

Google検索・AI Overview経由アクセスにおける閲覧傾向分析

Analysis of Browsing Trends in Google Search and AI Overview Referral Access

概要 (Abstract)

本レポートのデータはWebPita.comの機能を利用して観測をしている。

本レポートでは、Google検索画面およびGoogle AI Overview画面からのアクセスについて、Webアクセスログを用いた閲覧傾向の観測を行った。

観測では、Google referer を持つアクセスのうち、HTTP (200,206,304)ステータスかつ html / php / pdf / ディレクトリアccessを対象とし、画像・CSS・JavaScript・リダイレクト・エラーアクセス等を除外した。

また、User-Agent判定、アクセス挙動判定、およびISP判定を組み合わせることで、「実コンテンツ閲覧に近いアクセス」の抽出を試みた。

観測の結果、Google referer を持つアクセスのうち、「人間らしいアクセス」と推測される割合は約40%~60%程度で推移した。

一方で、Google referer を持つアクセスの中には、自動巡回やBot的特徴を持つアクセスも多数含まれていることが確認された。

また、PDF・技術情報を掲載する情報提供型サイトと、写真閲覧主体サイトでは、人間率に差異が確認され、コンテンツ特性がアクセス傾向へ影響している可能性が示された。

本観測は、「人間による操作」を完全に証明するものではなく、Webアクセスログを用いた閲覧傾向の観測結果として整理している。

Analysis of browsing trends in Google Search and AI Overview referral access using web access logs. Human-like access rates, User-Agent analysis, ISP classification, bot traffic analysis, and real content browsing conditions were observed.

第1章 本レポートの目的

Chapter 1 : Purpose of This Report

本レポートでは、Google検索画面およびGoogle AI Overview画面からのアクセスについて、Webアクセスログを用いた閲覧傾向の観測を行った。

観測では、Google referer を持つアクセスを対象として、「実コンテンツ閲覧に近いアクセス」がどの程度存在するかについて分析を実施している。

また、Google referer を持つアクセスの中に、自動巡回やBot的特徴を持つアクセスがどの程度含まれているかについても観測を行った。

第2章 観測対象アクセス

Chapter 2 : Target Access for Observation

観測では、Web接続ログより以下の条件に該当するレコードを対象とした。

| 項目 Target Item | 内容 Description |
|--------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 対象接続ログ Target Access Logs | HTTP (200,206,304)ステータスで、拡張子が html / php / pdf のファイル、およびディレクトリアクセスを対象とする。 HTTP (200,206,304) access to html / php / pdf files and directory access were targeted. |
| 除外条件 Exclusions | ※リダイレクト・エラー・画像・CSS・JavaScript等の取得を除外し、実コンテンツ閲覧に近いアクセスを観測対象とした。 Redirects, errors, images, CSS, JavaScript, and similar requests were excluded. |
| 対象のGoogle referer Target Google referer | <ul style="list-style-type: none">• https://gemini.google.com/• https://bard.google.com/• https://www.google.com/• https://www.google.co.jp/• android-app://com.google.android.googlequicksearchbox/ |

第3章 人間らしいアクセスの判定

Chapter 3 : Classification of Human-like Access

判定は、ユーザーエージェントおよびアクセス挙動による判定と、ISP判定の2段階で実施した。

完全な「人間による操作」の判定は困難であるため、本観測では「Hosting / Datacenter系ISPからの接続を除外することで、Botアクセスの多くを除外できる」という仮説に基づき集計を行っている。

一方で、AS Name取得等の外部照会はシステム負荷が高く、また全てのISP情報を保持しているわけではない。そのため前段として、ユーザーエージェントやアクセス挙動による簡易的な除外判定を実施している。

判定方法

Classification Method

| 判定分類 Classification | 判定内容 Check Item | 条件 Condition |
|-----------------------------------|------------------------------------|------------------------------------------------------------------------------------------|
| UAや挙動による判定 UA / Behavior Check | User-Agentチェック User-Agent Check | User-Agent文字列が極端に短い場合はNGとする。 Marked NG when the User-Agent string is extremely short. |

| 判定分類 Classification | 判定内容 Check Item | 条件 Condition |
|-----------------------------|--------------------------------------|----------------------------------------------------------------------------------------------------------|
| | User-Agentキーワード Bot Keywords | User-Agent文字列にbotらしき文字列が含まれる場合はNGとする。 Marked NG when bot-like keywords exist. |
| | ブラウザ要素 Browser Elements | ブラウザ要素（mozillaやchrome等）が存在しない場合はNGとする。 Marked NG when browser elements are missing. |
| | ブラウザVersion要素 Browser Version | ブラウザVersion要素が存在しない場合はNGとする。 Marked NG when browser version elements are missing. |
| | request_uriチェック request_uri Check | request_uri が空白の場合はNGとする。 Marked NG when request_uri is blank. |
| | 特定ファイルアクセス System Files | robots.txt / sitemap.xml にアクセスしたIPはNGとする。 IPs accessing robots.txt or sitemap.xml were marked NG. |
| | 高頻度アクセス High Frequency | 2秒以内に複数ページを閲覧しているIPはNGとする。 IPs accessing multiple pages within two seconds were marked NG. |
| ISP判定 ISP Classification | 日本国内IP Domestic Japanese IP | 主要レンタルサーバー接続であればNG判定とする。 Major hosting provider access was marked NG. |
| | 海外IP Overseas IP | AbuseIPDB の usage 情報を利用しISP判定を行う。 ISP classification was performed using AbuseIPDB usage information. |

