



DOKAI SEO

Las Bases del GEO

→ *Una forma moderna de decir SEO bien hecho*

La guía práctica para entender GEO sin siglas vacías, sin humo y con alguna referencia al sidral de por medio.

AUTORES — **Marcelo Rojas Guerra** · **Marc Mayol Orell**

dokaiseo.com · Primera edición · 2026

道改

Las Bases del GEO

Marcelo Rojas Guerra · Marc Mayol Orell

Primera edición, 2026. DOKAI SEO · Oviedo, Asturias, España.

Este libro no es una biblia. Es una guía práctica basada en experiencia real. Si algo ha cambiado desde que lo escribimos —y en este sector algo siempre cambia— la lógica de fondo sigue siendo la misma. Las siglas pasan; los principios, no.

Queda permitida la reproducción parcial con atribución a los autores. Para uso comercial o formativo, consultar en dokaiseo.com.

Índice

1. Un SEO + un AI Engineer = acercamiento al GEO	4
2. GEO: el nombre odian los SEO pero que capta clientes	9
3. Cómo piensa un LLM (y por qué no es un buscador)	16
4. La base que nunca cambia: SEO técnico y de contenido.	23
5. Contenido para modelos generativos: escribir para que te usen	30
6. Digital PR y menciones en terceros	38
7. Medir la visibilidad en IA: lo que los números sí te dicen.	46
8. GEO local	55

00

Un SEO + un AI Engineer = acercamiento al GEO

Lo que viene después no es un libro de gurús. Es una guía escrita por un especialista en SEO y un ingeniero de IA que han cometido suficientes errores como para tener algo útil que contarte.

道改 CEO · DOKAI SEO · AI IMPACTO

Marcelo Rojas Guerra

CEO · DOKAI SEO · AI IMPACTO · Oviedo, Asturias

Empecé estudiando logopedia en Oviedo. No es el origen que nadie espera de alguien que ahora dirige una agencia SEO, pero tiene más sentido del que parece: la logopedia te enseña que la comunicación importa, que las palabras tienen estructura y que si el mensaje no llega claro, no llega. Eso es exactamente lo que hace el SEO bien hecho.

Mientras montaba una pequeña web para dar visibilidad a la profesión, descubrí que podía hacer que Google la encontrase. Y desde ahí ya no paré. Cuatro años trabajando para varias agencias a la vez, aprendiendo de noche lo que aplicaba de día, gestionando cuentas en España y América hasta que

tuve claro cuál era el enfoque que quería defender: SEO orientado a negocio, no a posiciones.

Hoy dirijo DOKAI SEO y AI IMPACTO, con la que ayudo a agencias a escalar sus departamentos de SEO con IA y automatización. Y escribo este libro porque llevo tiempo viendo cómo se vende GEO sin saber SEO, y eso me parece un problema que merece respuesta.

“El SEO es comunicación. La comunicación son emociones. Y las emociones te conectan con tu cliente. Eso es lo que hacemos en DOKAI SEO: conectar el posicionamiento lo que realmente genera ingresos.”

— Marcelo Rojas Guerra · CEO DOKAI SEO

Generative AI Tech Lead · CANCOM · Mallorca

Marc Mayol Orell

Generative AI Tech Lead · Educator · Islas Baleares

Yo no soy SEO. Soy ingeniero de IA, y eso es exactamente por qué estoy aquí. Si este libro lo hubieran escrito dos SEOs, te habrían contado qué hacer para aparecer en los LLMs pero sin saber del todo por qué funciona. Yo sé por qué funciona. He construido los sistemas por dentro.

Trabajo como AI Engineer senior en proyectos internacionales, doy clases de IA generativa y he desarrollado PromptSuite, un sistema de gestión de prompts con control de versiones para equipos que trabajan con GPT-4o, Claude, Gemini y otros modelos a la vez. He implementado sistemas RAG en producción, he liderado proyectos de automatización con LangChain y APIs de Anthropic, OpenAI y Google, y he visto con mis propios ojos cómo un modelo decide qué fuente usar y cuál ignorar.

Hay mucho contenido sobre cómo “optimizar para la IA” escrito por gente que no ha abierto un paper sobre LLMs en su vida. Esto es lo contrario.

“Los LLM son tus mejores amigos cuando empiezas a entender cómo funcionan”

— Marc Mayol Orell · Generative AI Tech Lead

Por qué escribimos esto

No escribimos este libro porque tuviéramos tiempo libre (de hecho todo lo contrario, por eso hemos tardado en sacar esta información). Lo empezamos a escribir entre 2025 y 2026 porque hemos visto el mismo patrón repetirse demasiadas veces, empresas pagando por “servicios de GEO” a personas que no saben hacer SEO. Y **sin SEO, no hay GEO.**

El problema no es la etiqueta. El problema es que la etiqueta se ha convertido en un salvoconducto para vender lo mismo de siempre con menos rigor y más PowerPoint. **Queríamos escribir algo que te dijera exactamente qué hay detrás del término**, qué es real, qué es hype y qué le deberías exigir a cualquier profesional que te hable de visibilidad en IA.

Y sí, usamos el término GEO conscientemente. Porque comunica bien. Cuando dices “quiero que ChatGPT me cite cuando alguien busque lo que yo vendo”, la gente lo entiende mejor que cuando dices “quiero optimizar mi contenido para la fase de inferencia de los modelos de lenguaje de gran escala”. El lenguaje técnico es necesario cuando es necesario. El resto del tiempo, estorba.

“Usamos GEO como se usa la etiqueta “SEO Local” que quien lo escucha sabe de qué estamos hablando sin necesidad de especificar demasiado.”

— Marcelo Rojas Guerra

Cómo usar esta guía

Puedes leer esto de principio a fin —lo ideal— o saltar directamente al capítulo que resuelve tu problema concreto —lo que sabemos que probablemente vas a hacer, y está bien. Lo hemos estructurado para que funcione de las dos formas.

Cada capítulo tiene una parte de conceptos, una parte de aplicación práctica y un ejercicio al final. Los ejercicios tienen siempre una referencia a Asturias, que es de donde vengo yo —Marcelo— y, como cualquier asturiano sabe, cualquier concepto complejo se explica mejor con comida, paisaje o lluvia de fondo. Marc lo ha aceptado con la paciencia que le dan años de lidiar con errores de producción en plena madrugada.

No vas a encontrar fórmulas mágicas aquí. Vas a encontrar lógica, criterio y suficiente contexto para tomar decisiones informadas en un entorno que cambia rápido. Lo que hoy llamas “estrategia GEO” puede tener otro nombre dentro de dos años. Lo que no cambia es la base: ser claro, ser relevante, ser citado.

Una nota sobre las fuentes

Nos apoyamos en experiencia propia con clientes reales y en datos publicados por investigadores del sector: Ahrefs, Kevin Indig, Vercel, el Princeton GEO paper de Aggarwal et al. Donde citamos datos, citamos la fuente. Donde damos una opinión, la marcamos como tal. Es lo mínimo exigible en cualquier guía que aspire a ser tomada en serio.

CAPÍTULO 01

01

GEO: el nombre odian los SEO *pero que capta clientes*

Antes de entrar en materia, toca hacerse la pregunta incómoda: ¿es GEO algo real o es otra sigla inventada para vender un servicio que ya existía? Nos la hemos hecho nosotros también. Y la respuesta honesta es las dos cosas a la vez.

De dónde viene el término

GEO son las siglas de Generative Engine Optimization. El concepto fue acuñado en un paper académico de 2023 firmado por investigadores de Princeton, Georgia Tech y otros centros: Pranjal Aggarwal, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, Karthik Narasimhan y Ameet Deshpande. El paper se titula simplemente “GEO: Generative Engine Optimization” y estudia cómo optimizar contenido para que los motores de búsqueda generativos —los que responden con texto en lugar de listas de enlaces— lo incluyan en sus respuestas.

Ese es el origen. No un consultor de marketing. No una agencia buscando diferenciarse. Un grupo de investigadores intentando nombrar algo que empezaba a pasar en la web: que el canal donde la gente buscaba

información estaba cambiando de forma, y que las reglas de visibilidad en ese nuevo canal no eran exactamente las mismas.

“GEO surge del mundo académico, no del marketing. Y eso importa, porque significa que hay rigor detrás del concepto, aunque luego mucha gente lo haya usado para vender servicios sin ese rigor.”

— Marcelo

El paper analizó qué técnicas de optimización de contenido hacían que un motor generativo como SGE de Google incluyera una fuente en su respuesta. Probaron varias estrategias: citar fuentes, incluir estadísticas, añadir citas textuales, mejorar la fluidez del texto, simplificar el lenguaje y otras. Y midieron qué combinaciones funcionaban mejor. El resultado no sorprenderá a nadie que lleve años en SEO: el contenido bien estructurado, con datos, fuentes claras y lenguaje preciso era el que más se citaba.

► Lo que dice el paper, en resumen

Citar fuentes de autoridad, incluir estadísticas concretas, usar definiciones claras y mejorar la legibilidad del texto aumenta la probabilidad de ser citado por un motor generativo. Exactamente lo mismo que lleva años funcionando para los featured snippets de Google.

Por qué el sector lo adoptó tan rápido

Cuando el paper salió, el sector SEO llevaba meses viendo cómo Google lanzaba las AI Overviews, ChatGPT empezaba a generar tráfico real hacia

webs externas y la pregunta de “cómo aparezco en las respuestas de la IA” no tenía todavía un nombre claro. GEO lo dio. Y cuando un sector ansioso por nominar lo nuevo encuentra una etiqueta que viene con aval académico, la adopción es instantánea.

El problema no fue la etiqueta. El problema fue lo que pasó después: mucha gente adoptó el nombre sin leer el paper. Y sin leer el paper no entiendes que las técnicas que propone son prácticamente las mismas que ya se usaban para optimizar contenido para Google. Lo que cambia no es la metodología base. Lo que cambia es el canal donde esa metodología se aplica y la forma de medir los resultados.

Hay un tweet de John Mueller de Google de 2025 que resume esto mejor que cualquier paper: “Si estás optimizando para búsqueda, entonces en cierto modo también estás optimizando para las AI Overviews y el AI mode”. Lo dijo el propio responsable de comunicación técnica de Google. No hace falta añadir mucho más.

Por qué nosotros usamos GEO igualmente

Voy a ser directo: antes que nada somos parte del marketing. Y en marketing, el nombre vende. Si un cliente llega dispuesto a pagar 1.500 euros por GEO, adelante. No voy a ponerme a explicarle en esa reunión que GEO es básicamente SEO bien hecho con una capa nueva encima, porque entonces pierdo el cliente y se va con alguien que usa el vocabulario que él espera. Los clientes hoy no buscan SEO. Buscan GEO. Y en el mejor de los casos buscan SEO/GEO. El mercado ha hablado y yo no voy a llevarle la contraria.

Pero hay una segunda razón, más de fondo, que me gusta más: el término GEO marca una especialización real. Y las especializaciones merecen

nombre propio. Piensa en el SEO técnico: es SEO, sí, pero quien lo domina de verdad te está hablando de crawling, JavaScript rendering, log file analysis, Core Web Vitals, arquitectura de servidor... todo lo que te da la programación y más. Una persona puede saber un poco de todo eso. Pero se nota en dos minutos quién tiene profundidad real. Lo mismo pasa con el SEO de contenidos: también es SEO, pero quien lo domina te habla de topicalidad, clusterización, intención de búsqueda a escala, Information Gain... son mundos.

El GEO pasa exactamente lo mismo. Es una especialización dentro del SEO que abre su propio abanico: entender cómo se entrena un LLM, cómo funciona RAG en producción, cómo medir la visibilidad en IA, cómo estructurar contenido para que un modelo lo use y cómo interpretar lo que la IA dice de una empresa para ajustar la estrategia. Marc y yo no somos los mismos profesionales que éramos antes de profundizar en esto. Y esa profundidad tiene un precio que se justifica sola: porque bien aplicada multiplica la facturación orgánica de forma medible.

“Hay profesionales que saben un poco de todo. Levantan proyectos de cero, que es valioso. Pero hay profesionales con profundidad real que cogen una empresa que factura 100.000 euros al mes y la llevan a medio millón. No porque sean mágicos. Sino porque saben exactamente qué palanca mover y cuándo. El SEO básico escala sitios. La especialización escala negocios.”

— Marcelo

Y la capa de medición que trae el GEO no es un detalle menor. Saber qué dice ChatGPT de tu empresa, en qué contexto te menciona, cómo te describe frente a tu competencia y cómo evoluciona eso en el tiempo abre un abanico

de inteligencia estratégica que el SEO tradicional sencillamente no tenía porque el canal no existía. No sustituye las métricas clásicas. Las complementa. Y para empresas que quieren pasar de facturar 100.000 al mes a medio millón, esa información extra es la diferencia entre una estrategia que intuyes y una que mides y escalas con criterio.

