

令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

社会実装実施計画書

2025年3月31日

インターネット上の偽・誤情報技術の開発・実証事業 管理団体
ボストン・コンサルティング・グループ合同会社

Agenda

- 1. データグリッド
- 2. NEC
- 3. OP技術組合
- 4. DataSign
- 5. 関西テレビソフトウェア
- 6. エヴィクサー

令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

多様なメディアにおける最新のディープフェイクに追従した
偽・誤情報検出技術の開発・実証
社会実装実施計画書

2025年3月12日

株式会社データグリッド

目次

1. 社会実装実施計画の概要
 - 1-1. 本実証後の展望
 - 1-2. 事業化に向けた実効的な取組計画・事業化活動

1. 社会実装実施計画の概要

1-1 本実証後の展望

2025年度	2026年度	2027年度以降
SNS上の 偽・誤情報対策サービスの開発・実証	サービス普及に向けた本格展開	ビジネスモデルの確立
<p>ディープフェイク対策を含む、実行性ある偽・誤情報対策サービスのコアコンセプトを実装・実証する。</p> <p>メディア向けツールの開発・実証：</p> <ul style="list-style-type: none">ファクトチェック対象とする情報の選定やファクトチェック効率化を支援するシステムを開発・実証 <p>SNSユーザー向けツールの開発・実証：</p> <ul style="list-style-type: none">偽・誤情報と疑われる情報に対して、SNSユーザーが主体となるファクトチェックを支援するシステムを開発・実証 <p>各ステークホルダーとの実証を通じて、課題の抽出・改善を実施し、サービスの本格展開に向けた計画を策定する。</p>	<p>官公庁での利活用や対応SNSの拡充等により、サービス普及に向けた本格展開を開始する。</p> <p>システムの改良：</p> <ul style="list-style-type: none">2025年度実証で得たフィードバックをもとにした改善や複数SNSへの対応を強化新たな偽・誤情報のパターンへの対応策の実装 <p>偽・誤情報対策エコシステムの拡充：</p> <ul style="list-style-type: none">官公庁と連携した実証を実施し、公共分野での利活用を検証発信者の真正性保証技術等との連携 <p>廉価でのサービス提供等により実効性のある偽・誤情報対策の社会実装を最優先として事業展開する。</p>	<p>収益性を確保したビジネスモデル導入し、持続的な事業運営フェーズに移行する。</p> <p>収益モデルの確立：</p> <ul style="list-style-type: none">偽・誤情報対策にとどまらず、企業ブランドのリスク管理サービスとして応用海外市場への展開 <p>継続的技術改善：</p> <ul style="list-style-type: none">ユーザーデータやフィードバックに基づいた定期的な機能改善市場善、新機能の追加生成AIの進化への対応、セキュリティ対策、プライバシー保護の強化を継続的に実施 <p>展開地域や対象業界の拡張や継続した技術開発により、収益性のある偽・誤情報対策を普及する。</p>

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動

R5補正における課題及び今後の深化の観点

課題1：生成AI技術の急速な発展

- 画像、映像、音声に関するディープフェイク判別精度は本事業年度の目標を達成した一方で、本事業年度内においても新たな生成AIが次々と登場しており、そうした最新の生成AIによるディープフェイクに対しては、検出精度が劣化する例も確認されている
- そのため、持続的な事業運営に向けて、画像、映像、音声に関するディープフェイクの判定精度をさらに向上させることに加えて、低コストかつ短期間で最新のディープフェイクに対応する仕組みが不可欠である

課題2：ディープフェイク以外の偽・誤情報の多様な作成による被害

- 偽・誤情報は、当然ながら単に「ディープフェイク」技術によって作成されるものに限らず、その他にも多種多様な生成・加工手法が存在している。これにより、ネット上に拡散される虚偽のコンテンツの背景には複雑かつ多岐にわたる技術が絡んでおり、各手法に柔軟に対応可能な真偽判定技術の確立が急務となっている

課題3：メディアにおけるSNSファクトチェックの活性化

- 偽・誤情報が拡散し続ける現代の情報環境を踏まえ、偽・誤情報対策プロセス全体を一貫して支援するツールを開発し、広く提供する取り組みが必要とされている。これにより、各メディアが実効性の高い偽・誤情報対策の仕組みを効果的に整備し、正確な情報伝達と信頼性の向上を実現するための基盤を確立することが求められている

課題4：SNSユーザーによる主体的なファクトチェックの促進

- SNS利用者が自らの意思に基づいて積極的にファクトチェックを行えるよう支援する仕組みを整備するとともに、偽・誤情報を見極めるための判断力を養うメディア・リテラシーの向上を図る取り組みが、現代社会において非常に重要な課題として求められている

