

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**

2026/3/19

技10_SEARCHLIGHT株式会社

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

1-1.開発・実証のサマリ

アプローチする課題・目指す姿

- ・ ストリーミング動画は、社会的・政治的な情報流通の主要なチャネルとなりつつあり、公共的な議論や有権者の意思形成にも多大な影響を及ぼしている。特に動画形式による主張・評論・告発といったコンテンツは、感情的訴求力が高く、視聴者の印象形成に大きな役割を果たしている。しかし、ストリーミング動画コンテンツは検証コストが相対的に高く、多くのユーザーが利用する情報空間の健全性確保に、真偽検証を支援するツールの開発提供が寄与する可能性がある。
- ・ ユーザーベースが大きく、しかしデータへのアクセスが難しいストリーミング動画コンテンツに含まれる真偽の把握が困難なプラットフォームについて、広く対応した真偽検証支援ツールを提供。これにより、上記サービスの利用者が真偽不明の情報に早期対応・判断できる健全な情報流通環境の構築を目指す。【デバンキング対象の拡張】

技術区分	コンテンツの真偽判別支援技術、改ざん検知技術	実施体制 (下線：技術開発主体)	SEARCHLIGHT株式会社、AIQVE ONE株式会社
対象とするモデル種	動画		

技術開発の取組・成果

- ・ ストリーミング動画コンテンツの解析を目的とし、「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を開発しました。大規模言語モデル（Gemini 3 Flash等）を活用し、動画内の主張や検証のためのエビデンスを、論拠強度を加味して収集・自動照合するシステムを構築したほか、実在しないURLを排除する出典照合の高度化も行いました。
- ・ 主要報道機関等での実証実験において、検証業務の負荷軽減と効率化に対する有効性を確認しました。
- ・ また、上記技術の広告・ブランドセーフティ領域や企業の炎上監視サービスへの応用を行いました。

社会実装に係る取組・成果

- ・ 「AIが真偽を断定する」のではなく人間の判断を支援するという思想に基づき、報道機関等と対話を行い検閲リスクを回避する合意形成を図りました。
- ・ また、マルチテナント対応のSaaS基盤の構築や、安全性検証（Red Teaming）を通じ、組織導入に耐えうる信頼性を確保しました。
- ・ 成果として、2度の国政選挙において主要報道機関での試験導入を果たし有効性を実証したほか、広告・ブランドセーフティ領域や企業の炎上監視サービスへとビジネス展開を進め、持続可能な社会インフラとしての基盤構築のきっかけをつくりました

技術開発及び社会実装にあたっての課題・展望

【課題】

技術面では、陰謀論など「存在しない事実」を見抜く技術の確立と、外資プラットフォームに依存しない冗長性の確保が課題です。社会実装面では、対策市場が未成熟なため持続可能なビジネスモデル構築が急務です。社会実装にあたりベーシックなニーズは確認できたものの、平時での利用定着や広範な潜在顧客への販売チャネル強化が求められています。

【展望・対応事項】

今後は「プロ支援から社会インフラ」を掲げ改善を進めます。技術面では、主張の論理的飛躍を可視化する「論拠強度検証」や、AIが反証を生成する「ゴーストファクトチェック」機能を実装し検証精度を向上させます。社会実装面では、広告主へアプローチしブランドセーフティ市場を開拓します。SaaS基盤を活かして一般企業へ横展開して収益基盤を確立するとともに、将来的には海外での利用を目指し、持続可能な対策エコシステムの構築を目指します

代表者コメント



SEARCHLIGHT株式会社
代表取締役
瀬戸 亮

実証期間が短く、機能面でも社会実装面でも非常に難しい状況でしたが、参院選と衆院選という大きいイベントが発生したことにより、潜在的ニーズが顕在化し、多くのフィードバックを賜ることで、とても実りのある開発・実証となりました。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

2-1. 開発技術によりアプローチする課題

開発技術によりアプローチする課題

<概要>

本事業は、社会的影響力を増す動画ストリーミングサービスにおいて、事実検証の仕組みが追いついていない「検証の空白」課題に取り組むものです。具体的には、大手動画ストリーミングサービスのデータへのアクセスの困難性に起因する閉鎖性や検証作業の属人性に起因する「検証コストの高さ」を解消し、ステークホルダーが直面するリソース不足やブランド毀損リスクを低減することを目的とします。また、AIによる自動判定ではなく、「主張の構造化」と「出典の可視化」を通じた判断支援を行うことで、検閲リスクを回避しつつ情報の透明性を高める新たな社会的基盤の構築を目指します。

<詳細>

本実証事業において、開発技術（ストリーミング動画コンテンツの真偽検証支援ツール）が解決を目指す課題は、大きく分けて「情報環境の構造的課題」「実務上の運用課題」「技術・社会的課題」の3つの側面に整理できます。

1. 情報環境の構造的課題

現代の情報流通におけるチャンネルの変化と、それに対応しきれていない検証環境のギャップに関する課題です。

- 動画プラットフォームの影響力増大と検証の未整備
 - 近年、社会的・政治的な言論空間の主戦場がテキストから「動画ストリーミングサービス」へ移行しています。特に2024年の兵庫県知事選の事例に見られるように、有権者の投票行動においてSNSや動画サイトが既存マスメディア（新聞・テレビ）以上の影響力を持つケースが確認されています。
 - 動画コンテンツは感情訴求力が強く拡散力が強い一方で、テキスト情報に比べて真偽検証が十分に行われないまま拡散されやすい傾向があります。

2-1. 開発技術によりアプローチする課題

開発技術によりアプローチする課題

- Walled Garden（閉鎖的エコシステム）による外部検証の困難さ
 - 大手動画ストリーミングサービスは閉鎖的なエコシステム（Walled Garden）であり、外部の研究者や検証者がデータへアクセスする手段が限定されています。
 - これにより、偽・誤情報の拡散状況やアルゴリズムの影響を第三者が客観的に検証することが構造的に困難な状況にあります。
- 日本市場における対策ツールの空白
 - 欧米では動画プラットフォーム側によるラベリングや真偽検証機能の実装が進んでいますが、日本国内はその対象外となっているケースが多く、対策の空白地帯となっています。

