

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

**AIを活用した情報コンテンツの
真偽判別支援システムの開発・実証
成果報告書**

2026/3/19

技14_日本電気株式会社

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

AIを活用した情報コンテンツの真偽判別支援システムの開発・実証

アプローチする課題・目指す姿

- 複合的なコンテンツの組み合わせによる偽・誤情報への技術的対策については、マスメディア企業や自治体、その他企業において個別に対策・検討が進められているものの、手動分析に多くの工数を要し、業務負荷が増大しているという課題がある。
- 本開発では、AIを活用した偽・誤情報分析支援システムを用い、昨年度開発したシステムに対するユーザヒアリング等から得られた技術的課題を改善することで、①総合的な真偽判定の高度化を実現する。あわせて、SNS等多様な利用環境やユースケースに対応可能な機能拡張を行い、②マスメディア企業や自治体、その他企業・団体との連携による社会実装を目指す。

技術区分

①コンテンツの真偽判別支援・改ざん検知技術

対象とするモジュール種

文章、画像、音声、動画

実施体制 (下線:技術開発主体)

日本電気(株)、東京大学、NECソリューションイノベータ(株)、日本アルゴリズム(株)

技術開発の取組・成果

- コンテンツの真偽判別支援技術のうち、複合的なコンテンツを真偽判別する技術について開発した。令和6年度で開発した真偽判別支援技術をベースと、実証事業の中で明確になった課題を解決するための機能強化及び機能の追加。具体的な機能強化及び機能の追加として大きく以下をおこなった。
 - 総合真偽判別の高度化
 - 様々な利用ユーザの意図に応じた分析の実現とSNS等の分析対象情報の検知機能追加
- 技術開発におけるKPIとしては、令和6年度の開発技術で偽・誤情報の真偽判別に失敗する情報コンテンツに対して80%以上判別可能となったことを確認しKPIを達成した。

社会実装に係る取組・成果

- ビジネス視点として、本実証により社会実装の利用主体の拡大について定量的に評価するKPIを設定した。具体的には20社以上のマスメディア(放送局・通信社・新聞社)と2つ以上の自治体でのユースケースに沿って開発したシステムの実証・ヒアリングを行うことを目指した。また、展示会への出展を2件以上(例:マスメディア向けInterBEE、他業界向け展示会等)、学会での論文採択・発表1件以上(例:AI系学会のAAAI等)を行い、社会実装の拡大に向けたユーザ探索を積極的に進めることを目指すこともKPIとして設定していた。
- 上記において69社のマスメディア(放送局・通信社・新聞社)と3つの自治体へシステムの実証・ヒアリングを実施。展示会への出展を3件、会での論文採択を1件実施したためKPIは達成となった。

技術開発及び社会実装にあたっての課題・展望

- 今年度予定の開発項目を組込み、技術開発上のKPIとしていた誤判別率の向上は達成できたものの、ユーザの試用により、以下のような課題が確認された。
 - 専門用語に対応できない場合がある
 - レスポンスタイムが想定よりも長い
 - ユーザが望むユースケースである「生成AIによって加工・作成されたフェイク画像の検知」については正確な判別をおこなうことが難しい場合がある
- 上記を踏まえ、本事業の展望としては以下が考えられる。
 - ユーザの利用頻度が高いと想定される領域を強化したソリューション開発(専門用語集対応含む)
 - 偽誤情報判別ジョブの並列化
 - 生成AIによる画像フェイク検知の精度向上

代表者コメント



日本電気株式会社
メディアプラットフォーム
統括部
井上有二郎グループ長

昨年度の成果を踏まえ、本年度は技術面・社会実装面ともに活動の幅を広げることができた。特に、実証を通して多様な利用シーンでのニーズを把握し、それに応じた機能強化を進められたことは大きな収穫である。

また、メディアや自治体との連携、展示会や学会での発信を通じて、社会実装に向けた基盤を着実に築くことができた。

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

【社会的背景】

- 近年、生成人工知能（AI）技術の急速な発展とインターネット・ソーシャルメディア（SNS）の普及に伴い、偽情報・誤情報コンテンツ（テキスト・動画・画像・音声等の複合型）がインターネット上に蔓延している。これに伴い、これらの情報が一般国民の日常的な情報摂取環境に入り込み、各界に影響を与えるリスクが増大している。
- 特に、AI生成コンテンツによる民主主義プロセスへの不当な介入が顕在化し、社会不安を引き起こす懸念が顕著となっている。

【現状課題】

現状、情報の正確性が求められる分野における対応が急務となっている。

- マスメディア（放送局・通信社・新聞社等）
信頼性の高い情報発信が求められる立場において、偽情報の判別業務に人的リソースを集中させており、業務負荷が増大している。
- 自治体等
災害時などの緊急事態においては、公的機関の正確な情報発信が至上命題となる中で、偽情報の混入防止が深刻な課題となっている。

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

【本実証事業の目的】

- 本事業は、複合的な情報コンテンツ（テキスト・動画・画像・音声）に対する真偽判別支援技術の開発・実証を通じて、上述の事象への対策と社会実装を加速させることを目的とする。
- 具体的には、昨年度（令和6年度）にNECが実施した「AI活用による情報コンテンツ真偽判別支援技術」実証事業の中で明確になった課題を解決する機能の強化や追加を行う。

【短期・中長期における提供価値】

短期

ファクトチェック業務を行っている人々のファクトチェック分析支援を行うことで業務効率化による真偽判定業務のスピード向上に貢献

中・長期

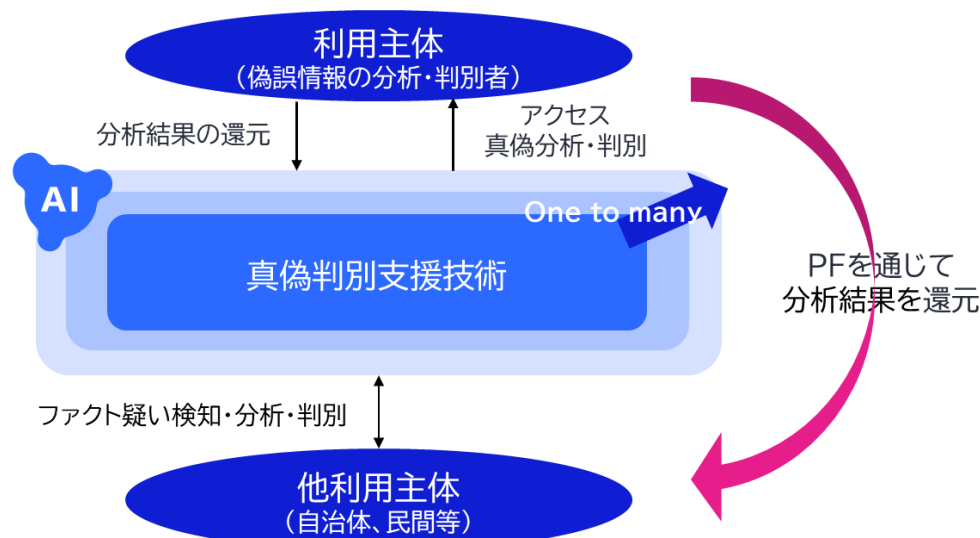
日本の生活者が偽誤情報か否か分かったうえで情報に触れることが出来る社会を実現する

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

【最終的に目指す姿】

- 専門的に真偽判別を行っている利用主体の分析結果やナレッジを他利用主体へ還元していく仕組みを本実証実験で構築するプラットフォームに追加していくことを検討する。サービス利用型のモデルで、利用主体を民間へも広げていくことで、各企業で自ら事業ができる自走可能なビジネスモデルを検討する。また、長期的にはB to B to Cでの国民のシステム利用を目標としていく。
- 提供形式は、One to Manyが可能なサービスプラットフォーム型を想定する。本領域は常にAI技術の最新化が必要であり、サービス型とすることで利用主体は意識することなく最新の価値を受けることが可能となる。事業側も個々の利用主体への対応が基本的に不要となり、高い費用対効果が見込める。また、サービス型とすることで、必要に応じて新規機能を短期間で提供することも可能となり、事業成長も行いやすい形になると考える。



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

【マスメディアにおける活用】

- 近年、マスメディアでは、取材情報だけでなくSNSなど、多様なインターネットソースの情報を番組内で用する機会が増加している。これに伴い、コンテンツの真偽判別作業は重要な業務となっているが、人力による判定は多大な工数を要する。
- 本実証で開発された技術により、以下の効果が期待される。

期待される効果	詳細
工数削減	コンテンツに適して自動化された真偽判別を行うことで、判定プロセスを効率化。
偽・誤情報判別率向上	位置情報の活用や自然言語処理技術の高度化により、判別精度を向上。
情報の信頼性維持	正確な情報の提供を通じて、社会に対するマスメディアの信頼性維持に貢献。

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

【自治体における活用】

- 災害時には、SNS上で大量の情報が流通する反面、偽情報や誤情報が拡散するリスクが高く、自治体の公式情報が埋もれるなどの問題が発生している。
- 本技術の活用により、自治体では以下の運用が可能となる。

期待される効果	詳細
迅速な情報選別	偽・誤疑いの情報を自動検知し、真偽を判定する仕組みを構築。
より効果的な情報発信	正確な情報を迅速に国民に伝えることで、公式情報が埋もれることを防ぐ。
社会秩序の維持	正確な情報の認知が早まることで、社会的な混乱の低減。

