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

多元統合型偽・誤情報検出技術の開発・実証

成果報告書

2026/3/19

技12_サン電子株式会社

目次

目次

1-1. 開発・実証のサマリ

アプローチする課題・目指す姿	• 近年、偽・誤情報リスクが急拡大する中、従来の単一技術では誤検知や精度低下といった限界が生じている。そのため、複数の検知技術を統合し、総合的な信頼性判断を可能にする新たなアプローチが不可欠となっている。 • 複数の改ざん・偽造検知技術を統合したシステムを構築し、統合検知結果を活用した信頼性判断支援技術を確立する。さらに、これらを用いた信頼性判断支援サービスの社会実装に向けた基盤を構築することを目指す。		
技術区分	コンテンツの真正性保証・信頼性判断支援技術	実施体制 （下線：技術開発主体）	サン電子株式会社、サイバーコマンド株式会社
対象とするモーダル種	画像、音声、動画		

技術開発の取組・成果

- 複数のツールが出力する異なる形式の判定結果を統合可能な形式に正規化し三段階判定を実装する統合判定ロジックを構築した。
- 統合ダッシュボードを開発し、評価・可視化・パラメータ最適化を一元的に行える環境を整備した。
- 統合ロジックにより判定精度が向上し、過検知や見逃しの低減を実現し、単独のツール製品では達成できない高精度・高信頼度の判定を実現した。
- 可視化UIと自動評価機能により、判定根拠の透明性が向上し、パラメータ調整や実務運用が効率化され、社会実装に耐える運用ロジックが確立した。

社会実装に係る取組・成果

- 市場ニーズの深堀りと社会実装モデルの策定を行った。
- 有識者・担当者ヒアリング（報道、プラットフォーマー、タレントマネジメント、ブランド保護、行政機関、教育機関、ツール提供会社等）を実施し、実務における具体的なペインポント（人件費、風評被害、ブランド毀損、法務コスト）を抽出した。
- ヒアリング対象者の視覚的な理解を促進し、課題やニーズを引き出す為、当初の計画にはなかったモックを作成した。
- モックを用いたヒアリングを実施した結果、根拠を示した判定支援の必要性を確認した。
- 市場ニーズを把握し、事業化における具体的な経済価値の特定、売上具体化に向けた具体性の高いビジネスモデルの構想を策定した。

技術開発及び社会実装にあたっての課題・展望

- 統合判定においては、重み付けの調整により過検知低減や見逃し低減といった判定特性の切り替えが可能である一方、統合性能の最適化とユーザニーズをどの程度反映させるべきかについて、明確な判断基準の必要性を感じた。
- 悪意をもって編集・加工が施されたコンテンツについては、元の著作物や発信者の意図とは異なる印象を受け手に与える可能性が高く、社会的影響が大きい。このようなケースでは、コンテンツ自体が真正であっても、文脈の切り取りや意図的な演出によって誤解を招くおそれがあり、単純な真偽判定のみでは社会的リスクを十分に評価できないことが大きな課題であることを認識した。
- 統合結果の性能指標とユーザニーズ（過検知許容度、見逃し許容度）を結び付ける評価軸を整理し、ユースケース別に適切な統合方針を選択できる設計を検討する。既に前処理や多段階評価は実施しており、今後は、コンテキスト判別の高度化により、単なる真偽判定を超えた、実効性の高いリスク評価システムの構築を目指していく。

代表者コメント

サン電子株式会社
グローバルDI事業部
事業部長　須藤慎二

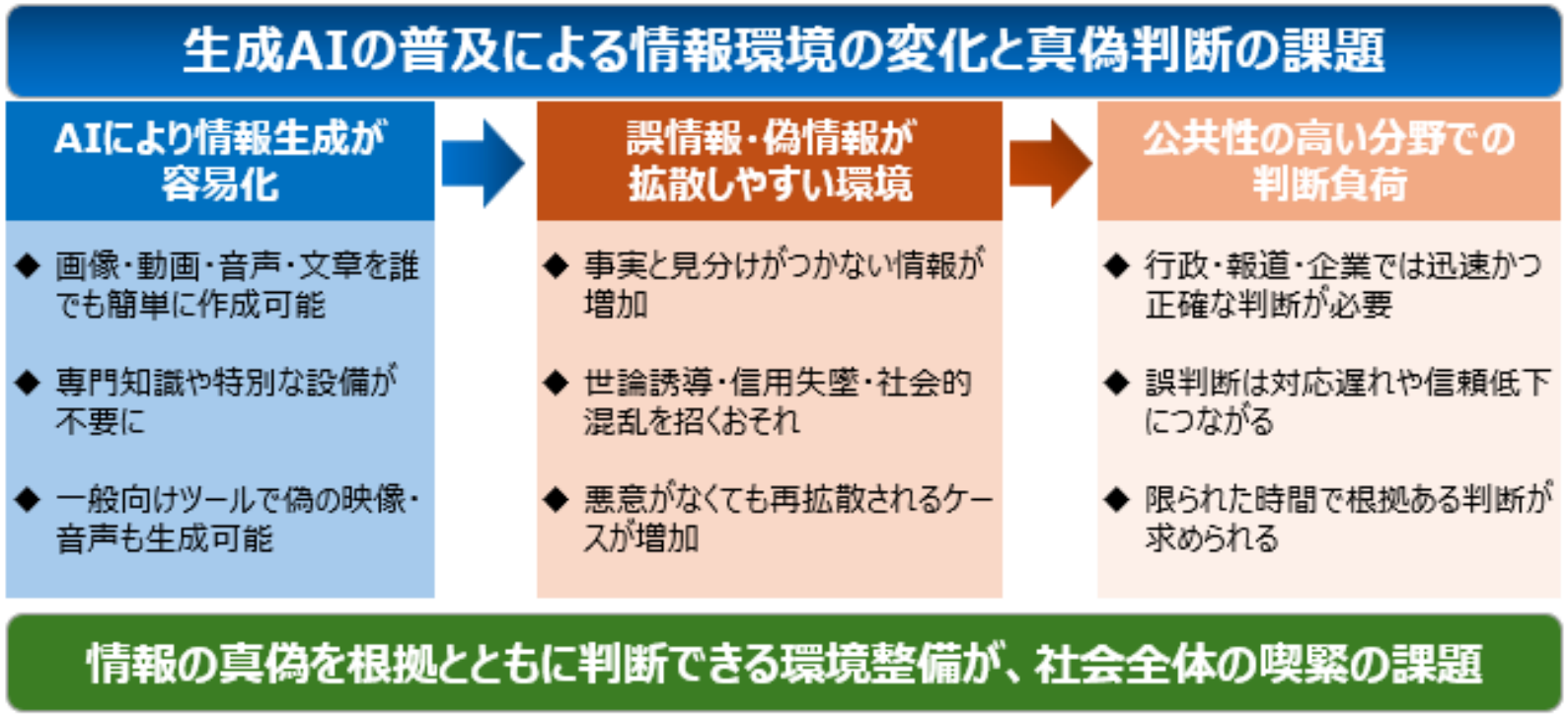
本プロジェクトでは、複数の検知技術を統合し三段階で判定する独自ロジックを確立。実証を通じ、精度の向上に加え「判定根拠の可視化」が実務に不可欠であると確認しました。今後は単なる真偽判定を超え、文脈や意図まで考慮した高度なリスク評価システムの構築を目指し、誰もが情報を安心して信じられる社会の実現に貢献してまいります。

目次

2-1. 開発技術によりアプローチする課題

背景となる社会的課題（偽誤情報の高度化・拡大）

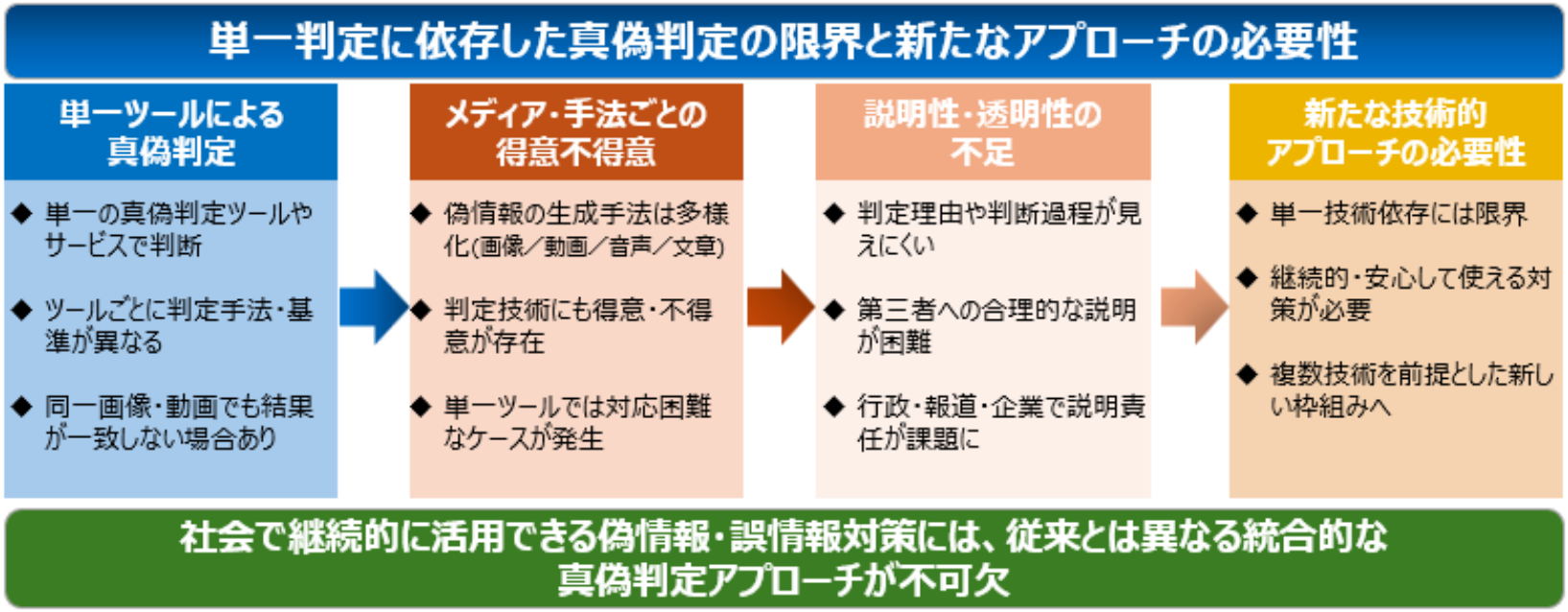
- 近年、AI技術の進展により、画像、動画、音声、文章といった情報コンテンツを、専門的な立場に限らず幅広い利用者が作成できる環境が拡大している。
- また、従来は高度な専門知識や専用設備を要していた高度な加工・生成行為についても、操作の自動化やモデルの高度化により、専門的技能を有しない者でも実行可能となっている。
- さらに、こうした生成機能は研究用途に限定されるものではなく、一般に公開・提供されているサービスや市販ソフトウェア等を通じて利用可能となっており、偽の映像・音声等の生成が特別な環境を要しない状況が生じている。



2-1. 開発技術によりアプローチする課題

既存技術の技術的課題（単独判定の限界）

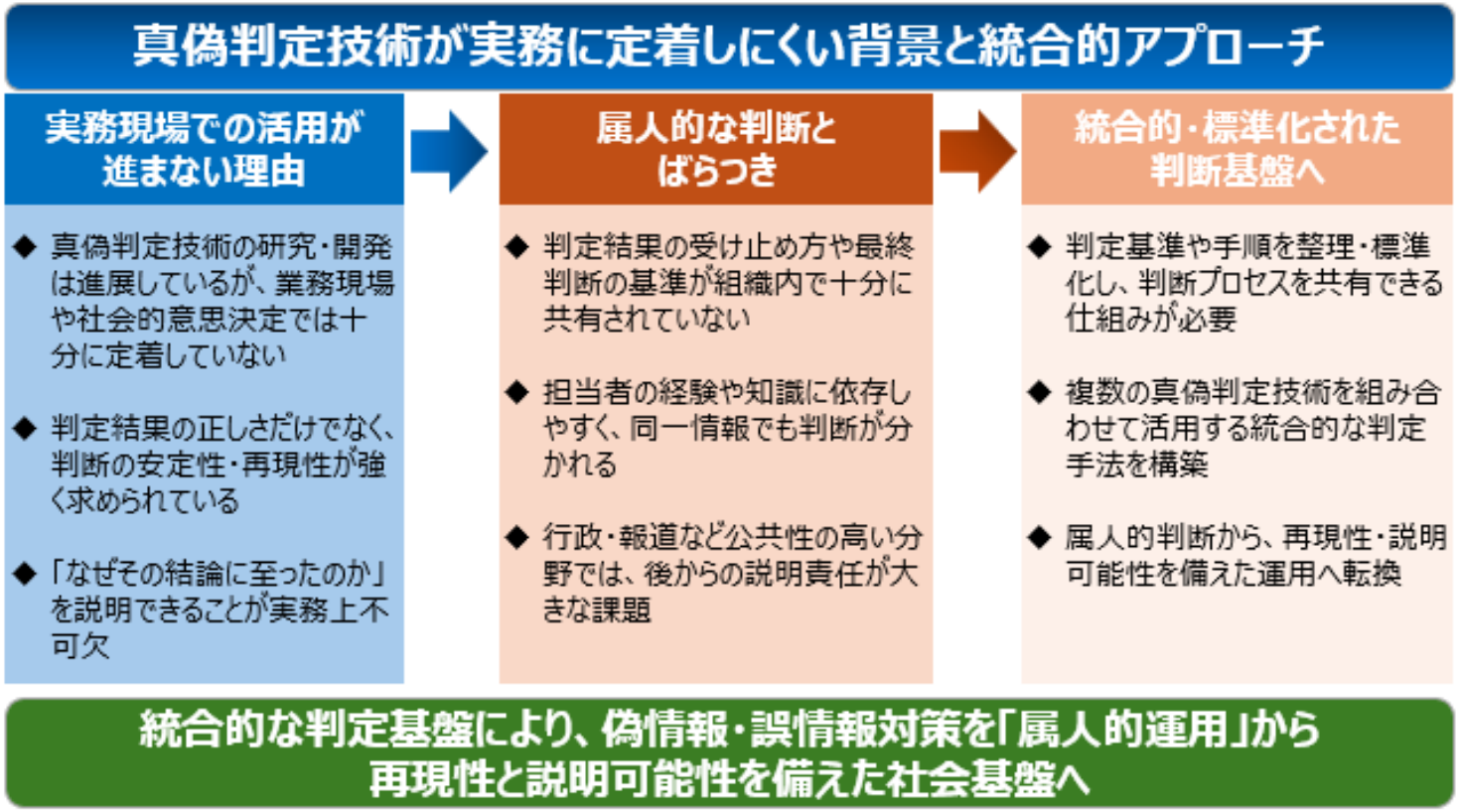
- 多くの偽・誤情報対策は、単一の真偽判定ツールに依存しているが、ツールごとに基準が異なるため同一情報でも結果が一致しない場合がある。
- また、画像・動画・音声・文章などメディアごとに生成手法や検出特性が異なり、単一ツールでは見逃しや誤判定のリスクが残る。
- さらに、判定根拠の可視化が不十分で、行政・報道・企業において説明責任の観点から課題となっている。このため、単一技術に依存しない新たな真偽判定アプローチが求められている。



2-1. 開発技術によりアプローチする課題

運用・社会実装上の課題（標準化・説明可能性）

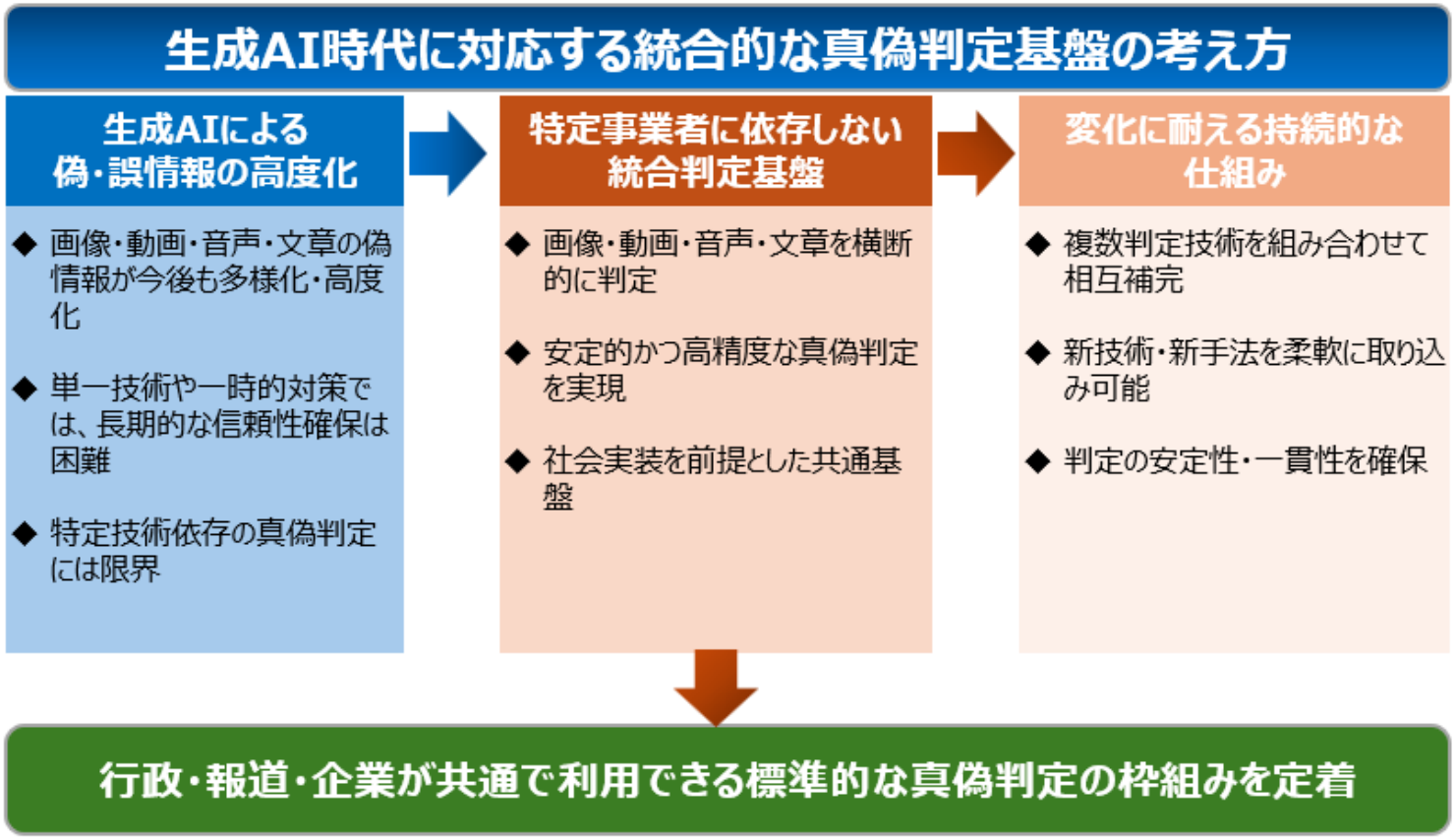
- 偽情報対策技術は進展しているが、実務現場では十分に活用・定着していない。
- その要因として、判定の正確性だけでなく、再現性や説明可能性が強く求められている点がある。
- 現状では判断基準や手順が十分に共有されておらず、担当者の経験に依存しやすい。
- このため、複数技術を組み合わせ、再現性と説明可能性を備えた統合的判定手法の整備が求められている。



2-2. 開発技術により目指す姿・ゴール

最終的に目指す全体像

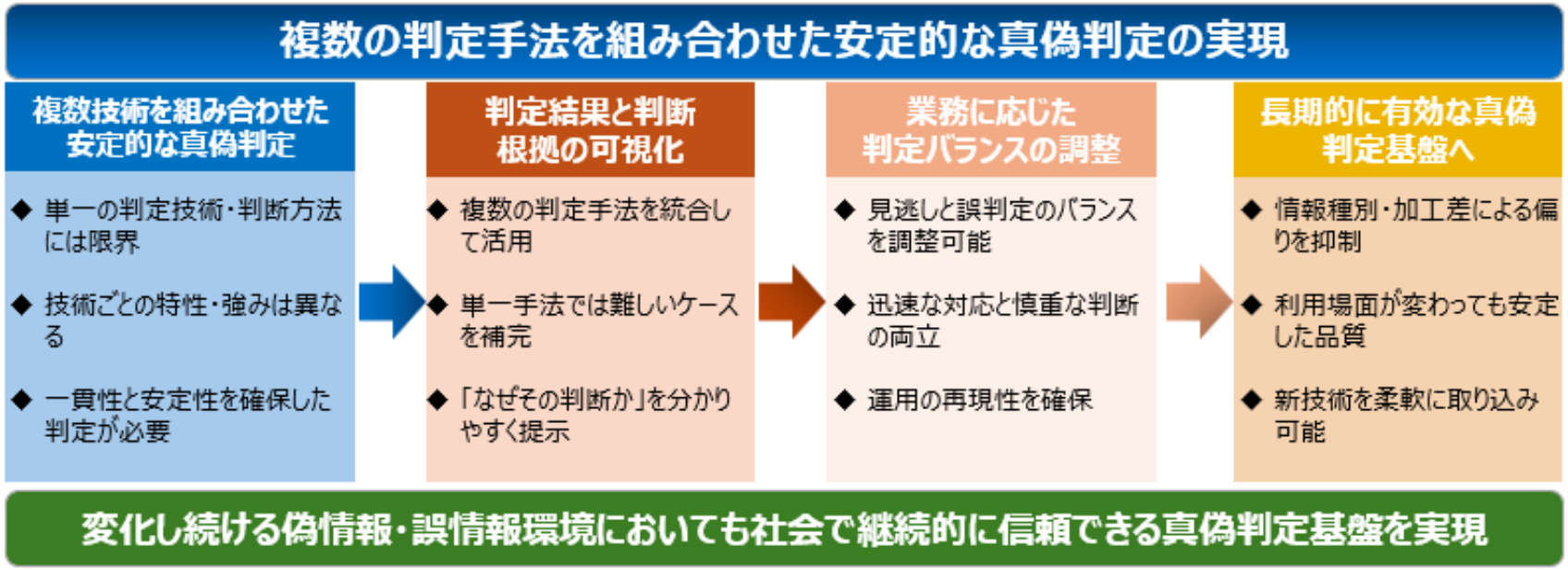
- 生成AIの発展により、偽情報・誤情報の作成手法は今後も高度化・多様化すると想定される。
- そのため、特定の技術や製品に依存した真偽判定では、長期的な対応が困難である。
- 本開発では、複数の判定技術を統合し、画像・動画・音声・文章に対応可能な判定基盤の構築を目指す。
- 最終的に、行政・報道・企業が共通して活用できる標準的な情報判断基盤の確立を目指す。



