WAF Hardening Report: Bing Indexing Fix and Probe Blocking

WAF 強化レポート: Bing インデックス問題の修正とプローブブロック

Prepared for / 宛先

Japan Activation Capital

ジャパン・アクティベーション・キャピタル

March 11, 2026 / 2026 年 3 月 11 日

English Version

See page 4 →

日本語版

15 ページへ →

eSolia

WAF Hardening Report: Bing Indexing Fix and Probe Blocking

Prepared for

Japan Activation Capital

March 11, 2026

Contents

Executive Summary

Prepared by: eSolia Inc.
Date: 6 March 2026
Site: japanactivationcapital.com
Zone: Cloudflare Pro plan
Status: Complete

The japanactivationcapital.com site — like all public websites — is under constant automated attack. Vulnerability scanners, credential harvesters, and bot networks probe the site continuously, looking for exposed configuration files, WordPress installations, debug endpoints, and other attack surfaces. After migrating to Cloudflare Workers (see migration report), real-time observability made the scale of this traffic visible for the first time: hundreds of malicious requests per hour that were previously invisible under Cloudflare Pages.

This report covers two actions we took in response:

1. **Hardened the WAF with probe-blocking rules.** We ported eSolia's battle-tested ruleset to the JAC zone, adding a second layer of defense on top of Cloudflare's managed rulesets. Together, the two layers stopped 679 probe requests in the first 6 hours — all of which previously reached the Worker and consumed invocations.
2. **Fixed a Bingbot access issue we introduced.** While tightening security, we discovered that a WAF skip rule for search engine crawlers was too narrow, which contributed to Bingbot being challenged. We broadened the rule, verified the fix in Workers logs, and confirmed that Bingbot now passes through without any obstruction. Googlebot has never been affected — Google has indexed the site without interruption throughout this entire process.

Background

Cloudflare's Default Protections

The JAC Cloudflare zone (Pro plan) includes managed rulesets that run automatically — Cloudflare's threat intelligence identifies and blocks known attack patterns, bot signatures, and malicious payloads without any configuration. In the 6-hour sample after this deployment, the managed rulesets independently blocked 338 requests. Extrapolated across a typical week, this is tens of thousands of blocked probes. The custom rules described in this report are an additional layer on top of these defaults, targeting patterns specific to the JAC site's technology stack.

Google vs. Bing: A Stark Contrast

It is worth noting that **Googlebot has had zero problems** with the JAC site at any point — before, during, and after the WAF changes. Google indexed the site immediately after the Workers migration, continued indexing throughout the WAF hardening work, and was never blocked or challenged. Google Search Console shows clean crawl data with no errors. This is the expected behavior for a properly configured site.

Bingbot, by contrast, has a well-documented history of indexing interruptions that are difficult to diagnose. This is not unique to the JAC site — it is an industry-wide pattern. eSolia has experienced similar Bing indexing problems on other client sites as well. The Cloudflare community forums have multiple threads about it, and Microsoft's own Q&A forums are full of webmasters reporting that Bing won't index their sites with vague error messages and no clear resolution path. One root cause identified in the Cloudflare community: even when Bing's own verification tool confirms a bot IP is legitimate, the reverse DNS records sometimes don't match what Cloudflare expects — a bug on Microsoft's side that resolves itself when Microsoft updates the PTR records.

The JAC site experienced Bing indexing problems on its previous host (Netlify) as well — that time with zero server-side visibility into what was happening. The issue was eventually resolved only by escalating directly to Microsoft. Their response when closing the ticket: they "can't provide details" on what was fixed.

The current issue — SBFM interfering with Bingbot despite Cloudflare's documentation stating that verified bots are exempt — is another instance of this pattern. Cloudflare's own community tutorial on SBFM acknowledges that "due to the high security threshold, false positives do sometimes happen." The fix (an explicit skip rule for `cf.client.bot`) is straightforward, but the underlying fragility on Microsoft's side means Bing indexing could break again in the future for reasons entirely outside our control. We'll continue to monitor via Bing Webmaster Tools and Workers observability logs.

