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 成果報告書

2025年2月28日 エヴィクサー株式会社

目次

1. 実証事業の概要	
1-1. 実証概要のサマリ	• • • 5
1-2. 事業の目的	••• 6
1-3. 開発する技術の概要	••• 7
1-4. 社会実装のための実証の詳細	•••• 11
1-5. 実施計画の詳細	•••• 13
2. 実証事業の成果	
2-1. 開発した技術・ツールの詳細	••• 18
2-2. 社会実装のための実証の結果	•••• 28
2-3. 本実証後の展望	•••• 40

目次

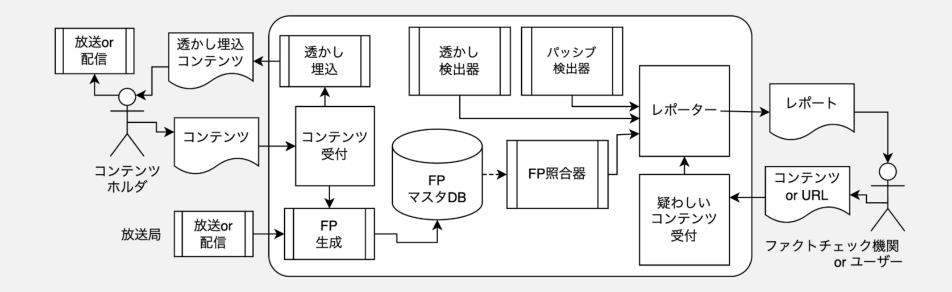
1. 実証事業の概要	
1-1. 実証概要のサマリ	••• 5
1-2. 事業の目的	••• 6
1-3. 開発する技術の概要	••• 7
1-4. 社会実装のための実証の詳細	•••• 11
1-5. 実施計画の詳細	•••• 13
2. 実証事業の成果	
2-1. 開発した技術・ツールの詳細	••• 18
2-2. 社会実装のための実証の結果	•••• 28
2-3. 本実証後の展望	•••• 40



実証概要のサマリ

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

実施体制	エヴィクサー株式会社	対象とする偽・ 誤情報	音声、映像コンテンツの音声トラック	
対策技術	音響透かしを用いた真正性保証・信頼性判断支援技術と音響フィンガープリントを用いた真 偽判別支援技術	目標	実証段階では技術の実証、システム構築と関係機関とのヒアリ ングを実現する。関係機関への採用を進め、中長期的に真正 性の確認を一般化し、正しい情報の流通を確保する。	
実証概要	ネット上の偽・誤情報に対して、偽画像への対策は早急に進められているが、音声に対しては対策が遅れていると思われる。本提案では音響透かしと音響フィンガープリントを用いた総合的なシステムを提供し、これにより音声情報や動画の音声に対して、情報の真正性、信頼性の判断を行うことが容易になると期待される。現時点で知られているAI生成データやディープフェイクに対して、真正性・信頼性を確保できるか実証を行い、ユーザビリティの良いシステムを構築することで、広く社会実装可能であることを実証する。中長期的には、メディア、公的機関(国・自治体)、プラットフォーム事業者等への普及の促進、ファクトチェック機関や国民への情報確認の促進を目標とする。広く公的な機関で採用がなされた後には、政治的な情報の発信、災害時の情報発信の際の信頼性担保に有効に働くと考えている。			





事業の目的

事業の目的

新たなシステム開発を行い、疑わしい音声、映像コンテンツの音声トラックに対する偽・誤情報への判断材料を提供することで適切な情報の流通を促す。

メディア、公的機関(国・自治体)、プラットフォーム事業者等に対してコンテンツの来歴情報等の埋め込みと検出機能を提供することで真正性の保証を支援する。

同時に音響フィンガープリントを利用したマスタデータベース(以下、マスタ DB)を構築することで、疑わしいコンテンツを照合する機能も提供することで 信頼性と真偽の判断を支援する。

透かしの検出もされず、マスタDBにもマッチしない疑わしいコンテンツに対しては、音響フィンガープリントを利用してパッシブ検出を行い、真偽判別を支援する。

ファクトチェック機関や国民に疑わしいコンテンツの検出機能の利用を促し、インターネット上の偽・誤情報への判断材料を与え、適切な情報・コンテンツの流通に貢献する。

偽・誤情報への技術的対策の課題の状況、今後目指す姿

- ・音響透かしを用いてコンテンツに来歴情報を埋め込み、真正性・信頼性を 保証する技術の課題は、実際のユースケースでの信頼性実証とユーザビリティの高いシステム構築である。これには、利用しやすいサーバーシステムを構築し、広く利用を促進する環境を整える必要がある。
- ・音響フィンガープリントによるコンテンツの真偽判断支援技術の課題は、信頼性のあるマスタDBの構築である。信頼性のあるURLからの配信等のケースでは、クローリング等の手法で大規模なDBを構築していくことも必要となる可能性がある。
- ・音響フィンガープリントによる真偽判別支援は、生成AIによって合成された 音声の検出能力の向上がカギとなる。音声生成AIエンジンによる合成音声 の検出試験を行い、検出能力を高める。

広く検出行為が行われ、適切なコンテンツ流通が実現するシステムを構築するために、プラットフォーム事業者やファクトチェック機関との連携を促進する。 また、システムコストを見極め、機能限定版の開放を目指して広く利用を促進する必要がある。