2-2. 開発技術により目指す姿・ゴール

技術的ゴール

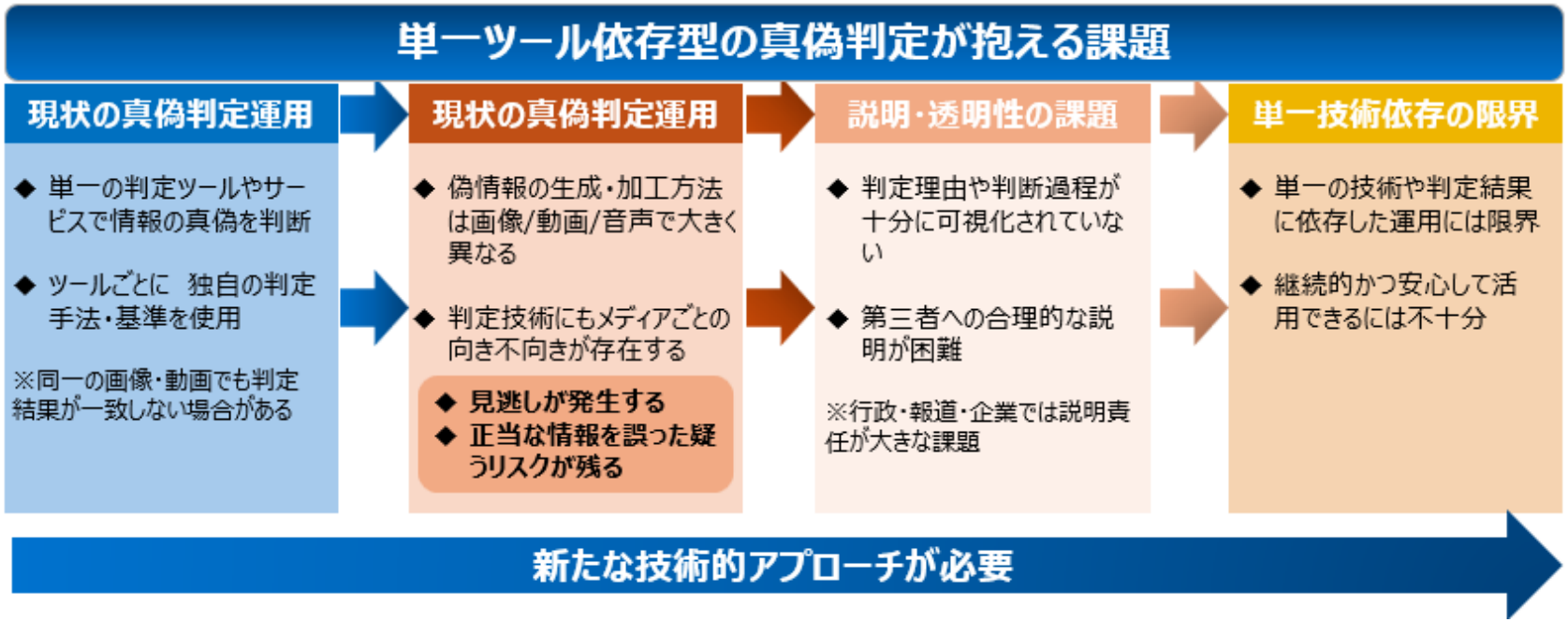
- 本技術は、複数の判定手法を組み合わせることで、単一技術に依存しない安定した真偽判定を実現することを目指す。
- また、判定結果だけでなく判断根拠を可視化し、利用者が合理的に意思決定できる環境を整備する。
- さらに、見逃しと誤判定のバランスを用途に応じて調整できる設計とする。
- 将来的な新技術も取り込み可能な柔軟な真偽判定基盤の構築を目指す。



2-2. 開発技術により目指す姿・ゴール

社会実装後に実現する姿

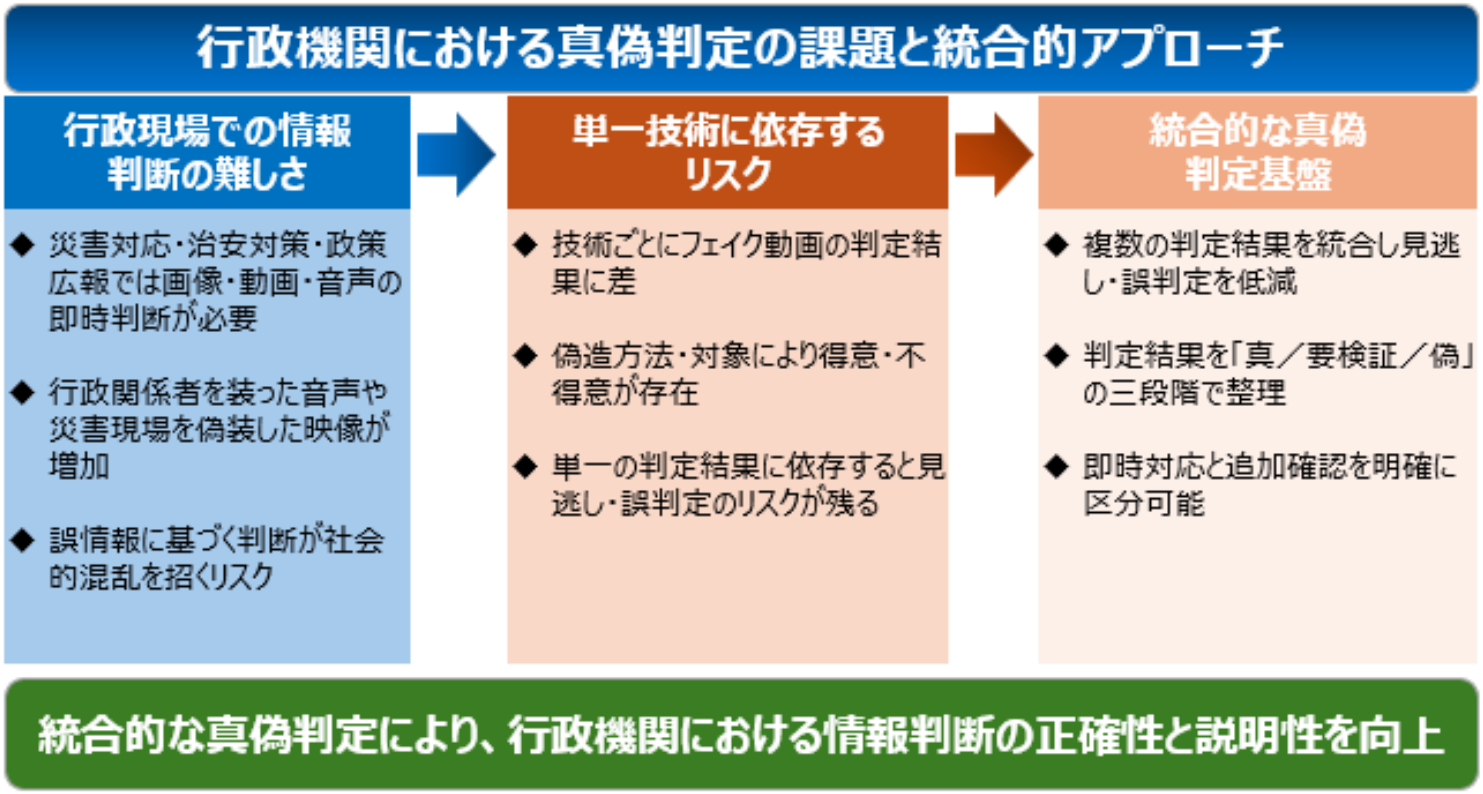
- 本技術により、真偽判断を個人の経験に依存した対応から、手順に基づく組織的な判断へ移行することを目指す。
- 判定結果と根拠を可視化することで、意思決定の質と説明可能性を向上させる。
- これにより、行政・報道・企業において判断の一貫性と再現性の確保が期待される。
- 最終的に、社会全体で合理的な情報判断が行える環境の整備に寄与する。



2-3. 開発技術により対処可能なユースケース

行政・公共機関におけるユースケース

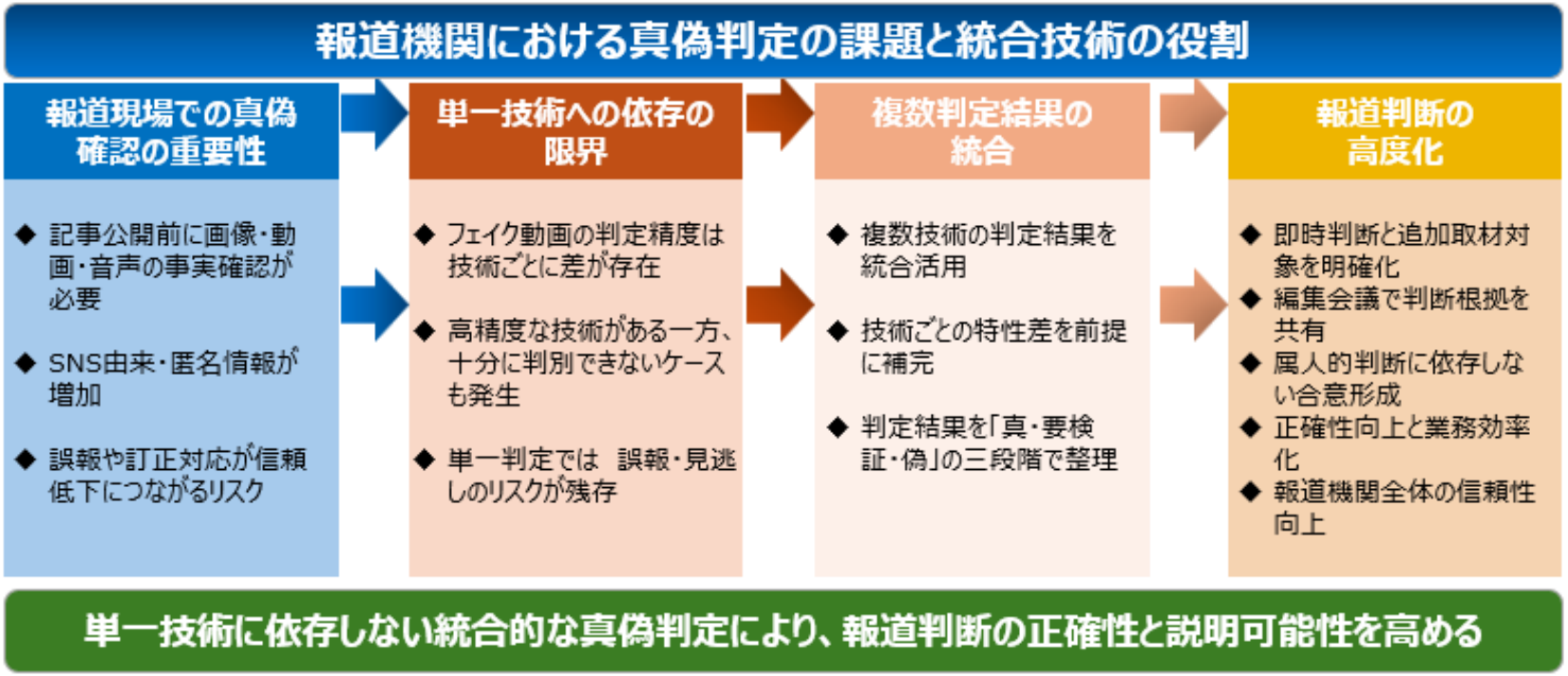
- 行政機関では、災害対応や政策広報において、画像・動画・音声の真偽を迅速に判断する必要がある。
- しかし、検証の結果、フェイク動画の判定は技術ごとに結果の差が大きく、単一ツール依存にはリスクがあることが確認された。
- 本技術は複数の判定結果を統合し、見逃しや誤判定の低減を目指す。
- さらに「真・要検証・偽」の三段階判定により、迅速かつ説明可能な情報判断を支援する。



2-3. 開発技術により対処可能なユースケース

報道機関・メディアにおけるユースケース

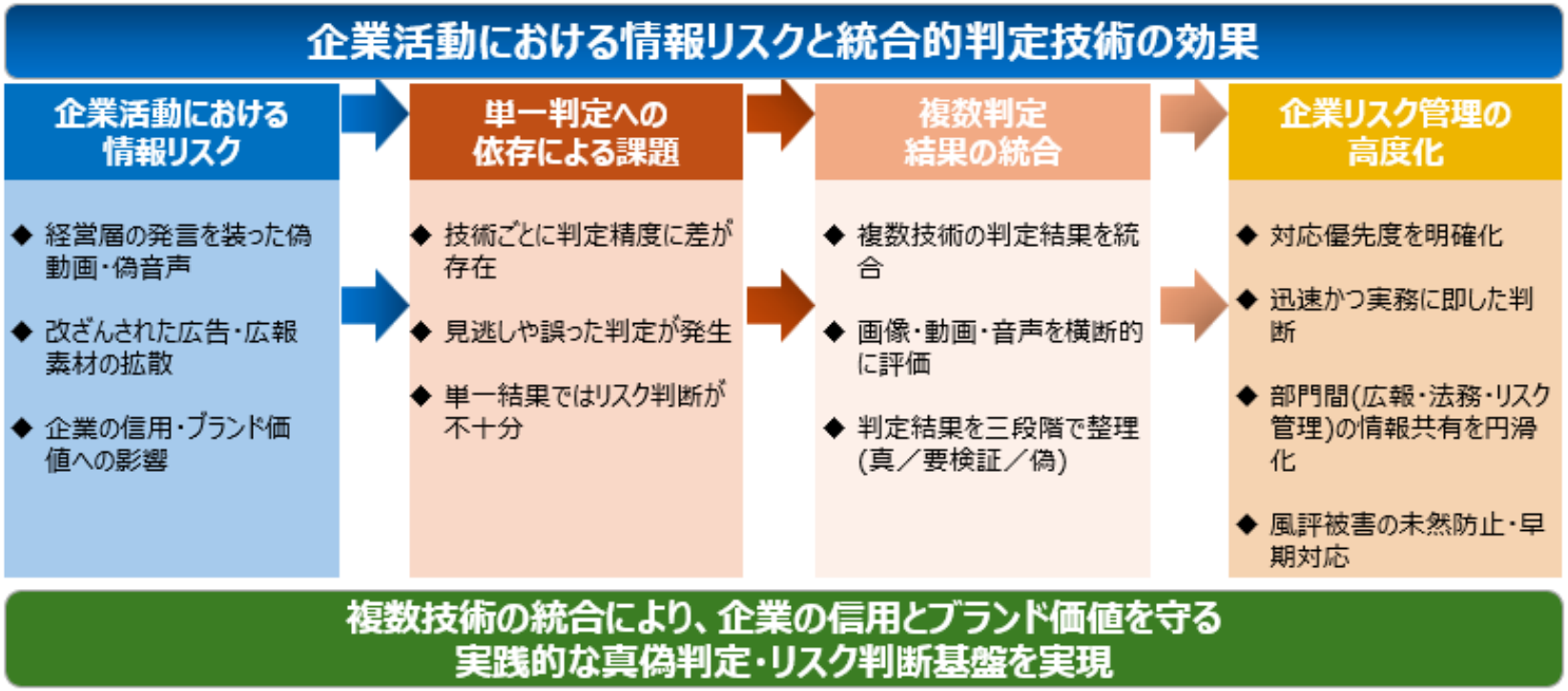
- 報道機関では、取材素材（画像・動画・音声）の真偽確認が重要だが、SNS由来情報の増加により誤報リスクが高まっている。
- 実証の結果、フェイク動画の判定精度には技術差があり、単一ツール依存には見逃しや誤判定のリスクがある。
- 本技術は複数の判定結果を統合し、「真・要検証・偽」の三段階で整理する。
- これにより判断根拠の共有と合意形成を支援し、報道の正確性と信頼性の向上に寄与する。



2-3. 開発技術により対処可能なユースケース

企業における風評・リスク管理のユースケース

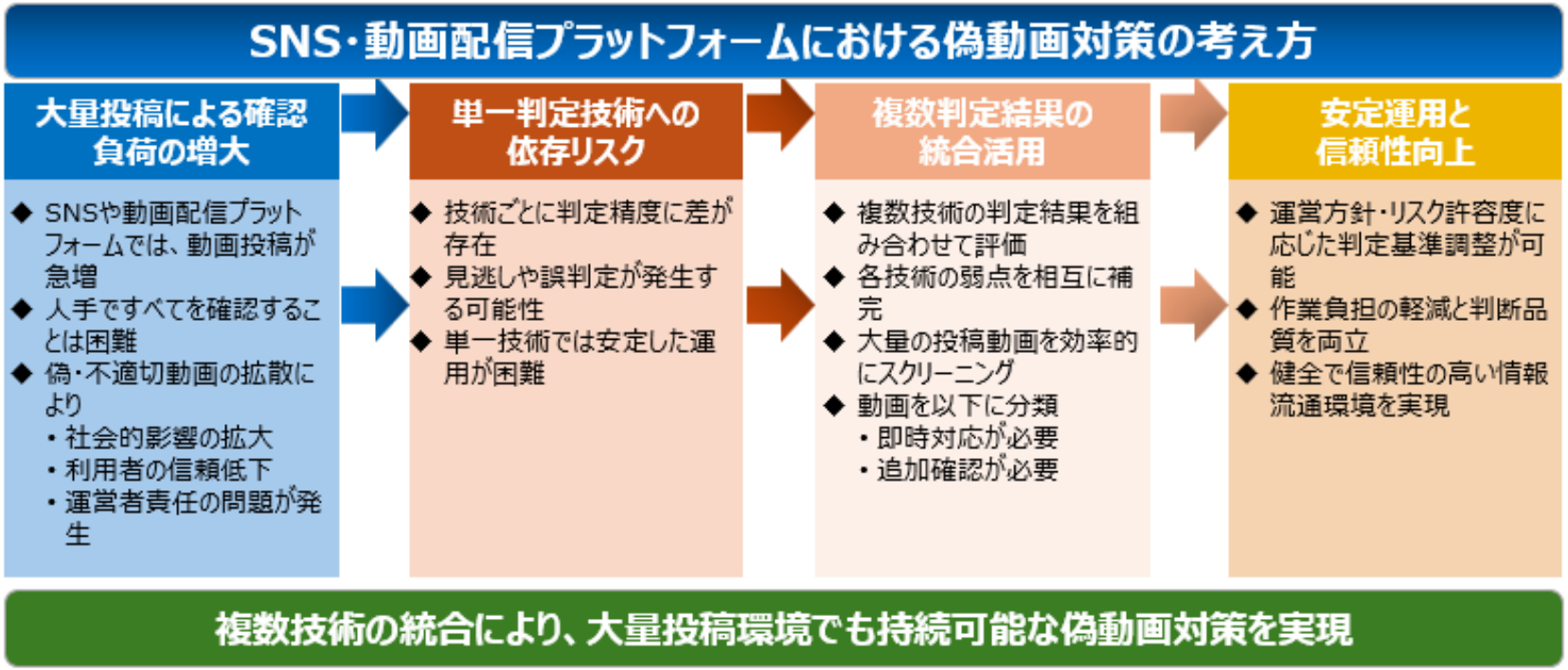
- 企業では、経営層の偽発言動画や改ざん広告の拡散により、ブランド価値や信用が損なわれるリスクが高まっている。
- 本実証では、フェイク動画の判定結果が技術ごとに異なり、単一技術だけでは十分に対応できない可能性が確認された。
- 本技術は、画像・動画・音声を横断的に評価し、複数の判定結果を組み合わせてリスクを把握する。
- これにより、広報・法務・リスク管理部門が連携した迅速な企業対応を支援する。



