Googleがスパム防止ガイドを更新、AIツールの生成ページも対象か？

Google は、ユーザー生成スパムの防止ガイドのドキュメントを更新した。

Google は、ユーザー生成スパムの防止ガイドのドキュメントを更新しました。
リンクで共有されたコンテンツのインデックス登録をブロックする方法と、Search Console で不正行為を監視する方法に関するガイダンスが追加されています。

ドキュメントを更新した理由は、検索向けではないユーザー作成ページがインデックス登録されるのを防ぎ、スパムを早期に検出できるように、プラットフォームを支援するためとのことです。

更新内容

更新内容を説明します。

ツールも対象

まず、冒頭の段落に記述が追加されました。

更新前：

Spammers often take advantage of open comment forms and other user generated content inputs and generate spammy content on an unsuspecting victim site. Hosting platforms may be similarly open to abuse; spammers may create a large number of sites that violate our spam policies and add little or no value to the web.

スパマーは、公開コメント フォームやその他のユーザー作成コンテンツの入力欄を利用して、無防備なサイトでスパム コンテンツを生成することがよくあります。ホスティング プラットフォームでも同様に不正行為が起きる可能性があります。スパマーによって、Google のスパムに関するポリシーに違反し、ウェブにほとんどまたはまったく価値をもたらさないサイトが多数作成されることもあります。

更新後：

Spammers often take advantage of open comment forms and other user generated content inputs and generate spammy content on an unsuspecting victim site. Hosting platforms **and tools that let users publish shared pages, generated content, or web apps** may be similarly open to abuse; spammers may create a large number of **pages** or sites that violate our spam policies and add little or no value to the web.

Googleがスパム防止ガイドを更新、AIツールの生成ページも対象か？

Google は、ユーザー生成スパムの防止ガイドのドキュメントを更新した。

スパマーは、公開コメント フォームやその他のユーザー作成コンテンツの入力欄を利用して、無防備なサイトでスパム コンテンツを生成することがよくあります。**ユーザーが共有ページや生成コンテンツ、ウェブアプリを公開できる**ホスティングプラットフォームや**ツールも**、同様に悪用される可能性があります。スパマーによって、Google のスパムに関するポリシーに違反し、ウェブにほとんどまたはまったく価値をもたらさないページやサイトが多数作成されることもあります。

ℹ️この記事を書いている時点では日本語ドキュメントは未更新のため、翻訳は僕による
強調した記述が追加されました。

プラットフォームだけではなくツールも対象であることが明記されました。
具体的なツール名には言及していませんが、普及してきた AI 関連のツールもきっと含まれているはずです。

検索向けのコンテンツのみをインデックス登録する
「検索向けのコンテンツのみをインデックス登録する」セクションが追加されました。

Only index content meant for Search

If your service lets users create shareable pages, generated outputs, or interactive tools, consider whether that content is meant to be found through search engines or only accessed directly by people with a link.

When pages are created for direct sharing, collaboration, or utility rather than public search discovery, apply the noindex rule (using a robots meta tag or X-Robots-Tag HTTP response header) or block crawling in robots.txt by default.

検索向けのコンテンツのみをインデックス登録する

サービスでユーザーが共有可能なページ、生成されたコンテンツ、インタラクティブなツールを作成できる場合は、そのコンテンツを検索エンジン経由で見つけてもらうことを意図しているのか、それともリンクを持っている人が直接アクセスすることだけを意図しているのかを検討します。

ページが公開検索での発見ではなく、直接共有、共同作業、またはツールとしての利用を目的に作成されている場合は、デフォルトで noindex ルールを適用します（robots メタタグまたは X-Robots-Tag HTTP レスポンスヘッダーを使用する）か、robots.txt でクロールをブロックします。

https://www.suzukikenichi.com/blog/google-updates-spam-prevention-guide-are-ai-generated-pages-also-covered/

Googleがスパム防止ガイドを更新、AIツールの生成ページも対象か？

Google は、ユーザー生成スパムの防止ガイドのドキュメントを更新した。

追加のモニタリング方法

サービスの不正使用をモニタリングする方法に次の 2 項目が追加されました。

- Verify your site or subdomain in Search Console so you receive email alerts if Google issues a manual action against your site or detects security issues. Also check the Page Indexing report and Search performance report regularly for unexpected spikes in indexed pages, impressions, or queries that don't match your platform's purpose.
- Track platform-level creation and indexing metrics, and set up alerts for anomalies such as spikes in newly published public pages or a small number of users generating a large share of indexable URLs.