Problems Identified

1. Bingbot Skip Rule Was Too Narrow

The existing rule used this expression:

```
(cf.client.bot and http.user_agent contains "bingbot")
```

This is problematic for two reasons:

- **Redundant:** `cf.client.bot` is Cloudflare's verified bot field — it is `true` only for bots confirmed via reverse DNS lookup (Googlebot, Bingbot, Yandex, etc.). Adding a user-agent check on top of this is unnecessary and can cause mismatches if Bing changes its user-agent string.
- **Too narrow:** Only Bingbot was covered. Googlebot and other legitimate search crawlers were not included in the skip rule, meaning they could be challenged or blocked by Super Bot Fight Mode (SBFM).

The working rule on esolia.co.jp uses just `(cf.client.bot)` — no user-agent filter.

2. No SEO-Critical File Protection

There was no skip rule for `/robots.txt`, `/sitemap.xml`, or `/favicon.ico`. If any other firewall rule or rate limit triggered, crawlers could be blocked from fetching the files they need for indexing.

A skip rule for `/llms.txt` existed but did not cover the standard SEO files.

3. No Probe Blocking

The JAC zone had no rules blocking common vulnerability probes. The eSolia zone has been running probe-blocking rules since mid-February, informed by real traffic observed in Workers logs. Without these, every probe request reaches the Worker, consuming invocations and polluting logs.

4. Unused Rules

An "Allow Authenticated APIs" rule existed (checking for `auth_token` cookies on POST requests), but the JAC site has no authenticated API endpoints. This rule served no purpose.

Changes Made

The custom firewall ruleset was replaced with 8 rules in the following order. Rule order matters — skip rules are placed early so they take effect before block rules.

Rule 1: Skip SBFM for Verified Bots

```
Expression:  (cf.client.bot)
Action:      Skip → http_request_sbfm
Logging:     Enabled
```

Lets all Cloudflare-verified search bots (Bingbot, Googlebot, Yandex, etc.) bypass Super Bot Fight Mode. This is the primary fix for the Bing indexing issue.

Rule 2: Allow SEO-Critical Files

```
Expression:  (http.request.uri.path eq "/robots.txt") or
             (http.request.uri.path eq "/sitemap.xml") or
             (http.request.uri.path eq "/sitemap-index.xml") or
             (http.request.uri.path eq "/favicon.ico") or
             (http.request.uri.path eq "/llms.txt")
Action:      Skip → http_ratelimit, http_request_firewall_managed, http_request_sbfm
             + remaining custom rules
Logging:     Enabled
```

Ensures crawlers can always fetch robots.txt and sitemaps regardless of any other rules. Also covers llms.txt (machine-readable site description for AI tools) and favicon.ico.

Rule 3: Geo-Block (Unchanged)

```
Expression:  (ip.geoip.country in {"CN" "IR" "KP" "RU" "SY"})
Action:      Managed challenge
```

Existing rule, unchanged. Presents a challenge to traffic from high-risk countries.

Rule 4: Block Probe Patterns Under API Paths (New)

```
Expression:  starts_with(http.request.uri.path, "/api/") and
             not (http.request.uri.path eq "/api/preview" or
                  http.request.uri.path eq "/api/exit-preview") and
             ((http.request.uri.path contains ".env") or
              (http.request.uri.path contains ".git") or
              (http.request.uri.path contains ".php") or
              ... 20+ additional patterns)
Action:      Block
```

Blocks vulnerability probes targeting `/api/` paths while explicitly allowing the two legitimate Prismic CMS preview endpoints (`/api/preview` and `/api/exit-preview`). Ported from the equivalent rule on esolia.co.jp.

Rule 5: Allow Prismic Preview API Endpoints (Updated)

```
Expression:  (http.host eq "japanactivationcapital.com" or
              ends_with(http.host, ".japanactivationcapital.com")) and
             (http.request.uri.path eq "/api/preview" or
              http.request.uri.path eq "/api/exit-preview")
Action:      Skip → http_ratelimit, http_request_firewall_managed, http_request_sbfm
             + remaining custom rules
Logging:     Enabled
```

Previously a generic "Allow API Endpoints" rule that skipped all `/api/` paths. Narrowed to only the two endpoints that actually exist. The `ends_with` pattern covers both the production domain and `staging.japanactivationcapital.com`.

Rule 6: Block Various Probes (New)

```
Expression:  (http.request.uri.path contains ".php") or
             (http.request.uri.path contains "wp-admin") or
             (http.request.uri.path contains "wp-login") or
             (http.request.uri.path contains ".env") or
             (http.request.uri.path contains ".git/config") or
             ... 60+ additional patterns
Action:      Block
```

Full probe-blocking rule ported from esolia.co.jp. Blocks requests for WordPress paths, PHP files, environment files, credential files, debug endpoints, development tool paths, and other common scanner targets. This is the same rule set that has been running on the eSolia zone since mid-February with no false positives.