目次

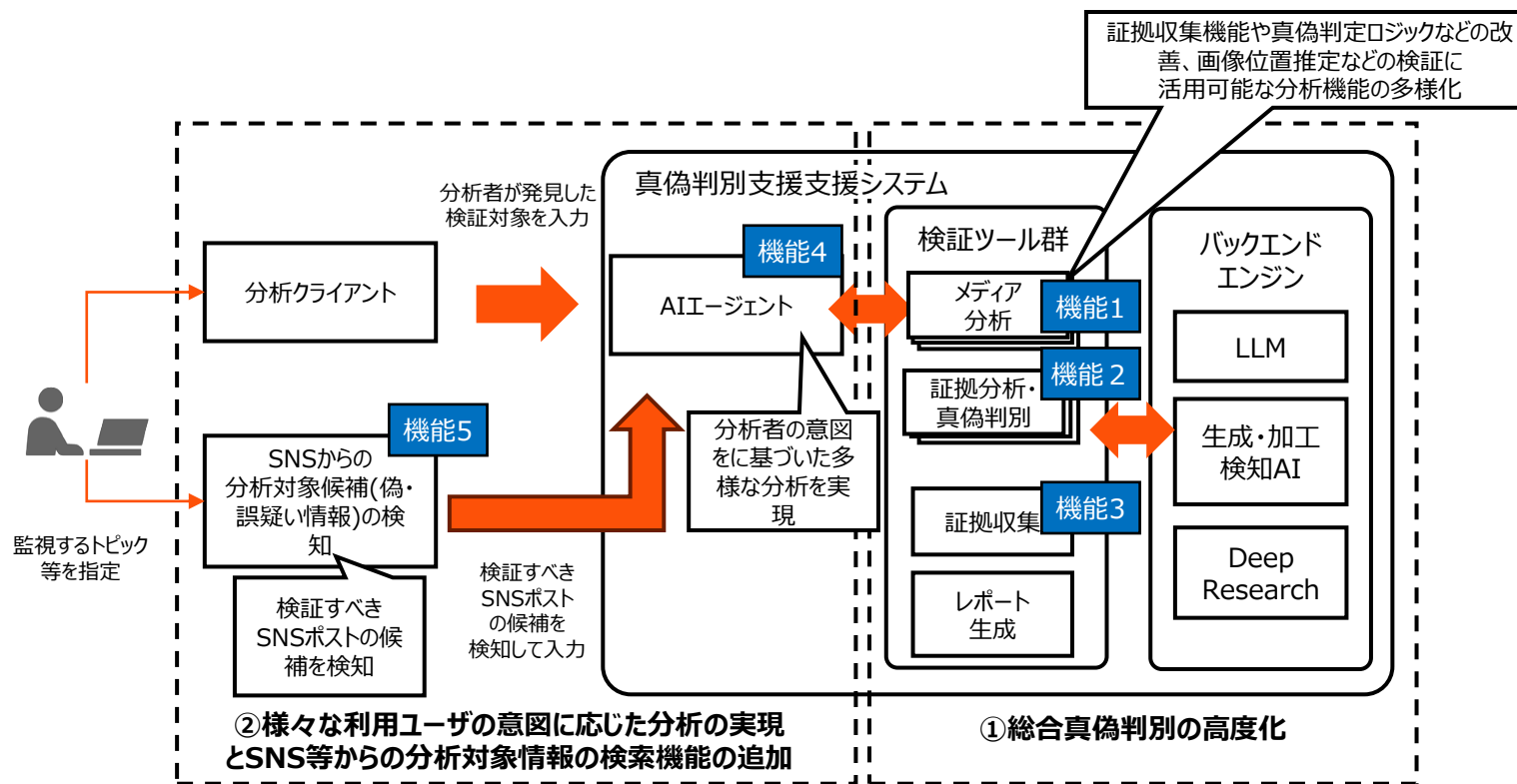
1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「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

インターネット上の偽・誤情報等への対策技術として、コンテンツの真偽判別支援技術のうち、複合的なコンテンツを真偽判別する技術について開発した。令和6年度で開発した真偽判別支援技術をベースとし、実証事業の中で明確になった課題を解決するための機能強化及び機能の追加した。

具体的な機能強化及び機能の追加として大きく、①総合真偽判別の高度化、②様々な利用ユーザの意図に応じた分析の実現と SNS 等の分析対象情報の検知機能追加をおこなった。次頁以降に、機能強化／機能追加した各ポイントについて記述する。



3-1. 技術開発の全体像

類似技術との差別化ポイント

既存の類似技術としては、以下にあげる2つの技術が考えられる。

技術名・サービス名	特長
①DeepResearch	AIエージェントを利用した高度な検索機能をAI各社がサービス提供している
②BCP※関連情報提供サービス	気象データ・道路／河川カメラ、SNS上の有益な情報から、災害時に必要なものを収集

※：Business Continuity Plan（事業継続計画）

これらに対し、本技術では以下の点について差別化することが可能と考えている。

- ① 既存技術は基本的に解析の基本技術として「検索」を軸としている。一方、本技術は検索のみならず検索だけでは収集できない情報の分析（位置情報推定やSNSマイクロ分析、ディープフェイク分析など）など必要な分析を自動的にを行い、それらを用いて総合的に判断することが可能。
- ② 既存技術は、位置が不明なSNSデータについては、投稿に付随する画像等を活用して位置を推定するが、画像中に文字情報（看板・標識）やランドマーク等が必要であり、特定可能なシーンが限定的。弊社保有の技術は投稿画像と上空画像（衛星画像・航空写真等）を照合することで位置を推定するため、高い網羅性で広範囲をカバー可能。

3-2. 技術開発の個別詳細

偽・誤情報判別支援システム 性能評価結果

- 性能評価結果
 - 評価方法: 第三者機関によって公開されている真偽判別レポートを用い、同等の検証が可能かを評価し、令和6年度の開発技術で誤判定したケースに対する改善度合いを評価
 - 分類評価: 5段階の判定結果を「正確」と「誤り」に2段階にまとめて判定結果が一致するかどうかを評価
 - レポートロジック評価: 収集された証拠や判定根拠のロジックが専門家のレポートと類似しているか主観評価
- 評価結果:
 - 分類評価: 16/17件を正しく判定 (94.1%)
 - レポートロジック評価 14/17件で類似 (82.4%)
- 誤判定したケースに関する分析
 - 誤分類やロジック誤りの主な要因はフェイク検知器の誤判定に起因することを確認

【KPI達成状況】

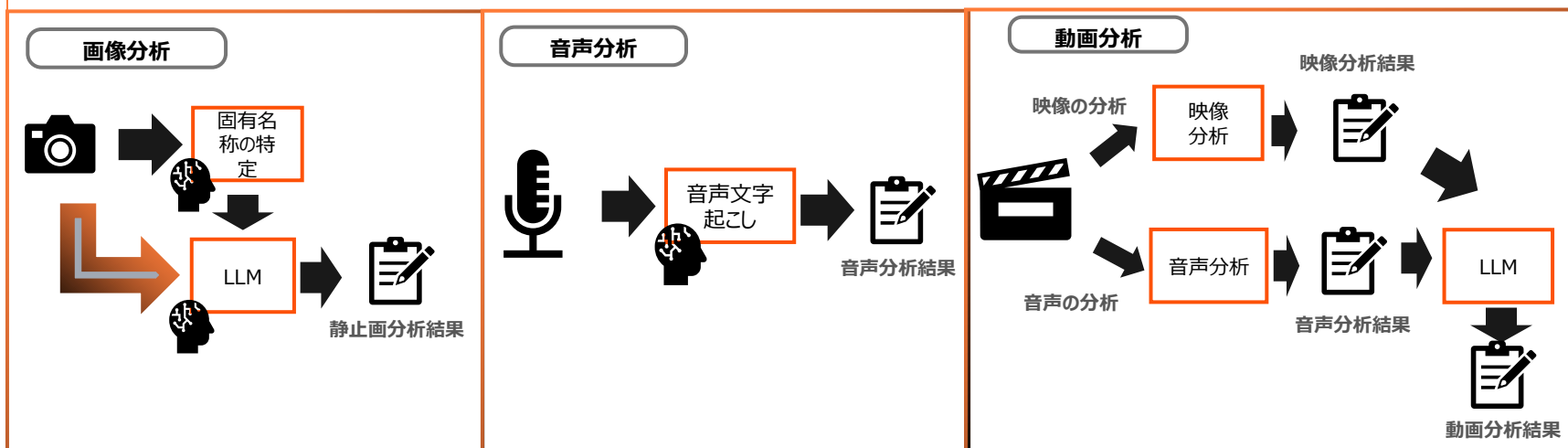
KPI	達成状況
令和6年度の開発技術で偽・誤情報の真偽判別に失敗する情報コンテンツに対して50%以上判別可能となることを目指す。	令和6年度の開発技術で偽・誤情報の真偽判別に失敗する情報コンテンツに対して80%以上判別可能となったことを確認しKPIを達成した

3-2. 技術開発の個別詳細

機能1 メディア分析技術 (1/2)

【概要】

- 真偽判別対象として入力されたマルチメディアコンテンツの内容を分析し、テキスト情報に変換する機能
- 画像・音声・動画の内容をそれぞれに対応した分析機能により分析を行い、それぞれ内容を説明するテキスト情報に変換
 - 画像分析：画像の内容を理解した分析結果をテキストで出力
 - 音声分析：音声を文字起こした結果をテキストで出力
 - 動画分析：動画の内容を理解した分析結果をテキストで出力
- 今年度は偽・誤情報判別の精度向上を目的として、**動画の内容理解をより精緻に行うための映像分析技術**アップデートを実施



3-2. 技術開発の個別詳細

機能1 メディア分析技術 (2/2)

【詳細】

- 従来技術：動画からサンプルしたフレームごとにLLMで分析し、フレームごとの結果を要約
- 新技術：動画からサンプルしたフレームを同時にLLMに分析させることで動画分析結果を得る
- 評価結果：
 - 動画を用いたQA用データセット NExT-QA[1]のテストデータを用いてQAタスクで性能を評価
 - 分析に利用するフレームを8-128で変化させて評価
 - 新技術についてはフレームのサンプル方法をランダムと等間隔のケースで評価
 - 従来技術と比較して少数フレームでも性能を発揮し、64Frame以下のケースで8-9%の性能向上を確認



	8 Frame	16 Frame	64 Frame	128 Frame
従来[等間隔]	75%	77%	79%	81%
改善[ランダム]	84%	85%	87%	83%
改善[等間隔]	85%	85%	86%	86%

[1] X. Junbin, et al., "NExT-QA: Next Phase of Question-Answering to Explaining Temporal Actions," in Proc. CVPR 2021.

