

ホワイトペーパー

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について AI 生成物のラベリングとディープフェイク対策

This paper discusses the growing threat of misinformation and disinformation in the age of generative AI and social media, focusing on their societal risks and the challenges of verifying the authenticity of audio content. It introduces Evixar Audio Forensics (EAF), a technological approach that embeds forensic information directly into audio, enabling real-time verification without affecting the listening experience. The paper also situates EAF within broader national and international policy frameworks, including initiatives by Japan's Ministry of Internal Affairs and Communications and the Hiroshima AI Process. Furthermore, it emphasizes that EAF should be positioned not only as a technical solution but also as a trust infrastructure for the future distribution of digital content.

瀧川 淳

(エヴィクサー株式会社 代表取締役社長 CEO)

日本開発工学会 第8回研究発表大会 発表より転載

2025 年 6 月



Hear Deeper.
耳をすまそう。

1. はじめに

近年、SNS の普及と生成 AI^(注1) の急速な進化が相まって、インターネット空間上の偽・誤情報やディープフェイクへの対応・対策が社会的課題として浮上している。とりわけ、音声・映像を伴う情報（コンテンツ）は、SNS を通じて拡散しやすく、災害時や選挙などに深刻な影響を与える事例が報告されている。本稿では、筆者が代表を務めるエヴィクサー社で開発した音響技術を用いたコンテンツの真正性保証や真偽判別手法について、新たなコンテンツテクノロジーの研究テーマとして取り上げ、2024 年 7 月から取り組む総務省の開発・実証事業における成果、そのビジネスニーズのあぶり出しと包括するパーパスについて報告する。

2. 偽・誤情報とディープフェイク

2.1 偽・誤情報とは

総務省が推進する ICT 活用リテラシー向上プロジェクトによれば、「偽情報」とは、個人、社会集団、組織または国に危害を与えるため、意図的・意識的に作られた虚偽の情報のことを指し、「誤情報」とは、危害を引き起こす意図で作成されたものではなく、勘違い・誤解により拡散した間違い情報のことと定義されている。さらに、関連する法律等も含めた相関は図1のように整理される。

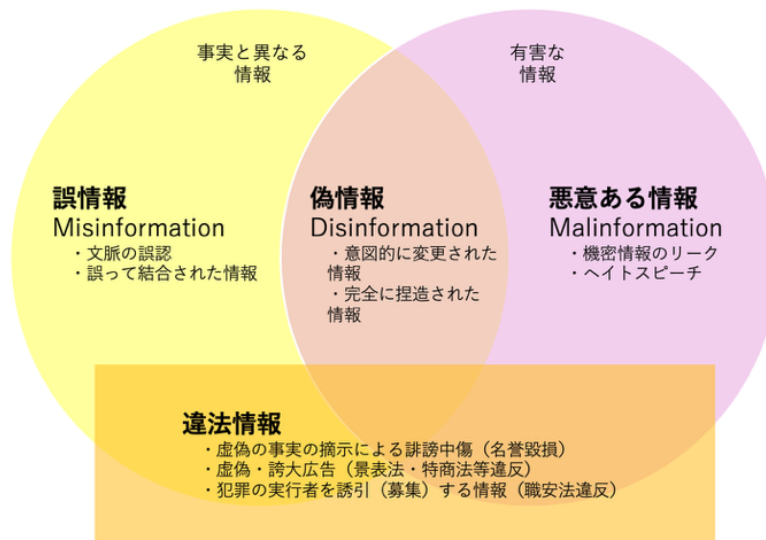


図1 情報の分類、偽・誤情報等の相関

2.2 ディープフェイクとは

ディープフェイク（deepfake）とは、「ディープラーニング」と「フェイク」を組み合わせた造語であり、機械学習アルゴリズムの一つであるディープラーニング（深層学習）を応用して、人物の顔や音声などを本物のように合成・置換することで、オリジナルとは異なる画像・動画・音声を人工的に生成する技術や行為、またはそれによって作成されたコンテンツを指す。特に2022年頃から、テキスト・画像・音声・動画など異なる種類のデ

ータを同時に処理し組み合わせ、より高度な情報処理を実現するマルチモーダルAIの台頭により、SNSなどで流通する現実の映像や音声・画像の一部を加工して、虚偽の情報を組み込み、あたかも本物や事実のように見せかけることがより精巧かつ容易になっている。