2. 実務上の運用課題

情報の信頼性を担保すべきステークホルダー（後述）が直面している具体的な業務課題です。

- 動画検証の「視聴コスト」とリソース不足
 - 報道期間・ファクトチェック団体など：テキスト記事と異なり、動画コンテンツの内容を確認するには「視聴」が必要です。長尺動画や複雑な政治的コンテンツの中から問題発言を特定し、裏取りを行う作業は極めて工数がかかり、急速に拡散する動画（急伸動画）のスピードに検証が追いついていない現状があります。
 - 広告主・広告事業者・広告配信事業者：ブランドセーフティの観点で不適切な動画（偽情報や権利侵害など）を排除したいニーズがありますが、既存ツールではメタデータ（タイトル、概要欄、視聴回数等）の検索しかできず、動画の中身（発言内容）までは自動解析できないため、目視確認に頼らざるを得ない状況です。
- ブランド毀損リスクの増大
 - 偽・誤情報が拡散された動画やチャンネルに広告が出稿されることで、広告主のブランド価値が毀損されるリスクが顕在化していますが、有効な回避手段が不足しています。

2-1. 開発技術によりアプローチする課題

開発技術によりアプローチする課題

3. 技術・社会的課題

技術的な難易度と、社会実装における倫理的な課題です。

- 非構造化データ（動画）からの論点抽出の難しさ
 - 単に文字起こしをするだけでなく、文脈の中から「主張（Claim）」や「因果関係」を特定し、それが「事実（Fact）」なのか「意見（Opinion）」なのかを切り分ける技術的ハードルが存在します。
- 「検閲」リスクと社会的受容性のジレンマ
 - 偽・誤情報対策は、一步間違えると「検閲」や「言論統制」と受け取られるリスクがあります。特にAIが「真偽を自動判定」することは、誤判定のリスクや透明性の観点から社会的な反発を招く恐れがあります。
 - したがって、「AIによる断定」ではなく、あくまで人間の「判断支援（構造化・出典提示）」に留めるという、技術と倫理のバランスの取れた設計が求められます。

2-2. 開発技術により目指す姿・ゴール

開発技術を通して目指す姿・ゴール

<概要>

本事業は、動画ストリーミングサービス上の情報流通において、AI技術を活用した「専門家のための透明な検証支援エコシステム」の構築を目指すものです。最終的なゴールは、AIが真偽を断定するのではなく、論理構造の可視化と出典の提示を通じて、「報道機関等の検証負荷軽減」および「広告主のブランドセーフティ確保」を実現することにあります。

また、本技術を「Walled Garden」に対する外部検証の標準的手段として確立し、将来的には災害・医療分野やグローバル市場への展開を通じて、信頼できる情報社会基盤（トラスト・インフラ）の一翼を担うことを目指します。

<詳細>

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

検閲や言論統制のリスクを回避しつつ、情報のプロフェッショナルが正確かつ迅速に検証を行える環境を実現します。

- 「ブラックボックス」から「透明な構造化」へ
 - 動画内の「事実（Fact）」と「意見（Opinion）」が未分化なまま拡散される現状に対し、AIが論理構造（因果関係・前提）を可視化することで、情報の透明性を向上させます。
 - AIによる自動判定ではなく、あくまで「根拠の提示」に徹することで、公平性を保ったまま検証プロセスを効率化します。

2-2. 開発技術により目指す姿・ゴール

開発技術を通して目指す姿・ゴール

- 認知セキュリティ（Cognitive Security）の強化
 - 感情的な煽りや情報の偏り（思い込み）が生じやすい動画コンテンツに対し、客観的なかつ根拠強度の高いエビデンスを提示する技術を確立します。
 - これにより、報道機関やファクトチェック団体は、外部からの認知攻撃や誤情報の拡散に対して、強固な対抗手段を持つことができます。

2. 【経済・実務】「Walled Garden」を超えた信頼性の担保

巨大プラットフォームの閉鎖性（Walled Garden）に阻まれていた外部検証を可能にし、経済活動における信頼性を回復します。

- 動画広告における「ブランドセーフティ」の確立
 - 従来はタイトル等のメタデータでしか判断できなかったリスク動画を、内容レベルで解析可能にします。
 - これにより、広告主や広告代理店は、偽・誤情報や不適切な動画への出稿を回避でき、ブランド毀損リスクを最小化できるようになります。
- 検証コストの大幅削減と業務効率化
 - 報道機関や広告主・広告事業者・広告配信事業者が「全編視聴」によって行っていたチェック作業を自動化により削減します。
 - 特に「急伸動画」の初動対応を迅速化し、真偽検証のサイクルを高速化することで、人的リソースの不足を解消します。

2-2. 開発技術により目指す姿・ゴール

開発技術を通して目指す姿・ゴール

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

日本特有の事情に対応しつつ、将来的には国際的な標準規格への接続を目指します。

- アジア・グローバル展開への足がかり
 - 「Walled Garden」による検証の困難さは世界共通の課題です。本技術を、国際的なファクトチェック基準（IFCN等）や海外の研究機関（香港大学など）との連携を通じて、グローバルな課題解決ツールへと発展させます。
- 他分野への技術転用（スケーラビリティ）
 - 政治分野だけでなく、正確性が命に関わる災害情報、医療情報、行政広報などの領域へ技術を横展開し、社会全体のレジリエンス（回復力）を高めることを目指します

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

<概要>

本技術は、ストリーミング動画コンテンツの「全編視聴」や「目視確認」に依存していた従来の確認業務を、AIによる構造化と出典照合によって代替・支援するものです。主なユースケースは、①報道機関・ファクトチェック団体における「真偽検証の高速化」、②広告主・広告事業者・広告配信事業者における「ブランドセーフティおよび法令適合性の確保」、③企業・プラットフォームにおける「動画リスクの自動監視」の3点に大別されます。特に、拡散スピードの速い「急伸動画」への初動対応や、メタデータ検索では検知できない動画内リスク（不適切な発言等）の排除において、業務効率の劇的な改善とリスク低減を実現します。

<詳細>

本技術は、動画コンテンツの内容を独自にチューニングしたAIが解析・構造化することで、従来は「人の目と耳」に頼らざるを得なかった確認作業を自動化・効率化します。主なユースケースは以下の3点です。