2-3. 開発技術により対処可能なユースケース

SNS・プラットフォーム運営者のユースケース

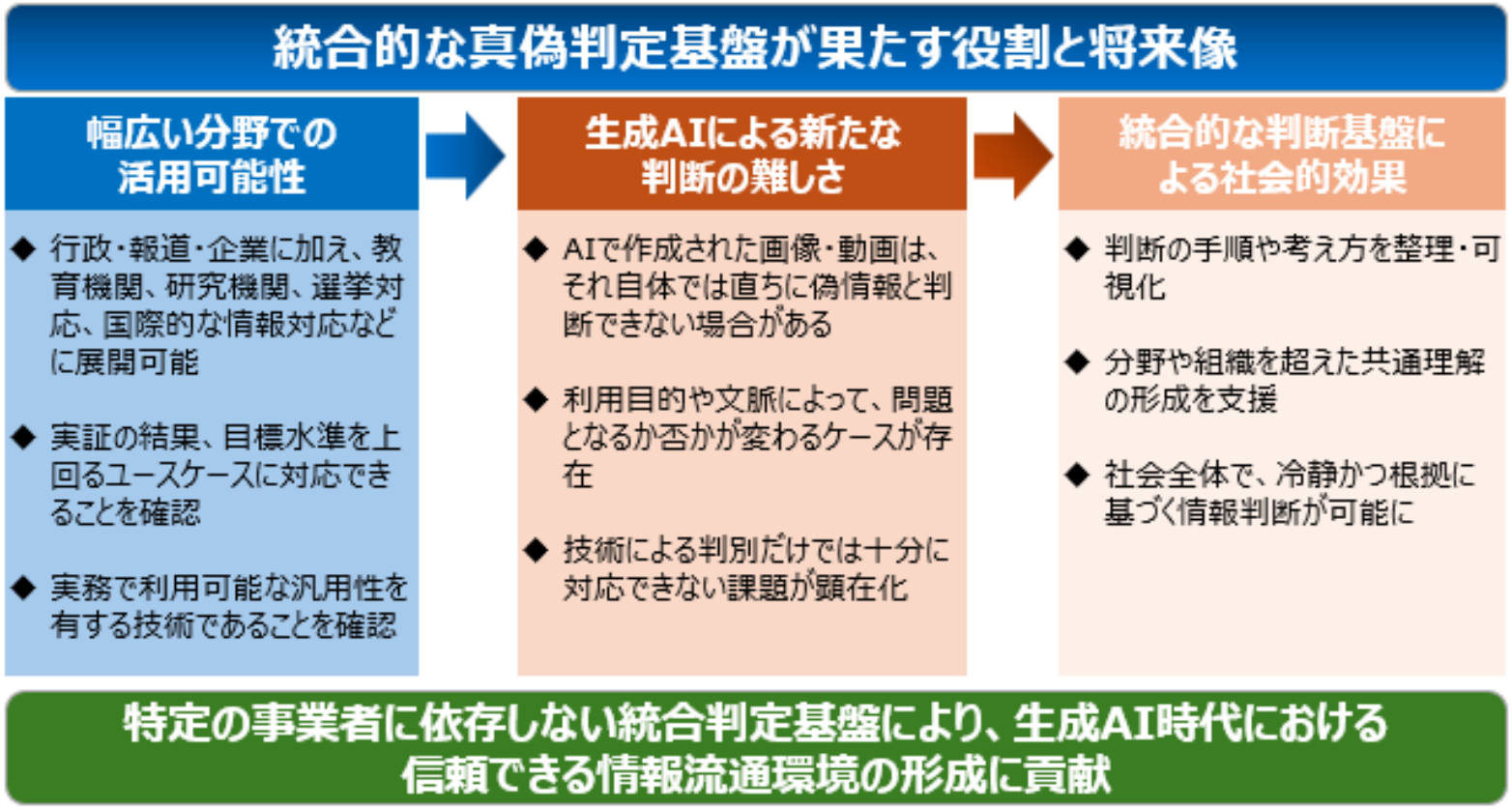
- SNSや動画配信プラットフォームでは大量の動画が投稿されるため、人手のみでの真偽確認は困難である。
- 実証の結果、フェイク動画の判定精度には技術差があり、単一技術では見逃しや誤判定のリスクが残る。
- 本技術は複数の判定結果を組み合わせ、大量動画を効率的にふるい分ける仕組みを提供する。
- これにより、対応優先度の整理と運用負担の軽減を図り、健全な情報流通環境の維持に寄与する。



2-3. 開発技術により対処可能なユースケース

社会全体への横断的ユースケースと効果

- 本技術は、行政・報道・企業に加え、教育・研究、選挙対応など幅広い分野での活用が期待される。
- 実証の結果、当初目標を上回るユースケースへの対応が可能であり、実務で利用できる汎用性が確認された。
- また、AI生成コンテンツは文脈によって問題となる場合もあり、技術判定に加え総合的な判断が重要である。
- 本技術は、分野を超えた共通の判断基盤を提供し、社会全体での適切な情報判断に貢献する。



目次

3-1. 技術開発の全体像

ツール全体および新規開発要素の概要図

真偽情報判定基盤の新規開発
— 単独判定から統合判定へ —

生成AI時代、単独ツールの限界を克服し、高精度・説明可能な統合判定基盤を構築して実施

生成AI時代の課題

- 真偽判定困難な誤情報、偽情報が拡散しやすい
- 真偽判定ツール間で結果不一致
- 判定根拠不透明で共通ルールの欠如

ツールA・B・C

ツールA	ツールB	ツールC
根拠簡略	中間曖昧	詳細不足
誤差変動	柔軟不足	傾向検知
定量評価	分類特化	媒体限定
数値出力	判定明快	詳細不足

真偽情報判定基盤の新規開発

判定結果

ツールA	ツールB	ツールC
Fake0.8	0.97	Suspicious

正規化 ※判定形式の統一化
異なる出力仕様を統一

ツールA	ツールB	ツールC
0.9	0.97	0.75

統合ロジック
ツールの特徴を活かした重み付け統合
Weighted統合方式
w1×0.9+w2×0.97+w3×0.75

目指す姿

- 単独判定から統合判定へ
- 説明可能かつ再現性のある社会実装基盤へ

統合判定結果

- 判定の高精度化
- 判定の安定化
- 履歴管理
- 出力形式の統一化
- 判定根拠の可視化

統合判定結果

- 統合スコア
- 信頼度
- 根拠内訳

単独の真偽判定ツールを統合することで、安定的な高精度判定と根拠の可視化を実現

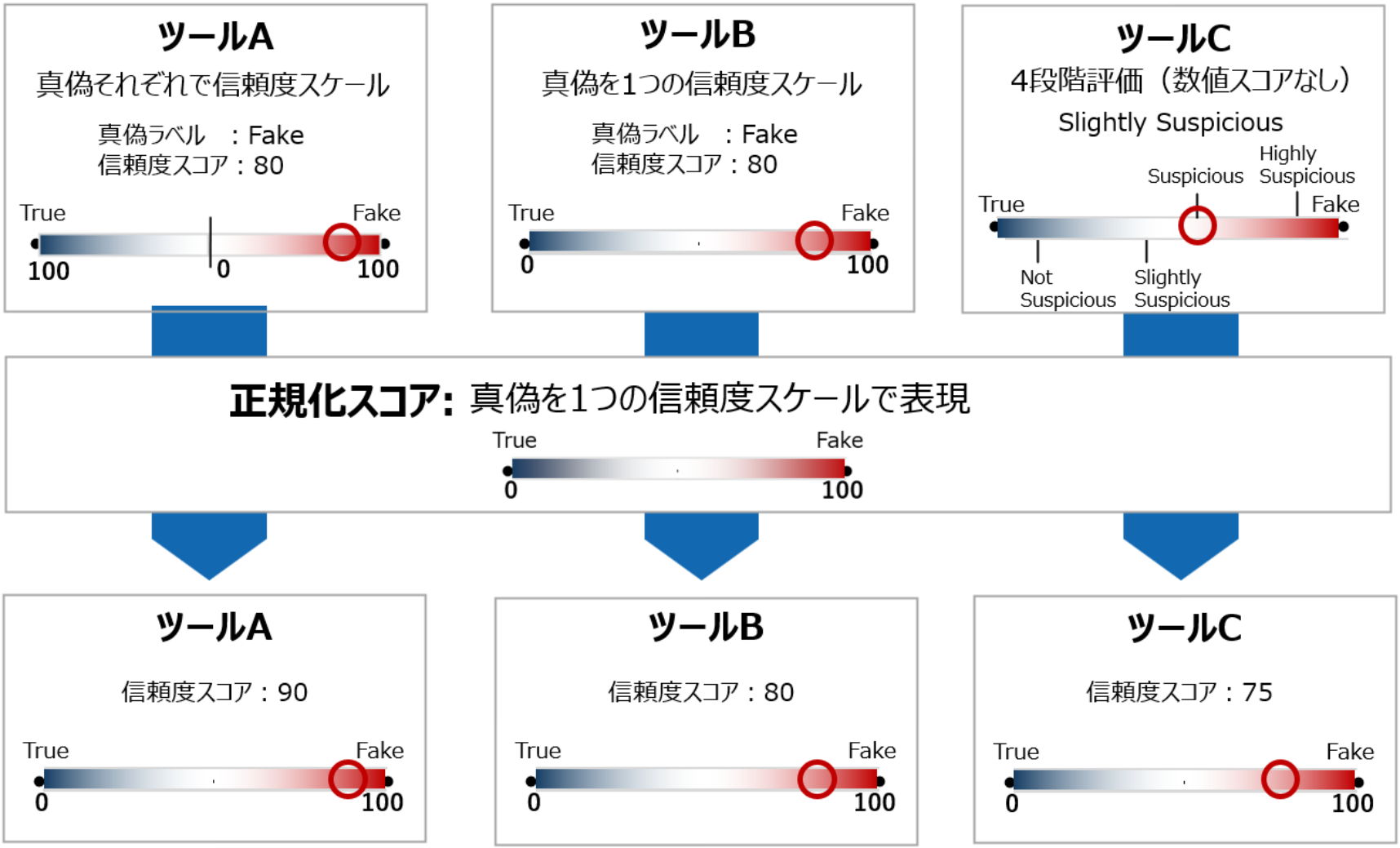
3-1. 技術開発の全体像

各ツールの特徴

項目	ツールA	ツールB	ツールC
主な対象	画像・動画・音声対応	画像・動画・音声対応	画像・動画・SNS
出力形式	真偽ラベル 信頼度スコア 判定理由	真偽ラベル 信頼度スコア 悪意の可能性 スコア判定理由	4段階評価
強み	最新生成モデルの検出	文脈分析 URL参照	ボット・偽アカウント分析
分析観点	動画（フレーム解析＋音声同期） 音声（スペクトル・時間構造解析） 画像（ピクセル構造・周波数解析） ＝＞複数生成技術観点での解析	複数生成技術観点での解析 （検出モデルはブラックボックス） 文脈分析 公式サイトとの照合	拡散構造・ネットワーク分析
想定用途	真偽判定補助	改ざん検出	世論操作リスク検知

3-1. 技術開発の全体像

異なる出力仕様を共通形式化



※本図に記載している真偽ラベルおよびスコアの数値は説明用の例示であり、最終的な評価結果や実測値を示すものではありません。

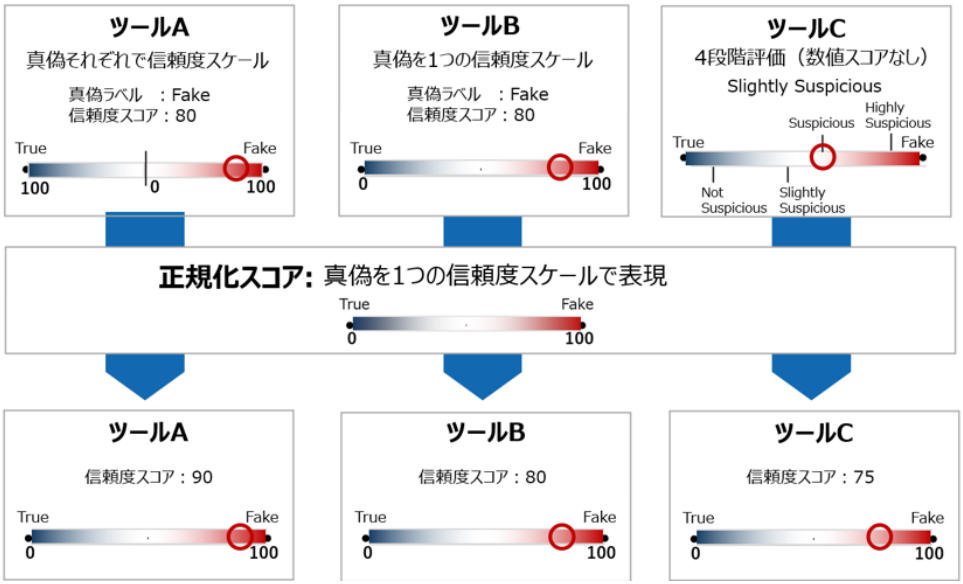
3-1. 技術開発の全体像

技術開発の全体像

統合判定アルゴリズムの設計

- 本開発・実証では、インターネット上の偽・誤情報に対応するため、画像・動画・音声の真偽判定技術の高度化と、複数ベンダーの検知結果を統合する評価基盤の構築に取り組んだ。特定の製品に依存せず、複数の海外ベンダーの判定結果を組み合わせることで、より高精度かつ安定した判定の実現を目的としている。

- 下図に示すように、各ツールが出力する異なる形式の判定結果を共通の信頼度スケールに正規化し、統合処理を行う構成とした。まず各ベンダーを同一重みで扱う Simple統合方式 により基礎的な統合効果を確認し、その後、検知特性を重みに反映する Weighted統合方式 を実装した。さらに判定閾値（t_low / t_high）を用いた三段階判定（真・要検証・偽）を導入し、即時判断が可能な情報と追加検証が必要な情報を区分できる仕組みを構築した。



3-1. 技術開発の全体像

技術開発の全体像

3-1. 技術開発の全体像

技術開発の全体像　各指標の使用意図

評価指標の選定方針

- 本実証では、統合方式の性能を多面的に評価するため、適合率、再現率、F1スコア、ROC曲線、PR曲線といった複数の評価指標を用いた。

適合率・再現率・F1スコアによる基本評価

- 適合率は、「偽と判定した情報のうち、実際に偽であった割合」を示し、過検知の抑制を評価する指標である。再現率は、「実際に偽である情報をどれだけ見逃さずに検出できたか」を示し、見逃しの低減を評価する指標である。
  F1スコアは、適合率と再現率の調和平均であり、両者のバランスを一つの値で評価できるため、総合的な性能指標として用いている。

ROC曲線による識別性能の評価

- ROC曲線は、判定閾値を変化させた際の真陽性率（実際に偽である情報をどれだけ見逃さずに検出できたか）と偽陽性率（実際には正しい情報を誤って偽と判定したか）の関係を示すものであり、判定基準に依存しない全体的な識別性能を把握するのに適した指標である。そのため、統合方式同士の基本的な性能比較に用いている。

PR曲線による実運用を想定した評価

- 一方、PR曲線は、適合率と再現率の関係を示す指標であり、誤検知と見逃しのバランスに着目した評価が可能である。偽・誤情報の検出においては、誤って正しい情報を偽と判定する場合や、偽情報を見逃す場合の影響が大きいため、実運用を想定した評価としてPR曲線を採用している。

複数評価指標の統合的活用

- 以上の理由から、本実証では、全体的な識別性能、誤検知および見逃しの傾向、ならびに両者のバランスを総合的に評価するため、これらの評価指標を組み合わせて用いることとした。

3-1. 技術開発の全体像

技術開発の成果

統合判定技術による精度・信頼性の向上

- 本技術開発・実証を通じて、複数のベンダーによる判定結果を統合することで、単独製品では達成が困難であった高精度かつ高信頼な真偽判定が可能であることを確認した。統合ロジックの適用により、過検知低減や見逃し低減し、判定結果の安定性が向上するなど、統合の有効性が実証された。

判定結果可視化UIによる運用基盤の整備

- また、判定結果と評価指標を可視化するUIを整備したことで、判定根拠の透明性が向上し、検証作業やパラメータ調整の効率化が実現した。これにより、担当者の経験や勘に依存した属人的な判断から、定量指標に基づく再現性のある運用へと移行する基盤が整備された。

社会実装に向けた技術基盤の確立

- 本開発で得られた成果は、行政・報道機関・企業など公共性の高い分野における真偽判定業務への適用可能性を示すものであり、今後の社会実装や対象領域の拡張に向けた基礎的な技術基盤として位置付けられる。

3-2. 技術開発の個別詳細

技術開発全体の位置づけ

- 本開発・実証期間においては、インターネット上で流通する偽・誤情報への対策技術として、画像・動画・音声を対象とした真偽判定技術の高度化および、複数の検知結果を活用した統合的な運用基盤の構築に取り組んだ。特定の製品やベンダーに依存することなく、複数の真偽判定ツールの判定結果を統合的に活用することで、実運用を見据えた高精度かつ安定した信頼性判断を実現することを目的とした。

- 本パートでは、実施計画書「3(1) インターネット上の偽・誤情報等への対策技術の開発」に基づき、判定結果の正規化、統合判定ロジックの設計・実装、評価・可視化基盤の整備といった具体的な技術開発の取組内容と、それぞれの実証を通じて得られた成果を個別に整理する。

- 本開発・実証においては、信頼性判断支援ユースケースへの対応範囲の拡大、統合判定による検知精度および安定性の向上、ならびに社会実装を見据えた運用基盤の確立を評価指標（KPI）として設定した。

- これらの評価指標は、本技術の有効性について、単一の性能指標のみで評価するのではなく、実運用を想定した適用範囲、判定性能、および社会実装に必要な基盤整備の3側面から総合的に評価することを目的として、段階的な技術開発および検証を実施した。

3-2. 技術開発の個別詳細

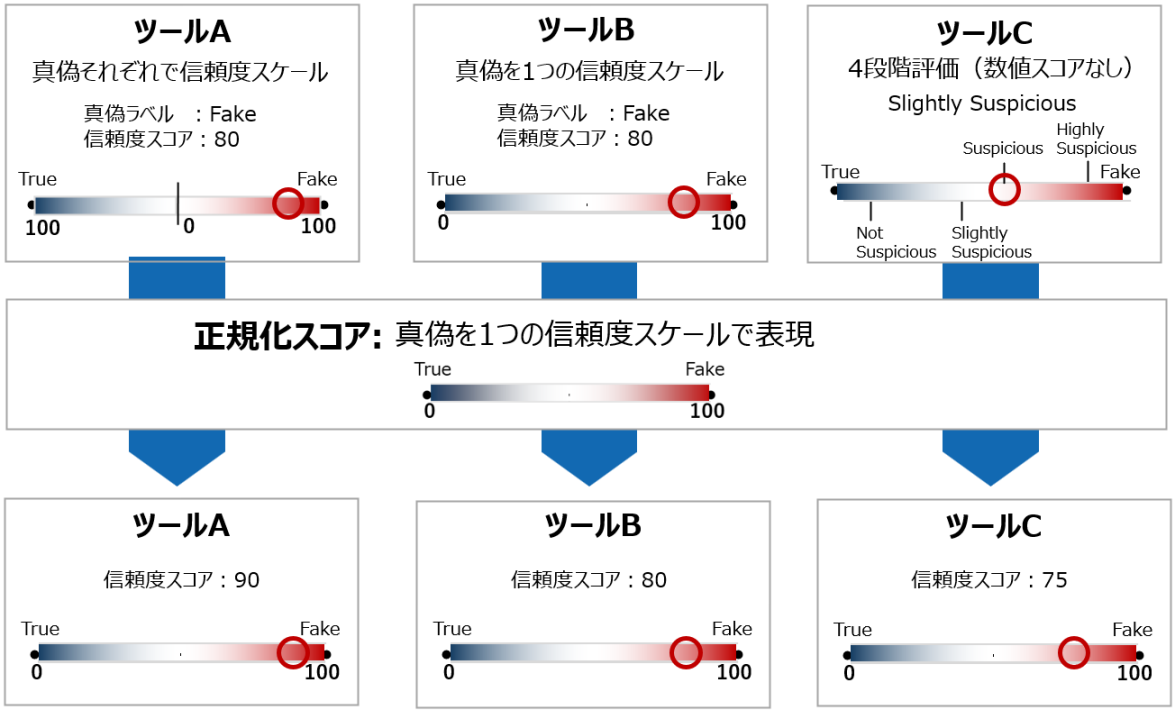
取組① 判定対象データの収集・整理

- 本実証では、画像・動画・音声といった複数コンテンツを対象に、真情報および偽誤情報データを収集・整理し、検証用データ基盤を構築した。
- まず、真情報・偽誤情報の区分が明示された一般公開の検証用データセットを収集し、生成手法や改ざん手法ごとに分類したデータを基礎データとして整備した。
- 加えて、汎用生成AIモデルを用いて独自に生成した画像・動画コンテンツを作成し、既存データセットでは十分に含まれていない生成パターンや状況表現を補完した。さらに、自ら撮影・収録した実データも用意し、実環境に近い条件を含むデータを整備した。
- 収集したデータには、解像度、圧縮方式、ファイル形式、生成手法、加工条件などの属性情報を付与し、複数の条件で整理・管理できる構造とした。これにより、異なるメディア特性や生成手法を含む多様な条件を扱うことができる検証用データ基盤を構築した。

3-2. 技術開発の個別詳細

取組② 複数真偽判定ツールの判定結果の正規化

- 各真偽判定ツールの判定結果は、スコアの範囲、信頼度の表現方法、判定の粒度がそれぞれ異なっており、そのままではツール間での横断的な比較や統合処理を実施することが困難であった。
- そこで本取組では、各ツールの出力仕様を整理した上で、判定結果を共通のスケールおよび評価軸に変換する正規化処理を実装した。
- 正規化処理によりツール間の出力差異を吸収し、ツールやコンテンツの種類にかかわらず、一貫した評価が可能となった。その結果、従来は単独ツールごとにしか行えなかった評価を、定量的かつ統一的に比較・統合する基盤を構築することができた。



※本図に記載している真偽ラベルおよびスコアの数値は説明用の例示であり、最終的な評価結果や実測値を示すものではありません。

3-2. 技術開発の個別詳細

取組③ 統合判定ロジック（Simple統合）の実装

- 複数の真偽判定ツールの判定結果を同等に扱う Simple統合方式 を実装し、各ツールの出力スコアを平均化することで統合判定を行う仕組みを構築した。本方式は、複数ツールの判定結果を統合するための基礎ロジックとして位置づけられるものである。
- Simple統合方式では、各ツールの判定結果を同一の重みで扱うため、統合処理の基本構造をシンプルに実装できる一方で、ツールごとの検知特性や信頼度の違いは反映されない構成となっている。
- この特性を踏まえ、各ツールの性能特性やメディア種別ごとの検知傾向を考慮した Weighted統合方式 を設計した。Weighted統合方式では、ツールごとの信頼度や条件別の性能指標を重みとして設定し、統合スコアに反映する仕組みとしている。
- これにより、複数ツールの特性を踏まえた統合判定を実現するための拡張可能な統合ロジックを構築した。