Lo que GEO añade que el SEO clásico no contemplaba

Dicho todo lo anterior, sería deshonesto decir que GEO es exactamente igual al SEO de siempre. Hay una capa que el SEO tradicional no tenía porque no existía el canal: la monitorización de cómo te percibe y te cita la IA.

En SEO clásico mides posiciones, impresiones, clics, conversiones. Eso sigue siendo válido. Pero ahora hay una variable nueva: ¿qué dice ChatGPT cuando alguien le pregunta por lo que tú vendes? ¿Te menciona? ¿En qué contexto? ¿Con qué precisión? ¿Cómo te describe? ¿Y qué dice de tu competencia?

Esa capa de monitorización y ajuste específico para entornos generativos es lo que nosotros llamamos GEO en la práctica. No es magia. Es una extensión lógica de lo que ya hacías. Pero requiere herramientas distintas, métricas distintas y, sobre todo, entender cómo funciona un LLM por dentro para saber qué ajustar. De eso trata el capítulo siguiente.

SEO clásico vs. GEO: lo que cambia y lo que no

	SEO clásico	GEO
Objetivo principal	Posición en SERP	Ser citado en respuestas de IA
Base técnica	Arquitectura, indexación, contenido	Igual — la base no cambia
Contenido	Relevante para la intención de búsqueda	Estructurado para que el LLM lo entienda
Medición	Posiciones, clics, conversiones	Menciones, contexto, precisión de la cita
Ciclo de resultados	Semanas / meses	Meses (entrenamiento) o días (RAG)
Digital PR	Links de autoridad	Menciones en fuentes que el LLM conoce
Novedad real	---	Capa de monitorización de visibilidad en IA

El hype y la realidad

Hay dos extremos que nos parecen igual de peligrosos. El primero: vender GEO como algo revolucionario que reemplaza al SEO y requiere empezar de cero. Eso es humo. El segundo: descartar GEO como una moda sin sustancia y no hacer nada. Eso es quedarse fuera de un canal que ya genera tráfico que convierte 23 veces mejor que el tráfico orgánico tradicional, según los datos de Ahrefs.

La posición honesta está en el medio: GEO es una evolución del SEO, no una ruptura. Requiere entender cómo funcionan los LLMs —que es exactamente

lo que veremos en el capítulo 02—, aplicar la misma base técnica de siempre y añadir una capa de monitorización específica. Ni más ni menos.

El tráfico que llega desde ChatGPT o Perplexity representa hoy solo el 0,5 % del total del tráfico web según Ahrefs, pero genera el 12,1 % de los leads. Los usuarios que llegan desde un chatbot ya han pasado por un proceso de consulta más profundo, llegan más preparados y convierten más. Para algunas empresas ese 0,5 % ya supone el 10 % de sus conversiones. El canal es pequeño hoy. La trayectoria no invita a ignorarlo.

► **El dato que cambia la conversación**

Ahrefs analizó datos reales de tráfico y encontró que el tráfico procedente de IA convierte unas 23 veces mejor que el tráfico orgánico tradicional. No es una promesa de futuro. Son números actuales.

✨ **MITOLOGÍA ASTURIANA**

El Cuélebre y el tesoro que nadie encuentra

En la mitología asturiana, el Cuélebre es una serpiente alada enorme que vive en cuevas y grutas ocultas en la montaña. Su función es guardar un tesoro. Un tesoro real, valioso, que existe de verdad. El problema no es el tesoro. El problema es que nadie sabe llegar hasta él.

El Cuélebre no sale a buscar visitantes. No pone un cartel en la carretera. No tiene Google Business Profile. Está ahí, con su oro, esperando que alguien lo encuentre por casualidad o por leyenda transmitida de boca en boca. Y la mayoría de las veces, el tesoro se queda en la cueva.

Muchas empresas son exactamente el Cuélebre: tienen un servicio excelente, un conocimiento real, un valor diferencial claro. Pero la IA no sabe que existen. No las menciona. No las cita. No las recomienda. El tesoro está ahí. Pero si

ChatGPT no sabe dónde está la cueva, para el usuario que pregunta, es como si no existiera.

El GEO es, en parte, aprender a salir de la cueva y dejar rastro en los sitios donde la IA aprende. No para el Cuélebre. Para el viajero que llega con la pregunta correcta y necesita que alguien le diga a dónde ir.

CAPÍTULO 02

02

Cómo piensa un LLM (y por qué no es un buscador)

Para influir en algo necesitas entender cómo funciona. No a nivel de paper académico, sino lo suficiente como para tomar decisiones que tengan sentido. Este capítulo es eso: lo que necesitas saber del interior de un LLM para que tu estrategia GEO no se base en suposiciones.

Un LLM no busca. Recuerda.

Lo primero que hay que entender es que un modelo de lenguaje grande —ChatGPT, Gemini, Claude, Perplexity— no funciona como Google. Google

sale a buscar en tiempo real. Un LLM, en su estado base, no busca nada. Responde desde lo que aprendió durante el entrenamiento.

El entrenamiento es el proceso en el que el modelo ingiere cantidades enormes de texto de internet, libros, artículos, foros, documentación técnica, y ajusta millones de parámetros internos para aprender patrones del lenguaje y del conocimiento del mundo. Cuando termina ese proceso, el modelo queda congelado en ese momento. Si tu empresa no existía, si tu contenido era débil o si nadie hablaba de ti en sitios que el modelo ingirió, sencillamente no estás en su memoria.

Esto tiene una implicación directa y que mucha gente no considera: los ciclos de reentrenamiento son lentos. Un modelo puede llevar meses o incluso años con un corte de conocimiento determinado. Si tu empresa es nueva o ha crecido mucho recientemente, puede que el modelo esté hablando de una versión de ti que ya no existe. Y lo hará con toda la confianza del mundo, porque él no sabe que está desactualizado.

“Yo siempre lo explico igual: un LLM es como un estudiante muy listo que se preparó un examen hace seis meses y ahora contesta desde la memoria. Si la información que estudió era buena, da buenas respuestas. Si estudiaste mal o no aparecías en sus apuntes, en el examen no existes.”

— Marc

► Qué significa esto para tu estrategia

Si quieres aparecer en la memoria del modelo a largo plazo, necesitas que fuentes de autoridad hablen de ti antes del próximo ciclo de entrenamiento. El Digital PR en medios y webs relevantes no es opcional: es el mecanismo por el que entras en los apuntes del estudiante.

RAG: cuando el modelo sale a buscar

Pero los LLMs modernos no trabajan solo desde la memoria. Existe una arquitectura llamada RAG, Retrieval-Augmented Generation, que en la práctica significa que el modelo, en el momento en que genera una respuesta, puede salir a consultar fuentes externas y usar esa información para completar o actualizar lo que sabe.

ChatGPT con búsqueda activada funciona así. Perplexity funciona así prácticamente siempre. Las AI Overviews de Google también. Cuando el sistema detecta que la pregunta requiere información fresca o muy específica, lanza una búsqueda, recupera fragmentos de páginas web y los integra en la respuesta final.

Esto cambia el juego de forma importante: ya no solo depende de si estabas en el entrenamiento. En el modo RAG, tu contenido puede ser consultado en tiempo real. Lo que significa que el SEO técnico y la calidad del contenido impactan directamente en si el sistema te recupera o no. El pipeline es siempre el mismo: primero recupera candidatos, luego filtra por relevancia y autoridad, finalmente genera la respuesta. Si no pasas el primer filtro, da igual todo lo demás.

“En RAG, el ranking clásico sigue siendo la puerta de entrada. Si tu contenido no pasa el filtro de recuperación, el LLM nunca lo verá. Da igual lo bien escrito que esté.”

— Marc

Cómo decide el modelo si buscar o no

No todos los modelos buscan siempre. ChatGPT, por ejemplo, tiene un clasificador interno llamado SONIC que evalúa la consulta del usuario y decide si lanza una búsqueda externa o responde desde la memoria. Si la probabilidad supera un umbral —en el caso analizado, 0,54— el modelo busca. Si no, responde solo desde lo que sabe.

Esto significa que para consultas estáticas o de conocimiento general, el modelo tira de su memoria entrenada y tu contenido actual no entra en la ecuación. Para consultas que el sistema percibe como recientes, locales o muy específicas, sí activa la búsqueda y ahí tu SEO técnico y tu contenido actual sí importan directamente.

Además, GPT-5 tiene implementado algo parecido al sistema QDF de Google —Query Deserves Freshness— con una escala del 0 al 5. Una consulta con QDF 5 requiere información del mes en curso. Una con QDF 0 es histórica y el modelo puede responder sin buscar. Saber en qué banda entra tu sector te dice mucho sobre cuánto peso tiene tu contenido actual frente a lo que el modelo ya tenía aprendido.

► QDF en la práctica

Precios, eventos, noticias del sector, lanzamientos de productos → QDF alto, el modelo busca. Historia de tu empresa, definiciones, conceptos de tu industria → QDF bajo, el modelo responde desde la memoria. Conocer esto te ayuda a priorizar dónde poner la energía.

El mito del chunking y por qué no debes obsesionarte con él

Desde que se popularizó el término GEO, apareció una corriente que dice que hay que trocear el contenido de la web en páginas pequeñísimas y ultra específicas para que el LLM lo ‘chunkee’ mejor. La lógica suena razonable: si los LLMs trabajan por fragmentos, démosles fragmentos perfectos.

El problema es que nadie sabe exactamente cómo chunkear problema cada sistema. ¿500 tokens? ¿1.000? ¿Con solapamiento del 20%? ¿Por estructura semántica o por cortes del DOM? Perplexity ha reconocido que usa fragmentos de apenas 7 tokens. Optimizar para eso es perder el tiempo. Estás ajustando tu contenido para un algoritmo que no conoces y que además cambia.

La recomendación de Danny Sullivan de Google en Search Off the Record fue directa: no trocees tu contenido pensando que así le gusta a los LLMs. Y añadió algo importante: aunque hoy te funcionara, mañana los sistemas cambian. El contenido consolidado en una URL sigue funcionando mejor. Google ya usa Passage Ranking desde 2020, lo que significa que el sistema extrae los fragmentos relevantes por ti. No necesitas hacer ese trabajo tú de forma manual.

“Trocear el contenido para la IA es como cortar la fabada en trocitos pequeños pensando que así se cocina mejor. La fabada necesita tiempo y calor, no fragmentación. El contenido bueno necesita profundidad y estructura, no atomización.”

— Marcelo

La regla práctica que usamos nosotros es más simple: si un humano puede entender un párrafo sin necesitar el contexto anterior, cualquier sistema de chunking funcionará bien con él. Escribe claro, estructurado y con lógica propia en cada bloque. El resto lo hace el sistema.

Los dos escenarios que debes tener claros

Resumiendo todo lo anterior, cuando piensas en GEO tienes que tener en mente dos escenarios distintos porque requieren estrategias distintas:

	Escenario 1 · Memoria	Escenario 2 · RAG
Cómo responde	Desde el entrenamiento previo	Consultando fuentes en tiempo real
Quién decide	El propio modelo	El clasificador de queries (SONIC)
Qué necesitas	Digital PR en fuentes que ingirió el modelo	SEO técnico + contenido de calidad indexado
Velocidad de impacto	Meses (próximo ciclo de entrenamiento)	Días o semanas si estás bien indexado
Tu mayor riesgo	No aparecer en sus datos de entrenamiento	No pasar el filtro de recuperación

Lo que esto significa en la práctica

Marc y yo llegamos a la misma conclusión desde sitios distintos. Yo desde el SEO: la base técnica y de contenido que siempre ha funcionado sigue siendo el filtro de entrada. Sin eso, no hay GEO posible. Marc desde la ingeniería: los sistemas RAG siguen dependiendo de ranking clásico para decidir qué documentos ver. Si no pasas ese filtro, el LLM nunca ve tu contenido.

La diferencia con el SEO clásico es que ahora hay una variable adicional: la memoria del modelo. Y esa memoria se construye con tiempo, con presencia en fuentes de autoridad y con contenido que otros citan. No se compra, no se instala en un fin de semana y no se arregla con un truco técnico. Se construye igual que se construye la reputación: despacio, con consistencia y con criterio.

En los capítulos siguientes vamos a ver exactamente cómo hacer eso: cómo estructurar el contenido para que los modelos lo entiendan y lo usen, cómo construir presencia en las fuentes que importan y cómo medir si todo eso está funcionando.

✨ MITOLOGÍA ASTURIANA

La Xana y el secreto que solo ella sabe

En la mitología asturiana, las Xanas son hadas que viven en fuentes, ríos y manantiales de la montaña. Son hermosas, sabias y guardan un conocimiento muy valioso: saben dónde están los tesoros ocultos, quién ha pasado por el bosque y qué ocurrió hace décadas en cada valle.