Full list of blocked patterns includes: `.php`, `wp-admin`, `wp-login`, `wp-content`, `wp-includes`, `wp-json`, `wp-sitemap`, `wp-config`, `xmlrpc`, `.env`, `.aws/`, `.git/config`, `.git-credentials`, `.docker`, `.netrc`, `.npmrc`, `credentials.json`, `/actuator`, `/configprops`, `/artisan`, `/_profiler`, `/_wdt`, `/_ignition`, `/_debugbar`, `/telescope`, `/horizon`, `/@vite`, `/.vite`, `/webpack-dev-server`, `/.vscode`, `/__debug__`, `/server-status`, `/server-info`, `docker-compose`, `Dockerfile`, `/api/v1/namespaces`, `/autodiscover`, `/graphql`, `/__cve_probe`, `secrets.*`, `sendgrid`, `phpinfo`, `aws-credentials`, `cgi-bin`, `ckeditor`, `ueditor`, `terraform`, `WEB-INF`, `heapdump`, `elmah.axd`, `master.key`, `debug/pprof`, `laravel-filemanager`, `finchat`, `/_next/`, `__env`, `env.txt`, `env.prod`, `environment.ts`, `webpack.config`, `vite.config`, `/ip`.

Rule 7: Block Specified User Agents (Unchanged)

```
Expression:  (http.user_agent contains "sqlmap") or
             (http.user_agent contains "nikto") or
             ... additional scanner user agents, plus empty user-agent
Action:      Block
```

Existing rule, unchanged. Blocks known vulnerability scanner user agents.

Rule 8: Block AI Scraper Bots (Updated)

```
Expression:  (http.request.uri.path ne "/robots.txt") and
             (http.request.uri.path ne "/llms.txt") and
             ((http.user_agent contains "CCBot") or
              (http.user_agent contains "Bytespider") or
              ... additional AI training-only user agents)
Action:      Block
```

Blocks bots used solely for AI model training — not search. GPTBot (OpenAI) and ClaudeBot (Anthropic) were subsequently removed from this rule because they power AI search answers (ChatGPT, Claude). Bots that remain blocked (CCBot, Google-Extended, Bytespider, etc.) are training-only crawlers that don't surface content in search results. Search engine crawlers (Bingbot, Googlebot) are not listed here and pass through normally.

Rules Removed

“Allow Authenticated APIs” — This rule checked for `auth_token` cookies on POST requests to the JAC domain. The JAC site is a read-only content presentation site with no authenticated API endpoints, no forms that write data, and no user login. The rule had no effect and was removed.

“Allow AI access to llms.txt” — This standalone skip rule for `/llms.txt` was merged into Rule 2 (Allow SEO-Critical Files) for cleaner organization.

“Allow Bingbot” — The overly narrow Bingbot-only skip rule was replaced by Rule 1, which covers all verified bots.

Staging Environment

The staging site at `staging.japanactivationcapital.com` is protected by Cloudflare Zero Trust Access (login required). The WAF rules apply to it as well:

- Rule 5 (Prismic preview) uses `ends_with(http.host, ".japanactivationcapital.com")`, so preview functionality works on staging
- Probe-blocking rules protect staging the same as production
- The Zero Trust Access policy is a separate layer — it requires authentication before any content is served, providing additional protection beyond the WAF

Early Results

Within the first hour after deployment, the Cloudflare firewall event logs already showed the rules in action. The following data covers the first 6 hours post-deployment (via Cloudflare GraphQL Analytics).

Firewall Actions Summary

Action	Source	Count
Block	Custom rules (our probe blocks)	341
Block	Managed ruleset (Cloudflare built-in)	338
Skip	Custom rules (verified bots + SEO files)	123
Managed challenge	Custom rules (geo-block)	83
Block	Hot (Cloudflare threat intel)	10

679 probe requests blocked in 6 hours — all of these would have previously reached the Worker and consumed invocations.

