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

画像・動画を中心としたSNS上の投稿の真偽判定システムの開発・実証
成果報告書 概要版

2026/3/19

技11_Sakana AI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

- Xを始めSNSは偽・誤情報の主戦場となっており、社会に大きな影響を与えている。この偽・誤情報の対策として、ファクトチェックがあるが、これは様々な目線でチェックを行うマニュアル作業であり、膨大な工数がかかっているほか、専門性が高く、かつ、昨今の生成AI技術の進歩によりその判定はより難しくなっている。膨大な偽・誤情報の中から、どの情報を対象に、いかに効率よく、標準的に、高度に検知するかが、偽・誤情報対策の鍵となっている。

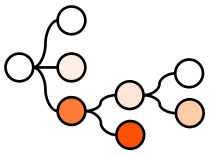
SNS上の膨大な情報	偽・誤情報の複雑、巧妙化	打ち手
限られたリソースで どの情報をチェックすればよいのか？	膨大なチェック工数 人間の目で見極めることができない	有効な打ち手がわからない

上記課題を踏まえ目指す姿・ゴール

- 本実証事業では、SNS空間で拡散される①偽・誤情報・ナラティブを可視化し、真偽判定するべき情報の見極めと、②その真偽判定の実施、③及び対策案の立案を支援するシステムを開発する。これにより、プラットフォーム事業者や一般ユーザが早期対応・判断できる健全な情報流通環境の構築を目指す。

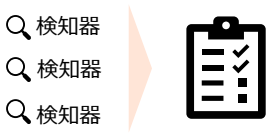
① SNS空間の可視化

階層ナラティブツリーによる分析



② 総合的な偽・誤情報判定

複数検知器の組み合わせ



③ 対策案の立案

ABM※によるシミュレーション

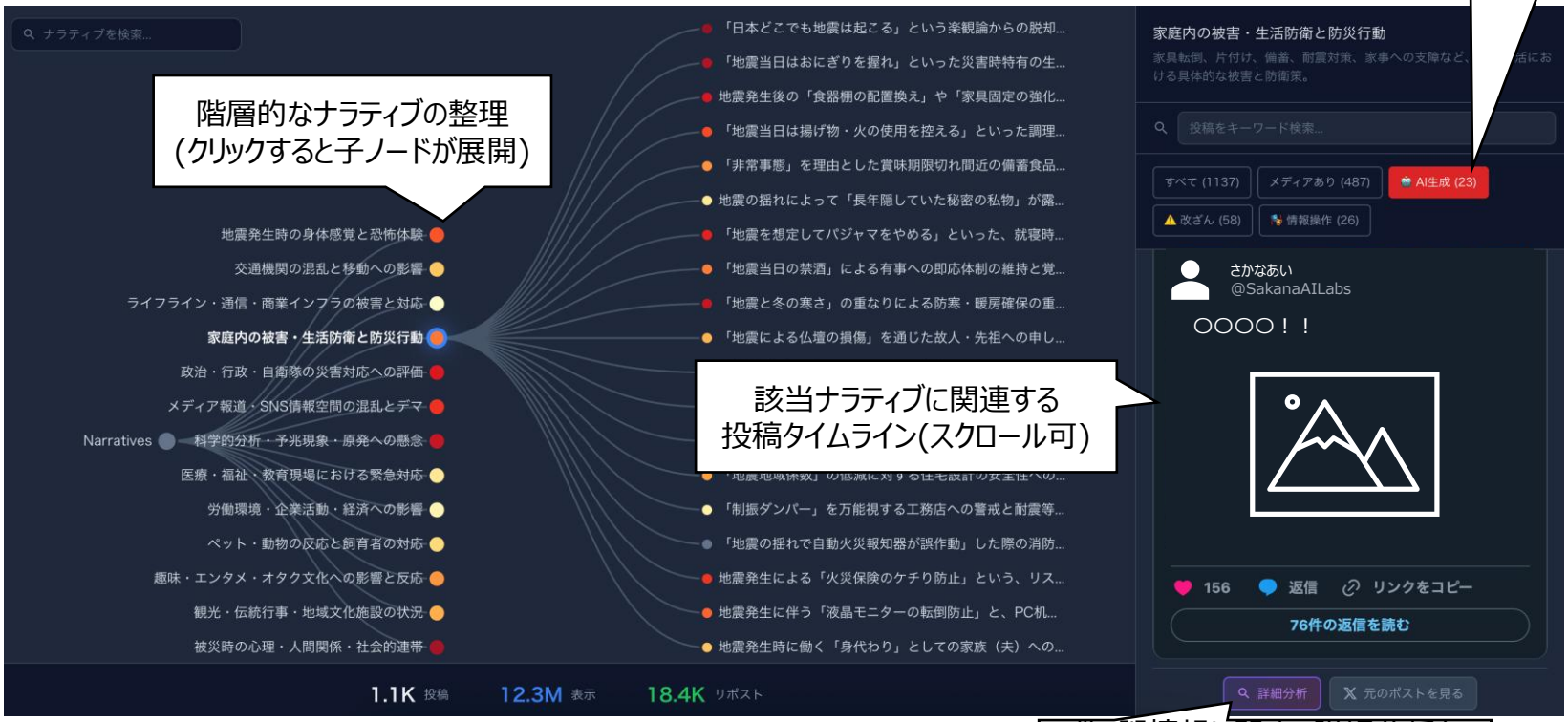
※Agent Based Modeling：本実証ではSNSのアカウントを生成AIで模擬してSNS空間上のSNSの動きを再現する取り組み