Pero la Xana no comparte ese conocimiento con cualquiera. Solo habla con quien se ha ganado su confianza: quien respeta el bosque, quien ha estado allí antes, quien ha dejado huella en su entorno. Con los extraños que llegan de repente pidiendo respuestas, la Xana guarda silencio.

Un LLM funciona de forma parecida. Su conocimiento viene de lo que ha leído, de las fuentes que ha procesado, de las voces que se repetían en sus datos. Si tu empresa nunca apareció en esas fuentes, si nadie habló de ti en los sitios que el modelo ingirió, la Xana no sabe que existes. Y cuando alguien le pregunta, guarda silencio o nombra a otro.

La estrategia GEO es, en parte, ganarse la confianza de la Xana antes de que alguien haga la pregunta. Dejar rastro en los lugares correctos, con la consistencia suficiente, para que cuando el modelo necesite recordar quién hace lo que tú haces, tu nombre esté en su memoria.

03

La base que nunca cambia: **SEO técnico y de contenido.**

Hay algo que se repite en cada conversación sobre GEO con clientes: esperan que les hablemos de cosas nuevas, de trucos específicos para la IA, de configuraciones especiales. Y nosotros siempre empezamos por lo mismo: ¿cómo está tu SEO técnico? La cara que ponen en ese momento lo dice todo.

Por qué el SEO técnico es la puerta de entrada

En el capítulo anterior vimos que el pipeline de un sistema RAG tiene tres fases: recuperar, filtrar y generar. La recuperación es el primer filtro y es el más crítico, porque si tu contenido no se recupera, nunca llega a la fase de generación. Y la recuperación depende de los mismos factores que han determinado el ranking en Google durante años: indexación limpia, arquitectura lógica, contenido que el rastreador puede leer.

Los crawlers de los LLMs no son muy distintos a Googlebot en lo fundamental. Siguen enlaces, leen HTML, procesan texto. Si tienes JavaScript que bloquea el contenido principal, si tu robots.txt está mal configurado, si tus páginas importantes no tienen enlazado interno suficiente, si tu web tarda diez segundos en cargar... el crawler puede ignorarte o indexarte mal. Y un contenido mal indexado es un contenido que no existe para el modelo.

Perplexity, en una entrevista reciente, confirmó que usan una forma de PageRank para priorizar fuentes. ChatGPT en modo búsqueda también usa señales de autoridad y relevancia para decidir qué documentos recuperar. Es decir, los mismos principios que definen la calidad en Google —lo que internamente llaman Q^* y T^* , calidad y topicalidad— siguen siendo los filtros que determinan si apareces o no en las respuestas de la IA.

“Cada vez que llega un cliente preguntándonos por GEO y le revisamos el site, encontramos lo mismo: problemas de indexación que llevan meses ahí. El GEO no es la solución a eso. Es una capa más encima. Sin la base, no hay donde construir.”

— Marcelo

► Lo que revisamos siempre primero

Cobertura de indexación en Search Console, errores de rastreo, páginas bloqueadas en robots.txt, contenido en JavaScript no renderizado, velocidad de carga y Core Web Vitals, estructura de enlazado interno. Si algo de esto está roto, es la primera prioridad antes de cualquier otra cosa.

Arquitectura que los modelos pueden entender

Más allá de la indexación, hay algo que separa un sitio que los LLMs usan bien de uno que ignoran aunque lo indexen correctamente: la claridad de la estructura. Un LLM en fase de inferencia no navega tu web como lo haría un humano. Recupera fragmentos. Y para que esos fragmentos tengan sentido de forma aislada, la arquitectura del contenido tiene que ser muy clara.

Esto significa varias cosas concretas. Una URL por tema, sin canibalizar. Jerarquía de cabeceras coherente: el H1 dice de qué va la página, los H2 dividen los subtemas principales, los H3 desarrollan cada subtema. Párrafos cortos con una sola idea cada uno. Listas cuando el contenido es enumerativo. Definiciones claras cuando introduces un concepto técnico. Nada de esto es nuevo. Es exactamente lo que lleva años funcionando para los featured snippets.

El enlazado interno también importa más de lo que parece en el contexto GEO. No solo para el rastreo: para la coherencia semántica. Cuando tus páginas se enlazan entre sí con anclajes descriptivos, le estás diciendo al modelo qué relación tienen esos contenidos. Eso construye la red de conocimiento que el sistema usa para entender de qué eres referente.

► La prueba del párrafo aislado

Coge cualquier párrafo de tu web y léelo sin contexto. ¿Tiene sentido solo? ¿Se entiende de qué trata sin necesitar leer lo anterior? Si la respuesta es no, ese párrafo no está preparado para ser chunkeado por un LLM. Reescríbelo hasta que sí.

Featured snippets y AI Overviews: la misma lógica

Uno de los argumentos más sólidos para entender GEO sin inventarse nada nuevo es este: el contenido que aparece en featured snippets de Google es prácticamente el mismo que aparece en las AI Overviews y en las respuestas de ChatGPT cuando busca en la web. No es casualidad. Es la misma lógica aplicada al mismo tipo de contenido.

¿Qué tienen en común los contenidos que se llevan los featured snippets? Responden una pregunta concreta de forma directa al principio. Están bien estructurados con cabeceras y listas cuando el tema lo permite. Usan lenguaje claro sin rodeos. Incluyen datos cuando el tema lo requiere. Citan fuentes cuando hablan de estadísticas o hechos verificables. Exactamente eso mismo es lo que el paper de GEO de Princeton identificó como las técnicas que aumentan la visibilidad en motores generativos.

La evolución de Google en los últimos diez años apunta siempre en la misma dirección: Hummingbird en 2013 para entender significados, BERT en 2018 para captar matices y contexto, MUM en 2021 para conectar formatos y lenguajes, AI Overviews en 2024 como capa generativa encima del ranking clásico. Cada paso es una capa nueva sobre la misma base. El contenido que ha sobrevivido a todos estos cambios tiene algo en común: es claro, es útil y está bien estructurado.

“Si llevas años optimizando para featured snippets y tu contenido aparece en ellos con regularidad, ya estás haciendo GEO sin saberlo. La diferencia es que ahora hay un canal más donde ese trabajo paga.”

— Marcelo

Contenido que los LLMs pueden usar: cinco principios

Hay cinco principios concretos que aplicamos cuando auditamos o creamos contenido pensando en visibilidad en IA. No son exclusivos del GEO. Son principios de buen contenido que en el contexto de los LLMs tienen más peso todavía.

1. Responde primero, desarrolla después

El modelo recupera fragmentos. Si la respuesta a la pregunta principal está enterrada en el quinto párrafo después de una introducción larga, el sistema puede recuperar el fragmento equivocado o no recuperarte en absoluto. Pon la respuesta directa al principio. Luego desarrolla el contexto, los matices y los ejemplos.

2. Datos y fuentes siempre que puedas

Los LLMs priorizan contenido que perciben como fiable. Un dato con fuente citada tiene más peso que una afirmación sin respaldo. No se trata de llenar el texto de estadísticas: se trata de que cuando hagas una afirmación relevante, la respaldes. Esto es exactamente lo mismo que lleva años funcionando para la E-E-A-T de Google.

3. Definiciones claras de los conceptos de tu sector

Si eres la fuente que mejor explica los términos técnicos de tu industria, el modelo aprenderá a referenciarte cuando alguien pregunta por esos términos. Glosarios, preguntas frecuentes bien respondidas, explicaciones de conceptos complejos en lenguaje accesible: todo eso construye autoridad semántica que los LLMs reconocen.

4. Sin ironía ni ambigüedad en el contenido estratégico

Los modelos de lenguaje procesan el texto de forma literal en la mayor parte de los casos. La ironía, el doble sentido y el lenguaje figurado pueden funcionar muy bien para conectar con lectores humanos, pero confunden a los sistemas de recuperación. En el contenido que quieres que la IA use —páginas de servicio, artículos de referencia, definiciones— opta por la claridad directa. La creatividad, para el blog y las redes.

5. Consistencia entre lo que dices en tu web y lo que dicen de ti fuera

Si tu web dice que eres especialista en SEO para e-commerce pero todos los artículos que has publicado fuera hablan de SEO local, el modelo tiene señales contradictorias sobre quién eres. Esa incoherencia diluye tu autoridad semántica. La consistencia entre lo que publicas, lo que dicen de ti y cómo te describes es más importante en GEO que en el SEO tradicional porque el modelo la usa para decidir en qué conversaciones te menciona.

✨ MITOLOGÍA ASTURIANA

El Nuberu y el camino que solo existe si está despejado

En la mitología asturiana, el Nuberu es el señor de las nubes y las tormentas. Viaja por el cielo montado en una nube negra y puede oscurecer el camino con la misma facilidad con que lo despeja. Los aldeanos sabían algo importante sobre el Nuberu: no podías negociar con él, pero sí podías prepararte para cuando pasara. Limpiar el camino, marcar los hitos, dejar las sendas transitables.

Un sitio web con problemas técnicos es como un camino lleno de barro bajo el Nuberu. Aunque el destino sea bueno, nadie llega. El crawler no llega, el modelo no indexa, el contenido no se recupera. El Nuberu pasa por encima y sigue de largo.

Limpiar el camino no es la parte emocionante del trabajo. No hay una diapositiva bonita que mostrar cuando arreglas el robots.txt o cuando corriges los errores de rastreo. Pero es lo que hace que todo lo demás funcione. El contenido más brillante del mundo no sirve de nada si el Nuberu no puede verlo desde su nube.

Antes de hablar de estrategia GEO, asegúrate de que el camino está despejado.

CAPÍTULO 04

04

Contenido para modelos generativos: **escribir para que te usen**

El capítulo anterior estableció que la base técnica no cambia. Este capítulo entra en el siguiente nivel: cómo estructurar y escribir el contenido para que un LLM no solo lo encuentre, sino que lo use en sus respuestas. Hay diferencia entre estar indexado y ser citado. Esto trata de lo segundo.

La diferencia entre estar indexado y ser citado

Puedes tener una web perfectamente indexada, con SEO técnico impecable, y aun así no aparecer en las respuestas de la IA. Porque hay un salto entre “este contenido existe y lo puedo recuperar” y “este contenido es lo suficientemente claro y autoritativo como para usarlo en mi respuesta”.

Ese salto es lo que este capítulo aborda. Y el criterio que separa un lado del otro no es misterioso: es el mismo que distingue un buen artículo de referencia de uno mediocre. Responde preguntas concretas. Usa lenguaje preciso. Respalda sus afirmaciones. Tiene estructura que facilita la lectura fragmentada. Y transmite autoridad porque quien lo escribe sabe de lo que habla.

Lo que cambia con los LLMs es que ese criterio ahora tiene consecuencias dobles: te afecta tanto en el ranking de Google como en si la IA te cita o no. Optimizar para uno optimiza para el otro. No son dos trabajos distintos. Son el mismo trabajo con más visibilidad de sus resultados.

“Cuando reescribo una página para que la IA la use mejor, siempre acaba mejor en Google también. No es coincidencia. Es que los criterios de calidad son los mismos. La IA solo los hace más obvios.”

— Marcelo

Cómo estructurar una página para que el LLM la use

Hay un patrón de estructura que funciona consistentemente bien tanto para los featured snippets de Google como para las respuestas de los LLMs. No es una fórmula rígida, pero sí una lógica que merece la pena interiorizar.

Abre con la respuesta directa

El primer párrafo de cualquier página importante debe responder la pregunta principal que esa URL intenta resolver. Sin introducción de contexto, sin “a lo largo de este artículo veremos”, sin calentamiento. La respuesta primero. El desarrollo después. Si alguien llega buscando “qué es el GEO”, el primer párrafo tiene que responder eso en dos o tres frases. El resto del contenido amplía, matiza y profundiza.

Usa cabeceras como preguntas o afirmaciones completas

Una cabecera como “Características” no le dice nada al modelo sobre el contenido que viene debajo. Una cabecera como “¿Qué características debe tener un contenido para ser citado por la IA?” ya es en sí misma una unidad

de información. El modelo puede recuperar esa cabecera y usarla directamente. Escribe las cabeceras pensando en que cualquiera de ellas podría aparecer como extracto en una respuesta de IA.

Párrafos cortos con una sola idea

Cada párrafo debería poder leerse de forma aislada y tener sentido completo. Si para entender un párrafo necesitas haber leído los tres anteriores, ese párrafo no está preparado para ser fragmentado por un sistema RAG. La regla práctica: si supera las cinco o seis líneas, probablemente haya dos ideas dentro que merece la pena separar.

Listas para contenido enumerativo

Cuando el contenido es una serie de puntos, pasos o características, las listas funcionan mejor que los párrafos continuos tanto para la lectura humana como para la recuperación por parte de los modelos. Una lista bien formada es fácil de fragmentar, de citar y de incluir en una respuesta estructurada. La clave es que cada ítem de la lista sea autocontenido: que se entienda sin leer los demás.