Agenda

1. データグリッド
- 2. NEC
3. OP技術組合
4. DataSign
5. 関西テレビソフトウェア
6. エヴィクサー

令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

AIを活用した情報コンテンツの真偽判別支援技術の開発・実証 社会実装実施計画書

2025年3月12日

日本電気株式会社

目次

1. 社会実装実施計画の概要

1-1. 本実証後の展望

1-2. 事業化に向けた実効的な取組計画・事業化活動

1. 社会実装実施計画の概要

1-1 本実証後の展望

2025年度	2026年度	2027年度以降
<ul style="list-style-type: none">➤ 一部ユーザーでの実用化➤ プラットフォーム検討/整備/構築➤ 適用ユーザー拡大検討	<ul style="list-style-type: none">➤ 実用化ユーザー拡大➤ プラットフォームの高度化	<ul style="list-style-type: none">➤ ビジネスとしての自走化➤ 更なる適用ユーザー拡大検討
<p>優先度が高い領域で専門ユーザーが業務利用</p> <ul style="list-style-type: none">• 一部ユーザーの利用が進む <p>プラットフォーム構築</p> <ul style="list-style-type: none">• 総務省24年度実証の他案件成果物（NEC以外）等との組み合わせによる総合的な偽・誤情報分析・検知プラットフォームの構築 <p>適用ユーザーの拡大に向けた実証</p> <ul style="list-style-type: none">• 適応ユーザー拡大のための必要機能整理・設計	<p>適用範囲が拡大、専門ユーザーの業務利用の範囲が拡大</p> <ul style="list-style-type: none">• 感度の高い / 親和性の高い利用主体（メディア・自治体等）での活用が広がる <p>プラットフォームの高度化</p> <ul style="list-style-type: none">• 広い事業者等で専門ユーザーが業務利用を行えるプラットフォームへ成長	<p>継続的な技術開発・投資に向けてビジネスモデルが確立</p> <ul style="list-style-type: none">• 先行事例の確立に伴い、幅広いユーザーでの活用が進み専門ユーザーの業務利用が拡大 <p>更なる適用ユーザー拡大</p> <ul style="list-style-type: none">• 非専門ユーザーへの適応拡大検討・実証

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動

R5補正における課題及び今後の深化の観点

社会実装に向けた技術課題

(ファクトチェック機関向け)

- 出展情報の取得方法のさらなる工夫
 - 他社記事を引用した2次、3次情報から1次情報に辿りつき、証拠出展情報の優先度をつけて、分析精度を高めたい。
 - 有料サイトや論文、社内情報を加味した真偽分析ができるか。

(マスメディア・放送局・新聞社・通信社)

- 速報用途での偽・誤情報の分析ニーズへの対応
 - 災害報道への対応
 - 位置情報に関する他マップツールなどとの風景照合や過去画像の流用確認による偽・誤情報検知ができるか。

(その他企業)

- 判定を目的とした用途への対応
 - ファクトチェック機関やマスメディアと異なり、分析よりも判定に重きをおいている。
 - 一方ファクトチェック機関からは、偽・誤情報に対してのリテラシーの低いまま、本システムを使用することで真偽結果を誤ってしまう懸念もあることから対象者へのレクチャーや、よりわかりやすいUIや判断理由の説明ロジックが必要となる。