第4章 観測対象サイト

Chapter 4 : Observed Websites

本観測では、コンテンツ特性の異なる2種類のWebサイトを対象として観測を行った。

| サイト Site | コンテンツ特性 Content Characteristics |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------|
| A | 業界情報・行政情報等を掲載する情報提供型サイト。HTMLおよびPDFによる技術情報・業界資料を掲載している。 Information-oriented website containing technical documents and PDF files. |
| B | 土木写真を中心とした写真閲覧型サイト。約4,000点の写真を掲載している。 Photo-browsing website containing approximately 4,000 civil engineering photos. |

第5章 観測結果

Chapter 5 : Observation Results

2026年5月5日～2026年5月13日の期間において、Google検索画面およびGoogle AI Overview画面からのアクセスを対象に観測を行った。

サイト別集計（全体）

| site_id | Google referer全件 All Referer | HTML/PHP/PDF Content Access | UA/挙動 NG UA/Behavior NG | ISP NG ISP NG | 人間らしいアクセス Human Access | 人間率 Human Rate |
|---------|---------------------------------|--------------------------------|----------------------------|------------------|---------------------------|-------------------|
| A | 4,102 | 3,791 | 1,421 | 775 | 1,595 | 42.1% |
| B | 5,535 | 2,416 | 433 | 663 | 1,320 | 54.6% |

日付別推移（site A）

Daily Trend : Site A

| 日付 Date | Google referer全件 All Referer | 対象件数 Target Count | UA/挙動NG UA/Behavior NG | ISP NG ISP NG | 人間らしいアクセス Human Access | 人間率 Human Rate |
|------------|---------------------------------|----------------------|---------------------------|------------------|---------------------------|-------------------|
| 2026/5/5 | 148 | 124 | 22 | 47 | 55 | 44.4% |
| 2026/5/6 | 205 | 178 | 47 | 52 | 79 | 44.4% |
| 2026/5/7 | 709 | 673 | 320 | 108 | 245 | 36.4% |
| 2026/5/8 | 597 | 558 | 215 | 109 | 234 | 41.9% |
| 2026/5/9 | 214 | 198 | 48 | 63 | 87 | 43.9% |
| 2026/5/10 | 202 | 185 | 44 | 57 | 84 | 45.4% |
| 2026/5/11 | 654 | 604 | 238 | 127 | 239 | 39.6% |
| 2026/5/12 | 647 | 595 | 235 | 99 | 261 | 43.9% |
| 2026/5/13 | 726 | 676 | 252 | 113 | 311 | 46.0% |

日付別推移（site B）

Daily Trend : Site B

| 日付 Date | Google referer全件 All Referer | 対象件数 Target Count | UA/挙動NG UA/Behavior NG | ISP NG ISP NG | 人間らしいアクセス Human Access | 人間率 Human Rate |
|------------|---------------------------------|----------------------|---------------------------|------------------|---------------------------|-------------------|
| 2026/5/5 | 671 | 272 | 32 | 93 | 147 | 54.0% |

| 日付 Date | Google referer全件 All Referer | 対象件数 Target Count | UA/挙動NG UA/Behavior NG | ISP NG ISP NG | 人間らしいアクセス Human Access | 人間率 Human Rate |
|------------|---------------------------------|----------------------|---------------------------|------------------|---------------------------|-------------------|
| 2026/5/6 | 520 | 207 | 15 | 55 | 137 | 66.2% |
| 2026/5/7 | 702 | 317 | 69 | 78 | 170 | 53.6% |
| 2026/5/8 | 650 | 301 | 76 | 76 | 149 | 49.5% |
| 2026/5/9 | 545 | 230 | 31 | 63 | 136 | 59.1% |
| 2026/5/10 | 551 | 211 | 19 | 59 | 133 | 63.0% |
| 2026/5/11 | 613 | 281 | 63 | 74 | 144 | 51.2% |
| 2026/5/12 | 642 | 318 | 76 | 81 | 161 | 50.6% |
| 2026/5/13 | 641 | 279 | 52 | 84 | 143 | 51.3% |

第6章 考察

Chapter 6 : Discussion

本観測では、Google検索画面およびGoogle AI Overview画面からのアクセスであっても、その全てが「人間による閲覧」であるとは限らない傾向が確認された。

- robots.txt / sitemap.xml への巡回
- Hosting / Datacenter系ISPからの接続
- 不自然なUser-Agent
- 高頻度アクセス

等、自動巡回やBot的特徴を持つアクセスが多数含まれていた。

また、観測対象サイトによって、「人間らしいアクセス」の割合に差異が確認された。

site A では、人間率は約40%台で推移し、UAや挙動による除外件数が多い傾向が見られた。一方、site B では50%~60%台で推移し、比較的安定した閲覧傾向が確認された。

この差異には、以下の要素が影響している可能性が考えられる。

- PDF掲載の有無およびそのボリューム
- 技術情報の専門度・量
- 検索クエリや検索導線の違い
- ページ構成（インデックスのされ方）
- その他、コンテンツ自体の特性

そのため、Google検索・AI Overview経由アクセスの分析では、単純なアクセス数のみではなく、実コンテンツ閲覧条件、User-Agent、ISP、アクセス挙動等を組み合わせた分析が重要であると考えられる。

第7章 まとめ

Chapter 7 : Conclusion

本観測では、Google検索画面およびGoogle AI Overview画面からのアクセスについて、Webアクセスログを用いた閲覧傾向の分析を行った。

その結果、Google referer を持つアクセスの中には、自動巡回やBot的特徴を持つアクセスが多数含まれていることが確認された。

また、サイト特性によって「人間らしいアクセス」の割合に差異が存在し、コンテンツ構成や掲載情報がアクセス傾向へ影響している可能性が確認された。

AI時代のアクセス解析では、単純なアクセス数のみではなく、User-Agent、ISP、アクセス挙動、実コンテンツ閲覧条件等を組み合わせた多面的なアクセス解析が重要になると考えられる。