¿Qué es llms.txt? Una guía SEO para el estándar web de IA

By rankstudio.net Published October 17, 2025

Resumen Ejecutivo

El archivo `/llms.txt` es un estándar web recientemente propuesto destinado a ayudar a los *modelos de lenguaje grandes* (LLMs) y a las herramientas de IA a descubrir, analizar e interpretar mejor el contenido de los sitios web. Análogo en espíritu al antiguo `robots.txt` para los rastreadores web, `llms.txt` actúa como un **mapa estructurado y curado de las páginas e información clave de un sitio** para agentes de IA. Los defensores argumentan que, debido a que los LLMs tienen ventanas de contexto limitadas y a menudo luchan por extraer contenido textual relevante de páginas web complejas, un `llms.txt` creado por humanos puede mejorar drásticamente la precisión de la IA al dirigir a los modelos directamente a los recursos de texto plano más importantes (Source: searchengineland.com) (Source: www.released.so). Los primeros en adoptarlo —incluidas plataformas de desarrolladores y algunas empresas tecnológicas— han comenzado a crear archivos `llms.txt`, y han surgido herramientas/generadores para ayudar en la implementación (Source: www.released.so) (Source: github.com).

Sin embargo, **el debate está lejos de resolverse**. Algunas voces de la industria advierten que `llms.txt` puede ser una solución prematura o innecesaria, argumentando que la optimización para motores de búsqueda (SEO) tradicional ya es suficiente para los casos de uso de la IA. Representantes de Google han declarado explícitamente que las Resúmenes de IA de Google se basan en el SEO estándar y *no* utilizarán `llms.txt` (Source: searchengineland.com). Del mismo modo, respetados profesionales del SEO señalan que los mecanismos existentes (por ejemplo, sitemaps XML o licencias Creative Commons) pueden satisfacer muchas necesidades sin un nuevo formato de archivo (Source: searchengineland.com) (Source: searchengineland.com). El análisis empírico muestra una adopción insignificante entre los 1.000 sitios web principales (efectivamente 0%) (Source: www.rankability.com) (Source: www.rankability.com), aunque comunidades más pequeñas reportan políticas relativamente altas de "permitir IA" en los sitios que *sí* lo implementan (Source: llmscentral.com). Sopesando las perspectivas de desarrolladores de IA, expertos en SEO, operadores de sitios web y defensores de la privacidad, este informe concluye que **`/llms.txt` es una innovación convincente en teoría, pero tiene un impacto práctico incierto**. Su valor probablemente dependerá de si los mantenedores de plataformas de IA realmente lo tienen en cuenta y de cómo los editores web equilibran los costos de crear metadatos de llms con los posibles beneficios de alcance de la IA.

Introducción y Antecedentes

A medida que la IA generativa y los modelos de lenguaje grandes (LLMs) como GPT de OpenAI y Gemini de Google se convierten en interfaces omnipresentes para la información, existe un creciente interés en hacer que la Web existente sea más **amigable para los LLM**. Actualmente, los sitios web están construidos principalmente para lectores humanos y motores de búsqueda tradicionales; los humanos navegan fácilmente por interfaces complejas, y Googlebot indexa páginas a través de enlaces y sitemaps. Pero los LLMs enfrentan una limitación crítica: *ventanas de contexto limitadas*. No pueden ingerir páginas web complejas enteras al por mayor y a menudo se distraen o confunden con barras de navegación, anuncios, scripts y otros elementos no textuales (Source: searchengineland.com) (Source: llms-txt.io). Como señala Jeremy Howard, el tecnólogo detrás de la propuesta de llms.txt:

> *"Los modelos de lenguaje grandes dependen cada vez más de la información de los sitios web, pero enfrentan una limitación crítica: las ventanas de contexto son demasiado pequeñas para manejar la mayoría de los sitios web en su totalidad. Convertir páginas HTML complejas con navegación, anuncios y JavaScript en texto plano amigable para LLM es difícil e impreciso." (Source: searchengineland.com)*

Esta limitación fundamental significa que un agente de IA que intenta responder a la pregunta de un usuario rastreando un sitio puede pasar por alto la información clave o malinterpretarla. Las técnicas tradicionales de SEO y diseño web enfatizan la usabilidad humana y la visibilidad en los motores de búsqueda, pero no abordan directamente las necesidades de los agentes de IA en tiempo de inferencia (Source: llms-txt.io). En la práctica, un LLM debe examinar el desorden de la página y aún así solo puede retener un extracto limitado. Por ejemplo, un desarrollador informó haber tenido que aplanar un sitio de documentación completo en un solo archivo de texto de 115.378 palabras (966 KB) para alimentar a un LLM con contexto completo (Source: searchengineland.com).

Para abordar esta brecha, **el archivo `/llms.txt` fue propuesto a finales de 2024** por Jeremy Howard (cofundador de Answer.AI y fast.ai) como una *extensión simpática* de los estándares de metadatos web. La idea es simple: en la raíz de un sitio web (al igual que con `robots.txt`), el webmaster puede colocar un archivo de texto plano Markdown llamado `llms.txt` que contenga:

- Un título H1 con el nombre del sitio o del proyecto (un elemento requerido).
- Una breve introducción o "resumen" en formato de cita en bloque, que proporcione contexto clave.
- Una o más secciones narrativas para explicar el sitio o su uso a una IA.
- Listas de viñetas bajo encabezados H2, cada una enumerando páginas importantes como enlaces Markdown `[Título](URL)` con descripciones opcionales.
- (Opcionalmente) Una sección "Opcional" separada para enlaces de menor prioridad que el LLM puede omitir si está restringido.

Dicho archivo tiene como objetivo funcionar como **"un mapa del tesoro para la IA"** (Source: www.linkedin.com). En lugar de obligar a la IA a analizar el HTML del sitio web, el `llms.txt` sirve como una **tabla de contenidos curada** que apunta a todo el contenido relevante. El archivo en sí está escrito en Markdown claro, eliminando scripts y navegación para que el LLM vea solo texto plano. En la práctica, un agente o herramienta de IA puede buscar "/llms.txt" y ver, por ejemplo, un título, un resumen de la empresa y luego secciones como "Productos" o "Documentos" con enlaces de viñetas. Esto le da al modelo acceso inmediato a las páginas y al contexto que sus creadores consideran más importantes.

La noción evoca esfuerzos históricos para hacer que la web sea "comprensible para las máquinas". De hecho, los críticos lo han comparado con la iniciativa de la **Web Semántica**, que estuvo inactiva durante mucho tiempo y que intentó anotar el contenido web para la interpretación de las máquinas (Source: news.ycombinator.com). La visión de Tim Berners-Lee de hace décadas de agentes "analizando todos los datos en la Web" en una "Web Semántica" de máquina a máquina nunca se realizó por completo (Source: news.ycombinator.com). El enfoque de `llms.txt` evita ontologías pesadas o esquemas RDF, confiando en cambio en texto plano. Como observó un defensor, evita la complejidad que aplastó el esfuerzo de la Web Semántica y utiliza "formatos sin estado" (Markdown, XML) para comunicarse con la IA (Source: news.ycombinator.com).

Fundamentalmente, `llms.txt` no se trata de *bloquear* o de control legal, sino de *guiar* a la IA. A diferencia de `robots.txt` (que utiliza reglas "Disallow: URL" para prohibir la indexación), `llms.txt` no tiene **directivas de bloqueo**. Es completamente opcional e instructivo: el propietario del sitio elige qué páginas destacar. Los implementadores enfatizan que es "más bien una elección sobre qué contenido debe mostrarse contextual o totalmente a una plataforma de IA" (Source: searchengineland.com). Efectivamente, le dice a un LLM "si quieres aprender sobre nuestro sitio, aquí es exactamente donde debes buscar". Por ejemplo, Howard y sus colaboradores describen el uso de un pequeño `llms.txt` para alimentar herramientas como Cursor o Claude con documentación precisamente curada, evitando la necesidad de que cada usuario recopile el contexto manualmente (Source: news.ycombinator.com).

Así, `/llms.txt` encarna una **visión colaborativa**: los sitios web colaboran explícitamente con "agentes" de IA de la misma manera que colaboran con los motores de búsqueda. Como lo resumió un comentario, *"LLMs.txt está a punto de cambiar la forma en que su contenido es visto, utilizado y protegido en el mundo de los modelos de lenguaje grandes"* (Source: llmsly.com). Desde este punto de vista, permite a los creadores de contenido "controlar su narrativa" informando a la IA con información autorizada (Source: www.linkedin.com). Los beneficios propuestos van desde una mayor precisión en las respuestas de la IA hasta un tráfico potencialmente medible desde interfaces de búsqueda impulsadas por IA. Los primeros experimentos de los profesionales han dado señales mixtas pero intrigantes: motores como los modelos de OpenAI aparentemente rastrean estos archivos, mientras que Google Search (hasta ahora) no los usa automáticamente (Source: searchengineland.com) (Source: searchengineland.com).

Sin embargo, la propuesta de `llms.txt` no es universalmente aceptada. Los críticos señalan *tensiones entre la elegancia y la practicidad*. Si bien `llms.txt` puede simplificar el rastreo de la IA, esencialmente duplica lo que el contenido bien diseñado ya debería hacer: ser accesible y claro para *todos* los lectores (humanos o IA). Como señaló un comentarista, "Esto no es una buena UX para las máquinas. Esto es un parche para una mala UX" – una tirita en lugar de arreglar los diseños imprecisos subyacentes (Source: news.ycombinator.com). Otros se preocupan de que, sin un proceso de estandarización robusto (por ejemplo, el registro formal de una URI o metaetiquetas conocidas), el formato pueda fragmentarse. Expertos de alto perfil también advierten que exigir a los propietarios de sitios que creen manualmente otro archivo los carga, dado que *ningún sistema de IA lo usa actualmente* (Source: searchengineland.com) (según Google) o aparentemente ha solicitado dicho archivo. Incluso existe el punto de vista de que las licencias web existentes (Creative Commons, etc.) podrían gobernar el uso de la IA de manera más limpia que un nuevo archivo de texto (Source: searchengineland.com).

En las secciones siguientes, profundizamos en **qué es `/llms.txt`, cómo se supone que funciona y por qué puede o no importar**. Examinamos la especificación técnica y el formato (tal como se propone actualmente), las herramientas para su generación y las diferencias con estándares relacionados como `robots.txt` y `sitemap.xml`. Revisamos el **estado actual de adopción**, incluidos estudios de caso (por ejemplo, empresas que prueban `llms.txt` para documentos de productos) y datos sobre cuántos sitios lo han implementado. Resumimos las perspectivas de desarrolladores de IA, especialistas en SEO y defensores de la privacidad, utilizando entrevistas y declaraciones publicadas. También discutimos cómo están reaccionando las plataformas de IA: algunas prueban activamente `llms.txt`, otras permanecen agnósticas (Source: searchengineland.com) (Source: searchengineland.com). Finalmente, exponemos las posibles implicaciones para el futuro: desde cómo las empresas gestionan su contenido digital hasta cómo evolucionarán los motores de búsqueda y generativos. A través de exhaustivas citas y análisis, el informe pretende responder: *¿Es `/llms.txt` realmente revolucionario para la búsqueda de IA, o solo otra pieza de desorden digital?* La evidencia inicial sugiere que puede ser importante para casos de uso específicos (como documentos de desarrolladores y sitios pequeños), pero su impacto general en el descubrimiento web convencional aún está por verse.