※本図に記載しているスコアおよび重み付けの数値は説明用の例示であり、最終的な評価結果や実測値を示すものではありません。

3-2. 技術開発の個別詳細

取組④ 統合判定ロジック（Weighted統合）の実装

設計思想・統合方針

- 本取組では、複数の真偽判定ツールの判定結果を統合するための方式として、各ツールの判定結果を同一条件で扱う Simple統合方式 に加え、ツールごとの検知特性や精度傾向を反映する Weighted統合方式 を設計・実装した。
- Weighted統合方式では、各ツールの判定結果を一律に扱うのではなく、ツールごとの検知特性やメディア種別ごとの特性を踏まえて影響度を調整し、統合スコアに反映する構成としている。
- 特に、画像・動画・音声といったメディア種別の違いや、動画におけるテロップ付与や音声差し替えなどの加工条件を考慮し、条件に応じた重み付けが可能な統合ロジックとして設計した。

評価指標・重み付けの算定方法

- Weighted統合方式では、重み設計の指標として ROC曲線 と PR曲線 を用いる構成とした。
- ROC曲線は真陽性率と偽陽性率の関係から識別性能を把握する指標であり、PR曲線は適合率と再現率の関係から誤検知と見逃しのバランスを評価できる。
- これらの指標を基にメディア種別の特性を考慮した重みを設定できる仕組みとし、重みは固定値ではなく、データや運用条件に応じて調整可能なパラメータとして実装した。
- これにより、将来的な再チューニングや新規ツール追加にも対応可能な設計としている。

3-2. 技術開発の個別詳細

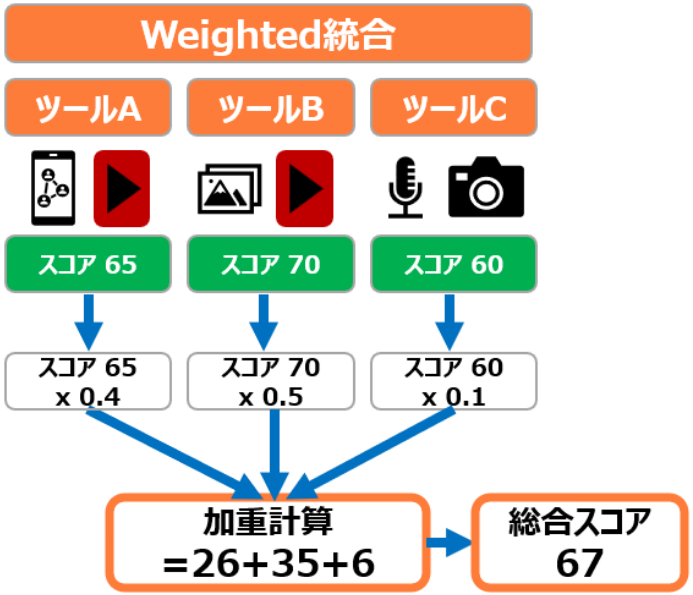
取組④ 統合判定ロジック（Weighted統合）の実装

得られた成果

- 本取組では、複数の真偽判定ツールの結果を統合する方式として、各ツールの特性を反映する Weighted統合方式 を実装した。Weighted統合方式では、ツールごとの検知特性やメディア種別ごとの特性を踏まえて重みを設定し、統合スコアに反映する仕組みとしている。

- また、動画コンテンツにおいては、テロップ付与や映像の切り貼り、音声差し替えなど多様な編集・加工が存在することを考慮し、単純なメディア種別だけでなく、加工条件を含めた特徴に応じて重みを調整できる統合設計とした。

- さらに、本統合方式では重みを固定値とせず、対象データの特性や運用条件に応じて調整可能なパラメータとして実装している。これにより、メディア種別や加工形態の違いに柔軟に対応しながら、統合判定ロジックを拡張できる構成としている。

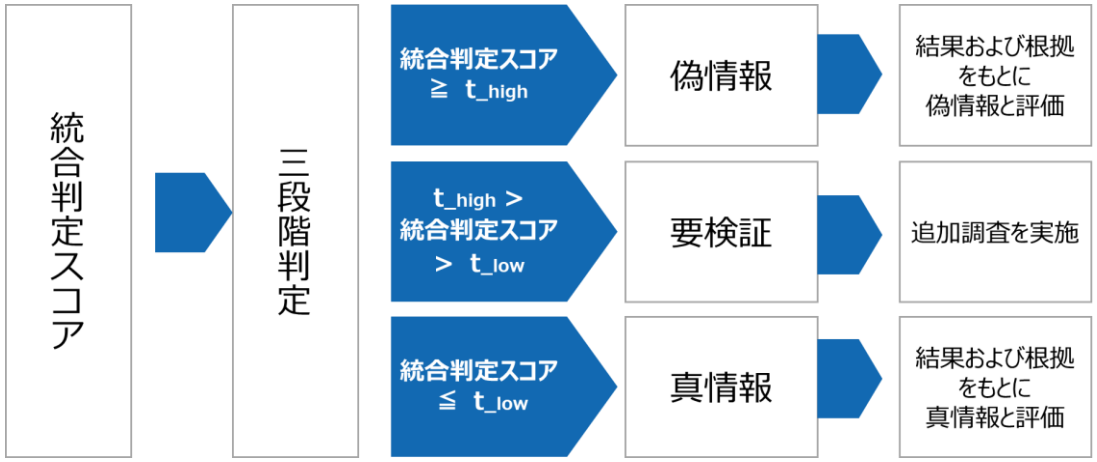


※本図に記載しているスコアおよび重み付けの数値は説明用の例示であり、最終的な評価結果や実測値を示すものではありません。

3-2. 技術開発の個別詳細

取組⑤ 三段階判定ロジックの設計

- 本技術では、統合判定ロジックとして、単純な真偽の二値判定にとどまらず、判定結果の確信度に応じた 三段階判定（真・要検証・偽） を導入した。
- 具体的には、複数ツールの判定結果を統合して得られる統合スコアに対し、2つの判定閾値（t_low / t_high）を設定し、「t_low 未満を真」、「t_high 超過を偽」、「その中間を要検証」とする判定方式としている。
- このように「要検証」領域を設けることで、AIによる自動判定結果を最終判断とするのではなく、実務担当者による追加確認や判断を前提とした運用を可能とする設計としている。
- また、本判定方式では、判定に用いる分析モデルや判定根拠を可視化するとともに、必要に応じて公開情報（公式URL等）と照合できる機能を備えることで、判定結果の根拠を確認できる構成としている。これにより、判定過程を第三者にも理解しやすい形で提示できる仕組みとしている。



3-2. 技術開発の個別詳細

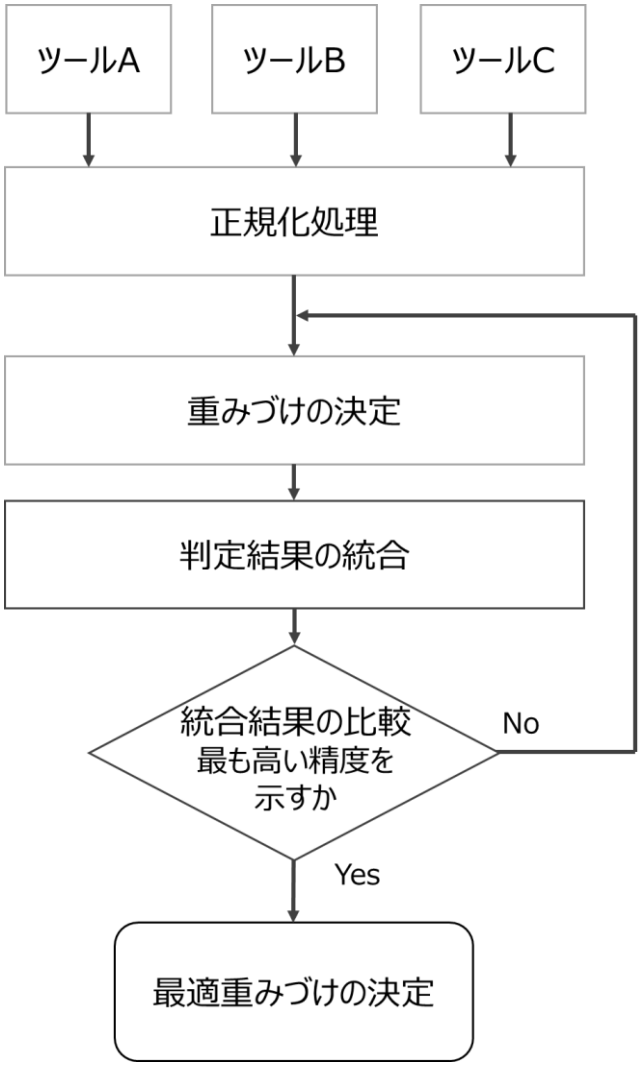
取組⑥ 評価指標・性能可視化機能の実装

- 本実証では、適合率、再現率、F1スコア、誤検知率、ROC曲線、PR曲線等の評価指標を算出するため、各指標を計算可能なツールを作成した。
- 判定結果および正解ラベルを入力することで、各評価指標が自動的に算出される構成とし、検証内容の更新に応じて評価結果を容易に再計算できる環境を整備した。
- これにより、検証の都度、判定性能を定量的に把握することが可能となり、判定ロジックやパラメータ調整による性能変化を客観的に比較・評価できるようになった。また、関係者間での結果共有や再検証が容易となり、継続的な評価・改善に資する評価基盤を構築した。

3-2. 技術開発の個別詳細

取組⑦ パラメータ調整・最適化検証

- 本技術では、統合判定方式のパラメータ調整および性能確認を行うための 評価指標・性能可視化機能 を実装した。
- 具体的には、正規化後の各ツールの判定結果を入力データとし、適合率、再現率、F1スコア、ROC曲線、PR曲線などの評価指標を算出できる仕組みとした。これにより、統合判定の挙動を定量的な指標として把握できる構成としている。
- また、重み付けパラメータを段階的に変更しながら評価指標を更新することで、パラメータ設定の違いによる判定結果の変化を確認できる設計とした。
- さらに、Simple統合方式およびWeighted統合方式の双方に同一の評価手順を適用できる構成とし、評価結果の推移や調整前後の差分を可視化する機能を備えた。これにより、パラメータ調整の過程を定量的に確認できる検証環境を構築した。



3-2. 技術開発の個別詳細

取組⑧ 判定方式の再現性および継続的改善に関する方針

- 本判定方式は、同一入力・同一設定条件下において同一結果が再現される設計としている。使用するモデルのバージョン、重み付けは固定管理されており、同一構成で処理を実行した場合には、統合スコアおよび判定区分が再現可能である。
- 判定閾値については、利用者の運用要件に応じて変更可能な設計としており、状況や要件に合わせて変更可能な値となっているが、同一閾値を適用すれば同一結果を再現できる構成となっている。
- また、各ツールの出力値、正規化後スコア、重み付け、最終スコア等を記録可能としており、判定過程の内訳を追跡できる。これにより、再現性を確保するとともに、監査対応を見据えた検証可能な設計としている。
- 本判定方式においては、将来的なモデル更新およびデータ分布の変化に対応するため、継続的な性能評価および重み付けの最適化の仕組みを開発・実装する方針である。
- 具体的には、モデルのバージョン更新時に既存検証データセットを用いた再評価を実施し、ROC曲線およびPR曲線の評価指標に基づき、更新前後の性能差を定量的に比較する仕組みを整備する予定である。一定の性能基準を満たす場合にのみ本番環境へ反映する運用とすることを想定している。
- また、実運用データの傾向を継続的に把握し、入力データの分布変化や誤検知・見逃し傾向の変化を検知した場合には、再評価を実施する体制を構築する予定である。必要に応じて、重み付けの再最適化を行う設計とする。

3-2. 技術開発の個別詳細

総括KPI（技術開発の達成度評価）

- 本開発・実証期間では、統合判定技術の社会実装可能性を総合的に評価する観点から、3つのKPIとして設定した。これは、技術の有効性を単一の性能指標のみで評価するのではなく、実運用を想定した適用範囲、判定性能、および社会実装に必要な基盤整備の3側面から総合的に評価することを目的としたものである。
- KPI①については主要ユースケース対応率の目標60%に対し67%を達成し、想定したユースケース群において統合判定技術が適用可能であることを確認した。
- KPI②については、画像を中心に精度向上および誤検知率低減の効果が確認され、画像においては精度が7.0%向上し、動画においては精度が7.5%向上した。また、誤検知率は最大で約50%の低減となるなど、統合判定による性能改善の有効性が確認された。
- KPI③については、判定結果正規化機能、統合判定ロジック、統合効果検証手法の確立、および社会実装実施計画の策定を計画どおり完了し、社会実装に向けた技術基盤を整備した。
- 以上より、本年度に設定したKPIは概ね達成され、統合判定技術が信頼性判断支援のための実用的な基盤として機能する可能性が示された。今後は、動画を含む多様なコンテンツへの適用性向上、評価データの拡充、ならびにユースケースごとの最適な重み付け・閾値設計の高度化を進めることで、適用領域の拡大と社会実装の具体化を図る予定である。

KPI項目	評価指標	本年度目標	達成度	評価
KPI① 信頼性判断支援ユースケースのカバレッジ	主要ユースケース対応率	60%	67%相当を達成	◎
KPI② 統合判定による精度・安定性向上	精度向上／誤検知率低減	精度 +5～+15% 誤検知率▲15%	精度 最大+7.5% 誤検知率 最大▲50%	◎
KPI③ 社会実装基盤の技術的完成度	基盤機能の設計・実装完了	基盤整備完了	計画通り完了	○

凡例
◎：目標達成（計画以上または想定通りの成果を確認）
○：概ね達成（社会実装に向けた基盤を確立）
△：一部未達（次年度以降に継続対応）

3-2. 技術開発の個別詳細

総括KPI（技術開発の達成度評価）

KPI①について

- 本開発では、社会的影響の大きい代表的な４つのユースケースを対象に、統合判定による真偽判定支援機能が有効に機能するかを検証した。
- 検証にあたっては、各ユースケースに対応した検証用データの収集を行い、収集したデータを用い各ユースケースに対し3つのモダリティと2つの判定対象の組合せによる計6項目について、統合判定ロジックの有効性を評価した。
- その結果、本年度目標として設定したKPI達成度60%に対し、各評価項目の総合達成度はこれを上回る水準となり、特にユースケースのカバレッジについては目標値を超過する成果が確認された。実施計画書に定めた業務事項およびKPIは達成水準に到達している。
- これらの成果は、次年度以降のユースケース拡張および社会実装段階への発展に向けた技術的基盤を形成するものである。

ユースケース	判定対象	動画	画像	オーディオ	対応項目数	カバレッジ
① メディア・報道関連	人	○	○	○	4/6	67%
	動物・物体・現象（自然現象）	○	-	-		
② 企業・組織の風評被害対策	人	○	○	○	3/6	50%
	動物・物体・現象（自然現象）	-	-	-		
③ 公的機関の情報環境保全	人	○	○	○	5/6	83%
	動物・物体・現象（自然現象）	○	○	-		
④ プラットフォーム事業者の技術基盤強化	人	○	○	○	4/6	67%
	動物・物体・現象（自然現象）	○	-	-		
				計	16/24	67%

凡例
○：対応（使用したデータ基盤がユースケースに対応）
－：非対応（使用したデータ基盤がユースケースに非対応）

3-2. 技術開発の個別詳細

総括KPI（技術開発の達成度評価）

KPI②について

- 本開発では、複数の真偽判定技術を統合することによる精度向上および誤検知低減を、技術的中核となるKPIとして設定した。
- 具体的には、単独で最も高い性能を示す真偽判定ツールを基準とし、その判定結果とWeighted統合による判定結果を比較することで、統合判定による性能向上の効果を評価した。
- 画像においては精度が最大で約+7.0%向上し、統合判定による性能改善が確認された。また、誤検知率については最大で約50%の低減が確認され、設定したKPI目標を上回る改善効果が確認された。
- これにより、複数の真偽判定ツールを統合することによる性能改善への有効性が確認された。
- 一方で、動画データにおいては見逃し低減を重視した条件では精度は7.5%の改善が見られたものの、過検知抑制領域では単独ツールが高い性能を示すケースが見られた。これは、テロップ付与等の複数の加工要素を含むことにより、各検知ツールの判定特性の違いが統合結果に影響した可能性が考えられる。そのため、今後は加工要素を含む評価データの拡充や統合時の重み付け調整等の検討を進める必要がある。

KPI③について

- 社会実装基盤に必要な要素を、「判定結果正規化機能の設計」、「統合判定ロジックの開発」、「統合効果検証手法の確立」、「実証結果に基づく社会実装実施計画書の作成」と定義し、社会実装に向けた基盤整備の達成度を評価するKPIとして設定した。
- これらの要素については、計画どおりこれらの要素に関する開発および計画策定を完了した。
- 一方で、実証を進める中で、動画加工（テロップ付与等）に対する検知安定性の向上、ユースケースごとの評価データの拡充、ならびにユーザニーズ（過検知抑制・見逃し低減等）に応じた重み付けおよび閾値設計については、社会実装基盤の構築とは切り分け、高度化・最適化を目的とした検討事項として位置付け、今後の課題として整理した。

目次

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

- 本開発・実証期間においては、統合判定技術の有効性と適用可能性を確認するため、検証用テストデータの整備、複数ツールによる実証テスト、統合方式の設計・評価、モダリティ差の分析、ならびにモックを用いたユースケース適用性の確認を段階的に実施した。
- まず、生成AIコンテンツおよび商用データセットより、真正／偽造コンテンツを収集・整理し、難易度の異なる検証データセットを整備した。
- その上で複数の真偽判定ツールを用いて評価を実施し、Simple統合およびWeighted統合による統合判定方式を設計・検証した。
- さらに、画像・動画といったモダリティごとの特性を踏まえて統合効果を分析した。
- モックを用いたユースケース検証を行い、単一ツールに依存しない判定および複数観点からの根拠提示が真偽判定の利用シーンに有効であることを確認した。

	検証データの整備	統合判定方式の検証	モダリティごとの評価	ユースケース検証
実施内容	・商用データ+生成AIコンテンツ収集 ・真偽/解像度/生成手法を整理	・Simple統合 / Weighted統合を設計 ・モダリティごとの最適重み付けの探索	・画像・動画での統合方式の有効性評価	・ユーザ体験を検証するモックを作成 ・報道/企業風評/災害/選挙の各ユースケースを机上検証
成果	・様々な判定難易度のデータセット ・画像32・動画64・音声4 	・統合による性能改善 ・Weighted統合の優位性の示唆	・画像：統合効果が明瞭 ・動画：加工や文脈評価が影響	・単一ツールに依存しない判定と複数観点からの根拠提示が可能

4-2. 検証及び調査の個別詳細

検証用テストデータの選定・準備

- 本開発・実証期間において、統合判定技術の有効性を検証するため、ディープフェイク画像および動画データの選定・準備を実施した。商用利用可能なデータセットおよび生成AIにより生成されたコンテンツを中心に、真正コンテンツとディープフェイクコンテンツを体系的に収集し、検証用テストデータとして整理した。
- データの整理にあたっては、コンテンツの性質、生成手法および検出難易度、データ特性（解像度・圧縮方式・ファイル形式等）を考慮し、多様な条件下で統合判定技術の性能を検証可能な構成とした。また、汎用生成AIモデルにより生成された高度偽造コンテンツを対象に含め、従来型ディープフェイクだけでなく最新の生成技術に対する適用可能性も評価対象とした。
- さらに、検証用テストデータはメディア報道、企業風評、災害情報、政治・社会情報等のユースケースを想定して整理し、実運用を想定した統合判定技術の適用範囲および課題を検証可能な構成とした。

整理観点	整理内容	検証目的
コンテンツの性質による分類	真正コンテンツ／ディープフェイクコンテンツを体系的に整理	判定結果の妥当性を確認し、検証結果のトレーサビリティを確保
生成手法・検出難易度	生成手法の違いおよび検出難易度の多様性の観点で整理	異なる偽造手法や難易度に対する統合判定技術の技術的適用範囲を評価
データ特性	解像度、圧縮方式、ファイル形式等の違いを考慮	データ特性の違いによる検出性能への影響を確認
メディア種別	画像、動画、音声のデータを選定	モダリティごとの検出特性および統合効果を評価
高度偽造コンテンツ	汎用生成AIモデルにより生成されたコンテンツを対象に含める	従来型ディープフェイクに加え、最新生成技術への適用可能性を検証
想定ユースケース	報道、企業風評、災害情報、政治・社会情報を想定	実社会における適用範囲および運用上の課題を確認

4-2. 検証及び調査の個別詳細

実証テスト計画の策定・実施　検証データセットの選定に関して

実証テストの概要

- 本実証テストにおいては、統合判定技術の有効性の実証を目的として、3つの検出ツールを用いた実証テストを実施した。

検証データ構成

- 検証用テストデータとしては、動画64件、画像32件、音声4件をピックアップした。この構成は、近年流通しているディープフェイク関連コンテンツの大半が動画・画像を中心としている実態を踏まえ、遭遇頻度の高いモダリティに重点を置いた評価を行うことを目的として設定したものである。

検証データ構成

- 動画については、生成AIによる映像生成に加え、テロップの付与や編集・加工を伴うコンテンツが多様に存在し、検出難易度やツール特性の差異が顕在化しやすいことから、十分な検証が行えるよう件数を多く設定した。画像についても、静的コンテンツとして流通量が多く、加工手法の多様性が高いことを考慮し、動画に次ぐ件数を確保した。