1)

3. SNSとアテンション・エコノミー

情報通信白書（令和5年度版／2023年）で、情報過多の社会においては、供給される情報量に比して、消費者が支払える注意・関心（アテンション）ないし消費時間が希少となるため、それらが経済的価値を持って市場（アテンション・マーケット）で流通するようになり、こうした経済モデルは、一般に「アテンション・エコノミー」と呼ばれると指摘されている。（図2参照）

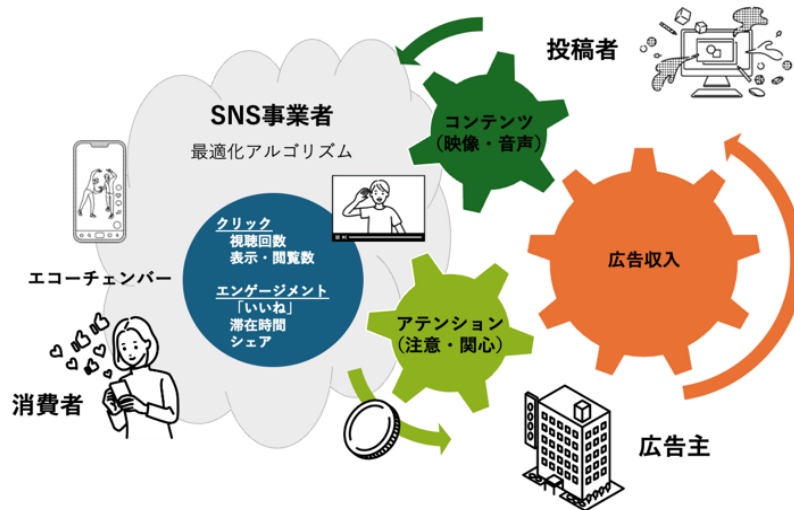


図2 アテンション・エコノミーの概念

SNSの運営者であるプラットフォームは、可能な限り多くの時間、多くのアテンションを獲得するため、データや最適化アルゴリズムを駆使してその利用者が「最も強く反応するもの」を予測する。この作用はエコーチェンバーと呼ばれ、消費者はSNSを利用する際、自分と似た興味関心を持つ他のユーザーをフォローする結果、意見をSNSで発信すると自分と似た意見が返ってくるという状況を加速させ、閉じた小部屋で音が反響する物理現象に例えられている。消費者は過激な意見に繰り返し触れる一方で多数の人が同じ意見を支持していると聞かされれば信じ込む人が生じる。さらに、SNSの無料サービスモデルは、ユーザーの視聴回数や接触回数が多ければ多いほど大きな対価を得られる仕組みで、真偽に関わらず注目を集めるコンテンツが優先される傾向に拍車を掛け、偽・誤情報が経済的インセンティブによって拡散される土壌となっている。災害時の偽情報拡散や選挙期間中の候補者なりすましディープフェイクの事例を鑑みれば、情報（コンテンツ）の真正性保証や真偽判別はもはや単なる技術的課題にとどまらず、社会インフラの課題として扱うべきは明らかであろう。

4. AIに関する国際的・日本政府の動向

4.1 世界初のAI包括的ルール「広島AIプロセス（注²）」

2023年5月、G7広島サミットにおいて合意された「広島AIプロセス」は、生成AIによるリスク対応の国際的枠組み形成の第一歩となった。各国は、AIによって生成されたコンテンツの透明性向上と信頼性担保に向けた指針策定を進めており、今後の国際標準化の基盤ともなり得るであろう。とりわけ、2023年12月にその成果として策定された全11条からなる「高度なAIシステムを開発する組織向けの広島プロセス国際指針」において、コンテンツに関する技術的な方針として、第7条で下記のような原則が定められている。