Tablas para comparativas y datos estructurados

Las tablas son uno de los formatos que mejor funcionan para los LLMs en modo RAG porque ya tienen la información estructurada de forma que el modelo puede procesarla directamente. Comparativas de productos, diferencias entre conceptos, rangos de precios, plazos... cualquier información que pueda ponerse en formato tabla, ponla en tabla.

El contenido que construye autoridad semántica

Ser citado puntualmente está bien. Ser la fuente de referencia que el modelo usa cuando alguien pregunta por tu sector es otra cosa. Eso segundo se construye con autoridad semántica: ser el sitio que mejor explica los conceptos de tu industria de forma consistente y en profundidad.

Hay tres tipos de contenido que construyen autoridad semántica de forma especialmente eficiente en el contexto GEO. El primero son las definiciones: páginas que explican claramente qué significa un término técnico o concepto de tu sector. Si eres la fuente que mejor define “Qué es el ROI en SEO” o “Cuál es la diferencia entre tráfico orgánico y de pago”, el modelo aprenderá a referenciarte para esas consultas. El segundo son las comparativas: análisis honestos de opciones, herramientas o enfoques distintos. El modelo usa este tipo de contenido cuando alguien pide una recomendación. El tercero son los casos con datos reales: resultados concretos, números, metodologías aplicadas. La IA prioriza el contenido que respalda sus afirmaciones con evidencia.

Lo que une estos tres tipos es algo que Marc siempre enfatiza desde el lado técnico: el modelo evalúa la coherencia entre lo que dices y lo que corroboran otras fuentes. Si escribes que eres referente en GEO pero no hay ningún otro sitio que lo confirme, el modelo tiene una sola fuente y la trata con cautela. Si hay tres, cinco, diez fuentes distintas que te mencionan en ese contexto, la señal se amplifica.

► El contenido que más rápido impacta en visibilidad GEO

Páginas de preguntas frecuentes bien respondidas, glosarios de términos del sector, comparativas honestas de herramientas o metodologías, y estudios de caso con datos reales. Son los formatos que los LLMs recuperan y citan con más frecuencia en respuestas informativas.

Reescribir tu ‘Sobre nosotros’ para que la IA lo entienda

La página ‘Sobre nosotros’ o ‘Quiénes somos’ es una de las más ignoradas en SEO y una de las más importantes para GEO. Es la página donde le dices al mundo, y al modelo, exactamente quién eres, qué haces, para quién lo haces y por qué importa. Si está escrita con frases vagas como “Somos una empresa comprometida con la excelencia”, el modelo no sabe nada de ti que pueda usar en una respuesta.

Una buena página ‘Sobre nosotros’ para GEO responde estas preguntas con precisión: ¿Qué hace exactamente tu empresa? ¿Para qué tipo de cliente? ¿En qué sector o nicho específico? ¿En qué geografía opera? ¿Cuánto tiempo lleva haciéndolo? ¿Qué resultados concretos ha conseguido? ¿Quiénes son los responsables y qué trayectoria tienen? Cada una de esas preguntas respondida con datos concretos es una pieza de información que el modelo puede usar cuando alguien pregunte por empresas de tu tipo en tu zona.

“Le pedí a ChatGPT que me describiera DOKAI SEO antes de reescribir la web. Lo que dijo era vago y generico. Después de reescribir las páginas clave con datos concretos, la descripción cambió completamente. El

modelo literalmente cambia cómo te presenta cuando le das mejor información con la que trabajar.”

— Marcelo

Lo que no funciona aunque parezca lógico

Hay tres errores que vemos con mucha frecuencia en contenido que supuestamente está optimizado para GEO y que en la práctica no funciona o incluso perjudica.

El primero es saturar el contenido de palabras clave pensando que el modelo las detecta como el SEO clásico. Los LLMs no funcionan por frecuencia de términos de la misma forma que BM25. Entienden contexto y semántica. Un texto natural y fluido que trata un tema en profundidad funciona mejor que uno lleno de repeticiones forzadas.

El segundo es crear contenido artificialmente corto para ‘facilitar el chunking’. Ya lo vimos en el capítulo anterior: nadie sabe cómo chunka cada sistema, y el contenido consolidado con profundidad real sigue funcionando mejor. Una página delgada sin sustancia no tiene ventaja competitiva sobre una página densa y bien estructurada.

El tercero, y quizás el más sutil, es publicar mucho contenido mediocre pensando que el volumen compensa la calidad. En el contexto GEO esto es especialmente dañino: el modelo aprende patrones. Si el patrón que aprende de tu dominio es ‘contenido superficial sobre muchos temas’, eso es lo que recordará cuando alguien pregunte por tu sector. La profundidad en pocos temas construye más autoridad semántica que la superficialidad en muchos.

► La prueba de la autoridad real

Si le preguntas a ChatGPT por el tema en el que quieres ser referente y no apareces entre las primeras fuentes citadas, hay dos posibilidades: o tu contenido no llega al nivel de profundidad necesario, o no tienes suficientes menciones externas que corroboren tu autoridad. Normalmente es las dos cosas.

✨ MITOLOGÍA ASTURIANA

El Trasgu y el desorden que nadie ve

El Trasgu es uno de los personajes más conocidos de la mitología asturiana. Es un duende travieso que vive en las casas y se dedica a hacer el desorden cuando nadie lo ve: mueve objetos de sitio, esconde cosas, revuelve lo que estaba ordenado. Los aldeanos sabían que el Trasgu no era malicioso, solo caótico. El problema es que su caos hacía muy difícil encontrar lo que se necesitaba cuando se necesitaba.

El contenido mal estructurado es el Trasgu de tu web. La información está ahí, en alguna parte, pero nadie —ni el usuario ni el modelo— puede encontrarla cuando la necesita. Los conceptos están mezclados. Las respuestas aparecen en el quinto párrafo. Las cabeceras no dicen nada. Los párrafos tienen tres ideas cada uno y ninguna se entiende sin el contexto del anterior.

Expulsar al Trasgu de tu web no requiere escribir más. Requiere ordenar lo que ya tienes. Respuesta primero, contexto después. Cabeceras que informan. Párrafos que se entienden solos. Estructura que cualquiera —humano o modelo— pueda seguir sin perderse.

El buen contenido no es el que más dice. Es

05

Digital PR y menciones en terceros

Puedes tener el mejor contenido del mundo, perfectamente estructurado y con SEO técnico impecable. Si nadie habla de ti fuera de tu web, el modelo tiene una sola fuente. Y una sola fuente no construye autoridad. Esto trata de cómo conseguir que las fuentes que los LLMs respetan te mencionen antes de que alguien haga la pregunta.

Por qué el modelo aprende de lo que dicen de ti

Los LLMs se entrenan con texto de internet. No solo con tu web. Con todo lo que hay publicado sobre tu sector, tu empresa, tu nombre y tu trabajo en cualquier fuente que el sistema ingirió durante el entrenamiento. Eso significa que lo que dicen de ti en medios, blogs especializados, podcasts con transcripción, foros de tu sector, newsletters o directorios de autoridad forma parte del perfil que el modelo tiene de tu empresa.

Si ese perfil externo es escaso o inexistente, el modelo tiene muy poco con qué trabajar cuando alguien pregunta por empresas de tu tipo. Y cuando tiene poco con qué trabajar, tira de lo que mejor conoce: los actores con más presencia en sus datos de entrenamiento, que casi siempre son los más grandes o los más activos en generar presencia externa. No los mejores necesariamente. Los más visibles en el ecosistema de fuentes que el modelo ingirió.

Marc lo explica desde la arquitectura: el modelo asigna un peso implícito a cada fuente de información durante el entrenamiento. Una mención en un medio de referencia de tu sector tiene más peso que diez menciones en directorios de baja calidad. La coherencia entre fuentes también cuenta: si cinco sitios distintos te describen de la misma forma como especialista en lo mismo, esa consistencia refuerza la representación que el modelo construye de ti.

“El Digital PR para GEO no es una novedad. Es lo que siempre ha sido el link building para SEO, con una diferencia: ahora importa tanto el enlace como el texto que lo rodea. Lo que dice la fuente sobre ti es tan relevante como el hecho de que te mencione.”

— Marcelo

Qué fuentes priorizan los modelos

No todas las menciones tienen el mismo valor. Los modelos aprenden con más fuerza de unas fuentes que de otras, y aunque no tenemos acceso a los datos exactos de entrenamiento, sí podemos inferir la jerarquía a partir de lo que sabemos sobre cómo se construyen esos corpus.

Medios y publicaciones especializadas del sector

Las publicaciones de referencia de tu industria son la fuente de mayor peso. Si tu sector tiene medios con trayectoria y credibilidad, aparecer en ellos —ya sea en artículos, entrevistas, op-eds o análisis— es la forma más eficiente de construir presencia en los datos de entrenamiento. En el sector SEO en español esto incluye medios como Marketing4eCommerce, Semrush Blog, Ahrefs Blog en español, newsletters especializadas con audiencia real y podcasts con transcripción disponible.

Wikis, Wikipedia y fuentes enciclopédicas

Wikipedia es uno de los corpus más usados en el entrenamiento de LLMs por su estructura clara, su neutralidad y su cobertura temática. Si tu empresa, tu fundador o tu metodología tiene relevancia suficiente para tener una entrada en Wikipedia o Wikidata, esa presencia tiene un peso desproporcionado en la memoria del modelo. No es fácil de conseguir porque Wikipedia tiene criterios de notoriedad exigentes, pero si la ocasión se da, es la mención de mayor impacto.

Foros y comunidades de tu sector

Reddit, foros especializados, comunidades de práctica y grupos de discusión forman parte de los datos de entrenamiento de los principales modelos. Una discusión en la que tu empresa o metodología aparece como referencia válida dentro de una comunidad experta tiene valor real. No porque te enlacen necesariamente, sino porque el contexto en el que te nombran es de pares que se reconocen entre sí.

Podcasts y contenido en vídeo con transcripción

Perplexity sincroniza YouTube directamente. Otros modelos usan transcripciones de podcasts y entrevistas en vídeo como parte de sus datos. Aparecer como invitado en podcasts de tu sector donde hablas de tu metodología con profundidad es una de las formas más subestimadas de construir presencia en los LLMs. El audio no importa directamente: importa el texto que lo acompaña o lo transcribe.

Directorios y agregadores de autoridad

Clutch, G2, Trustpilot o los directorios específicos de tu sector son fuentes que los modelos conocen y usan cuando alguien pregunta por recomendaciones de empresas. Tener un perfil completo, con reseñas reales y descripción precisa de tus servicios en estos directorios, es una capa de presencia externa que vale la pena mantener actualizada.

Cómo construir presencia externa con criterio

La estrategia de Digital PR para GEO no es distinta en su esencia de la que ya funciona para SEO: conseguir que fuentes relevantes te mencionen de forma consistente y en el contexto correcto. Lo que cambia es el énfasis. Para SEO el objetivo principal era el enlace. Para GEO el objetivo principal es el contexto: qué dice esa fuente sobre ti, cómo te describe, con qué términos te asocia.

Hay tres vías que funcionan especialmente bien en la práctica. La primera es el guest posting estratégico: escribir artículos de opinión o análisis en publicaciones de tu sector donde demuestres profundidad real en el tema por

el que quieres ser reconocido. No content de relleno. Contenido que la comunidad vaya a guardar y citar. La segunda es la aparición en roundups y comparativas: cuando alguien hace una lista de “mejores agencias de SEO en España” o “herramientas de GEO”, estar en esa lista con una descripción precisa de lo que haces es exactamente el tipo de mención contextualizada que los modelos usan. La tercera es el PR de datos: publicar estudios, encuestas o análisis propios con datos reales que otros medios quieran citar. Un estudio propio que se cita en cinco sitios distintos genera la misma red de menciones consistentes que antes tardarías meses en construir con link building clásico.

► **La pregunta que orienta la estrategia**

Cuando alguien le pregunta a ChatGPT por las mejores empresas de tu tipo en tu zona, ¿quién aparece y por qué? Busca esa respuesta ahora mismo. Las empresas que aparecen tienen algo en común: presencia externa consistente en las fuentes que el modelo conoce. Eso es exactamente lo que tienes que construir.

Los ciclos de entrenamiento y por qué esto no es instantáneo

Hay que ser honesto sobre los tiempos. Si publicas un artículo hoy en un medio de tu sector, esa mención no va a aparecer mañana en las respuestas de ChatGPT. Los modelos que responden desde la memoria tienen ciclos de reentrenamiento que pueden durar meses. Eso significa que la estrategia de Digital PR para GEO es una inversión a medio plazo, no una palanca de impacto inmediato.

