

WebPitaのGoogle系AI Web訪問判別方法

WebPita's Verification Methodology for Google AI Web Access

発行日: 2026-05-16

概要

Googleは、AI用途専用の完全な識別情報を公開していません。そのため、Google系AIアクセスの判定には一定の誤差が含まれます。

WebPitaでは、Google公式IPレンジ、User-Agent、Refererを組み合わせ、AI関連アクセス群として高精度に分類しています。

また、観測対象ログは、HTTPステータスが（200, 206, 304）であり、拡張子が `html`、`php`、`pdf` のファイル、およびフォルダへのアクセスに限定しています。

1. User-Agentのみで判定しない理由

WebPitaでは、Google系AI User-Agentに一致したアクセスであっても、その全てをAI関連アクセスとして扱っているわけではありません。

実際の観測では、Google系AI User-Agentに一致し、HTTPステータスやURI条件を満たしたアクセスであっても、WebPitaの判定条件を通過する件数は**53.6%~69.4%**となったケースが確認されています。

これは、User-AgentのみでAIアクセスと断定せず、Google公式IPレンジ、Referer、アクセス対象ファイル、接続挙動などを組み合わせて、AI関連アクセスの可能性が高いものだけを分類しているためです。

このようにWebPitaでは、単純なUser-Agent一致ではなく、複数条件による絞り込みを行うことで、偽装アクセスやAI関連性の低いアクセスを除外しています。

【WebPitaのUA判別・検証結果】

以下は、WebPitaが実際のWebアクセスログに対してGoogle系AI User-Agent判定を行った結果です。

観測期間	観測対象ログ件数	Google系AI UA一致のログ件数	判定通過数	判定通過率
2026-05-05 - 2026-05-13	3,062 件	599 件	321 件	53.6%
2026-05-05 - 2026-05-13	2,406 件	294 件	204 件	69.4%

2. User-AgentとGoogle公式IPレンジの組み合わせによる判定

WebPitaでは、以下のUser-AgentとGoogle公式IPレンジ分類の組み合わせを、AI関連アクセスとして1次分類しています。

User-Agent	判定対象のGoogle公式IPレンジ
Google Web Preview	Google Public IP Ranges / User-initiated Fetchers
google-agent	GoogleProducer / Google Public IP Ranges / User-triggered Fetchers / Google User-triggered Agents
Gemini	GoogleProducer / Google Public IP Ranges / User-triggered Fetchers / Google User-triggered Agents
google-lens	Google Public IP Ranges / User-initiated Fetchers / User-triggered Fetchers / Google
GoogleOther	Google Public IP Ranges / User-triggered Fetchers / Google
GoogleOther-Image	Google Public IP Ranges / User-triggered Fetchers / Google
GoogleOther-Video	Google Public IP Ranges / User-triggered Fetchers / Google
GoogleProducer	GoogleProducer / Google Public IP Ranges / User-triggered Fetchers / Google
Google-Extended	Google Public IP Ranges / User-triggered Fetchers / Google
Google-Read-Aloud	Google Public IP Ranges / User-initiated Fetchers / User-triggered Fetchers / Google
GSA	Google Public IP Ranges / User-initiated Fetchers / User-triggered Fetchers / Google
Google-CloudVertexBot	GoogleProducer / Google Public IP Ranges / User-initiated Fetchers / User-triggered Fetchers / Google / User-triggered Agents
Google-NotebookLM	Google Public IP Ranges / User-initiated Fetchers / User-triggered Fetchers / Google

3. Referer判定

Refererが空白（リファラなし）の場合は、User-AgentとGoogle公式IPレンジの組み合わせによる判定を継続します。

Refererが存在する場合は、Google系Refererであることを確認します。Google系Refererではない場合は、AI関連アクセスではないものとして除外します。

これにより、Google系User-Agentを名乗るだけの偽装アクセスや、Google系AIアクセスとは関係のないRefererを持つアクセスを確実に除外します。

4. 観測対象ログ

WebPitaでは、AIが実際に参照した可能性のあるページアクセスを分析するため、観測対象を以下に限定しています。

- HTTPステータスが（200, 206, 304）のアクセス
- html、php、pdfファイルへのアクセス
- フォルダURLへのアクセス

画像、CSS、JavaScript、リダイレクト、404エラーなどは、原則としてAI関連ページ訪問の分析対象から除外しています。

5. 補足事項

Googleは、AI関連アクセスの仕様、User-Agent、IPレンジ分類を継続的に変更しています。そのためWebPitaでは、Google公式IPレンジ情報の更新状況や実際の観測ログをもとに、判定ロジックを継続的に見直しています。

本判定は、GoogleがAI専用の完全な識別情報を公開していない前提での観測ベース判定です。そのため、AIアクセスを完全に断定するものではなく、AI関連アクセスの可能性が高いアクセス群を高度に分類するための判定方法です。

English Summary

WebPita's Verification Methodology for Google AI Web Access

Since Google does not publish a fully dedicated, definitive identification dataset exclusively for AI-related traffic, determining genuine AI crawler access inherently involves a certain margin of error. To

overcome this, WebPita filters and classifies AI-related traffic by cross-referencing three key data points: **Google Official IP Ranges**, **User-Agent (UA) strings**, and **Referer headers**.

Furthermore, to ensure data integrity, the target logs for this observation are strictly limited to successful page views—specifically HTTP statuses 200, 206, and 304, for files with .html, .php, or .pdf extensions, as well as directory URLs.

1. Why User-Agent Alone is Insufficient

WebPita does not classify an access as an AI visit based solely on a matching Google AI User-Agent. In actual observations, even when an access matches a Google AI UA and satisfies the specific HTTP status and URI criteria, the percentage of requests that successfully pass WebPita's comprehensive verification conditions was found to range from **53.6% to 69.4%**.

This indicates that relying solely on User-Agent strings is insufficient to confirm genuine AI access. By combining UA criteria with Google's official IP ranges, referers, target file types, and connection behavior, WebPita effectively isolates high-probability AI-related traffic while eliminating spoofed requests and low-relevance non-AI access.

WebPita Verification Results

Observation Period	Total Target Logs	Google AI UA Matched Logs	Verified AI Accesses	Pass Rate
2026-05-05 - 2026-05-13	3,062	599	321	53.6%
2026-05-05 - 2026-05-13	2,406	294	204	69.4%

2. Continuous Adaptation

Google frequently updates its AI access specifications, User-Agent structures, and IP range data. Consequently, WebPita continuously refines and updates its verification logic based on real-world log analysis and official documentation updates to maintain high precision in AI traffic analytics.