Agenda

1. データグリッド
2. NEC
- 3. OP技術組合
4. DataSign
5. 関西テレビソフトウェア
6. エヴィクサー

令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

3 Originator Profile技術研究組合 社会実装実施計画書

2025年2月28日

Originator Profile
技術研究組合

目次

1. 社会実装実施計画の概要
 - 1-1. 本実証後の展開（案）
 - 1-2. 事業化に向けた実効的な取組計画・事業化活動

1. 社会実装実施計画の概要

1-1 本実証後の展開（案）

2025年度	2026年度	2027年度以降
一部ユーザーでの実用化	適用範囲の拡大	ビジネスとしての自走化
<p>本実証で検証系サイトにOP実装した報道機関、自治体の通常Webサイト（本番系）へのOP実装を目指す。</p> <p>次に、2024年の実証で検討したニュースの配信の技術仕様を元に、ニュースアグリゲーター、プラットフォームへのOPを利用した記事配信の実証を行う。</p> <p>さらに発信元の識別をするブラウザ機能拡張の改善と主要ブラウザ対応を行い、一般市民のOP拡張機能のダウンロードとその利用を促進するとともに、W3Cなど標準化団体への提案を開始する。</p> <ul style="list-style-type: none">自治体WebサイトへのOP実装とその安定的運用をサポートするためのCMS提供企業とOPCIPの協業スマートニュースアプリやNTTグループへのニュースの再配信の実証ブラウザ機能拡張の主要ブラウザ対応ブラウザ機能拡張の一般市民の利用促進W3Cなど標準化団体への提案開始	<p>鳥取県での実績をもとに南海トラフ地震対策地域の29自治体の中で経済的インパクトの大きい地域を検討し、当該自治体へのOP実装と当該自治体に本社を置く、または創業地とする企業のサイトへのOP実装、当該地域の一般市民のブラウザOP拡張機能アプリの利用促進を横展開していく。</p> <p>さらに、これら自治体にCMSを提供する企業とOP技術研究組合が協業し、WebサイトへのOP実装のビジネスモデルを確立させていく。</p>	<p>自治体や企業サイトでのOP実装実績をもとに、広告関連団体や、企業のデジタル広告受発注業務でのOP利用を促進していく。またWebサイトでのOP利用やデジタル広告取引におけるOP利用への課金設定をOP普及促進とともに検討していく。</p>

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動

R5補正における課題及び今後の深化の観点

社会実装に着手したうえで見えてきた課題と、それを踏まえた今後の社会実装を深める方向性

• OP技術の高度化

現状のテキストと画像へのOP対応に加え、以下のコンテンツへの対応やその再配信対応を図っていく。

- テキストと画像コンテンツの再配信対応
- ストリーミング動画などへの対応
- 生成AIコンテンツへの対応
- ブラウザに加えてスマホアプリでのOP情報の閲覧を可能とする

• OP技術のユースケース拡大

- 自治体や企業のWebサイトのスムーズなOP対応のために
 - 様々なCMSへのOPプラグインの開発のためにCMSベンダーやクラウドサーバー提供ベンダーとの協業
 - 企業や自治体Webサイト製作を業務とする広告会社やシステムベンダー、制作会社などとの協業体制の整備
- OP技術の広告システムとの連携のために
 - 広告関連団体とOP利用促進に関する合意形成を図っていく。

• 国際標準化の推進

- 国際標準化に向けた活動進捗の成果詳細は成果報告書に記載。

Agenda

1. データグリッド
2. NEC
3. OP技術組合
- 4. DataSign
5. 関西テレビソフトウェア
6. エヴィクサー

令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

個人の署名によるコンテンツの真偽表明データベース 社会実装実施計画書

2025年3月3日

株式会社DataSign

目次

1. 社会実装実施計画の概要

1-1. 本実証後の展望

1-2. 事業化に向けた実効的な取組計画・事業化活動

1. 社会実装実施計画の概要

1-1 本実証後の展望

2025年度	2026年度	2027年度以降
一部ユーザーでの実用化	一般への開放	サステイナブルな仕組みとして自走
<p>著名人・ジャーナリストを対象とした実証に基づいた改善、機能追加等を行い、一部社会実装を行う。 また、法人やすべてのインターネット利用者を対象とした社会実装を行うための実証を行う。</p> <ul style="list-style-type: none">システムをサステイナブルに運営するためのスキーム、ビジネスモデル等を検討する著名人・ジャーナリストに対する属性証明のためのガバナンスモデルの確立法人および一般の生活者が一部の属性を証明した上で真偽情報を表明できるようにする信頼性のさらなる確保に向けたブロックチェーンの適用領域を検討し実証を行う	<p>本システムをサステイナブルに運営する主体により、運用が開始される。 プラットフォーム事業者、広告配信事業者等によるファクトチェックやコンテンツモデレーションに活用が行われる。</p> <ul style="list-style-type: none">システムをサステイナブルに運営するためのスキーム、ビジネスモデルの確立法人および一般の生活者に対する属性証明のガバナンスモデルの確立メディア事業社、プラットフォーム事業者、広告配信事業者での活用が開始される災害時の救助要請等、公共分野を含めた横展開の可能性を検討する	<p>サステイナブルな取り組み（ビジネス含む）に不可欠なシステムとなる。 本システムを横展開した取り組みの運用が開始され、さまざまな領域での活用の検討が行われる。</p> <ul style="list-style-type: none">メディア事業社、プラットフォーム事業者、広告配信事業者において、事業に不可欠なシステムとして導入される本システムのOSSコミュニティが発生し、様々な領域で活用の検討、実装がされはじめる公共分野で本システムを利用した実証がおこなわれ、公共分野においてサステイナブルに運用するためのスキームの検討する