El Estándar `/llms.txt`: Detalles Técnicos y Propósito

La propuesta y especificación de `/llms.txt` están documentadas de manera más completa por sus creadores en [llmstxt.org] y los repositorios de GitHub asociados (Source: llmstxt.org) (Source: github.com). En resumen, un archivo `llms.txt` es un documento **Markdown** de texto plano, ubicado en la raíz de un sitio web (por ejemplo, `https://example.com/llms.txt`). Utiliza la sintaxis Markdown para presentar contenido estructurado, lo que lo hace tanto *legible para humanos* como analizable por máquinas. El formato evita intencionalmente el anidamiento arbitrario o las etiquetas desconocidas, a favor de una disposición bien definida de encabezados, párrafos, citas en bloque y listas. El elemento mínimo requerido es simplemente un encabezado de nivel superior (H1) que contenga el título del sitio o proyecto (Source: llmstxt.org). Más allá de eso, la especificación define los siguientes componentes, en orden:

- **Título H1 (obligatorio)** – El nombre del proyecto o sitio web (por ejemplo, el nombre de una empresa). Esto ancla la identidad del archivo.
- **Resumen de texto plano (opcional)** – Una **cita en bloque de Markdown** que contiene una breve descripción o declaración de visión. Este "discurso de ascensor" proporciona contexto de antemano.
- **Secciones introductorias (opcional)** – Cualquier número de párrafos o listas (pero *no* encabezados adicionales) que proporcionen detalles sobre el sitio o instrucciones para interpretar los enlaces posteriores. Pueden ser texto plano, listas de viñetas, etc.
- **Secciones de enlaces H2 (opcional)** – Cero o más subsecciones, cada una encabezada por un H2. Cada H2 va seguido de una lista de viñetas de enlaces (anclas Markdown `[texto](URL)`), opcionalmente con notas delimitadas por dos puntos. Estas compartimentan el contenido del sitio por categoría. Por ejemplo:

  ```
  ## Documentación
  - [Referencia de API](https://example.com/api): Documentación detallada de la API para desarrolladores.
  - [Guías](https://example.com/guides): Tutoriales paso a paso.
  ```

  Dichas secciones se tratan como "listas de archivos" de URL en la especificación; los LLMs o herramientas pueden iterar a través de ellas.
- **Sección "Prioridad Baja" Opcional** – Se recomienda (pero no es obligatorio) que una sección final titulada "Opcional" enumere las páginas de menor prioridad, para que un LLM pueda omitirlas si su ventana de contexto es limitada.

Esta estructura tiene como objetivo imitar la forma en que los humanos podrían resumir la arquitectura de información de un sitio. El archivo *en sí* está escrito en Markdown específicamente *porque* Markdown es fácilmente analizable tanto por LLMs como por humanos (Source: llmstxt.org) (Source: golevels.com). El formato es lo suficientemente inequívoco como para que las herramientas automatizadas lo procesen utilizando un análisis de texto simple (incluso métodos basados en expresiones regulares o XML, como muestra el ejemplo de FastHTML) (Source: llmstxt.org) (Source: github.com). Críticamente, la especificación enfatiza que el contenido de `llms.txt` debe ser conciso y relevante; no debe simplemente volcar el contenido completo de las páginas sin crítica. En cambio, **destaca** las URL y los hechos que el propietario del sitio considera más importantes para que la IA los ingiera.

Por ejemplo, la [especificación oficial de llmstxt.org] (y la [descripción de AnswerDotAI en GitHub]) proporciona un modelo ilustrativo:

```
# Título del Sitio de Ejemplo

> Este es un resumen conciso del propósito del sitio web y sus ofertas clave. Podría mencionar la industria, los producto

Las siguientes secciones enumeran las áreas de contenido más importantes de este sitio para que la IA las considere.

## Guías

- [Primeros Pasos](https://example.com/start): Una introducción para nuevos usuarios.
- [Documentación de API](https://example.com/api): La referencia completa de la API.
- [Preguntas Frecuentes](https://example.com/faq): Preguntas frecuentes.

## Proyectos

- [Proyecto Alfa](https://example.com/alpha): Información detallada sobre el Proyecto Alfa.
- [Proyecto Beta](https://example.com/beta): Resumen del Proyecto Beta.

## Opcional

- [Blog](https://example.com/blog): Noticias y actualizaciones (omitir si es limitado).
```

Este ejemplo demuestra el uso previsto: una IA que lee `llms.txt` ve un resumen y luego listas claramente estructuradas de URL relevantes con etiquetas o notas cortas. Con esto, los modelos pueden precargar resúmenes de páginas clave en lugar de rastrear todo el sitio a ciegas.

Un aspecto clave de `llms.txt` es que **no intenta reemplazar los estándares web**, sino complementarlos para el uso de la IA. Por ejemplo, podría funcionar *implícitamente* como un sitemap adicional (listando páginas) pero con contexto descriptivo. La especificación explícitamente *no* define reglas restrictivas; más bien, es informativa. Como señala un explicador, `llms.txt` es "similar a robots.txt... pero también ofrece un beneficio adicional: el aplanamiento completo del contenido" (Source: searchengineland.com). En otras palabras, mientras que robots.txt dice a las máquinas *qué no rastrear*, `llms.txt` les dice *qué rastrear* (y por qué). Es más parecido a un **mapa del sitio curado por humanos** extendido, combinado con documentación. De hecho, una guía lo llama formalmente *"el nuevo robots.txt para la era de los LLM"* (Source: www.released.so), enfatizando que **guía** a los LLM para evitar conjeturas.

En el lado práctico, la propuesta de llms.txt y las herramientas relacionadas prevén que las páginas web con contenido útil también ofrezcan "versiones Markdown limpias" de esas páginas (por ejemplo, en la misma URL pero con una extensión `.md`) (Source: llmstxt.org). Esta sugerencia es como proporcionar HTML preprocesado para máquinas, pero no es estrictamente requerida por el estándar llms.txt en sí. El principal producto de esta iniciativa es el archivo llms.txt, que también puede listar enlaces opcionales (en sus secciones) a dichos recursos Markdown si están disponibles. Algunos proyectos, como FastHTML, han ido más allá al convertir programáticamente sus páginas específicas de mm a Markdown y luego referenciarlas en las listas de llms.txt (Source: github.com). El ejemplo de FastHTML es instructivo: su `llms.txt` se expandió automáticamente en archivos "llms-ctx.txt" y "llms-ctx-full.txt" que incorporan el texto de las páginas enlazadas, adaptados a las necesidades de contexto XML del modelo Claude (Source: github.com).

En resumen, `llms.txt` es una **convención** —aún no un estándar formal de la IETF— sobre cómo publicar metadatos del sitio consumibles por IA. Prescribe un nombre de archivo y un formato específicos, pero deja mucha flexibilidad a los propietarios del sitio. La esperanza es que, al anunciar y documentar esta convención (a través de llmstxt.org y GitHub), los desarrolladores y las empresas comiencen a adoptarla voluntariamente. Si suficientes proveedores de contenido lo hacen, los desarrolladores de IA (o las herramientas de usuario final) podrían verificar programáticamente `yourwebsite.com/llms.txt` como una fuente conocida y buena de contenido en la página.

Relación con los Estándares Existentes (Robots.txt, Sitemaps, etc.)

Para evaluar la importancia de `llms.txt`, es crucial contrastarlo con los estándares web más establecidos que sirven a los motores de búsqueda. La comparación más natural es `robots.txt`, que ha gobernado el comportamiento de los rastreadores web desde la década de 1990. Si bien tanto `robots.txt` como `llms.txt` comparten la idea de un archivo conocido en la raíz del sitio, sus funciones divergen bruscamente. `robots.txt` es un **conjunto de comandos** para robots web: les dice a los motores de búsqueda (a través de directivas como `User-agent` y `Disallow`) qué partes del sitio *no* deben ser raspadas o indexadas. En contraste, `llms.txt` **no se trata de bloquear**. Proporciona una guía positiva —esencialmente una tabla de contenido rápida— sobre *qué incluir* en el contexto de un LLM. Como explica Search Engine Land, "los archivos robots.txt funcionan bien para los rastreadores y no necesitan cambios para el propósito de los LLM" (Source: searchengineland.com), porque el caso de uso de robots.txt (gobernar los permisos de rastreo) es ortogonal al de llms.txt (mejorar la ingesta de contenido).

Otro análogo útil es el sitemap XML (`sitemap.xml`). Un sitemap es simplemente una lista de URLs formateadas en XML, opcionalmente con metadatos como fechas de última modificación o prioridades, destinado enteramente a los motores de búsqueda. No contiene contexto descriptivo ni resúmenes; simplemente enumera páginas para su descubrimiento. Por el contrario, `llms.txt` es como un **mapa del sitio contextual**. Todavía lista enlaces, pero en un formato anotado y legible por humanos. Una guía de marketing señala que "a diferencia de un sitemap.xml (que es solo una lista de URLs), llms.txt proporciona contexto y estructura para cada enlace" (Source: golevels.com). En cierto modo, se puede ver `llms.txt` como la fusión de los conceptos de un sitemap y alguna forma de página "Acerca de": enumera las páginas clave y explica qué son.

Podemos resumir algunas distinciones clave en la siguiente tabla:

ASPECTO / ARCHIVO	ROBOTS.TXT	SITEMAP.XML	LLMS.TXT
Propósito	Controlar la indexación del rastreador (excluir páginas) (Source: searchengineland.com)	Informar a los bots de búsqueda de todas las URLs y metadatos del sitio	Guía curada de contenido importante para LLM (Source: llms-txt.io) (Source: golevels.com)
Tipo de Contenido	Directivas de texto plano (ej. `Disallow:`)	XML con entradas `<url>`	Markdown: encabezados, listas, enlaces, texto
Audiencia/Agente	Rastreadores de motores de búsqueda (Googlebot, etc.)	Rastreadores de motores de búsqueda	Sistemas de IA y agentes basados en LLM
Diferencia Clave	Dice a los bots *qué omitir*	Lista *todas las páginas a incluir*	Destaca *en qué enfocarse*
¿Legible por humanos?	Sí (comandos simples)	No (formato XML de máquina)	Sí (Markdown plano con descripciones) (Source: golevels.com)
Ejemplo de Uso	`Disallow: /private/` bloquea la ruta	`<loc>https://example.com/page.html</loc>`	`- [FAQ] (https://exa.com/faq): temas comunes`

(Fuentes: Consulta de propuestas de llms.txt y guías de SEO (Source: searchengineland.com) (Source: golevels.com) (Source: llms-txt.io).)