El impacto inmediato viene del modo RAG: si la fuente que te menciona está bien indexada y el modelo la consulta en tiempo real cuando alguien hace una búsqueda relacionada, esa mención puede impactar antes. Pero no puedes contar con eso como única vía. La estrategia correcta trabaja las dos capas: construye presencia externa de calidad de forma consistente para los ciclos de entrenamiento futuros, y asegúrate de que tu contenido propio está bien optimizado para el modo RAG que opera en tiempo real.

La analogía que uso con los clientes es la de la reputación profesional. No se construye de una semana para otra. Se construye apareciendo en los sitios correctos, diciendo cosas que merecen ser recordadas, de forma consistente en el tiempo. El modelo aprende igual que aprende una persona que lee sobre tu sector: por exposición repetida en fuentes que considera fiables.

“La fabada de Casa Fermín en Oviedo no se hizo famosa porque pusieran un anuncio un día. Se hizo famosa porque durante décadas, cada persona que la probó se lo contó a alguien. El Digital PR es exactamente eso: conseguir que las fuentes correctas cuenten tu historia antes de que alguien pregunte.”

— Marcelo

Consistencia de marca en todas las fuentes

Hay un error que subestimamos hasta que lo vemos en las respuestas de la IA: la inconsistencia en cómo nos describimos en distintas fuentes. En nuestra web decimos una cosa. En un directorio ponemos otra descripción. En LinkedIn el perfil dice algo diferente. En el podcast usamos otro término para lo mismo.

Para un humano eso es tolerable: entiende el contexto y reconcilia las diferencias. Para un modelo que aprende patrones de texto, las señales contradictorias diluyen la representación. Si en diez fuentes distintas te describen de diez formas distintas, el modelo no tiene un perfil claro de quién eres. Y cuando no tiene un perfil claro, tiende a omitirte o a asociarte con términos genéricos que no te diferencian de nadie.

La solución es tan poco glamurosa como efectiva: definir con precisión cómo te describes —qué haces, para quién, con qué resultado, en qué geografía— y usar esa descripción de forma consistente en todas las fuentes donde apareces. Tu bio de LinkedIn, tu perfil en directorios, tu presentación en podcasts, tu página ‘Sobre nosotros’, el pie de tus artículos de invitado. Mismo mensaje, adaptado al formato. No el mismo texto copiado: la misma esencia.

✨ MITOLOGÍA ASTURIANA

La Güestia y el rastro que dejas cuando no estás

La Güestia es una de las leyendas más inquietantes de Asturias. Es una procesión de almas que recorre los caminos rurales de noche, normalmente precedida por el sonido de una campana. Los aldeanos sabían que no debían cruzarse con ella, pero también sabían algo curioso: la Güestia solo recorría los caminos que alguien había transitado en vida. Los senderos que nadie había pisado no existían para ella.

El Digital PR funciona de forma parecida. Los modelos solo recorren los caminos que otros han transitado antes. Si no hay rastro tuyo en las fuentes que el modelo conoce, si nadie ha hablado de ti en los sitios donde el modelo aprendió, no hay camino que recorrer. Cuando alguien hace la pregunta, el modelo busca en su memoria y encuentra... silencio.

Cada mención en un medio relevante, cada aparición en un podcast de tu sector, cada artículo firmado en una publicación con autoridad es una huella en el camino. Una señal de que has estado ahí, de que otros te reconocen, de que cuando alguien pregunte por lo que haces, hay un rastro que seguir.

No esperes a que pregunten para empezar a dejar huella. Para cuando llegue la pregunta, el camino tiene que estar ya transitado.

道改 **DOKAI SEO** · dokaiseo.com · Oviedo, Asturias, 2026

06

Medir la visibilidad en IA: *lo que los números sí te dicen.*

Medir la visibilidad en IA es más complicado de lo que parece y más útil de lo que muchos creen. Este capítulo explica qué se puede medir con rigor, qué herramientas existen, qué limitaciones tienen y cómo construir un sistema de seguimiento que te dé información real para tomar decisiones.

El problema de fondo: los LLMs no son deterministas

Antes de hablar de cómo medir, hay que entender por qué medir visibilidad en IA es técnicamente difícil. Y la razón es ésta: los modelos de lenguaje no dan la misma respuesta dos veces. Por diseño, hay aleatoriedad en cómo generan texto. Eso significa que si lanzas la misma pregunta diez veces, puedes obtener diez respuestas distintas. Unas veces te mencionan, otras no. Unas veces en primer lugar, otras en tercero.

Esto tiene consecuencias directas para cualquier herramienta que prometa decirte ‘tu posición en ChatGPT’. Un estudio de SparkToro de enero de 2026

analizó casi 3.000 prompts ejecutados por 600 voluntarios en ChatGPT, Claude y Google AI. El resultado: menos de 1 entre 100 probabilidades de obtener la misma lista de marcas dos veces. Menos de 1 entre 1.000 de obtenerla en el mismo orden. Otra investigación de GPT-4o encontró que con temperatura cero —la configuración supuestamente más determinista— el modelo solo produjo respuestas idénticas el 3% de las veces.

Añade a eso la volatilidad de las fuentes: un estudio de Profound sobre 80.000 prompts comparó las respuestas de junio y julio de 2025 y encontró que entre el 40 y el 59% de los dominios citados habían cambiado en ese mes. En periodos de seis meses, la volatilidad sube al 70-90%.

“Cuando un cliente nos pide ‘su posición en ChatGPT’, la respuesta honesta es que eso no existe como concepto fijo. Lo que sí existe es una tendencia: ¿apareces con frecuencia o casi nunca? ¿El contexto en el que te mencionan es el que quieres? Eso sí se puede medir con sentido.”

— Marcelo

Qué se puede medir con rigor

Dicho todo lo anterior, hay cosas que sí se pueden medir de forma útil si se hace con el método correcto. La clave está en ejecutar volumen suficiente de prompts y en medir tendencias, no snapshots puntuales.

Frecuencia de mención

Cuántas veces apareces en un conjunto de N prompts relacionados con tu sector. Si lanzas 50 preguntas distintas sobre tu industria y apareces en 30 de las respuestas, tienes una frecuencia del 60%. Eso sí es comparable en el tiempo: si el mes siguiente apareces en 40, estás creciendo. Si apareces en 20, algo ha cambiado. El número absoluto no importa tanto como la tendencia.

Contexto de la mención

No solo si apareces, sino cómo te describen. ¿Te asocian con los términos correctos? ¿El contexto en el que apareces es el que quieres? ¿Te mencionan como referente, como opción entre muchas, o de pasada? Esta dimensión cualitativa es igual de importante que la frecuencia y requiere lectura humana de las respuestas, no solo conteo automático.

Comparativa con competidores

Los mismos prompts que lanzas para ti, los lanzas para tus tres o cuatro competidores principales. ¿Con qué frecuencia aparecen ellos? ¿Cómo los describe el modelo en comparación? Esta comparativa te dice dónde estás en el ecosistema de tu sector desde el punto de vista del modelo, que es lo que importa cuando alguien hace la pregunta.

Precisión de la información

Cuando el modelo habla de tu empresa, ¿lo que dice es correcto? ¿Hay errores factuales? ¿Te atribuye servicios que no ofreces o te falta alguno importante? Detectar estos errores y corregirlos —reforzando la información

correcta en tu web y en fuentes externas— es una de las acciones más directamente accionables de toda la estrategia GEO.

► El mínimo de prompts para tener datos útiles

Investigadores de Anthropic recomiendan ejecutar entre 3 y 5 repeticiones por prompt para mejorar la consistencia estadística. Para un seguimiento mensual útil, trabaja con al menos 30-50 prompts distintos cubriendo las consultas más relevantes de tu sector. Por debajo de ese volumen, los datos tienen demasiado ruido para sacar conclusiones fiables.



DATO CURIOSO DE ASTURIAS

Asturias tiene más sidrerías por habitante que ningún otro lugar del mundo

Asturias produce alrededor de 80 millones de litros de sidra al año, lo que la convierte en la región con mayor concentración de sidrerías por habitante del planeta. Solo en Villaviciosa, conocida como la capital de la manzana asturiana, hay más lagares activos que en algunos países enteros.

Lo interesante desde el punto de vista del GEO: si le preguntas a ChatGPT por las mejores sidrerías de Asturias, el modelo conoce las más famosas porque aparecen en guías, blogs de viajes y artículos de prensa. Las demás, aunque sean igual de buenas, son invisibles. Exactamente lo mismo pasa con tu empresa en tu sector.

La calidad no basta. La visibilidad en las fuentes correctas es lo que decide quién aparece en la respuesta.

Herramientas disponibles y sus limitaciones

El mercado de herramientas para medir visibilidad en IA creció muy rápido en 2025. Hay opciones que van desde soluciones enterprise hasta scripts que puedes montar tú mismo. Ninguna es perfecta. Todas tienen limitaciones que conviene entender antes de interpretarlas.

Prompt trackers especializados

Herramientas como Profound, Brandwatch AI, Mention o Semrush AI Toolkit ofrecen dashboards que ejecutan prompts periódicamente y muestran evolución de menciones. Son útiles para tener una visión tendencial sin tener que hacer el trabajo manual. Su limitación principal es que ninguna publica con transparencia cuántos prompts ejecutan por ciclo ni cómo gestionan la variabilidad intrínseca del modelo. Toma los números como referencia de tendencia, no como datos absolutos.

Seguimiento manual con plantilla propia

La opción más fiable y más económica al mismo tiempo. Diseñas un banco de 30 a 50 prompts relevantes para tu sector, los ejecutas cada mes en los modelos que te interesan —ChatGPT, Perplexity, Gemini— y registras los resultados en una hoja de cálculo. Frecuencia de mención, contexto, competidores citados, precisión. En tres meses tienes una línea de tendencia real. En seis meses, datos suficientes para tomar decisiones.

Google Search Console para AI Overviews

Search Console ya muestra datos de impresiones y clics procedentes de las AI Overviews de Google en algunos mercados. Es la fuente más directa y fiable para medir el impacto del canal generativo de Google porque viene de datos reales de usuarios, no de simulaciones. Si aún no tienes activado el filtro de AI Overviews en Search Console, es el primer sitio donde mirar.

Analytics para tráfico referral de IA

ChatGPT, Perplexity y otros modelos ya aparecen como fuentes de tráfico en Google Analytics 4 y en otras herramientas de analítica. Revisar mensualmente el tráfico referral proveniente de estos dominios te da el dato más tangible de todos: visitas reales que llegaron desde la IA. Combina esto con las tasas de conversión de ese tráfico y tienes el argumento de negocio más sólido que puedes presentar a un cliente.

Los KPIs que tienen sentido

No todos los números que puedes medir merecen un informe. Estos son los que en nuestra experiencia tienen más utilidad práctica para guiar decisiones de estrategia GEO.

KPI	Qué mide	Por qué importa
Frecuencia de mención (%)	Veces que apareces sobre total de prompts	Tendencia de visibilidad en el tiempo

Share of voice vs. competidores	Tu frecuencia frente a la de tu competencia	Tu posición relativa en el sector
Calidad del contexto	Cómo te describe el modelo	Si el mensaje es el que quieres transmitir
Precisión factual	Si los datos que da el modelo son correctos	Errores que hay que corregir en fuentes
Tráfico referral de IA	Visitas reales desde ChatGPT, Perplexity...	Impacto real en negocio, no en simulaciones
Conversión del tráfico IA	Cuánto convierte ese tráfico vs. otros canales	El argumento de ROI más sólido que existe

Cómo presentarlo a un cliente

Uno de los retos reales del GEO es que los clientes están acostumbrados a los informes de SEO: posición número uno en esta keyword, 3.000 visitas orgánicas este mes, CTR del 4,2%. Los datos de visibilidad en IA son más difusos por naturaleza y eso hace que sea más difícil comunicar el valor del trabajo.

La forma que mejor funciona en nuestra experiencia es anclar la conversación en tráfico real y conversión real. Si puedes mostrar que el tráfico proveniente de ChatGPT tiene una tasa de conversión tres o cuatro veces superior al tráfico orgánico clásico —que es lo que los datos del sector muestran consistentemente— la justificación del trabajo se hace sola. No necesitas explicar cómo funciona un LLM. Solo necesitas mostrar que ese canal genera leads de mayor calidad.

Para la parte de frecuencia de mención y contexto, el formato que mejor funciona es la comparativa temporal: así te describía ChatGPT hace tres

meses, así te describe ahora. El cambio es visible y atribuible al trabajo realizado. Eso conecta mejor con un cliente que una tabla de porcentajes abstractos.

“El mejor informe de GEO que he presentado era una sola diapositiva: la respuesta de ChatGPT sobre nuestra industria en enero y la de abril. En enero el cliente no aparecía. En abril estaba en el segundo párrafo. No necesité explicar nada más.”