- サイトまたはサブドメインを Search Console で所有権確認すると、Google がサイトに手動による対策を実施した場合やセキュリティの問題を検出した場合に、メールで通知を受け取れるようになります。また、ページのインデックス登録レポートと検索パフォーマンスレポートを定期的に確認し、インデックス登録済みページ数や表示回数の予期しない急増、プラットフォームの目的に合わない検索クエリがないか調べます。
- プラットフォーム全体でコンテンツの作成やインデックス登録に関する指標を追跡し、新たに公開されたページの急増や、少数のユーザーがインデックス登録可能な URL の大部分を生成しているといった異常について、アラートを設定します。

◇ ◇ ◇

共有ページや生成コンテンツを公開できるサービスの運営者は、検索向けではないページをデフォルトで noindex にしておくのが無難です。

Marie Haynes さんは X の投稿で、AI ツールやアプリ構築プラットフォームに対する手動対策が出てくる可能性を指摘しています。

Google は具体的なツール名を挙げていないので実際にそうなるかは現時点では分かりませんが、可能性はありそうです。

ChatGPTの使い方は世代で違う？ 日本の年齢別データから見えること

OpenAI は、個人が ChatGPT をどう使っているかを示すインタラクティブなデータダッシュボードを更新した。日本を含む国ごとに、年齢層や話題別の利用状況を確認できる。

OpenAI は 2026 年 10 月 26 日、個人が ChatGPT をどう使っているかを示すインタラクティブなデータダッシュボード「Signals individual data」を更新しました。
OpenAI の経済研究チームによる「OpenAI Signals」の一部で、日本を含む国ごとに、年齢層や話題別の利用状況を確認できます。

調査方法 概要

データは、2024 年 7 月から 2026 年 9 月まで毎月 100 万件ずつ抽出したメッセージのサンプルに基づいています。
対象は 18 歳以上と申告したアカウントで、プランは Free、Go、Plus、Pro です。アカウントを削除したユーザーや学習へのデータ利用を許可していないユーザー、Enterprise と Codex の利用は含まれません。
そのため、ビジネスや技術用途は実際より少なく出ている可能性が高いと OpenAI は述べています。
メッセージは自動で分類され、公開データには差分プライバシーによる保護が施されています。
話題は次の 7 つです。

- 実用的な助言（Practical guidance）
- 文章作成（Writing）
- 情報探し（Seeking information）
- 自己表現（Self-expression）
- 技術的な手助け（Technical help）
- マルチメディア（Multimedia）
- その他・不明（Other）

詳しい数値は、インタラクティブなグラフや公開 CSV で確認できます。

https://www.suzukikenichi.com/blog/does-chatgpt-usage-differ-by-generation-what-japans-age-data-shows/

ChatGPTの使い方は世代で違う？ 日本の年齢別データから見えること

OpenAI は、個人が ChatGPT をどう使っているかを示すインタラクティブなデータダッシュボードを更新した。日本を含む国ごとに、年齢層や話題別の利用状況を確認できる。

日本の年齢別の利用状況

日本の年齢別メッセージシェアは次のとおりです。

年齢	2024 年 7 月	2025 年 9 月	2026 年 9 月
18～24 歳	40.2%	29.9%	23.8%
25～34 歳	29.4%	33.5%	33.0%
35～44 歳	16.5%	17.9%	20.3%
45～54 歳	9.4%	12.1%	14.5%
55～64 歳	3.4%	4.7%	6.2%
65 歳以上	1.2%	1.9%	2.1%

2024 年 7 月と 2026 年 9 月を比べると、**35 歳以上のすべての年齢層で割合が上がって**います。
ただし、毎月一貫して増えたわけではありません。
35 歳以上の合計は 30.5% から 43.1% に上がり、18～34 歳の合計は 69.6% から 56.8% に下がりました。
それでも、メッセージの過半数は引き続き 35 歳未満が占めています。
最も大きな層は 25～34 歳です。
2025 年中は 18～24 歳と首位が入れ替わっていましたが、2025 年 11 月からは 25～34 歳が首位を保っており、割合は 32.2～34.2% で推移しています。