Probes Caught by Our Custom Rules

Path	Count	What it is
`/autodiscover/autodiscover.xml`	3	Exchange/Outlook autodiscover probe
`/.well-known/acme-challenge/plugins.php`	3	PHP injection attempt
`/.tmb/plugins.php`	3	PHP injection attempt
`/wp-includes/*/plugins.php`	9+	WordPress plugin probes (multiple variants)
`/wp-content/themes/plugins.php`	3	WordPress theme probe

Cloudflare's managed ruleset independently caught WordPress `wlwmanifest.xml` probes across multiple path variants (12+ blocks). The two rulesets are complementary — our custom rules catch application-specific patterns that the managed rules don't, and vice versa.

Verified Bots Passing Through

The skip rules are working correctly — 123 verified bot requests (Googlebot, Bingbot, etc.) and SEO-critical file requests passed through without being challenged or blocked.

Worker Invocation Logs

Cross-referencing with Workers observability telemetry, the requests reaching the Worker in the last hour were clean: legitimate page views (`/` , `/en/` , `/en/newsroom/` , `/team/*`), SvelteKit data fetches (`__data.json`), and standard assets (`apple-touch-icon.png`). No `.php` , `.env` , `wp-admin` , or other probe paths appeared in the Worker logs after the rules took effect.

Current Status

- **Google: fully operational.** Googlebot continues to index the site without interruption. Google Search Console shows clean crawl data with no errors — exactly as it has throughout the migration and WAF hardening work. Google has never been affected by any of these changes.
- **Bingbot: access verified, indexing still stalled.** We have confirmed via Workers observability logs and WAF skip rule hits that Bingbot is now reaching the site without any obstruction. The WAF issue on our side is resolved. However, Bing has not resumed indexing. This is particularly frustrating because Bing Webmaster Tools now reports **zero errors** — all previous issues flagged by Bing (duplicate H1 tags, meta descriptions too short, missing descriptions, etc.) have been addressed and resolved. Every diagnostic test in their UI passes, the site scan tool returns all-green results, and there are no "cannot be indexed" warnings remaining. The site's `robots.txt` and `sitemap.xml` are valid and accessible. There is simply no technical reason on our side for the indexing failure. This is consistent with the industry-wide Bing indexing problems described in the Background section — clean diagnostics, zero errors, yet no indexing. It seems improbable that a company the size of Microsoft would ship a search engine with such fundamental reliability issues, but the evidence across the industry is clear. Attempts to open a support ticket with Microsoft have been unsuccessful — their automated support system loops through the same intake questions repeatedly without creating a ticket. This matches the broader pattern of Bing support issues. We will continue to monitor and attempt to escalate, but this is on Microsoft's side, not ours.
- **Vulnerability scanner traffic** is being blocked at the Cloudflare edge, as confirmed by the early results above
- **No impact on legitimate visitors** — all rules target bot traffic, probe paths, or scanner user agents

Contact Us

eSolia Inc. Shiodome City Center 5F (Workstyling) 1-5-2 Higashi-Shimbashi, Minato-ku Tokyo 105-7105, Japan

Phone	03-4577-3380
Email	rick.cogley@esolia.co.jp
Web	https://esolia.co.jp/en
Hours	Monday-Friday, 9:00-18:00 JST

WAF 強化レポート: Bing インデックス問題の修正とプローブブロック

宛先

ジャパン・アクティベーション・キャピタル

2026 年 3 月 11 日

目次

概要

作成: 株式会社イソリア
日付: 2026 年 3 月 6 日
サイト: japanactivationcapital.com
ゾーン: Cloudflare Pro プラン
ステータス: 完了

japanactivationcapital.com は、すべての公開ウェブサイトと同様に、自動化された攻撃に常にさらされています。脆弱性スキャナー、認証情報収集ボット、ボットネットワークが継続的にサイトを探査し、設定ファイルの露出、WordPress のインストール、デバッグエンドポイント、その他の攻撃対象を探しています。Cloudflare Workers への移行（移行レポート参照）により、リアルタイムのオブザーバビリティが初めてこのトラフィックの規模を可視化しました。Cloudflare Pages では見えなかった悪意のあるリクエストが、毎時数百件発生していました。

