技術資料

Site Safety Checker

URL安全性チェッカー — AI多角分析＋6軸レーダーチャート

作成日	2026年2月12日
バージョン	1.2（事例調査手法追加版）
開発者	特定非営利活動法人リハビリコラボレーション

目次

1. プロジェクト概要
2. システムアーキテクチャ
3. 画面遷移フロー
4. 分析フロー
5. 6軸スコアリングアルゴリズム
6. 検出対象カテゴリ（19種 + 7種）
7. 感度調整アルゴリズム
8. クライアント側URL構造分析
9. Gemini AIプロンプト設計
10. 事例調査手法と活用方法
11. セキュリティ設計
12. 技術仕様
13. ファイル構成

1. プロジェクト概要

1.1 目的

Site Safety Checkerは、URLを入力するだけでウェブサイトの安全性をAIで多角的に分析するツールです。詐欺サイト・危険サイトの特徴パターンを検出し、6軸レーダーチャートで信頼性を可視化します。

1.2 対象ユーザー

- 怪しいURLを受け取った一般ユーザー（SNS、メール、SMS等）
- 通販サイトの安全性を確認したい消費者
- 投資・副業案内の信頼性を判断したい方

1.3 主要機能

機能	説明
URL分析	URLを入力→クライアント側構造分析＋AI分析で総合評価
テキスト分析	ページ内容をコピペ→AI分析（ログイン必要ページ等に対応）
6軸レーダーチャート	信頼性を6次元で可視化（Canvas描画）
19カテゴリ詐欺検出	闇バイト・投資詐欺・フィッシング等の既知パターンマッチ
7種広告規制違反検出	景表法・薬機法等の横断的な法令違反チェック
3段階感度調整	高感度/標準/低感度でスコアリング閾値を調整
BYOK（Bring Your Own Key）	ユーザー自身のGemini APIキーを使用、開発者サーバー不要

> 本ツールの分析結果はAIによる参考情報であり、サイトの安全性を保証するものではありません。

2. システムアーキテクチャ

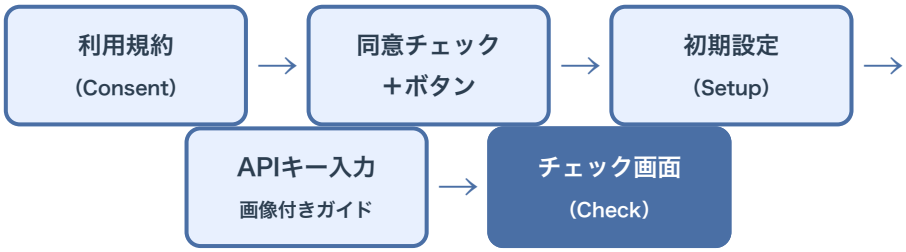
2.1 全体構成

2.2 技術スタック

レイヤー	技術	詳細
フロントエンド	Vanilla HTML/CSS/JS	外部依存なし、IIFE パターン
チャート描画	Canvas API	Retina対応6軸レーダーチャート自前実装
プロキシ	Cloudflare Workers	エッジコンピューティング、グローバル配信
AI分析	Gemini 2.5 Flash	構造化出力（JSON Schema）、BYOK
ホスティング	GitHub Pages	HTTPS自動対応、noindex

3. 画面遷移フロー

3.1 初回利用フロー

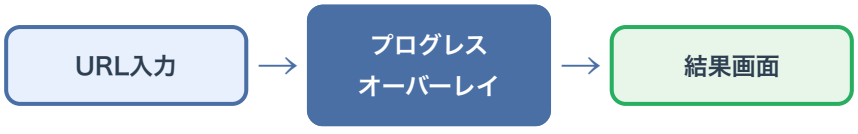


3.2 2回目以降

3.3 全画面遷移図

画面	ID	遷移先	トリガー
利用規約	screenConsent	Setup or Check	「同意して始める」ボタン
初期設定	screenSetup	Check	「保存して始める」ボタン
チェック	screenCheck	Results / Settings	「チェック」or 歯車ボタン
結果	screenResults	Check	「別のURLをチェック」ボタン
設定	screenSettings	Check / Consent	「保存」「戻る」「利用規約を表示」