2026 年 9 月の話題別データでは、年齢層ごとの特徴も見えます。

- **18～24 歳:**全体でのシェアは 23.8%。文章作成では 29.4% と高くなっている。マルチメディアでは 18.5%
- **25～34 歳:**すべての話題で最大のシェアを占めており、割合は 29.3～35.2%
- **45～54 歳:**全体でのシェアは 14.5% だが、技術的な手助けでは 16.8%、情報探しでは 16.2% と高くなっている。文章作成では 10.9%
- **55～64 歳と 65 歳以上:**情報探しのシェアが全体より高く、それぞれ 7.4%（全体は 6.2%）、2.9%（全体は 2.1%）

なお、話題別のグラフが示しているのは「各話題のメッセージを、どの年齢層が送ったか」です。
「その年齢層が最もよく使う話題」を示しているわけではありません。
また、これは構成比なので、ユーザー数やメッセージ数そのものが増えたか減ったかは判断できません。

https://www.suzukikenichi.com/blog/does-chatgpt-usage-differ-by-generation-what-japans-age-data-shows/

ChatGPTの使い方は世代で違う？ 日本の年齢別データから見えること

OpenAI は、個人が ChatGPT をどう使っているかを示すインタラクティブなデータダッシュボードを更新した。日本を含む国ごとに、年齢層や話題別の利用状況を確認できる。

日本のデータから言えそうなこと

ここからは、上記のデータをもとにした推測を含みます。

ChatGPT は若い世代のツールから、幅広い世代のツールに移りつつある可能性あり

2 年余りで 35 歳以上のシェアが約 3 割から 4 割強に広がりました。

ただし、構成比の変化なので、中高年の利用が増えたのか、若い世代の伸びが相対的に鈍ったのかは、このデータだけでは区別できません。

中高年にとって、ChatGPT は検索に近い使われ方をしている可能性あり

45 歳以上の各層は、情報探しで全体より高いシェアを占めています。

これまで検索エンジンで調べていたことを ChatGPT に聞く層が、中高年にも広がっているのかもしれません。

SEO や GEO の観点では、中高年向けの情報を扱うサイトほど、AI 経由で情報が探される流れを意識しておく価値がありそうです。

また、技術的な手助けが 16.8% と他世代よりも高いのは、コーディング利用に使う職種が多いせいかもしれません。

若い世代は「調べる」より「作る」に使う傾向あり

18〜24 歳は文章作成でシェアが高く、45〜54 歳は文章作成でシェアが低くなっています。

大学生を中心にレポートや論文を書くのに使っているのではないかと推測します。

このように、世代によって、ChatGPT に期待する役割が違っているようです。

中心的な利用層は 25〜34 歳

すべての話題で最大のシェアを占めています。

仕事と私生活の両方で ChatGPT を幅広く使っている層だと考えられます。

◇ ◇ ◇

個人的には、中高年の「情報探し」における ChatGPT の存在感が気になります。

検索からの置き換えとまでは言えませんが、自サイトの読者がどこで情報を探しているのか、年齢層も踏まえて今後の施策を考えておきたいところです。

https://www.suzukikenichi.com/blog/a-diagnostic-tool-to-check-whether-gptbot-claudebot-and-other-ai-crawlers-are-blocked/

robots.txtだけでは分からない、GPTBotやClaudeBotなどAIクローラーのブロック状況を調べる診断ツール

AI ボット(クローラー)が、自分のサイトをクロールできるか・できないかを調査するツールの紹介

この記事では、AI ボット(クローラー)が、自分のサイトをクロールできるか・できないかを調査するツールを紹介します。

SEO & LLM Diagnostic & Reference Tools というツールで、Vanessa Fox(ヴァネッサ・フォックス)氏が開発しました。

ℹ️すずきメモ:フォックス氏は、現在は独立した SEO コンサルタントだが、Google 在籍時に Search Console の前身である Google Webmaster Central/Web Master Tools を構築した人物。また、検索エンジンの標準規格である sitemaps.org の立ち上げにも尽力した。