Lo anterior destaca que los estándares existentes satisfacen diferentes necesidades. La optimización SEO tradicional (a través de HTML adecuado, metaetiquetas, datos estructurados, sitemaps, etc.) sigue siendo fundamentalmente para usuarios humanos y los algoritmos de Google (Source: llms-txt.io) (Source: llms-txt.io). `llms.txt` reconoce explícitamente que *esos métodos son insuficientes para la IA*. De hecho, como señala un análisis, los LLM "tienen una capacidad finita para procesar información a la vez" y "el contenido optimizado por palabras clave no siempre proporciona la comprensión completa que los LLM necesitan" (Source: llms-txt.io). En otras palabras, un sitio fuertemente optimizado para SEO podría clasificarse bien en Google pero aún así confundir a una IA, haciendo que pierda contexto o ingiera información inútil. `llms.txt` se ofrece como un suplemento —*no un reemplazo*— para las prácticas de SEO (Source: llms-txt.io) (Source: llms-txt.io). Un buen SEO (páginas rápidas, encabezados claros, etc.) sigue siendo necesario para la visibilidad general, mientras que `llms.txt` aseguraría adicionalmente que la *IA* vea la esencia de su contenido.

Otras ideas relacionadas en la industria apoyan esta división. Por ejemplo, algunos han sugerido añadir etiquetas especiales `<meta name="LLM">` o sugerencias de encabezado HTTP para indicar contenido amigable para la IA. Un estratega de SEO incluso propuso un enlace `rel="llm"` o un perfil MIME para Markdown amigable para LLM (Source: news.ycombinator.com). Estas propuestas comparten el objetivo de señalar contenido relevante a la IA, pero difieren en la implementación. `llms.txt` fue elegido (al menos inicialmente) como un archivo simple en la raíz para evitar requerir cambios en el diseño HTML o la configuración del servidor HTTP. Los defensores de `llms.txt` argumentan que un archivo de texto independiente es una solución de baja fricción: cualquier sitio que aloje contenido estático puede insertar un archivo Markdown sin riesgo de romper la presentación del sitio.

Es importante destacar que el gigante de búsqueda web Google ha opinado sobre este ecosistema proliferante. En un informe de Search Engine Land de julio de 2025, Gary Illyes de Google (del equipo de Search Central) dijo explícitamente que **Google no procesará archivos llms.txt**: "Las descripciones generales de IA de Google se basan en el SEO estándar; no necesita llms.txt ni ningún archivo especial" (Source: searchengineland.com). Illyes reafirmó en una discusión pública que Google "no apoya LLMs.txt y no planea hacerlo" (Source: searchengineland.com). En cambio, Google instruye a los webmasters a *simplemente usar SEO normal* para ser visibles en las funciones de "Resumen de IA" impulsadas por IA. En contraste, algunos productos de IA de startups más pequeñas (como los motores de OpenAI o Claude) parecen estar explorando o incluso leyendo activamente estos archivos. Por ejemplo, un desarrollador web informó que el rastreador de OpenAI estaba accediendo a los puntos finales /llms.txt de sus sitios cada pocos minutos (Source: searchengineland.com). Así, en la actualidad, parece que llms.txt puede ser relevante para herramientas de IA especializadas, pero no para la indexación de búsqueda principal.

En resumen, **llms.txt ocupa un nuevo espacio**: está explícitamente destinado no a los motores de búsqueda, sino a los agentes de IA. Complementa en lugar de reemplazar a `robots.txt` o `sitemap.xml`. Está *inspirado en* estas convenciones más antiguas (de ahí que a veces se le llame el "robots.txt para IA" (Source: www.released.so), pero su guía es de una naturaleza diferente. Si los LLM y las empresas adoptarán esta convención es una cuestión central (abordada más adelante), pero técnicamente llena un nicho único: hacer que el contenido complejo del sitio sea fácilmente consumible por la IA generativa.

La Razón de Ser: Por Qué `/llms.txt` Podría Importar

Comprender la importancia de `llms.txt` requiere examinar las motivaciones y los beneficios anticipados desde múltiples ángulos: para los propietarios de contenido, para los desarrolladores de IA y para los usuarios finales.

1. Control sobre la interpretación de la IA: El beneficio más citado es dar a los propietarios de sitios web cierto control sobre cómo la IA utiliza su contenido. En el panorama actual, los grandes modelos de IA suelen entrenarse con raspados web masivos y sin categorizar (por ejemplo, Common Crawl) o recuperan páginas ad-hoc sin guía (Source: privacyinternational.org). Autores y empresas han expresado su preocupación de que este proceso pueda tergiversar o malinterpretar su contenido, o que la IA pueda responder preguntas de los usuarios sin dar la "citación" o el contexto adecuados. Al proporcionar `llms.txt`, un sitio puede *destacar* las páginas y datos exactos que *quiere* que las IA lean. Esto puede asegurar, por ejemplo, que se incluyan descripciones de productos o términos legales, mientras que se omiten páginas sin importancia (como menús de navegación, inicio de sesión o páginas de error). Según los autores de la propuesta, esta transparencia puede ser una forma de gestión de derechos de contenido: los sitios web pueden señalar eficazmente qué contenido permiten que un LLM "ingiera" para responder consultas (Source: llmscentral.com) (Source: www.released.so). Desde este punto de vista, llms.txt se convierte en una contraparte del debate en curso sobre los datos de entrenamiento de la IA y los derechos de autor. Como señala Search Engine Land, los creadores de contenido lo ven como "cierta garantía de un mayor control por parte del propietario, en términos de qué y cuánto debe ser accedido" (Source: searchengineland.com).

2. Mejora de la calidad de las respuestas de la IA: Cuando un LLM tiene acceso directo a una base de conocimientos concisa, la calidad de su generación mejora. Si un asistente de IA responde preguntas sobre su sitio o dominio, usted quiere que tenga fuentes autorizadas de las que extraer información. El análisis de HTML sin procesar puede producir "alucinaciones" o omisiones sin contexto. Por el contrario, un archivo `llms.txt` bien elaborado resume los hechos clave y enlaza información actualizada. Los profesionales han informado que, después de alimentar a un LLM con el contenido listado en `llms.txt`, la IA proporciona respuestas más precisas y relevantes sobre el sitio. Por ejemplo, un profesional probó un archivo `llms.txt` para una empresa llamada Enhance Media utilizando tres modelos (ChatGPT, Gemini, Claude) y descubrió que *los tres pudieron resumir correctamente el negocio solo a partir de ese archivo* (Source: www.linkedin.com). El formato estructurado del archivo ayudó a los modelos a centrarse rápidamente en los puntos más importantes. De manera similar, los creadores de FastHTML descubrieron que el contexto cuidadosamente curado (a través de un archivo `llms.txt` expandido) producía "resultados dramáticamente mejores" de Claude y otras herramientas que el raspado no dirigido (Source: news.ycombinator.com).

3. Eficiencia técnica: Los rastreadores a gran escala (especialmente para modelos de IA más pequeños) consumen muchos recursos. Las empresas de LLM deben equilibrar la frecuencia con la que vuelven a rastrear los sitios para obtener datos frescos. Una oferta de `llms.txt` puede servir como una **baliza de frescura**: puede permitir que un rastreador de IA verifique un solo archivo en busca de actualizaciones en lugar de rastrear todo el sitio. De hecho, como se informó en [33], al menos un sistema de OpenAI estaba sondeando los `llms.txt` de los desarrolladores cada 15 minutos en busca de actualizaciones. Este tipo de flujo de trabajo optimizado puede reducir la carga innecesaria tanto en la IA como en los servidores web. También puede garantizar que la versión del contenido a la que está expuesta la IA sea la versión oficial y aplanada proporcionada por el sitio, no un fragmento parcial o desactualizado. En efecto, `llms.txt` podría servir como una especie de "API" para el contenido de sitios estáticos, aunque sin la estructura formal de una llamada a la API.

4. Igualar el campo de juego: Los sitios más pequeños y las nuevas startups pueden ver llms.txt como una forma de competir por la atención en la búsqueda impulsada por IA. Algunos analistas han establecido un paralelismo con las primeras estrategias de SEO: en los inicios de la web, las pequeñas empresas utilizaban `robots.txt`, metaetiquetas y sitemaps para destacarse ante los motores de búsqueda. Ahora, si los agentes de IA se convierten en nuevos "curadores" de contenido, cualquier sitio puede usar `llms.txt` para destacarse también ante ellos. Este ángulo democratizador se menciona explícitamente en los materiales promocionales: al añadir `llms.txt` e incluso compartirlo en plataformas como GitHub, *"darás forma a cómo la IA trata tu contenido"* (Source: llmsly.com). La idea es que los sitios web con visión de futuro pueden obtener una ventaja reputacional al ser los primeros en asociarse con la IA.

5. Precedente de "robots" de IA: Ya, algunas herramientas de IA se presentan como agentes que rastrean la web. Por ejemplo, Claude Projects (una integración IDE) puede tomar archivos de documentación en contexto. Dichas herramientas a menudo requieren que los usuarios las dirijan a documentos o datos clave. `llms.txt` puede automatizar ese proceso. Al ofrecer un archivo ancla conocido, los propietarios de sitios pueden inscribirse automáticamente en estos ecosistemas de IA emergentes. Es similar al papel inicial de `robots.txt`: al principio, pocos sitios lo usaban, pero a medida que Googlebot y otros aprendieron a verificarlo, se convirtió en estándar. Los primeros en adoptar `robots.txt` (alrededor de 1994-95) lo hicieron para guiar a los rastreadores de AltaVista o Google. Hoy, los diseñadores de `llms.txt` esperan que los "arquitectos de la IA" (algunos equipos líderes de IA) hagan algo similar. De hecho, los creadores a menudo destacan que los desarrolladores de Anthropic están promoviendo `llms.txt` en su documentación, y que empresas como Mintlify crearon soporte para ello (Source: www.released.so). En resumen, `llms.txt` es importante para sus defensores porque **aborda directamente un cuello de botella técnico de los sistemas de IA actuales**. Promete una forma sencilla de hacer que la Web sea más "compatible con LLM", lo que potencialmente facilita el trabajo de la IA y mejora sus respuestas.

Adopción, Respuesta de la Industria y Casos de Estudio

¿Qué tan ampliamente se está utilizando `llms.txt` en la práctica, y quién le está prestando atención? Desde que la idea surgió a finales de 2024, la adopción ha sido limitada y desigual, pero ciertos grupos de actividad son notables.

En primer lugar, **las empresas tecnológicas y las plataformas de documentación** han mostrado interés. En noviembre de 2024, la plataforma de documentación Mintlify anunció soporte integrado para `llms.txt` para proyectos publicados en su sitio (Source: www.released.so). Esto significó que, prácticamente de la noche a la mañana, *miles* de documentaciones de proyectos de software se volvieron accesibles a través de `llms.txt`. La publicación del blog de Jens Schumacher señala: "De un solo golpe, hicieron que la documentación de miles de herramientas para desarrolladores fuera compatible con LLM, como Anthropic y Cursor" (Source: www.released.so). Los proyectos de herramientas para desarrolladores cuya documentación se ejecuta en Mintlify (por

ejemplo, muchas bibliotecas de código abierto) adquirieron así archivos llms.txt sin necesidad de una acción individual por parte de los mantenedores. De manera similar, algunas empresas tecnológicas están creando explícitamente `llms.txt`. En [15], Radu Stoian afirma que **Anthropic** (la empresa detrás de la IA Claude) y otros no especificados solicitaron públicamente archivos llms.txt para sus sitios: “Líderes de IA como Anthropic… lo han iniciado… han construido sus modelos con la expectativa de encontrar este archivo” (Source: www.linkedin.com). Hemos verificado de forma independiente que https://www.anthropic.com/llms.txt (o el enlace equivalente generado estáticamente) realmente existe y enumera docenas de páginas en el sitio de Anthropic (Source: llmstxtgen.com).