開発する技術の概要 (1/4)

開発技術

音響透かしを用いて来歴情報等をコンテンツへ埋め込む真正性保証・信頼性判断支援技術の開発を行う。

音響透かしとは、音声信号に暗号化を施した文字情報などを埋め込む技術であり、音声そのものに聴こえないようにデータを埋め込む技術のため、ファイルヘッダへのメタデータ埋め込み等に比べるとデータの除去が難しく、高い対攻撃性能がある。

本実証では、音響透かしの技術を利用し、音声データおよび映像 コンテンツの音声部分に対して、来歴情報等のメタデータを埋め込む手法についての開発を行う。

本事業全体における当該技術の役割

ファクトチェック機関等は、信頼性が疑わしいコンテンツに対して音響透かしの検出を行うことで、そのコンテンツが真正なものであるか、元の素材からの改変が行われていないかを確認し、信頼性を判断することが可能となる。



開発する技術の概要 (2/4)

開発技術

音響フィンガープリントによる真偽判断支援技術の開発。

音響フィンガープリント技術は、音声データを信号処理により分析して、異なる内容をもつ2つの音声データ同士の区別が可能であるような最小限の情報を抽出する技術である。この抽出された情報は、ちょうど指紋から個人を特定できるように、個々の音声に対して特徴的な一致を示すことができるため、音響フィンガープリントと呼ばれる。この技術を用いることで、任意の機器で再生されたマルチメディアデータの内容を認識することが可能となる。

本実証では、音声または映像の音声部分から音響フィンガープリントを作成してマスタDBを作成しておき、疑わしいコンテンツをマスタDBと照合することによって、その真偽を判断する。

マスタDBの作成のため、製作者はあらかじめ音声映像データをアップロードする。あるいはテレビ放送局など恒常的に放送・配信される内容であれば、放映されると同時に音響フィンガープリントを作成し、音響フィンガープリントをシステム上のDBに保管する。

ファクトチェック機関等は信頼性が疑わしいコンテンツをシステムの音響フィンガープリントDBと照合することで、信頼性、真偽判断を行う。

本事業全体における当該技術の役割

信頼性が疑わしいコンテンツに対してフィンガープリントを算出してマスタDBとの照合を行うことで、そのコンテンツが真正なものであるか、元の素材からの改変が行われていないかを確認し、信頼性を判断することが可能となる。



開発する技術の概要 (3/4)

開発技術

音響フィンガープリントを用いた、パッシブな真偽判別支援技術の開発を行う。

この技術は、生成AIによって作成されたと疑われるコンテンツの音 声区間を音響フィンガープリント技術によって解析し、その不自然さ 判定することで、生成AIの使用の有無を判別する技術である。

透かしの検出もされず、マスタDBにもマッチしない疑わしいコンテンツに対しては、このパッシブ検出を行い、真偽判別を支援することが可能になる。前述の音響透かしと音響フィンガープリントのDBを用いた真偽判断と組み合わせて、パッシブ検出もサポートした複合的手法として開発を行う。

本事業全体における当該技術の役割

前述の音響透かしと音響フィンガープリントによるマスタDBを用いた真偽判断の方法は、透かしの埋め込み作業やデータベースへの登録が必須となるが、これらは放送局などの公共放送で使用されることを想定している。

一方、公共的な情報発信を旨としない一般国民が動画サイトなどで発信したデータの場合には、それに対する偽情報による攻撃が行われた際には、上位の技術では対抗できないため、パッシブ検出をサポートした複合的手法を提供する。

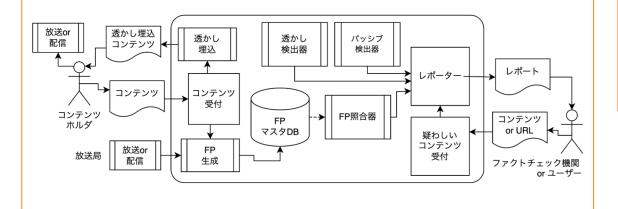


開発する技術の概要 (4/4)

開発技術

本提案において実用化を目指すものは、「音響透かしを用いた、 来歴情報等をコンテンツに埋め込む真正性保証・信頼性判断支 援技術」、「音響フィンガープリントによるマスタDBを用いた真偽判 断支援技術」、「音響フィンガープリントによるテレビ放映DBを用い た真偽判断支援技術」、「音響フィンガープリントによるパッシブな 真偽判別支援技術」の4つの技術を内包したWEBシステムであ る。

これらの基礎技術を用いた偽・誤情報対策技術の開発と、その有効性の検証を行い、社会実装に向けたユーザビリティーの高いシステムの設計、構築を行う。



本事業全体における当該技術の役割

本事業で構築するシステムはWEBシステムとしてクラウド上に構築されるため、かつてのオンプレ全盛時のような当初から莫大なコストがかかる構造とはなっておらず、事業の進行に合わせてシステムをスケールさせ、最適なコストで運用が可能である。使用される要素技術も全て自社開発のものであるため、検証が進む過程で要素技術に遡っての調整も可能であり、システム全体の開発費用が主たる技術的なコストとなるように事業設計されている。