1. 【報道・調査】真偽検証の高速化・効率化

急速に拡散する動画（急伸動画）に対し、報道機関や専門団体が迅速に事実確認を行うための支援を行います。

- 検証プロセスのコスト削減とスピード向上
 - 課題：従来の動画検証は、記者が全編を視聴して問題発言を探す必要があり、膨大な工数がかかっていました。
 - 解決策：AIが動画内の「主張」と「論理構造」を自動抽出し、関連する出典（ソース）を即座に提示します。これにより、リサーチの初動負荷を劇的に下げ、検証記事の公開までのスピードを向上させます。

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

- 「急伸動画」の即時検証
 - 課題: 政治的イベント時などに多く再生され、またたく間に世間に広まった動画の検証が追いつかない現状があります。
 - 解決策: 公開24時間以内に急拡散している動画をシステムが「検証対象」としてサジェストし、拡散の初期段階での介入を可能にします。

2. 【広告・マーケティング】ブランドセーフティと法令適合性の確保

広告主・広告事業者・広告配信事業者が、動画コンテンツの内容リスクを管理するために活用します。

- 動画コンテンツの法令適合性検証（薬機法・景表法等）
 - 課題: 動画アフィリエイト等において、動画内の表現が法令に違反していないか（例：誇大広告や禁止表現）を目視でチェックする必要があり、限界がありました。
 - 解決策: 動画内容を解析することで、ウェブサイトのLP（ランディングページ）のチェックツールと同様に、法令や業界規則への適合性を自動判定します。これにより、当事者による審査業務効率と精度を向上させます。
- ブランド毀損リスクの回避（ブランドセーフティ）
 - 課題: 既存ツールではタイトル等のメタデータ検索しかできず、動画の中身で「偽情報」や「不適切な発言」を検知できないため、意図せず不適切な動画に広告が出稿されるリスクがありました。
 - 解決策: 動画の内容ベースで信頼性を評価し、偽・誤情報や権利侵害を含むリスクの高い動画・チャンネルへの出稿を未然に防ぎます。

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

3. 【リスク管理】プラットフォーム・企業の炎上監視

企業広報において、動画コンテンツを起点としたリスクを早期に検知します。

- 動画投稿の自動炎上監視
- 課題：テキスト投稿と異なり、動画投稿の内容監視は自動化が難しく、有人監視に頼らざるを得ないため、コストと対応速度に課題がありました。
- 解決策：動画コンテンツの内容を解析し、炎上の火種となる「過激な主張」や「偽情報」が含まれていないかを自動でモニタリングします。これにより、有人監視の負荷を下げつつ、リスクの早期発見・対応が可能になります。

ユースケース領域	ターゲット	提供価値
真偽検証	報道機関、ファクトチェック団体等	「検証コストの削減」 長尺動画からの論点抽出・裏取り作業の自動化
法令適合性	広告主・広告事業者・広告配信事業者	「動画内容の解析」 メタデータでは検知できない法令違反・リスクの検知
監視・リスク管理	企業広報	「自動化による監視網羅性」 有人監視ではカバーできない投稿数の常時監視

このように、「報道の質」だけでなく、「広告・ビジネスの安全性」や「企業リスク管理」という経済合理性の高い領域にも明確なユースケースが存在することが、本事業の強みです。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

3-1. 技術開発の全体像

技術開発に係る取組・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) 等の目標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。

これらの成果により、動画プラットフォームの閉鎖性 (Walled Garden) に起因する検証の空白地帯を解消し、健全な言論空間と経済活動を支える技術基盤の目処をつけることができました。

3-1. 技術開発の全体像

技術開発に係る取組・成果の全体像

なお、技術開発のアウトプットとしては以下の3サービスです

- 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。

検出された指摘事項				
違反判定	箇所/時間	原文	修正案/アドバイス	根拠
医療 違反	title	たった1回で激変、劇 的改善	「激変」「劇的」といった表現は、効果を誇張し患者に過度な期待を抱かせるため「誇大広告」に該当します。また「たった1回で」という表現は効果を保証す	医療広告ガイドライン（誇大広告の禁止）、医療広告規制におけるウェブサイト等の車側解説書

- ユーザーはチェックしたい動画コンテンツのURLを入力することで、解析結果を得ます。

3-1. 技術開発の全体像

技術開発に係る取組・成果の全体像

- 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。



- ユーザーはチェックしたい動画コンテンツのURLを入力することで、解析結果を得ます。
- ストリーミング動画コンテンツだけでなく、テキストや音声データをインプットした場合も、同フォーマットで解析を行います。

3-1. 技術開発の全体像

技術開発に係る取組・成果の全体像

- 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。

The screenshot displays the SEARCHLIGHT interface. On the left is a dark sidebar with navigation options: ダッシュボード, 投稿一覧 (selected), and 収集設定. The main area is titled '投稿一覧' and includes a search filter for '全てのPF', a 24-hour time range, and risk level filters (高, 中, 低, なし). Below the filters, a '検知フィード' section shows 635 detected items. A table lists the top results:

PF	タイトル / AI分析	センチメント / リスク	指標	投稿日時
🔔	[Redacted]	なし → ポジティブ	📊 71	2026/1/22 13:35:29
🔔	[Redacted]	なし → ポジティブ	📊 363	2026/1/22 13:06:30
🔔	[Redacted]	なし → ポジティブ	📊 384	2026/1/22 12:36:58
🔔	[Redacted]	なし → ポジティブ	📊 793	2026/1/22 12:35:18

- ユーザーはモニタリングしたいコンテンツを収集するためのキーワードを設定します。
- また、必要に応じて一次情報を提供することで、投稿内容との差分を説明します。