1-2. 技術開発の取組・成果

機能① SNSの言論空間の可視化

- SNSの言論空間の可視化方法として、階層ナラティブツリーマップを提案。インプレッション数などの単一的な指標では捉えられない有機的な言論空間を高解像度かつ網羅的に補足することができる。これにより、人手による全量確認が不可能な状況下においても、全体像の即時把握と、対処すべき情報の優先順位付けを可能にした。
- それぞれのナラティブは単純なタグやカテゴリではなく、社会に波及力を持つ「具体的な論調」であり、 これを探索するためのノベルティサーチ技術※を開発した。

※ ノベルティサーチ技術とは、Xの投稿データを再帰的にサンプリングし、新規性の高いナラティブを効率的に探索・抽出する技術

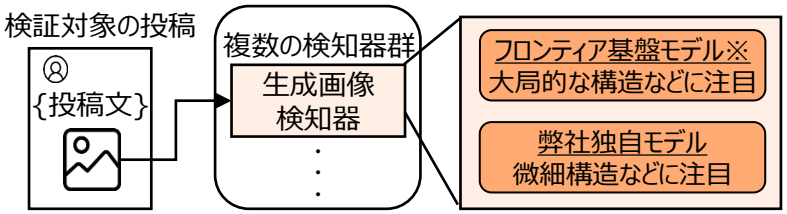


1-2. 技術開発の取組・成果

機能② 偽・誤情報の真偽判定のための各種検知機能

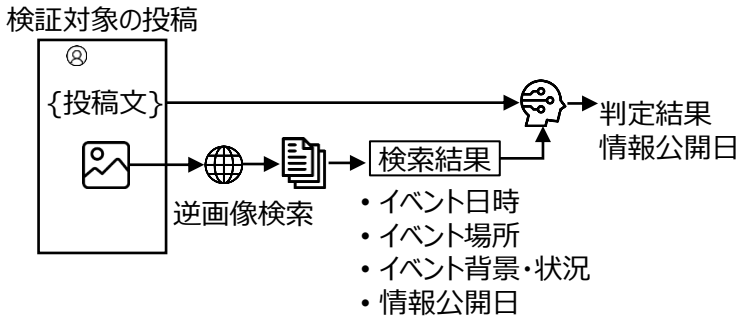
- 巧妙・複雑化する偽・誤情報を判定するために、様々な角度から検証する検知器を開発。