（引用はじめ）

「7. 技術的に可能な場合は、電子透かしやその他の技術等、ユーザーがAIが生成したコンテンツを識別できるようにするための、信頼できるコンテンツ認証及び来歴のメカニズムを開発し、導入する＝これには、適切かつ技術的に実現可能な場合、組織の高度なAIシステムで作成されたコンテンツのコンテンツ認証及び来歴メカニズムが含まれる。来歴データには、コンテンツを作成したサービス又はモデルの識別子を含めるべきであるが、ユーザー情報を含める必要はない。組織はまた、透かし等を通じて、特定のコンテンツが高度なAIシステムで作成されたかどうかをユーザーが判断できるツールやAPIの開発に努めるべきである。組織はさらに、可能かつ適切な場合には、利用者がAIシステムと相互作用していることを知ることができるよう、ラベリングや免責事項の表示等、その他の仕組みを導入することが奨励される。」

（引用終わり）

4.2 日本政府の取り組み

2024年4月、日本政府（経済産業省／総務省）は、生成AIの普及を始めとする近年の技術の急激な変化等に対応すべく、有識者等と議論を重ね、関連する既存のガイドラインを統合・アップデートし、「AI事業者ガイドライン」を発表するとともに、偽・誤情報に関する総合的な対策を提言²⁾し、「普及啓発」「技術的対応」「制度的対応」を三本柱としている。（図3参照）

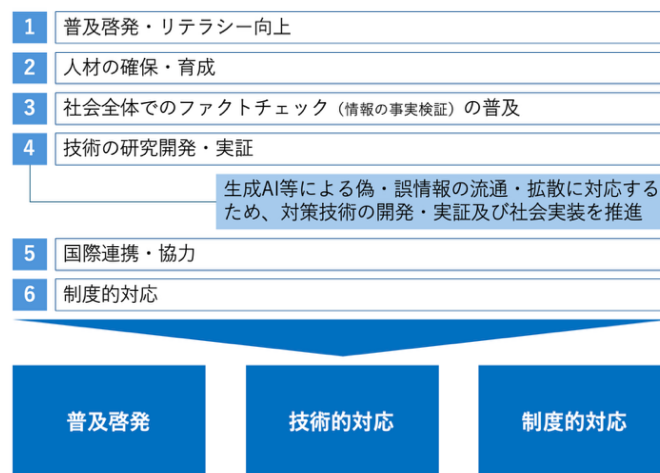


図3 総務省による「デジタル空間における情報流通の健全性確保のための総合的対策」

本稿では以降、上記の④技術の研究開発・実証の一部に位置付けられた「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6年度／2024年）^{（注3）}」に関し、エヴィクサー社の取り組みについて報告する。

5. 音響透かし^{（注4）}と音響フィンガープリント^{（注5）}を用いた偽・誤情報対策クラウドシステム

エヴィクサー社は、偽画像への対策は進められているが音声に対しては対策が遅れているとの問題意識に基づき、独自の音響技術を応用し、音声および映像コンテンツの音声トラックを対象として、「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」を提案した。（図4 参照）

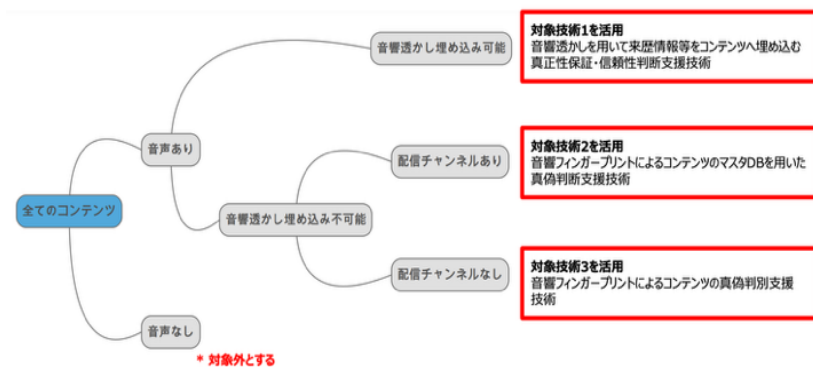


図4 2024年度に提案した三つの技術とコンテンツの相関

図4に示した三つの技術を束ねて Evixar Audio Forensics（以下「EAF」）と呼称する総合的なシステムとして開発すれば、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仮説を立てた。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し、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自身による情報の確認促進を目標とした。

