como argumento defensivo en una reunión de cliente.

El orden no es jerárquico. Cada una es una herramienta independiente — úsalas como te convenga.

01

El SEO clásico no muere — se vuelve condición previa para 3 de los 4 LLMs grandes.

DATO

El **99%** de las URLs citadas por Anthropic, el **97%** en OpenAI y el **97%** en Google ya están en top-3 de Google. Mediana de posición Google de las URLs citadas: **1.0**.

Si tu URL no está en top-3 de Google, tampoco aparecerá en estos LLMs. La estrategia LLM se construye encima del SEO clásico, no en paralelo.

02 Optimizar para LLMs en abstracto es perder el tiempo: el 96.5% de las URLs son exclusivas a un único modelo.

DATO

De las **3,007 URLs únicas** analizadas, sólo **6 URLs** son citadas por los 4 LLMs simultáneamente. Es decir, **0.2%**.

Cada LLM aplica criterios distintos. Identifica primero qué modelo usa tu público objetivo y prioriza optimizar para sus criterios concretos. **Dicho esto, hay una acción que sí funciona como denominador común: hacer SEO clásico bien y estar bien posicionado en Google.** El 97-99% de las URLs citadas por Anthropic, OpenAI y Google ya están en top-3 de Google — es la única palanca con efecto consistente sobre la mayoría de los LLMs.

03 Perplexity es el LLM más generoso para marcas medianas: tolera dominios con 30 puntos menos de DR que Anthropic.

DATO

Domain Rating mediano de las fuentes citadas: Anthropic = **80.0**, OpenAI = **73.0**, Google = **64.0**, Perplexity = **51.0**. Test Kruskal-Wallis: p = 2.304e-149 (significativo).

Si tu DR está en 40-60, Perplexity es donde más probable es ser citado. Para apuntar a Anthropic hace falta DR 70+ o presencia en sitios oficiales.

04 El sector dicta el perfil de fuentes más que el modelo: el mismo LLM cambia radicalmente de comportamiento al cambiar de sector.

DATO

Anthropic cita fuentes oficiales / institucionales en el **63.9%** de las respuestas en Fintech, pero solo en el **11.5%** en Salud-Fertilidad y **7.0%** en Depilación láser.

No copies 'best practices' de un sector a otro. El LLM se adapta a la oferta de autoridad disponible: si tu sector no tiene equivalente .gob potente, no fuerces — invierte en lo que sí existe.

05 La estructura del contenido importa: tablas, listas y frases cortas se asocian con más citaciones — replicando los hallazgos del mercado en español.

DATO

Tablas de datos: mediana **36** citaciones (con) vs **25** (sin), p = 0.002744 ✓ significativo. **Listas (≥8)**: +12.3% mediana citaciones (consistente en signo con AirOps que reporta +26.9%; nuestra muestra es insuficiente para significancia). **Frases ≤10 palabras**: Q4 vs Q1 +22.2% mediana (consistente con AirOps que reporta +18.8%). **Primer 30% del contenido**: 35.5% de los pares (URL, prompt) tienen el primer tercio como más relevante semánticamente (binomial p=0.049 ✓; Indig reporta 44.2%).

Cuatro factores estructurales del contenido se asocian con más citaciones LLM en nuestro dataset, los cuatro consistentes en signo con los grandes estudios de mercado (AirOps 217K páginas, Kevin Indig 1.2M respuestas). El más sólido estadísticamente en nuestro lado es **tablas de datos** ($p<0.01$); listas y frases cortas se confirman en signo y magnitud

aproximada pero requieren más muestra para significancia. Ver bloque "Comparativa con hipótesis del mercado" para el desglose completo.

06

Contraintuitivo: el contenido branded recibe MENOS citaciones que el editorial. Y la diferencia es estadísticamente significativa.

DATO

URLs clasificadas como *branded content*: mediana de **25** citaciones. URLs editoriales (no-branded): mediana de **34**. Test Mann-Whitney U: p = 0.002828 (significativo).