ユーザーにとっても、上記のようなシステムの構築の考え方である ため最適化された費用で提供することが可能である。また、コンテンツホルダーにとっては自身の信頼性という金額では換算しにくい 非常に大きな資産を保護する効果を得られるため、費用対効 果は高いものになっている。



社会実装のための実証の詳細 (1/2)

社会実装のゴール

現時点で知られているAI生成データやディープフェイクに対して、真正性・信頼性を確保できるか実証を行い、ユーザビリティの良いシステムを構築することで、広く社会実装可能であることを実証する。

中長期的には、メディア、公的機関 (国・自治体)、プラットフォーム事業 者等への普及の促進、ファクトチェック機 関や国民への情報確認の促進を目標と する。

広く公的な機関で採用がなされた後には、政治的な情報の発信、災害時の情報発信の際の信頼性担保に有効に働くと考えている。

社会実装に向け、必要な検討事項

技術的対策だけでなく、社会的観点からも偽・誤情報への対策を強化する必要がある。たとえば、次のような取り組みが考えられる:

- ●ファクトチェック機関との協力:独立したファクトチェック機関と連携し、情報の正確性を検証する体制を整備する。
- ●セミナーの開催: 偽・誤情報に対する認識を広めるため、企業や自治体向けにセミナーを開催する。
- ●クラウドサービスの提供:情報の正確性を確認するためのクラウドベースの検証サービスを提供する。

現時点での課題として、以下の点が挙げられる:

- ●技術の普及には時間とコストがかかる
- ●誤情報と悪意ある情報の区別が難しい
- ●善意ある情報を尊重しながら、悪意ある情報を排除する方法の確立 これらの課題を克服するためには、法律、技術、社会の各側面からの包括的な取り組みが必要 である。将来的には、法律的枠組みの整備(より迅速・的確なインターネット上の偽・誤情報の 削除手続の整備等)と技術的進歩(より確実な偽・誤情報の特定・証明手段の開発等)の 融合によって、偽・誤情報に対する効果的な対策が実現することが期待される。



社会実装のための実証の詳細 (2/2)

実証内容

- (ア) 技術検証に関する内容:
 - ① 同社の特定チャンネルの全番組につき、音響透かしを埋め込んで放送・配信
 - ② 同社が放送・配信する複数チャンネルにつき、音響フィンガープリントの生成
 - ③ 音響透かしを埋め込んだ放送・配信につき、一般 視聴者環境のスマートフォンで音響透かしを検出可能かどうかの検証
 - ④ 対象チャンネルにつき、一般視聴者環境のスマートフォンで音響フィンガープリントによるマッチングが可能かどうかの検証
- (イ) 普及啓発に関する内容:
 - ① テレビ放送局/放送スポンサー等にとっての導入 インセンティブ検討
 - ② 本取り組みに関する関係者へのヒアリング

実証パートナー

ダイバーシティメディア (東北ケーブルテレビネットワーク)

想定ゴール

現時点で知られているAI生成データやディープフェイクに対して、真正性・信頼性を確保できるか実証を行い、ユーザビリティの良いシステムを構築することで、広く社会実装可能であることを実証する

とりわけ、メディアにとっての導入インセンティブを検討する

開発する下記三点の技術からもたらされるアウトプットの法的 留意点の整理

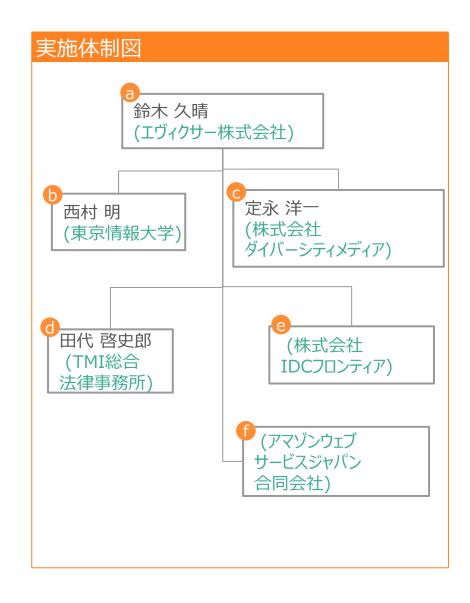
- 音響透かしを用いて来歴情報等をコンテンツへ埋め込む 真正性保証・信頼性判断支援技術
- 音響フィンガープリントによるコンテンツのマスタDBを用いた 真偽判断支援技術
- 音響フィンガープリントによるコンテンツの真偽判別支援

TMI総合法律事務所

開発システムからもたらされるアウトプットが「情報の真正性、信頼性の判断」材料とされる場合、法的根拠を明確にした上で、「ノーティス・アンド・テイクダウン」等の手続きの有効性を整理する



実証の実施体制



団体名 役割及び責任

a エヴィクサー 株式会社

(実施主体)本事業全般の管理・統括業務 音響透かし、音響フィンガープリント技術の開発・提供

b 西村 明 (東京情報大学)