2024年度の成果の一部として、EAFでアウトプットされるレポート例を紹介したい。（図5参照）

音響透かし（電子透かし）の検出レポートでは、検査対象の動画の音声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秒単位の精度で確認することが可能である。図5で、青色部分は音響透かしが検出された部分、灰色部分は音響透かしが検出されなかった部分である。このグラフの灰色部分の存在により、オリジナルの動画の音声に何らかの変更が加えられたことが示唆されている。合成音声判定のレポートでは、事前のDB登録などは不要であり、検査したい音声の含まれた動画がEAFにアップロードされると、内部アルゴリズムにより非自然性を検知し、合成音声が使用されているかどうかを判定する。アップロ

ードされた動画のうち、一部分でも合成音声の可能性のある箇所があれば、「合成音声の可能性あります」と表示される。

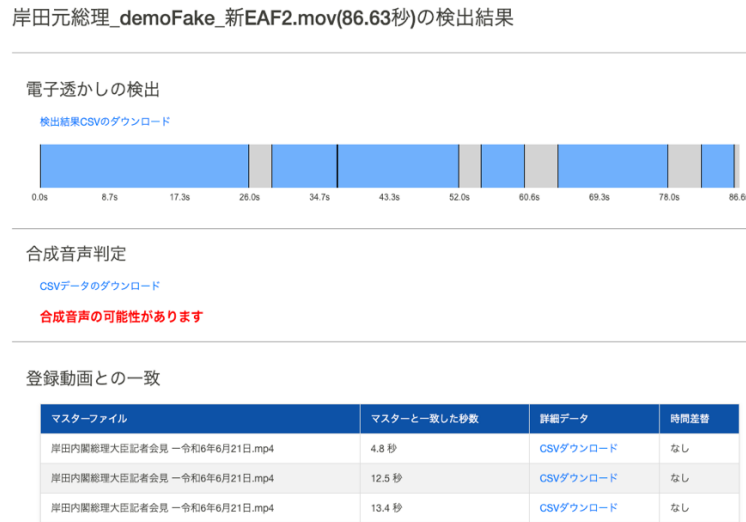


図5 EAFのレポート画面イメージ

また、ケーブルテレビ局との協力により、実際の放送・配信環境下で実証実験を行った。番組音声に透かしを埋め込み（ラベリング）、視聴者がスマートフォンを通じて透かしを検出することで、コンテンツの真正性を確認しつつ、一般消費者が参加出来るインセンティブを模索する内容（図6参照）とした。この実証スキームは六ヶ月以上にわたって運用され、事故なく稼働した。

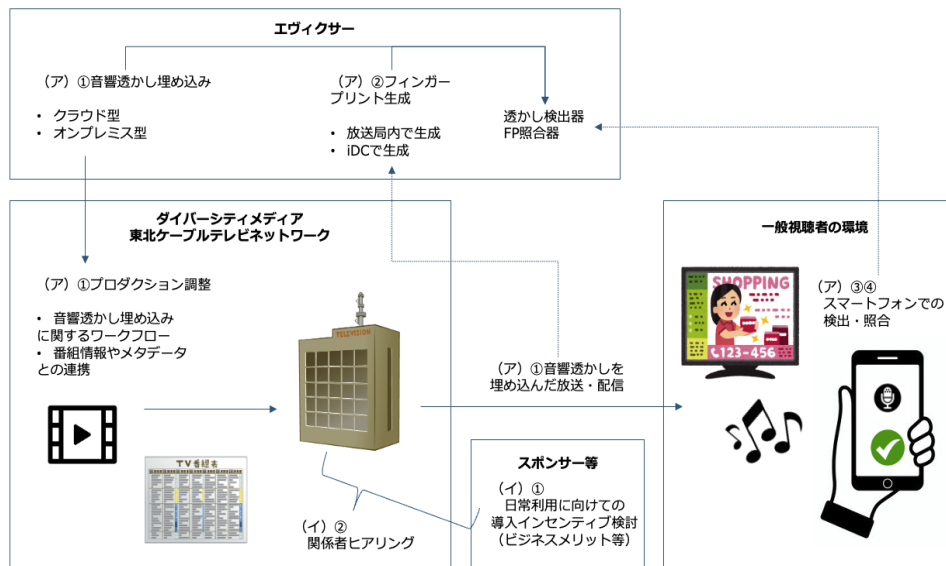


図6 2024年度にケーブルテレビ局で行った実証スキーム

6. おわりに

2024 年度の開発・実証により技術的な有効性は一定レベルにおいて確認されたものの、実際の社会実装に向けては複数の課題が存在する。第一に、放送・配信事業者に対する導入のインセンティブ設計が必要であり、放送収益モデルや広告主との連携が重要となる。第二に、一般消費者への認知と習慣化を促す仕組みが求められ、UX のさらなる向上と教育的アプローチが必要である。第三に、真正性情報の国際的な相互運用性を担保するため、C2PA^(注6) など国際標準を目指す団体や規格との整合性を高める開発が求められる。