3.4 分析中フロー



プログレス中は「キャンセル」ボタンで中断可能

4. 分析フロー

4.1 URL分析モード

ステージ	進捗	処理内容	タイムアウト
1. URL構造分析	5%	クライアント側でTLD、ブランド偽装、IDN等をチェック	即座
2. サイト取得	15%	Worker経由でHTML+ヘッダー取得	15秒
3. コンテンツ解析	35%	DOMParserでテキスト/フォーム/スクリプト抽出	即座
4. AI分析	55%	Gemini APIで6軸スコア＋詐欺パターン分析	60秒
5. 結果統合	95%	クライアント＋AIスコアブレンド、リスクレベル判定	即座

4.2 部分結果の段階的表示

各ステージが失敗しても、それ以前のステージの結果を表示します。

失敗箇所	表示される結果
Worker取得失敗	URL構造分析のみ（ドメイン信頼性＋技術的安全性）
AI分析失敗	URL構造分析＋HTML抽出結果（4次元はデフォルト50点）
全成功	6軸完全スコア＋詐欺カテゴリ＋詳細所見＋AI総評

4.3 テキスト分析モード

ログイン必要ページ等、Worker経由で取得できない場合に使用。ページ内容をコピペして直接AI分析に送信します。

5. 6軸スコアリングアルゴリズム

5.1 評価次元

軸	キー	評価内容	スコア源
ドメイン信頼性	domain_trust	URL構造、TLD、ブランド偽装、SSL	クライアント40% + AI 60%
コンテンツ安全性	content_safety	不審キーワード、煽り表現、緊急性	AI 100%
運営者透明性	operator_transparency	特商法表記、会社概要、連絡先	AI 100%
主張の信頼性	claim_credibility	誇大広告、非現実的保証	AI 100%
詐欺パターン非合致	scam_pattern	19カテゴリの既知パターンとの非類似度	AI 100%
技術的安全性	tech_safety	SSL、難読化スクリプト、隠しフォーム	クライアント40% + AI 60%

5.2 スコア統合アルゴリズム

各軸は0〜100のスコアで表現されます。

```
// ブレンド比率
domain_trust  = clientScore * 0.4 + aiScore * 0.6
tech_safety   = clientScore * 0.4 + aiScore * 0.6
他4次元       = aiScore * 1.0

// 総合平均
average = (domain_trust + content_safety + operator_transparency
         + claim_credibility + scam_pattern + tech_safety) / 6

// リスクレベル判定
average >= 80  → safe     (安全)
average >= 60  → low      (低リスク)
average >= 40  → medium   (中リスク)
average >= 20  → high     (高リスク)
average <  20  → critical (危険)

// エスカレーションルール（感度依存）
criticalDims = 閾値以下の次元数
if (criticalDims >= 2 || scam_pattern <= 閾値) → 最低 high
if (criticalDims == 1) → safe の場合 low にナッジ
```

5.3 リスクレベル表示

レベル	アイコン	色	説明
safe	✅	#27AE60	特に問題は検出されませんでした
low	🟢	#8BC34A	軽微な注意点があります
medium	⚠️	#F39C12	複数の注意点が見つかりました
high	🚨	#E74C3C	詐欺の可能性があります
critical	⛔	#C0392B	非常に危険なサイトです

6. 検出対象カテゴリ

6.1 詐欺・違法サイトカテゴリ（19種）

#	カテゴリ	主な検出パターン
1	闇バイト	「高額報酬」「即日払い」「簡単作業」、Telegram/Signal誘導
2	投資詐欺	「元本保証」「年利30%」、金融庁無登録業者、ポンジスキーム
3	フィッシング	ブランド偽装URL、ログインフォーム、個人情報入力要求
4	偽通販	極端な割引、会社概要不備、不自然な日本語
5	健康詐欺	薬機法違反、「確実に治る」、医療広告規制違反
6	被害回復詐欺	「被害金を取り戻せます」、二次被害
7	サポート詐欺	偽Microsoft警告、遠隔操作ソフト導入要求
8	ロマンス詐欺	出会い系、軍人・医師なりすまし、送金要求
9	違法オンラインカジノ	日本向け違法賭博サイト
10	偽造品・コピー品販売	ブランドコピー、スーパーコピー
11	架空請求・ワンクリック詐欺	「登録完了」「退会費用」
12	副業・タスク詐欺	「スマホで月50万」、初期費用要求
13	なりすまし広告詐欺	著名人の偽広告、AIディープフェイク
14	仮想通貨・暗号資産詐欺	偽取引所、ICO詐欺、ウォレット接続要求
15	情報商材詐欺	「秘密のノウハウ」、高額コンサル
16	闇金・違法貸金業	「ブラックOK」「審査なし」、貸金業登録番号偽造
17	著作権侵害・海賊版	漫画・映画・ソフトの違法配信
18	還付金詐欺・偽行政サイト	偽マイナポータル、還付金手続き
19	霊感商法・疑似科学	スピリチュアル詐欺、水素水、マイナスイオン、EM菌