実証内容への技術的、社会的アドバイス

相談役

C株式会社 ダイバーシティメディア

技術的対策の有効性等に関する検証

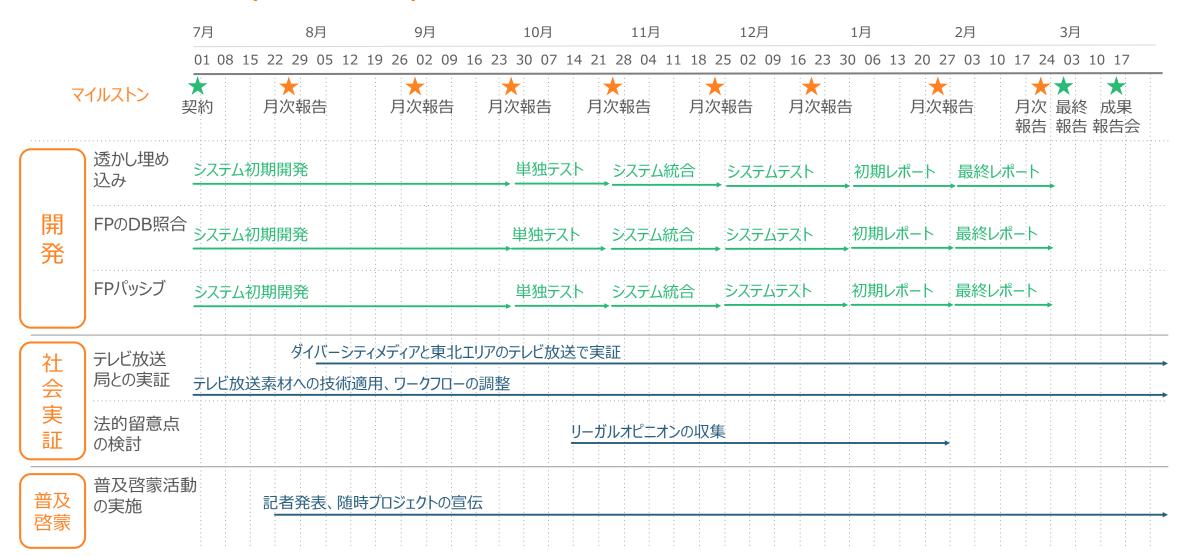
dTMI総合 法律事務所

法的留意点のアドバイス

e 株式会社 IDCフロンティア

②株式会社 データセンターの提供

実施計画 (スケジュール)





実証スケジュール (月次マイルストーン 1/2)

計画(当該期間に実施予定の事項)

~7/22週

- テレビ放送局と実証内容の調整
- 初期環境構築
- 機材調達

~8/26週 (夏季休暇想定で5w)

- テレビ放送局との記者発表(7/23、山形)
- テレビ放送局での実証開始
- システム初期開発

~9/23逓

- テレビ放送局での実証(前月に続き)
- システム開発

~10/21週

- テレビ放送局での実証(前月に続き)
- システム単独テスト



実証スケジュール (月次マイルストーン 2/2)

計画(当該週に実施予定の事項

~11/18週

- テレビ放送局での実証(前月に続き)
- リーガルオピニオンの収集開始
- テレビ放送局関係者へのヒアリング
- システム統合

~12/16週

- テレビ放送局での実証(前月に続き)
- リーガルオピニオンの収集(前月に続き)
- テレビ放送局関係者へのヒアリング(前月に続き)
- システム統合テスト

~1/20週 (年末年始想定で5w)

- テレビ放送局での実証(前月に続き)
- リーガルオピニオンの収集(前月に続き)
- テレビ放送局関係者へのヒアリング(前月に続き)
- 初期レポート

~2/17调

- 成果報告書の作成開始
- テレビ放送局での実証(前月に続き)
- テレビ放送局関係者へのヒアリング(前月に続き)

最終報告

成果報告書の作成

目次

1. 実証事業の概要	
1-1. 実証概要のサマリ	••• 5
1-2. 事業の目的	••• 6
1-3. 開発する技術の概要	••• 7
1-4. 社会実装のための実証の詳細	••• 11
1-5. 実施計画の詳細	•••• 13
2. 実証事業の成果	
2-1. 開発した技術・ツールの詳細	•••• 18
2-2. 社会実装のための実証の結果	•••• 28
2-3. 本実証後の展望	•••• 40



開発した技術・ツールの詳細 (1/10)

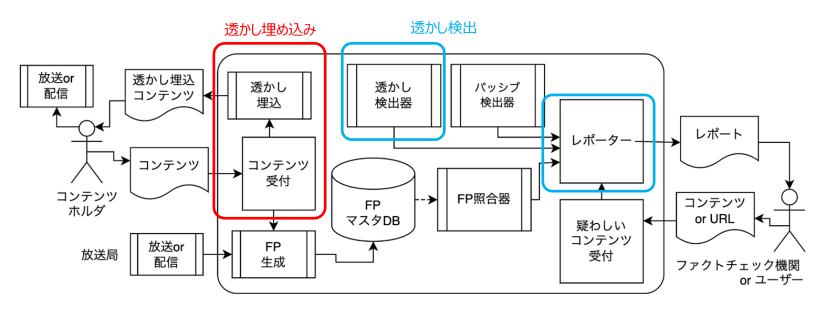
開発技術の詳細内容

音響透かし検出

音響透かしの技術を利用し、音声データおよび映像コンテンツの音声トラックに対して、来歴情報等のメタデータを透かしとして埋め込む手法についての開発を行った。

放送局などのコンテンツホルダは、放送するコンテンツにあらかじめ透かし埋め込み処理を行ってから放送を行う。

ファクトチェック機関等は、信頼性が疑わしいコンテンツに対して透かしの検出を行うことで、そのコンテンツが真正なものであるか、元の素材からの改変が行われていないかを確認し、信頼性を判断することが可能となる。