— Marcelo

✨ MITOLOGÍA ASTURIANA

El Hombre del Saco y las cosas que no quieres que te cuenten

En Asturias, como en muchos lugares de España, el Hombre del Saco era la figura que los mayores usaban para advertir a los niños: ‘si te portas mal, viene el Hombre del Saco y se te lleva’. La amenaza no era nunca específica. No sabías exactamente cómo era ni qué hacía. Solo sabías que no querías averiguarlo.

Hay una versión corporativa del Hombre del Saco en GEO: los datos que no quieres ver. El informe que muestra que tu competidor aparece en el 70% de las respuestas y tú en el 8%. Que el modelo te describe como ‘una empresa generalista de marketing digital’ cuando llevas años intentando posicionarte

como especialista. Que hay tres errores factuales en lo que la IA dice de ti y uno de ellos es el precio de tu servicio principal.

La tentación es no mirar. Hacer la estrategia sin medir, publicar contenido sin auditar cómo te describe la IA, seguir como si el canal no existiera. El problema es que el cliente que pregunta a ChatGPT por tu sector sí mira. Y lo que ve no es lo que tú elegirías mostrarle.

Los datos de visibilidad en IA no son cómodos siempre. Pero son los que te dicen exactamente dónde actuar.

07

GEO local

El GEO no es solo para grandes marcas o negocios nacionales. El negocio local tiene una oportunidad real y concreta en este canal, y las reglas son distintas a las del GEO genérico. Aquí el que controla los datos gana. Y los datos locales tienen dueño.

Cómo responde ChatGPT a las consultas locales

Cuando alguien le pregunta a ChatGPT ‘mejores fontaneros en Oviedo’ o ‘agencias de marketing en Gijón’, el modelo no lo sabe todo de memoria. En la mayoría de casos activa la búsqueda RAG porque son consultas con QDF alto —requieren información local, específica y actualizable— y consulta fuentes externas para construir la respuesta.

Lo que hace ChatGPT en ese momento es muy concreto: lanza una búsqueda a través de su herramienta de búsqueda web, recibe un JSON estructurado con resultados que incluye un bloque específico de negocios locales llamado `businesses_map`, y usa esa información para construir la

respuesta. Ese bloque de negocios locales viene principalmente de Foursquare.

En febrero de 2025, OpenAI anunció una asociación con Foursquare para alimentar las búsquedas locales en ChatGPT a través de su Places API. Foursquare tiene más de 100 millones de puntos de interés en más de 200 países. Un análisis de 50 prompts de búsqueda local en varias ciudades españolas encontró que entre el 60 y el 70% de los negocios que ChatGPT mostraba en primer lugar provenían de Foursquare. Eso no es un detalle técnico menor. Es la palanca más directa del GEO local.

“Durante meses estuvimos hablando de GEO local como si fuera lo mismo que el GEO genérico pero con ciudad añadida. Cuando vi el análisis de Foursquare me di cuenta de que no: el GEO local tiene su propio ecosistema de fuentes, y quien controla esas fuentes controla quién aparece.”

— Marcelo

Foursquare: la fuente que pocos trabajan

Foursquare no es la red social de check-ins que muchos recuerdan de 2012. Hoy es una empresa de datos de localización B2B cuya Places API alimenta a decenas de aplicaciones, plataformas y ahora a ChatGPT. Tener un perfil completo y actualizado en Foursquare es la acción de GEO local con mayor

impacto directo que existe ahora mismo para negocios con presencia física o geográfica.

La cobertura de Foursquare no es homogénea. En ciudades grandes como Madrid, Barcelona o Sevilla hay miles de perfiles bien documentados. En ciudades medianas como Oviedo, Gijón o Santander la densidad es menor y la competencia por aparecer también. En municipios pequeños, muchos negocios directamente no existen en la base de datos. Eso significa que para un negocio en una ciudad media, el coste de entrada es bajo y la recompensa potencial alta: si eres de los pocos bien documentados en Foursquare para tu categoría en tu ciudad, eres de los pocos que ChatGPT conoce.

► **Cómo verificar si estás en Foursquare**

Entra en foursquare.com, busca tu negocio por nombre y ciudad. Si apareces, revisa que el nombre, dirección, categoría, teléfono y descripción sean correctos y completos. Si no apareces, crea el perfil. Es gratuito y es actualmente una de las acciones de GEO local con mayor retorno por tiempo invertido.

NAP consistency: el principio que no ha cambiado

NAP son las siglas de Name, Address, Phone: nombre, dirección y teléfono. La consistencia de estos tres datos en todas las fuentes donde aparece tu negocio es uno de los principios más antiguos del SEO local y sigue siendo igual de relevante en el contexto GEO.

El motivo es el mismo que en el SEO local clásico: los modelos aprenden de múltiples fuentes. Si tu negocio aparece como 'Carpintería Fernández' en Google Business Profile, como 'Carpinteria Fernandez SL' en Foursquare,

como ‘Talleres Fernández’ en un directorio local y con una dirección diferente en otro, el modelo tiene señales contradictorias sobre quién eres y dónde estás. Esa contradicción reduce la confianza del sistema en la información y puede hacer que el modelo te omita en favor de competidores con datos más coherentes.

La solución es auditórea: revisar todas las fuentes donde apareces —Google Business Profile, Foursquare, Yelp, TripAdvisor si aplica, directorios sectoriales, tu propia web, LinkedIn de empresa— y asegurarte de que el NAP es idéntico en todas. Mismo nombre exacto, misma dirección exacta, mismo formato de teléfono. No es trabajo glamuroso. Es trabajo que funciona.

Google Business Profile y los LLMs

Google Business Profile sigue siendo la fuente de datos locales más importante que existe, y su relación con los LLMs es doble. Por un lado, los datos de GBP alimentan directamente las AI Overviews de Google cuando alguien hace una búsqueda local. Por otro, Google indexa el contenido de los perfiles de GBP y ese contenido puede ser recuperado por sistemas RAG de otros modelos.

Un perfil de GBP bien trabajado para GEO local incluye estos elementos: categoría principal y categorías secundarias precisas —no genéricas—, descripción del negocio clara y con los términos por los que quieres que te encuentren, horarios actualizados, fotos recientes y representativas, respuestas a todas las reseñas, y publicaciones periódicas que mantengan el perfil activo. Las reseñas tienen un peso especial: son el único contenido generado por terceros dentro de tu perfil, y los modelos las usan para calibrar la reputación de un negocio.

► Lo que más impacta en GBP para GEO

La descripción del negocio bien escrita con términos específicos, las reseñas recientes con respuestas del propietario, y la coherencia entre lo que dice GBP y lo que dice tu web. El volumen de reseñas importa, pero la calidad y la recencia importan más.

Directorios locales y fuentes sectoriales

Más allá de Foursquare y Google Business Profile, hay un ecosistema de directorios y fuentes locales que los modelos conocen y usan. La prioridad depende del sector y la geografía, pero hay algunos que tienen presencia consistente en los datos de entrenamiento de los principales modelos.

- **Yelp:** especialmente relevante para restauración, servicios al consumidor y ocio
- **TripAdvisor:** turismo, restauración y alojamiento
- **Trustpilot y Google Reviews:** cualquier negocio con componente de servicio al cliente
- **Directorios sectoriales específicos:** colegios profesionales, asociaciones de sector, cámaras de comercio
- **Prensa local digital:** menciones en medios locales con trayectoria y presencia online consolidada

La estrategia no es estar en todos. Es estar bien en los que tu cliente potencial usa y que los modelos conocen. Para un restaurante en Oviedo, TripAdvisor y Google Business Profile son prioritarios. Para una clínica dental, los directorios de salud y el portal del Colegio de Odontólogos. Para una

agencia de SEO, las plataformas de reviews de servicios B2B como Clutch o G2, y los medios del sector.

El contenido local que los modelos usan

El GEO local no es solo datos estructurados y directorios. El contenido de tu web con referencia local explícita también importa, especialmente para los sistemas RAG que consultan tu sitio en tiempo real.

Una página de servicio que menciona explícitamente la ciudad o la región donde operas, que explica por qué conoces el mercado local, que incluye referencias a clientes o proyectos en esa zona —sin datos confidenciales, con el nivel de detalle que el cliente permita— es un documento que el modelo puede usar cuando alguien pregunta por servicios de tu tipo en tu zona. No es suficiente con tener la ciudad en el título de la página. El contenido tiene que demostrar presencia y conocimiento local de verdad.

Esto conecta directamente con algo que Marcelo aplica desde el primer día en DOKAI: la agencia está en Oviedo, trabaja con empresas en Asturias y en toda España, y eso se dice explícitamente en el contenido. No porque sea un truco de GEO. Sino porque es la realidad, y la realidad bien explicada es exactamente lo que los modelos buscan.

“Cada vez que alguien nos pregunta si tiene sentido poner la ciudad en el contenido, la respuesta es sí, siempre que sea verdad. Si trabajas en Oviedo, dílo. Si tienes clientes en Gijón, menúciónalo. La geografía específica es información útil para el modelo, no keyword stuffing.”

— Marcelo



DATO CURIOSO DE ASTURIAS

Oviedo tiene más esculturas por habitante que casi cualquier ciudad del mundo

Oviedo es conocida entre los urbanistas como 'la ciudad de las esculturas'. Con más de un centenar de obras de arte público repartidas por sus calles, tiene una de las densidades de escultura urbana más altas del mundo. Hay obras de Botero, Miró, Dalalí, Chillida y docenas de artistas más. La Regenta, los personajes de Woody Allen, La Maternidad... cualquier local sabe dónde están.

Pero si le preguntas a ChatGPT por las esculturas de Oviedo, el modelo solo conoce las más famosas: las que tienen página en Wikipedia, las que aparecen en guías de viaje digitales, las que han salido en medios. Las demás, aunque estén ahí en medio de la calle desde hace veinte años, son invisibles para el modelo.

Tu negocio local es igual. Puedes llevar décadas siendo una referencia en tu barrio, en tu ciudad, en tu sector. Si no estás en las fuentes que el modelo conoce —Foursquare, GBP, medios locales, directorios del sector— para el usuario que pregunta a la IA, no existes. La visibilidad digital no es opcional. Es el mapa sin el que nadie te encuentra.



MITOLOGÍA ASTURIANA

El Busgosu y el que cuida el bosque que conoce

El Busgosu es un espíritu del bosque de la mitología asturiana. Se le describe como un ser peludo y salvaje que vive en lo más profundo de los bosques asturianos y que solo conoce el territorio que ha recorrido. Dentro de su bosque, lo sabe todo: dónde están los senderos, qué árboles hay, dónde beben los animales. Fuera de su bosque, está tan perdido como cualquiera.

Los LLMs funcionan de forma similar con el conocimiento local. Conocen bien los territorios que han recorrido: las ciudades con mucha presencia digital, los negocios que aparecen en muchas fuentes, los sectores con mucho contenido publicado. En cambio, en los territorios poco digitalizados, en las ciudades medianas con poca presencia en fuentes de autoridad, en los sectores con poco contenido online, el modelo está tan perdido como el Busgosu fuera de su bosque.

Eso crea una oportunidad. En un mercado local poco digitalizado, ser el primero en trabajar bien tu presencia en las fuentes que los modelos conocen te da una ventaja desproporcionada. No tienes que competir con cien empresas bien posicionadas. Tienes que ser el primero que el modelo aprende a reconocer en ese territorio.

El Busgosu defiende su bosque. Tú defiende el tuyo: tu ciudad, tu sector, tu especialidad. Quien primero marque el territorio en los datos que el modelo ingiere, ese aparece cuando alguien hace la pregunta.

08

De la teoría *a la práctica*

Ocho capítulos después, tienes el marco completo. Ahora toca la parte en la que muchos libros fallan: pasar del conocimiento a la acción. Este último capítulo es una checklist de 40 puntos, un plan de acción por fases y los errores que vas a ver siempre, para que al menos los veas venir.

Antes de la checklist

Una advertencia honesta: esta checklist no es un algoritmo que ejecutas una vez y listo. El GEO es un trabajo continuo porque el ecosistema cambia, los modelos se actualizan y la competencia también se mueve. Lo que sí funciona es tener una base sólida y revisarla periódicamente.

Los 40 puntos están organizados en cinco bloques que siguen la misma lógica que el libro: base técnica, contenido, presencia externa, local si aplica, y medición. Dentro de cada bloque hay puntos de prioridad alta —los que no

puedes saltarte— y puntos de optimización que marcan la diferencia a medio plazo.

► **Cómo usar esta checklist**

Haz una primera pasada marcando qué tienes y qué te falta. Lo que te falta en los puntos de prioridad alta es tu plan de acción inmediato. Lo que te falta en los de optimización es tu hoja de ruta de los próximos tres meses.