Más allá de los desarrolladores, **consultorías y agencias** han comenzado a recomendar llms.txt. Por ejemplo, un autor de un blog orientado a los negocios lo llama “tu nueva arma secreta” para la optimización de IA (Source: llmsly.com). Otros sitios web centrados en SEO y artículos de LinkedIn aclaman a `llms.txt` como “esencial para las marcas” en la era de la IA (Source: www.linkedin.com), dándole una gran visibilidad dentro de los círculos de marketing. Un número significativo de empresas más pequeñas y proveedores de servicios (desde agencias de SEO hasta proveedores de IA) han publicado blogs sobre cómo implementar `llms.txt` en los sitios de sus clientes. Este entusiasmo es en parte exploratorio: muchos ven el contenido de IA como la próxima frontera de visibilidad y están tratando a llms.txt como una buena práctica a probar.

Sin embargo, cuando examinamos el *uso real*, la situación es mixta. Un directorio de archivos llms.txt de origen colectivo, [llmstxt.site], da fe de cientos de sitios web donde se ha detectado `llms.txt` (Source: llmstxt.site). Este directorio enumera docenas de sitios de ejemplo y sus recuentos de tokens. Por ejemplo, la popular herramienta de diseño **Framer** tiene un `llms.txt` con aproximadamente 1.821 tokens (tamaño de texto) (Source: llmstxt.site). La empresa fintech **Klarna** (en su subdominio de documentación) tiene 17.387 tokens en su `llms.txt` (Source: llmstxt.site). Incluso un sitio de contenido aparentemente grande, Weather.com (The Weather Company), aparece con un llms.txt (¿en blanco?) (0 tokens) (Source: llmstxt.site), lo que sugiere que podría haber creado el archivo pero lo dejó vacío. A menor escala, muchos blogs personales, educativos y tecnológicos han implementado `llms.txt`, ocasionalmente con miles o incluso cientos de miles de tokens. Por ejemplo, un blog de astrología “LookUpTheStars” informa de un `llms.txt` con ~385.221 tokens (Source: llmstxt.site). En el otro extremo, algunos archivos `llms.txt` tienen solo unos pocos cientos de palabras (por ejemplo, Ideanote.io tenía 1.106 tokens) (Source: llmstxt.site). Nuestra encuesta del directorio llmstxt.site revela una adopción *experimental* generalizada: empresas de varios tamaños, desde productos de software hasta comercio electrónico de nicho, han creado estos archivos (a menudo convirtiendo sitemaps existentes o listas de enlaces manuales). Muchos parecen seguir el formato de la especificación con precisión, mientras que algunos tienen implementaciones incompletas o ascendentes (ejemplos de consejos para el analizador están disponibles en foros de la comunidad).

Para obtener una visión más amplia de la adopción, se han reportado dos análisis por parte de terceros. Uno es un “Informe de la Industria” de un sitio llamado *LLMS Central*, que afirma haber analizado 2.147 sitios web en 15 industrias a principios de 2025 (Source: llmscentral.com). Sus estadísticas principales son que el **68% de los sitios “permiten” el entrenamiento de IA** (con políticas totalmente abiertas o selectivas), el 23% “permiten todo”, el 45% tienen “políticas selectivas”, el 18% bloquean todo y solo el 14% no tiene `llms.txt` en absoluto (Source: llmscentral.com). Interpretan esto como que la mayoría de los sitios están publicando alguna guía para los LLM. En particular, en su muestra de empresas de tecnología y software (n=387), informan que el 95% tiene algún tipo de política explícita de `llms.txt` (Source: llmscentral.com). Sin embargo, estas cifras deben tomarse con cautela. El informe no revela cómo se eligieron los sitios ni si simplemente buscaron cualquier mención de `llms.txt`. Es posible que su conjunto de datos esté enriquecido con empresas ya involucradas en IA/tecnología, lo que sesga los porcentajes al alza.

En marcado contraste, una firma de análisis SEO, *Rankability*, publicó un “Informe mensual de adopción de LLMS.txt” centrado en los **1.000 principales sitios web comerciales** por tráfico (Source: www.rankability.com). Encontraron una adopción prácticamente **nula**: una tasa de adopción del 0,3% (efectivamente 3 de cada 1000) (Source: www.rankability.com). Afirman sin rodeos “Adopción actual cero” (Source: www.rankability.com), con un extenso escaneo automatizado que arrojó casi ningún resultado positivo. Por industria, sus datos muestran un 0,00% de adopción en los sectores de comercio electrónico, redes sociales, finanzas, atención médica y gobierno, con solo un 0,73% de adopción en el sector educativo (lo que sugiere que quizás 7 de cada 1000 son universidades o valores atípicos similares) (Source: www.rankability.com). En resumen: entre los sitios más grandes del mundo, prácticamente ninguno implementa `llms.txt` a mediados de 2025. Esto implica que el estándar sigue siendo de nicho.

¿Por qué tal discrepancia? Parece que la adopción se ha concentrado entre **sitios más pequeños o con orientación tecnológica**, y prácticamente ninguna entre las principales marcas convencionales. La lista de los 500-1000 principales comprende gigantes globales (Amazon, YouTube, etc.) con equipos de SEO arraigados; evidentemente, aún no ha penetrado en esos círculos. En comparación, los sitios pequeños y medianos, las bases de conocimiento y las herramientas para desarrolladores

se han volcado a él. Los datos de Rankability sugieren que se encontraron exactamente uno o dos *casos aislados* entre 1000 (probablemente sitios pequeños que se clasificaron justo dentro de los 1000). Mientras tanto, el informe de LLMS Central probablemente muestreó empresas al menos parcialmente involucradas en discusiones sobre IA, de ahí sus cifras de adopción más altas. Esta brecha entre la "comunidad entusiasta" y el "mercado masivo" será importante para evaluar cuánto impacto real puede tener `llms.txt`.

Dadas estas cifras, es justo decir que **`llms.txt` tiene chispa pero no (todavía) llama**. Importa en ciertos ecosistemas (especialmente en la documentación de software y los comentarios de agencias de SEO) pero no de forma generalizada en toda la web. Dicho esto, las tendencias de adopción podrían acelerarse si plataformas importantes como Google o Bing de Microsoft deciden aprovecharlo. Alternativamente, puede seguir siendo una optimización opcional para un subconjunto de propietarios de sitios. A continuación, exploramos algunos ejemplos detallados del uso de `llms.txt`, así como las reacciones de los desarrolladores de herramientas de IA.

Caso de estudio: Documentación técnica

Un caso de uso temprano y lógico es la documentación técnica de software. La documentación para desarrolladores a menudo ya genera contenido HTML a partir de marcado (por ejemplo, Markdown) y generalmente se esfuerza por ser legible tanto por máquinas como por humanos. También se benefician enormemente de respuestas precisas. La biblioteca FastHTML discutida anteriormente es un ejemplo: sus desarrolladores crearon entradas `llms.txt` para ayudar a las IA orientadas a desarrolladores. Otro ejemplo destacado es la **documentación para desarrolladores de Klarna** (la empresa europea de pagos). Según el directorio llmstxt, la documentación de Klarna (alojada en docs.klarna.com) incluye un `llms.txt` con aproximadamente 17.387 tokens (Source: llmstxt.site).

De manera similar, un proyecto de GitHub "pgai/llms.txt" indica que el proyecto Postgres AI (Timescale) añadió un archivo `llms.txt` a su repositorio, sugiriendo una implementación en un producto de base de datos real (Source: github.com). Las API corporativas, las bibliotecas de código abierto y las plataformas en la nube (el directorio enumera entradas para AWS, Azure docs, etc.) también han comenzado a adoptar el formato. Estos usos tienen sentido: es probable que los usuarios técnicos se beneficien de tener resúmenes de documentación claros y legibles por IA.

Caso de estudio: Sitios de servicios y profesionales

No toda la adopción se da en la alta tecnología. Por ejemplo, el directorio SEO enumera a **HoodChefs** (un servicio de alquiler de cocinas) con 44.494 tokens (Source: llmstxt.site), y un sitio web de un **concesionario de automóviles** "AutoChampion24" en Alemania con 6.750 tokens (Source: llmstxt.site). Estas entradas demuestran que incluso las pequeñas empresas ven potencial. "GalaxxiaMarketing" (una firma de marketing brasileña) tiene 676 tokens (Source: llmstxt.site), aparentemente presentando sus servicios a través de llms.txt. También se han detectado sitios religiosos y espirituales, blogs personales y proveedores de e-learning. La existencia de un sitio como "lookupthestars.com" con 385k tokens (Source: llmstxt.site) es notable: parece ser un sitio de temática astrológica que adoptó completamente el estándar.

Es difícil verificar la motivación comercial de cada `llms.txt` ad hoc, pero es probable que muchos lo hicieran por curiosidad o para experimentar con el SEO. Las contribuciones de la comunidad a los directorios de llmstxt sugieren que se han creado plugins de WordPress para autogenerar llms.txt, y los desarrolladores en los foros mencionan ocasiones en que sus bots de tutoría de IA vieron por primera vez soporte para llms-txt.

Respaldos de la industria

Algunos actores importantes al menos han **reconocido** el concepto. El blog de Cloudflare (mayo de 2025) analiza cómo sus nuevos servicios AI Gateway se integran con otras herramientas de IA, aunque no menciona directamente llms.txt (Source: www.cloudflare.net). Más relevante es Anthropic: su portal de documentación ahora incluye un enlace visible al archivo "LLMS.txt", y han "publicado en X" sobre su soporte (Source: www.released.so). En resumen, las **empresas orientadas a la IA** son al menos curiosas. En contraste, las empresas tecnológicas o de medios convencionales han guardado silencio. No tenemos conocimiento de ningún informe de adopción de llms.txt por parte de Google, Amazon (más allá de los que figuran en el directorio público) o Facebook.

Métricas y análisis

Existen pocos datos sobre la *efectividad* de llms.txt. Una métrica aproximada proviene de un autor de LinkedIn que examinó las analíticas de Google Search Console. Afirmó que Google ya había indexado un archivo llms.txt de un sitio de prueba (Source: www.linkedin.com), aunque Google dice que no los *usará*. Otro rastro citado son los registros del servidor: un webmaster notó que los rastreadores de *OpenAI* hacían ping a los archivos llms.txt de sus sitios cada 15 minutos para verificar su frescura (Source: searchengineland.com). Esta anécdota sugiere que al menos algunos servicios avanzados de búsqueda/IA están prestando atención.