開発した技術・ツールの詳細 (2/10)

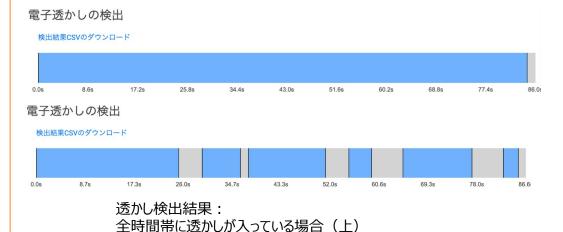
R5補正における技術開発の成果

音響透かし検出

音響透かしの検出レポートでは、検査対象の動画の全体に対して音響透かしが入っている箇所と入っていない箇所をグラフで確認することができた。

下図で、青色の部分が音響透かしが入っている部分、灰色の部分が音響透かしが入っていない部分である。このグラフの灰色部分の存在により、アップロードされた動画の音声が、放映された透かし入り音声ではない可能性が示唆される。

透かし検出結果の詳細データについては下記のようなcsv形式でもダウンロードすることができ、専門家による詳細な検証も可能である。



部分的にか透かしが入っていない場合(下)

fromTime,toTime,wm_value,name

1.3,21.4,10,"embedded_石破内閣総理大臣記者会見一令和6年10月1日 [b_Ebl-pMNVA].mp4"

3

csvデータの例



開発した技術・ツールの詳細 (3/10)

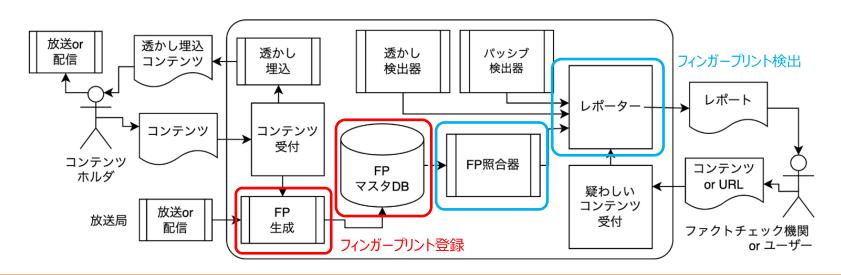
開発技術の詳細内容

大規模マッチング(マスターDBマッチング)

この技術は、音声または映像の音声部分から音響フィンガープリントを作成して、あらかじめマスタDBを作成しておき、疑わしいコンテンツをマスタDBと照合することによって真偽を判断する技術である。

マスタDBの作成のため、製作者は音声映像データをクラウド上にアップロードする。

登録フィンガープリントの検出では、検査対象の動画と、あらかじめDB登録が行われている動画を比較し、一致する区間を検出することが可能である。また、検査対象の動画が複数の動画からの切り貼りにより作成されている場合でも、複数の動画の複数の区間をそれぞれ全て検出することが可能である。





開発した技術・ツールの詳細 (4/10)

R5補正における技術開発の成果

大規模マッチング(マスターDBマッチング)

EAFのCMSから動画をアップロードすると、動画の音声に透かしが埋め込まれると同時にフィンガープリントデータが生成され、マスタDBに登録される。

この処理は1時間の動画でも数分程度で完了する。

登録フィンガープリントの検出結果の詳細データについてもcsv形式でダウンロードすることができ、 専門家による詳細な検証も可能である。

大規模マッチングでは、マッチングした動画の「ファイル名」、「時間データ」の2つのメタデータが検出できる。提出した動画の0.7~59.3秒が、マスターの「年頭記者会見」動画の728.8~787.3秒と一致したことがわかる。



マスタDBに登録されたFPとのマッチングでは、 ほとんどの区間で一致し、 時間の編集も行われていなかったことがわかる。

登録動画との一致

マスターファイル	マスターと一致した秒数	詳細データ	時間差替
石破内閣総理大臣年頭記者会見一令和7年1月6日 [o7lAtutSql4].mp4	58.6 秒	CSVダウンロード	なし



開発した技術・ツールの詳細 (5/10)

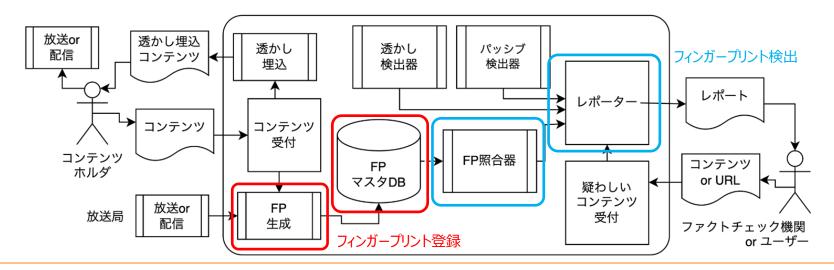
開発技術の詳細内容

大規模マッチング(TV放送マッチング)