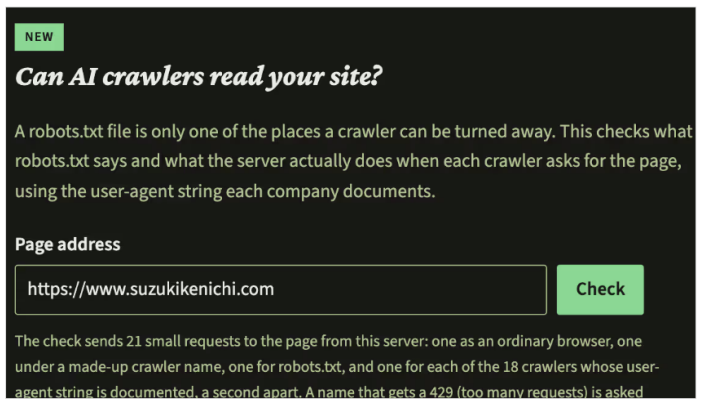
ツールの使い方と見方

GPTBot や ClaudeBot、Google-Extended など 20 種近くの AI クローラーが、任意のサイトのページを読めるかどうかを調べることができます。

次の監査項目が対象です。

- **robots.txt の記述:** 各クローラーを許可しているか、ブロックしているかを確認する
- **サーバーの実際の応答:** 各クローラーの User-Agent で実際にリクエストを送り、ページを返すか(200)、拒否するか(429 など)を確認する。robots.txt だけでは分からない、CDN やサーバー側のブロックを見つけられる。
- **HTTP ヘッダーの内容:** X-Robots-Tag などの noarchive や noindex といった HTTP レスポンスヘッダーの値も確認する。

使い方はシンプルです。



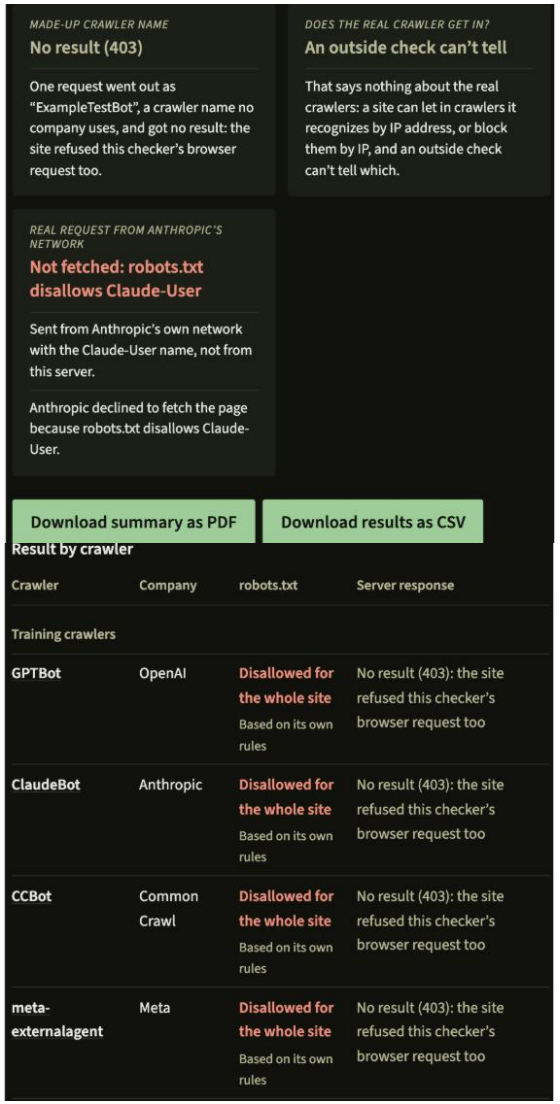
https://www.suzukikenichi.com/blog/a-diagnostic-tool-to-check-whether-gptbot-claudebot-and-other-ai-crawlers-are-blocked/