音声データの位置付け

- 一方、音声は、ディープフェイクとしての流通事例や社会的影響が現時点では動画・画像と比較して限定的であること、また検出手法や評価指標が発展途上である点を踏まえ、本実証では補完的な位置付けとして少数のサンプルにより傾向把握を行う方針とした。

ユースケース別検証の考え方

- さらに、各データはユースケースごとに整理し、幅広い社会的ケースへ対応できるかどうかについても併せて検証した。なお、本実証ではユースケースごとのデータ数が十分ではないため、AUC等の指標を用いた定量的な性能比較は行わず、モックを用いた適用性の確認を中心に評価を行った。

4-2. 検証及び調査の個別詳細

実証テスト計画の策定・実施　判定の統合に関して

統合方式の設計

- 統合方式として Simple統合 と Weighted統合 の2方式を採用した。Simple統合は各検出ツールの結果を単純平均する方式であり、ツールの判定結果を均等に扱う構成とした。Weighted統合では、評価指標に基づき重みを調整し、統合スコアを算出する方式とした。

Weighted統合における評価指標

- Weighted統合の評価指標として、統合性能を把握するため AUC と PRAUC を用いた。AUCはROC曲線の面積であり、閾値に依存せず判定性能を比較できる指標である。PRAUCはPR曲線の面積であり、適合率と再現率の関係から見逃しと誤検知のバランスを評価する指標である。

複数指標による重み調整

- ユーザニーズに応じた重み調整を可能とするため、再現率・適合率・F1スコアも参考指標として取り入れた。これにより、単一指標に依存せず、複数の観点から統合方式の設計およびチューニングを行える構成とした。

4-2. 検証及び調査の個別詳細

実証テスト計画の策定・実施　判定の統合に関して

判定結果の取得および統合方法

- ツール間の連携はAPIによる自動統合ではなく、各ツールの判定結果を個別に取得し、机上で統合処理を実施した。これにより、API仕様や実装差異などのシステム要因が検証結果に影響することを避け、判定ロジックおよび統合方式に着目した検証環境を構築した。

検証環境の再現性確保

- 検証の再現性を確保するため、使用したデータセットおよびツールのバージョンを記録し、検証環境のトレーサビリティを確保した。また、検証手順を整理することで、追加検証や再評価が可能な実証基盤を整備した。

実証環境の整備

- 以上の取組により、複数ツールを前提とした統合判定方式の実証テストを実施するための検証環境を構築し、今後の技術検討および実運用に向けた評価手法を整理した。

4-2. 検証及び調査の個別詳細

（参考）検証用テストデータの詳細

表4.2.1 データセットごとのデータ数

データソース	F/R	動画	画像	音声	合計
データソースA	Fake	8	4	2	14
	Real	8	4	2	14
データソースB	Fake	4	4	-	8
	Real	8	4	-	12
データソースC	Fake	2	4	-	6
	Real	8	4	-	12
データソースD	Fake	2	4	-	6
	Real	8	4	-	12
データソースE	Fake	16	-	-	16
	Real	-	-	-	-
Fake計		32	16	2	50
Real計		32	16	2	50
モーダル合計		64	32	4	100

表4.2.2 ユースケースごとのデータ数

ユースケース	F/R	動画	画像	音声	合計
① メディア・報道関連	Fake	8	4	2	14
	Real	8	4	2	14
② 企業・組織の風評被害対策	Fake	8	4	-	12
	Real	8	4	-	12
③ 公的機関の情報環境保全	Fake	8	4	-	12
	Real	8	4	-	12
④ プラットフォーム事業者の技術基盤強化	Fake	8	4	-	12
	Real	8	4	-	12
Fake計		32	16	2	50
Real計		32	16	2	50
モーダル合計		64	32	4	100

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施に関する総評　統合判定の結果に関して

- 統合判定については、複数ツールの判定結果を用いた統合方式を採用し、単一ツール判定と比較することで、ツール間の相互補完効果を確認した。その結果、統合により各種指標において性能が改善する傾向が認められ、統合判定技術の有効性を定量的に示すことができた。
- 画像および動画の双方において、閾値に依存しない全体的な判定性能を示すAUCと、見逃しと誤検知のバランスを示すPRAUCのいずれについても、単一ツールより高い値となる重み付け条件が確認された。このことから、統合判定により閾値設定に依存しない全体的な判定性能が向上することが確認された。
- モダリティごとに検出特性や得意・不得意が異なるため、単純な統合ではなく、各ツールおよび各モダリティの特性を考慮した統合設計が重要であることが明らかとなった。特に、単一ツールの性能が著しく低い場合には、統合性能に悪影響を及ぼす可能性があるため、統合対象の選定や重み付けの調整が必要である。
- 統合方式の設計においては、評価指標の重み付けを調整することで、過検知低減を重視する設定や、見逃し低減を重視する設定など、利用目的や運用ニーズに応じた判定特性の調整が可能であることを確認した。これにより、利用シーンに応じた柔軟な運用が可能であることが示された。
- 一方で、統合判定において得られる性能指標上の最適性と、ユーザが重視する運用上のニーズ（過検知許容度や見逃し許容度）を、どの程度の比重で反映させるべきかについては、明確な基準を定めるには至っていない。
  このため、統合結果の判定性能とユーザニーズとのバランスの取り方については、今後の検討課題である。

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施　モダリティによる違い

- モダリティごとの単体判定および統合方式による性能比較を、画像および動画それぞれの特性を踏まえ、複数の評価指標に基づく定量的比較を行った。単独最良ツールを単独ツールの中で最もよい精度を示したものと定義した。
- 画像においては、精度、適合率、再現率、F1スコアのすべての評価指標において、Weighted統合が最も高い性能を示し、単独性能が最も高いツールBと比較しすべての指標で5%以上の性能の向上が見られ、特に誤検知率は50%低減しており、画像における偽造表現の特徴がツールごとに補完的であり、重み付けによる統合が有効に機能したことを示唆している。
- 動画においては、表4.2.4に示す通り、見逃し低減を重視した条件では、Weighted統合により精度、再現率およびF1スコアが最良単独ツールを上回る結果となり、統合による性能向上が確認された。一方で、誤検知率については最良単独ツールと同程度であった。一方、過検知抑制を重視した条件では、単一ツールが高い性能を示すケースが見られ、統合による改善効果は限定的であった。これは、動画は生成AIによる映像生成に加え、テロップの付与や切り貼りといった編集・加工が施されるなど、複数の加工形態を含んでいる点が影響していると考える。
- こうした多様な表現要素を含む動画に対しても、各ツールはそれぞれ異なる観点で判定を行っており、統合判定による優位性は確認された一方で、ツール毎の加工や編集の定義の揺らぎにより、評価対象や判定ロジックの差が統合結果に影響を与える場合があることも明らかとなった。
- このことから、動画モダリティにおいては、単純な統合方式を適用するのではなく、コンテンツに含まれる加工内容や特徴に応じて、より細かな分類を行った上で統合判定を行うことが重要であると考えられる。

表4.2.3　単体・統合方式の性能比較（画像）

判定方式	精度	適合率	再現率	F1スコア	誤検知率
ツールA	0.875	0.800	1.000	0.889	0.250
ツールB	0.906	0.882	0.938	0.909	0.125
ツールC	0.625	0.583	0.875	0.700	0.625
Simple統合	0.875	0.929	0.813	0.867	0.0625
Weighted統合	0.969	0.929	1.00	0.967	0.0625
Weighted統合 – Simple統合比	0.107	0.929	0.231	0.116	0.000
Weighted統合 – 最良単独ツール比	0.070	0.929	0.066	0.064	-0.500

表4.2.4単体・統合方式の性能比較（動画）

判定方式	精度	適合率	再現率	F1スコア	誤検知率
ツールA	0.703	0.697	0.719	0.708	0.313
ツールB	0.828	0.839	0.813	0.825	0.156
ツールC	0.500	0.500	1.000	0.667	1.000
Simple統合	0.500	0.500	1.000	0.667	1.000
Weighted統合	0.891	0.857	0.938	0.896	0.156
Weighted統合 – Simple統合比	0.781	0.714	-0.063	0.343	-0.844
Weighted統合 – 最良単独ツール比	0.075	0.022	0.154	0.085	0.000

※本表の比較はSimple統合あるいは最良単独ツールを基準とし、Weighted統合との相対比較結果を示しており、値は基準値に対する増減率を表す。

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施　動画の統合結果に対する考察

- 動画において、過検知抑制を重視した条件では、Weighted統合は最良ツールと比較して一部の評価指標において性能の低下が確認された。
- これは主に、各ツールの検知対象および判定観点の違いにより、統合処理において判定結果が相互に希釈されたためと考えられる。具体的には以下の３点が影響した。
  - 高度偽造コンテンツの判定精度差：高度な生成AIによる偽造動画に対しては、ツールAが高い検知性能を示した一方、その他のツールでは検知精度が十分でないことが確認された。そのため、単一ツールでは正しく検知できていたケースにおいても、各ツールの判定結果が相互に影響し、場合によっては判定結果が大きく異なり、真偽の判定が逆転するケースも確認された。
  - 加工や編集の定義の揺らぎ：生成AIによる映像生成だけでなく、テロップ付与、切り貼りなどの編集加工が含まれる場合がある。これらの加工や編集の扱いについてはツールごとに定義が異なるため、例えば、ディープフェイク生成技術そのものの痕跡検知を主目的とするツールでは「真正」と判定される一方、編集の有無やその意図まで評価するツールでは「改変あり」と判定されるなど、同一動画に対して判定結果が大きく異なるケースが確認された。
  - コンテキストの評価の精度：ツールBは、発言内容や文脈を踏まえたコンテキスト評価に強みを持ち、内容の妥当性を評価する機能を備えている。一方で、その他のツールは主に映像や音声の特徴量を基に判定を行うため、文脈情報を含めた評価は限定的である。このような判定観点の違いにより、同一コンテンツに対して異なる判定が出るケースがあった。
- 以上３点のような要因により、あるケースでは得意分野を持つツールの結果が他ツールの結果によって希釈され、また別のケースではツールごとの定義差により判定結果が分散するといった状況が発生し、統合結果の評価指標が低下するケースが確認された。
- これらの要因は特定できているため、今後は以下の対応を行うことで、統合の改善を図る。
  - 高度偽造コンテンツ検知に高い性能を持つツールが、特定のモデルにより高信頼度で偽造と判定した場合には、そのカテゴリにおいて高い性能を有するツールの結果を優先して統合する。
  - 動画加工や編集の定義を整理し、評価対象の統一を行う。
  - コンテキスト評価が有効なケースでは、文脈評価を行うツールの結果を中心に統合する。

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施　モダリティによる違い

- 音声モダリティについては、検証対象とした3社のうち、1社が音声データに対応しておらず、当該ツールは評価対象外とした。残る2社の検出ツールにより4件の音声データを評価した結果、1社は4件中2件を正しく判定し、もう1社は4件中3件を正しく判定した。
- 音声データの件数が限定的であるため、定量的な性能評価として一般化することは困難であるが、ツール間で判定精度に差が生じる傾向が確認された。
- 本実証において音声は補完的な位置付けでの評価に留めており、今後は検証データの拡充および評価手法や評価の検討が必要である。

表4.2.5　音声に対する各ツールの真偽判定結果

Case		ツールA	ツールB	ツールC
項番	本当の真偽			
1	Fake	Real	Real	未対応
2	Fake	Fake	Real	未対応
3	Real	Real	Real	未対応
4	Real	Real	Real	未対応

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施　高度偽造コンテンツの検知

- 高度偽造コンテンツに対する検知性能を評価するため、誤検知率を低く抑える高閾値条件で評価を実施した。
- 高度偽造コンテンツに対する単体判定の結果、検出性能にはツール間で大きな差異が確認された。具体的には、ツールAの精度は93.75%と高い水準を示した一方で、ツールBは68.75%、ツールCは37.5%にとどまった。この結果から、高度偽造コンテンツに対する対応力は、ツールごとに大きく異なることが明らかとなった。
- 特に、ツールAが高い精度を示した要因については、完全合成生成コンテンツの検出に特化したモデルと、部分的な改変を検出するモデルの双方を利用していることが確認された。これにより、シーン全体が生成されたコンテンツと、既存素材に対して一部改変が施されたコンテンツの双方に対応可能な構成となっており、幅広い高度偽造表現に対して安定した判定性能を発揮したものと考えられる。
- 高度偽造コンテンツに対しては、単一の判定手法やツールに依存した運用では見逃しや誤判定が生じるリスクが高いことが確認された。生成AI技術の進展に伴い、この種の高度偽造コンテンツは今後さらに増加することが想定されるため、ツールごとの得意・不得意を踏まえた統合判定や、多角的な評価アプローチの必要性が示唆される。

表4.2.5　高度偽造コンテンツに対する各ツールの真偽判定結果

Case 項番	ツールA	ツールB	ツールC
1	Fake	Fake	Real
2	Fake	Fake	Real
3	Fake	Real	Fake
4	Fake	Fake	Real
5	Fake	Real	Fake
6	Fake	Fake	Real
7	Real	Fake	Real
8	Fake	Fake	Real
9	Fake	Fake	Real
10	Fake	Real	Fake
11	Fake	Fake	Real
12	Fake	Fake	Fake
13	Fake	Real	Real
14	Fake	Real	Real
15	Fake	Fake	Fake
16	Fake	Fake	Fake
精度	93.75	68.75	37.5

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施に関する総評　統合技術の実用的価値と信頼性に関する考察

公開情報を参照する判定ツール

- 一部のツールでは、公的機関や事業者の公式サイト、注意喚起情報などのインターネット上の公開情報を参照し、既に流通している真贋情報を評価する仕組みが用いられている。これらのツールは、参照元情報を明示することで、利用者が判定根拠を確認しやすい特徴を持つ。

公開情報を用いた真贋確認

- 例えば、インタビュー映像や発言については、公式に公開されている情報源を参照することで正規情報であることを示すことができる。また、ファクトチェック機関が公開する情報を基に、偽・誤情報として指摘されているコンテンツの根拠を提示することも可能である。

前処理としての活用

- このような公開情報の参照を前処理として活用することで、真贋が既に明らかなコンテンツを統合処理の前段階で識別・分類できる。これにより、統合処理の対象を絞り込み、処理負荷の低減やレスポンス時間の短縮が期待できるとともに、判定結果の安定性向上にも寄与する構成とした。

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施に関する総評　統合技術の実用的価値と信頼性に関する考察

コンテキスト評価の概要

- 一部のツールでは、発言内容が当事者の立場や過去の言動と整合しているか、また発言の文脈が保たれているかといった観点から、悪意の有無や利用文脈を考慮したコンテキスト評価を行う仕組みが用いられている。

発言内容と文脈の整合性評価

- 例えば、映像中の人物が実在の政治家で映像自体が真正であっても、発言内容が公式方針や過去の発言と大きく異なる場合には、誤解を招く可能性のある表現として評価対象となる。また、発言の一部のみを切り取る編集や、内容と異なる意味を持つテロップ・字幕の付与についても、文脈を歪める表現として評価される。

文脈欠如による誤認リスク

- 記者会見やインタビュー映像では、発言の前後が省略されることで本来の趣旨とは異なる印象を与える場合や、発言時の条件や前提が欠落する場合がある。このようなケースでは、映像や音声の真正性だけでは判断が難しく、発言者の立場や意図を踏まえた多面的な評価が必要となる。

4-2. 検証及び調査の個別詳細

定量的・定性的評価の実施に関する総評　統合技術の実用的価値と信頼性に関する考察

コンテキスト評価の意義

- コンテキスト評価では、映像や音声の真正性だけでなく、「誰が・どのような立場で・どのような意図で発信したか」といった背景を踏まえて判断を行う。これにより、コンテンツの潜在的なリスクや注意点を理解しやすくし、判定結果への納得感を高める効果が期待できる。

AI判定技術を補完する役割

- 発言の切り取りやテロップによる印象操作などの編集・加工は、生成AIの検出技術のみでは把握が難しい場合がある。コンテキスト評価はこうしたケースを補完し、情報の信頼性判断を実務的に支援する役割を担う。

実用化に向けた知見の整理

- 本取組では、統合判定技術の有効性に加え、ツール特性を踏まえた活用方法や社会実装に向けた課題を整理した。これにより、ユーザ理解や説明性を重視した評価の重要性を示すとともに、今後の技術発展に向けた基盤となる知見を得た。

4-2. 検証及び調査の個別詳細

モックの作成と利用模擬検証

プロトタイプ（モック）の作成

- 本開発・実証期間では、統合判定技術を用いたサービス提供時の画面構成や操作フローを確認するため、実証用プロトタイプアプリ（モック）を作成した。

検証対象とした機能範囲

- 本モックは、外部ツール連携や課金ロジックなどの複雑な管理機能を除き、「画面とデータ表示の成立性」および「ユーザの機能認知」を確認するプロトタイプとして位置付けた。ユーザが画面操作を通じて、機能の動作、情報が得られるタイミング、操作と結果の関係を直感的に理解できるかを確認した。

想定ユースケースによる検証

- 作成したモックを用い、メディア報道前確認、企業風評監視、災害時情報検証、選挙期間中の政治情報検証といった情報信頼性判断に関するユースケースを想定した疑似検証を実施した。

業務フローを想定した机上検証

- 各ユースケースでは実際のユーザ操作は行わず、想定業務フローを机上で再現し、判定結果の表示内容、根拠情報の提示、注意喚起表示などが利用シーンに適した形で提示されるかを確認した。

サービス運用時のリスク表示

- 判定精度の限界やプライバシー・著作権に関する注意点など、サービスの限界や潜在的リスクを画面上で認知させる設計についても確認し、実サービス運用時の期待値調整に資する構成であることを整理した。

4-2. 検証及び調査の個別詳細

モックのUIおよび機能　メディア分析

① メディアアップロード画面

- ユーザが分析対象となるコンテンツ（画像・動画・音声など）をアップロードするための画面として設計した。
- モックでは、ファイルアップロードを起点とした疑似的な分析フローを再現しており、アップロード後に分析画面へ遷移する一連の操作を確認できる構成とした。

- 本画面は、ユーザが「どのようなコンテンツを分析対象として扱えるか」を直感的に理解することを目的としている。
- 実際のサービスリリースにおいては、ファイルアップロードに加え、URL指定によるコンテンツ分析も可能とする想定であり、その拡張性を見据えたUI構成としている。
- 本モックでは、実ファイル処理や外部解析処理は行わず、以降の分析結果表示は事前に登録したデータを参照する疑似動作としている。

図4.2.2 メディアアップロード画面

4-2. 検証及び調査の個別詳細

モックのUIおよび機能　メディア分析

② メディア分析選択画面

- アップロード後に表示される分析条件選択画面として、分析モードおよびコンテンツの取扱い条件を指定できるUIを実装した。

- 分析モードについては、「見逃し低減」「過検知抑制」を選択できる構成とし、統合判定の特性をユーザが意識的に切り替えられる設計とした。本画面は、ユーザが想定するユースケースに応じて分析モードを選択できる構成とすることで、「どのような目的で分析を行うか」によって、適切な判定特性が異なることを理解させることを狙いとしている。

- あわせて、判定対象コンテンツの著作権が「自身に帰属するもの」か「企業に帰属するもの」かを選択できる項目を設けた。これにより、単に分析を実行するのではなく、ユーザが分析行為に伴う前提条件やリスクを認識した上で操作を行うことを促すUI構成としている。

- 本画面は、ユーザに対して「分析設定の選択が判定結果に影響を与えること」および「コンテンツの取扱いに注意が必要であること」を理解させることを狙いとしている。

図4.2.3 メディア分析選択画面

4-2. 検証及び調査の個別詳細

モックのUIおよび機能　メディア分析

③ 著作権に関する注意喚起ポップアップ

- メディア分析選択画面において、著作権が企業に帰属するコンテンツを選択した場合、著作権に関する注意喚起をポップアップ表示する機能を実装した。
- 本ポップアップでは、分析対象が著作物に該当する可能性があること、および分析行為自体が利用条件や契約内容によっては留意を要する点を明示する構成とした。
- これにより、ユーザが無意識にリスクのある操作を行うことを防ぎ、判定精度だけでなく、法的・運用的観点も含めた注意喚起を行うUI設計であることを確認した。本機能は、サービスの限界点やリスクを画面上で認知させ、実運用時の期待値調整や責任分界の明確化に寄与することを狙いとしている。

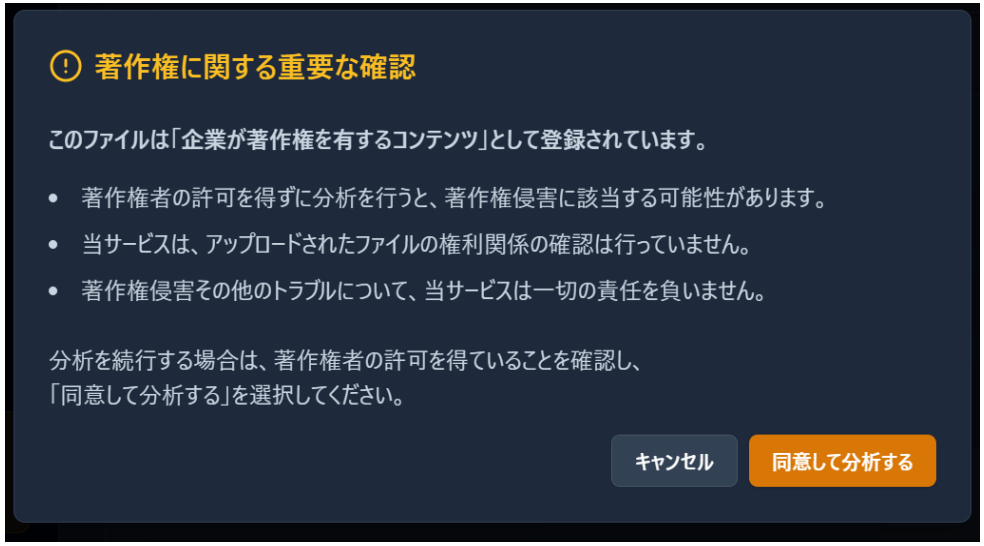


図4.2.4　著作権に関する注意喚起に関するポップアップ

4-2. 検証及び調査の個別詳細

モックのUIおよび機能　メディア分析結果

- 本画面は、メディア分析の結果をユーザに提示し、統合判定の考え方および各ツールの判定結果の関係性を理解させることを目的とした結果表示画面として設計した。
- 複数の分析ツールによる判定結果を統合した最終的な判定結果を画面上部に表示する構成とした。あわせて、Fakeスコアを数値および視覚的な指標として表示し、判定結果の強度や傾向を直感的に把握できるよう配慮した。
- メディアに関する基本情報を一覧で表示する構成とした。これにより、ユーザが「どのコンテンツを、いつ、どの条件で分析したか」を即座に確認できるようにした。分析結果と対象コンテンツの紐付けを明確にし、誤認や取違えを防止する効果を狙いとしている。
- 統合判定結果の下に、各分析ツールごとの判定結果をサマリ形式で表示する構成とした。各ツールの判定傾向や結果の違いを一覧で確認できるようにすることで、単一ツールではなく複数ツールを用いる意義を可視化した。