本レポートでは、これに対して実施した 2 つの対策を報告します。

1. **WAF をプローブブロックルールで強化。** イソリアの実績あるルールセットを JAC ゾーンに移植し、Cloudflare のマネージドルールセットの上に第 2 の防御層を追加しました。両層あわせて、デプロイ後最初の 6 時間で 679 件のプローブリクエストをブロックしました。これらはすべて以前は Worker に到達し、呼び出し回数を消費していたものです。
2. **当方が導入した Bingbot アクセス問題を修正。** セキュリティ強化の過程で、検索エンジンクローラー用の WAF スキップルールの範囲が狭すぎたことが判明し、Bingbot がチャレンジを受ける一因となっていました。ルールを拡張し、Workers ログで修正を検証し、Bingbot が妨害なくアクセスできることを確認しました。Googlebot は一切影響を受けていません。Google はこの一連の作業を通じて、サイトのインデックスを中断なく継続しています。

背景

Cloudflare のデフォルト保護

JAC の Cloudflare ゾーン（Pro プラン）にはマネージドルールセットが含まれており、設定不要で自動的に稼働します。Cloudflare の脅威インテリジェンスが既知の攻撃パターン、ボットシグネチャ、悪意のあるペイロードを検出・ブロックします。今回のデプロイ後 6 時間のサンプルでは、マネージドルールセットが単独で 338 件のリクエストをブロックしました。1 週間に換算すると数万件規模のプローブをブロックしている計算になります。本レポートに記載するカスタムルールは、これらのデフォルト機能の上に追加する層であり、JAC サイトの技術スタックに固有のパターンを対象としています。

Google と Bing の明確な対比

Googlebot は JAC サイトで一度も問題が発生していません。 Workers 移行前、移行中、WAF 変更後を通じて、Google は一貫してサイトをインデックスし続けています。Google Search Console にはクロールエラーがなく、クリーンなデータが表示されています。適切に設定されたサイトとして、これが期待される動作です。

一方、Bingbot にはインデックス中断の歴史があり、診断が困難な既知の問題です。これは JAC サイト固有の問題ではなく、業界全体のパターンです。イソリアは他のクライアントサイトでも同様の Bing インデックス問題を経験しています。Cloudflare コミュニティフォーラムだけでも複数のスレッドが存在しています。Microsoft 自身の Q&A フォーラムにも、Bing がサイトをインデックスしないという報告が曖昧なエラーメッセージとともに多数寄せられ、明確な解決策が示されないケースが目立ちます。Cloudflare コミュニティで特定された原因の 1 つは、Bing 自身の検証ツールではボット IP が正当と確認されるにもかかわらず、リバース DNS レコードが Cloudflare の期待と一致しない場合があること — これは Microsoft 側のバグであり、Microsoft が PTR レコードを更新すると自然に解消されます。

JAC サイトは以前のホスト（Netlify）でも Bing のインデックス問題を経験しています。その際はサーバーサイドの可視性がまったくなく、最終的に Microsoft に直接エスカレーションしました。チケットクローズ時の Microsoft の回答は「詳細は提供できない」というものでした。

今回の問題 — Cloudflare のドキュメントでは検証済みボットは除外されるとされているにもかかわらず、SBFM が Bingbot に干渉した件 — も同じパターンの一例です。Cloudflare 自身の SBFM に関するコミュニティチュートリアルでも「セキュリティ閾値が高いため、誤検知が発生する場合がある」と認めています。修正（`cf.client.bot` の明示的なスキップルール）は単純ですが、Microsoft 側の根底にある脆弱性のため、今後も当方の制御範囲外の理由で Bing のインデックスが停止する可能性があります。Bing Webmaster Tools と Workers のオブザーバビリティログで引き続き監視します。

特定された問題

1. Bingbot スキップルールの範囲が狭すぎた

既存のルールは以下の式を使用していました。

```
(cf.client.bot and http.user_agent contains "bingbot")
```

これには 2 つの問題があります。

- **冗長:** `cf.client.bot` は Cloudflare の検証済みボットフィールドで、リバース DNS ルックアップにより確認されたボット（Googlebot、Bingbot、Yandex 等）のみ `true` を返します。その上にユーザーエージェントの確認を追加するのは不要であり、Bing がユーザーエージェント文字列を変更した場合に不一致が生じる可能性があります。
- **範囲が狭い:** Bingbot のみが対象でした。Googlebot やその他の正規の検索クローラーはスキップルールに含まれておらず、Super Bot Fight Mode (SBFM) によりチャレンジやブロックを受ける可能性がありました。

esolia.co.jp で正常に機能しているルールは `(cf.client.bot)` のみを使用しており、ユーザーエージェントフィルターはありません。

2. SEO 重要ファイルの保護がなかった

`/robots.txt`、`/sitemap.xml`、`/favicon.ico` に対するスキップルールがありませんでした。他のファイアウォールルールやレート制限が発動した場合、クローラーがインデックスに必要なファイルの取得をブロックされる可能性がありました。