3-2. 技術開発の個別詳細

機能2 ディープフェイク検知技術 (1/2)

【概要】

- 偽・誤情報判別支援システムでは機能: 静止画、動画、音声を入力として、それぞれが四種類の生成・加工物であるか判定した結果と生成・加工物である確率を出力
 - 顔ディープフェイク検知器: 動画・静止画に含まれる顔がディープフェイクであるか判定
 - 音声ディープフェイク検知器: 音声ディープフェイクであるか判定
 - 画像部分加工検知器: 静止画が部分加工されているか判定
 - 風景画像ディープフェイク検知器: 静止画が生成AIにより生成されたものか判定
- 令和7年度開発では**顔ディープフェイク検知器**と**風景画像ディープフェイク検知器**をアップデートし性能を向上

顔ディープフェイク



Face Swap: 顔の入れ替え[1]



Reenactment: 表情の操作 [1]

音声ディープフェイク

“文章”⇒“音声”

Text to speech:
テキストを音声化

“音声”⇒“音声”

Speech to speech:
他者の声に変換

部分加工



Copy-move: オブジェクトを複製[2]



Inpainting: オブジェクトの除去[2]



Splicing: 他画像の貼り付け [2]

風景画像ディープフェイク



GAN modelによる生成 [3]



Diffusion modelによる生成 [4]

[1]: Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward, Applied intelligence, 2023

[2]: IML-VIT: Benchmarking Image Manipulation Localization by Vision Transformer, AAAI, 2024

[3]: Large Scale GAN Training for High Fidelity Natural Image Synthesis, ICLR, 2019

[4]: <https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0>

3-2. 技術開発の個別詳細

機能2 ディープフェイク検知技術 (2/2)

【詳細】

- 顔ディープフェイク検知器のアップデート
 - 最新の難関国際学会で発表かつ、ライセンス上商用利用可能なモデル(ForensicsBench[1])を選定し評価
 - 公開顔ディープフェイクデータ (KoDF[2]) を用いて精度評価
 - 昨年度利用していたCADDM[3]より精度改善を確認したため方式を更新
- 風景ディープフェイク検知器のアップデート
 - 最新の難関国際学会で発表かつ、ライセンス上商用利用可能なモデル(AIDE[4], ForensicsBench[1], AeroBlade[5])を選定し評価
 - 災害(洪水、地震、火事等)画像に限定した自作データセットおよび公開データセット ([6, 7]) を用いて評価
 - AIDEが昨年度方式(UFD[6])と得意・不得意なケースを補完できる関係にあるため、両方式を併用

顔DF検知評価結果

手法	Accuracy	Precision	Recall
CADDM	0.751	0.258	0.242
ForensicsBench (GPT-4.1 Low)	0.863	0.741	0.303
ForensicsBench (GPT-4.1 High)	0.915	1.00	0.5

組み込み済みDF検知器

風景DF検知評価結果 (自作のみ)

生成手法	UFD(Acc)	AIDE (Acc)
Amazon Titan Image Generator	0.524	0.772
Microsoft Copilot	0.540	0.864
Stable Diffusion-base1.0	0.663	0.588
Ppaine Landscape	0.556	0.682

参考文献リスト

- [1] J. Wang et al., "Forensics-Bench: A Comprehensive Forgery Detection Benchmark Suite for Large Vision Language Models," in Proc. CVPR 2025.
- [2] K. Patrick et al., "KoDF: A Large-Scale Korean DeepFake Detection Dataset," in Proc. ICCV 2021.
- [3] S. Dong, et al., "Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization," in Proc. CVPR 2023.
- [4] S. Yan, et al., "A Sanity Check for AI-generated Image Detection," in Proc. ICLR 2025.
- [5] J. Ricker, et al., "AEROBLADE: Training-Free Detection of Latent Diffusion Images Using Autoencoder Reconstruction Error," in Proc. CVPR 2024.
- [6] U. Ojha, et al., "Towards Universal Fake Image Detectors that Generalize Across Generative Models," in Proc. CVPR 2023.
- [7] S. Wang, et al., "CNN-generated images are surprisingly easy to spot...for now," in Proc. CVPR 2020.
- [8] T. Hemlata et al., "Automatic speaker verification spoofing and deepfake detection using wav2vec 2.0 and data augmentation," in Proc. Odyssey 2022.
- [9] L. Xiaohong et al., "PSCC-Net: Progressive spatio-channel correlation network for image manipulation detection and localization," IEEE Transactions on Circuits and Systems for Video Technology, 2024.

検知器タイプ	エンジン
顔DF検知器	ForensicsBench (GPT-4.1 High)
音声DF検知器	SSL Anti-spoofing[8]
画像部分加工検知器	PSCC Net[9]
風景DF検知器	UFD + AIDE

3-2. 技術開発の個別詳細

機能2 文脈外利用(Out of Context) 判定技術

【概要】

- 昨年度版では検知できなかった文脈外利用※ (Out of Context: OoC) をシステムで検知できるように改善を実施
 - ※複数のモーダル間それぞれは真の情報であるが、文脈の異なる情報を組み合わせることで嘘や誤りとなるようなタイプの偽・誤情報

【詳細】

- OoCの検知に利用できる調査ツールを拡充・改善するとともに、調査結果をOoC判別に利用するようにシステムプロンプトを改善
 - 「時間」の矛盾：逆画像検索
 - 「場所」の矛盾：逆画像検索、画像位置推定技術
 - 「人物」の矛盾：Google Vision APIによるエンティティの特定
- 性能評価
 - 評価用データ：News CLIPpingsデータセット[1]（実際のニュースの見出しと画像を組み替えOoCを発生させたデータセット）
 - 評価方法：
 - 評価用データを偽・誤情報判別支援システムに入力し、OoCを理由に偽情報であると判定できた割合を評価
 - 比較として、昨年度版システムおよび今年度版システムに調査観点としてOoCを対象とする指示のあり/なしで結果を比較
 - 評価結果：
 - 昨年度版からOoCによる偽情報の検知精度を向上。観点の指示があれば高精度に検知できることを確認

方式	精度
昨年度版	0%
今年度版（指示なし）	20%
今年度版（指示あり）	80%

OoCデータ例

フランスリールでのサッカーファンによる警察との騒動の写真



検知結果（抜粋）

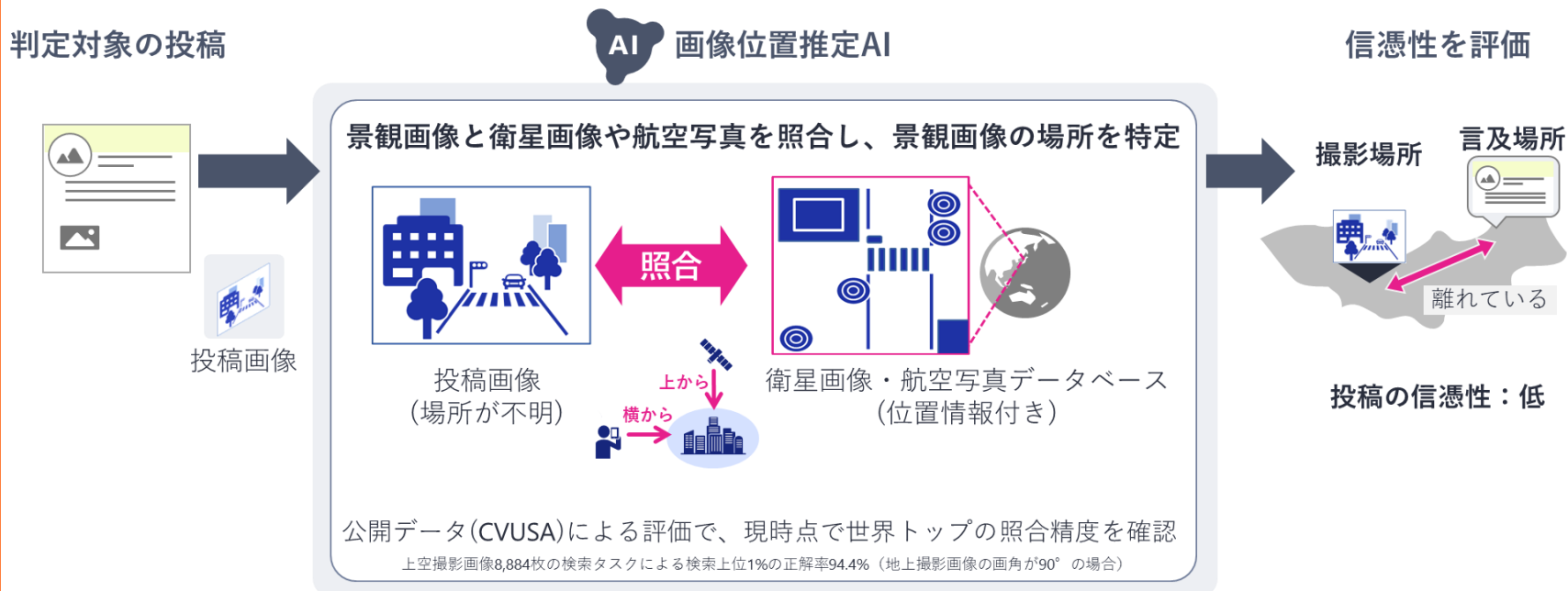
提供された画像が「フランス・リールでのイングランドサッカーファンと暴動警察の遭遇」を直接示しているとは断言できず、逆に**画像の最古使用例は2010年のロンドン学生デモの暴動**であることから、この画像はリール騒乱を示しているという説明は**誤用か誤解**