- 統合判定結果の下に、各分析ツールごとの判定結果をサマリ形式で表示する構成とした。各ツールの判定傾向や結果の違いを一覧で確認できるようにすることで、単一ツールではなく複数ツールを用いる意義を可視化した。
- 統合判定が個別結果を基に算出されていることを把握できる設計とし、結果の透明性を高めることを狙いとしている。
- 各ツールの分析結果について、より詳細な情報を確認できる詳細分析画面へのリンクを設けた。本画面から詳細画面へ遷移することで、ユーザが判定結果の根拠や内訳を確認し、自身の判断を補完できることを狙いとしている。

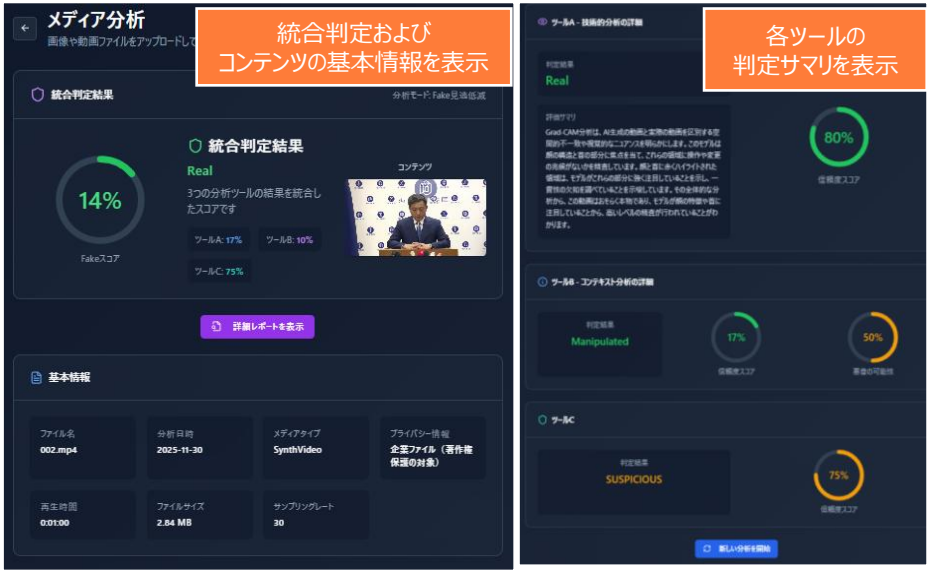


図4.2.5　メディア分析結果画面

4-2. 検証及び調査の個別詳細

モックのUIおよび機能　詳細分析レポート

- 詳細分析画面は、メディア分析結果画面から遷移し、各分析ツールの判定結果を個別に確認するための画面として設計した。統合判定結果のみでは把握しきれない判定の内訳や傾向を確認し、判定結果の根拠を理解させることを主な目的としている。

- 詳細分析画面では、検出ツールごとにタブを切り替えるUI構成を採用した。ユーザは、同一コンテンツに対するツールごとの判定結果を、同一画面内で比較しながら確認できる構成となっている。
-
- ツールごとの得意・不得意や判定傾向を把握できるため、統合判定結果を鵜呑みせず、参考情報として適切に活用する姿勢を促す効果がある。

- 本画面を通じて、最終的な判断はユーザ自身が行うものであるという位置付けを明確にし、サービスの限界点や責任分界を認識させる狙いがある。

図4.2.6　メディア詳細分析結果画面

4-2. 検証及び調査の個別詳細

ユースケース別の有効性検証　情報信頼性判断ユースケースにおける有効性検証

① 報道機関における利用シーンと有効性検証

- 本検証では、報道機関が記事や映像を公開する前に、その内容が正しいかを確認する場面を想定した。
- 偽・誤情報を公開してしまうことは報道機関の信頼性低下につながるため、「見逃しをできるだけ減らす」ことを重視した判定が重要となる。そのため、本ユースケースでは、複数の判定結果を統合する際に、偽情報を検出しやすい方向に判定閾値を設定・調整することが有効であると評価した。
- また、インターネット上の公開情報を参照し、「どの情報を根拠に判断したのか」を利用者に明示できる機能は、高い説明責任が求められる報道機関にとって有用であると評価した。

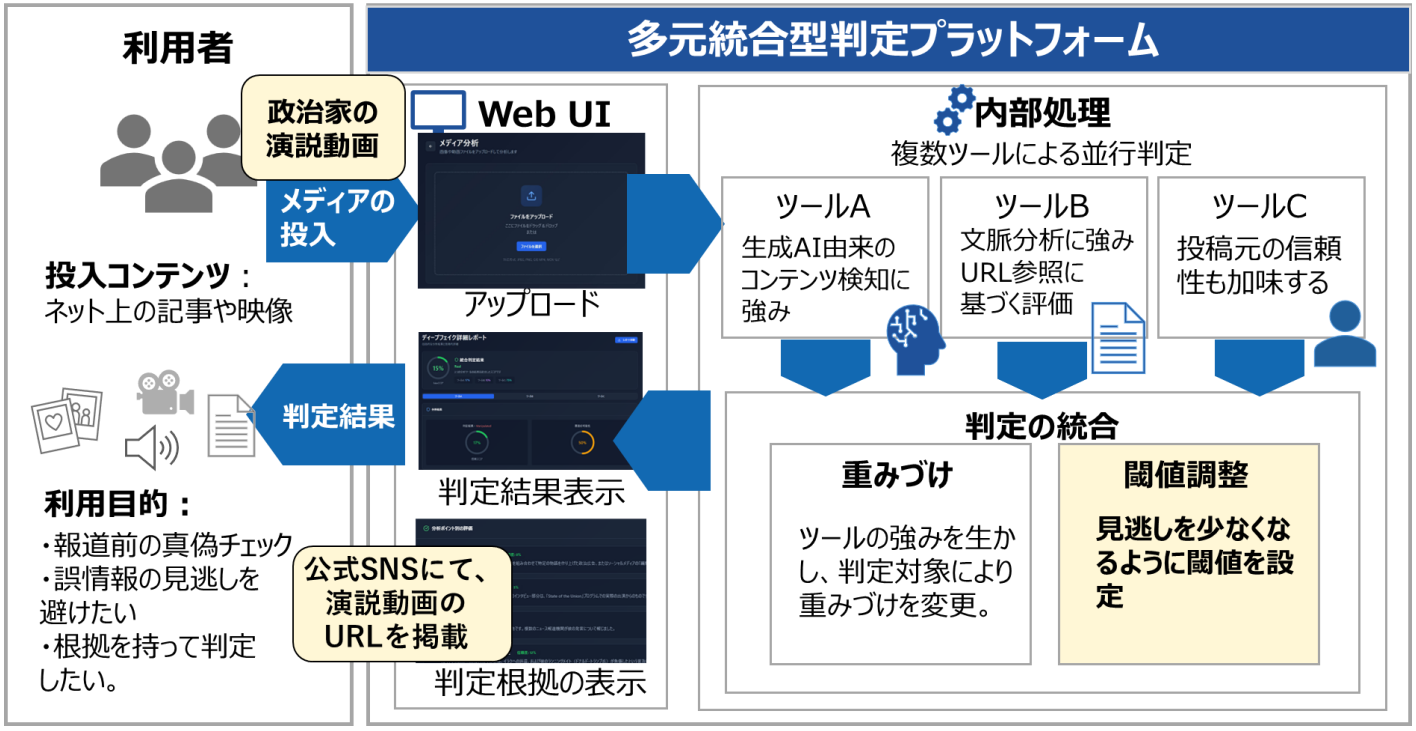


図4.2.7　報道機関における利用シーンと統合判定フロー

4-2. 検証及び調査の個別詳細

ユースケース別の有効性検証　情報信頼性判断ユースケースにおける有効性検証

② 企業における利用シーンと有効性検証

- 本検証では、企業に関する風評被害や誤情報の検出を目的とし、自社や自社製品に関するコンテンツを継続的に確認する利用シーンを想定した。
- モックでは、判定対象となるコンテンツが著作物に該当するかを確認を促す機能を提供しており、判定行為そのものが著作権侵害となるリスクの低減に寄与すると考えられる。一方で、正規のコンテンツであっても、意図しない形で切り抜きや加工が施される可能性が想定されるため、こうしたケースにも対応可能な仕組みの構築が重要であると考えられる。

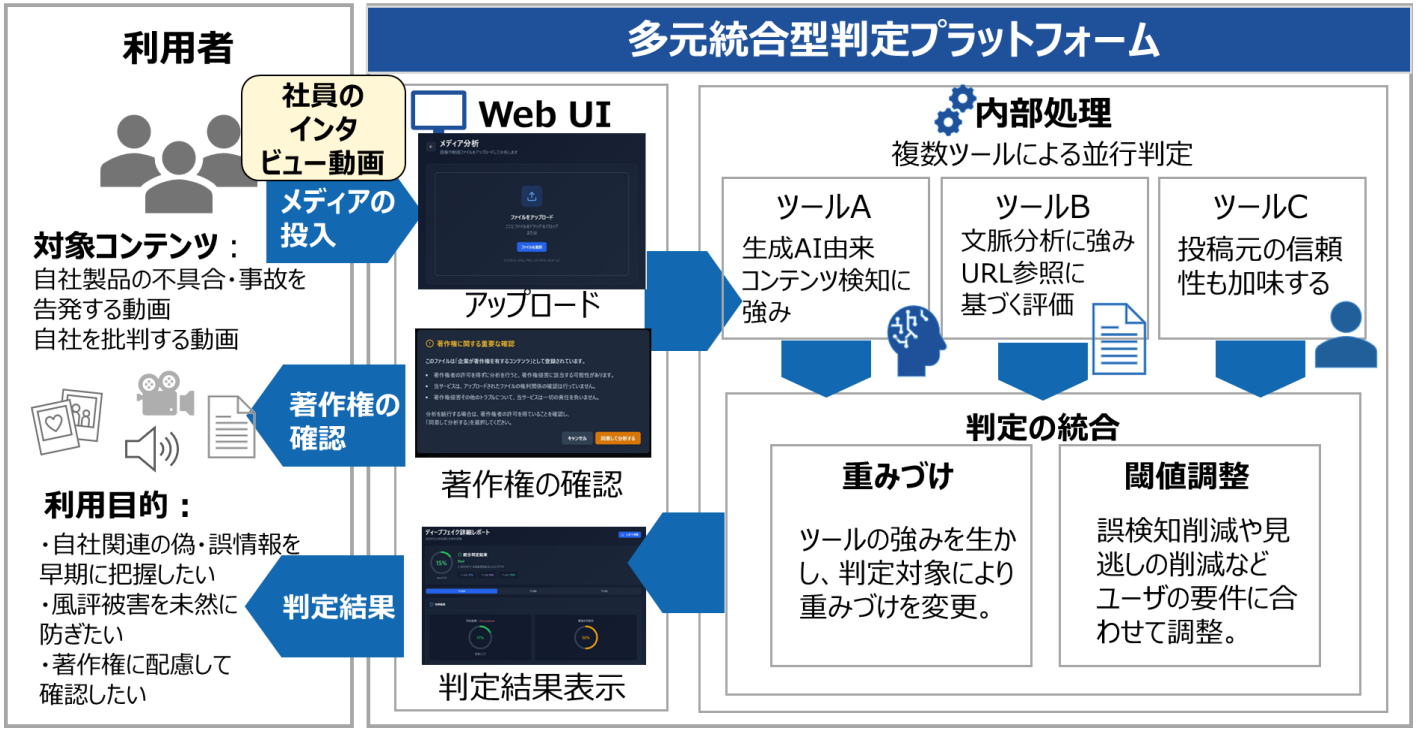


図4.2.8　企業における利用シーンと統合判定フロー

4-2. 検証及び調査の個別詳細

ユースケース別の有効性検証　情報信頼性判断ユースケースにおける有効性検証

③ 災害発生時における利用シーンと有効性検証

- 本検証では、災害発生時に流通するSNS上の情報の真偽確認を想定した。
- 災害時においては、誤情報の拡散が社会的混乱や二次被害につながる恐れがある一方で、過検知により本来拡散されるべき重要な情報が十分に共有されない可能性も考えられる。そのため、誤検知の低減と見逃しの低減の両立を意識したチューニングが重要であると考えられる。また、情報の拡散元を検査し、情報の信頼性を検証することも有効であると考えられる。

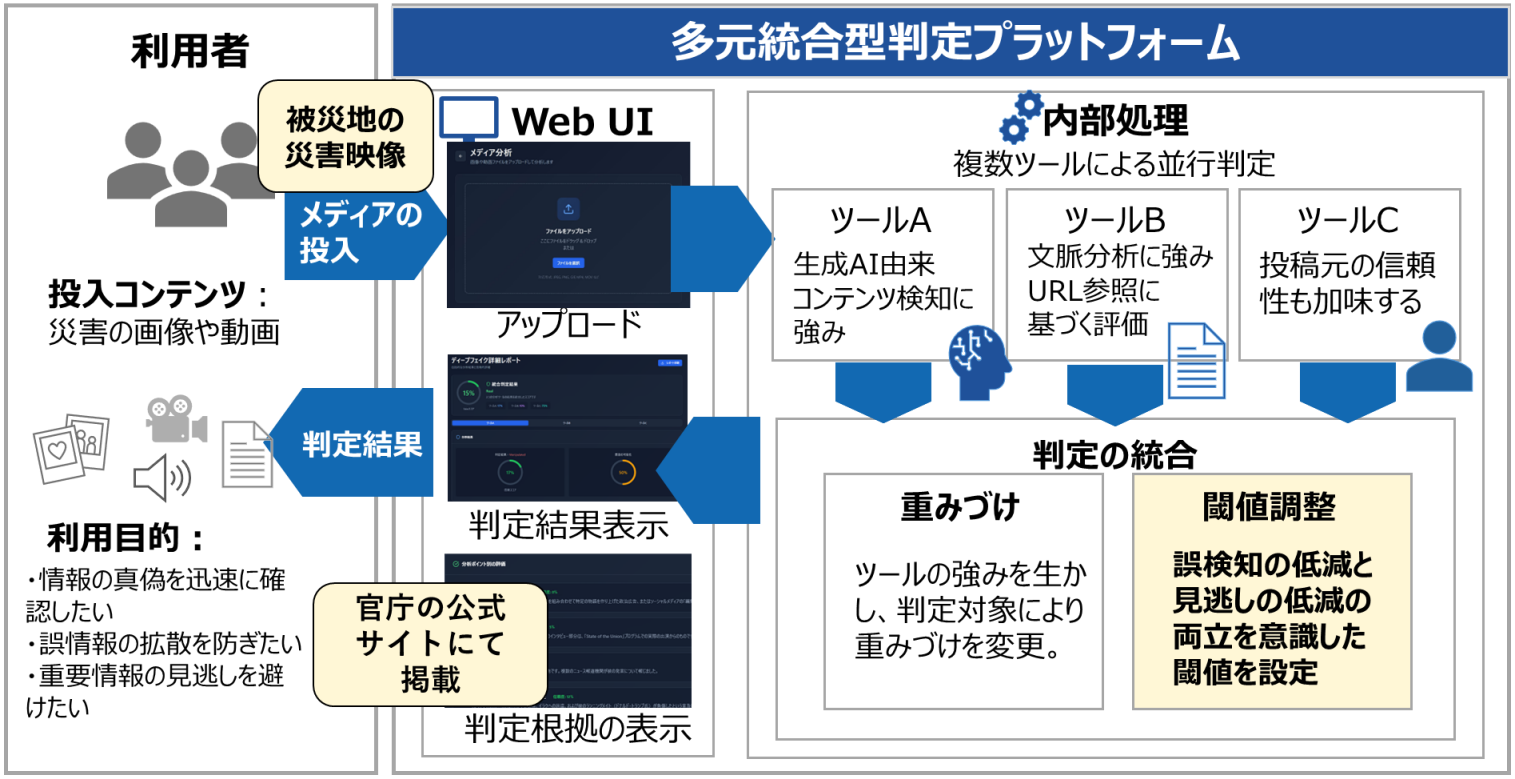


図4.2.9　災害発生時における利用シーンと統合判定フロー

4-2. 検証及び調査の個別詳細

ユースケース別の有効性検証　情報信頼性判断ユースケースにおける有効性検証

④ 選挙期間中における利用シーンと有効性検証

- 選挙期間中に流通する政治関連情報を対象とし、選挙期間中の政治情報検証ユースケースを想定した。
- 選挙期間中の情報については、社会的影響が大きく、誤った情報の拡散が民主的な意思決定に的な意思決定に影響を与える可能性があることから、特に慎重な情報信頼性判断が求められる利用シーンである。一方で、内容の真偽だけでなく、文脈や切り取りによる印象操作の可能性も考慮する必要があり、コンテンツの構成や利用意図を踏まえた多面的な評価の重要性が確認された。そのため、コンテキストの観点を含めた評価が重要な要素となると評価した。

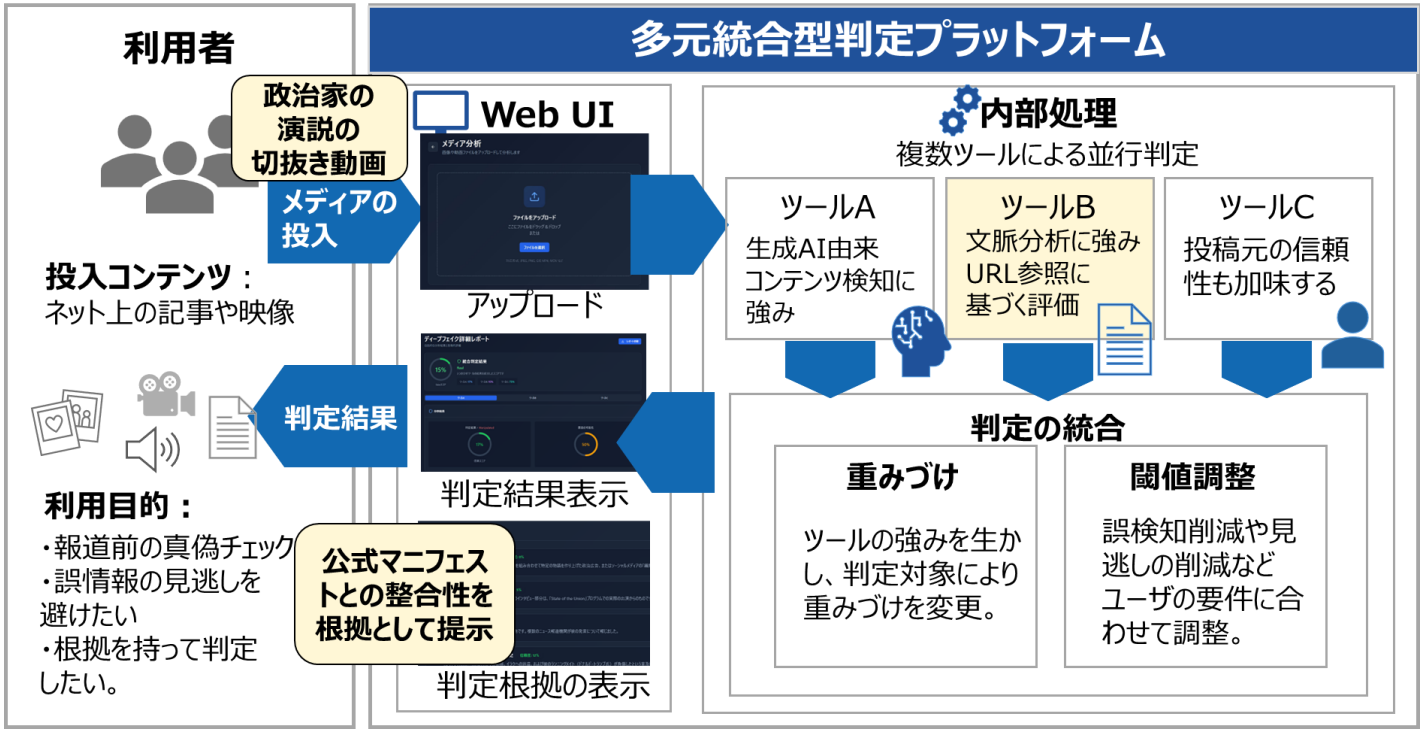


図4.2.10　選挙期間中における利用シーンと統合判定フロー

目次

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

取組と成果の全体像

アクション	結果 / 気づき
公共・企業へのヒアリング調査	「判断支援」への課題や期待効果の確認
持続可能なビジネスモデルの構築検討	実用サービス展開の妥当性の確認
国内展開戦略・市場性の基本評価	他団体との連携と展開ロードマップの想定
多元統合型判定による検知精度の実証	高度偽造コンテンツに対する有効性
事業化における経済価値の特定	監視工数削減等の具体的ROIの可視化