今後は、技術的課題の解決・性能向上、リードユーザーの獲得、標準化機関（SDO）への提案、特許戦略の検討・実施等を同時並行的に進めていく。また、本取り組みの過程で、新しいビジネスニーズがあぶり出されている。具体的には、個人認証やタイムスタンプ等のトラストサービス提供事業者との事業提携、電話通話の音声に関する真正性保証・真偽判別を実現するため通信キャリアとの協業を進めている。総務省の提言通り、社会実装に向けては、技術的対応だけでは不十分で総合的な対策との連携を視野に入れる。

エヴィクサー社は 2023 年に CI を刷新し、「音を使いこなして、人の感じる力を進化させる」と図 7 のようにパーパスを掲げた。本取り組みは、まさに人間を取り巻く環境が同時並行的に多様化・不確実化・機械化していく中で求められる課題解決であり、視聴覚障害者にも大きな影響を与えるため、その対策はアクセシビリティの観点でも重要である。他事業で手がける視聴覚障害者向け情報保障サービス^(注7) との相乗効果も生み出すことで、パーパスへのアプローチを複層的に行っていく。



図 7 エヴィクサー社のパーパス

注釈

(注1) 生成 AI AI (Artificial Intelligence) とは人工知能と訳され、その中でも生成 AI とは特に大量のデータから学習し、テキスト・画像・音声・音楽・動画など、新しいコンテンツを生成できる AI を指す。

(注2) 広島 AI プロセス 2023 年の第 49 回先進国首脳会議 (G7 広島サミット) で立ち上げられた、人工知能 (AI) を巡るルール形成を協議する国際枠組みの一つで、各国の閣僚級が協議を行ったのち、2023 年 12 月に G7 首脳が、「全 AI 関係者向けの広島プロセス国際指針」を含む国際指針等を承認した。

<https://www.soumu.go.jp/hiroshimaiprocess/index.html>

(注3) 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 6 年度/2024 年) 総務省は、2024 年度、生成 AI に起因する偽・誤情報を始めとしたインターネット上の偽・誤情報の流通リスクに対応するため、事業者や研究機関等を対象に技術開発主体の公募を行った。エヴィクサー社の 2024 年度における本取り組みの詳細は総務省 Web サイト上で公開されている。

https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/taisakugijutsu_fy2024.html

(注4) 音響透かし 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る。音声そのもの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、高い攻撃耐性がある。本取り組みでは攻撃耐性を向上させるために、メタデータの埋め込みは高い周波数に埋め込む音響モデム方式ではなく、広い周波数帯域に拡散して埋め込む方式を採用する。

(注5) 音響フィンガープリント 短い時間区間ごとの音声を信号処理により分析して、異なる内容をもつ二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音声信号の内容を認識することが可能となる。

(注6) C2PA 「Coalition for Content Provenance and Authenticity」の略で、デジタルコンテンツの信頼性を確保するための技術標準化団体。特に、AI 生成画像やディープフェイクなどの偽情報や誤情報の拡散を防ぐことをミッションとしている。

<https://c2pa.org/>

(注7) 視聴覚障害者向け情報保障サービス エヴィクサー社では映画館における音声・字幕ガイドサービス「HELLO! MOVIE (ハロームービー)」を提供している。

<https://hellomovie.info/>

参考文献

- 1) 国際大学グローバル・コミュニケーション・センター (2024) 『Innovation Nippon 2024 偽・誤情報、ファクトチェック、教育啓発に関する調査研究』
- 2) デジタル政策フォーラム (2024) 『AI ガバナンスの枠組みの構築に向けて Ver2.0』