robots.txtだけでは分からない、GPTBotやClaudeBotなどAIクローラーのブロック状況を調べる診断ツール

AI ボット(クローラー)が、自分のサイトをクロールできるか・できないかを調査するツールの紹介

New York Times(ニューヨーク タイムズ)の調査結果です。
多くの AI クローラーを NYT はサイト全体でブロックしています。





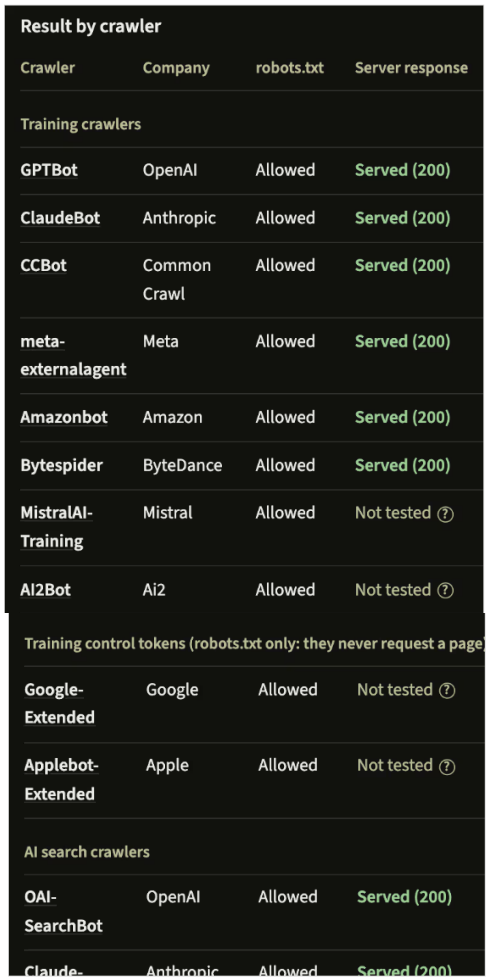
余談ですが、NYT は Google-Extended もブロックしています。
しかし、AI contribution pilot にほぼ間違いなく参加しているはずです。

https://www.suzukikenichi.com/blog/a-diagnostic-tool-to-check-whether-gptbot-claudebot-and-other-ai-crawlers-are-blocked/

robots.txtだけでは分からない、GPTBotやClaudeBotなどAIクローラーのブロック状況を調べる診断ツール

AI ボット(クローラー)が、自分のサイトをクロールできるか・できないかを調査するツールの紹介

僕のブログは、どの AI クローラーもブロックせずクロール許可しています。
ツールは正しく判定できています。



Result by crawler

Crawler	Company	robots.txt	Server response
Training crawlers			
GPTBot	OpenAI	Allowed	Served (200)
ClaudeBot	Anthropic	Allowed	Served (200)
CCBot	Common Crawl	Allowed	Served (200)
meta-externalagent	Meta	Allowed	Served (200)
Amazonbot	Amazon	Allowed	Served (200)
Bytespider	ByteDance	Allowed	Served (200)
MistralAI-Training	Mistral	Allowed	Not tested
AI2Bot	Ai2	Allowed	Not tested
Training control tokens (robots.txt only: they never request a page)			
Google-Extended	Google	Allowed	Not tested
Applebot-Extended	Apple	Allowed	Not tested
AI search crawlers			
OAI-SearchBot	OpenAI	Allowed	Served (200)
Claude-	Anthropic	Allowed	Served (200)

結果は PDF または CSV でダウンロードできます。
また、チェック対象の AI クローラーの開発元や User Agent 名、役割などの詳細はスプレッドシートにまとめられています(このスプレッドシートだけでも役立つリソース)。

◇ ◇ ◇

自身で意図して AI クローラーをブロックしているのであれば問題ありません。
しかし、レンタルサーバーや CDN、プラグインなどの設定で意図せず AI クローラーがブロックされているケースもありえます。
ツールを使って、管理サイトの AI クローラー制御状況を確認してみてください。

https://www.suzukikenichi.com/blog/google-search-query-growth-hits-a-5-year-high-driven-by-ai-mode-alphabet-q3-2026-earnings/

Google検索のクエリ成長率が過去5年で最高に、AI Modeが牽引

Google の持ち株会社である Alphabet は 2026 年第 3 四半期決算を発表した。

Google の持ち株会社である Alphabet は 2026 年第 3 四半期決算を発表しました。
この記事では、Google 経営陣による説明と公表数値に基づいて、検索関連の重要ポイントをまとめます。