3-2. 技術開発の個別詳細

機能2 メディア分析技術（画像位置推定技術）

【概要】

- 真偽判別対象として入力されたSNSの投稿に対して添付画像から撮影場所を推定し、投稿文中で示唆されている場所とを比較しSNSの投稿の信憑性を評価する機能。
- 地上で撮影された投稿画像と上空から撮影された衛星画像や航空写真とを照合し、投稿画像の撮影場所を推定。
 - 投稿画像に映っている建物だけではなく、横断歩道や道路などのパターンなどを複合的にAIが解釈。



3-2. 技術開発の個別詳細

機能2 メディア分析技術（画像位置推定技術）

【詳細】

• 評価：

• ①画像位置推定技術の単体評価

- 評価方法：Xなどから、新宿区内を対象に撮影場所が既知の画像を収集して計115枚の評価用データセットを構築し、画像位置推定の精度を評価。
- 評価指標：テスト画像の撮影場所と推定結果との誤差距離が100m以内の場合正解とし、推定結果の上位N件以内に正解が含まれている割合（N位正解率）で評価。

• 評価結果：

	1位正解率	5位正解率	10位正解率	50位正解率
	35.7%	44.4%	45.1%	53.9%

• ②実運用を考慮した評価（投稿文と投稿画像の位置の不整合の検出）

- 評価方法：投稿文と投稿画像の位置に不整合がある投稿22件と投稿文と投稿画像の位置が同じ投稿22件の評価用データセットを構築し、画像位置推定を活用したSNSの投稿の信憑性を判定する機能を評価。
- 評価指標：投稿文と投稿画像の位置に不整合があるかどうか正しく判定できた割合（正解率）
- 評価結果(正解率)：63.6%

予測

投稿文と画像が同じ位置 投稿文と画像が異なる位置

正解

投稿文と画像が同じ位置(22)	6	16
投稿文と画像が異なる位置(22)	0	22

3-2. 技術開発の個別詳細

機能2 位置推定技術を用いた偽・誤情報判定 (1/2)

【概要】


- 画像位置推定技術の他、SNSアカウントの活動位置特定技術[1]を偽誤情報分析へ適用し効果的なユースケース検討と効果検証を行った。

【詳細】


- Web検索などでは証拠情報が収集できず判定できなかったケースに対して判定可能とするために、位置情報の推定を行いテキストに含まれる位置情報との整合を検証。
 - SNSアカウントの活動位置推定の適用例：入力「今まさに新宿で火災が発生している！」。
 - 活動位置推定なし：ニュース情報が無く根拠不明。
 - 活動位置推定あり+新宿近辺のアカウントの場合：ニュース情報は無いが現地人の可能性が高く速報情報の可能性がある。
 - 活動位置推定あり+新宿近辺でないアカウントの場合：ニュース情報は無く現地の情報である可能性も低い。
- 2つの位置推定技術を組み合わせることでより高品質な位置推定が可能になることを確認。

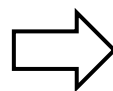
	近辺地域アカウント	無関係地域アカウント
関連地域画像	正確	誤り
無関係地域画像	誤り	誤り


SNSアカウントの活動位置推定技術で判定可能になる例



今まさに新宿で火災が発生している！







この画像は新宿のものですが、アカウントは新宿近辺で活動していません偽・誤情報の可能性があります

3-2. 技術開発の個別詳細

機能2 位置推定技術を用いた偽・誤情報判定 (2/2)

- 評価
 - SNSアカウントの活動位置特定技術を偽誤情報判別支援システムへ適用し、判定精度を評価。
- 評価データ
 - SNSアカウントのロケーション情報に対して近辺/無関係ロケーションの投稿文をLLMで生成。
 - 10アカウントに対し近辺/無関係ロケーションの投稿を生成し計20件の投稿を生成。
- 評価方法
 - 推定された位置と近辺のロケーション投稿を正確な情報、無関係のロケーション投稿を偽情報とし、偽誤情報判別支援システムによる判定結果と比較して精度を評価。
- 評価結果
 - 位置推定なしの場合は全て根拠不明となり判定に失敗。
 - 無関係なロケーション投稿に対しては位置推定を行うことで偽情報の検出ができることを確認
 - 位置推定ありでも近辺ロケーション投稿の判定は根拠不明が多く精度向上は不十分
 - 位置推定以外の確証が得られず根拠不明と判断されてしまった
 - レポート内では正確寄りの内容が記載されているため、証拠としての重要度を上げるなどにより精度が上がる可能性が高い。

評価結果

	近辺ロケーション投稿	無関係ロケーション投稿
位置推定なし	0.0	0.0
位置推定あり	0.2	1.0

3-2. 技術開発の個別詳細

機能3 証拠収集能力を改善するDeepResearch技術

【概要】

- 偽・誤情報の検証に用いる証拠情報としてWeb上からの証拠収集において、従来より1次情報を収集可能な検索機能を開発した。

【背景】

- 昨年度の有識者ヒアリングでWeb上からの証拠収集は1次情報（初出の情報）を発見できることが調査工数の削減につながるとの意見をいただいた。

【詳細】

- 証拠検索機能の強化のためにAIエージェント(※1)を用いたDeepResearch(※2)を導入するとともに、1次情報に到達するための改善を実施。
 - 検索された情報に対して、引用文の有無を判別して1次情報かどうかを判断し、1次情報の量が不十分な場合はクエリ生成～検索を再度実行。
- 評価結果
 - 評価用データ：発見してほしい情報源（ウェブページから引用されている1次情報）と検証対象を対応付けた独自データ37件。
 - 評価方法：証拠検索により発見してほしい情報源を発見できた割合を評価。
 - 評価結果：改善方式により昨年度と比較して30%程度の精度向上を確認。



証拠検索システム	証拠内に情報源を含む割合
昨年度開発版	0/37
OpenDeepResearch版	3/37
OpenDeepResearch改善版	10/37

- ※1：AIが目標を理解し、自律的に計画・実行するソフトウェア
- ※2：AIが人間の指示に従い、Web上の情報などから調査対象に関する情報を自律的に収集・分析するソフトウェア

3-2. 技術開発の個別詳細

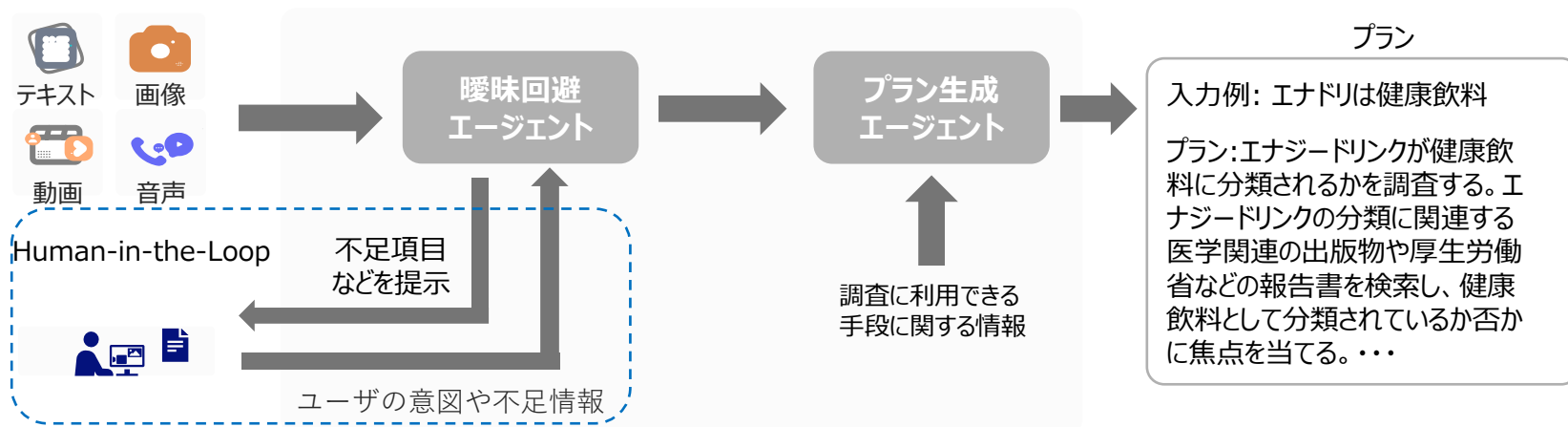
機能4 AIエージェント技術 (1/3)

【概要】

- 入力される情報コンテンツに対して適切な偽・誤情報判別プロセスを実行するAIエージェントを開発

【詳細】

- 真偽判別エージェントは複数のAIエージェントを組み合わせることで、偽・誤情報判別を行うための①プランニング、②調査を行い、最終的な調査レポートを出力する。
 - ①プランニング機能は曖昧回避エージェントとプラン生成エージェントにより構成される。
 - 曖昧回避エージェント: ユーザの入力から不足情報や曖昧な表現・検証対象をHuman-in-the-Loopで補完。
 - プラン生成エージェント: 検証対象に対して利用できる調査手段を考慮して調査方針を決定。
 - ①プランニング機能によるプランニングの妥当性評価。
 - 評価方法: 人工データ (42件) で昨年度版とレポートの判定ロジックやプランに対するの妥当性を主観評価。
 - 評価結果: 42件中41件で昨年度版よりも妥当性が高いと評価 (97.6%)。