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動

R5補正における課題及び今後の深化の観点

- 身元証明書の発行における課題
 - どのような主体が誰に対して、どのような基準で身元証明書を発行し、その発行者をどのように信頼できるかという点に関して、フィードバックの中でも指摘があった。
 - Boolcheckに対し偽誤情報を名前や所属等の属性を公開して投稿したい方に対し、どのように身元証明書を発行するかという点について運用ルールが必要である。
 - 今回の実証では、株式会社DataSignまたは公益社団法人日本ジャーナリスト協会が身元証明書の発行主体となり実証を行い、発行主体の実在性については、サーバー証明書（OV証明書）を活用することにより、既存のウェブのガバナンスに則った信頼性を確保できたが、以下の観点については課題となった。
 - 発行主体の身元証明書発行業務がスケールしない。
 - 発行主体が虚偽の身元証明書を発行しないことをどのように担保するか。
 - 発行主体が適切な者に対して身元証明書を発行しているか。
 - 発行主体による身元証明書発行対象者の適切性判断基準をどのように設定しているか。
- 真偽情報への署名/投稿における課題
 - 「署名する」という今までにない操作が必要となるため、実際に利用してもらう際に、投稿者の理解が進まなかった。具体的には「Boolcheckにログインするにはどうするのか？」といった質問やなど、通常のSNSに投稿する際のUXとは異なることによる混乱が見られた。
 - 今回の実証では所属を含めた身元証明を前提としたため、所属組織の公式なコメントとして見られてしまう可能性があり、気軽に投稿が出来ないというフィードバックもあり、所属組織に依存しない身元証明の方法も必要であることが確認できた。
 - 選択的開示により、顔写真、名前を秘匿して投稿できる機能の実装も行ったが、1回目は顔写真無しで、2回目は顔写真ありで投稿した場合、証明書による名寄せによって、1回目の投稿に対して顔写真が紐づいてしまう。所属と苗字が必須となっているので、この事象は今回の実証においては大きな問題とはならないが、この投稿では性別だけ公開したい、別の投稿では名前を公開しても良い、といったことを実現するためには新たな技術を用いる必要がある。
- 真偽情報の確認における課題
 - BoolcheckではURLに対して真偽情報を投稿することができるように構成したが、真偽情報閲覧者が当該URLにアクセスした際に表示されるコンテンツが真偽情報投稿時点とは異なる可能性があるため、Wayback Machine等の仕組みを利用し、真偽情報投稿時のコンテンツを残す必要がある。
 - また、URL全体に対しての真偽情報のため、例えば、メディアサイトに訪れた際に、その記事自体は真の情報であったとしても、そのページに表示される広告がそれ自体が偽・誤情報であったり、偽・誤情報に誘導するものである等の可能性があるが、現状のBoolcheckではそのような広告が表示される段階でアラートを表示するような機能が無い。
 - Boolcheckでは、身元が検証できる真偽情報と、匿名の真偽情報を分けて表示しており、それぞれの件数が確認できるが、件数を確認できるだけであり、その情報が偽・誤情報であること（無いこと）を示しているわけではない。最終的には真偽情報閲覧者がその判断をすることとなるが、件数だけでは無く、偽誤情報を総合的に判断してそのURLが偽・誤情報か否かを表示するようにしても良いのではないかとのご意見もいただいた。
 - そもそも情報閲覧者がBoolcheckの存在を知っていることが前提のため、メディアやプラットフォームに組み込む必要がある。