画像・動画の生成・加工の検知

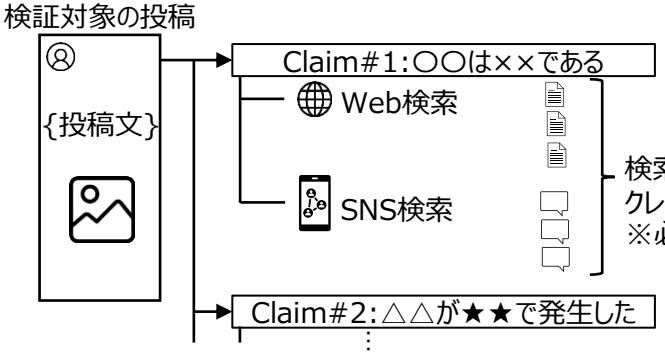


※ フロンティア基盤モデルとは、GPT-5やGemini-3など、世界最高水準の性能を有する最先端の汎用大規模言語モデルのこと

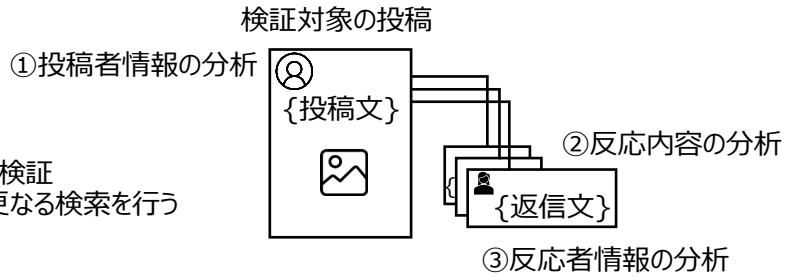
画像・動画の文脈チェック



真偽判別



ユーザー分析/反応分析

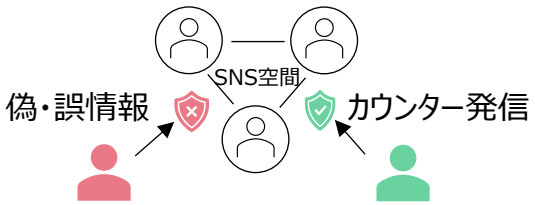


1-2. 技術開発の取組・成果

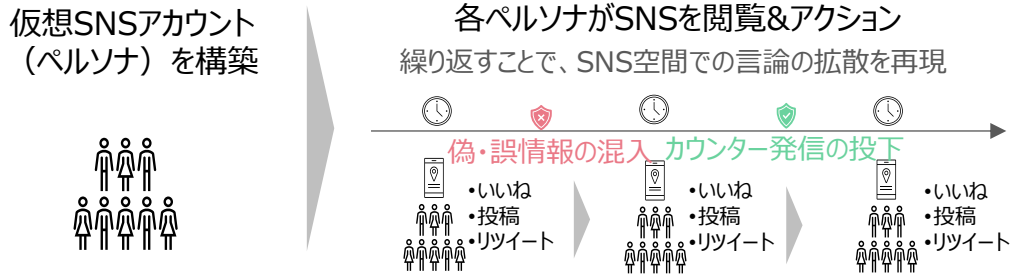
機能③ 対策案立案のためのSNS空間のシミュレーション

- 偽・誤情報の拡散を抑制・鎮静化させる対策として、正確な情報を戦略的に発信し、誤った認識を打ち消す「カウンター発信」の有効性を検証。そのために、SNS言論空間およびユーザーの振る舞いを精緻に再現するシミュレーション基盤を開発した。(弊社独自フレームワークShachiを活用)

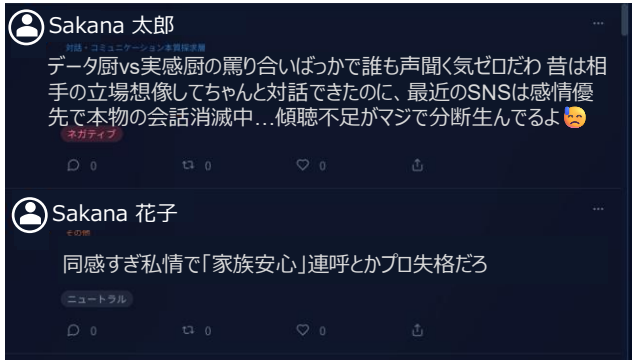
カウンター発信のイメージ



SNS言論空間のシミュレーション方法



ペルソナの反応例
（SNSのリアルなやり取りを再現できる）



（一部記載表現を変更）

カウンター発信の効果検証
（ペルソナの反応を集計し、複数のカウンター発信案の効果を比較できる）

目次

2-1. 社会実装に係る取組・成果

実証実験の結果

- 実効性の高い偽・誤情報対策システムを構築するため、最前線で偽・誤情報対応にあたる有識者に対し、広範かつ段階的なヒアリングとレビューを実施した。
- 現場の課題と業務要件を精緻に反映させて開発した本プロダクトは、中曽根平和研究所 情報空間のリスク研究会様より高く評価され、実務におけるその有効性が確認された。