3-2. 技術開発の個別詳細

技術開発の個別詳細

実施計画書に掲げた6つの取り組み項目について、以下の通り開発を完了いたしました。

=====

1. プロジェクト立ち上げ・システム設計方針の策定

計画：LLMの選定、システムアーキテクチャの設計、倫理設計方針の確立。

- LLM基盤の確立：当初想定モデルに加え、Google CloudのVertex AI（Geminiシリーズ）をメインエンジンに採用しました。開発終盤では最新の「Gemini 3 Flash Preview」へ全面移行し、最大出力トークン数を65,536へ拡大することで、長尺動画の分析精度を向上させました。
- アーキテクチャの刷新：スケーラビリティを確保するため、バックエンドDBをSupabaseからGoogle Cloud SQLへ移行し、認証基盤もCloud Identity Platformへ刷新しました。また、長時間処理に対応するため、SSE（Server-Sent Events）からCloud Run Jobsを用いた非同期バッチ処理＋ポーリング方式へと構成を変更し、堅牢性を高めました。

2. 教師データの収集・整備

計画：モデル構築に必要なデータの収集、過去の真偽検証データの活用。

- 評価用データセットの構築：「Video Veritas LLM Evaluation System」のために、標準データセット（19件）、セキュリティ攻撃パターン（11件）、検索品質評価用データ（8件）など、計38件以上のゴールデンデータセットを整備しました。
- APIによるデータ取得：APIに加え、信頼性の高いデータ収集プロバイダを併用するハイブリッド構成を実装し、取得の安定性を確保しました。

3-2. 技術開発の個別詳細

技術開発の個別詳細

3. 主張自動抽出モデルの開発

計画：文脈解析による主張単位の抽出、信号語・意見分類のタグ付け。

- 文脈解析と優先度判定：解析対象から「主張（Claim）」を抽出する際、社会的影響度や検証可能性に基づき「優先度（Priority）」を自動判定するロジックを実装しました。LLMによる判定のばらつきを検証し、安定して高優先度の主張を特定できることを確認しています。
- テキスト分析への拡張：動画だけでなく、ニュース記事やSNS投稿などのテキスト入力に対しても同様の主張抽出・検証を行える機能を新規追加し、汎用性を高めました。

4. 論理構造可視化機能の開発

計画：因果関係・前提関係の判定、推論グラフUIの開発。

- 論拠強度検証（Argument Strength）：単なる因果関係の表示にとどまらず、トゥールミンモデル（Toulmin Model）に基づき、「データ→主張」の論理的整合性や論拠の妥当性をスコアリング（0-100点）する高度な検証機能を設計いたしました。
- タイムラインと構造化UI：抽出された主張を時系列およびトピックごとに整理して表示するUIを構築しました。クリックすることで前後の文脈（コンテキスト）を確認できる機能も実装し、ユーザーの理解を支援しています。

5. 出典探索・照合機能の開発

計画：明示的出典の抽出、類似主張の検索による補助情報提示。

- 「Web Search Grounding」の高度化：検索の精度を高めるため、SERP APIを用いたWeb検索・URL検証システムを構築しました。これにより、実在しないURLの提示を排除し、ページの実在性を確認した上で出典として提示する仕組みを確立しました。
- マルチソース検索：検索API及び検索クエリの自動最適化を行うことで、日本語特有の表現や最新ニュースに対しても高い適合率での出典提示を実現しました。

3-2. 技術開発の個別詳細

技術開発の個別詳細

6. フィードバック設計・実装

計画：ユーザー判断ログの収集、再学習への活用、人間参加型ループの構築。

- 保守的な検証方針の実装：フィードバックに基づき、見逃し（False Negative）を防ぐために「False Positive（疑わしきは検証対象とする）優先」のポリシーをシステムに反映しました。また、陰謀論や信頼性が極端に低いコンテンツに対しては、リスクレベルに応じた警告を表示する機能を実装しました。
- 定量的な評価システムの構築（DeepEval）：「DeepEval」フレームワークを導入し、ファクトチェック品質（FactCheckQuality）やセキュリティ耐性（SecurityResistance）を定量的なスコア（閾値0.75以上等）で管理するCI/CDパイプラインを構築しました。これにより、モデル更新時の品質劣化（レグレッション）を防ぐ体制を整えました

3-2. 技術開発の個別詳細

技術開発の個別詳細

<サマリ>

開発項目	実施内容・成果概要
①システム基盤	Gemini 3 Flashへの移行による長文解析対応と、Cloud SQL/Cloud Run Jobsによるスケーラブルな非同期処理基盤の構築完了。
②主張抽出	社会的影響度に基づく「優先度判定」の実装。テキスト入力への対応によるマルチモーダル化。
③構造化・可視化	抽出された主張を時系列およびトピックごとに整理して表示するUIを構築
④出展照合	Web Search Groundingの強化。SERP APIによるURL実在確認プロセスを導入し、ハルシネーション（架空の出典提示）を抑制。
⑤品質評価	DeepEvalを用いた自動評価システムの構築。「False Positive優先」ポリシーの適用による見逃し防止策の徹底。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

4-1. 検証及び調査の全体像

検証及び調査に係る取組・成果の全体像

本事業では、開発した「真偽検証支援技術」を実際の社会課題解決に適用するため、報道・企業・教育の各分野で実証を行いました。

特筆すべき成果として、2025年・2026年の国政選挙において主要報道機関、地方紙及びファクトチェックメディアへの試験導入を果たし、選挙期間中の動画検証業務における有効性を実証しました。

また、ビジネス領域ではシエンプレ社・トライビート社との連携により、大手企業向けの炎上監視、レピュテーションリスクコントロールサービスを開始しました。

一方、計画していたIFCN理事との連携による国際標準化や、教育現場への展開に関しては、ビジネス領域における社会実装を優先した結果、実証期間中において明確な成果を得ることができませんでした。

衆院選のYouTube動画、投稿者の7割が匿名 政党発信の4.6倍視聴

衆院選速報2025 [+フォローする](#)

2025年2月13日 15:00 [公開済記事]

保存

📧 📧 📧 📧 📧 📧



国会議事堂ではスマホで写真や録音者撮影する観客が多い（1月27日、東京都千代田区）

動画配信サイト「YouTube」で衆院選関連動画を調査したところ、匿名の投稿者による動画が視聴数の7割を占めたことがわかった。テレビやネット動画を第三者が再編集した「切り抜き」が中心だ。また再生数の7割が10秒～1分程度のショート動画だった。専門家は「意見の偏りや中毒性の高さに留意した方がよい」としている。