Otras métricas podrían incluir cambios en las respuestas a las consultas o en el tráfico de referencia. Al momento de escribir esto, dichos datos en su mayoría no son públicos. En teoría, se podría rastrear el tráfico de las interfaces de chat de IA (a través de etiquetas UTM especiales o "referencias" de API), pero pocos propietarios de sitios tienen dicho seguimiento implementado. Algunos artículos de SEO sugieren usar API personalizadas para monitorear el tráfico impulsado por LLM, pero los ejemplos concretos son escasos (la guía de golevels.com lo discute conceptualmente). Las primeras señales en los resultados de búsqueda también pueden indicar uso. Una publicación de LinkedIn de un consultor de SEO mostró resultados de búsqueda de Google destacando un archivo `llms-full.txt` en los resultados, insinuando la indexación (Source: distinctly.co), pero no está claro si eso es oficial o un error.

Adopción por región o sector

Los datos de Rankability muestran que la **educación** es el único sector con una presencia medible (0,73%) en los principales sitios (Source: www.rankability.com). Esto podría deberse a que universidades o proyectos académicos están experimentando con el formato. En contraste, sectores como el comercio electrónico, las redes sociales, las finanzas, la atención médica y el gobierno tuvieron un 0% en el top 1000 (Source: www.rankability.com). El informe de LLMS Central (aunque menos autoritario) indica que las empresas de **tecnología/software** son líderes en adopción, con un "95% con políticas explícitas" dentro de ese segmento (Source: llmscentral.com). Esto coincide con la intuición: los editores de tecnología son los primeros bancos de pruebas de la tecnología de IA.

Críticas, preocupaciones y perspectivas alternativas

Para mantener el equilibrio, debemos abordar las razones por las que `/llms.txt` podría *no* prosperar o ser problemático. Han surgido varias críticas de desarrolladores, expertos en SEO y escépticos. Las organizamos aquí:

A. Duplicación de esfuerzos y preocupaciones sobre la UX: Los críticos observan que si un sitio ya está bien estructurado y tiene páginas de "ayuda" o "acerca de", añadir llms.txt puede ser redundante. Como señaló una discusión en Hacker News: *"Esto no es una buena UX para máquinas. Esto es un parche para una mala UX para ayudar a los LLM... Algunos sitios web tienen el mismo parche para humanos en forma de una sección de 'Ayuda' o 'Acerca de'"* (Source: news.ycombinator.com). En otras palabras, idealmente un sitio bien diseñado ya debería hacer accesible la información central, y un lector (humano o bot) debería encontrarla de forma natural. Si el contenido real del sitio fuera más simple o más textual (por ejemplo, a través de una "vista de lectura"), una IA podría no necesitar llms.txt. Esta crítica esencialmente dice: "Arregla el sitio web, no encubras sus defectos". También advierte que llms.txt es una especie de atajo que podría desincentivar la mejora del diseño subyacente del sitio (como meter contenido en una cita en bloque de SEO en lugar de crear una interfaz utilizable).

B. Alcance limitado (entrenamiento vs. inferencia): Es importante aclarar que llms.txt afecta principalmente el uso de sitios web por parte de la IA en el *momento de la inferencia*, no el entrenamiento inicial del modelo. Muchos propietarios de contenido quieren controlar cómo se utiliza su contenido para entrenar nuevos modelos (un debate legal y ético), pero llms.txt, tal como se especifica, no impone ni registra directamente los permisos de entrenamiento. Simplemente ayuda a un LLM a *obtener* contenido para responder consultas. Como argumenta Search Engine Land, las diferencias clave giran en torno a la indexación frente al uso: *"Robots.txt trata sobre la gestión del rastreo, mientras que la discusión sobre derechos de autor trata sobre cómo se utilizan los datos"* (Source: searchengineland.com). Los críticos podrían decir: si una empresa no quiere que su sitio aparezca en las salidas de IA en absoluto, llms.txt no detiene a nadie (solo guía). Por el contrario, si la empresa ya licencia contenido explícitamente (por ejemplo, con Creative Commons), llms.txt añade poco. El artículo GEO de Konstantinos Zoulas de 2023 sugiere que las licencias Creative Commons (CC0, CC-BY, etc.) podrían gobernar el uso de la IA más directamente que las directivas de robots o llms (Source: searchengineland.com). Esta visión implica que llms.txt resuelve solo el síntoma (descubrimiento de datos), no el núcleo del problema de los derechos de contenido.

C. Falta de estandarización y aplicación: Actualmente, `/llms.txt` es una propuesta voluntaria sin ningún RFC o registro formal. Como el propio Jeremy Howard admitió en Hacker News, no ha sido registrado bajo el registro URI .well-known de la IANA (un paso requerido para el estado de estándar oficial) (Source: news.ycombinator.com). Sin una decisión formal o un respaldo de toda la industria, no hay garantía de que el software lo busque de manera confiable. Los críticos señalan que incluso robots.txt no se aplica estrictamente —es una convención— y Google ha demostrado que puede ignorar "robots.txt" si es necesario por razones legales. Con llms.txt aún más en flujo, algunos argumentan que podría desvanecerse si los actores clave se mantienen al margen. (La postura de Google de ignorarlo ya puede haber mermado el entusiasmo.)

D. Potencial de uso indebido o manipulación: Como con cualquier señal similar al SEO, uno podría preocuparse por el spam o la "manipulación" de llms.txt. En principio, un sitio malicioso podría crear un llms.txt que contenga enlaces engañosos o maliciosos, o enterrar URLs de seguimiento o de anuncios. Sin embargo, debido a que llms.txt no inyecta automáticamente contenido en los datos de entrenamiento de la IA, este riesgo es limitado. Es más bien un riesgo de que un sitio sin escrúpulos pueda rellenar su llms.txt con enlaces irrelevantes solo para empujar a los usuarios (a través de respuestas de IA) hacia ellos. La especificación actual no especifica ninguna validación o limitación de velocidad. ¿Cómo sabría una herramienta de IA si un llms.txt es legítimo? Esta es una pregunta sin resolver. En la práctica, dado que el formato es legible por humanos y presumiblemente curado, el abuso flagrante probablemente sería detectado y desacreditado por la comunidad antes de que proliferara.

E. Impacto en el rendimiento de los sitios web: Otra preocupación (en su mayoría hipotética) es si el rastreo y la entrega de estos archivos de texto potencialmente grandes podrían sobrecargar los servidores web. Como se ha señalado, algunos archivos `llms.txt` alcanzan cientos de kilobytes o incluso megabytes, comparable a una pequeña página HTML. Si un sistema de IA los consulta con frecuencia (como cada 15 minutos, según indicó un registro (Source: searchengineland.com), esto podría imponer una carga no trivial. Los operadores de sitios deben ser conscientes de ello, aunque este problema es paralelo al concepto preexistente de sondeo de «sitemap.xml». Los servidores siempre podrían almacenar en caché y limitar la velocidad; es un detalle técnico, pero uno que debe ser implementado por los administradores web si `llms.txt` gana terreno.

F. Confusión sobre nombres y versiones: Existe cierta ambigüedad en la terminología: la propuesta original utiliza «`llms.txt`», pero muchas publicaciones (y artículos de LinkedIn) lo escriben como «LLMS.txt» (con mayúsculas o LLMs en plural). La comunidad generalmente se ha decantado por «llms.txt» (nombre de archivo en minúsculas). Además, diferentes herramientas hablan de `llms-full.txt` (que contiene el texto completo concatenado de las páginas) frente a `llms.txt` (que lista enlaces). Esto puede confundir a los recién llegados. La estandarización o la nomenclatura podrían evolucionar, pero por ahora esta confusión puede disuadir la adopción casual.

G. Enfoques alternativos (sin archivo nuevo): Finalmente, la crítica más fundamental: *¿Necesitamos siquiera un archivo nuevo?* Algunos expertos en SEO argumentan que los mismos objetivos podrían lograrse revitalizando ideas más antiguas. Por ejemplo, las primeras discusiones de OpenAI mencionaron el uso de «noindex» o «nofollow» en robots para diferenciar la búsqueda regular del uso de IA (Source: searchengineland.com). Otros proponen señales completamente en banda: por ejemplo, Google (a mediados de 2023) sugirió simplemente usar enlaces normales y prácticas de SEO para que la IA (como los propios Overviews de Google) encuentre contenido de forma natural (Source: searchengineland.com). También existe el concepto de una cabecera HTTP o un elemento que identifique un archivo o formato para los LLM, en lugar de un archivo de texto sin formato. Algunos comentaristas dicen que esto sería más semánticamente «web» que inventar otro tipo de archivo. Los defensores de llms.txt generalmente responden que nada impide usar múltiples enfoques (cabecera *y* llms.txt), pero esto sigue siendo un área de discusión.

En resumen, las críticas se centran en la **practicidad y la necesidad**: Si Google (y Bing) obtienen todo el contenido a través de métodos antiguos, llms.txt puede ser excesivo. Si los desarrolladores de IA pudieran simplemente raspar mejor el HTML o usar incrustaciones de índices de búsqueda existentes, quizás no lo necesiten estrictamente. Al mismo tiempo, los partidarios señalan que estos problemas no han disuadido los experimentos iniciales o la formación de estándares. Si estas preocupaciones resultan fatales o superables probablemente dependerá del uso concreto y del impulso de la comunidad.

Datos y Análisis

Un análisis exhaustivo de `/llms.txt` requiere no solo descripciones, sino también información basada en datos. Sin embargo, a mediados de 2025 el ecosistema aún es incipiente. A continuación, resumimos los datos disponibles y los hallazgos cuantitativos:

- **Adopción en los rankings de tráfico web:** El estudio de Rankability es uno de los pocos análisis de adopción reportados públicamente. Encuestó los 1.000 sitios web más visitados (a nivel global) a mediados de 2025 y encontró un uso del **0%** de `llms.txt` (solo ~0.3% según un recuento, redondeando a 0%) (Source: www.rankability.com). Desglosándolo por sector, reportó un 0.00% de adopción en cada categoría principal de la industria (comercio electrónico, redes sociales, finanzas, etc.), excepto un pequeño pico del 0.73% en Educación (Source: www.rankability.com). Esto sugiere que, entre los pesos pesados de la Web, prácticamente ninguno ha implementado `llms.txt`. En términos prácticos, si buscas en Google cualquier sitio grande (por ejemplo, Wikipedia, CNN, Amazon), no encontrarás ningún `llms.txt` a menos que alguien lo haya configurado explícitamente solo para probar. (Cabe destacar que la definición de «adopción» de Rankability probablemente requería una respuesta HTTP 200 para /llms.txt. Algunos sitios que devolvían 404 o un error se considerarían como no adopción.)

- **Adopción entre los sitios encuestados:** En contraste, un análisis diferente de un conjunto más amplio de 2.147 sitios web (el informe «LLMS Central») afirmó que el 86% de los sitios tenían **algún contenido llms.txt** (68% permitiendo el entrenamiento de IA de forma total o selectiva, y solo el 14% sin ninguno) (Source: llmscentral.com). Su metodología no es completamente transparente, pero agruparon las políticas de los sitios como «Permitir todo», «Selectivo», «Bloquear todo» o «Sin archivo». Ver una categoría como «Permitir todo» (23%) implica que estos sitios tienen un llms.txt que declara explícitamente permitir el uso de LAI. Si se toma al pie de la letra, este informe sugiere que **más de dos tercios de los sitios de tamaño mediano** de su muestra publicaron un llms.txt. También encuentra que las empresas tecnológicas son especialmente ávidas: el 95% de las empresas de tecnología/software que encuestaron tenían un archivo llms (Source: llmscentral.com), frente a porcentajes menores en otras industrias. Sin embargo, sin conocer su selección de muestra, esto puede reflejar un sesgo de autoselección (quizás rastrearon sitios que ya mencionaban la IA en sus blogs).