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

- 公共機関や企業へのヒアリングを通じ、公共への適応性や活用可能性を把握するとともに懸念点解消に向けて下記の基本課題の調査に取り組んだ。
  - 現行の偽・誤情報対策プロセスや課題把握
  - 導入障壁や懸念事項(コスト/運用負荷/法的懸念等)の明確化
  - 最初に実装すべきユースケースや期待効果の確認
- 持続可能なビジネスモデルを構築するために、実用的なサービスの展開を見据えた取組を行った。
  - 仮説設定した本技術の導入フェーズ別のビジネスモデルの妥当性検討
    (顧客視点での価格受容性や社会への普及等の見極め)
- 国内展開市場を狙っていくための取組として、展開戦略の基本検討・市場性/事業性の基本評価を行った。
  - 社会実装の加速を目的とした他事業者とのパートナーシップや協業の検討
    （他団体、企業との連携としてコンソーシアムの参加の必要性を認識した）
  - 自社視点での持続可能な事業(ビジネス)として社会への普及の時間軸と収益性評価(キャズムの克服検討)
- 検知精度の有意な向上
  - 統合判定技術の導入により、根拠を示した判定支援の必要性を確認した。特に生成AIで作成したディープフェイク等の高度な偽造コンテンツに対する頑健性が向上した。
- 事業化における具体的な経済価値の特定
  - ターゲット層（大手IPホルダー等）において、現状年間3,000万円程度発生している監視・法務コストを本サービスで代替・効率化できるという具体的なROI（投資対効果）を算出した。
  - 「3,000万円規模の投資判断（取締役決裁）」が可能な市場ニーズが存在することを特定し、売上具体化に向けた確度の高いビジネスモデルを構築できた。

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

実施した主な内容

- 市場ニーズの深掘りと社会実装モデルの策定。
- 有識者・担当者ヒアリング（報道、プラットフォーマー、タレントマネジメント、ブランド保護、行政機関、教育機関等）を実施し、実務における具体的なペインポイント（人件費、風評被害、ブランド毀損、法務コスト）を抽出した。
- ヒアリング結果により、検知後のワークフロー（法務・警察連携）までを考慮したサービスの需要を検討する。
- 課題やニーズを引き出す為、実施計画書では想定していなかったモックを作成した。
- ツール提供企業3社を対象に、既存および想定されるユースケース、ならびに実装（ビジネス）モデルに関するヒアリングを実施した。

- 市場ニーズの深掘りおよび実務における具体的なペインポイントを正確に抽出するため、有識者・実務担当者へのヒアリングを実施した。その際、広報の建前を意識した回答に留まらない「本音」の情報を収集することを重視し、組織名を非公開とする条件で計20団体以上からの協力を得ている。

エンターテインメント/IPホルダー	大手広告代理店、芸能事務所、ゲーム・玩具、金融・保険、総合商社
官公庁/地方自治体	中央省庁、防衛関連、地方自治体、法執行機関、教育・研究機関
報道機関/メディア	公共放送局、ネットニュース配信事業者、全国紙新聞社、専門雑誌社
SNS/オンラインプラットフォーム運営者	システムインテグレータ、ソーシャルリスニング事業者、ネットワークインフラ

5-2. 社会実装に向けた取組の個別詳細

ターゲットへのヒアリングの詳細

経営層のペインポイント（CEOなりすましリスク）

- 検証1　広報リスクから「経営リスク」へ（CEOなりすまし被害の甚大化）

- 検証内容
  - 現状の課題
    従来は「炎上対策」という広報レベルの課題認識であったが、有識者ヒアリングを通じ、生成AIによる「経営者（CEO）のなりすまし動画」が、企業の存続に関わる重大な経営リスクに変貌していることが確認された。
  - 具体的な脅威
    CEOの偽動画による「虚偽の決算発表」や「架空の投資案件への誘導（投資詐欺）」は、真偽が判明するまでの数時間で株価乱高下を引き起こし、億単位の時価総額毀損を招く恐れがある。
    結論として本システムのターゲットは、現場担当者だけでなく、「経営層」および「IR部門」であり、導入予算は広報費ではなく役員決裁の危機管理費から拠出されるべき性質のものと予想される。

5-2. 社会実装に向けた取組の個別詳細

ターゲットへのヒアリングの詳細

日本特有の「反論の心理的障壁」

- 検証2　日本企業特有の「沈黙」を生む心理的障壁と解決策

- 検証内容
  - 現状の課題（広告代理店）
    欧米企業と比較し、日本企業は「不確実な状態での反論」を極端に避ける傾向がある。明らかな偽・誤情報であっても、「万が一、事実だったらどうするのか」「客観的な証拠はあるのか」という社内確認に時間を要し、結果として初動が遅れる（あるいは沈黙してしまう）。
  - 判定スコアの役割（エビデンス）
    現場が求めているのは、単なるアラートではなく、「AI判定スコア：偽誤確率 100%」という客観的な数値である。
    この数値が説明責任の担保となり、広報担当者が自信を持って「これはフェイクです」と否定するための「心理的な許可証」として機能することを期待していることを確認した。
    公式サイトや注意喚起情報等のインターネット上の公開情報を参照し、これを明示することで真偽を評価する技術が存在するが、上記の観点からも担当者が判定根拠を理解する支援となることを確認した。

5-2. 社会実装に向けた取組の個別詳細

ターゲットへのヒアリングの詳細

現場運用の限界（完全検知、自動判定の不可能性）

- 検証3　完全自動化の限界と「判断支援ツール」へ

- 検証内容
  - 現状の課題（SNSプラットフォーマー）
    攻撃側（生成AI）の進化速度は凄まじく、単一の検知技術で「100%の精度」を永続的に保証することは不可能である。
    「誤検知（False Positive）」のリスクをゼロにできない以上、AIに全権を委ねる自動判定（ブラックボックス化）は、企業にとって逆にリスクとなる。
  - あるべき姿（判断支援）
    特定のツールに依存しない、技術の進化（いたちごっこ）に柔軟に対応し続けられる統合プラットフォーム。
    社会実装において必要なのは、「完全自動に判断してくれるツール」ではなく、人間が最終判断を下すための材料（素材の改ざん箇所、拡散元、過去の類似事例）を提示する「高度な判断支援システム」である。

5-2. 社会実装に向けた取組の個別詳細

ターゲットへのヒアリングの詳細

悪意ある「当たり屋」対応のコスト構造と削減効果

- 検証4　現場の対応で発生するコスト規模の調査

- 検証内容
  - 現状の課題（IPホルダー、ブランド保護）
    特定の企業を標的に、不祥事やデマを繰り返し捏造する「当たり屋」的アカウントが存在する。これらは愉快犯だけでなく、PV稼ぎ（インプレッション収益）や恐喝目的の場合もあり、対応は長期化する。
  - 見えないコスト
    これらへの対応には、24時間の有人監視、弁護士への相談、プロバイダへの削除申請手続きなど、膨大なリソースが割かれている。
    試算では、中規模以上の上場企業において年間1,500万～3,000万円相当のコストが発生しており、本システムによる自動化・省力化のROI（投資対効果）は極めて高い。

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

得られた成果

- 検知精度の有意な向上（見込み含む）
  統合判定技術の導入により、根拠を示した判定支援の必要性を確認した。特に生成AIで作成したディープフェイク等の高度な偽造コンテンツに対する頑健性が向上した。
- 事業化における具体的な経済価値の特定
  ターゲット層（大手IPホルダー等）において、現状年間3,000万円程度発生している監視・法務コストを本サービスで代替・効率化できるという具体的なROI（投資対効果）を算出した。
- 「3,000万円規模の投資判断（取締役決裁）」が可能な市場ニーズが存在することを特定した。

インタビュー先企業情報から推測した風評被害対応にかかる人件費の内訳予想

区分	主な業務内容	担当者数	週あたり工数（合計）	年間合計工数	概算費用
日常監視・一次検証	24時間体制に近い監視（外注含む有人チェック）、SNS巡回、一次レポート作成	一般4名	59時間	2,950時間	**950万円**
2次検証・技術調査	疑わしい動画・画像の素材解析、ツール提供会社とのやり取り、IT部門協力	係長2名	20時間	1,000時間	**437万円**
判断支援・内部調整	判定結果の解釈、広報方針の策定、各部署（IR等）との調整	課長2名	22時間	1,100時間	**658万円**
法務・警察・SNS連携	削除申請手続き、弁護士への相談、警察への被害相談、証拠保全	課長2名	22時間	1,100時間	**658万円**
経営判断・危機管理	重大事案の決裁、緊急対策会議の実施、記者会見準備、IR声明承認	部長/役員1名	6時間	300時間	**217万円**
合計		**11名(兼務)**	**129時間**	**6,430時間**	**2,920万円**

役職別単価の設定（時給換算/法定福利費含む：令和５年賃金構造基本統計調査より）

役員・部長級：7,245円

課長級：5,980円

係長級：4,370円

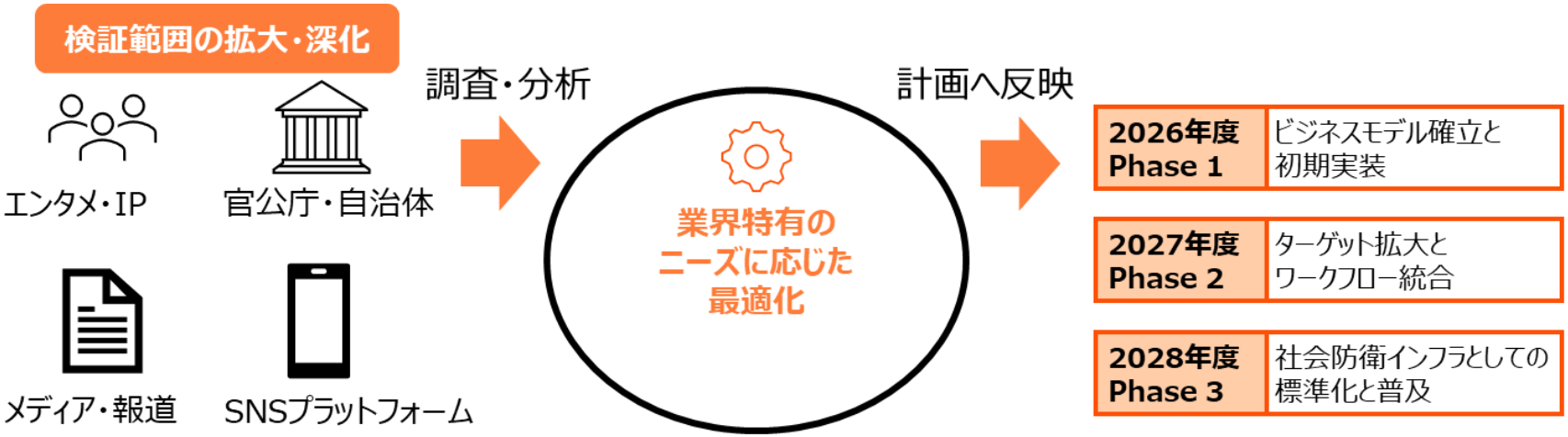
一般職：3,220円

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

実証を通じた社会実装ロードマップの確立

4つのユースケースに基づき、各業界特有のニーズ（コンテンツごとの判定基準等）に合わせたカスタマイズの必要性を特定した。



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

① 意思決定支援を核とした「多元統合」技術の確立

- 生成AIの高度化に伴い、機械による完全自動判定には限界があるという現状は、本事業が当初から想定していた通りである。この実証を通じて、ツールの役割を『人間が迅速かつ確信を持って決断を下すための高度な判断材料の提供』とするアプローチの正当性と不可欠性が改めて確認された。

- 高度な判断支援とエビデンスの可視化
  単なる判定スコアだけでなく、「改ざんの痕跡」を視覚的に提示することで、実務担当者の心理的負担を軽減し、客観的根拠不足による「沈黙のリスク」※を打破する。

- マルチモーダルな多元統合解析
  単一の検知ロジックでは限界がある日々高度化する偽情報に対し、画像・動画・音声といった複数の検知エンジンを組み合わせた「多元統合」を実施。各エンジンの得意領域で死角を補完し合い、単独判定では到達し得ない高精度かつ高信頼な判定基盤を実現を目指す。
  現状の対応範囲では、一部のツールを利用して画像、動画を組み合わせた判定が実現している一方で、音声＋動画・画像の組み合わせに対しての判定については、音声コンテンツ自体を補完的な位置付での評価に留めており、加えて動画・画像とリンクさせての評価は来年度の課題と認識している。

※沈黙のリスク： 客観的な判定根拠が不足しているために、企業が公式な初動対応を躊躇し、結果として情報の拡散を許してしまうリスクのこと。

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

- 成果指標（KPI）の転換
  単なる機械的な検知精度の追求に留まらず、「実務担当者が躊躇なく次のアクションに移れたか」という実効性に主眼を置いた評価軸を採用する必要性を実感した。組織としての対応スピードの最大化を本質的な価値として位置づける。

- 「総体としては偽」となるケース
  個別のコンテンツは真であっても、組み合わせや文脈によって「偽」となるケースが存在する。
  - 社会的リスクの認識
    コンテンツ自体が真であっても、悪意ある編集や文脈の切り取りによって、発信者の意図とは異なる印象を与える「社会的リスク」は極めて大きいと認識。単純な真偽判定のみでは、こうしたリスクを十分に評価できないことが本実証を通じて明確になった。
  - 来年度の展望
    今後は、単なるコンテンツの真贋を超え、「誰が・どのような意図で（悪意のあり・なし）」発信したかを考慮する文脈（ナラティブ）判別技術を高度化させることを検討。具体的には、「素材は真だが付随するテキストが偽」である場合などの振り分けルールや、ユースケースに応じた「悪意・不適切」の定義の策定を評価軸に組み込む展望を計画する。

- これにより、AIが断定するのではなく、人間が最終判断を下すための「高度な判断支援システム」としての社会実装を目指す。

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 社会実装を支える「運用ルール」と法的安全性の担保

技術の進化が攻撃側の高度化に後手に回るリスクを想定し、技術と「運用ルール（ガイドライン）」をセットで提供するビジネスモデルを確立する。

- 実務フローの標準化とプリセットの用意
  偽情報検知後の削除申請や法的通知、対抗論（カウンターナラティブ）※の作成など、混乱が生じやすい現場のために対応フローを標準化。ステークホルダー別に「何を、いつ、どのトーンで」伝えるべきか、即座に参照できる回答案（プリセット）を用意し、組織の対応に一貫性を持たせる。

- 非弁行為リスクの徹底回避
  ユーザーヒアリングで浮き彫りとなった弁護士法違反（非弁行為）への懸念に対し、法曹界と連携して「システムによる証拠収集」と「弁護士による法的判断・介入」の境界線を明確化。企業が安心して導入できる法的ルールを確立する。

- 心理的・物理的コストの削減
  高度な専門知識を要する判断を技術が代替・支援し、対応内容をルール化することで、担当者が個人の判断で重責を負う状況を解消。事前に合意された運用ルールに従うことで、意思決定のスピードを最大化する。

※**対抗論（カウンターナラティブ）：** 偽情報に対し、事実に基づいた「正しい文脈」を提示して影響力を中和する対抗措置。削除等の法的措置が完了するまでの間、即効性のある被害抑止策（火消し）として機能する。

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 社会インフラ化へ向けた「防衛エコシステム」の構築

一企業のサービスを超え、デジタル空間の真正性を担保する「社会的な防衛インフラ」としての地位確立を目指す。

- エコシステム（技術・法律・経済補償等）
  産学連携（神戸大学等）による技術の信頼性確保に加え、保険業界との連携による「偽・誤情報対策保険」の提供を模索。対策コストや時価総額毀損リスクをカバーし、被害発生時の経済的補償を可能にする持続可能な仕組みの構築を目指す。

- 対処機能の統合と被害最小化
  単なる検知に留まらず、プラットフォーム管理者への通報や証拠保全を迅速化。被害の最小化という実利を提供し、法的措置へのスムーズな移行を支援するエコシステムを実現する。

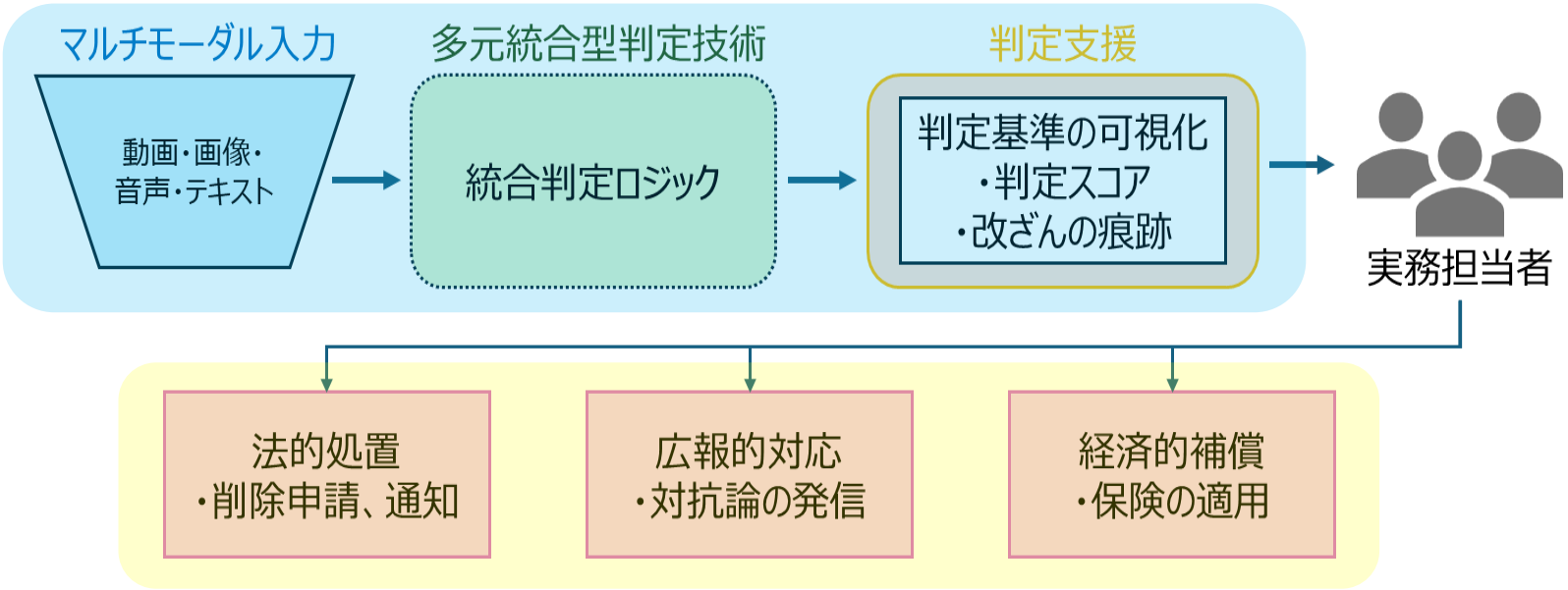
- 情報の信頼性確認インフラへの進化
  災害時、選挙時、あるいは平時の企業広報において、専門知識のない担当者でも迷わず決断を下せる環境を提供。デジタル社会の健全な発展を支える、不可欠な「情報の信頼性確認インフラ」へと進化させる。

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

多元統合型 偽・誤情報対策システムの構造

現時点で対応可能な範囲



追加で開発・検討の必要がある範囲

- 本事業の立ち位置と将来の課題解決アプローチ
  - 本取組で浮き彫りになった風評被害やブランド毀損等の重層的なペインポイントに対し、本技術は「人間が確信を持って意思決定するための高度な判断支援」により、最もクリティカルな初動対応を解決します。判定結果の「客観的エビデンス（心理的な許可証）」と「判定根拠の可視化」を提供することで、組織全体の対応スピードを最大化できる有効性を確認しました。今後はこの統合判定ロジックを核に、法的適応を考慮した運用ルールを整備し、社会を支える不可欠な「情報の防波堤」としての地位確立を目指します。

目次

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

普及啓発活動の基本方針

本実証で開発した「多元統合型偽・誤情報検出技術」の成果を広く社会に浸透させ、その重要性を啓発するため、広報活動・メディア対応、および関連イベント・セミナーでの発表・展示に積極的に協力する。

活動の目的

- 技術成果の社会浸透と啓発

単一技術では困難な「相互補完」による高精度な検知（精度向上5-15%）等の実証成果を広め、対策技術の必要性を周知する。

- 社会的理解促進と継続的改善

実効性の高い「判断支援」モデルへの理解を促し、活動を通じて得られるフィードバックを技術・運用ルールの継続的改善に反映させる。

- 情報環境の健全性向上への貢献

産官学の連携を強化し、我が国のデジタル社会における信頼性の高い情報流通環境の実現を支援する。

普及啓発活動のアジェンダ

プレスリリース（2025年9月～） 社会的認知度の向上と公的信頼の獲得。

有識者・実務担当者向けヒアリング・デモンストレーション（継続実施） 有用性の提示と理解促進。

イベント・セミナーでの発表・展示（2025年10月～2026年3月）

- 危機管理産業展/SEECAT（10月）
- IDF（デジタル・フォレンジック・コミュニティ）第22回シンポジウム（12月）
- くまもとサイバーセキュリティシンポジウム2025（12月）
- 保安電子通信技術セミナー・展示会（2月）
- 総務省主催 成果発信イベント（3月予定）

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

【情報発信】プレスリリースを通じた社会的認知の獲得

総務省事業採択の公表

① 実施内容

プレスリリース（2025年9月1日）：[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」への採択について]を公式発表。事業体制を公開した。

