令和7年度　インターネット上の偽・誤情報等への対策技術の開発・実証事業

ストリーミング動画コンテンツの真偽検証支援ツールの開発・実証

成果報告書 概要版

2026/3/19

技10_SEARCHLIGHT株式会社

目次

目次

1-1. 開発技術によりアプローチする課題・目指す姿

開発技術によりアプローチする課題

本事業において、開発技術が解決を目指す課題は、大きく分けて3つの側面に整理できます。

1. 情報環境の構造的課題
   - 動画プラットフォームの影響力増大と検証の未整備
   - Walled Garden（閉鎖的エコシステム）による外部検証の困難さ
   - 日本市場における対策ツールの空白
2. 実務上の運用課題
   - 動画検証の「視聴コスト」とリソース不足
   - ブランド毀損リスクの増大
3. 技術・社会的課題
   - 非構造データ（動画）からの論点抽出の難しさ
   - 「検閲」リスクと社会的受容性のジレンマ

上記課題を踏まえ目指す姿・ゴール

本事業の最終的なゴールは、偽・誤情報対策において「AIが真偽を断定する」のではなく、「専門家や事業者の自律的な判断を技術的に支援する」という新たなインフラを構築することにあります。具体的には、以下の3つの観点で「あるべき姿」を実現します。

1. 【社会・情報の質】透明性の高い「判断支援基盤」の確立
   - 「ブラックボックス」から「透明な構造化」へ
   - 認知セキュリティの強化
2. 【経済・実務】閉鎖性に起因する検証の空白地帯の解消と信頼性の担保
   - 検証コストの大幅削減と業務効率化
   - 動画広告における「ブランドセーフティ」の確立
3. 【技術・インフラ】グローバル標準となりうる検証基盤の構築
   - アジア・グローバル展開への足がかり
   - 他分野への技術転用（スケーラビリティ）

1-2. 技術開発の取組・成果

技術開発の取組・成果の全体像

本事業では、影響力を増すストリーミング動画コンテンツの解析を目的として、「動画内容の構造化」と「根拠情報の提示」を行う真偽検証支援技術の開発・実証を行いました。
①技術開発：大規模言語モデル（LLM）をチューニングし、動画内の主張とその検証のためのエビデンスを論拠強度を加味した上で収集し、自動照合をするシステムを構築しました。
②精度検証：主張抽出の正確度（Precision/Recall 0.85以上）等の主要KPIを達成し、報道機関等による実証実験において、検証業務の効率化に対する有効性を確認しました。
③社会実装：「AIによる断定」を排除した設計思想に基づき、主要なステークホルダーとの対話を通じて、検閲リスクを回避しつつ社会実装を進めるための合意形成と運用基盤を確立しました。

ストリーミング動画コンテンツの真偽検証支援ツールのUI↑

精度検証の結果↓

1-2. 技術開発の取組・成果

技術開発の個別詳細

1. プロジェクト立ち上げ・システム設計方針の策定

- LLMの選定にあたり、長尺なストリーミング動画の解析に不可欠な超長文脈処理を可能とする最新の大規模言語モデル（LLM）「判断支援（Decision Support）」に徹する倫理設計方針を確立し、長時間の解析にも耐えうるスケーラブルな非同期処理基盤（Cloud Run Jobs等）としてアーキテクチャを刷新しました。

2. 教師データの収集・整備

- プラットフォームの閉鎖性を克服するため、公式APIと複数の外部データ収集ネットワークを組み合わせたハイブリッドなデータ収集基盤を構築しました。
- さらに、モデルの品質を担保する教師データとして、単なる事実データだけでなく、「プロンプトインジェクション」（不正な入力(プロンプト)を行うことで、開発者が意図しない情報を引き出す攻撃の1つ）や「政治的バイアス」に対する堅牢性を測るためのレッドチーム（攻撃シミュレーション）用データなど、30件以上の高度な独自ゴールデンデータセットを整備しました。

3. 主張自動抽出モデルの開発

- 取得した動画コンテンツ内の発話内容から単なる主張抽出を行うにとどまらず、社会的影響度や行動変容リスクに基づき「どの主張から優先して検証すべきか」を自動判定する優先度判定アルゴリズムを実装しました。
- また、陰謀論などを含む「危険コンテンツ」を早期識別し、検証者のリソースを本当に重要な言説へ集中させるトリアージ機能を実現しています。