- **Tamaños y contenido de los archivos:** Al observar el contenido real de los archivos llms.txt, vemos una tremenda variación. El ejemplo de la Tabla 2 a continuación muestra algunos recuentos de tokens representativos para algunos sitios (del directorio llmstxt.site). Este tipo de números dan una idea de la escala. En particular, algunos sitios de documentación técnica dan como resultado archivos llms enormes: por ejemplo, *M-Source* (una empresa de bases de datos) tiene **328.716 tokens** listados (Source: llmstxt.site), y *LookupTheStars* tiene **385.221 tokens** (Source: llmstxt.site). (Para contextualizar, el límite de contexto de GPT-4 es de alrededor de 32k tokens, por lo que un solo llms.txt de 300k tokens necesitaría ser dividido en fragmentos.) Otros son más ligeros en tokens: el llms.txt de Ideanote.io tiene 1.106 tokens (Source: llmstxt.site), HoodChefs 44.494 tokens (Source: llmstxt.site), Framer 1.821, Klarna 17.387, etc. Un caso atípico extremo es *X-CMD*, cuyo archivo llms-full tiene 590.515 tokens (Source: llmstxt.site) (lo que implica un sitio colosal o posiblemente una peculiaridad de cómo se genera). La variabilidad indica que los sitios interpretan de manera diferente cuánto incluir.

- **Información sobre rastreo y tráfico:** Hay pocos datos públicos sobre el tráfico. Una tabla del sitio de informes de SEO [33] destaca que las solicitudes de Googlebot para llms.txt ocurren *cero* veces («Google no rastreará tu LLMS.txt» (Source: searchengineland.com). Por el contrario, el usuario Ray Martinez informó en los registros de su sitio que «OpenAI rastrea mi archivo LLMs.txt en algunos sitios… haciendo ping a nuestros servidores cada 15 minutos en busca de novedades» (Source: searchengineland.com). Este análisis de registros sugiere que, al menos para sus sitios, **los sistemas de OpenAI están revisando activamente llms.txt con frecuencia** (quizás asumiendo que deberían hacerlo). John Mueller de Google dijo de manera similar en un hangout anterior de Search Console que «ningún sistema de IA está utilizando actualmente el archivo LLMS.txt» (Source: searchengineland.com) (cita de seroundtable). En resumen, la única información empírica que tenemos es anecdótica: la búsqueda de Google lo ignora, algunos laboratorios de IA lo consultan.

- **Correlación con el rendimiento SEO:** No existen datos agregados creíbles que vinculen llms.txt con una mejora en el ranking de búsqueda o el tráfico. Google dice explícitamente que el SEO normal es adecuado (Source: searchengineland.com), lo que implica que no encontraron ninguna ventaja. Queda por ver si, por ejemplo, la inclusión de llms.txt afectará positivamente a los fragmentos o «respuestas» en las interfaces de chat de IA. En principio, si un asistente de IA cita directamente el contenido de llms.txt, un comercializador astuto intentará detectarlo y optimizarlo en consecuencia. Pero a mediados de 2025, esto sigue siendo hipotético.

- **Soporte de herramientas LLM:** Más allá de Google, productos LLM notables han comenzado a reconocer llms.txt. La documentación de *Anthropic* (Claude) lo incluye; el MCP (plugin de múltiples contextos) de *LangChain* admite la lectura de llms.txt desde IDEs (Source: github.com). Algunos frameworks de chatbot de código abierto basados en LLM ahora tienen código estándar para buscar llms.txt. La mera existencia de un repositorio de GitHub (AnswerDotAI/llms-txt) y pruebas CI automatizadas indica interés por parte de los desarrolladores. Por otro lado, plataformas importantes como ChatGPT (front-end

de OpenAI) no han anunciado soporte formal, aparte de la indexación en el backend. Informes de analistas de Distinctly (noticias de SEO) han señalado una captura de pantalla de ChatGPT extrayendo contenido de un «llms-full.txt» (Source: distinctly.co), pero faltan detalles y esto podría ser un caso aislado.

Estos puntos de datos pintan el panorama: **emergente pero menor**. Decenas o cientos de sitios más pequeños tienen llms.txt, pero no una masa crítica. Si la adopción se graficara a lo largo del tiempo, podríamos ver un lento aumento entre los sitios de nivel medio a finales de 2024 y durante 2025, estabilizándose. Un punto de inflexión probablemente requeriría que una o más plataformas de IA dominantes declararan «sí, usamos llms.txt». De lo contrario, podría seguir siendo una práctica recomendada de nicho.

A continuación, se presenta una tabla que resume algunas estadísticas y ejemplos de adopción:

MÉTRICA / CATEGORÍA DE SITIO	VALOR / EJEMPLOS	FUENTE
Sitios del Top-1000 que usan llms.txt	~0% (0.3%)	(Source: www.rankability.com)
Empresas de tecnología/software (encuestadas)	95% (sitios en esas categorías tienen políticas de llms en un informe)	(Source: llmscentral.com) (Source: llmscentral.com)
Permitir todo (todo el contenido abierto)	23% de los sitios (según un informe)	(Source: llmscentral.com)
Políticas selectivas (algunas páginas)	45% de los sitios	(Source: llmscentral.com)
Bloquear todo (no se permite el uso de IA)	18% de los sitios	(Source: llmscentral.com)
Sin archivo `llms.txt`	14% de los sitios	(Source: llmscentral.com)
Sitios de ejemplo con llms.txt	Framer.com (1.821 tokens), Klarna docs (17.387), M-Source (328.716) (Source: llmstxt.site) (Source: llmstxt.site)	(Source: llmstxt.site) (Source: llmstxt.site)
Tamaño de llms más grande reportado	~385.221 tokens (lookupthestars.com)	(Source: llmstxt.site)
Frecuencia de rastreo de OpenAI	~cada 15 minutos (registro del sitio)	(Source: searchengineland.com)
Solicitudes de Googlebot para llms.txt	Ninguna reportada; Google dice que no rastreará llms.txt	(Source: searchengineland.com)

Tabla: Cifras seleccionadas relacionadas con la adopción y el uso de llms.txt.

Perspectivas y Opiniones de Expertos

Para comprender completamente lo que está en juego con `/llms.txt`, consideramos lo que varios expertos y partes interesadas han dicho —a veces en voz alta— sobre la propuesta.

- **Jeremy Howard (Answer.AI, fast.ai):** *Proponente y autor de la propuesta de llms.txt.* Argumenta principalmente desde la perspectiva de la usabilidad para desarrolladores. En los hilos de discusión, Howard enfatizó que el objetivo es ayudar a «los usuarios finales a usar la información de los sitios web con la ayuda de la IA» (Source: news.ycombinator.com). Dio ejemplos concretos: cuando lanzó la biblioteca FastHTML, muchos usuarios potenciales se quejaron de que las herramientas de IA (cursor, etc.) no podían responder preguntas al respecto porque los modelos eran posteriores a sus conocimientos. Su solución: curar manualmente la documentación una vez en un llms.txt para que las herramientas de IA la tengan disponible fácilmente en el momento de la inferencia. Howard enmarca llms.txt como una ayuda para el usuario final/la comunidad en lugar de una preocupación por el scraping: «llms.txt no está realmente diseñado para ayudar con el scraping; está diseñado para ayudar a los usuarios finales a usar la información de los sitios web con la ayuda de la IA» (Source: news.ycombinator.com). También subraya que proporcionar llms.txt ahorra esfuerzo a *todos*: en lugar de que cada ingeniero elija individualmente el contexto para los prompts, el propietario del sitio lo hace una vez. En entrevistas y publicaciones de blog, menciona con frecuencia casos de uso para documentación de desarrolladores, y el hecho de que muchos documentos de fast.ai/nbdev ahora auto-generan markdown para satisfacer esta necesidad (Source: github.com).

- **Analistas de SEO/Marketing (ej. SearchEngineLand, Agencias de SEO expectantes):** En general, las publicaciones de SEO han adoptado una visión cautelosamente optimista. El artículo de SEL de marzo de 2025 de Roger Montti examinó llms.txt y señaló tanto a «creadores de contenido interesados» como a «detractores» (Source: searchengineland.com). La postura de Montti es neutral-a-curiosa; explica la especificación y sugiere que «aumenta el control por parte del propietario» . Roger destaca el ángulo de ahorro de recursos (los LLM se centran en la inteligencia, no en el rastreo). Mientras tanto, otros en la comunidad SEO promocionan llms.txt como algo imprescindible para las marcas. Por ejemplo, el artículo de Radu Stoian en LinkedIn lo titula sin rodeos «no negociable para tu marca» (Source: www.linkedin.com). Dichos artículos prometen una narrativa de marca mejorada e incluso afirman que Google está indexando llms.txt ahora. Sin embargo, al ser un blog no verificado, estos deben leerse con escepticismo. Voces más mesuradas (fuera de SEL) sugieren que llms.txt es una técnica incremental de «SEO de IA»: una posible optimización pero poco probable que supere al SEO tradicional (Source: llms-txt.io).

- **Ingenieros de Búsqueda de Google:** Las declaraciones más claras han venido de Google mismo, aunque indirectamente. En un evento de Google Search Central en julio de 2025, Gary Illyes (Analista de Búsqueda) lo dejó explícito: *«Para que tu contenido aparezca en AI Overview, simplemente usa prácticas normales de SEO... También dijo que Google no rastreará el archivo LLMS.txt.»* (Source: searchengineland.com). En efecto, el mensaje de Google es: *Ignora llms.txt en términos de ranking de búsqueda – no lo usamos.* Esto fue secundado por John Mueller, quien dijo en un Webmaster Hangout que «ningún sistema de IA está utilizando actualmente el archivo LLMS.txt» (Source: searchengineland.com). Estas afirmaciones significan que, desde la perspectiva de Google, llms.txt no tiene ninguna influencia en el SEO. Esto puede desanimar a los editores que se preocupan principalmente por la «Googleabilidad». También plantea una pregunta más importante: incluso si llms.txt es beneficioso para el encuentro de tu contenido con *alguna* IA, si esa IA no es la que domina las búsquedas (Búsqueda de Google), el impacto en el tráfico real podría ser pequeño.

- **OpenAI (desarrolladores de ChatGPT):** OpenAI no ha comentado públicamente sobre llms.txt, pero la evidencia limitada sugiere que al menos lo han *probado* o permitido su uso. El análisis de registros de Ray Martinez es una prueba irrefutable de que parte de la infraestructura de OpenAI está consultando llms.txt en busca de cambios (Source: searchengineland.com). Esto sugiere que los agentes de OpenAI han reconocido llms.txt en la práctica y lo tratan como un «punto final de frescura». Sin embargo, los portavoces de OpenAI no han anunciado ninguna postura política. De forma anecdótica, los usuarios de herramientas como el plugin «Navegar con Bing» de ChatGPT o agentes de terceros intentan aprovechar llms.txt, pero no hay documentación oficial disponible.