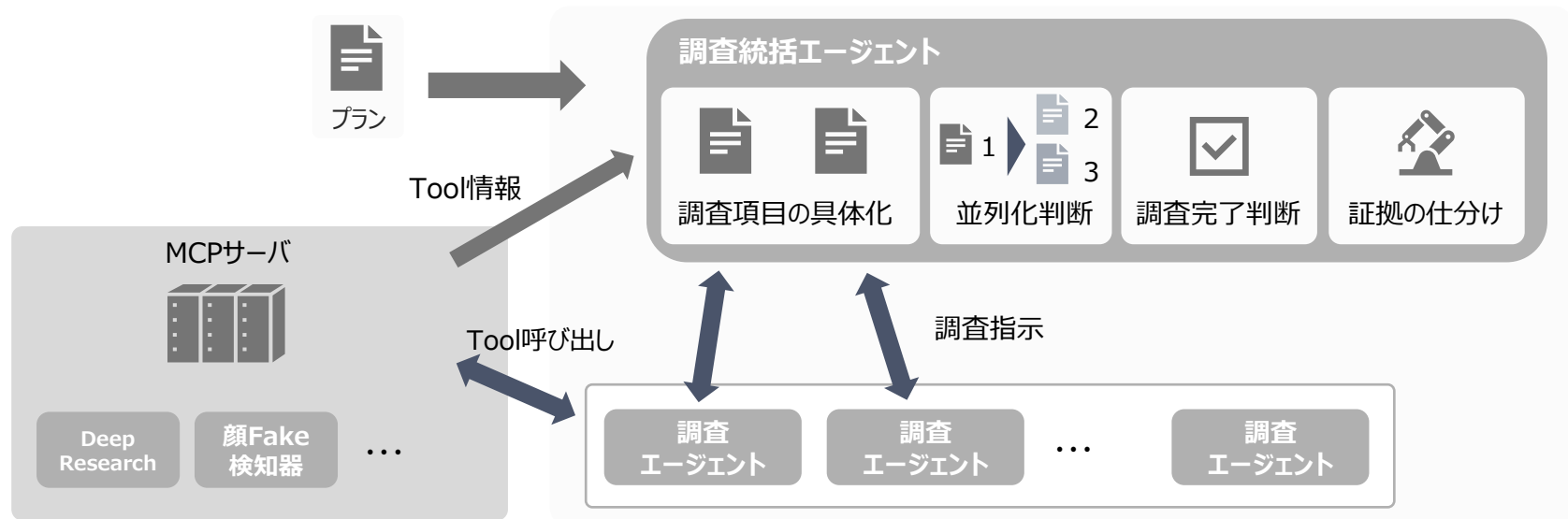


3-2. 技術開発の個別詳細

機能4 AIエージェント技術 (2/3)

- ②調査機能は、調査統括エージェントおよび調査エージェントにより構成される。
 - 調査統括エージェント：調査プランを受け取り、複数の調査エージェントへ調査指示と調査状況の管理を行う。
 - 調査エージェント：各種調査ツールを管理し、調査指示に従って調査を実行。
- ディープフェイク検知器やDeepResearch機能といった各種調査ツールはMCPサーバとして構成され、MCPサーバ（※1）を追加することで新たな調査ツールを容易に追加し、適切に偽・誤情報判別プロセスの中に組み込み可能である。

※1：Model Context Protocolサーバの略称。
AIが外部のツールやデータにアクセスするための共通的な規約MCPに準拠し各ツールをAIから柔軟に呼び出し可能としたサーバを指す



3-2. 技術開発の個別詳細

機能4 AIエージェント技術 (3/3)

- AIエージェントによる検証はプランニングエージェントにより、ユーザとの対話を通してプランニングを行い、プランニングの結果やエージェントにより実行された調査ツールによる調査結果、および総合的な判断結果が最終出力される。

プランニングエージェントによる Human-in-the-Loopを用いたプランニング

ファクトチェックする対象をどのように指定しますか？

自由入力

ファクトチェック対象を入力してください。

動画

ファイルが正常にアップロードされました

他に入力がありますか？

いいえ

以下ニュースを取得しました。

NEWS

LIVE

00:00 / 0:15

のスーパーからパンが消失！?

agent

映像の真偽と内容の信頼性を検証してほしい

検証対象の動画・画像・音声ファイルについて、具体的にどのような事実を確認希望でしょうか？例えば、映像の真偽、登場する人物の発言の正確性、事件の背景情報など、調査内容を具体的に教えてください。

検証対象
コンテンツ

ユーザと対話を行い曖昧性を排除してプランニング

システムにより生成されたレポート

ニュース動画の信頼性と真偽検証報告

調査プラン

はじめに

2026年1月9日時点で、「全国のスーパーからパンが消失した」という内容の日本語ニュース動画が話題となっている。本報告では、この動画の真偽と信頼性を検証するため、メディアのフォレンジック分析（画像検出・音声のディープフェイク検出）、画像の改変有無、動画の音声調査、及び日本の公式・信頼できるニュースソースやスーパーからの情報を基に検証を行った。

1. メディアの本物性と改変検証

1.1 顔画像ディープフェイク検出

新キャスターの顔の連続フレーム中、5フレームについて検証を実施。そのうち5フレームは非常に高確率（62.9%〜99.9%）でフェイク（AI生成や合成）と判定された。[証拠1-5]

1.2 画像・動画フレームの部分加工検出

顔の表情・姿勢等の詳細判定ではなく、フレーム全体の画像レベルでは不自然な加工や改ざんは低いと判定された。得点は低く（90.3%偽造率）、ピクセルレベルでの不正加工は検出されなかったため動画の一部は合成の可能性が高い。[証拠1-17]

1.3 音声のディープフェイク検出

動画内の音声は非常に高い確率（約99.97%）で人工生成や改変が疑われるフェイク音声であると判定された。[証拠1]

1.4 メディアの総合評価

映像の顔部分と音声は高確率でフェイクである一方、画像そのものは大規模な改ざんの形跡が確認されなかった。つまり、このニュース動画は合成・加工によって作られた偽造映像であると強く推定される。

2. 全国のパン消失の事実検証

2.1 公式及び信頼できる情報源の状況

当該動画に関連する「全国のスーパーからパンが消失した」という報告は、国内の大手新聞社、公共放送、スーパー各チェーンの公式発表等で確認できなかった。

2026年1月現在、食料品の価格については安定的であり、特にパンの品揃や全国的な品切れに関する公式な警告・報告は一切なされていない。むしろ一部店舗の一時的な品揃や棚が空いているように見える画像・動画はSNS上の風評的記事である可能性が高い。

2.2 SNS投稿と動画内の引用

動画内でのSNSユーザー「太郎 (@taro)」の投稿としてパンの籠が空になっている画像が示されているが、この画像の出典に関する明確な出所や検証はできず、またこの画像のみでは全国的な品切れを示す証拠とは認められない。

3. 総合判断

- 当該ニュース動画自体は顔と音声が高確率にて人工生成・加工された偽造動画である。
- 画像の一部は加工したが、合成や編集で構成された映像と判断される。
- 「全国のスーパーからパンが消失した」という主張は、信頼できる公的・大手報道機関や公式販売店からの情報と照合せず、証拠も乏しく誤情報・フェイクニュースの類である可能性が高い。

4. 結論

本ニュース動画は編集・人工生成された偽造動画（ディープフェイク）であり、「パンが全国のスーパーから消失した」という報道内容は偽偽である。

参考文献・情報源

(1) 顔画像フェイク検知結果
(2) 部分加工フェイク検知結果
(3) 音声フェイク検知結果
(各検証は提供された診断ツールより取得)

最終判定: False (偽偽)

全証拠リストを表示

プランに従って実行された調査ツールごとの結果

総合判断結果

3-2. 技術開発の個別詳細

機能5 SNSからの分析対象候補（偽・誤疑い情報）の検知

【概要】

- 偽・誤情報の拡散傾向に類似している投稿や偽・誤情報を含む投稿で頻繁に使われる言い回しなどを基に投稿を選別する機能。



3-2. 技術開発の個別詳細

機能5 SNSからの分析対象候補（偽・誤疑い情報）の検知

【詳細】

- 下記の3つの技術を活用した投稿フィルタリングAIにより偽・誤情報疑いの投稿を選別。
 - 一定間隔毎にその時々々のトピックを抽出、過去に抽出したトピックとを比較して新規情報を含む投稿を検知。
 - 偽・誤情報を含む投稿で頻繁に使われる言い回しやフレーズなどを含む投稿を検知（小規模言語モデル）。
 - リツイート数や引用数、表示数などの単位時間当たりの増加数に基づき極端に早く拡散された投稿を検知。
- 評価：
 - 評価方法：自民党総裁選時に収集したXの投稿250万件を用いて投稿フィルタリングAIの性能を評価。
 - 評価指標：
 - 削減率：投稿フィルタリングAIにより真偽判定支援システムに回す投稿件数をどれだけ減らせたか。
 - 再現率(Recall)：偽・誤情報疑いのラベルを付与した375件(※)をどれだけ真偽判定支援システムに回せたか。
 ※国際ファクトチェックネットワーク（International Fact-Checking Network、IFCN）が定める5原則に基づき、人手で偽・誤情報疑いのラベル付けを行った。偽・誤情報を含む投稿では訂正などがコメントで行われることが多いため、収集期間内にコメントが付いた投稿27,000件を対象とし、その中から偽・誤情報で使われがちなキーワードを含む1,200件を抽出してラベル付けを実施した。
- 評価結果：

拡散速度のパラメータ	削減率	再現率
100人/日で拡散されうる	99.3% (1.78万/248万)	40.8% (153/375)
1,000人/日で拡散されうる	99.6% (9,890/248万)	28.5% (107/375)