`/llms.txt` のスキップルールは存在しましたが、標準的な SEO ファイルはカバーしていませんでした。

3. プローブブロックがなかった

JAC ゾーンには一般的な脆弱性プローブをブロックするルールがありませんでした。eSolia ゾーンでは Workers ログで観測された実際のトラフィックに基づいたプローブブロックルールを 2 月中旬から運用しています。これらのルールがなければ、すべてのプローブリクエストが Worker に到達し、呼び出し回数を消費してログを汚染します。

4. 未使用のルール

「Allow Authenticated APIs（認証済み API の許可）」ルール（POST リクエストの `auth_token` Cookie を確認）が存在しましたが、JAC サイトには認証済み API エンドポイントがないため、このルールは機能していませんでした。

実施した変更

カスタムファイアウォールルールセットを以下の 8 ルールに置き換えました。ルールの順序は重要です。スキップルールはブロックルールより前に配置し、先に評価されるようにしています。

ルール 1: 検証済みボットの SBFM スキップ

```
式:        (cf.client.bot)
アクション: スキップ → http_request_sbfm
ログ:      有効
```

Cloudflare が検証したすべての検索ボット（Bingbot、Googlebot、Yandex 等）が Super Bot Fight Mode をバイパスできるようにします。これが Bing インデックス問題の主要な修正です。

ルール 2: SEO 重要ファイルの許可

```
式:        (http.request.uri.path eq "/robots.txt") or
           (http.request.uri.path eq "/sitemap.xml") or
           (http.request.uri.path eq "/sitemap-index.xml") or
           (http.request.uri.path eq "/favicon.ico") or
           (http.request.uri.path eq "/llms.txt")
アクション: スキップ → http_ratelimit, http_request_firewall_managed,
           http_request_sbfm + 残りのカスタムルール
ログ:      有効
```

他のルールに関係なく、クローラーが常に robots.txt やサイトマップを取得できることを保証します。llms.txt（AI ツール向けの機械可読サイト説明）と favicon.ico もカバーします。

ルール 3: 地理ブロック（変更なし）

```
式:        (ip.geoip.country in {"CN" "IR" "KP" "RU" "SY"})
アクション: マネージドチャレンジ
```

既存のルールで変更なし。高リスク国からのトラフィックにチャレンジを提示します。

ルール 4: API パス配下のプローブパターンのブロック（新規）

```
式:        starts_with(http.request.uri.path, "/api/") and
           not (http.request.uri.path eq "/api/preview" or
                http.request.uri.path eq "/api/exit-preview") and
           ((http.request.uri.path contains ".env") or
            (http.request.uri.path contains ".git") or
            (http.request.uri.path contains ".php") or
            ... 20以上の追加パターン)
アクション: ブロック
```

`/api/` パスを狙う脆弱性プローブをブロックしつつ、2 つの正規の Prismic CMS プレビューエンドポイント（`/api/preview` と `/api/exit-preview`）を明示的に許可します。esolia.co.jp の同等ルールから移植しました。

ルール 5: Prismic プレビュー API エンドポイントの許可（更新）

```
式:        (http.host eq "japanactivationcapital.com" or
            ends_with(http.host, ".japanactivationcapital.com")) and
           (http.request.uri.path eq "/api/preview" or
            http.request.uri.path eq "/api/exit-preview")
アクション: スキップ → http_ratelimit, http_request_firewall_managed,
           http_request_sbfm + 残りのカスタムルール
ログ:      有効
```

以前はすべての `/api/` パスをスキップする汎用的な「Allow API Endpoints」ルールでしたが、実際に存在する 2 つのエンドポイントのみに絞りました。`ends_with` パターンにより、本番ドメインと `staging.japanactivationcapital.com` の両方をカバーします。

ルール 6: 各種プローブのブロック（新規）

```
式:        (http.request.uri.path contains ".php") or
           (http.request.uri.path contains "wp-admin") or
           (http.request.uri.path contains "wp-login") or
           (http.request.uri.path contains ".env") or
           (http.request.uri.path contains ".git/config") or
           ... 60以上の追加パターン
アクション: ブロック
```

esolia.co.jp から移植したプローブブロックルールです。WordPress パス、PHP ファイル、環境ファイル、認証情報ファイル、デバッグエンドポイント、開発ツールパスなどのスキャナー標的をブロックします。eSolia ゾーンで 2 月中旬から運用しており、誤検知はありません。