6.2 広告・表示規制違反（7種）

#	違反パターン	関連法令
1	根拠なきNo.1表示	景品表示法（優良誤認）

2	効果のない商品の効能表示	薬機法・景表法
3	二重価格表示	景品表示法（有利誤認）
4	ステルスマーケティング	景品表示法（2023年10月〜）
5	打消し表示の不備	景品表示法
6	定期購入の表示不備	特定商取引法
7	グリーンウォッシュ	景品表示法

7. 感度調整アルゴリズム

7.1 3段階設定

設定	用途	特徴
高感度	疑わしいサイトを見逃さない	誤検知（偽陽性）が多くなる
標準（推奨）	バランスの取れた判定	デフォルト設定
低感度	誤検知を減らす	見逃し（偽陰性）が多くなる

7.2 閾値パラメータ

パラメータ	高感度	標準	低感度	効果
criticalDim閾値	20	15	10	この値以下の次元をcriticalと判定
warnDim閾値	35	30	20	この値以下の次元をwarnと判定
scamPattern閾値	35	30	20	詐欺パターン軸の警戒閾値

7.3 プロンプト補正

感度設定に応じてGemini APIへのプロンプトに指示を追加します。

設定	追加プロンプト
高感度	「少しでも疑わしい点がある場合は積極的に低いスコアを付けてください。安全側に判定を誤るより、危険側に誤る方が望ましいです。」
標準	（追加なし）
低感度	「明確な根拠がある場合のみ低いスコアを付けてください。曖昧な場合は高めのスコアにしてください。」

8. クライアント側URL構造分析

8.1 分析項目と減点

チェック項目	減点（domain_trust）	減点（tech_safety）
HTTP（SSL未使用）	—	-30
IPアドレスURL	-40	—

不審なTLD（.xyz, .top等 28種）	-20	—
過剰サブドメイン（5階層以上）	-15	—
ブランドタイポスクワッティング（24社）	-30	—
IDNホモグラフ攻撃	-25	—
過剰ハイフン（4つ以上）	-10	—
異常なポート番号	—	-15
過剰パス深度（6以上）	-10	—
不審パスキーワード	-15	—

8.2 監視対象ブランド（24社）

Amazon, 楽天, Yahoo, Google, Apple, Microsoft, Facebook, Instagram, Twitter, PayPal, Netflix, docomo, SoftBank, メルカリ, PayPay, SMBC, MUFG, みずほ, ゆうちょ, イオン, ファミリーマート, ローソン, ユニクロ 等

9. Gemini AIプロンプト設計

9.1 プロンプト構造

セクション	位置	内容
回答ルール	冒頭	反ハルシネーション6ルール（LLMが最初に読む位置に配置）
感度指示	冒頭直後	高感度/低感度時のスコアリング方針（標準時は省略）
19カテゴリ定義	中段	各カテゴリの特徴パターン一覧（法令違反を内包）
7種広告規制違反	中段	横断的な法令違反パターン
分析対象データ	末尾	URL、クライアント分析結果、抽出テキスト、ヘッダー情報

9.2 構造化出力（JSON Schema）

Gemini APIの `response_mime_type: "application/json"` と `response_schema` を使用し、確実にパース可能なJSONを取得します。

フィールド	型	説明
scores.*	integer (0-100)	6軸スコア
detected_categories[]	array	検出された詐欺カテゴリ（名前＋信頼度＋説明）
ad_violations[]	array	検出された広告規制違反
findings[]	array	詳細所見（タイトル＋深刻度＋説明）
summary	string	AI総合評価（日本語200文字以内）