Agenda

1. データグリッド
2. NEC
3. OP技術組合
4. DataSign
- 5. 関西テレビソフトウェア
6. エヴィクサー

令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

放送波を活用した災害時における偽・誤情報対策技術の実証 社会実装実施計画書

2025年2月28日

関西テレビソフトウェア株式会社

目次

1. 社会実装実施計画の概要

1-1. 本実証後の展望

1-2. 事業化に向けた実効的な取組計画・事業化活動

1. 社会実装実施計画の概要

1-1 本実証後の展望

2025年度	2026年度	2027年度以降
実運用に向けて完成度を高める	既存システムとの連携により実用化	偽・誤情報の対象を拡張（横展開）
機能性向上 <ul style="list-style-type: none">実運用に向けたシステム構成の検討、精度向上と利便性を高める。 放送設備を使用した検証 <ul style="list-style-type: none">本放送への使用に向けて、実際に放送設備の検証を行う。 受信機開発 <ul style="list-style-type: none">アプリケーション開発が容易になる受信機のプロトタイプを開発する。 セキュリティ検証 <ul style="list-style-type: none">本手法のセキュリティについて第三者企業からの評価を受ける。	災害時のユースケースを広めニーズを探る <ul style="list-style-type: none">安否情報との連携 等災害時流通システムとの連携 等 本放送への継続的重畳時の影響調査 <ul style="list-style-type: none">既存の放送内容や冗長構成の影響調査放送運行システムの設定等の影響調査 受信環境に依存しない機能の構築 <ul style="list-style-type: none">信頼性基盤としての有効性を維持できるか調査複数手段による情報配信ローカルネットワークでの情報補完 受信機普及に向けた課題解決 <ul style="list-style-type: none">受信機購入インセンティブ設計中継局の接近情報配信	平常時においても有効活用 <ul style="list-style-type: none">サイバーセキュリティ（以下、展開例）<ul style="list-style-type: none">偽・誤情報を信じたユーザーによるシステム制御を行うことを防ぐ。スマートシティ（以下、展開例）<ul style="list-style-type: none">偽・誤情報による、都市観測システムの分析精度の低下を防ぐ。 情報のベーシックインカム（最終目標） <ul style="list-style-type: none">積極的な経済活動を支える通信に対して、地域の安心・安全にかかわるデータは放送で平等に配信。受信したデータを根拠に経済活動や行政手続きが行える信頼性を保つ。<ul style="list-style-type: none">安全保障制度として提言セキュリティを保つガイドラインの設定日本発の国際規格として展開

1-2 事業化に向けた実効的な取組計画・事業化活動

R5補正における課題及び今後の深化の観点

見えてきた課題と今後の方向性

- 画像だけでは偽・誤情報とは言い切れない（例：画像の内容は真実だが、日時や場所を誤解している等）。
 - テキストなど複合的に判断する必要性。
- 長期間運用の概念検証は成功したが、実運用するためには速度パフォーマンスなどさらなる改善が必要。
 - データベースやキャッシュシステムの採用。
- ブロックチェーンを使用するための暗号資産の管理を改善する必要。
 - 偽・誤情報を登録する組織と煩雑な暗号資産の管理を分離する必要がある。
- 受信機開発の必要性
 - 受信機と検証アプリケーションをシームレスに転送連携できることが重要。
 - ✓ その他多種多様の受信機と共存するためのガイドラインや規格提言の必要性。
- セキュリティ評価を得る必要性
 - 導入の候補として幅広い業界から注目されるためには、双方向通信よりもセキュアであることのアピールが重要。
 - 安全な運用、受信機量産の判断材料となるガイドライン策定の必要性。
- 技術周知の必要性
 - IPDCは現状の放送送信インフラがそのまま使え、受信機がどれだけ増えても送信にかかるコストが増加しないメリットの周知。
 - 平常時も使えるニーズを発掘することで需要を拡大する必要がある。
 - ✓ 偽・誤情報問わずユースケースの拡大が必要。