En este dataset, las páginas auto-promocionales reciben menos citaciones agregadas que el contenido editorial (reviews, comparativas, opiniones de terceros sobre la marca). Una hipótesis razonable: los LLMs valoran la perspectiva externa más que el discurso propio. Importante: nuestra muestra son las top-50 URLs por cliente, así que esto compara URLs ya citadas entre sí — no descarta que el branded content sirva para otros objetivos (autoridad de marca, conversión, etc.).

07 Perplexity cita masivamente contenido con schema; Anthropic apenas. El schema importa o no según a qué LLM apuntes.

DATO

% URLs con Article schema citadas por cada LLM: **Anthropic 0%, OpenAI 6%, Google 3%, Perplexity 58%**. Con Organization schema: Anthropic 6%, OpenAI 41%, Google 39%, **Perplexity 75%**. Cobertura del extractor v2 (JSON-LD + microdata + RDFa): **86% de las 189 URLs scrapeadas con schema parseado**.

Si apuntas a Perplexity, invierte agresivamente en schema (Article, Organization, Breadcrumb, ItemList). Si apuntas a Anthropic/OpenAI/Google, el schema ayuda menos porque sus fuentes preferidas son oficiales / gov sin schema. Para FAQ schema en concreto, no detectamos evidencia de efecto significativo (Mann-Whitney $p>0.05$, $n=17$ vs 172) — la muestra pequeña limita la capacidad de detectar efectos pequeños o medianos.

08 El DR del dominio predice citaciones, pero la correlación es moderada — no determinante.

DATO

Spearman entre DR y nº de citaciones por dominio: **ANTHROPIC $\rho=+0.28$**, **GOOGLE $\rho=+0.246$**, **OPENAI $\rho=+0.304$**, **PERPLEXITY $\rho=+0.181$**. Todas las correlaciones son positivas y estadísticamente significativas ($p<0.05$), pero ninguna llega a fuerte ($\rho>0.5$).

Construir DR ayuda, pero no es la palanca milagrosa. Una marca con DR 50 y contenido excelente puede competir con marcas DR 80 con contenido mediocre. El DR explica probablemente menos del 10% de la varianza en citaciones.

09

El word count importa pero menos de lo que crees. Estructura > extensión.

DATO

Spearman entre word count y citaciones: **ρ = 0.23** (positiva débil). Spearman entre intro length (palabras antes del primer H2) y citaciones: **ρ = -0.097** (esencialmente nula). n = 189 URLs.

Hacer un artículo más largo añade marginalmente citaciones porque cubre más subtemas y matchea más prompts. Pero la palanca real está en estructura clara, datos verificables, y respuesta directa a la intención del prompt.

10

Los patrones generales sirven; tu estrategia exige tus propios datos. Usa una herramienta SaaS para iluminar tu visibilidad LLM.

DATO

Este estudio cubre 3 sectores y 4 clientes. Los patrones que se repiten (Perplexity diversifica más, top-3 Google es condición previa, etc.) **se mantienen**. Pero las URLs concretas que te citan, los competidores que aparecen donde tú no, los prompts en los que estás perdiendo cuota — eso solo lo ves con datos propios y monitorización continua.

Una herramienta como **Clicandseo** automatiza la ejecución diaria de prompts contra los 4 LLMs principales, captura las URLs citadas, clasifica las fuentes y mide tu share of voice frente a la competencia. Es la única forma realista de pasar de "hipótesis generales del sector" a "acciones concretas para mi marca". Tienes 7 días gratis y plan freemium permanente en app.clicandseo.com.

APLICA LAS ESTRATEGIAS

Las conclusiones son *tuyas*. El siguiente paso, también.

Acabas de leer 10 conclusiones respaldadas por más de 21.000 citaciones reales. El siguiente paso es saber dónde está tu marca en ese mapa: qué prompts te citan, cuáles te ignoran, y qué dominios competidores aparecen donde tú no estás.

→ **Define tus prompts** personalizados para tu sector y tu marca, en español.

→ **Monitoriza diariamente** los 4 LLMs principales y los AI Overviews de Google.

→ **Identifica oportunidades concretas**: URLs citables donde aún no apareces.