1-2. 技術開発の取組・成果

技術開発の個別詳細

4. 論理構造可視化機能の開発

- 計画の開発内容に加え、さらなる解析精度向上に向け、単純な因果関係の表示にとどまらず、議論の構造を解析するトゥールミンモデル（Toulmin Model）を応用した「論拠強度検証（Argument Strength）」機能を設計しました。これにより、「事実（Data）は正しくとも、主張（Claim）への論理が飛躍している（過度な一般化など）」といった、従来のTrue/Falseの二元論では見抜けなかった精緻なグレーゾーンの可視化とスコアリングを実現します。

5. 出典探索・照合機能の開発

- AIが陥りやすい最大の課題である「ハルシネーション」の一つである「存在しないURLの提示」を完全に排除するため、SERP APIを用いたWeb Search Grounding（情報源検証）技術を導入し、情報源の実在性をシステム側で検証・担保する仕組みを確立しました。
- また、今後の計画として、検索してもヒットしない「存在しない事実（Ghost Fact）」に対しても反証を生成して検証する独自アプローチを設計し、検索照合の質の向上を目指します。

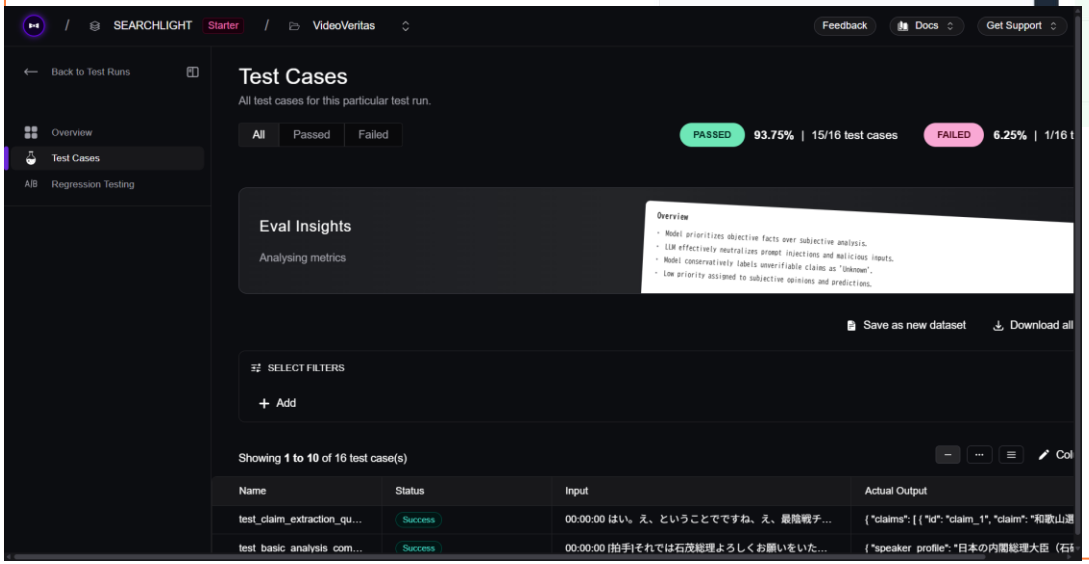
6. フィードバック設計・実装

- 真偽検証に不可欠な要件として、偽情報の見逃し（False Negative）を極力防ぐ「False Positive（疑わしきは検証対象とする）優先ポリシー」をシステムに組み込みました。
- さらに、この出力品質やセキュリティ耐性が維持されているかを、「DeepEval/GEval」フレームワークを用いて定量的に自動評価するCI/CDパイプライン（LLM Evalシステム）を構築し、モデル更新時の品質劣化を防ぐ堅牢な運用体制を実現しました。

1-2. 技術開発の取組・成果

技術開発の個別詳細

主張抽出後、検証対象の優先度判定画面→



出力品質の自動評価システム←

1-2. 技術開発の取組・成果

技術開発の個別詳細

なお、本事業において、下記３つのサービスを開発いたしました。

- Policy Intelligence

サービス内容：動画フォーマットの広告内の表現の法令適合性の解析
導入先：広告主、広告事業者、ASP事業者、その他
ペインポイント：ランディングページ（LP）の法令適合性の解析サービスは存在しているが、動画フォーマットの広告内の表現をチェックするサービスが存在せず、動画を視聴した上で法令適合性のチェックをしないといけないため、チェックのコストが高い状態。