調査は日本経済新聞とスタートアップのサーチライト（東京・豊島）で実施した。公示日の1月27日から2月1日に投稿され、政党名や党首名などのキーワードを含む再生回数上位2000本の動画を分析した。投稿者アカウントは約800件、総再生回数は計3億6000万回にのぼった。

2026.02.24

Web/SNSモニタリングの監視対象を「主要動画プラットフォーム・新興SNS」へ拡大

総務省事業採択団体のSEARCHLIGHT社と連携し、動画内の「音声・発言」までAIが判定する深層監視を開始



デジタル・クライシス対策のシエンプレ株式会社（本社：東京都港区、代表取締役社長：佐々木 寿郎、以下シエンプレ）は、総務省「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」に採択されたSEARCHLIGHT株式会社（本社：東京都豊島区、代表取締役：瀬戸 亮、以下SEARCHLIGHT）と連携いたしました。

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

1. 【報道・ファクトチェック】国政選挙での実戦投入と有用性の実証

実施計画上「Tier1（報道機関）およびTier2（ファクトチェック団体）」と定義した各想定利用団体と連携し、実際の選挙報道プロセスへの技術投入を行いました。

- 国政選挙における実戦投入
 - 2025年7月 参議院議員選挙および2026年2月 衆議院議員選挙において、本「真偽検証支援ツール」を実戦投入しました。
 - 選挙期間中に急増する政治家や候補者の動画発言、および切り抜き動画に対し、AIによる「主張の構造化」と「出典照合」を実施し、報道現場の真偽検証業務を支援しました。
- 主要メディアでの試験導入・連携
 - 毎日新聞、共同通信、日本テレビ、TBS、NHKといった主要報道機関内の一部ユーザに加え、FIJ（ファクトチェック・イニシアティブ）提携メディアでの試験導入を実現しました。
- 利用者へのヒアリングやアンケートにより、「動画検証の視聴コスト削減」や「主張の検証に必要な情報の収集コストの削減」などの点において一定の評価を得たものの、よりマクロなトレンドの解析や他のSNSへの対応を求める声があり、想定したユースケースより多くのニーズに応える必要があることが分かりました。
- また、日経新聞と衆議院選挙期間中のストリーミング動画コンテンツに関する共同調査を行いました。
 - 2026年2月3日「衆議院選挙のYouTube再生数、7割が匿名投稿 政党発信の4.6倍」
 - 2026年2月20日「「アンチ」政党・候補者動画が再生数稼ぐ 衆院選、平均より6割多く」
 - 2026年2月24日「拡散動画で月収200万円 衆院選YouTube、再生数の55%が匿名投稿者」
 - 2026年2月26日「衆院選動画の匿名投稿者、半数は開設1年未満 「バズる」政治家次々と」
 - 2026年3月3日「維新がYouTube広告に衆院選演説、選挙運動なら公選法違反 動画調査」
 - 2026年3月5日「野放しの中傷動画、選挙工作の温床に アルゴリズム規制は欧州先行」

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

2. 【リスク管理・広告】炎上監視・ブランドセーフティの事業化

計画段階で掲げていた「広告主・ASP事業者」向けのユースケースを拡張し、企業のレピュテーションリスク管理（炎上監視）領域でのビジネス展開を開始しました。

- 炎上監視
 - デジタル・クライシス対策企業のシエンプレ株式会社と連携し、実際のクライアント企業に対し、動画コンテンツの自動リスク検知サービスの提供を開始しました。
 - 従来は有人監視に頼らざるを得なかったプラットフォーム上の投稿について、AIによる自動監視を実現し、監視漏れのリスク低減とコスト削減に寄与しています。
 - なお、実証期間中に2件の案件が実際に開始されております。
- レピュテーションリスクコントロールサービス
 - デジタルマーケティング企業の株式会社トライビートと連携し、大手企業に対し、SNS横断でのレピュテーションリスクコントロールサービスの提供を開始いたしました。
 - コンテンツ内の「不適切な内容」や「偽情報」、「公式情報との乖離」を常時解析・把握することが、大手企業のレピュテーションリスクをコントロールするニーズと合致することを確認いたしました。

3. 【調査報道・法令】動画コンテンツの適法性検証

技術の応用範囲を広げ、ジャーナリズムおよび法令順守の観点での検証を実施しました。

- 報道機関との共同調査
 - 動画内の表現が法令やガイドラインに適合しているかをAIで検証する手法を確立し、その成果を報道機関と協力し、調査報道として公開する予定です。これにより、技術が単なる「チェックツール」を超え、社会的な監視機能の一翼を担えることを目指します。

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

4. 【教育・国際連携】リテラシー教育と国際標準化への接続

技術の社会受容性を高めるため、教育現場への展開と、グローバルなファクトチェック基準への準拠を進めるべく関係者とのコミュニケーションを行いました。

- 教育現場への展開
 - 実施計画に基づき、株式会社インフォハント（代表：安藤氏）と連携し、教育現場での実証展開に向けた検討を行いました。
 - 学生や生徒が情報の真偽を見極めるための「補助ツール」として本技術を活用し、メディアリテラシー教育における有用性を検証予定です。
- 国際標準化へのアプローチ
 - 国際的なファクトチェックの規格統一を目指し、IFCN（国際ファクトチェックネットワーク）理事へのアプローチを行いました。
 - 「Walled Garden」問題は世界共通の課題であるため、本技術のアプローチを国際標準（Global Fact等）に接続し、アジア発の検証プラットフォームとして展開するための足がかりの構築を目指します。

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

5. 【技術検証】定量的評価システムの確立とKPI

- 定量的評価システムの構築
 - 開発フェーズにおいて、38件以上のゴールデンデータセット（正解データ）を整備し、「DeepEval」フレームワークを用いた自動評価システムを構築しました。
 - 主張抽出精度（Precision/Recall）において、目標としていたKPI（0.85以上）を達成いたしました。
- セキュリティ・堅牢性の検証
 - プロンプトインジェクションや政治的バイアスを含む11パターンの「Red Teaming（攻撃シミュレーション）」を実施し、システムが偏った出力を行わないことを定量的に実証しました（Security Resistance スコア 0.75以上）。
- 実証プロセス・ユーザビリティ
 - 利用期間の短さから、有効なフィードバックを得ることができませんでした。
 - 定性的な評価としては、当初想定のとおり、長尺の動画コンテンツをすべて視聴することなく内容を把握することができるという点を評価いただきました。
 - UIの操作性/負担感に関しては、特段の問題は無いようでした。

The screenshot displays two side-by-side views of a web application. The left view, titled 'fine-tuning後比較', shows a list of evaluation results for BERTScore. The right view, titled 'Test Cases', shows a summary of test results with a 'PASSED' rate of 93.75% and a 'FAILED' rate of 6.25%.