Agenda

1. データグリッド
2. NEC
3. OP技術組合
4. DataSign
5. 関西テレビソフトウェア
- 6. エヴィクサー

令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

音響透かしと音響フィンガープリントを用いた 偽・誤情報対策クラウドシステムの開発 社会実装実施計画書

2025年2月28日

エヴィクサー株式会社

目次

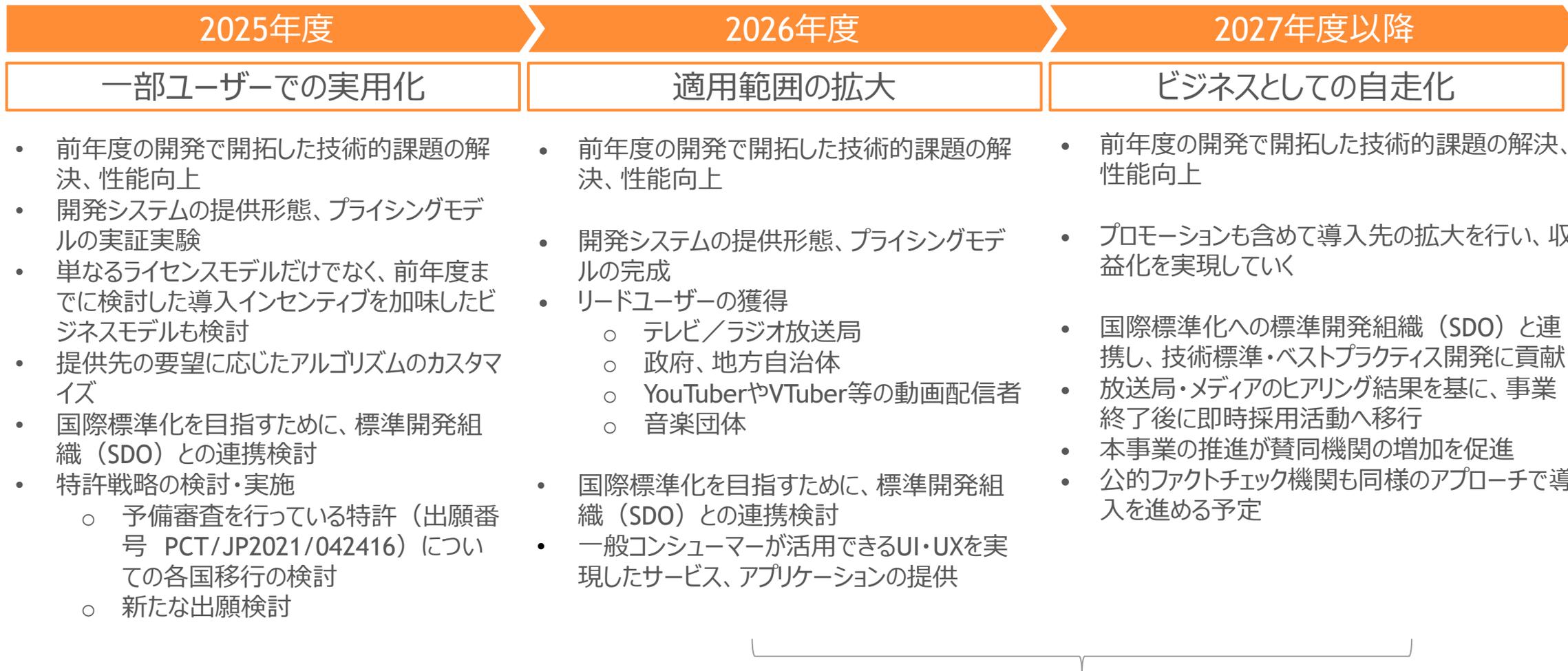
1. 社会実装実施計画の概要

1-1. 本実証後の展望

1-2. 事業化に向けた実効的な取組計画・事業化活動

1. 社会実装実施計画の概要

1-1 本実証後の展望



社会実装を優先し、外部環境によって
アジャイル的に項目は前後することも想定

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動 (1/8)

課題及び今後の深化の観点

法的観点から必要となる取組みの分析・評価

本事業によって開発した技術を用いて、あるコンテンツが偽・誤情報を含むものであることを明らかにすることができたとして、そのような偽・誤情報への対策・対応を行うには、適用される法令等に基づき、最終的には法的手段を講じる必要がある。そこで、開発した技術及びこれによるコンテンツの真偽検証の結果が、被害者による偽・誤情報に対する法的措置の支援という観点からも有用なものとなり得るかどうかについて、法的に必要な要件や手続に即して検討を行った。この観点からは、特に生成AIを用いた偽・誤情報による被害が発生した際に、裁判所又はプラットフォームに対して法的請求を行うための要件を明確にし、それを効果的に実行する手続きについて整理する必要がある。

1.1 法的請求の要件

1.1.1 総論

日本の現行の法令においては、偽・誤情報を作成し、公表しただけでは、違法な行為とは評価されない。例えば、(a)偽・誤情報の作成や公表が他人の著作権を侵害する場合、(b)偽・誤情報を他人に示し、それが真実であると誤信させて財物を取得する場合（詐欺に該当する場合）、(c)偽・誤情報の頒布が、他人の名誉や信用を毀損する場合（名誉毀損や信用毀損に該当する場合）等に、違法な行為として、刑事罰の対象となったり、民事上の差止請求や損害賠償請求の対象となったりする可能性が生じる。すなわち、対象となったコンテンツに偽・誤情報が含まれていることは、適用される法令の要件を満たすための事実（いわゆる構成要件や要件事実）の一つであり、本事業によって開発した技術やこれによる真偽の検証結果は、その立証手段の一つという位置付けとなると考えられる。以下では、偽・誤情報に対し、民事訴訟における法的請求によって対応する場合を念頭に、本事業によって開発した技術の用途、位置付けを検討する。

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動 (2/8)

課題及び今後の深化の観点

1.1.2 裁判手続の場合

本事業における偽・誤情報対策の技術によって情報の真偽が判定できたとしても、前述のとおり、そのような情報の真偽は、差止めや損害賠償のような法的請求を行うための要件の一つでしかないため、情報が偽物又は虚偽であることを主張するのみでは、裁判所においてその差止請求や損害賠償請求が認められるものではない。

裁判所において有効なものとして認められる法的請求になるか否かは、個別のケースに適用される法律の要件に従って判断することになる。例えば、

- (1) 名誉毀損のケースでは、名誉毀損の真実性の抗弁の法的要件において、
- (2) 著作権侵害のケースでは、著作権侵害における類似性や依拠性の判断において、
- (3) 詐欺罪のケースでは、人を欺く行為や被害者の錯誤において、
- (4) 信用毀損のケースでは、虚偽の風説や虚偽の事実の判断において、偽・誤情報対策の技術が有益な判断材料を提供し得ると考えられる。