本システムに登録がなされていない動画に関しても、全国のテレビ放映の音声のフィンガープリントをDBに収集しておくことで、一致区間の検出が可能である。検査対象の動画が複数の動画からの切り貼りにより作成されている場合でも、複数の動画の複数の区間をそれぞれ全て検出することが可能である。

FPの生成には、弊社独自の技術によるFP生成機が用いられた。

本事業の目的のために小型化・組み込み化され改良されたFP生成機が作成され、東京と大阪の2ヶ所のデータセンターに設置され、両地域のすべてのテレビ番組に対応するように設定された。





開発した技術・ツールの詳細 (6/10)

R5補正における技術開発の成果

大規模マッチング(TV放送マッチング)

マッチング対象のデータベースは過去30日間のテレビ(東京7局)の放送音のフィンガープリントのDB。

テストに使用した動画は、石破総理のスピーチの順序を入れ替えて、意図的に時間の編集を加えた動画。動画を切り貼りして、発信者が意図しない文脈を生成されることを想定している。

マスタDBに登録されたFPとの比較では、ほとんどの区間で一致しているが、時間編集ありと判別されている。検出結果の詳細データについてもCSV形式でダウンロードすることができ、専門家による詳細な検証も可能である。このように、全ての音声がマスタと一致しても、時系列の入れ替えを検出できることが示された。

テレビ放送音との一致

テレビデータとの分析CSVデータのダウンロード

ch	一致した放送日時開始	一致した放送日時終了	素材認識開始秒	素材認識終了秒
NHK	2025/01/06 14:11:19	2025/01/06 14:11:21	0.48	3.36
NHK	2025/01/06 14:11:22	2025/01/06 14:11:35	4.32	16.8
NHK	2025/01/06 14:11:36	2025/01/06 14:11:39	17.76	20.64
NHK	2025/01/06 14:11:40	2025/01/06 14:11:46	21.6	28.32
NHK	2025/01/06 14:11:48	2025/01/06 14:12:05	30.24	46.56
NHK	2025/01/06 14:12:06	2025/01/06 14:12:09	47.52	51.36
NHK	2025/01/06 14:12:10	2025/01/06 14:12:17	52.32	59.04
NHK	2025/01/06 19:04:22	2025/01/06 19:04:24	55.456	57.376
ETV	2025/01/06 20:49:26	2025/01/06 20:49:28	55.296	57.216
NHK	2025/01/07 05:02:53	2025/01/07 05:02:55	55.104	57.024



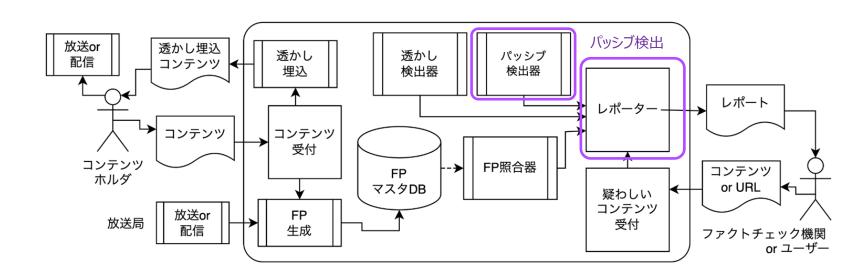
開発した技術・ツールの詳細 (7/10)

開発技術の詳細内容

生成AIによる音声かどうかのパッシブ検出

検査する音声が、生成AIや合成技術で作られた音声であるかの判別を行う技術である。 事前のDB登録などは必要としない。

合成音声判定では、検査したい音声が含まれた動画をサーバーにアップロードすることで、合成音声の可能性の有無を表示する。 提出された動画のうち、一部分でも合成音声の可能性のある箇所があれば、「合成音声の可能性があります」と表示する。





開発した技術・ツールの詳細 (8/10)

R5補正における技術開発の成果

生成AIによる音声かどうかのパッシブ検出

会見の様子をそのまま切り取った、オリジナル動画の場合、 全ての区間で合成音声は検出されないため、「合成音声ではありません」と 表示される。

全ての音声を合成音声に差し替えた偽動画の場合、 全ての区間で合成音声が検出されたため、「合成音声の可能性があります」 と表示される。

一部の音声を合成音声に差し替えた偽動画の場合、 合成音声が検出された区間が一部であっても、「合成音声の可能性があります」と表示される。



合成音声判定

合成音声ではありません

合成音声判定

合成音声の可能性があります



開発した技術・ツールの詳細 (9/10)

開発技術の詳細内容

4種類の検出方法を統合した利用しやすいUIを持つクラウドシステム

前述の4種類の検出方法を統合したシステムをクラウド上で提供するためのWEBシステムを開発した。

WEBシステム全体としては、主に以下の機能を提供するものである。

- 1. コンテンツのアップロード機能
- 2. アップロードされたコンテンツへの音響透かしでの来歴情報等の埋め込み機能
- 3. 来歴情報等を埋め込んだコンテンツのダウンロード機能
- 4. アップロードされたコンテンツの音響フィンガープリント生成機能
- 5. 生成された音響フィンガープリントのマスタDBへの登録機能
- 6. 疑わしいコンテンツの来歴情報検出、レポート機能
- 7. 疑わしいコンテンツのマスタDBとの照合、信頼性レポート機能
- 8. 疑わしいコンテンツの音響フィンガープリントを用いたパッシブな真偽判断、レポート機能
- 1~5は放送局・配信者が使用する機能であり、
- 6~8はファクト機関またはユーザーが使用する機能である。