9.3 反ハルシネーションルール

1. 分析対象のテキストに**書かれていない内容**を「存在する」と断定しない
2. URLの見た目だけで実在する企業との関連を断定しない
3. 「〜の可能性がある」と「〜である」を明確に区別する
4. 判断材料が不十分な場合は中間スコア（40-60）を付ける
5. 検出カテゴリは**明確な根拠**がある場合のみ追加する
6. 所見には必ず**具体的な根拠**（テキスト引用等）を含める

10. 事例調査手法と活用方法

10.1 調査の目的

本ツールの検出精度は、Gemini AIプロンプトに埋め込まれた**実際の事例データ**に大きく依存します。一般的なキーワードマッチではなく、公的機関が公表した統計・逮捕事例・行政処分・実際の手口フレーズを体系的に収集し、19カテゴリの詐欺パターン定義と7種の広告規制違反パターンに反映しています。

> 調査は7つの専門リサーチプロセスを**並列実行**し、各分野の公的情報源からデータを網羅的に収集しました。

10.2 調査対象機関と収集データ

調査対象機関	収集したデータ	活用先カテゴリ
警察庁（National Police Agency）	特殊詐欺統計（件数・被害額）、闇バイト逮捕事例、賭博事犯検挙数、闇金検挙事件数、還付金詐欺認知件数	Cat1闇バイト、Cat9違法カジノ、Cat16闇金、Cat18還付金詐欺
金融庁（Financial Services Agency）	SNS型投資詐欺被害統計、無登録業者警告リスト（KuCoin/Bybit/MEXC等5社）、金融商品取引法の断定的判断提供禁止規定	Cat2投資詐欺、Cat14仮想通貨詐欺、誤検知防止（正当な金融サービス識別）
消費者庁（Consumer Affairs Agency）	景表法措置命令事例（シボローカ/MIHORE/メラット等）、事業者名公表（協栄商事/和等）、ステマ処分事例（祐真会/RIZAP等6件）、タスク詐欺相談件数	Cat5健康詐欺、Cat12副業詐欺、Cat15情報商材、広告規制違反7種全て
IPA（情報処理推進機構）	偽ウイルス警告相談件数（2024年Q1: 1,385件）、ネットバンキング乗っ取り事例、サポート詐欺手口分析	Cat7サポート詐欺
国民生活センター（PIO-NET）	インターネット通販相談年間27万件、水素水相談2,260件、開運商法相談年間1,200〜1,500件、占いサイト相談年間2,000件超、情報商材相談6,593件	Cat4偽通販、Cat5健康詐欺、Cat15情報商材、Cat19霊感商法・疑似科学
JC3（日本サイバー犯罪対策センター）	偽ECサイト通報47,278件（2023年）、フィッシング通報年間171.8万件、なりすましブランドTop10	Cat3フィッシング、Cat4偽通販
JFC（日本ファクトチェックセンター）	2024年検証330件、AIディープフェイク事例（堀江貴文AI音声→2.2億円被害）、災害偽情報（能登震災）	Cat13なりすまし広告詐欺

厚生労働省	医療広告ガイドライン違反（2023年度: 1,098サイト/6,328件）、あはき法・柔整法違反事例	Cat5健康詐欺・医療広告違反
税関（財務省）	知的財産侵害物品差止33,019件（2024年、過去最多）	Cat10偽造品・コピー品販売
最高裁判所	「給料ファクタリングは貸金業法適用の貸付」判決（2023年）、漫画村賠償額17.3億円（2024年）	Cat16闇金、Cat17著作権侵害

10.3 並列リサーチプロセス

効率的なデータ収集のため、以下の7つの専門リサーチプロセスを同時並行で実行しました。

#	リサーチ領域	収集対象	主要情報源
1	闇バイト・特殊詐欺	首都圏連続強盗事件、Telegram経由の犯罪者募集手口、隠語辞典	警察庁、報道各社
2	投資詐欺・ロマンス詐欺・フィッシング	SNS型投資詐欺統計、有名人偽装事例、フィッシング通報データ、偽装ブランドランキング	金融庁、JC3、フィッシング対策協議会
3	景表法違反事例	措置命令データベース（2022〜2024年）、No.1表示集中取締り14社、ステマ規制6件	消費者庁、公正取引委員会
4	10カテゴリの具体的検出フレーズ	各カテゴリで実際に使用される勧誘文言・脅迫文言・煽り表現の網羅的収集	国民生活センター、IPA、消費者庁注意喚起
5	偽通販・闇金・仮想通貨・情報商材	大型詐欺事件（ジュビリーエース650億円等）、逮捕事例、手口の進化パターン	警察庁、金融庁、消費者庁
6	霊感商法・疑似科学・スピリチュアル詐欺	水素水措置命令、マイナスイオン措置命令、占いサイト被害事例、消費者契約法改正	国民生活センター、消費者庁
7	消費者庁の施策・ファクトチェック	定期購入ダークパターン（年間9万件）、タスク詐欺10億円、JFC検証330件	消費者庁、JFC