4.4 BERTScoreでの確認

- 評価モード: `mode:bert_score_official` (bert_score ライブラリ、bert_base: multilingual-cased、lang:ja)
- GPT-5.2ベースライン:
 - macro: P 0.821 / R 0.812 / F1 0.816
 - micro: P 0.820 / R 0.812 / F1 0.818
 - micro件数: n=1896, m=1387
 - 参照: `outputs/metrics_bert_score_official_gpt52_base_low_test.json`
- 未チューニング Gemini 2.5 Pro:
 - macro: P 0.858 / R 0.822 / F1 0.839
 - micro: P 0.864 / R 0.823 / F1 0.843
 - micro件数: n=876, m=1387
 - 参照: `outputs/metrics_bert_score_official_vertex_base25pro_v1_nomax_test.json`
- fine-tuned Gemini 2.5 Pro:
 - macro: P 0.871 / R 0.878 / F1 0.874
 - micro: P 0.870 / R 0.882 / F1 0.876
 - micro件数: n=1354, m=1387
 - 参照: `outputs/metrics_bert_score_official_vertex_tuned_v1_redux_nomax_test.json`

Test Cases

All test cases for this particular test run.

Overview: PASSED 93.75% | 15/16 test cases | FAILED 6.25% | 1/16

Eval Insights

Analysing metrics

SELECT FILTERS

Showing 1 to 10 of 10 test case(s)

Name	Status	Input	Actual Output
test_claim_extraction_qu...	Success	00:00:00 はい、え、ということですね、え、最終報告...	{ "claims": [[{ "id": "claim_1", "claim": "和歌山道
test_basic_analysis_com...	Success	00:00:00 用事それぞれは石茂経理よろしくお願いをいた...	{ "speaker_profile": "日本の内閣総理大臣 (石

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

5-1. 社会実装に向けた取組の全体像

社会実装に係る取組・成果の全体像

本事業では、技術の「社会的受容性」を最優先事項とし、報道機関や専門家との対話を通じた「判断支援型AI」としての合意形成を徹底しました。

実装面では、マルチテナント対応によるSaaS基盤の確立や、Red Teamingによる公平性の定量的実証を行うことで、組織導入に耐えうるセキュリティと信頼性を確保しました。

また、報道分野にとどまらず、広告・ブランドセーフティ領域でのビジネス化を進め、経済性と公共性を両立した持続可能な社会インフラとしての基盤を構築しました。

5-2. 社会実装に向けた取組の個別詳細

社会実装に向けた取組の個別詳細

本事業では、偽・誤情報対策技術というセンシティブなテーマを扱うため、技術開発と並行して「社会的受容性の醸成」と「持続可能な運用モデルの確立」を両輪で推進しました。具体的には、「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。

- Tier 1（報道機関）とのクローズドな信頼構築
 - 全国紙、放送局に対し、開発初期段階から非公開（オフレコ）の説明会・体験会を実施しました。
 - 「真偽判定はあくまで人間が行い、AIは判断材料（出典・論理構造）の提示に徹する」という「判断支援（Decision Support）」の設計思想を共有し、報道現場の倫理基準に合致する仕様の合意を得ました。
- Tier 2（専門家・ファクトチェック団体）とのクロスレビュー
 - FIJ（ファクトチェック・イニシアティブ）やファクトチェックメディアへのヒアリングなどを通じ、開発技術のクロスレビューを実施しました。
 - フィードバックに基づき、判定基準（レーティング）や見逃しを防ぐための「False Positive（疑わしきは検証対象とする）優先ポリシー」をシステムに実装し、専門家が納得できる品質基準を確立しました。

5-2. 社会実装に向けた取組の個別詳細

社会実装に向けた取組の個別詳細

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。

- SaaS型・マルチテナント基盤の構築
 - 当初の単体ツール提供から発展させ、複数の組織（メディア企業、研究機関、企業広報）が独立した環境で利用できる「マルチテナント機能」を実装しました。
 - これにより、顧客ごとのデータ分離（セキュリティ）を確保しつつ、組織単位でのライセンス契約が可能なSaaSビジネスモデル（サブスクリプション型）への移行を予定しています。
- 「ブランドセーフティ」市場への展開
 - 報道機関向けのツール提供に加え、広告主やASP事業者向けに、動画広告の法令適合性や炎上リスクを判定するソリューションを展開しました。
 - 「全編視聴」が必要だった動画チェック業務を自動化するニーズに応え、経済合理性のあるBtoBサービスとしての市場を開拓しました。

5-2. 社会実装に向けた取組の個別詳細

社会実装に向けた取組の個別詳細

3. 技術的信頼性と国際標準への接続

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体制を整えました。

- Red Teaming（レッドチーミング）による安全性検証
 - 悪意ある入力（プロンプトインジェクション）や政治的バイアス誘導に対する堅牢性を検証するため、11パターンの攻撃シナリオを用いた「Red Teaming」を実施しました。
 - 定量評価（Security Resistanceスコア 0.75以上）において、特定の思想への偏りや感情的な操作に惑わされないことを実証し、公平な運用が可能であることを証明しました。
- 国際的な技術標準・ネットワークへの接続
 - 国際的なファクトチェックの規格統一を目指し、IFCN（国際ファクトチェックネットワーク）理事へのアプローチを行いました。
 - 募集を開始したGlobal Trust Challenge (<https://www.globalchallenge.ai/>) への応募を予定しています。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

6-1. 普及啓発活動の全体像

普及啓発活動に係る取組・成果の全体像

本事業では、偽・誤情報対策技術に対する「検閲」や「言論統制」といった社会的懸念を払拭するため、報道機関および専門家を対象とした「クローズドな信頼醸成プロセス」を徹底しました。