開発した技術・ツールの詳細 (10/10)

R5補正における技術開発の成果

4種類の検出技術を統合したクラウドシステム

右図は透かし入り放映動画の作成および検出のためのインターフェース画面である。下記の動画への透かし音声の埋め込みや検出は全てWEBブラウザからクラウドにアクセスして操作を行うことが可能である。

透かしの埋め込みやフィンガープリントの作成などの技術的な操作は全て自動化されており、ユーザーは簡単な操作でシステムを使用することができる。

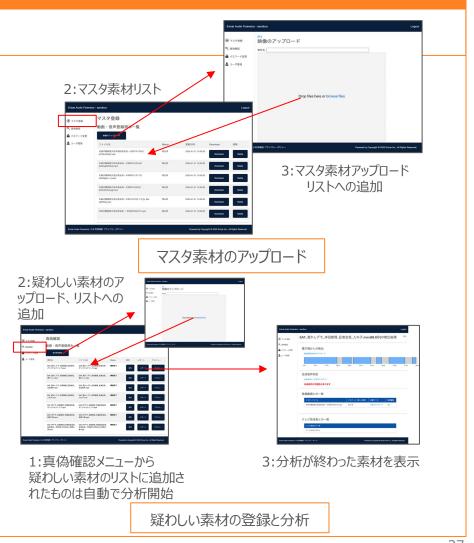
上記のようなフェイクニュースに対処するために、

放送局のようなメディア事業者は、本WEBシステムを以下のように利用する。

- 1. 放映前に自社の番組映像ファイルをシステムにアップロードする
- 2. 自動で編集・登録処理が行われる。
- 3. 透かしが埋め込まれた番組映像ファイルをシステムからダウンロードし、放送する。

ファクトチェック機関や国民は、本WEBシステムを以下のように利用する。

- 1. 疑わしいコンテンツのファイルをシステムにアップロードする。
- 2. システムから以下のレポートを得て、真正性、信頼性、真偽を判断する。





社会実装のための実証の結果 (1/12)

R5補正における社会実装の成果

放送メディアによる運用実証の結果

本システムでは、透かし埋め込みやフィンガープリントのDB登録の際に、テレビなどの放送メディアにおいて、放映前の動画データ取扱が必要となる。そのため、本システムが一般に認められ社会に普及するためには、放送メディア各社に対してシステムの導入にあたって何らかのメリットを提示し、導入コストをペイできるようなインセンティブを提供するのが有効であると考えられる。

以下は、山形県のケーブルテレビ放送局であるダイバーシティメディア及びゼクサバースの協力により行われた、透かし入り音声の放映の実証結果である。これにより、透かし入り音声が放送コンテンツとして有効であり、実用性を持って運用することが可能であることが実証され

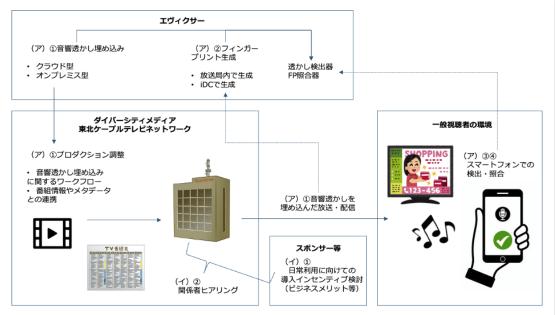
た。

●関係各社

ゼクサバース: chオーナー、番組の制作(調達) ダイバーシティメディア: 放送局(ケーブルテレビ) エヴィクサー(子会社AiryMedia): システムの提供

●備考

2024年7月23日に山形県山形市にて、上記3社で、合同記者会見を行った。





社会実装のための実証の結果 (2/12)

R5補正における社会実装の成果

放送メディアによる運用実証の手順

●手順

ゼクサバースが番組の買い付け、制作を行い、ダイバーシティメディアに納品する。

ダイバーシティメディアが本システムのCMSにmxf形式の素材をアップロードし、透かしを埋め込み、同時にDBへの登録を行う。ダイバーシティメディアは透かし入りの動画を放送サーバーに埋め込み、番組がケーブルテレビで放送される。

視聴端末(視聴者のテレビなど)で番組を受信し、スマホアプリなどで検出することにより、透かし音を用いた付加的なサービスを提供できることを実証する。

●検証内容

放送サーバーが受け入れられる素材が作成できることの確認放送時のスタッフの運用負荷の見極め

●検証結果

透かし入り番組の放送と検出は問題なく稼働した。ZEXA TVは事前に1ヶ月分の番組を登録しておく形で運用されており、本システムによるCMSベースのワークフローとは相性がよかった。

「ダイバーシティメディアが本システムのCMSにmxf形式の素材をアップロードし、透かしを埋め込み、同時にDBへの登録を行う。ダイバーシティメディアは透かし入りの動画を放送サーバーに埋め込み、番組がケーブルテレビで放送される。」という一連の手順に関する、素材あたりの費用イメージについて具体的に知見が得られた。