1-2 事業化に向けた実効的な取組計画・事業化活動 (3/8)

課題及び今後の深化の観点

(1) 名誉毀損のケース

まず、名誉毀損の真実性の抗弁は、公然と事実を摘示して人の名誉を毀損する行為がされたとしても、その行為が、①公共の利害に関する事実にかかり、②目的が専ら公益を図ることにあつて、③重要な部分において真実であることの証明があった場合には、違法性が阻却され、名誉毀損による不法行為が成立しないとするものである（刑法230条の2第項）。

本事業における偽・誤情報対策の技術は、上記③の法的要件の判断において、有益な判断材料を提供すると考えられる。例えば、AIによる生成音声を使って音声の改竄が行われていることが音響透かしの技術によって判明した場合、当該生成音声そのものによって又は当該生成音声を根拠にしてなされた名誉毀損は、真実性の抗弁が認められない方向に評価されうる。

(2) 著作権侵害のケース

著作権侵害における類似性は、後発の作品が既存の著作物と同一又は類似していることをいい、著作権侵害が認められるための要件の一つとされている。本事業における偽・誤情報対策の技術は、類似性の判断において、有益な判断材料を提供すると考えられる。例えば、無断でアップロードされている動画の音声の大部分が元の動画と同一であることが、音響透かしや音響フィンガープリントの技術によって判明した場合、元の動画との類似性が認められ、著作権侵害が肯定される方向に評価されうる。

また、著作権侵害の要件の一つに、既存の著作物に依拠して創作されたものであることが挙げられるが、無断でアップロードされた動画から、権利者が埋め込んだ音響透かしが検出されれば、著作物の一部をコピーしていることが明らかになることから、依拠していることの立証にとって有益であると考えられる（但し、生成AIが自ら学習したコンテンツから、他社の著作物との類似性が認められる作品を生成した場合に、そのAIの利用者に依拠性が認められるか否かについては、個々の事案によって異なる判断となり得る。）。

1-2 事業化に向けた実効的な取組計画・事業化活動 (4/8)

課題及び今後の深化の観点

(3) 詐欺罪のケース

刑法246条に定める詐欺罪の成立には、①人を欺く行為、②被害者の錯誤、③被害者による交付行為、及び④財物又は財産上の利益の移転という4つの構成要件を満たす必要があるとされている。本事業における偽・誤情報対策の技術は、上記①又は②の構成要件の判断において、有益な判断材料を提供すると考えられる。例えば、被害者が著名人の動画の発言を信じて投資詐欺に遭ったケースにおいて、その著名人の音声改竄されていることを音響透かしや音響フィンガープリントの技術によって検証することができれば、人を欺く行為があったこと、及び、被害者の錯誤があったことの立証に役立つと考えられる。

(4) 信用毀損のケース

刑法233条に定める信用毀損罪の成立には、①虚偽の風説を流布すること又は偽計を用いること、及び②人の信用を毀損することという2つの構成要件を満たす必要がある。また、不正競争防止法2条1項21号に定める信用毀損行為に該当するには、①競争関係にある者が、②客観的事実に反する虚偽の事実を、③告知又は流布して、④他人の、⑤営業上の信用を害する、という5つの要件を満たす必要がある。本事業における偽・誤情報対策の技術は、刑法上の信用毀損罪における上記①の「虚偽の風説」、又は、不正競争防止法上の信用毀損における上記②の「虚偽の事実」の判断において、有益な判断材料を提供すると考えられる。例えば、企業の代表者が自社の評判を下げるような発言をしている動画が、別の真正な動画を改竄して作成されたものであることが、音響透かしや音響フィンガープリントの技術によって判明した場合、その動画の内容が虚偽の風説又は虚偽の事実であることを立証するうえで役に立つと考えられる。

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動 (5/8)

課題及び今後の深化の観点

以上のように、本事業における偽・誤情報の検証技術を用いた判定結果を取りまとめた報告書は、裁判における立証方法として用いられることが想定される。証拠としての価値を担保するためには、

- ①用いられる技術の要点、
 - ②これらの技術によってコンテンツの真偽を判断し得る根拠・論拠、
 - ③真偽判定のために用いられたデータ及びその収集方法、
 - ④上記③の②への当てはめ、
 - ⑤結論を整理した報告書とすること、
- が考えられる。

このような新しい技術を用いた立証方法については、裁判例の蓄積等もないため、今後の実務的な動向を踏まえ、報告書の内容を更新するなどの対応も必要となると思われる。

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動 (6/8)