- **Anthropic (desarrolladores de Claude):** Se cree ampliamente que Anthropic es compatible con llms.txt. Su equipo de documentación lo añadió tempranamente, y los ingenieros de Anthropic han mostrado interés en la estandarización. Claude Projects (el plugin IDE de código de Claude) trata a llms.txt como un ciudadano de primera clase: los usuarios que cargan una base de código pueden especificar un llms.txt. Un fragmento de la comunidad en GitHub muestra instrucciones para configurar Claude Desktop/Cursor para leer llms.txt (Source: gist.github.com), lo que implica un soporte integrado. Kohl Marcus (en Distinctly news) mencionó que “Aimee Jurenka muestra a ChatGPT accediendo a contenido de un archivo llms-full.txt” (Source: distinctly.co), por lo que presumiblemente fue a través de los frameworks de Anthropic. Todo esto indica que al menos los productos de IA avanzados (como Claude) se están tomando llms.txt en serio.

- **Expertos Académicos y en Privacidad:** Las organizaciones preocupadas por la privacidad de los datos señalan que llms.txt aborda la narrativa del scraping. Privacy International, en una explicación sobre los LLM, subraya que *"cuanto más lenguaje escrito [los LLM] puedan obtener, mejor"* y que el web scraping suele ser "indiscriminado" (Source: privacyinternational.org). Aunque no mencionan llms.txt específicamente, la implicación es que cualquier cosa que haga que el scraping sea más dirigido (es decir, guiado por los propietarios) podría alinearse con la gobernanza de datos. Ninguna ley de privacidad formal reconoce llms.txt, pero defensores como Jay Graber (CEO de Bluesky) que lideran los debates sobre los derechos de los creadores de IA han señalado que llms.txt y otras iniciativas (como la "Declaración de Bletchley") forman parte de las normas emergentes para el control de datos en la IA. En resumen, algunos ven llms.txt como un gesto constructivo hacia el respeto de la propiedad del contenido, incluso si no es vinculante.

- **Críticos y Pragmatistas:** Muchos programadores y expertos en SEO abordaron llms.txt de manera pragmática. En Hacker News y blogs, los comentaristas expresaron escepticismo: uno señaló que si la UX de un sitio es buena, una "página de instrucciones" podría ser suficiente y llms.txt sería innecesario (Source: news.ycombinator.com). Otros dijeron que mantener un archivo adicional es una sobrecarga; preferirían depender de `rel=search` o de enfoques basados en API. Desde una perspectiva de estándares, un comentarista señaló que quizás una etiqueta `<link rel="llm">` o la negociación de tipo de contenido HTTP podrían ser más elegantes que un archivo de texto (Source: news.ycombinator.com). Estas sugerencias reflejan un deseo de soluciones que se integren sin problemas en la arquitectura web existente, en lugar de añadir un silo paralelo.

A pesar de estas opiniones encontradas, **el hilo conductor** es: `llms.txt` plantea la pregunta "¿Debería la Web adaptarse a la IA?". Muchas voces entrevistadas se enorgullecen de ser los primeros en adoptar. Los defensores argumentan que permite a los sitios web *unirse a la conversación* en lugar de ser minas de datos pasivas (Source: www.linkedin.com), mientras que los detractores dicen que interrumpe la interfaz uniforme de la Web. En última instancia, la mayoría lo ve como un *experimento*: una idea que vale la pena probar ahora, con la retroalimentación de la comunidad guiando si se convierte en un estándar de facto o se desvanece.

Consideraciones de Implementación y Herramientas

Para un propietario de sitio web que considere añadir `llms.txt`, surgen preguntas prácticas: *¿Cómo crearlo? ¿Qué contenido incluir? ¿Cómo mantenerlo?* Afortunadamente, han surgido varias herramientas y guías para abordar estas cuestiones.

- **Guías y Ejemplos:** El sitio de la comunidad llmstxt (llmstxt.org) presenta ejemplos de archivos llms.txt y una guía paso a paso. También hay numerosos artículos de blog y repositorios de GitHub con implementaciones de ejemplo de llms.txt. Los consejos clave incluyen: empezar con la página de inicio/título, escribir un resumen conciso (de 1 a 3 frases) en una cita en bloque, y luego listar las páginas cruciales. Algunos blogs de SEO recomiendan añadir información de la empresa (contacto, dirección), preguntas frecuentes, documentación para desarrolladores, páginas de productos – básicamente todo lo que una IA útil podría necesitar para responder a las consultas de los usuarios (Source: llmsly.com) (Source: golevels.com). A menudo se sugiere mantener el archivo por debajo de unos pocos megabytes; una publicación mencionó que los archivos llms.txt pueden variar desde unos pocos KB hasta cientos de KB (Source: searchengineland.com). El formato es flexible: se pueden usar imágenes (como enlaces), viñetas o párrafos cortos. Algunos sitios incluso dividen el contenido de llms en varios archivos: la variante *llms-full.txt* puede contener secciones completas de texto si es necesario.

- **Herramientas Existentes:** Varias herramientas de código abierto ayudan a generar o validar llms.txt. Por ejemplo:

  - **Generador de llms.txt (llmstxtgen.com):** Una aplicación web donde pegas tu sitemap o lista de URLs; rastrea y genera un borrador de llms.txt en segundos. La captura de pantalla [10] muestra la salida autogenerada de una herramienta (para anthropic.com).
  - **Utilidades CLI:** El repositorio de GitHub (AnswerDotAI/llms-txt) incluye scripts como `llms_txt2ctx` que pueden combinar llms.txt y markdown enlazado en un archivo de contexto consumible por máquinas (Source: github.com). Otros (como la herramienta de Firecrawl referenciada en [66]) pueden rastrear y ensamblar contenido en listas de marcado.
  - **Plugins de CMS:** Existen plugins para WordPress y otros CMS que generan llms.txt a partir de menús o publicaciones del sitio (como se insinúa en [59]). Estos permiten actualizaciones dinámicas a medida que cambia el contenido del sitio.
  - **Integraciones IDE/LLM:** Herramientas como `mcpdoc` de LangChain pueden extraer un llms.txt automáticamente al configurar la IA, para que los desarrolladores no tengan que recuperarlo manualmente (Source: github.com). Esto demuestra que los frameworks de LLM están empezando a reconocer el archivo.

- **Mantenimiento:** Dado que los sitios cambian, llms.txt necesita actualizaciones. A diferencia de sitemap.xml (que puede automatizarse), llms.txt se cura más manualmente. Sin embargo, algunos flujos de trabajo lo crean a partir de datos existentes del sitio: por ejemplo, un script puede escanear menús de navegación para listar URLs, o compilar archivos README. El proyecto de documentación de Ethereum, por ejemplo, utiliza un proceso de CI para reconstruir llms.md cada vez que cambian los documentos (como parte de su generación de sitio estático). En general, se recomienda revisar llms.txt cada vez que haya cambios importantes en el contenido del sitio, ya que los enlaces o resúmenes obsoletos podrían engañar a la IA. La monitorización implica simplemente verificar la disponibilidad de ese único archivo (por ejemplo, comprobaciones de salud del sitio).
- **Alojamiento y Rendimiento:** Como con cualquier activo estático, la mejor práctica es servir llms.txt con caché habilitada (control de caché HTTP) y compresión gzip, ya que suele ser texto. Los archivos llms.txt grandes (cientos de KB) pueden sobrecargar el ancho de banda si se rastrean con demasiada frecuencia, por lo que un almacenamiento en caché adecuado ayuda. Algunos han sugerido alojar llms.txt en una CDN o exponerlo a través de .well-known/llms.txt para que los proxies puedan almacenarlo en caché globalmente.

Estudios de Caso en Profundidad

FastHTML (Framework de Hipermedia): La experiencia del proyecto FastHTML es ilustrativa. FastHTML es una pequeña biblioteca para crear APIs y documentación. Sus desarrolladores reconocieron que los modelos de lenguaje típicos (como Claude) no tenían conocimiento de FastHTML (fue lanzado después de su fecha límite de entrenamiento). Para compensar, crearon un llms.txt para su sitio de documentación. Luego, usando `llms_txt2ctx`, generaron dos versiones de archivos de contexto: `llms-ctx.txt` (contenido principal) y `llms-ctx-full.txt` (extendido con enlaces opcionales) (Source: github.com). Esto les permitió proporcionar a Claude una vista concisa pero completa de la documentación cada vez que respondía preguntas. El resultado: informaron de respuestas asistidas por IA dramáticamente mejores en su IDE y bots de documentación, sin que cada usuario tuviera que copiar enlaces manualmente. Esto demuestra que `llms.txt` sirve a la “cola larga” del contenido (la documentación de FastHTML no fue indexada por Google, según [4]). Su caso muestra cómo un proyecto modesto puede aprovechar llms.txt para hacerse “buscable por IA” desde el primer día.

Anthropic (Empresa de IA): La adopción de llms.txt por parte de Anthropic es más simbólica que específica de un caso. Como una importante empresa de IA, se podría argumentar que tienen menos necesidad de ser "encontrables por IA", pero aun así han creado llms.txt por transparencia y para enviar una señal a la comunidad. Su llms.txt lista introducciones a sus productos (Claude), documentos de investigación, canales para desarrolladores y más (la salida [10] muestra páginas como “Claude en Slack”, “API”, “Clientes”). Su participación otorga credibilidad: que un líder de la industria incluya llms.txt sugiere que vale la pena tomarlo en serio. También es probable que retroalimente los propios modelos de Anthropic (si lo indexan internamente).

Institución Académica (ejemplo): Algunas universidades tienen grandes sitios web con catálogos de cursos, investigación, etc. Un ejemplo es “Juris Education”, que tiene un llms.txt considerable listado (22.885 tokens) (Source: llmstxt.site). La razón puede ser ayudar a los futuros estudiantes o a los tutores/chatbots de IA a recopilar información del curso rápidamente. Muchas universidades experimentaron con portales de IA para preguntas y respuestas de estudiantes, y llms.txt podría servir como recurso de backend.

Gobierno y Regulaciones: Hasta ahora, no parece haber directrices gubernamentales oficiales sobre llms.txt. Sin embargo, resuena con los debates políticos. Por ejemplo, el artículo de la Directiva de Derechos de Autor de la UE sobre la minería de texto y datos establece excepciones para la investigación, lo que implica que los sitios web no necesitarían optar explícitamente por ese uso si está dentro del alcance. LLms.txt se encuentra en un área gris: es metadatos voluntarios para el uso de datos de IA, no una licencia vinculante. Algunos legisladores abogan por mecanismos más aplicables (por ejemplo, leyes sobre bots de web scraping). Ningún gobierno conocido ha exigido algo como llms.txt.