10.4 収集データの活用方法

収集した事例データは、Geminiプロンプト内の以下の4つの要素として体系的に組み込まれています。

プロンプト要素	活用方法	具体例
統計データ（冒頭に配置）	各カテゴリの被害規模を示し、AIの判断の重み付けを支援。数値により深刻度を認識させる。	Cat2: 「2024年被害: 10,237件・1,271.9億円（前年比179.4%増）」 Cat8: 「2024年1-9月被害: 271億円、平均被害額1,242.7万円」

検出フレーズ（箇条書きで列挙）	実際の詐欺サイトで使用されている文言をそのまま記載。AIがサイト本文とパターンマッチングする際の参照基準。	Cat1: 「人から物を受け取るだけ」「報酬35万円 資金調達」 Cat12: 「いいねを押すだけ」「スクショを撮るだけ」
逮捕・処分事例（実名+年で記載）	具体的な事業者名と処分内容を含めることで、類似パターンの認識精度を向上。	Cat5: 「シンゲンメディカル社逮捕(2019年)」「ステラ漢方社逮捕(2020年)」 Cat16: 「BPMH社13人逮捕(2025年)、85億円」
手口の構造（フロー形式で記載）	詐欺の段階的な手口を記述し、サイトが手口のどの段階に該当するかをAIが判断できるようにする。	Cat2: 「偽広告→LINEグループ誘導→『先生』が指導→サクラ利益報告→高額投入→連絡断絶」 Cat12: 「SNS広告→LINE登録→少額タスク→高額送金要求→出金不可」

10.5 具体的な活用フロー（例：健康詐欺の検出）

以下に、Cat5（健康詐欺）を例として、調査データがどのように検出に活用されるかを示します。

Step 1: 公的機関の事例収集

消費者庁の措置命令データベースから、健康食品・サプリメントに対する景表法違反事例を収集。
例: シボローカ措置命令（2024年3月）— 「食事制限や運動なしで短期間で痩身効果」と表示するも合理的根拠なし。

▼

Step 2: 検出パターンとしてプロンプトに組込み

措置命令の対象となった表現をプロンプトのCat5検出フレーズに反映。
`「食事制限不要」「飲むだけで」「楽ヤセ」「短期間で薄毛改善」`
併せて処分事例: `「シボローカ/フラボス措置命令(2024年3月)」「MIHORE措置命令(2024年10月)」`

▼

Step 3: AIがサイトを分析

分析対象サイトに「飲むだけで-5kg」「食事制限不要で楽に痩せる」等の表現がある場合、
Gemini AIはプロンプト内の検出フレーズ・措置命令事例と照合し、Cat5（健康詐欺）を検出。
`claim_credibility` と `scam_pattern` に低スコアを付与。

▼

Step 4: 結果レポートに根拠を提示

検出結果の所見（findings）に、サイト本文からの引用と共に
「景品表示法違反の措置命令事例（シボローカ等）と類似する表現パターン」として根拠を提示。

10.6 誤検知防止のための事例活用

収集した事例データは検出だけでなく、**正当なサービスの誤検知防止**にも活用されています。

正当なサービス	誤検知リスク	事例に基づく判別基準
正規の金融サービス	「投資」「利益」等のキーワードでCat2に誤検出	金融庁登録番号の体系（「関東財務局長(金商)第○○号」）が正しいか、金商法に基づくリスク説明があるかで判別
正規の健康食品販売	「健康」「サプリ」等でCat5に誤検出	機能性表示食品届出番号の有無、「個人の感想です」等の免責表記、特商法表記の完備で判別