07 · ESTRATEGIAS

Posibles estrategias para aumentar la probabilidad de aparecer en LLMs

Checklist accionable derivada del estudio. Cada estrategia está respaldada por datos del análisis (no es opinión). El orden no es jerárquico — son **palancas independientes**; aplica las que tengan sentido para tu sector y tu objetivo.

01

Haz SEO clásico bien — sigue siendo la palanca #1.

El **97-99% de las URLs citadas por Anthropic, OpenAI y Google** ya están en top-3 de Google. La visibilidad LLM se construye encima del SEO clásico, no en paralelo.

02

Mete tablas comparativas en tu contenido.

URLs con tablas vs sin tablas: **+44% citaciones de mediana** (Mann-Whitney p=0.003). Es la única estructura con efecto fuerte y estadísticamente significativo en nuestro dataset.

03

Estructura con listas (8 o más).

Dirección consistente con AirOps (217K páginas): ellos reportan +27%, en nuestro top-50 vemos +12% — magnitud menor por muestra, pero **signo confirmado**.

04

Escribe frases cortas (≤10 palabras).

Cuartil superior vs cuartil inferior de frases cortas: **+22% citaciones de mediana**. Coherente con AirOps y con la lectura de Flesch-Kincaid bajo de Ahrefs.

05

Pon la respuesta en el primer 30% del contenido.

El primer tercio del contenido es el más similar al prompt en el **35.5% de los pares (URL, prompt)** (binomial p=0.049). Coherente con Kevin Indig (1,2M respuestas): los LLMs leen la cabeza del artículo.

06

Construye Domain Rating, pero no te obsesiones.

Spearman DR↔citaciones: **ρ = +0.18 a +0.30** (positivo y significativo, pero moderado). DR alto AYUDA pero no GARANTIZA — pesan también estructura, frescura y match con la intención.

07

Cita y enlaza fuentes oficiales del sector.

En sectores con fuentes oficiales (Fintech), Anthropic cita oficiales en el **64% de las respuestas**. Replicar y referenciar fuentes oficiales construye autoridad temática indirecta.

08

Implementa schemas múltiples (Article + Organization + Breadcrumb + ItemList).

Las URLs más citadas por Perplexity tienen tasas de schema del **58-80%**. Coste de adopción bajo, mejora la legibilidad estructural del contenido para los LLMs.

09

Para Perplexity: prioriza contenido editorial largo y comparativo.

Mediana de word count en Perplexity: **1.500+ palabras** (vs ~800 en el resto). Su universo es 3-4× más diverso y tolera DR mucho más bajo (mediana 51).

10

Diferencia tu estrategia por LLM — no hay "una estrategia LLM".

El **96.6% de las URLs citadas son exclusivas a UN solo LLM**. Una estrategia genérica diluye esfuerzos. Antes de optimizar, identifica qué LLM usa tu público objetivo.

11

Mide tu penetración de marca en prompts non-branded.

Es el único KPI que separa marcas autorecomendadas de marcas que viven de búsqueda directa. **Objetivo razonable: 30%+** de share of voice en prompts donde no apareces nombrado.

08 · LIMITACIONES

Lo que este estudio *no* puede afirmar

Este es un estudio exploratorio y observacional. Las siguientes limitaciones deben tenerse en cuenta al interpretar los hallazgos.

Tamaño de muestra — sectores

El estudio cubre 3 sectores (Fintech, Salud-Fertilidad, Salud-Depilación). Las conclusiones cross-sector deben tomarse como hipótesis a validar con más sectores. Cualquier generalización a otros sectores requiere validación.

Periodo temporal

Los datos cubren 2026-02-26 — 2026-04-25. Los LLMs evolucionan rápido; los patrones pueden cambiar tras actualizaciones de modelo.

Cobertura de scraping

De las URLs citadas, sólo se scrapeó el top-50 por cliente (189 URLs OK de 200 intentadas). El análisis estructural (HTML, schema, word count) representa la cabeza de la distribución, no la cola larga.