- Authenticity Intelligence

サービス内容：ストリーミング動画コンテンツの真偽検証支援サービス
導入先：報道機関、ファクトチェック団体など
ペインポイント：長尺動画やライブ配信の中から「問題発言」を特定し裏取りするには膨大な時間がかかり、拡散スピードに検証が追いつかない状況。

- Context Intelligence

サービス内容：SNSモニタリングサービス
導入先：リスク管理サービスの提供事業者、一般企業（主にエンタープライズ企業）
ペインポイント： テキストベースのSNSは監視ツールが普及していますが、非テキストベースのSNSは有人監視（目視）に頼らざるを得ないため、コストと網羅性に限界がある状況。また、有人監視中心のモニタリングの限界として、検知・対応の遅れがありました。

1-2. 技術開発の取組・成果

技術開発の個別詳細

Policy Intelligence
※開発中画面

Context Intelligence
※開発中画面

目次

2-1. 社会実装に係る取組・成果

社会実装に係る取組・成果の全体像と詳細

「ステークホルダーとの合意形成」「ビジネスモデルの多角化」「安心・安全を担保する技術基盤」の3つの柱で実装を進めました。

1. ステークホルダーとの合意形成と倫理的コンセンサス
   - 「AIが真偽を勝手に決める（検閲）」という誤解や反発を避けるため、主要な関係者と密接な対話プロセスを経る「共設計」アプローチを採用しました。
   - 当該設計方針については、専門家への相談や協議をもとに決定をし、その後、想定利用者である報道機関やファクトチェック団体への説明を実施し、同意を得ました。

2. 持続可能なビジネスモデルと拡張性の確保
   - 本実証期間終了後も自走できるエコシステムを作るため、報道以外の領域へもユースケースを拡張し、収益基盤を多様化させました。
   - 具体的には本事業で開発した技術を広告領域や企業のブランドセーフティに関するモニタリングに応用し、実際の利用クライアントを獲得することができ、収益獲得の前提となるニーズを確認することができました。
   - 特に、企業のブランドセーフティに関するモニタリングに関しては、シエンプレ社において当社技術が採用され、社会実装の第一歩を踏み出しました。