2026 Q3 決算発表から読み解く Google 検索の現状

- **クエリの伸び:**
  AI への投資が検索利用とクエリ数の増加につながっており、商業目的のクエリも伸びている。CEO の Sundar Pichai（スンダー・ピチャイ）氏によると、クエリ数の前年比成長率は過去 5 年で最高。
  ※具体的な成長率は示されていない
- **一つの検索体験:**
  AI Overviews、AI Mode、従来の検索を横断するシームレスな体験を構築している。モバイルへの移行時と同じく、ユーザー体験を優先し、その後に収益化を進める方針。
- **検索のしかたの変化:**
  AI Mode のクエリは従来の検索より平均 3 倍長い。世界の AI Mode 検索の 5 回に 1 回は、音声、画像、文書を使っている。また、Google のビジュアル検索は月間 500 億回を超えた。
  ※月間 500 億回の数字は AI Mode だけの検索回数として示されたものではない
- **AI Mode のクエリの特徴:**
  多くは購入前の情報収集など、アッパーファネルのクエリであり、新たな価値の高い商業的意図を捉えられるという。CBO の Philipp Schindler（フィリップ・シンドラー）氏は、AI Mode のクエリの大半が検索全体に上乗せされる「追加分」だと説明した。
- **AI 検索面の広告:**
  Google の利用者調査では、ユーザーはこれらの検索面の広告を受け入れ、関連性が高いと感じているという。AI Mode 内の商品リストでは、広告と自然掲載商品のクリック率がすでにほぼ同じ。Google は AI Overviews と AI Mode の広告をより多くのユーザーとクエリへ拡大しており、今回、AI Mode 広告の初の米国外展開と、米国でのマルチターン広告を発表した。
  ※決算発表では対象国の詳細には触れていないが、広告が新たに展開されるのは、オーストラリア、カナダ、インド、ケニア、マレーシア、ニュージーランド、ナイジェリア、パキスタン、フィリピン、シンガポールの10か国。2026年第4四半期には、ブラジル、メキシコ、イギリスを含む追加の地域にも拡大予定。マルチターン広告（multiturn ads）はフォローアップ質問の回答で掲載される広告
- **検索の収益:**
  「Google 検索とその他」の広告収益は前年比 12% 増の 633 億ドル。成長への貢献が最も大きかった業種は小売で、次が金融だった。
- **第 2 四半期から収益成長率が鈍化した理由:**
  最大の要因は為替だった。加えて、EU の少額輸入品に対する関税免除（デミニミス）の廃止がアジア太平洋の一部小売広告主に影響した。メディア・エンタメでは、ワールドカップで例年より好調だった第 2 四半期から広告支出が平常化した。収益化水準がまだ低い AI Mode のクエリが大きく増えたことも、成長率に影響した。

https://www.suzukikenichi.com/blog/google-search-query-growth-hits-a-5-year-high-driven-by-ai-mode-alphabet-q3-2026-earnings/

Google検索のクエリ成長率が過去5年で最高に、AI Modeが牽引

Google の持ち株会社である Alphabet は 2026 年第 3 四半期決算を発表した。

- **第 4 四半期の見通し:**
  現在の為替水準では、為替の逆風が第 3 四半期よりやや強まり、主に検索と YouTube 広告に影響する見込み。また、2025 年後半に検索の成長が加速したため、前年同期との比較条件が厳しくなる。CFO の Anat Ashkenazi（アナト・アシュケナージ）氏は、この比較要因について、第 3 四半期は約 3 ポイント、第 4 四半期は 2 ポイント程度と説明した。
- **広告の品質とツール:**
  第 3 四半期に商品リスト広告を過去 10 年で最大規模で改良し、満足度は 19% 上昇、不満は半分以下になった。Gemini を使って広告主が生成したキャンペーン素材は、第 3 四半期に 15 億点を超え、前年同期の 7 倍以上となった。AI Max では、複数の検索キャンペーンにまたがる予算や ROI 目標を一つの A/B テストで検証する機能を導入する。Telstra の事例では、長いクエリや多言語クエリを捉え、コンバージョンが 52% 増え、注文あたりの費用が 36% 減った。
- **エージェント機能と買い物支援:**
  予約、フォーム入力、日程調整、購入などを支援するエージェント機能は、企業向けの Gemini agent で先行して整備し、検索と Gemini アプリへ広げる方針。Universal Commerce Protocol（UCP）は、Google 上で数十万のブランドの商品を直接購入できるようにしている。Universal Cart も提供を開始したが、こちらはユーザーに代わって行動するエージェント機能とは区別され、買い物客が自分で管理・判断するための実用的なツールとしての位置付け。
- **検索の拡大機会:**
  経営陣はエージェント機能を検索にとって「追加的なもの」「拡大の機会」と捉えている。シンドラー氏は、エージェントでも複数の選択肢を提示する余地があり、広告機会につながると説明した。ピチャイ氏は、ナレッジグラフなど、長年蓄積した世界に関する理解が強みになると述べた。