3-2. 技術開発の個別詳細

機能5 SNSからの分析対象候補（偽・誤疑い情報）の検知のUI開発と偽誤情報分析への連携

【概要】

- 偽・誤情報疑いの検知技術を実行し偽・誤情報判別支援システムと連携したUIを開発

【詳細】

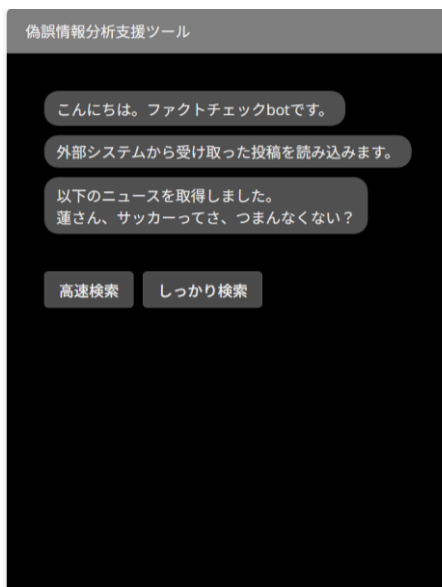
- ユーザ任意のトピックで偽・誤情報疑いの検知技術を実行し、選別された投稿の内ユーザ任意の投稿に対して偽・誤情報判別支援システムを実行できるシステム
 - 新規トピック作成ボタン: 任意のトピックを入力し投稿の抽出を開始しトピック一覧に追加
 - トピック一覧: 抽出中のトピックを表示し複数トピックを並列で実行している場合は切替可能
 - 投稿表示画面: 投稿フィルタリングAIにより選別された投稿を順次表示
 - 各投稿はクリックすることで投稿内容を入力した状態で偽・誤情報判別支援システムを開く

新規トピック
作成ボタン

クリックにより選択した投稿を自動で入力した
状態で偽・誤情報判別支援システムを開く

投稿表示画面

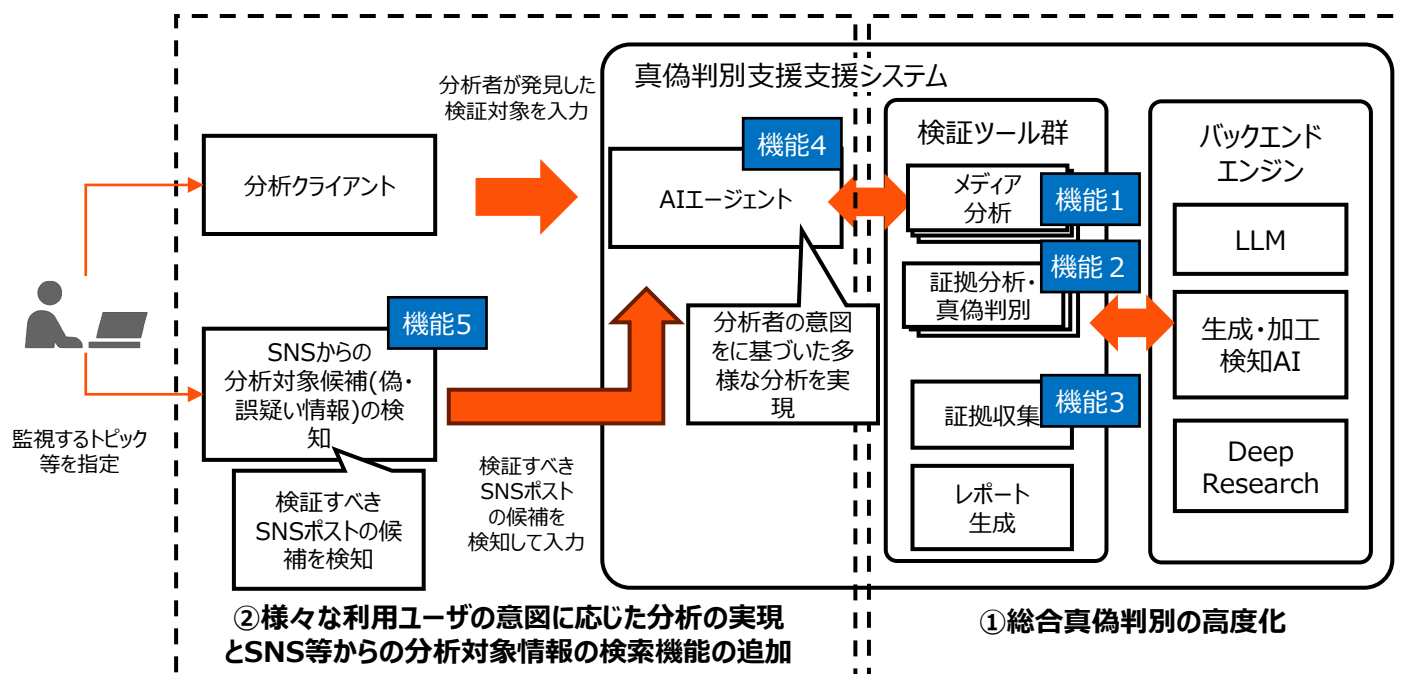
トピック
一覧



3-2. 技術開発の個別詳細

技術開発に係る取組・成果一覧

- 機能1：メディア分析機能として動画の内容をテキスト化するエンジンを改善しQAタスクの精度で8-9%向上
- 機能2：
 - 顔および風景画像に対するディープフェイク検知技術を最新技術にアップデートし精度向上
 - 文脈外利用の判定を可能とし、明示的な指示を与えることで検知精度を80%向上
 - 画像位置推定技術を組み込み、投稿文にある位置と画像の位置が異なる偽情報を100%検知可能であることを確認
- 機能3：Web上からの証拠検索機能としてDeepResearch技術を開発し、証拠発見精度を30%向上
- 機能4：入力情報コンテンツやユーザの指示に従ってプランニング～偽情報分析を行うAIエージェントを開発
- 機能5：SNSから偽・誤疑い情報を拡散傾向や言い回しなどをもとに選別する機能を開発し、SNS投稿全体の99%以上の投稿を削減しながら、含まれる偽・誤情報疑いのうち40%程度を検出可能であることを確認



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

- 検証及び調査に係る取組・成果の全体像は下記のとおり

検証及び調査に係る取組・成果の全体像		
実証の想定	共通	<ul style="list-style-type: none"> ユースケースに沿って開発したシステムの<u>実証・ヒアリングの実施</u>することで、<u>更なるニーズ・課題の明確化とユースケースの深堀</u>を実施。
	マスメディア	<ul style="list-style-type: none"> <u>報道などの速報性への対応</u>など実業務に沿った検証・ヒアリングを行い、<u>実用性を評価</u>。
	自治体	<ul style="list-style-type: none"> 特に災害時の対応にて、<u>偽・誤情報疑い情報の認識から自治体公式の情報発信</u>までの一連の業務において<u>有効性があるか検証・ヒアリング</u>。
実証概要		ステップ1としてマスメディアと自治体へヒアリングを実施。 ステップ2としてマスメディアと自治体で実証を実施。
実証成果	マスメディア	ステップ1：63社での放送局にヒアリングを実施。 ステップ2：6つの放送局で実証を実施。
	自治体	ステップ1：2つの自治体でのヒアリングを実施。 ステップ2：1つの自治体での実証を実施。

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ1】

- ヒアリングした一部の声は以下の通り

対象	ヒアリングからの声
放送局	<ul style="list-style-type: none"> 製品化されたら利用したい。 報道での用途がありそう。 視聴者投稿をチェックするなどの人員削減につながりそう。 災害時に画像の場所検証ができるのはありがたい。 自社内で保有している機密情報（外部環境に情報を持ち出し不可）も利用したい。 参考文献が出てくるのはありがたい。最終判断は人で行いたい。 速報性高い情報の評価を行いたい。
自治体	<ul style="list-style-type: none"> 偽誤情報であるという報告を上げるために、根拠を色々なソースから検索して集めなくてはいいが、レポートでその根拠一覧をまとめてくれるのはとてもありがたい。 画像から位置情報推定は防災に使えると思う。特に自治体と使っているものと組み合わせられるといい。 フィルタリングAIによる偽誤情報疑いの提供はありがたい。今どんな情報が盛り上がっているかを検知するのが難しいと感じている。 防災など特有の用語などが多く存在するため、専門用語も理解たうえで偽誤情報判別をしてほしい。

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 放送局】

- 放送局における検証結果の個別詳細は以下の通り

放送局における検証結果

実証期間 2月上旬～中旬に2週間

<p>偽情報判断の 精度・妥当性</p> <p>実証後ヒアリングで 得られた意見</p>	<ul style="list-style-type: none"> テキストのみの場合、流通情報があれば偽情報検知の精度は高い。 ファイル添付時の映像や音声解析、有名なフェイク画像や動画でReal判定となる場合があり更なる精度向上が必要。 ツール上の「一次情報」の定義を検討しなおしたほうが良いと感じるケースがあった。 適切なソースに基づく根拠は明確で満足できる、画像フェイク検知は複数ツールで確率が示されるため納得感が高い。
<p>期待する機能の 改善と過不足</p>	<ul style="list-style-type: none"> 不要な機能は特になし。 ソース入力の簡素化を希望。 複数人で作業する前提の機能（履歴保存、参照等）があると嬉しい。

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 放送局】

放送局における検証結果

ユーザーインター
フェイス（UI・
UX）の操作性

実証後ヒアリングで
得られた意見

- 画像や動画の入力はやや手間がかかるため、API連携などで簡単にチェックできると良い。
- ファイル選択やテキスト入力の反映がわかりづらい箇所があり、画面の使い方にも改善余地がある。
- 情報源リストや詳細表示の視認性向上が必要だが、慣れれば使いやすい。
- 応答スピードは現状長く、テキストは30秒以内、画像・動画は1～2分程度が望ましい。
- 一方、コピーや入力は容易で、画面遷移やボタン配置も直感的で迷わないという意見もあり、5分前後なら許容できるとの声もあり。

偽情報判断業
務の効率化への
貢献度

- 既出情報と新規情報の確からしさを区別して簡単に示せるなら、業務効率化につながる可能性がある。
- 判定精度よりも、根拠が扱いやすいことが重要。
- 情報収集の負荷軽減になる。