開発段階からステークホルダーを巻き込む「共設計」アプローチにより、技術の透明性を担保しつつ、実利ベースでのソリューション開発を行いました。

また、ユーザーマニュアルの整備等を通じて、技術の正しい理解と安全な利用を促進する基盤を確立しました。

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

本事業では、技術の社会的受容性を高めるため、不特定多数への性急な露出を避け、情報のプロフェッショナル（報道機関・専門家）との信頼構築を優先する「段階的啓発戦略」を採用しました。具体的には、以下の3つのレイヤーで活動を展開しました。

1. ステークホルダーとの「共設計」による信頼醸成（Tier 1/Tier 2）

「AIが真偽を勝手に決める」という誤解を防ぐため、開発段階から主要メディアや専門家を巻き込み、彼らの意見を反映させることで「自分たちのためのツール」という認識を形成しました。

- 報道機関（Tier 1）向けクローズドブリーフィング
 - 新聞社・テレビ局に対し、クローズドなヒアリングや説明会を実施しました。
 - 技術の限界（AIはあくまで支援）や倫理的配慮（検閲意図の欠如）を丁寧に説明し、現場記者の懸念を払拭するとともに、フィードバックをUI/UX改善に直結させる「相互信頼型プロセス」を実践しました。
- 専門家（Tier 2）とのクロスレビュー
 - ファクトチェック団体やジャーナリスト、研究者に対し、プロトタイプを用いたクロスレビューを実施しました。
 - 「陰謀論的解釈の含有率」や「ソースの信頼性評価」といった繊細な判定基準について、専門家の知見を取り入れることで、ツールとしての公平性と社会的正当性を担保しました。

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

2. 教育現場への展開とリテラシー向上

技術を単なる業務ツールとしてだけでなく、次世代のメディアリテラシーを育成するための「教材」として位置づけました。

- 教育プログラムへの技術提供
 - 株式会社インフォハント（代表：安藤氏／総務省地域情報化アドバイザー）と連携し、教育現場での利用の検討を行いました。
 - 学生や生徒が、AIが提示した「出典」や「論理構造」を元に自ら真偽を考えるプロセスを体験することで、ファクトチェックの重要性を学ぶ機会を創出します。

3. ユーザー向け資料・マニュアルの整備

専門知識がないユーザーでも技術の意図を正しく理解し、安全に利用できる環境を整備しました。

- 操作マニュアルとガイドラインの作成
 - 開発後期（2025年11月）において、エンドユーザー向けの操作説明書を作成し、システム内から常時参照できるよう実装しました。
 - 単なる操作手順だけでなく、「判断支援」というツールの位置づけやリスクへの注意点を明文化し、利用者のリテラシーを補完する仕組みを整えました。
- 成果の発信準備
 - 本実証事業に関する毎日新聞の取材を受け、本開発に関する発信を行いました。
 - 外部のPR会社と連携し、普及啓発活動の詳細な実行計画を作成しました。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

7-1. 技術開発及び社会実装における課題・展望

技術開発及び社会実装にあたっての今後の課題およびそれらを踏まえた今後の展望

<概要>

● 課題

- 偽・誤情報対策市場の不存在と、行政介入の限界により、現時点の対策の担い手を十分に確保できない状況。また、情報空間が外資系プラットフォームに依存しており、ガバナンスを効かせにくい構造的課題がある。

● 展望

- 本事業を通じて、「民間・アカデミア主導による対策基盤」の有効性が示されたため、今後は、「論拠強度検証」や「ゴーストファクトチェック」等の技術高度化を進めるとともに、広告市場（ブランドセーフティ）と連携したエコシステムを構築することで、国内における持続可能な偽・誤情報対策ソリューション市場を創出する。

<詳細>

本実証事業を通じて、ストリーミング動画コンテンツに対する検証技術の有効性は確認されましたが、これを社会インフラとして定着させるためには、技術面・構造面・市場面の複合的な課題解決が求められます。

1. 今後の課題

① 偽・誤情報対策ソリューション市場の未成熟と参入障壁

現在、日本国内には偽・誤情報対策のソリューション市場が事実上存在しておらず、収益化のモデルが確立されていません。そのため、民間企業の参入が進まず、有効な国産ソリューションが誕生しにくい「空白」が生じています。国が主導しすぎると「検閲・事前統制」のリスクが生じるため、行政の役割はリテラシー教育等に限定せざるを得ません。一方で、担い手の一つであるファクトチェック団体は、ヒト・モノ・カネのすべてのリソースが不足しており、持続的な対策を行う体力が欠如しています。

7-1. 技術開発及び社会実装における課題・展望

技術開発及び社会実装にあたっての今後の課題およびそれらを踏まえた今後の展望

② 外資プラットフォーム依存と「Walled Garden」の壁

日本の情報空間は、行政や国内企業のガバナンスが及びにくい外資系プラットフォーム（巨大PF）上に依存しています。PF側は閉鎖的なエコシステム（Walled Garden）を構築しており、外部からの検証データへのアクセスが制限されているため、正確な情報流通状況の把握や対策が困難です。また、API仕様の変更などが検証ツールの安定稼働に直結するため、外部依存度を下げる技術的な冗長性の確保も継続的な課題です。

③ 「存在しない事実（Ghost Fact）」への対抗技術の確立

OSINTをベースとした検証は「存在する情報」の照合には強いものの、「そのような事実は存在しない（証拠がない）」ことを証明する、いわゆる「悪魔の証明」は苦手とします。特に陰謀論や根拠のない噂話に対しては、検索してもヒットしないために「判定不能」となったり、AIが事実ではないことをもっともらしく作り上げてしまったりするリスクがあります。そのため、こうした実体のない情報（Ghost Fact）を精度よく見抜くための検証技術が必要です。

2. 今後の展望・ロードマップ

① 「産学官連携」による新たな対策エコシステムの構築

前述の「市場の不存在」と「行政介入の限界」というジレンマを解消するため、本実証のような「国が支援し、民間・アカデミアが実装する」という枠組みをさらに強化すべきと考えます。具体的には、本技術を想定利用者である報道機関やファクトチェック団体へ安価に提供（SaaS化・マルチテナント化）してリソース不足を補完しつつ、他の領域での転用可能性を高め十分に収益化することで、経済合理性のある持続可能な市場形成を目指します。