Implicaciones y Direcciones Futuras

De cara al futuro, el éxito o fracaso de `llms.txt` probablemente dependerá de algunos factores clave:

- **Adopción por Plataformas de IA:** Si los principales modelos o herramientas de IA llegan a reconocer y confiar en llms.txt, su adopción podría dispararse. Por ejemplo, si OpenAI lo apoyara oficialmente (por ejemplo, a través de ChatGPT instruyendo a GPT sobre un enlace llms.txt), o si Google cambiara de rumbo e indexara llms.txt, eso crearía un fuerte incentivo. Por el contrario, si los desarrolladores de IA prefieren depender de índices de búsqueda o embeddings (como Bing Chat ya utiliza los resultados de búsqueda internamente), la demanda de llms.txt podría seguir siendo limitada. El hecho de que Google lo

desestime actualmente sugiere que la “búsqueda de IA” convencional tardará en adoptarlo. Pero el panorama puede cambiar rápidamente: la última vez que comprobamos (junio de 2025), Google dijo que el SEO normal era suficiente (Source: searchengineland.com), pero un año después eso podría cambiar si el comportamiento del usuario se desplaza hacia los resúmenes de IA.

- **Ecosistema de Herramientas y Frameworks:** El crecimiento de herramientas para desarrolladores en torno a llms.txt podría facilitar su adopción. Por ejemplo, si GitHub Pages genera automáticamente llms.txt, o si WordPress y otros CMS lo incluyen por defecto, una avalancha de nuevos sitios podrían estar “listos para llms.txt” de la noche a la mañana. Ya hemos visto los comienzos: existe un plugin de WordPress, algunos generadores de sitios estáticos tienen complementos. Si los principales sistemas de gestión de contenido integran soporte, la adopción podría aumentar independientemente de los grandes actores de búsqueda.

- **Estandarización:** Pasar de propuesta a estándar normalmente requiere consenso y registro. Los autores insinuaron la posibilidad de registrarlo como una URI conocida (por ejemplo, `/.well-known/llms.txt`) si el estándar se afianza (Source: news.ycombinator.com). Tal movimiento facilitaría la orientación para los bots. Además, la publicación de un RFC o una nota del W3C podría consolidar el formato. Si llms.txt obtiene un respaldo formal, eso podría señalar un “estado oficial”, fomentando una adopción más amplia (al igual que RSS se volvió ubicuo una vez estandarizado).

- **Enfoques Alternativos:** Es posible que surjan soluciones mejores. Por ejemplo, Google podría desarrollar su propio “sitemap de IA” o metaetiquetas para controlar la indexación de IA, dejando obsoleto a llms.txt. O los asistentes de IA podrían usar señales contextuales (marcado schema.org, datos de Knowledge Graph, esquemas de asistentes de voz) para obtener información de manera más semántica. Existe una discusión en curso sobre estándares como las características de SERP o las “pistas de prompt de IA” incrustadas en HTML. En el peor de los casos, llms.txt podría convertirse en *una* de muchas propuestas similares, y quizás ser reemplazado por un protocolo más elegante.

- **Influencia Regulatoria:** Si los reguladores exigen a las empresas de IA que respeten robots.txt (como parte de la regulación de los scrapers), una extensión lógica podría ser exigir el respeto de las directivas de llms.txt. Esto podría ocurrir a través de la autorregulación de la industria o la ley, especialmente a medida que se intensifican los debates sobre los datos de entrenamiento de IA y los derechos de autor. Por ejemplo, si la UE o un país legislara que los sistemas de IA deben respetar las preferencias de uso de contenido publicadas por los propietarios de sitios web, podrían mencionar explícitamente llms.txt como un canal reconocido. Esto es especulativo pero dentro del ámbito de la gobernanza emergente de la IA.

- **Efectos de Red en el Descubrimiento de Contenido:** Estamos solo en las primeras etapas del “descubrimiento de contenido impulsado por IA”. Si uno o dos asistentes de IA populares comienzan a usar las listas de llms.txt por defecto, los usuarios podrían empezar a verlo indirectamente. Por ejemplo, si las respuestas de Gemini o Claude citan regularmente contenido de una página llms.txt, los equipos de contenido inteligentes lo notarán y optimizarán sus archivos. Esto es similar a cómo reaccionaron los SEO cuando los fragmentos destacados comenzaron a extraerse de estructuras HTML particulares (luego modificaron su contenido para alimentar los fragmentos). Con el tiempo, una buena práctica de llms.txt podría generar beneficios parciales de AI-SEO no capturados en las métricas tradicionales.

- **Mejores Prácticas de la Comunidad:** El ecosistema de `llms.txt` evolucionará a través de la experiencia compartida. A medida que los primeros adoptantes publiquen sus experiencias, se desarrollarán las mejores prácticas de la comunidad. Los recursos de GitHub y los blogs ya están documentando lo que se debe y no se debe hacer (por ejemplo, sugerencias sobre cómo estructurar las citas en bloque para que no confundan a un LLM). En los próximos meses, esperamos que aparezcan herramientas de linting para llms.txt (que verifiquen enlaces rotos, claridad, etc.). También podrían surgir convenciones de versionado (así como robots.txt no tiene una versión oficial, llms.txt podría fijar la especificación o permitir variaciones como llms-full.txt).

En conclusión, **el futuro de `llms.txt` es incierto**. Muchos observadores han señalado que ninguna tecnología por sí sola puede garantizar cómo evolucionará la IA, ya sea que el sector del “comportamiento del contenido” se consolide en torno a los editores (como llms.txt) o permanezca descentralizado. Por ahora, llms.txt se encuentra en un rincón de nicho pero activo de la web de la IA. Si tiene éxito, podría dar lugar a una nueva capa de estándares de archivos web; si no, podría retirarse silenciosamente como un experimento interesante.

Conclusión

Nuestra investigación de `/llms.txt` encuentra que es una propuesta bien definida con objetivos específicos: hacer que los sitios web sean más accesibles para los grandes modelos de lenguaje mediante un mapa de contenido creado por humanos. Las especificaciones técnicas (usando Markdown, listas de enlaces, etc.) son claras y relativamente fáciles de implementar. Estudios de caso tempranos en documentación de software han demostrado que llms.txt *puede* mejorar el rendimiento de los agentes de IA en tareas de nicho (Source: searchengineland.com) (Source: www.released.so). Sin embargo, al mismo tiempo, existe una medida igual de escepticismo. Los principales motores de búsqueda han proclamado públicamente hasta ahora que ignorarán este archivo (Source: searchengineland.com), y el escaneo empírico sugiere que los sitios principales aún no lo han adoptado de manera apreciable (Source: www.rankability.com).

¿Importa? Por ahora, la respuesta es: *Depende de tus prioridades*. Si eres un editor de tecnología, desarrollador o especialista en marketing con conocimientos de SEO que quiere experimentar con cada optimización de vanguardia en la era de la IA, `llms.txt` parece digno de explorar. Impone un costo relativamente bajo, es reversible, y si las herramientas de IA comienzan a soportarlo ampliamente, te habrás adelantado a la curva. Importa particularmente para dominios donde las preguntas y respuestas impulsadas por IA pueden impulsar el soporte técnico o la incorporación de usuarios: documentación para desarrolladores, APIs, manuales de productos, etc.

Sin embargo, si te centras únicamente en la búsqueda tradicional o tienes recursos limitados, entonces `llms.txt` puede considerarse opcional. El consenso del equipo de SEO de Google es que el "SEO normal" cubre la aparición en los resultados de IA (Source: searchengineland.com). Las organizaciones desinteresadas en el entrenamiento de IA de sus datos (o que se oponen a ello) podrían preferir mecanismos legales más concretos (licencias, bloqueos de robots) en lugar de una lista amigable. Como implicaba el informe de LLMS Central, muchos propietarios de contenido ven `llms.txt` como parte de la **transparencia del entrenamiento de IA**, pero si una IA realmente lo respeta (o lo compensa) sigue siendo en gran medida no probado.

De cara al futuro, el efecto más inmediato de llms.txt es iniciar conversaciones vitales entre los webmasters sobre el diseño de contenido para la IA. Al probar esta nueva herramienta, la comunidad puede descubrir dónde los LLM tienen éxito o fallan al digerir sitios reales. Informa a ambas partes (desarrolladores de sitios y de IA) sobre lo que funciona. En ese sentido, llms.txt ya ha tenido cierto impacto: hizo que los entrenadores de IA fueran conscientes de los problemas de la ventana de contexto, y que los expertos en SEO fueran conscientes de que los motores de búsqueda aún no son agentes de IA, etc.

En última instancia, la narrativa en torno a llms.txt se hace eco de discusiones más amplias sobre el futuro de la Web: ¿Ejercerán los creadores de contenido un control explícito sobre el uso de sus datos por parte de la IA, o la Web seguirá siendo un corpus de texto pasivo? ¿Veremos una "web de IA" con nuevos mini-estándares superpuestos a HTML (como ahora existen las convenciones AJAX y JSON), o la IA simplemente se superpondrá a la infraestructura existente (anotaciones semánticas, rastreo mejorado)? El veredicto aún no se ha emitido.

Lo que está claro es que **`llms.txt` importa en la medida en que la industria y la comunidad decidan que lo hace**. Si se ve como análogo a cómo `robots.txt` y `sitemap.xml` ganaron tracción, entonces su importancia crecerá tan pronto como suficiente contenido y suficientes sistemas de IA converjan en él. Todavía es pronto, y por cada beneficio técnico interno que se reclama, existen preocupaciones igualmente importantes sobre su necesidad y viabilidad.

En nuestra opinión, `llms.txt` es un experimento proactivo y constructivo: busca anticipar la desinformación relacionada con la IA en la web. Nuestra investigación sugiere que es una solución bien intencionada que aborda desafíos técnicos reales (Source: searchengineland.com) (Source: www.released.so). Su éxito futuro dependerá tanto de la adopción técnica (por parte de las plataformas de IA) como de la adopción por parte de la comunidad (por parte de los propietarios de sitios). Apoyamos su exploración continua; después de todo, un enfoque con desventajas insignificantes combinado con una pequeña ventaja en la fidelidad de la IA parece una apuesta que vale la pena. Si se convierte en parte del conjunto de herramientas estándar para internet, o solo en una nota a pie de página en la historia de la evolución de la Web, solo el tiempo (y los datos) lo dirán.

Referencias: Todas las afirmaciones y cifras anteriores se extraen de las fuentes citadas en el texto (Source: searchengineland.com) (Source: searchengineland.com) (Source: llmscentral.com) (Source: www.rankability.com) (Source: llms-txt.io) (Source: www.released.so) (Source: llmstxt.site) (Source: www.kdjingpai.com), y de informes adicionales de la industria y comentarios de expertos según se detalla. Cada cita identifica la fuente de la información descrita.

Tags: llms-txt, modelos-lenguaje-grandes, ia-generativa, optimizacion-ia, seo, robots-txt, estandares-web, optimizacion-contenido

DISCLAIMER

This document is provided for informational purposes only. No representations or warranties are made regarding the accuracy, completeness, or reliability of its contents. Any use of this information is at your own risk. Unknown shall not be liable for any damages arising from the use of this document. This content may include material generated with assistance from artificial intelligence tools, which may contain errors or inaccuracies. Readers should verify critical information independently. All product names, trademarks, and registered trademarks mentioned are property of their respective owners and are used for identification purposes only. Use of these names does not imply endorsement. This document does not constitute professional or legal advice. For specific guidance related to your needs, please consult qualified professionals.