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 放送局】

実際検証いただいた内容①：

ファクトチェック対象	詳細
画像	<p>画像の真偽、メッセージの内容、詐欺広告かどうか等を確認してほしい</p>

応答結果	詳細
応答時間	380秒
応答結果	適切な結果が表示された。
安定性	フリーズやエラー等は発生せず。
総合評価	応答時間が非常にかかった

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 放送局】

実際検証いただいた内容②：

ファクトチェック対象	詳細
テキスト	<p>総務省は6日、2025年の家計調査を発表しました。1世帯あたりの年間支出額でみるとギョーザは浜松市がラーメンは山形市が全国で1位となりました。</p> <p>この調査は、都道府県庁所在地と政令指定都市を対象に品目ごとに1世帯あたりの年間支出額を調べたものです。それによりますと、ギョーザは1位が3年連続で浜松市、2位宇都宮市、3位宮崎市でした。また、外食としてのラーメンは1位が4年連続で山形市、2位新潟市、3位は宇都宮市となりました。</p> <p>また2025年に1世帯が食料、住居、教育、娯楽など消費全体に支出した金額は1か月平均で31万4001円でした。前の年と比べると物価の変動を除く実質で0.9パーセント増え、3年ぶりの増加となっています。</p>
応答結果	詳細
応答時間	177秒
応答結果	結果の内容については違和感なし
安定性	フリーズやエラー等は発生せず。
総合評価	入力しているテキストの分量に対し、感覚的には待機時間が長く感じた

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 放送局】

実際検証いただいた内容の操作画面②：

The screenshot displays a web application interface with a dark theme. At the top, there is a navigation bar with the text '偽装情報分析支援ツール' on the left, '処理をキャンセルする' in the center, and '設定' on the right. Below the navigation bar, a grey banner contains the text '除く実質で0.9パーセント増え、3年ぶりの増加となっています。'.

The main content area is divided into two columns. The left column is titled '0 結論と要約' and contains the following text:

本件の主張は事実と整合しており、「真」であると判断される。

総務省が2026年2月6日に公表した2025年（暦年）の家計調査結果について、以下の点が公式発表および大手報道によって確認できる。

- 発表日時と調査枠組み
 - 総務省統計局は2026年2月6日に2025年1年間の家計調査結果を公表している。この調査は、都道府県庁所在地と政令指定都市を対象に、品目別に1世帯当たり年間支出額を算出する形式で行われている[1][2][5]。
- ギョーザ支出ランキング
 - 2025年の「ギョーザ」品目（主にスーパー等での購入額）について、1世帯当たり年間支出額の全国1位は浜松市であり、これは3年連続の1位である。
 - 2位は宇都宮市、3位は宮崎市と報じられている[5][6]。
- ラーメン（外食）支出ランキング
 - 「外食としてのラーメン」への年間支出額では、山形市が4年連続で全国1位となっている[5][6]。
 - 2位は新潟市、3位は宇都宮市と報じられており、山形市の年間支出額は25,102円で、2位以下に6000円以上の差を付けている[5]。
- 世帯あたり平均消費支出と増減率
 - 2025年における、2人以上の世帯が食料・住居・教育・娯楽等を含む消費全体に支出した金額は、1か月平均31万4001円。
 - 物価変動を除いた実質ベースで、前年より0.9%増となり、3年ぶりの実質増加であることが示されている[5][6]。

The right column is titled '参考文献一覧' and contains two entries under the heading 'Web Search Tool':

Web Search Tool

サマリ: このウェブページは、総務省統計局が提供する家計調査に関する情報を掲載しています。家計調査の概要、調査方法、調査結果の公開状況、関連FAQなどが利用者に向けて説明されています。また、家計調査のデータは、有料データベースのe-Statを通じて提供されており、最新データの公開スケジュールや過去の調査内容の変更点も案内されています。利用者は調査データを活用して経済動向の分析や政策立案などに役立てることができます。

URL: <https://www.stat.go.jp/data/kakei/2.html>

引用: このウェブページは、総務省の家計調査に関する情報を掲載しています。有料データベースのe-Statを通じて提供される家計調査の統計データ、調査概要、調査実施に関する変更点などが記載されています。また、FAQも用意されており、利用者の疑問に対応しています。

立場: 肯定
情報源レベル: 1次
検索クエリ:

Web Search Tool

サマリ: このウェブページは、総務省統計局が提供する家庭調査に関する速報情報を掲載しています。2025年7月から10月までの調査結果が公表されており、家計の動向を把握するためのデータが提供されています。詳細な統計データは、総務省の公式統計サイトやe-Stat（政府統計の総合窓口）から閲覧可能です。また、問い合わせ先の連絡情報も記載されており、調査に関する質問や不明点は総務省統計局に直接問い合わせることができます。サイト全体は日本語ですが、多言語でのICT政策情報の案内も一部含まれています。

URL: https://www.soumu.go.jp/mcmu/news/s-news/21houke07_01000218.html

引用: <https://www.stat.go.jp/data/kakei/sokuhou/tsuki/index.html>
<https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00200561&tstat=000000330001>

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 自治体】

- 自治体における検証結果の個別詳細は以下の通り

自治体における検証結果

実証期間	12月上旬に1週間	
	偽情報判断の精度・妥当性	<ul style="list-style-type: none"> ツールの偽情報判断結果は、<u>専門用語など事柄の判別が不十分な場合があります。</u> <u>正しく判定された例では、根拠が明確でおおむね満足。ただし動画検索では、根拠となる情報を十分に活用していない事例があった。</u>
実証後ヒアリングで得られた意見	期待する機能の改善と過不足	<ul style="list-style-type: none"> <u>精度向上が最も重要。</u>操作インターフェースはGoogle検索やLINEのような一般的なアプリに近づけてほしい。 不要な機能は特にないが、調査方針や英文メッセージの表示は一般向けには非表示が望ましい。 <u>解析の所要時間（予想）の表示機能を追加してほしい。</u>

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 自治体】

自治体における検証結果


	ユーザーインターフェース (UI・UX) の操作性	<ul style="list-style-type: none"> • 情報入力は悪くないが、入力欄やボタンはもう少し大きい方がよい。 • 画面遷移は概ね直感的だが、<u>入力完了時に「OKボタン」があると分かりやすい。</u> • <u>結果表示画面は章立てがしっかりしていて読みやすいが、最下部の表示は検証ごとにテキストが揺れている。</u> • <u>応答スピードは早いほど良い。理想は30秒以内、180秒なら許容。</u>
実証後ヒアリングで得られた意見	偽情報判断業務の効率化への貢献度	<ul style="list-style-type: none"> • <u>検証に30分かかっていた作業が、ツール利用で5分程度に短縮できると推察される。</u>
	その他	<ul style="list-style-type: none"> • <u>導入検討についてはツールに興味はあり。「正答率〇〇%」といった形で、スペックが明らかになった上で利用・導入を検討したい</u>

4-2. 検証及び調査の個別詳細

検証及び調査の個別詳細

【ステップ2 自治体】

実際の操作画面の内容：

ファクトチェック対象	詳細
画像	<p>「女川のクマと3体」 (女川町役場が公開後フェイクと判明し公開中止された画像を加工したもの) を入力。</p> 
テキスト	<p>「アップロードした画像は本物ですか？フェイクですか？」など、画像の真偽と公開経緯について質問。期待値は「正しくない」。</p>

応答結果	詳細
応答時間	ツールからの質問まで約60秒、回答から結果表示まで約185秒。
応答結果	生成AIによる偽造画像で誤情報と判明した旨が、自然な日本語で分かりやすく表示され、期待通りの判定だった。
安定性	フリーズやエラー等は発生せず。
総合評価	スムーズに利用できた。

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

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

【本実証KPI】

- ビジネス視点に関わる社会実装に向けた取組の個別詳細

本実証KPIと結果

ビジネス視点	KPI	<ul style="list-style-type: none"> ビジネス視点として、本実証により社会実装の利用主体の拡大について定量的に評価するKPIを設定した。具体的には20社以上のマスメディア（放送局・通信社・新聞社）と2つ以上の自治体でのユースケースに沿って開発したシステムの実証・ヒアリングを行うことを目指す。また、展示会への出展を2件以上(例：マスメディア向けInterBEE、他業界向け展示会等)、学会での論文採択・発表1件以上(例：AI系学会のAAAI等)を行い、社会実装の拡大に向けたユーザ探索を積極的に進めることを目指すこともKPIとして設定していた。
	結果	<ul style="list-style-type: none"> 上記において63社のマスメディア（放送局・新聞社）と2つの自治体へシステムの実証・ヒアリングを実施。<u>展示会への出展を3件、学会での論文採択を1件実施したためKPIは達成となった。</u>
ユーザ視点	KPI	<ul style="list-style-type: none"> ユーザ視点として、幅広い分析ニーズに対応できたかを定量的に評価するKPIを設定する。具体的には複数の現実的な分析パターンを設定し、分析パターンに沿った真偽判別が3つ以上実施できることを目指すことをKPIとしていた。
	結果	<ul style="list-style-type: none"> <u>このKPIに対し3つの真偽判定パターンを実施しKPI達成となった。</u> ①AIによる生成・加工されたメディアによる偽情報の判定 ②画像・映像の文脈違いによる偽情報の判定 ③事実無根の情報による偽情報の判定

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

【本実証KPI】

- ユーザー視点に関わる社会実装に向けた取組の個別詳細

ユーザー視点

- | | |
|-----|--|
| KPI | <ul style="list-style-type: none"> ユーザー視点として、幅広い分析ニーズに対応できたかを定量的に評価するKPIを設定する。具体的には複数の現実的な分析パターンを設定し、分析パターンに沿った真偽判別が3つ以上実施できることを目指すことをKPIとしていた。 |
|-----|--|