② 得られた成果と意義

- 社会的認知度の向上

  国レベルの重要プロジェクトであることを周知し、対策技術への関心を喚起した。

- 活動の基盤構築

  公的信頼を得ることで、その後の有識者ヒアリングやイベント展示における協力体制の円滑化を図った。

サン電子株式会社
RELEASE

2025年9月1日

【報道関係各位】

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」採択について

サン電子株式会社は、この度、総務省公募の令和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」（注1）において「多元的合成偽・誤情報検出技術の研究開発と実証事業」が採択されたことをお知らせいたします。

本事業は、サイバーコマンド株式会社、および、株式会社ザナリクスと連携し、社会問題となる偽・誤情報が拡散抑制に向けて、①複数の偽・誤情報判定プロダクトの結果を統合するアルゴリズムの高度化、②複数OSでの評価・検証基盤の整備、③中長期的には実運用環境における実証を通じ、精度・頑健性・説明可能性を備えた技術の確立を目指します。産学官が連携し、安心して情報にアクセスできる社会インフラの実現に寄与してまいります。

注1：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（令和7年度）」
https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/tasakugijutsu_fy2025.html

【サイバーコマンド株式会社について】
サイバーコマンド株式会社（本社：大阪府大阪市、代表取締役：大内 学）は、日本社会をサイバー犯罪・テロや各種セキュリティ事故から守ることをミッションとし、セキュリティ人材育成、セキュリティエンジニアリング対策、リスク評価、先端技術を融合したソリューションの企画・開発・提供を一貫して行っています。

【株式会社ザナリクスについて】
株式会社ザナリクス（本社：兵庫県神戸市、代表取締役：白石善明・神戸大学教授）は、国立大学法人神戸大学発の研究成果を基盤として設立されたベンチャー企業です。サイバー攻撃対策の社会実装を主たる事業とし、偽・誤情報分析技術をはじめとするサイバーセキュリティ関連において、研究成果の実用化、システム開発、人材育成コンサルティングを一貫して行っています。

【本件に関する報道関係者からのお問い合わせ先】
サン電子株式会社 グローバルDX事業部 井口、満園
〒105-0013 東京都港区浜松町二丁目2番12号 旧 浜松町ビル4階
TEL：03-3525-8191 FAX：03-6260-4886

サン電子ホームページ HOME 株主・投資家情報 IRニュース

https://contents.xj-storage.jp/xcontents/AS02371/a96c0d55/9210/4326/9548/acbfed18d8f4/140120250901551029.pdf

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

【実務啓発】有識者・実務担当者向けヒアリング・デモンストレーション

成果サマリーを用いた技術理解促進とヒアリング

① 内容・取組

- 報道機関、中央官庁、タレントマネジメント、IPホルダー、教育関係者等の関係者に対し、モックを用いて多元統合型偽・誤情報検出技術の理解を促進する。
- 実際の統合判定の挙動（モック）を用いたデモンストレーションを実施し、単なる数値的な精度だけでなく利用シーンのイメージを視覚化することで資料を用いた説明の次のステップに進み、さらなるペインポイントを抽出した。

② 得られた成果

- 「判断支援」への価値転換
  完全自動判定の限界を提示し、人間が最終判断を下すための材料を提示する「高度な判断支援システム」としての必要性について社会的理解を促進した。
- 日本特有の「沈黙のリスク」解消
  客観的な「判定スコア」が、日本企業特有の心理的障壁（不確実な状態での反論回避）を打破し、迅速な公式否定を行うための「心理的な許可証」となることを実証した。
- 実務フローの標準化と法的懸念の払拭
  検知後の「SNS事業者への削除申請」や「警察通報用レポート作成」の自動化ニーズを特定。同時に、ツール利用に伴う「非弁行為（弁護士法違反）」のリスク回避など、法曹界との連携による運用ルールの重要性を啓発した。
- 経済的合理性（ROI）の提示
  大手IPホルダー等において年間3,000万円規模に達する監視・対処コストが、本技術の導入により代替・効率化できるという具体的な投資対効果（取締役決裁ラインの妥当性）の目安であるという回答を導き出した。

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

【外部連携】シンポジウムへの出展を通じた専門家・自治体との対話

対象イベント

- 危機管理産業展（SEECAT）、IDFシンポジウム、くまもとサイバーセキュリティシンポジウム、保安電子通信技術セミナー・展示会等

実施時期　2025年10月～2026年2月

① 内容・取組

- 地方自治体、法執行機関、メディア関係者、IT専門家に対し、各分野の実務者が抱える将来的な懸念事項について意見交換を実施した。

② 得られた成果

- 法執行機関との対話
  捜査現場における偽・誤情報の深刻化に対し、将来的な「デジタル証拠の真正性確保」や「捜査工数の増大」といった、技術が解決すべき具体的な課題を抽出した。
- メディア担当者との対話
  現状は「疑わしいものは放送しない」という厳格な判断基準を運用しているが、将来的に対策技術が進化・普及すれば、技術的エビデンスに基づく「新たな放送判断基準」へと変容し得る可能性を確認した。
- 重要性の啓発
  地域社会や専門家コミュニティにおいて、特定の製品に依存しない「評価基盤」としての多元統合型判定技術の必要性を浸透させた。

6-2. 普及啓発活動の個別詳細

普及啓発活動の個別詳細

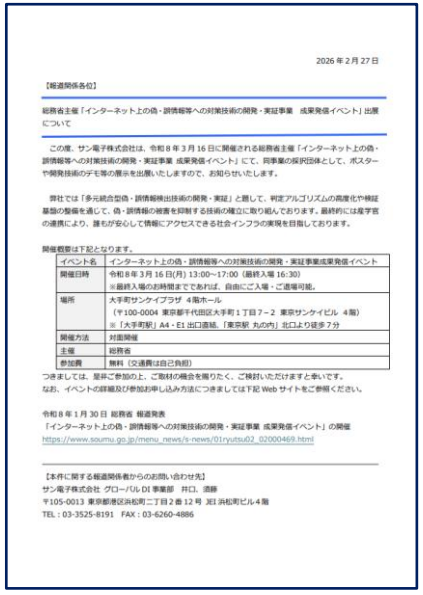
- 【外部連携】シンポジウムへの出展を通じた専門家・自治体との対話
  対象イベント
- 総務省主催 成果発信イベント（3月予定）

① 実施内容

プレスリリース（2026年2月27日）：[総務省主催「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 成果発信イベント」出展 について]を公開し、当日の来場を呼び掛けた。

② 得られた成果

- 未定



2026 年 2 月 27 日

【報道関係各位】

総務省主催「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　成果発信イベント」出展について

この度、サン電子株式会社は、令和 8 年 3 月 16 日に開催される総務省主催「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 成果発信イベント」にて、同事業の採択団体として、ポスターや開発技術のデモ等の展示を出展いたしますので、お知らせいたします。

弊社では「多元統合型偽・誤情報検出技術の開発・実証」と題して、判定アルゴリズムの高度化や検証基盤の整備を通して、偽・誤情報の複雑を抑制する技術の確立に取り組んでおります。最終的には産学官の連携により、誰もが安心して情報にアクセスできる社会インフラの実現を目指しております。

開催概要は下記となります。

イベント名	インターネット上の偽・誤情報等への対策技術の開発・実証事業成果発信イベント
開催日時	令和 8 年 3 月 16 日(月) 13:00~17:00（最終入場 16:30） ※最終入場のお時間までであれば、自由にご入場・ご退場可能。
場所	大手町サンケイプラザ 4 階ホール （〒100-0004 東京都千代田区大手町 1 丁目 7－2 東京サンケイビル 4 階） ※「大手町駅」A4・E1 出口直結、「東京駅 丸の内」北口より徒歩 7 分
開催方法	対面開催
主催	総務省
参加費	無料（交通費は自己負担）

つきましては、是非ご参加の上、ご取材の機会を賜りたく、ご検討いただけますと幸いです。
なお、イベントの詳細及び参加お申し込み方法につきましては下記 Web サイトをご参照ください。

令和 8 年 1 月 30 日 総務省 報道発表
「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 成果発信イベント」の開催
https://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000469.html

【本件に関する報道関係者からのお問い合わせ先】
サン電子株式会社 グローバル DI 事業部　村口、酒勝
〒105-0013 東京都港区浜松町二丁目 2 番 12 号 旭 浜松町ビル 4 階
TEL：03-3525-8191　FAX：03-6260-4886

サン電子ホームページ　HOME 株主・投資家情報 IRニュース

https://contents.xj-storage.jp/xcontents/AS02371/c275b76f/0804/4192/91e6/f7a1488ef30a/140120260227571885.pdf

目次

7-1. 技術開発及び社会実装における課題・展望

技術開発にあたっての今後の展望

検証データの拡充と整理

- 今後は、ユースケース別およびモダリティ別（画像・動画・音声・文章）に検証データの拡充を進めるとともに、コンテンツの内容や加工内容（切り抜き、字幕付与、音声差し替え等）に基づく分類を行う。
  これにより、実際の流通実態をより正確に反映した評価環境を構築し、統合判定ロジックの妥当性および限界をより精緻に検証できる基盤の整備を目指す。

統合方式および重み付け方針の高度化

- 統合結果の性能指標とユーザニーズ（過検知許容度、見逃し許容度）を結び付ける評価軸を整理し、ユースケース別に適切な統合方針を選択できる設計を検討する。また、コンテンツの内容や加工内容に応じた事前分類や多段階評価の導入を検討し、統合判定の安定性および説明性の向上を図る。なお、コンテンツの内容の文脈や加工意図を考慮したコンテキスト判別の高度化についても統合判定の重要な要素として位置付け、検討を継続する。

音声および文章モダリティの検証強化

- 音声および文章ディープフェイクの流通動向を踏まえ、対応ツールの拡充および評価手法の整理を進める。現状では、補完的位置づけにとどめているが、実運用を見据えた評価対象として段階的に検証を強化し、音声や文章に強みを持つツールを統合基盤へ取り込むことで、マルチモダリティにおいて一貫した高精度判定を実現する基盤を目指す。

段階的な実証フェーズへの移行

- モックによる疑似検証に加え、限定的な環境においてテストユーザによる操作検証を実施する。実際の利用を想定した検証を通じて、操作性、業務適合性、判断支援効果といった観点から評価を行い、実運用への移行を見据えた課題整理と改善を段階的に進めていく。

7-1. 技術開発及び社会実装における課題・展望

技術開発にあたっての今後の展望

検証データの拡充と評価基盤の整備

- ユースケース、モダリティ別に検証データを拡充し、内容や加工類型ごとの整理を行い、実態に即した評価環境を構築する。あわせて、調整用データと独立した評価用データを整備し、過学習の影響を排除した客観的な性能評価を実現する。これにより、データ分布の多様性を確保し、統合判定の汎化性能および評価の信頼性の向上を図る。

統合方式および重み付けの高度化

- 統合結果の性能指標とユーザニーズ（過検知許容度、見逃し許容度）を結び付ける評価軸を整理し、ユースケース別に適切な統合方針を選択できる設計を検討する。あわせて、内容や加工に応じた事前分類や多段階評価を導入し、統合判定の安定性および説明性の向上を図る。なお、コンテンツの文脈や加工意図を考慮したコンテキスト判別の高度化についても統合判定の重要な要素として位置付ける。

モダリティ拡張およびローカライズ対応の推進

- 音声偽造コンテンツの流通動向を踏まえ、対応ツールの拡充および評価手法の整理を進める。また、日本語ニュースやSNS等を対象とした評価データを整備し、日本語特有の表現や文脈を考慮した検知およびコンテキスト評価の高度化を進めることで、日本語圏での実利用を想定した判定精度の向上を図る。

継続的性能評価および重み付け最適化の仕組みの整備

- 将来的なモデル更新およびデータ分布の変化、ならびに新規ツールの導入に対応するため、統合判定の性能を継続的に評価し、重み付けを最適化する運用基盤を整備する。具体的には、モデルのバージョン更新および新規ツール追加時に既存の検証データセットを用いた再評価を実施し、ROC曲線およびPR曲線に基づく評価指標により性能を定量的に評価する仕組みを導入する。また、実運用データの傾向を継続的に把握し、入力データの分布変化や誤検知・見逃し傾向の変化が確認された場合には再評価を実施し、必要に応じて重み付けの再最適化を行うことで、統合判定の性能および信頼性の維持・向上を図る。

段階的な実証フェーズへの移行

- モックによる疑似検証に加え、限定的な環境においてテストユーザによる操作検証を実施し、操作性、業務適合性および判断支援効果の確認を行うことで、実運用に向けた段階的な実証を進める。

7-1. 技術開発及び社会実装における課題・展望

技術開発にあたっての今後の展望

- 実用型プロトタイプの開発
  本年度のモック開発は最小限の機能実装に留めたが、その検証結果に基づき、中長期的な計画として「実用型プロトタイプ」に必要な機能要件を定義した。

限定的プロトタイプと実用型プロトタイプの要件一覧

機能要件	非機能要件
ファイルアップロード（疑似処理） 分析履歴一覧 サマリレポート 詳細レポート（ツール別タブカード切替を含む） 判定結果のDB参照と表示（蓄積・表示） クレジット管理（仮環境） ユーザー認証（仮環境） 判定結果データベースの構築 クレジット課金ロジック・決済処理DB ユーザ管理DB セキュアログイン パフォーマンス・スケール要件 ストレージ連携・実ファイル処理 API連携（外部/内部問わず） リアルタイムでの共同編集やチーム間のコミュニケーションなどのコラボレーション機能	UIデザイン・画面表示 認証・認可（ログイン/権限） パフォーマンスチューニング セキュリティ強化（暗号化、監査ログ） 耐障害性・スケーラビリティ バックグラウンド処理、キュー、再実行制御など 完全な例外処理の網羅

表示説明 ：モック実装済み要件

7-1. 技術開発及び社会実装における課題・展望

社会実装にあたっての今後の展望

目指すべき今後の展望

1. 社会普及と収益実現の時間軸のギャップ（キャズム）を克服し、持続可能な事業として早期に収益化を実現するための施策を展開する。

- 具体的なROI（投資対効果）の提示

  年間3,000万円規模で発生している監視・法務コストを、本サービスで大幅に削減できるという具体的な試算に基づき、確度の高いビジネスモデルの構築を目指す。

- 付加価値の明確化による収益化

  単なる「検知ツールの提供」にとどまらず、高単価なライセンス契約に見合う「説明責任を果たすツール」として価値を明確化し、売上の具体化を推進する。

2. 「判断支援」への価値転換と説明責任の強化ツールに対する過度な期待値（完全検知・自動判定）ギャップを解消し、社会実装を実現する。

- 高度な判断支援システムへの進化

  AIに全権を委ねる自動判定（ブラックボックス化）ではなく、最終的に人間が確信を持って意思決定を下すための材料（改ざん痕跡や根拠スコア）を提供する「判断支援」をサービスの核に据える。

- 組織的な説明責任の完遂

  「なぜその判断に至ったのか」の根拠を可視化することで、公共性の高い分野（行政・報道等）において担当者が自信を持って公式発表を行える環境を整備する。

7-1. 技術開発及び社会実装における課題・展望

社会実装にあたっての今後の展望

3. 実務ワークフローの統合と運用の標準化検知そのものだけでなく、事後対応（対処・連携）を含むユーザー側の運用体制構築を支援。

- 対処・連携フローの自動化・効率化
  検知後の「SNS事業者への削除申請」や「法務・警察への連携用レポート作成」といったユーザー側の運用体制の未整備の解消を目指す。
- 運用モデル（手順書）の確立
  具体的な実装パートナーとのPOC（概念実証）を通じて実務上のボトルネックを理解し、現場の担当者が迷わず操作できる「運用モデル」や「手順書」を作成。

4. 運用・制度・社会接続の検証を構築。単なる判定技術の進歩のみならず、官民連携による社会的なインフラとしての地位確立を目指す。

- 専門家ネットワークとの連携
  IDF（デジタル・フォレンジック・コミュニティ）やくまもとサイバーセキュリティシンポジウム等で構築したネットワークを活かし、官民連携による「情報の信頼性基盤」としてのデファクトスタンダード化を目指す。
- 情報の信頼性基盤としての標準化
  特定の製品に依存しない「集合知としての判定」を社会的に定着させ、デジタル社会の健全な発展を支える「情報の防波堤」としての地位の確立を目指す。

7-1. 技術開発及び社会実装における課題・展望

社会実装にあたっての今後の課題

判定ロジックのローカライズと自社開発の必要性

- 現在の統合プラットフォームは外部の判定技術を基盤としているが、国内特有の文脈（日本の商習慣、エンタメ文化、特有のネットスラング等）に即した高精度な判定を実現するためには、判定ロジックのローカライズの必要性も認識している。

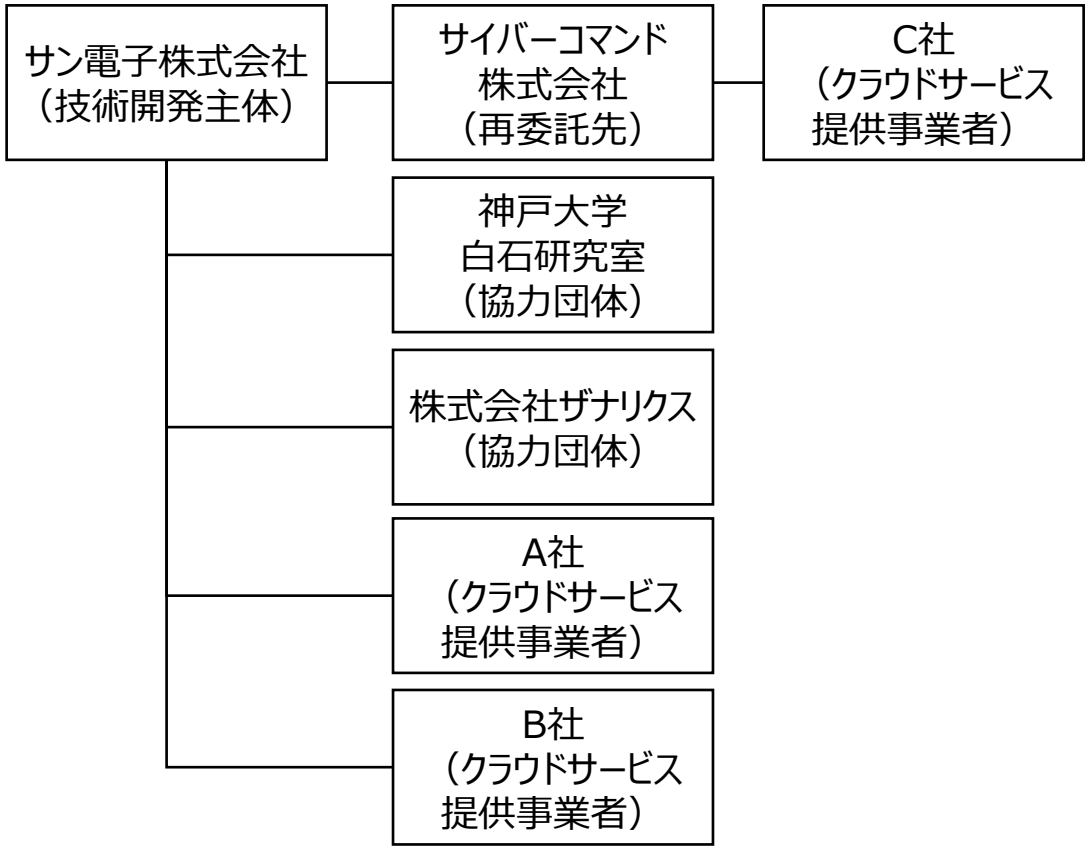
サン電子自社による判定ツールの開発検討

- 外部ツールの統合による強みを維持しつつ、並行してサン電子自社による判定ツールの開発を検討する。これにより以下の価値を創出できると期待する。
  - 技術のブラックボックス化解消
    自社開発により判定根拠の透明性を高め、行政や報道機関が求める「説明責任」をより強固に支援する。
  - 機動的なアップデート
    国内で発生する新たな偽造手法に対し、外部ベンダーに依存せずに自社で即座に検知モデルを更新・適用できる体制を構築する。
  - コスト構造の最適化
    自社技術を組み込むことで、API利用料等の外部依存コストを抑制し、中堅・中小企業でも導入可能な「合理的コスト」でのサービス提供を実現する。

目次

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

- サン電子株式会社
  本プラットフォーム開発プロジェクトの実行、及び、イスラエル技術プロバイダー社のサービスの導入/仕入れ
- サイバーコマンド株式会社
  プロジェクト全体支援・社会実装支援、及び、米国技術プロバイダー社のサービスの導入・仕入れ
- 株式会社ザナリクス
  複数ツール提供会社の統合技術と日本語環境での最適化処理において、実用的なシステム構築の観点から専門的なアドバイスを行う
- 神戸大学/白石研究室
  偽・誤情報検知精度向上に関する学術的見地からの支援を提供し、本実証の技術的信頼性と社会的価値の向上に貢献する
- A社、B社、C社は社名非公開
  Deepfake対策サービス提供事業者

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
【開発・検証】								
ツール提供会社との調整		→	→	→	→	→	→	→
デートデータの選定・準備		→	→					
判定結果正規化機能の開発				→	→			
精度向上統合判定ロジックの開発				→	→	→		
実証テスト計画の策定・準備			→	→	→			
ユースケース別の有効性検証				→	→	→		
定量的・定性的評価の実施					→	→		
【社会実装】								
社会実装に向けた調査・分析				→	→			
ターゲットユーザーへのヒアリングと要件整理			→	→	→	→		
サービス構成案とビジネスモデル案の策定			→	→	→	→		
国内展開戦略と市場性評価			→	→	→	→		
【普及啓発活動】								
広報活動・メディア対応への協力		→	→	→	→	→	→	→
イベントセミナーでの発表・展示				→	→	→	→	
総務省の普及啓発活動への協力			→	→	→	→	→	→