BLOQUE 1 · Base técnica

- **[PRIORITARIO]** Indexación revisada en Search Console: sin errores críticos ni páginas importantes excluidas
- **[PRIORITARIO]** robots.txt correcto: no bloquea páginas que deben ser rastreadas
- **[PRIORITARIO]** Contenido principal renderizado en HTML, no dependiente de JavaScript para cargarse
- **[PRIORITARIO]** Core Web Vitals en verde o en proceso de corrección activo
- **[PRIORITARIO]** Arquitectura de URLs limpia: una URL por tema, sin duplicados ni canibalización
- **[PRIORITARIO]** Enlazado interno coherente con anclajes descriptivos
- **[OPTIMIZACIÓN]** Datos estructurados schema.org implementados en páginas clave: Product, Article, FAQPage, LocalBusiness
- **[OPTIMIZACIÓN]** Sitemap XML actualizado y enviado a Search Console
- **[OPTIMIZACIÓN]** Hreflang correcto si hay versiones en múltiples idiomas

BLOQUE 2 · Contenido

- **[PRIORITARIO]** Páginas principales responden la pregunta clave en el primer párrafo
- **[PRIORITARIO]** Jerarquía de cabeceras coherente: H1 define el tema, H2 dividen subtemas, H3 desarrollan
- **[PRIORITARIO]** Página 'Sobre nosotros' responde: qué haces, para quién, dónde, cuánto tiempo llevas, qué resultados
- **[PRIORITARIO]** Páginas de servicio sin ironía ni ambigüedad: lenguaje directo y preciso
- **[PRIORITARIO]** Las afirmaciones relevantes van respaldadas con datos o fuentes citadas
- **[OPTIMIZACIÓN]** Página de preguntas frecuentes con respuestas directas y completas
- **[OPTIMIZACIÓN]** Glosario o definiciones de los términos clave de tu sector
- **[OPTIMIZACIÓN]** Al menos una comparativa honesta de opciones, herramientas o enfoques de tu industria
- **[OPTIMIZACIÓN]** Al menos un caso de estudio con datos reales y metodología descrita
- **[OPTIMIZACIÓN]** Párrafos que se entienden de forma aislada, sin necesitar contexto previo
- **[OPTIMIZACIÓN]** Tablas para información comparativa o estructurada en lugar de texto continuo

BLOQUE 3 · Presencia externa y Digital PR

- **[PRIORITARIO]** Descripción de empresa consistente en todas las fuentes externas: misma esencia, mismo posicionamiento
- **[PRIORITARIO]** Perfil completo y actualizado en LinkedIn de empresa
- **[PRIORITARIO]** Al menos una aparición como autor o entrevistado en un medio relevante de tu sector
- **[PRIORITARIO]** Perfil en al menos un directorio de autoridad de tu sector (Clutch, G2, o equiv.)
- **[OPTIMIZACIÓN]** Presencia activa en foros o comunidades donde tu cliente potencial hace preguntas
- **[OPTIMIZACIÓN]** Aparición en al menos un roundup o comparativa del tipo 'mejores empresas de X en Y'
- **[OPTIMIZACIÓN]** Aparición como invitado en podcast del sector con transcripción disponible
- **[OPTIMIZACIÓN]** Al menos un estudio o dato propio publicado que otras fuentes puedan citar
- **[OPTIMIZACIÓN]** Menciones en medios locales o regionales si el negocio tiene componente local

BLOQUE 4 · GEO local (si aplica)

- **[PRIORITARIO]** Perfil en Foursquare creado o reclamado, con nombre, dirección, categoría y descripción completos
 - **[PRIORITARIO]** Google Business Profile completo: categorías precisas, descripción con términos clave, horarios actualizados
 - **[PRIORITARIO]** NAP idéntico en todas las fuentes: GBP, Foursquare, web, directorios
-

- **[PRIORITARIO]** Reseñas en GBP con respuestas del propietario a todas ellas
- **[OPTIMIZACIÓN]** Contenido de web menciona explícitamente la ciudad o región donde opera el negocio
- **[OPTIMIZACIÓN]** Perfil actualizado en TripAdvisor, Yelp u otros directorios según sector
- **[OPTIMIZACIÓN]** Schema LocalBusiness implementado en la página de contacto o inicio

BLOQUE 5 · Medición

- **[PRIORITARIO]** Banco de 30-50 prompts diseñados y documentados para seguimiento mensual
- **[PRIORITARIO]** Primera ejecución completada: línea base de frecuencia de mención y contexto
- **[PRIORITARIO]** Tráfico referral de IA visible y separado en Analytics (ChatGPT, Perplexity, Gemini)
- **[PRIORITARIO]** Errores factuales detectados en cómo describe la IA tu empresa y plan de corrección activo
- **[OPTIMIZACIÓN]** Seguimiento mensual sistematizado con plantilla propia o herramienta especializada
- **[OPTIMIZACIÓN]** Comparativa de share of voice vs. tres competidores principales cada trimestre
- **[OPTIMIZACIÓN]** Conversión del tráfico IA medida y comparada con otros canales en el informe mensual

Plan de acción por fases

La checklist dice qué hacer. El plan de fases dice en qué orden. Intentar hacer todo a la vez es la forma más rápida de no hacer nada bien. Estas tres fases funcionan como punto de partida —ajústalas a tu situación concreta.

Fase	Qué hacer	Resultado esperado
Mes 1 Base	Auditoría técnica completa. Corregir errores críticos de indexación. Primera pasada de contenido en páginas clave. Configurar Foursquare y GBP si aplica. Lanzar banco de prompts inicial.	Base sólida lista. Línea base de visibilidad en IA documentada. Sin cuellos de botella técnicos que bloqueen el trabajo posterior.
Mes 2-3 Contenido	Reescritura estratégica de páginas principales: Sobre nosotros, servicios, FAQs. Crear o mejorar al menos una comparativa y un caso de estudio. Auditoría de consistencia de mensaje en fuentes externas.	Contenido que el LLM puede usar y citar. Coherencia de perfil entre web y fuentes externas. Primeras mejoras de frecuencia de mención visibles.
Mes 4-6 Autoridad	Ejecutar plan de Digital PR: guest posts, apariciones en podcasts, presencia en roundups. Publicar estudio o dato propio. Revisitar medición mensual y ajustar según tendencias detectadas.	Menciones externas crecientes en fuentes de autoridad. Mejora de share of voice frente a competidores. Tráfico referral de IA medible y con tasa de conversión documentada.

Los errores que verás siempre

Después de trabajar GEO con clientes de distintos sectores y tamaños, hay un patrón de errores que se repite con una consistencia casi cómica. No porque la gente sea descuidada, sino porque son los errores que parecen razonables hasta que ves por qué no funcionan.

Empezar por la medición antes que por la base

El cliente quiere ver su 'posición en ChatGPT' antes de haber arreglado los errores de indexación. Es comprensible: la medición parece lo nuevo y lo interesante. El problema es que medir sin base sólida solo genera datos que no puedes mejorar porque el canal está roto. La base primero, siempre.

Confundir frecuencia de mención con posición garantizada

Aparecer en el 60% de los prompts de tu sector no significa que aparezcas en el 60% de las consultas reales de usuarios. Los prompts del banco de seguimiento son una muestra controlada. La realidad es más variable. Los datos de frecuencia son útiles para medir tendencias, no para prometer resultados específicos a clientes.

Crear contenido nuevo en lugar de mejorar el existente

Cuando la auditoría de contenido revela páginas que el modelo no usa bien, la tentación es crear páginas nuevas. Casi siempre es mejor mejorar las que ya existen: tienen historial, tienen enlaces, tienen autoridad acumulada.

Reescribir bien una página existente tiene más impacto que crear cinco páginas nuevas desde cero.

Digital PR sin coherencia de mensaje

Conseguir menciones externas con descripciones contradictorias de lo que haces es peor que no tener menciones. Si cada fuente te describe de una forma distinta, el modelo no construye un perfil claro. Antes de cualquier campaña de PR, define el mensaje y úsalo consistentemente.

Esperar resultados en semanas

El GEO tiene dos velocidades: RAG, donde los cambios en contenido indexado pueden impactar relativamente rápido, y memoria del modelo, donde los ciclos de reentrenamiento pueden tardar meses. Un cliente que espera ver su nombre en ChatGPT dos semanas después de empezar el trabajo va a llevarse una decepción. Gestionar expectativas de tiempos es parte del trabajo desde la primera reunión.



DATO CURIOSO DE ASTURIAS

La sidra asturiana no se sirve, se escancia

En Asturias, la sidra natural no se vierte en el vaso desde arriba como cualquier bebida. Se escancia: el escanciador levanta la botella por encima de la cabeza y deja caer el chorro desde la mayor altura posible hasta el culín que sostiene casi

a la altura del suelo. El objetivo es oxigenar la sidra en ese instante para que el gas natural se libere y el sabor se abra.

Si no lo haces bien, si sirves la sidra de cualquier manera, el resultado es plano. La misma sidra, el mismo vaso, pero sin el ritual correcto pierde lo mejor de sí misma.

El GEO funciona igual. Puedes tener un negocio excelente, un servicio real, un valor diferencial claro. Pero si no lo presentas de la forma correcta —con estructura, con datos, con consistencia, con presencia en las fuentes adecuadas— el modelo no lo oxigena. No lo ve. No lo cita. El contenido llega plano.

Escancia bien. El ritual importa tanto como la sidra.

✨ MITOLOGÍA ASTURIANA

El Pelayo y la batalla que empieza por conocer el terreno

En el año 722, Don Pelayo lideró la batalla de Covadonga en las montañas de Asturias, considerada el punto de partida de la Reconquista. Con un ejército pequeño frente a fuerzas mucho mayores, Pelayo no ganó por tener más recursos. Ganó porque conocía el terreno mejor que nadie. Sabía dónde estaban los desfiladeros, cómo usar la montaña a su favor, cómo podía convertir una desventaja en una ventaja.

El GEO, en su esencia, es lo mismo. No se trata de tener el presupuesto más grande ni el equipo más numeroso. Se trata de conocer el terreno: cómo piensan los modelos, dónde aprenden, qué fuentes respetan, cómo deciden qué citar. Ese conocimiento del terreno es exactamente lo que este libro intenta darte.

Los que llegan al GEO con el volumen y el presupuesto pero sin entender cómo funciona el canal cometen los mismos errores una y otra vez. Los que entienden la lógica del terreno pueden hacer mucho con poco si priorizan bien.

Ahora conoces el terreno. La batalla es tuya.

Para terminar

Escribimos este libro porque queríamos que existiera algo honesto sobre GEO en castellano. Sin promesas de posicionamiento instantáneo. Sin siglas inventadas para justificar precios. Sin trucos que funcionan dos semanas y luego no.

Lo que hemos intentado transmitir es una sola cosa: el GEO es una evolución lógica del SEO aplicada a un canal nuevo. Quien tiene la base bien construida, quien entiende cómo piensan los modelos y quien trabaja la presencia externa con criterio tiene todo lo necesario para aparecer donde sus clientes toman decisiones hoy.

El canal va a seguir cambiando. Los modelos van a mejorar, las herramientas van a evolucionar, aparecerán nuevas siglas. Lo que no va a cambiar es la lógica de fondo: quien tiene contenido claro, relevante y respaldado por fuentes de autoridad va a seguir ganando visibilidad. Lleva siendo verdad desde Hummingbird en 2013. Sigue siendo verdad hoy.

“Si después de leer esto abres Search Console, miras cómo te describe ChatGPT y empiezas a mejorar una

página, el libro ha cumplido su función. El conocimiento que no se aplica no vale nada. El que sí se aplica multiplica la facturación orgánica. Eso es exactamente lo que hacemos en DOKAI.”

— Marcelo Rojas Guerra

Gracias por leerlo.

Marcelo Rojas Guerra · Marc Mayol Orell

道改 **DOKAI SEO** · dokaiseo.com · +34 684 622 306 · Oviedo, Asturias, 2026

道改 **DOKAI SEO** · dokaiseo.com · Oviedo, Asturias, 2026



Glosario

Todos los términos que aparecen en el libro, explicados sin rodeos. Si en algún momento del libro usamos una sigla y no quedó del todo claro, aquí está la respuesta.

Los números de capítulo indican dónde se usa el término con más detalle.



AEO — Answer Engine Optimization → Cap. 01

Variante del término GEO usada por algunos autores para referirse a la optimización de contenido orientada a motores que responden preguntas directamente, como los LLMs. En esencia es lo mismo que GEO con distinto nombre.

AI Overviews → Cap. 02 · 03

Bloque de respuesta generativa que Google muestra en la parte superior de algunos resultados de búsqueda, generado por Gemini a partir de contenido

indexado. Funciona con una arquitectura RAG: recupera fragmentos de páginas web y los usa para construir una respuesta sintética.

Autoridad semántica → *Cap. 04 · 05*

Grado en que un dominio o una página es reconocido como referente en un tema o conjunto de temas relacionados. Se construye con contenido en profundidad, coherencia temática y menciones externas de fuentes relacionadas con ese tema.