7-1. 技術開発及び社会実装における課題・展望

技術開発及び社会実装にあたっての今後の課題およびそれらを踏まえた今後の展望

② 技術的な深化：論拠強度検証とゴーストファクトチェック

単なる「真偽（True/False）」の二元論を超えた、より高度な判断支援機能を実装します。

- 論拠強度検証（Argument Strength Verification）：ツールミンモデルに基づき、主張を支える「データ」「論拠」「裏付け」が論理的に十分かをAIがスコアリングする機能を実装します。これにより、「嘘ではないが論理が飛躍している」といったグレーゾーンの情報を可視化します。
- ゴーストファクトチェック: 「存在しない事実」に対し、AIがあえて反論（Counter-argument）を生成して検証を行う手法を導入し、陰謀論やデマに対する検証精度を飛躍的に向上させます。

③ グローバル標準への接続と「認知セキュリティ」への昇華

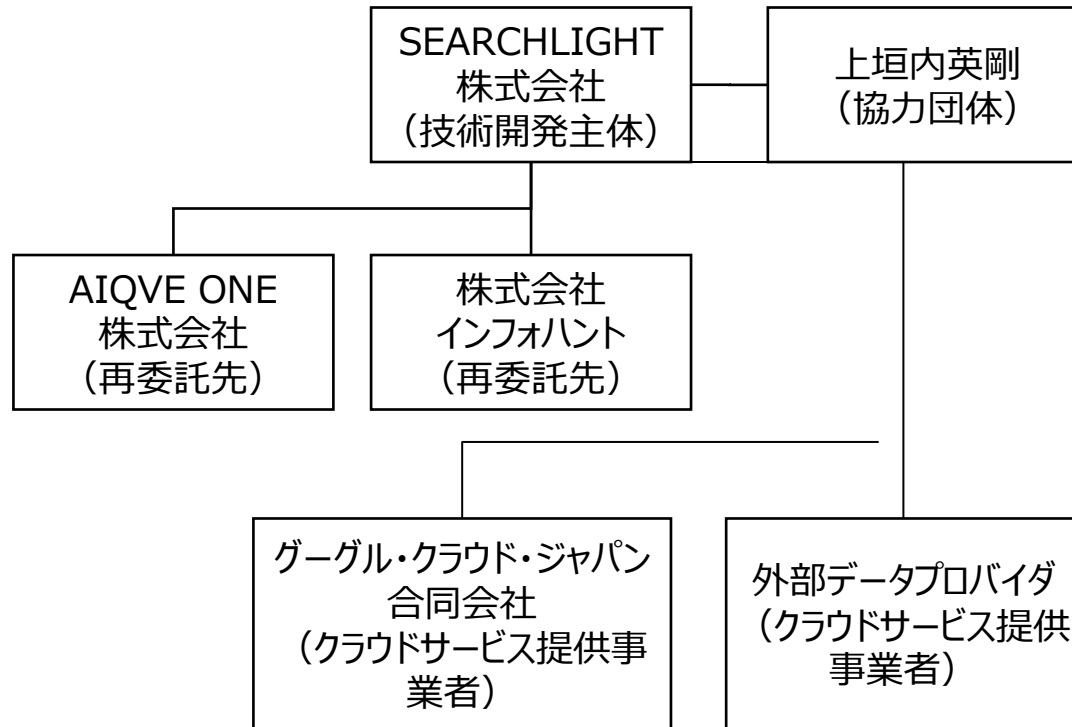
日本特有の「マスメディアへの信頼性が高い」土壌を活かしつつ、表現の自由を尊重した検証文化を根付かせます。将来的には、IFCN（国際ファクトチェックネットワーク）等の国際基準に準拠した技術仕様を確立し、アジア圏におけるWalled Garden対策の標準モデルとして展開します。これにより、単なるツール提供にとどまらず、社会全体の「認知セキュリティ（Cognitive Security）」を担保するインフラとしての地位を確立します。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

8-1. 実施体制及び役割分担

本事業の実施体制図



各団体の役割・業務範囲

SEARCHLIGHT株式会社：

- 本実証全体の統括
- 本実証における開発のリード

AIQVE ONE株式会社：

- 本実証の開発全般

株式会社インフォハント：

- 本実証の協力者やヒアリング対象者等を選定支援

上垣内英剛

- 本実証における開発に関する助言指導

グーグル・クラウド・ジャパン合同会社

- Google Cloudのクラウドサービス (Google Cloud) 提供

外部データプロバイダ

- 検索API等のデータ収集ツール

8-2. 全体スケジュール

主な実施事項	令和7年						令和8年	
	8月	9月	10月	11月	12月	1月	2月	3月
(1) 対策技術の開発								
1. プロジェクト立ち上げ	→							
2. 教師データ収集・整備	→	→	→	→	→			
3. 主張自動集出モデル開発	→	→	→					
4. 論理構造可視化機能開発			→	→	→	→		
5. 出典探索・照合機能開発			→	→	→	→		
6. フィードバック設計・実装					→	→	→	
(2) 有効性等に関する検証及び調査								
1. 検証用データ整備・評価データ作成	→	→						
2. 技術的有効性検証			→	→	→	→	→	
3. 社会実装に向けた実証・課題整理			→	→	→	→	→	
4. 検証結果の分析・評価			→	→	→	→	→	

8-2. 全体スケジュール

主な実施事項	令和7年						令和8年	
	8月	9月	10月	11月	12月	1月	2月	3月
(3) 社会実装に向けた取組								
1. ステークホルダーとの対話・ヒアリング	→							
2. ビジネスモデル検討・策定		→						
3. グローバル接続性・標準化検討			→					
(4) 成果報告書及び社会実装実施計画書の作成								
1. 成果報告書書作成					→			
2. 社会実装計画書作成					→			
3. 報告書納品							→	
(5) 普及啓発活動への協力								
1. 普及啓発戦略策定			→					
2. Tier1/Tier2向けクローズド報告会・ブリーフィング実施					→			
3. 普及資料・レポート作成					→			
4. 成果報告会実施							→	