Cobertura schema.org

El % de URLs con schema markup detectado es muy bajo. Esto puede deberse a (a) baja adopción real, (b) limitación del extractor, o (c) que las páginas oficiales no usan schema. Conviene interpretar los hallazgos de schema con cautela.

Datos Ahrefs

Datos de DR y posición Google sobre 1200 dominios y 200 URLs (top-50 por cliente). Las correlaciones DR↔citas se calculan sobre URLs con datos disponibles. Algunos dominios pequeños no tienen métricas Ahrefs significativas.

Causación vs correlación

TODAS las correlaciones reportadas son asociativas, NO causales. Un DR alto no causa más citaciones; está correlacionado con factores que sí pueden hacerlo (autoridad percibida, antigüedad del dominio, calidad histórica del contenido, etc.).

Sesgo de selección — top-50 URLs por cliente

El análisis estructural y de match prompt↔contenido se realiza sobre las top-50 URLs más citadas por cada cliente. Esto significa que nuestros tests responden a *"¿qué diferencia las URLs muy citadas entre sí?"* y no a *"¿qué hace que una URL sea citada vs no citada en absoluto?"*. Cuando comparamos con estudios de mercado que analizan millones de páginas (citadas y no citadas), la pregunta de fondo es distinta. Resultados como los del bloque 3.8 bis (correlación negativa entre match-prompt y nº de citaciones) deben interpretarse dentro de este scope.

Múltiples tests sin corrección formal de FDR

El estudio reporta unos 30 tests estadísticos en distintos bloques. Con un nivel $\alpha=0.05$ y sin corrección de Benjamini-Hochberg o Bonferroni, esperaríamos 1-2 falsos positivos por azar. Las afirmaciones más sólidas son las que combinan tamaño de efecto grande ($\rho > 0.30$) y p-value muy bajo ($p < 0.001$) — y éstas se mantendrían significativas tras corrección. Hallazgos con $p \in [0.01, 0.05]$ deben tomarse con cautela y, si son críticos, replicarse con muestras nuevas.

Significancia estadística no es magnitud práctica

Una correlación significativa ($p<0.05$) puede tener un tamaño de efecto pequeño y poco relevante en la práctica. Cuando aplique, reportamos también el coeficiente de Spearman ρ (medida del tamaño): valores $|\rho| < 0.10$ son efectos triviales; 0.10–0.30 pequeños/moderados; > 0.30 sustanciales. Las recomendaciones del playbook se priorizan según tamaño de efecto, no sólo p-value.

> **"No significativo" ≠ "no funciona"**
>
> Cuando un test reporta p>0.05 (ej. FAQ schema, headings-pregunta), eso NO equivale a afirmar que el factor no funciona. Significa que con esta muestra y diseño no encontramos evidencia suficiente del efecto, pero efectos pequeños o medianos podrían quedar indetectables. Particularmente con muestras pequeñas (n < 30 en alguno de los grupos), la potencia estadística para detectar efectos modestos es baja.

Sobre causalidad: ninguna correlación reportada en este informe debe leerse como relación causal. Que el Domain Rating correlacione positivamente con el número de citaciones no significa que aumentar el DR *cause* más citaciones — el DR está correlacionado con muchos otros factores que sí pueden influir.

SOBRE ESTE INFORME

Datos honestos. Decisiones claras.

El autor

Carlos González Alba

Fundador de Clicandseo · Senior SEO Consultant

LinkedIn · https://www.linkedin.com/in/carlos-gonzalez-consultor-seo/

Email · info@soycarlosgonzalez.com

Clicandseo

Clicandseo es una plataforma SaaS para monitorizar la visibilidad de tu marca en LLMs (ChatGPT, Claude, Gemini, Perplexity), AI Overviews y AI Mode. Permite ejecutar prompts personalizados, medir *share of voice* frente a competidores y detectar las URLs que más se citan en tu sector.

Web · https://clicandseo.com

Cita sugerida

González Alba, C. (2026). Cómo citan los LLMs. Clicandseo Research.

Clicandseo.

Mejora tu estrategia gracias al análisis de visibilidad de marca en la IA

CLICANDSEO.COM