ブロック対象のパターン一覧: `.php`、`wp-admin`、`wp-login`、`wp-content`、`wp-includes`、`wp-json`、`wp-sitemap`、`wp-config`、`xmlrpc`、`.env`、`.aws/`、`.git/config`、`.git-credentials`、`.docker`、`.netrc`、`.npmrc`、`credentials.json`、`/actuator`、`/configprops`、`/artisan`、`/_profiler`、`/_wdt`、`/_ignition`、`/_debugbar`、`/telescope`、`/horizon`、`/@vite`、`/.vite`、`/webpack-dev-server`、`/.vscode`、`/__debug__`、`/server-status`、`/server-info`、`docker-compose`、`Dockerfile`、`/api/v1/namespaces`、`/autodiscover`、`/graphql`、`/__cve_probe`、`secrets.*`、`sendgrid`、`phpinfo`、`aws-credentials`、`cgi-bin`、`ckeditor`、`ueditor`、`terraform`、`WEB-INF`、`heapdump`、`elmah.axd`、`master.key`、`debug/pprof`、`laravel-filemanager`、`finchat`、`/_next/`、`__env`、`env.txt`、`env.prod`、`environment.ts`、`webpack.config`、`vite.config`、`/ip`。

ルール 7: 指定ユーザーエージェントのブロック（変更なし）

```
式:        (http.user_agent contains "sqlmap") or
           (http.user_agent contains "nikto") or
           ... その他のスキャナーユーザーエージェント、および空のユーザーエージェント
アクション: ブロック
```

既存のルールで変更なし。既知の脆弱性スキャナーのユーザーエージェントをブロックします。

ルール 8: AI スクレイパーボットのブロック（更新）

```
式:        (http.request.uri.path ne "/robots.txt") and
           (http.request.uri.path ne "/llms.txt") and
           ((http.user_agent contains "CCBot") or
            (http.user_agent contains "Bytespider") or
            ... その他の AI 学習専用ユーザーエージェント)
アクション: ブロック
```

AI モデル学習専用のボットをブロックします。GPTBot（OpenAI）と ClaudeBot（Anthropic）は AI 検索の回答（ChatGPT、Claude）でコンテンツを表示するため、このルールから除外しました。引き続きブロック対象のボット（CCBot、Google-Extended、Bytespider 等）は、検索結果にコンテンツを表示しない学習専用クローラーです。検索エンジンクローラー（Bingbot、Googlebot）はここに記載されておらず、通常通り通過します。

削除されたルール

「Allow Authenticated APIs（認証済み API の許可）」 — JAC ドメインへの POST リクエストで `auth_token` Cookie を確認するルール。JAC サイトは読み取り専用のコンテンツ配信サイトであり、認証済み API エンドポイント、データ書き込みフォーム、ユーザーログインはありません。機能していなかったため削除しました。

「Allow AI access to llms.txt（AI の llms.txt アクセス許可）」 — `/llms.txt` の単独スキップルール。整理のためルール 2（SEO 重要ファイルの許可）に統合しました。

「Allow Bingbot（Bingbot の許可）」 — 範囲の狭い Bingbot 専用スキップルール。すべての検証済みボットをカバーするルール 1 に置き換えました。

ステージング環境

`staging.japanactivationcapital.com` のステージングサイトは Cloudflare Zero Trust Access で保護されています（ログインが必要）。WAF ルールはステージングにも適用されます。

- ルール 5（Prismic プレビュー）は `ends_with(http.host, ".japanactivationcapital.com")` を使用しているため、ステージングでもプレビュー機能が動作します
- プローブブロックルールは本番と同様にステージングを保護します
- Zero Trust Access ポリシーは別レイヤーとして機能し、コンテンツ配信前に認証を要求するため、WAF を超えた追加の保護を提供します

初期結果

デプロイ後 1 時間以内に、Cloudflare のファイアウォールイベントログでルールが機能していることを確認しました。以下はデプロイ後最初の 6 時間のデータです（Cloudflare GraphQL Analytics 経由）。