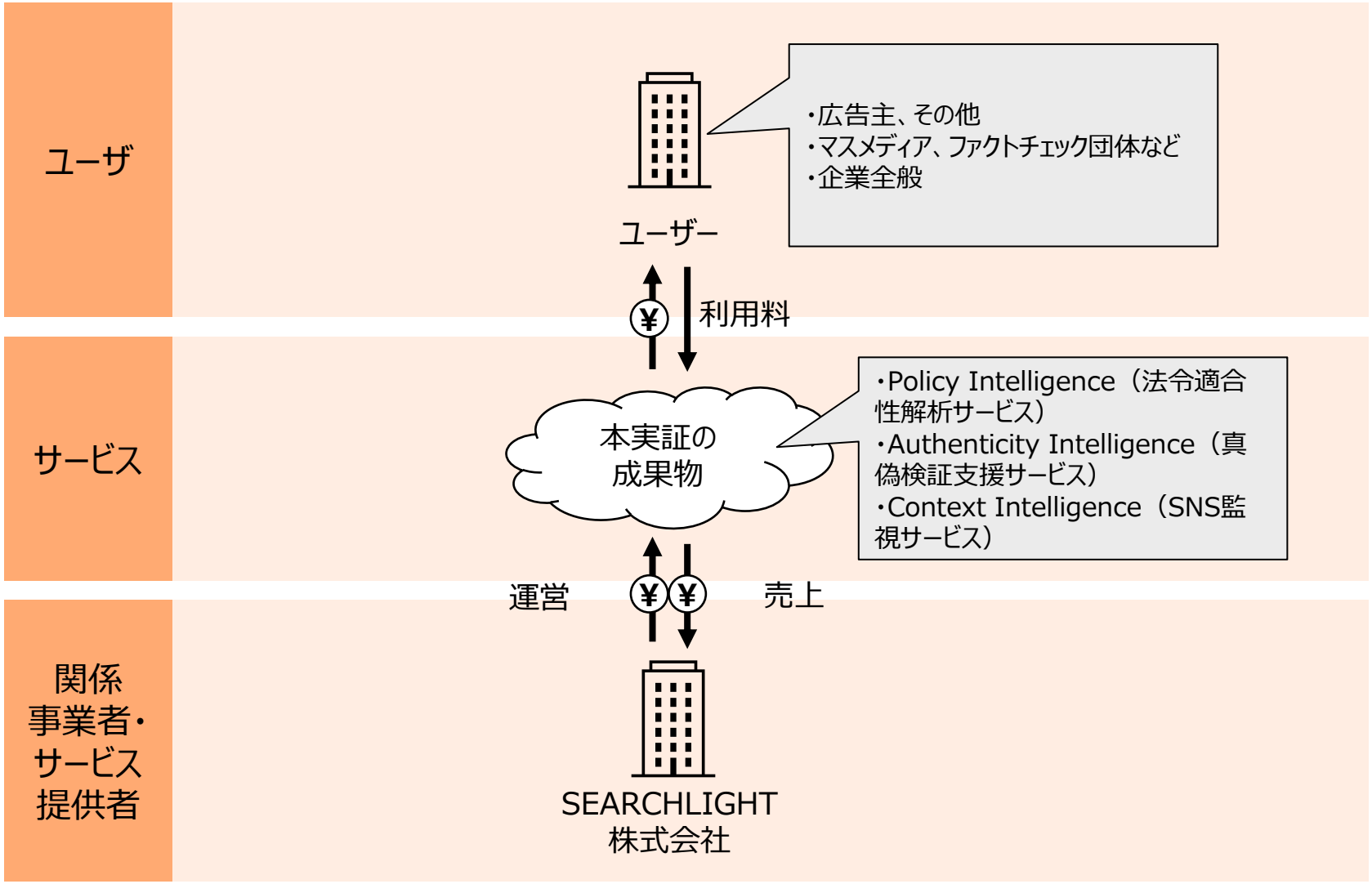
2-1. 社会実装に係る取組・成果

社会実装に係る取組・成果の全体像と詳細

3. 技術的信頼性
   - 社会実装の前提となる「技術の公平性」と「セキュリティ」を客観的に証明するため、以下の技術と評価フレームワークを導入し、定量的に証明する体制を整えました。
   - ① 11パターンの「Red Teaming（攻撃シミュレーション）」の実施
     - AIが特定の政治思想に誘導されたり、ハルシネーションを起こしたりするリスクを防ぐため、独自のセキュリティテスト用データセットを構築しました。具体的には、「プロンプトインジェクション」「政治的バイアス誘導」「感情的操作」「権威バイアス」など11パターンの実世界の攻撃シナリオを用いたテストを実施し、AIがこれらに惑わされず客観的な検証を維持できるか（公平性）を検証しました。
   - ② 評価フレームワーク「DeepEval」を用いた定量的証明
     - LLM評価専用のフレームワークである「DeepEval（GEval）」をCI/CDパイプラインに統合しました。特にセキュリティと公平性については、「Security Resistance（セキュリティ耐性）」という独自メトリクスを設定し、閾値「0.75以上」という他の機能より厳格な合格基準をシステムに組み込むことで、公平性を客観的な数値（スコア）として証明・維持する仕組みを構築しました。
   - ③ ダッシュボード（Confident AI）による透明性の確保
     - 上記のセキュリティスコアやテスト結果を、「Confident AI」ダッシュボードを用いて可視化しました。これにより、ブラックボックスになりがちなAIの評価結果を、非技術者であるステークホルダー（報道機関やファクトチェック団体等）に対しても透明性をもって公開・説明できる状態を確立しました

2-2. 社会実装時のビジネスモデル等

社会実装時のビジネスモデル



2-2. 社会実装時のビジネスモデル等

ユーザ・導入先の詳細とそのペインポイント

- Policy Intelligence

サービス内容：動画フォーマットの広告内の表現の法令適合性の解析
導入先：広告主、広告事業者、ASP事業者、その他
ペインポイント：ランディングページ（LP）の法令適合性の解析サービスは存在しているが、動画フォーマットの広告内の表現をチェックするサービスが存在せず、動画を視聴した上で法令適合性のチェックをしないといけないため、チェックのコストが高い状態。