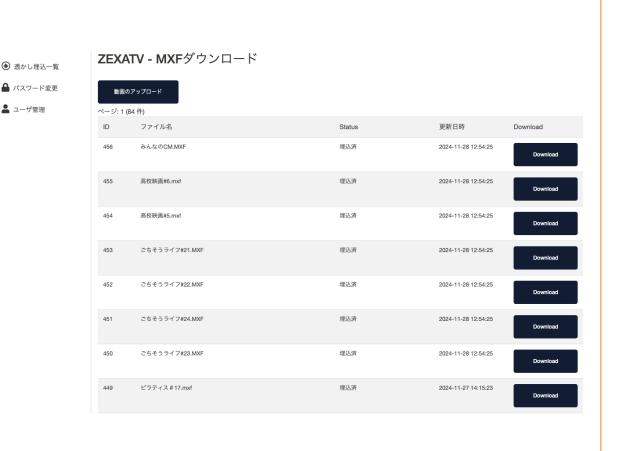
社会実装のための実証の結果 (3/12)

R5補正における社会実装の成果

放送メディアによる運用実証の手順

下記は、実際に使用されている、動画の音声部分に透かしを埋め込むCMSのシステム画面である。放送局が、放送用の動画ファイルをCMSにアップロードすると自動的に透かしが埋め込まれる。埋め込み処理にかかる時間は、元の動画素材の時間の1/5程度であった。その後、透かしの入った動画ファイルをダウンロードして放送を行う。

処理後の動画には可聴周波数範囲に聴き取ることのできない 透かしが入っており、番組ごとに異なる透かし番号が入っているため、スマホアプリなどで透かしを検出することにより放映されている番組を特定することが可能である。





社会実装のための実証の結果 (4/12)

R5補正における社会実装の成果

• 放送メディアによる運用実証の結果

ダイバーシティメディアでは、本実証のために「ZEXA TV」というチャンネルを開設して放映を行った。ZEXA TVは、新たなインタラクティブな 視聴体験を実装したチャンネルであり、放送される番組やCMにスマートフォンをかざして専用のブラウザから音声をキャッチすると、様々な情報を届けたり、クーポンやプレゼントなどか当たる「ZEXA TVくじ」を引くことができるチャンネルである。 下記が実際に放映された番組表である。これら全ての番組に音響透かしが埋め込まれ、放映された。

システムの導入としては、2024年7月からベータ版を提供し、運用上の問題点を観測し、仕様に少しずつ反映していった。放映された透かし埋め込み番組の本数は下記の通りである。

7月:30分番組99本

8月:30分番組62本、60分番組4本、120分番組1本

9月:30分番組48本、60分番組5本

10月:30分番組37本、60分番組7本

11月:30分番組36本、60分番組7本

透かし入り放送は、技術的な検証のみならず、 付加的なサービスの提供という社会的実証としても、 十分に放送メディアに導入可能であるということを実証することができた。





社会実装のための実証の結果 (5/12)

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。



社会実装のための実証の結果 (6/12)

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。

2-2

社会実装のための実証の結果 (7/12)

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って、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 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。)。

2-2

社会実装のための実証の結果 (8/12)

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、その著名人の音声が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。



社会実装のための実証の結果 (9/12)

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。

2-2

社会実装のための実証の結果 (10/12)

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が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。

2-2

社会実装のための実証の結果 (11/12)

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ので留意が必要である。

2-2

社会実装のための実証の結果 (12/12)

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。コンテンツに一意の 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。



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	適用範囲の拡大	ビジネスとしての自走化
前年度の開発で開拓した技術的課題の解決、性能向上開発システムの提供形態、プライシングモデ	前年度の開発で開拓した技術的課題の解決、性能向上	前年度の開発で開拓した技術的課題の解決、 性能向上
・ 開発ラステムの提供が高、フライラングモデルの実証実験・ 単なるライセンスモデルだけでなく、前年度までに検討した導入インセンティブを加味したビ	開発システムの提供形態、プライシングモデルの完成リードユーザーの獲得	プロモーションも含めて導入先の拡大を行い、収益化を実現していく
 では検討した等人インセンティンを加味したと ジネスモデルも検討 提供先の要望に応じたアルゴリズムのカスタマ イズ 国際標準化を目指すために、標準開発組 	・ カードユーケーの返得○ テレビ/ラジオ放送局○ 政府、地方自治体○ YouTuberやVTuber等の動画配信者	 国際標準化への標準開発組織(SDO)と連携し、技術標準・ベストプラクティス開発に貢献 放送局・メディアのヒアリング結果を基に、事業終了後に即時採用活動へ移行
織(SDO)との連携検討	○ 音楽団体	 本事業の推進が賛同機関の増加を促進 公的ファクトチェック機関も同様のアプローチで道

- 特許戦略の検討・実施
 - 予備審査を行っている特許(出願番号 PCT/JP2021/042416)についての各国移行の検討
 - 新たな出願検討

- 国際標準化を目指すために、標準開発組織(SDO)との連携検討
- 一般コンシューマーが活用できるUI・UXを実現したサービス、アプリケーションの提供
- 公的ファクトチェック機関も同様のアプローチで導入を進める予定

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