弁護士の被害回復支援	「被害金回収」でCat6に誤検出	日本弁護士連合会登録番号の有無、振り込め詐欺救済法に基づく法的手続きの範囲内かで判別
正規の占いサービス	「占い」「鑑定」でCat19に誤検出	料金体系の明示、特商法表記の完備、解約方法の明示、「娯楽目的」の位置づけで判別

> 詐欺サイトが正当なサービスの特徴を装うケース（偽の登録番号、コピペされた免責文等）にも対応するため、「運営者情報の有無」より「サービス内容の実現可能性」を重視する判定ルールを設けています。

11. セキュリティ設計

11.1 Worker セキュリティ

対策	詳細
CORS オリジン制限	GitHub Pages + localhost のみ許可。ワイルドカード(＊)廃止
認証	/fetch, /models/* は `X-API-Key` ヘッダー必須
SSRF防止	プライベートIPv4/IPv6、10進整数IP、16進IP、8進表記をすべてブロック
リダイレクト追跡	各ホップでSSRFチェック、最大5回
HTMLサイズ制限	200KB上限
タイムアウト	HTML取得 10秒、Gemini API 60秒
パス検証	`models/[a-zA-Z][\w.-]*:generateContent` のみ許可
リクエストサイズ	Gemini APIリクエスト 500KB上限

11.2 フロントエンドセキュリティ

対策	詳細
APIキー保存	localStorage（BYOK設計の制限、利用規約で明記）
APIキー形式検証	`/^AIza[A-Za-z0-9_-]{35}$/` で形式チェック
XSS防止	DOM出力は `textContent` 使用（innerHTML最小限・ユーザー入力直接挿入なし）
Selector Injection防止	localStorage値のホワイトリスト検証
Worker URL秘匿	デフォルトURLはlocalStorageに保存しない、設定UIでは空欄表示

11.3 SSRF防止の詳細

ブロック対象	例
ローカルホスト	localhost, 127.0.0.1, ::1, 0.0.0.0
プライベートIPv4	10.x.x.x, 172.16-31.x.x, 192.168.x.x
CGNAT	100.64.0.0/10

リンクローカル	169.254.x.x, fe80::*
IPv6 ユニークローカル	fc00::/7, fd00::/8
IPv4マップドIPv6	::ffff:127.0.0.1
10進整数IP	2130706433 (= 127.0.0.1)
16進IP	0x7f000001
8進表記	0177.0.0.1

12. 技術仕様

12.1 デプロイ情報

項目	値
フロントエンドURL	https://mizuking796.github.io/site-safety-checker/
Worker URL	https://site-safety-checker.mizuki-tools.workers.dev
リポジトリ	mizuking796/site-safety-checker（public）
AIモデル	Gemini 2.5 Flash（gemini-2.5-flash）
SEO	noindex + robots.txt Disallow: /

12.2 ブラウザ対応

機能	必要API	対応ブラウザ
コア機能	fetch, Promise, localStorage	Chrome 60+, Safari 12+, Firefox 55+
タイムアウト	AbortSignal.timeout()	Chrome 103+, Safari 16.4+, Firefox 100+
レーダーチャート	Canvas 2D, devicePixelRatio	全モダンブラウザ

12.3 localStorage キー

キー	型	説明
`ssc_config`	JSON	`{ apiKey: string, workerUrl?: string }`
`ssc_consent`	string	ISO日付（例: "2026-02-12"）
`ssc_sensitivity`	string	"high" | "standard" | "low"

12.4 カラースキーム

用途	カラーコード	プレビュー
Primary	#4A6FA5	
Background	#F5F7FA	
Safe	#27AE60	

Warning	#F39C12	
Danger	#E74C3C	

13. ファイル構成

ファイル	行数	説明
index.html	309	5画面シェル（Consent/Setup/Check/Results/Settings）
css/style.css	726	全スタイル（レスポンシブ対応）
js/app.js	1,652	全ロジック（IIFE、9モジュール構成）
worker/worker.js	283	Cloudflare Worker（CORS+認証+SSRF防止）
worker/wrangler.toml	3	Worker デプロイ設定
worker/README.md	39	Worker デプロイ手順
img/*.png	4枚	APIキー取得ガイド画像
robots.txt	3	クローラー拒否設定
.gitignore	2	REFERENCE.md, .DS_Store除外
合計		**2,970行 + 画像4枚**