課題及び今後の深化の観点

1.1.3 プラットフォームの場合

YouTubeやX（旧Twitter）のような動画コンテンツのプラットフォームに対して、問題の動画の削除等の請求を行う場合、偽・誤情報であることによって、当該プラットフォームの運営者により、その請求が有効なものとして認められ、問題のコンテンツが削除されるなどの対応が講じられる場合がある。例えば、YouTubeでは、誤った情報に関するポリシーやなりすましに関するポリシーが定められており、これらに違反したコンテンツについては、YouTubeの審査チームに対して報告を行うことができる。また、Xでは、合成又は操作されたメディアに関するポリシーが定められており、偽装アカウント、偽装行為、偽装コンテンツは、アプリ内の専用報告手続により報告することができる。このように、本事業における偽・誤情報対策の技術によって情報の真偽が判定できた場合、プラットフォームに対して削除等の請求を行うことで、対応できる場合がある。

しかし、プラットフォームにおいては、問題のあるコンテンツの削除を求めるなどの請求の方法が、それぞれの利用規約等に定められており、本事業における偽・誤情報対策の技術を生かした請求の根拠・証拠を提出することができない可能性がある。例えば、YouTubeやX（旧Twitter）のような動画コンテンツのプラットフォームにおいては、ポリシーの違反を報告する際に、「Spam or misleading」といった類型を選択できるにとどまり、その具体的な根拠を提出することができない。そのため、仮に音響透かし又は音響フィンガープリントの技術によって音声改ざんされたものであるという検証結果が得られたとしても、それをプラットフォームに提供する手段が設けられていない場合がある。

1-2 事業化に向けた実効的な取組計画・事業化活動 (7/8)

課題及び今後の深化の観点

1.2. 証拠収集と立証の課題

生成AIで作成された偽動画や偽情報に対して法的措置を講じる場合、その相手方（権利侵害の主体）の特定に加えて、違法な権利侵害行為を立証するための証拠収集とその信憑性の確保が大きな課題である。単に生成AIを使って偽動画や偽情報が作成されたことを主張するだけでは、偽動画や偽情報の違法性が認められる可能性は低いため、証明力の高い証拠を収集し、必要な事実を立証することが求められる。証拠収集と立証に当たっては、次のポイントを考慮する必要がある。

● 法的請求を行うための要件：法令に基づいて裁判手続により差止めや損害賠償を請求する場合と、プラットフォームの運営者に対し、利用規約等に基づいて問題のコンテンツの削除等を求める場合とでは、それぞれの請求が認められるための要件が異なる。個別のケースに適用される法令や利用規約等の条項を確認し、本事業の技術で得られる検証結果が有用かどうかを分析する必要がある。

● 法的請求を行うための手続：法令に基づいて裁判手続により差止めや損害賠償を請求する場合と、プラットフォームの運営者に対し、利用規約等に基づいて問題のコンテンツの削除等を求める場合とでは、それぞれの請求が認められるための手続が異なる。それぞれ、民事訴訟法や民事保全法に基づく裁判手続、プラットフォーム運営者が定める利用規約等に定める手続を確認し、必要な要件を満たした証拠の提出方法を検討する必要がある。特にプラットフォーム運営者に対する請求の場合には、提出できる証拠が限定されている可能性があるため留意が必要である。

1. 社会実装実施計画の概要

1-2 事業化に向けた実効的な取組計画・事業化活動 (8/8)

課題及び今後の深化の観点

2. 技術的観点から必要となる取組みの分析・評価

偽・誤情報対策においては、法的対応だけでなく技術的対応も重要な役割を果たす。以下のような技術的ソリューションが検討されている。

2.1 コンテンツ来歴管理

コンテンツ来歴管理は、誰がその情報を作成し、どのようなプロセスを経て公開されたかを検証するための技術である。たとえば、公共性のあるメディアから放映された音声や動画に透かし音を埋め込み、これを第三者機関で検証することで、その信憑性を保証することが可能である。

2.2 信頼性の確認

コンテンツの発信者やプロセスが適正であるかをエンドユーザーが確認できる仕組みを提供する。コンテンツに一意的 IDなどを付与し、エンドユーザーが簡単な操作でその正当性を確認できるようにする。これにより、偽情報の拡散を防ぐだけでなく、情報の信頼性を高めることができる。

2.3. 必要な技術的対策

- ・各発信者が自ら作成した情報コンテンツに自身の情報の付与を行い、信頼性や発信者の実在性を保証する方法。
- ・閲覧者が付与された情報を閲覧することで、情報コンテンツの信頼性や発信者の実在性の判断を行う方法を検討
- ・来歴情報等の付与を行うことで、信頼性を保証する方法及び閲覧者が付与された来歴情報等を閲覧することで、情報コンテンツの信頼性の判断を行う方法。