B

BERT → *Cap. 03*

Modelo de lenguaje de Google presentado en 2018 que introdujo la comprensión bidireccional del contexto en las búsquedas. Permitió a Google captar matices y relaciones entre palabras que antes pasaban desapercibidos. Precursor directo de los LLMs modernos.

BM25 → *Cap. 02*

Algoritmo clásico de recuperación de información, evolución de TF-IDF, que puntúa la relevancia de un documento para una consulta basándose en frecuencia de términos y longitud del documento. Sigue siendo parte del pipeline de recuperación de muchos sistemas modernos, incluyendo los que alimentan a los LLMs.

businesses_map → *Cap. 07*

Bloque específico dentro del JSON de respuesta que ChatGPT genera al hacer una búsqueda local. Contiene los negocios locales que el sistema recupera, principalmente a través de la Places API de Foursquare.

C

Chunking → *Cap. 02*

Proceso por el que un sistema divide un texto en fragmentos más pequeños para procesarlos o indexarlos. Los sistemas RAG chunkean los documentos antes de almacenarlos en la base vectorial para poder recuperar fragmentos relevantes en lugar de documentos enteros. No existe un estándar universal de tamaño de chunk.

Corpus de entrenamiento → *Cap. 02 · 05*

Conjunto masivo de texto con el que se entrena un LLM. Incluye páginas web, libros, artículos, foros y otras fuentes textuales. La calidad y composición del corpus determina qué conoce el modelo y qué peso asigna a distintas fuentes.

CTR — Click Through Rate → *Cap. 01*

Porcentaje de usuarios que hacen clic en un resultado después de verlo. En el contexto GEO, el tráfico que llega desde LLMs tiene tasas de conversión muy superiores al tráfico orgánico tradicional, aunque el CTR absoluto desde IA sigue siendo bajo en comparación.

Cuélebre → *Cap. 01*

Serpiente alada de la mitología asturiana que guarda tesoros ocultos en cuevas. En el libro se usa como metáfora del negocio con valor real pero sin visibilidad: el tesoro existe, pero si nadie sabe dónde está la cueva, es como si no existiera.

D

Digital PR → *Cap. 05*

Conjunto de acciones orientadas a conseguir menciones, enlaces y referencias en medios, publicaciones y comunidades digitales de autoridad. En el contexto GEO, el objetivo no es solo el enlace sino el contexto: que la descripción que las fuentes hacen de tu empresa sea la que quieres que el modelo aprenda.

Datos estructurados / Schema.org → *Cap. 08*

Marcado en código HTML que describe explícitamente a los motores de búsqueda y a los sistemas de IA el contenido de una página: tipo de entidad, propiedades, relaciones. Facilita que los sistemas interpreten correctamente la información sin tener que inferirla del texto.

E

E-E-A-T → *Cap. 03 · 04*

Experience, Expertise, Authoritativeness, Trustworthiness. Marco de Google para evaluar la calidad del contenido basado en la experiencia demostrada, la experiencia del autor, la autoridad del dominio y la confiabilidad de la información. Los mismos criterios que los LLMs usan implícitamente para ponderar fuentes.

Embeddings → *Cap. 02*

Representaciones matemáticas de texto en forma de vectores numéricos que capturan el significado semántico. Dos textos con significados similares tienen embeddings cercanos en el espacio vectorial. Los sistemas RAG los usan para encontrar fragmentos de contenido semánticamente relevantes para una consulta.

F

Featured snippet → *Cap. 03*

Bloque de respuesta destacada que Google muestra en la posición cero de los resultados de búsqueda, extrayendo un fragmento de contenido de una página web. El contenido que aparece en featured snippets es prácticamente el mismo que aparece en las AI Overviews y en las respuestas RAG de los LLMs.

Foursquare → *Cap. 07*

Empresa de datos de localización B2B cuya Places API alimenta las búsquedas locales de ChatGPT desde febrero de 2025. Tiene más de 100 millones de puntos de interés en más de 200 países. Entre el 60 y el 70% de los negocios que ChatGPT muestra en primer lugar en búsquedas locales provienen de esta fuente.

Frecuencia de mención → *Cap. 06*

Métrica que mide cuántas veces aparece una marca o empresa en un conjunto de prompts ejecutados sobre un LLM. Se expresa como porcentaje: si apareces en 30 de 50 prompts, tu frecuencia es del 60%. Es una métrica de tendencia, no de posición absoluta.



GEO — Generative Engine Optimization → *Cap. 01*

Término acuñado en el paper de Princeton de 2023 (Aggarwal et al.) para describir la optimización de contenido orientada a motores de búsqueda generativos que responden con texto en lugar de listas de enlaces. En la práctica es una especialización del SEO que añade una capa de monitorización y ajuste específico para entornos de IA generativa.

Google Business Profile (GBP) → *Cap. 07*

Ficha de negocio de Google que aparece en Maps y en los resultados de búsqueda local. Alimenta directamente las AI Overviews de Google para consultas locales y su contenido puede ser recuperado por sistemas RAG de otros modelos.

Güestia → *Cap. 05*

Procesión de almas de la mitología asturiana que recorre los caminos rurales de noche. Solo transita por los senderos que alguien ha recorrido en vida. En el libro se usa como metáfora del Digital PR: los modelos solo recorren los caminos donde has dejado rastro.

H

Hummingbird → *Cap. 03*

Actualización del algoritmo de Google en 2013 que introdujo la búsqueda semántica al interpretar el significado de las consultas en lugar de solo las palabras clave individuales. Primer paso hacia los sistemas de IA generativa que conocemos hoy.

I

Information Gain → *Cap. 03*

Concepto de Google que premia el contenido que aporta información nueva y original sobre un tema, en lugar de repetir lo que ya existe. Relacionado con

BERT y con los criterios E-E-A-T. Los LLMs tienden a preferir fuentes con perspectivas únicas sobre fuentes genéricas.



LLM — Large Language Model → *Cap. 02*

Modelo de lenguaje de gran escala entrenado con cantidades masivas de texto para generar, resumir, traducir y razonar sobre lenguaje natural. ChatGPT, Claude, Gemini y Llama son ejemplos. A diferencia de un buscador, un LLM responde desde su entrenamiento o desde fuentes consultadas en tiempo real mediante RAG.

LLMs.txt → *Cap. 08*

Archivo de texto propuesto como estándar para que los sitios web indiquen a los crawlers de LLMs qué contenido pueden usar y cómo. Su impacto real sigue siendo objeto de debate y no hay evidencia sólida de que los principales modelos lo procesen de forma consistente.



MUM — Multitask Unified Model → *Cap. 03*

Modelo de Google presentado en 2021 capaz de procesar información en múltiples idiomas y formatos simultáneamente. Representó un salto

significativo hacia la comprensión multimodal e interlingüística que caracteriza a los LLMs actuales.

N

NAP — Name, Address, Phone → *Cap. 07*

Nombre, dirección y teléfono de un negocio. La consistencia de estos tres datos en todas las fuentes digitales donde aparece el negocio es un principio fundamental del SEO local y del GEO local. Las señales contradictorias de NAP reducen la confianza del modelo en la información y pueden provocar que te omita.

Navboost → *Cap. 03*

Sistema de Google que mide cómo interactúan los usuarios con documentos específicos para consultas específicas. No es un algoritmo de machine learning sino una tabla que almacena datos de comportamiento de usuarios —clics largos, tiempo de permanencia— y los usa como señal de relevancia.

Nuberu → *Cap. 03*

Señor de las nubes y las tormentas en la mitología asturiana. En el libro representa los problemas técnicos de un sitio web: si el camino está lleno de barro —errores de indexación, JavaScript que bloquea contenido, robots.txt mal configurado— el Nuberu pasa por encima y sigue de largo.

PageRank → *Cap. 02 · 03*

Algoritmo de Google que asigna un valor de autoridad a cada página web basándose en la cantidad y calidad de los enlaces que recibe. Sigue siendo una de las señales más importantes del ranking. Perplexity ha confirmado que usa una forma de PageRank para priorizar fuentes.

Passage Ranking → *Cap. 02*

Tecnología de Google implementada en 2020 que permite indexar y recuperar pasajes individuales dentro de una página, en lugar de tratar la página como una unidad indivisible. Elimina la necesidad de crear páginas cortas para facilitar la recuperación: el sistema extrae los fragmentos relevantes por su cuenta.

Prompt tracker → *Cap. 06*

Herramienta que ejecuta prompts predefinidos contra LLMs periódicamente para medir si una marca aparece en las respuestas, con qué frecuencia y en qué contexto. Debido a la naturaleza no determinista de los modelos, sus datos deben interpretarse como tendencias, no como posiciones absolutas.

Q

Q* — Quality Star → *Cap. 03*

Señal interna de Google que mide la calidad percibida de un dominio. Es mayoritariamente estática y se asocia a dominios más que a URLs individuales. Se nutre de PageRank, autoridad del dominio, confiabilidad, popularidad y estructura del contenido. Revelada en el juicio antimonopolio contra Google.

QDF — Query Deserves Freshness → *Cap. 02*

Sistema por el que Google, y ahora también los LLMs, determinan si una consulta requiere información reciente. En GPT-5 se implementó con una escala de 0 a 5: 0 para consultas históricas sin necesidad de actualización, 5 para consultas que requieren información del mes en curso.

R

RAG — Retrieval-Augmented Generation → *Cap. 02*

Arquitectura que combina la recuperación de información en tiempo real con la generación de texto de un LLM. En la fase de inferencia, el sistema consulta fuentes externas —páginas web, bases de datos—, recupera fragmentos relevantes y los integra en la respuesta generada. Es el mecanismo por el que ChatGPT con búsqueda activada o Perplexity funcionan.

RankBrain → *Cap. 03*

Sistema de machine learning de Google introducido en 2015 para interpretar consultas complejas y llevar la búsqueda semántica a otro nivel. Primera aplicación a gran escala de deep learning en el algoritmo de ranking de Google.

S

Schema.org → *Cap. 08*

Vocabulario estándar de datos estructurados creado por Google, Bing, Yahoo y Yandex para que los webmasters describan el contenido de sus páginas de forma que los sistemas puedan interpretarlo sin inferencia. Se implementa en JSON-LD dentro del HTML.

SEO — Search Engine Optimization → *Cap. 01 · 03*

Conjunto de técnicas orientadas a mejorar la visibilidad de un sitio web en los resultados orgánicos de los motores de búsqueda. En el contexto de este libro, es la base imprescindible sin la cual el GEO no puede funcionar.

Share of voice → *Cap. 06*

Métrica que mide la proporción de menciones de una marca frente al total de menciones del sector en un canal determinado. En GEO, el share of voice en LLMs se calcula comparando tu frecuencia de mención con la de tus principales competidores sobre los mismos prompts.

SONIC → *Cap. 02*

Clasificador interno de ChatGPT que evalúa cada consulta del usuario y decide si el modelo debe buscar información en fuentes externas o responder desde la memoria de entrenamiento. Si la probabilidad calculada supera el umbral de 0,54, el modelo activa la búsqueda RAG.

T

T* — Topicality Star → *Cap. 03*

Señal dinámica del ranking de Google que mide la relevancia de un documento frente a una consulta. Se compone del sistema ABC: Anchors (enlaces entrantes), Body (contenido textual) y Clicks (comportamiento de usuario). Revelada en el juicio antimonopolio contra Google.

Temperatura (LLM) → *Cap. 06*

Parámetro que controla la aleatoriedad en la generación de texto de un LLM. Con temperatura 0 el modelo debería ser más determinista, pero en la práctica sigue produciendo variaciones. GPT-4o con temperatura 0 produce respuestas idénticas solo el 3% de las veces.

Trasgu → *Cap. 04*

Duende travieso de la mitología asturiana que vive en las casas y revuelve el orden cuando nadie lo ve. En el libro representa el contenido mal estructurado: la información está ahí pero nadie —ni el usuario ni el modelo— puede encontrarla cuando la necesita.



Vector / Base vectorial → *Cap. 02*

Representación matemática de texto en forma de números. Las bases vectoriales almacenan embeddings de fragmentos de texto para permitir búsquedas por similitud semántica. Son el componente de almacenamiento y recuperación en los sistemas RAG.

Visibilidad en IA → *Cap. 06*

Frecuencia y prominencia con la que una marca aparece en las respuestas de los modelos de lenguaje cuando los usuarios hacen preguntas relacionadas con su sector. Es la métrica central del GEO y requiere medición con volumen suficiente de prompts para ser estadísticamente significativa.



Xana → *Cap. 02*

Hada de la mitología asturiana que vive en fuentes y ríos y guarda un conocimiento valioso sobre el entorno. Solo comparte ese conocimiento con quien se ha ganado su confianza. En el libro representa la memoria del LLM: si no has dejado rastro en las fuentes correctas antes de que alguien haga la pregunta, la Xana guarda silencio.