事前・中間ヒアリング

偽・誤情報に係る課題を聴取し、業務要件と開発方向性を決めることを目的に次の協力団体様に対して幅広くヒアリングを実施

- **LINEヤフー株式会社様**
- **ファクトチェック有識者団体**
- **(株)Classroom Adventures様**

レビュー

本実証システムは情報の正確性を極めて重視するユーザーを対象としているため、評価者にはシンクタンクや学識者といった専門家を選定した。具体的には、国家安全保障に係る調査、分析、政策提言を行う中曽根平和研究所様の複数の研究員に依頼し、有識者レビューを実施。

有識者レビュー結果：公益財団法人 中曽根平和研究所 情報空間のリスク研究会 様

- 従来は把握が困難であったSNS上の情報空間を、体系的に理解できるようになった。
- 膨大なSNSデータに対し、人間の力では困難だった情報の抽出・整理が容易になった。特に画像、動画を一気通貫で取り扱える点がよい。
- プレバンキング（偽・誤情報の拡散防止措置）としても期待できる。

2-1. 社会実装に係る取組・成果

普及啓発活動

- 安全保障およびインテリジェンス関連のイベントや研究会へ参画し、各界・諸団体との連携強化を図った。
- 当初、本システムの主要ユーザーはメディア事業者やファクトチェック団体を想定していたが、一連の活動を通じ、安全保障分野の研究者、政策立案者、実務者からも高い関心が示され、導入候補先の外延を広げることに成功した。
- また、ユーザー属性ごとに「関心事」や「分析の切り口」が異なる実態も明らかとなり、今後のシステム導入アプローチをより精緻に具体化することに繋がった。

防衛装備庁技術シンポジウム2025

出典：防衛装備庁 技術シンポジウム2025
(https://www.mod.go.jp/atla/research/ats2025/index.html)

Landpower Forum in Japan 2025

イベント名称	Landpower Forum in Japan（ＬＦＪ）
開催日時	令和７年１２月１７日（水）０９：３０～１７：００ 令和７年１２月１８日（木）０９：３０～１５：１５
場所	東京ドームシティ・プリズムホール （東京都文京区後楽１丁目３-６１）
主催	陸上自衛隊（陸上幕僚監部、教育訓練研究本部）
協力団体	公益財団法人陸修偕行社、 一般社団法人日本防衛装備工業会、 一般社団法人日本ＵＡＳ産業振興協議会
展示規模	７６ブース（８８社）
入場条件	完全事前登録制（入場審査通過者のみ） ※ＷＥＢ入場申請の内容によって、入場不可の場合があります ※入場料無料

出典：Landpower Forum in Japan 2025
(https://k3rws.stage.ac/LFJ2025/)

令和7年度防衛産業参入促進展

出典：令和7年度防衛産業参入促進展
(https://www.atla-event.com/2025su/)

中曽根平和研究所 情報空間のリスク研究会

出典：中曽根平和研 情報空間のリスク研究会
(https://www.npi.or.jp/research/2026/01/06125815.html)

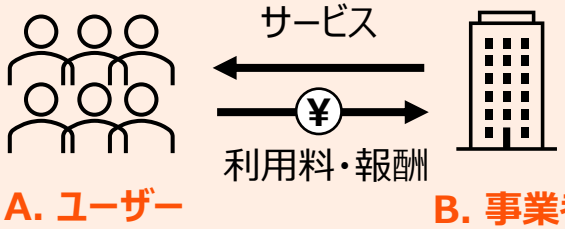
中曽根平和研究所 公開ウェビナー

出典：中曽根平和研 公開ウェビナー
(https://npi.or.jp/event/2026/01/29135036.html)