【参照ソース】

- Alphabet Investor Relations – 2026 Q3 Earnings Call
- Q3 2026 earnings call: Remarks from our CEO

◇ ◇ ◇

AI がクエリ数の成長を加速させている一方で、AI Mode の低い収益化水準が検索広告の収益成長率に影響しているようです。
今後は AI 検索面での広告拡大と、エージェント機能の展開を進めるとのことです。
AI によって検索の機会は広がっていますが、クエリの増加がサイトへの流入増加につながるとは限りません。
AI 検索を意識した SEO では、購入前の情報収集も含めたユーザーの幅広い疑問に応えながら、流入や成果の変化を継続的に見ていくことが大切です。
また、動画や画像といったマルチモーダル コンテンツへの注力も必須です。

https://www.suzukikenichi.com/blog/one-setting-to-opt-out-of-ai-training-google-openai-and-other-major-ai-companies-to-support-a-shared-signal/

AI学習の拒否を一括設定、GoogleやOpenAIなど主要AI企業が共通シグナルに対応へ

Google、OpenAI、Microsoft の 3 社は、AI 学習へのコンテンツ利用を拒否するための robots.txt の共通シグナルに対応すると、それぞれ発表した。

Google、OpenAI、Microsoft の 3 社は 2026 年 10 月 30 日、AI 学習へのコンテンツ利用を拒否するための robots.txt の共通シグナルに対応すると、それぞれ発表しました。

概要

この共通シグナルを使えば、サイト運営者は 1 回の設定で、対応するすべての AI 事業者に学習の拒否を伝えられます。
これまでは AI 企業のクローラーごとに個別のルールを書く必要がありました。

たとえば、Google Gemini の学習利用を拒否したいなら、Google-Extended を robots.txt でブロック（Disallow）します。
OpenAI の AI モデルの学習利用を拒否したいなら、GPTBot を robots.txt でブロックします。

しかし、共通シグナルを利用すれば、こうした個別の設定の手間がなくなります。

仕様は、Anthropic の Kevin Kelley 氏が IETF に提出した Internet-Draft「Machine-Readable Training Preference Signal」にまとめられています。
「AI 学習」の定義は、IETF の AIPREF（AI Preferences）ワーキンググループの語彙ドラフト「draft-ietf-aipref-vocab-08」から引用されています。

書き方と仕組み

書き方はシンプルです。

robots.txt に User-Agent: _usage_ai-training のグループを作り、いつもどおり Disallow と Allow をパスごとに書くだけです。

```
User-Agent: _usage_ai-training
Disallow: /
Allow: /press-kit/
```

User-Agent で指定する「_usage_ai-training」は固定です。
この値が、各 AI 企業の AI モデルが従う共通のシグナルであることを示します。

2 行目の Disallow: / でサイト全体のコンテンツの学習利用を拒否しています。
3 行目の Allow: /press-kit/ は拒否の例外設定です。
/press-kit/ ディレクトリ配下のコンテンツに関しては学習利用を許可しています。

AI学習の拒否を一括設定、GoogleやOpenAIなど主要AI企業が共通シグナルに対応へ

Google、OpenAI、Microsoft の 3 社は、AI 学習へのコンテンツ利用を拒否するための robots.txt の共通シグナルに対応すると、それぞれ発表した。

なお、robots.txt における新しいディレクティブや文法は追加されていません。
つまり、robots.txt の書き方そのものは変えず、既存のルールの枠組みにそのまま収まる形にした仕組みです。
たとえば、Content-Usage: train-ai=n のような新しいディレクティブを追加したわけではありません（AIPREF の初期では、例に出したディレクティブが AI 学習を拒否するための新たな命令として議論されていた）。
このシグナルに対応していないパーサーやクローラーから見ると、_usage_ai-training のグループは「知らないクローラー向けのルール」にすぎません。
自分宛てではないので読み飛ばします。
その結果、

- Googlebot や Bingbot などのクロール制御は、これまでとまったく同じように動きます
- 構文エラーになったり、ほかのルールの解釈が変わったりすることはありません