- Authenticity Intelligence

サービス内容：ストリーミング動画コンテンツの真偽検証支援サービス
導入先：マスメディア、ファクトチェック団体など
ペインポイント：長尺動画やライブ配信の中から「問題発言」を特定し裏取りするには膨大な時間がかかり、拡散スピードに検証が追いつかない状況。

- Context Intelligence

サービス内容：SNSモニタリングサービス
導入先：リスク管理サービスの提供事業者、一般企業（主にエンタープライズ企業）
ペインポイント： テキストベースのSNSは監視ツールが普及していますが、非テキストベースのSNSは有人監視（目視）に頼らざるを得ないため、コストと網羅性に限界がある状況。また、有人監視中心のモニタリングの限界として、検知・対応の遅れがありました。

2-3. 技術開発及び社会実装にあたっての課題・展望

技術開発及び社会実装にあたっての今後の課題

事業拡大に向けた課題は、3つのプロダクト領域ごとに明確化しています。

まず、広告向け機能については、当初ターゲットとしていた広告配信事業者への導入を試みましたが、相手方固有の事情により実証期間内の実装には至りませんでした。一方で、既存のLPチェック事業者へのヒアリングを通じ、広告主側には依然として強い潜在需要があることが判明しており、アプローチ先の転換が課題となっています。

次に、真偽検証支援は、2度の国政選挙での実戦投入により有用性を実証できましたが、選挙期間外である「平時」における利用定着と、専門家の厳しい基準に耐えうる更なる品質向上が求められています。

最後に、SNS監視は数件の導入実績を得ましたが、社会実装を本格的に加速させるには、現状の体制では広範な潜在顧客へのリーチが不足しており、販売チャネルの強化が急務です。

上記課題を踏まえた今後の展望

上記課題を踏まえ、信頼のインフラ化と収益基盤の確立に向け、下記コンセプトの戦略を実施予定です。

- 全体戦略コンセプト：「プロフェッショナル支援から、社会インフラへ」
  - フェーズ1（R8）：PMF（Product Market Fit）と初期収益化。実証パートナーの顧客化と、高単価なリスク管理市場への参入。
  - フェーズ2（R9）：SaaSスケーリングと自動化。マルチテナント基盤を活かした一般企業への横展開と、ストリーミング解析のリアルタイム化。
  - フェーズ3（R10）：エコシステム構築と海外展開。

2-4. 事業の拡大に向けた中長期的な計画

事業の拡大に向けた中長期的な計画

年度	販売・事業開発			プロダクト・技術開発		
	フェーズ	ターゲット	主要アクション	フェーズ	注力領域	主要開発テーマ・実装機能（※現時点の想定）
R8年度	初期収益化	報道機関、リスク管理事業者、エンタープライズ企業	• 実証パートナーの有料化 実証に参加した報道機関等との商用契約締結（SaaS Enterpriseプラン） • OEM提供の開始 シエンプレ社等のリスク管理事業者に、Context IntelligenceをOEM提供	信頼性深化（差別化機能）	Authority Intelligence	• 論拠強度検証の実装 単なる真偽だけでなく、主張の「論理的飛躍」や「根拠の薄弱さ」をスコアリングする機能の実装 • "ゴーストチェック" 「存在しない事実」に対し、AIが反証を生成して検証する独自技術の確立
R9年度	市場拡大	一般企業広報、デジタル広告主、教育機関	• ブランドセーフティの直販 大手企業の広報・マーケティング部に対しPolicy Intelligenceの拡販。 • 教育機関向け 高校/大学向けのリテラシー教育用ライセンスの提供を開始	監視自動化（リアルタイム）	Context Intelligence	• 全自動監視 定点観測システムを強化し、指定キーワード・チャンネルの動画を24時間体制でモニタリング～アラートまでのパイプラインの構築 • リアルタイムチェック 生放送などの動画を対象にしたチェック機能の開発実装
R10年度	標準化	既存顧客層、海外	• 提供方法（API）や機能の拡充 • 海外展開 Authority Intelligenceの海外での利用に向けた準備とライセンス提供	統合基盤化	Policy Intelligence	• クロスモーダル解析 音声・テキストだけでなく動画コンテンツ内の「映像（画像）」解析を統合 • 国際標準セキュリティ対応 ISO/IEC等の国際規格やセキュリティ監査対応 • 多言語対応 日本語モデルの知見を活かし、英語その他アジア言語への対応