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

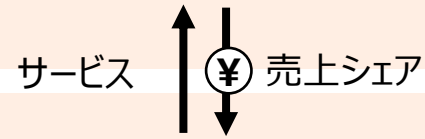
ユーザ



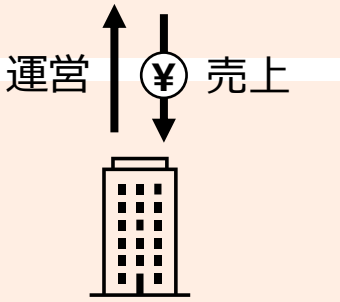
本実証で策定した仮説

A.ユーザー	B.事業者
一般ユーザー	メディア事業/報道機関
情報利用者	ファクトチェック団体
政府、官公庁	安全保障シンクタンク

サービス



関係事業者・サービス提供者



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

- 事業者ごとの取り組みとペインポイントの分析・整理により、実効性の高い開発方針と導入アプローチを具体化した。

	①メディア事業	② ファクトチェック団体	③ 安全保障関連組織 / 研究者
事業特性	情報を伝える立場として、**正確性、即応性、新規性**が求められる。	情報空間の健全性を保つ立場として、**偽・誤情報を探す**ことが求められる。	情報を利用する立場として、**正確な情報に基づき、分析、提言**をすることが求められる。
ペインポイント	事実確認（裏取り）においては、**多角的な証拠収集と緻密な検証**が不可欠。一方で、情報の鮮度を保つための**即応性**も求められる。特に**災害時**において、この両立は重要。	「嘘をつくのは一瞬だが、検証には数日かかる」という非対称性の中で、拡散力が高く社会的に**有害な偽・誤情報を優先的に特定**（トリアージ）することが難しく、経験による差が出やすい。	**画像・動画**を含んだ情報の分析は多くの工数を要する。 また、**情報の出所**や**拡散経路**の特定に加えて、大衆の**感情**をどう操作しようとしているかといった様々な分析視点が求められる

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

- 本実証事業開始時点よりも多くの事業者への導入可能性が確認された一方で、事業者ごとに分析の観点や切り口が異なることも明らかとなった。今後は共通基盤とカスタマイズ領域の線引きを明確化し、汎用性と個別最適のバランスを定義した上でプロダクション化を進める必要がある。

上記課題を踏まえた今後の展望

- 想定導入先との協議を通じて、多様なニーズを効率的に満たす「共通基盤」と「カスタマイズ領域」の境界を明確化・標準化し、スケーラブルなプロダクトとして製品化・商用化を推進していく。

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

- 本実証を通じて関係を構築した各分野（メディア・ファクトチェック・安全保障）の導入見込み先と協議を重ね、成果物の社会実装を推進する。初期導入における運用知見を製品開発（共通基盤と専用モジュールの高度化）にフィードバックしつつ、段階的にビジネスモデルを固め、最終的にはデジタル空間の信頼性を支える社会インフラとしての定着を目指す。

フェーズ1（2026年度）

先行パートナーへの導入と運用実績の確立

- メディア事業、ファクトチェック有識者団体、安全保障研究機関などの既存の導入見込み先に対し、PoCに着手する。
- 実際の業務フロー内での活用実績（サクセスケース）を積み上げつつ、現場からのフィードバックに基づき、UI/UXや分析機能の最適化を行う。

フェーズ2（2027年度）

サービスモデルの確立と普及拡大

- フェーズ1の知見を基に、「共通基盤＋用途別モジュール」のパッケージ製品としての仕様を確定させ、SaaS型等のスケーラブルな提供モデルを整備する。
- 確立した導入事例を武器に、各業界内のより広範な事業者（地方局、他シンクタンク等）へと導入を拡大し、収益基盤を強化する。

フェーズ3（2028年度）

社会インフラ化と適用領域の拡張

- API連携等を通じて、SNSプラットフォーム事業者や外部セキュリティツールとの相互運用性を確保し、偽・誤情報対策のエコシステム（社会インフラ）を構築する。
- また、蓄積された脅威データや検知技術を応用し、民間企業のブランド保護（リスク管理）など、より広い市場へも事業を展開する。