ただし、対応していないパーサーが、その行を無視するのか、エラーとして扱うのか、前後のルールの解釈が崩れるのかは、AI 企業の実装しだいでばらつく可能性があります。
ドラフトは「実際に使われているパーサーで問題なく動くこと」を設計目標に挙げていて、作者の Kelley 氏もメーリングリストの投稿で、パーサーごとの挙動の違いに頼らない方式を選んだと説明しています。
ちなみに、User-Agent の名前の先頭に _ を付けているのは、柔軟すぎるマッチングをするパーサーが、ほかのクローラー名と誤って一致させるのを防ぐためとのことです。

少し脱線しましたが、押さえておきたいポイントは次のとおりです。

- **対象は生成 AI モデルの学習だけ：**
  ここでいう「AI 学習」は、コンテンツを使って生成 AI モデルのパラメータを更新することを指す。分類、ランキング、スコアリングなど、生成以外の用途だけに使うモデルの学習は含まれない。
- **User-Agent: * は対象外：**
  ワイルドカードのグループを学習の意思表示として読んではならないと定められている。_usage_ai-training のグループがなければ、意思表示はなかったものとみなされる。
- **クロール制御とは別物：**
  このトークンが示すのはコンテンツの使い道であり、クロールしてよいかどうかではない。クローラーごとのルールはこれまでどおり機能する。
- **具体的な指定が優先される：**
  企業独自のトークンのような具体的な指定を、共通トークンより優先させてもかまわない。また、サイト運営者と AI 事業者が個別に契約（ライセンス契約など）を結んでいる場合は、その内容をシグナルより優先させてもかまわない。

なお、どちらのドラフトもまだ Internet-Draft の段階で、正式な標準ではありません。
Kelley 氏のドラフトは個人提出のもので、IETF の承認はなく、有効期限は 2027 年 5 月 3 日です。

https://www.suzukikenichi.com/blog/one-setting-to-opt-out-of-ai-training-google-openai-and-other-major-ai-companies-to-support-a-shared-signal/

AI学習の拒否を一括設定、GoogleやOpenAIなど主要AI企業が共通シグナルに対応へ

Google、OpenAI、Microsoft の 3 社は、AI 学習へのコンテンツ利用を拒否するための robots.txt の共通シグナルに対応すると、それぞれ発表した。

各社の対応

各 AI 企業の現時点での対応は次のとおりです。

Google

Google は、サイトが共通トークンで学習を拒否していて、Google-Extended を明示的に指定していなければ、Google-Extended を Disallow したものとして扱います。

Google-Extended は今後も使え、明示的に許可したサイトは、Gemini アプリのグラウンディング経由で回答に表示され、クリックを得られます。

対応は 2027 年の予定です。
Google は、この取り組みが EU AI 法の GPAI 行動規範にも沿っているとしています。

OpenAI

OpenAI は今後数か月以内に導入する予定です。
詳細はそのときに公表するとしています。

既存のクローラー制御も引き続き有効です。

Anthropic

Anthropic は、ドラフトの公開を知らせる Kelley 氏のメーリングリスト投稿によると、すでに対応の実装を始めています。
既存のクローラー制御を置き換えるのではなく、補うものだとしています。

Microsoft

Microsoft (Bing) は共通シグナルに加えて独自の拡張を用意し、2027 年 2 月 1 日から適用します。
ページ単位での学習設定に加え、学習とは別に、コンテンツを AI の回答の根拠に使う「グラウンディング」の設定を追加し、従来の NoArchive と NoCache を置き換えます。

NoArchive や NoCache を使って AI の学習・引用を制御しているサイトは、2027 年 3 月 1 日までに対応が必要です。

詳しくは別記事で後日解説します。

https://www.suzukikenichi.com/blog/one-setting-to-opt-out-of-ai-training-google-openai-and-other-major-ai-companies-to-support-a-shared-signal/

AI学習の拒否を一括設定、GoogleやOpenAIなど主要AI企業が共通シグナルに対応へ

Google、OpenAI、Microsoft の 3 社は、AI 学習へのコンテンツ利用を拒否するための robots.txt の共通シグナルに対応すると、それぞれ発表した。

◇ ◇ ◇

大手 AI 事業者が、AI 学習を拒否するための robots.txt の共通トークンに足並みをそろえつつあります。
ただし、仕様はまだドラフトの段階で、対応時期も各社で異なります。
今後の標準化の動きと、各社の実装状況を見守りたいところです。

進展があればお伝えします。