- | | |
|------|--|
| 結果詳細 | <ul style="list-style-type: none"> これに対して以下の<u>3つの真偽判定パターンに対する検証を実施しKPIを達成することができた。</u> |
|------|--|

真偽判定パターン	判定方法例
AIによる生成・加工されたメディアによる偽情報の判定	生成・加工検知器AIを用いて判定
画像・映像の文脈違いによる偽情報の判定	画像検索や位置推定による画像の文脈の特定を行い、文脈違いを判定
事実無根の情報による偽情報の判定	Web上からの情報の検索・収集を行い、偽情報の主張との比較検証をすることで判定

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

【さまざまな業界へのヒアリングの実施】

- 自治体・放送局だけでなく消防庁など新しい業界へのヒアリングを行いニーズを発見した。
- 具体的にヒアリングを実施したのはマスメディア（放送局・通信社・新聞社）、自治体、消防、警察である。

業界	詳細内容
放送局	特に画像・動画について精度が求められ、この精度向上が十分に取れば実装は考えられるのではとの声をいただいた。また、自社内APIとの連携を行いたいとの声もあった。
自治体	機能として偽誤情報疑いの検出は必須であるという声が複数散見された。また、 <u>画像位置推定</u> については災害や事故で利用したいという声もいただいた。
消防	事故の初期で、 <u>SNSであげられた情報を確認するためその信頼性を担保したい</u> との課題が発見された。
警察	本システムについては、 <u>現場に行く前に心構えができる</u> という点で前向きに捉えられる場合もあったが、一方で「 <u>現場対応が必須である以上、ツールの必要性は限定的ではないか</u> 」といった意見も見受けられた。

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

【市場拡大における社内連携の開始】

- 更なる市場拡大を考え、自社内でも他組織と連携しながら自治体や消防などにアプローチ。
- また、各市場での既存システムに本ツールを組み込む可能性も検討しており、幅広い業界での活用が期待される。
- 他社との連携も視野に検討を実施。

【製品化社内体制の整備】

- 製品化に向けた社内体制を整備中。
- NECの価値創造モデル「BluStellar(ブルーステラ)※」に製品登録化を対応中。登録を行うことで、NECがすでにパイプを持つ幅広い業界への提案が可能となる。

※ BluStellar(ブルーステラ)とは

- お客さまを未来へ導く、価値創造モデル。NECの先進的なテクノロジーとビジネス変革の知見・経験を体系化した“BluStellar”で社会課題・経営課題を解決へ導きます。

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

- 普及啓発活動に係る取組・成果の全体像は以下の通り

展示会名	普及啓発活動の概要	
某県警展示会	目的	警察マーケットのニーズ確認
	実施内容	警察マーケットにおける「偽誤情報対策」のニーズ調査を目的に出展。法執行機関の全業務領域での活用イメージを提示し、現状の課題感と導入に向けた具体的なユースケースの抽出を実施。
危機管理産業展 (RISCON TOKYO2025)	目的	自治体・自衛隊の防災ニーズ確認
	実施内容	偽・誤情報判別支援ユースケースをNECの災害対応ソリューションとともに展示。
InterBEE	目的	放送局のニーズ深掘り（国内最大の放送関連展示会）
	実施内容	総務省実証の昨年度→今年度の機能差分を説明。昨年度の課題に対し今年度追加した機能の妥当性をヒアリング。自治体ニーズに合わせて新規追加した機能への意見も収集。
言語処理学会の第三十二回年次大会	実施内容	2026年3月9日～13日に開催予定の言語処理学会第三十二回年次大会に本事業で開発した偽誤情報判定向けのエージェントAI技術について技術および評価結果の論文の投稿を完了し、発表を予定。

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

- 普及啓発活動の個別詳細は下記のとおり

展示会名	普及啓発活動の個別詳細
某県警展示会	<ul style="list-style-type: none"> 警察マーケットへのニーズ確認のため出展。 警察業務に特化したソリューションを展示。偽誤情報が法執行業務（捜査・警備・広報等）の全域に及ぼす影響を定義し、具体的な活用シナリオを提示しながらヒアリングを実施。 深掘りした課題（事例）として以下を挙げ、現場での実効性を確認。 <ul style="list-style-type: none"> SNSの虚偽投稿による公衆への不安煽り（詐欺・扇動）への対応。 悪戯の110番通報等によるリソースの不当な割き出し。 現時点では限定的な需要であることが分かった。しかしヒアリングを通じて特定の事例に対する具体的なニーズを掘り起こすことができたため、今後の重要性の高まりを見据え、引き続き市場動向をウォッチ。
危機管理産業展 (RISCON TOKYO2025)	<ul style="list-style-type: none"> 災害への対策（防災）として自治体や自衛隊などのニーズ確認のため出展。 NECが持つ災害対応に関するソリューションを複数展示する中で、偽・誤情報判別支援ソリューションを活用したユースケースを展示。 SNS分析サービス各社の特徴を活かし、デマ可能性の高い情報については、<u>デマ可能性情報を「ファクトチェック」カテゴリとして一元的に可視化し、デマ打消し等の業務に寄与できるのではないかという仮説の元ヒアリングとニーズの深堀を実施。</u> 消防についての需要が見られそうな為<u>26年度市場拡大を見据え個別ヒアリングなどを実施予定。</u>

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

展示会名	普及啓発活動の個別詳細
InterBEE	<ul style="list-style-type: none"> • InterBEEが放送関係では国内最大の展示会である為、主に放送局関係のニーズの深堀の為に出席。 • 昨年度の総務省実証から今年度の総務省実証での機能差分の説明と昨年度の実証をふまえた課題を今年度の機能で追加で入れ込んだことに対するヒアリングを実施。 • <u>また自治体でニーズがあり新しく追加した機能に対してのご意見等を頂いた。</u> • <u>偽誤情報に関するセミナー「生成AI時代の偽情報リスクと対策」を実施。内容としては生成AIの進化は、偽情報の拡散を加速させ、社会に深刻な影響を与えている現状から、AIが生成する偽情報の現状とリスクと対策のオーバービューと併せてNECの取り組みについてを講演。</u> • <u>3月中旬に実施されるNEC独自企画 AfterInterBEE Zoomウェビナーでも講演予定。放送局のみならず、新聞社へも案内実施。</u>
言語処理学会の第三十二回年次大会	<ul style="list-style-type: none"> • <u>2026年3月9日～13日に開催予定の言語処理学会第三十二回年次大会に論文の投稿を完了し、発表を予定している。</u> • <u>論文の内容としては本事業で開発した偽誤情報判定向けのエージェントAI技術について技術および評価結果をまとめ投稿した。会期中には発表を通して有識者との社会実装に向けた議論を実施予定。</u>

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

【課題】

2章で述べた開発項目を組み込み、KPIとしていた誤判別率の向上は達成できたものの、ユーザの試用により以下のような課題が確認された。

- 入力として専門用語が与えられた際、正しく対応できないことがある
- レスポンスタイムが想定よりも長い
- ユーザが望むユースケースである「生成AIによって加工・作成されたフェイク画像の検知」については正確な判別をおこなうことが難しい場合がある

【今後の展望】

上記の課題を受け、今後のソリューションの展望としては以下の事項が考えられる。

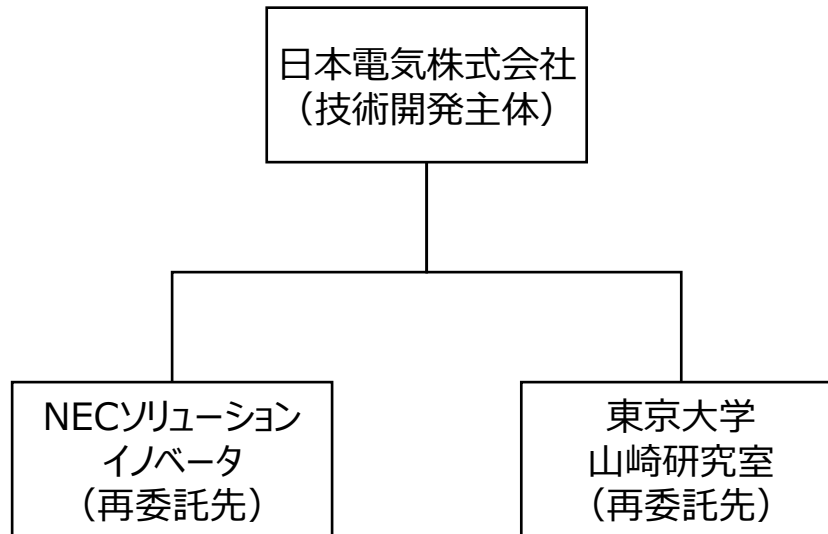
- ユーザの利用頻度が高いと想定される領域を強化したソリューション開発（専門用語集対応含む）
- ジョブの並列化を可能にし、ジョブの結果通知前に次のジョブを実行させることを可能とする。
（結果を待つ間に次のジョブを投入開始できる）
- 生成AIによる画像フェイク検知の精度向上

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

- 日本電気株式会社
 - 実証事業代表者
 - 実証事業全体まとめ
 - 技術開発主体
 - 事業計画立案
 - 報告業務をはじめとする事業全般の管理・統括業務
- NECソリューションイノベータ
 - ネットワーク・実証環境構築
 - 判別支援アルゴリズム実装担当
- 東京大学山崎研究室
 - 生成加工検知AIの開発技術アドバイザー

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
マイルストーン			▼RISCON	▼InterBEE			学会発表 ▼	
技術開発	方式検討	設計 → 試作 ↑ 評価 ↓		各種機能結合		総合評価		
実証、普及啓蒙活動	実証内容検討	実証内容ブラッシュアップ		(自治体) 実証		(メディア) 実証		
		各種展示会対応						