ファイアウォールアクションの概要

アクション	ソース	件数
ブロック	カスタムルール（プローブブロック）	341
ブロック	マネージドルールセット（Cloudflare 組み込み）	338
スキップ	カスタムルール（検証済みボット + SEO ファイル）	123
マネージドチャレンジ	カスタムルール（地理ブロック）	83
ブロック	Hot（Cloudflare 脅威インテリジェンス）	10

6 時間で **679 件のプローブリクエストをブロック** しました。これらはすべて以前は Worker に到達し、呼び出し回数を消費していたものです。

カスタムルールで検出されたプローブ

パス	件数	内容
`/autodiscover/autodiscover.xml`	3	Exchange/Outlook 自動検出プローブ
`/.well-known/acme-challenge/plugins.php`	3	PHP インジェクション試行
`/.tmb/plugins.php`	3	PHP インジェクション試行
`/wp-includes/*/plugins.php`	9+	WordPress プラグインプローブ（複数パターン）
`/wp-content/themes/plugins.php`	3	WordPress テーマプローブ

Cloudflare のマネージドルールセットは WordPress の `wlwmanifest.xml` プローブを複数のパスパターンで独立してブロックしました（12 件以上）。2 つのルールセットは相互補完的に機能しています。カスタムルールはマネージドルールが検出しないアプリケーション固有のパターンを検出し、その逆も同様です。

検証済みボットの通過確認

スキップルールは正常に機能しています。123 件の検証済みボットリクエスト（Googlebot、Bingbot 等）と SEO 重要ファイルへのリクエストが、チャレンジやブロックを受けずに通過しました。

Worker 呼び出しログ

Workers オブザーバビリティのテレメトリデータと照合すると、直近 1 時間に Worker に到達したリクエストはすべて正当なものでした。ページビュー（`/`、`/en/`、`/en/newsroom/`、`/team/*`）、SvelteKit のデータフェッチ（`__data.json`）、標準的なアセット（`apple-touch-icon.png`）のみ。ルール適用後は `.php`、`.env`、`wp-admin` などのプローブパスは Worker ログに記録されていません。

現在の状況

- **Google: 完全に正常稼働。** Googlebot はサイトのインデックスを中断なく継続しています。Google Search Console にはクロールエラーがなく、クリーンなデータが表示されています。Workers 移行および WAF 強化作業を通じて、Google は一度も影響を受けていません。
- **Bingbot: アクセスは検証済み、インデックスは停滞中。** Workers オブザーバビリティログと WAF スキップルールのヒット記録により、Bingbot が妨害なくサイトにアクセスしていることを確認しました。当方の WAF に起因する問題は解決済みです。しかし、Bing はインデックスを再開していません。特に不可解なのは、Bing Webmaster Tools で**エラーがゼロ**になっていることです。Bing が以前指摘していたすべての問題（重複する H1 タグ、メタディスクリプションの文字数不足、説明の欠落等）は対処済みです。UI 上のすべての診断テストに合格、サイトスキャンツールはすべてグリーン（問題なし）を返し、「インデックス不可」の警告も残っていません。`robots.txt` と `sitemap.xml` は有効でアクセス可能です。当方側にインデックス失敗の技術的原因はありません。これは「背景」セクションで説明した業界全体の Bing インデックス問題と同じパターンです。診断はクリーン、エラーはゼロ、にもかかわらずインデックスされない状況です。Microsoft ほどの規模の企業がこれほど基本的な信頼性の問題を抱えた検索エンジンを運用しているとは考えにくいですが、業界全体のエビデンスは明白です。Microsoft へのサポートチケット作成を試みましたが、自動サポートシステムが同じ質問を繰り返しループするだけでチケットが作成されません。これはより広範な Bing サポートの問題と一致するパターンです。引き続き監視とエスカレーションを試みますが、これは Microsoft 側の問題であり、当方の問題ではありません。
- **脆弱性スキャナーのトラフィック** は上記の初期結果で確認された通り、Cloudflare エッジでブロックされています
- **正規の訪問者への影響なし** — すべてのルールはボットトラフィック、プローブパス、またはスキャナーのユーザーエージェントを対象としています

お問い合わせ

株式会社イソリア 〒105-7105 東京都港区東新橋 1-5-2 汐留シティセンター 5 階 (Workstyling)

電話	03-4577-3380
メール	rick.cogley@esolia.co.jp
Web	https://esolia.co.jp
営業時間	月～金、9:00～18:00