

コンテンツの信頼性を高める来歴情報提示技術



日本放送協会 放送技術研究所 ネットサービス基盤研究部

おおたけ こう
大竹 剛

1. はじめに

インターネット上では、フェイクニュースなどの偽情報・誤情報がSNS上で拡散し社会問題となっている。国政選挙や大規模な自然災害の発生時においては、偽情報・誤情報を含むコンテンツが数多く流通するため、ユーザーは膨大なネット上の情報から真実の情報を見極めるのが困難となり、安全・安心が脅かされる恐れがある。生成AIを用いると、誰でも簡単に放送局になりすました高品質な偽動画を作成することが可能であるため、放送局が発信する「正確な情報」と、偽情報の見分けが困難な状況となっている。その結果、ユーザーが「真実である」と信じてSNSで偽動画を拡散する、という事態が生じている。従来の電波による放送サービスの場合、誰が発信者であるかを知ることは容易であるが、インターネットによる放送サービスの場合、誰もが発信者になり得るオープンな世界であるため、発信者の特定は困難である。

NHKでは、インターネットにおいても放送と同様に信頼できる情報を確実に届けるため、来歴情報提示技術の研究開発を進めている。

2. 来歴情報提示技術

コンテンツの信頼性をユーザーが判断できるようにするため、コンテンツの出どころや制作過程などの来歴情報をコンテンツに埋め込み、ユーザーに提示する来歴情報提示技術が2018年ごろから検討され始めた。来歴情報が偽・誤情報対策として有効に機能するためには、インターネット上に流通する多くのコンテンツに来歴情報が付与されるとともに、来歴情報に関する認知が拡大することが求められる。そこで、来歴情報提示技術に関する共通技術仕様の策定

及び普及を目的として、コンテンツの出どころと認証に関する国際標準化団体C2PA (Coalition for Content Provenance and Authenticity) が設立され、技術仕様が公開されている。C2PAではコンテンツに対し、制作者が撮影・編集・配信などの各工程において来歴情報及びデジタル署名を付与する。ユーザー側では、デジタル署名の検証によりコンテンツと来歴情報の改ざん有無を確認するとともに、来歴情報や署名者の情報を基に、コンテンツの信頼性を判断する(図1)。

◆C2PAの概要

C2PAは、Adobe、Microsoft、BBCなどによって2021年に設立された。放送局やカメラメーカー、IT企業など、コンテンツの制作や配信に関わる、インターネットにおける信頼できる情報発信に関心のある国内外の企業が多数参加しており、NHKも2023年5月に加入した。

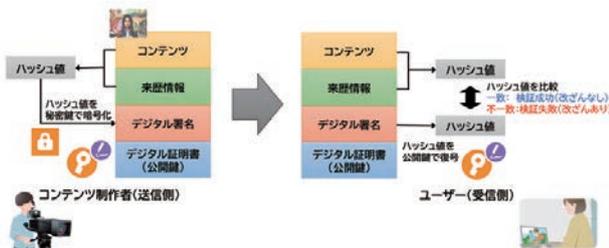
C2PAの技術仕様では、来歴情報の付与方法や検証・提示方法などが定められている。コンテンツ制作における「撮影」、「編集」、「配信」など、コンテンツに対するアクションを誰がいつ行ったかを、デジタル署名を用いて改ざん検知可能な形式でコンテンツのメタデータとして埋め込む。来歴情報が付与されたコンテンツに新たなアクションが加えられる際は、新しく生成された来歴情報を追記することで、ユーザーにコンテンツが届くまでのすべての過程を示す来歴情報を届けることができる。デジタル署名を用いた来歴情報付与・検証の仕組みを図2に示す。デジタル署名はコンテンツと来歴情報から計算されたハッシュ値をコンテンツ制作者の秘密鍵を用いて暗号化したものである。来歴情報にはデジタル署名のほか、秘密鍵に対応する公開鍵(デジタル



■ 図1. C2PAにおける来歴情報付与・提示の流れ



証明書)が含まれており、ユーザーは受信したコンテンツに含まれる公開鍵を用いてデジタル署名を復号して得られたハッシュ値と、コンテンツと来歴情報から計算したハッシュ値とを比較することで、コンテンツ及び来歴情報の改ざんを検知できる。なお、C2PA仕様は、デジタル署名を検証することにより、コンテンツや来歴情報が改ざんされていないこと及び来歴を付与した人・組織・撮影機材・編集ソフトなどのエンティティが本物である(なりすましなどの詐称がされていない)ことを保証する仕組みであり、来歴情報の有無がコンテンツの真偽そのものや、来歴を付与したエンティティが信頼できるかどうかを判断するものではないため、ユーザー自身による判断が必要である。

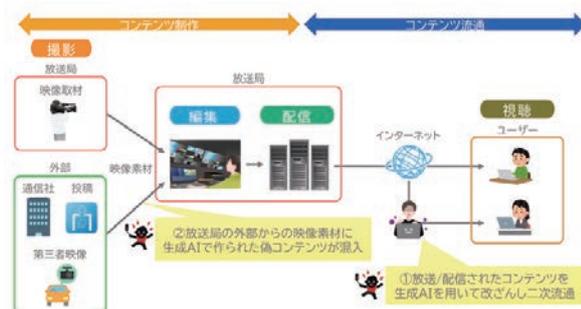


■図2. 来歴情報付与・検証の仕組み

3. 生成AIが放送業界にもたらす脅威

近年の生成AI技術の進展により、高品質な偽動画を誰でも簡単に作成することが可能となった。これにより、コンテンツ制作と流通それぞれのワークフローにおいて脅威をもたらすと考えられる(図3)。

コンテンツ流通のワークフローでは、放送局が放送・配信したコンテンツを第三者が生成AIを用いて改ざんし、放送局になりすましたコンテンツがSNS上に二次流通するなどのリスクがある。一方、コンテンツ制作のワークフローにおいて、放送局が提供するニュース番組は、自身が取材した映像だけでなく、視聴者による投稿映像や通信社から伝送される映像など、外部からの映像を用いて制作される。



■図3. 制作・流通のワークフローにおける脅威

これらの映像素材に生成AIで作られた偽コンテンツが混入し、誤った情報を放送・配信してしまうといったリスクがある。

4. 来歴情報提示技術の研究開発事例

先に述べたコンテンツ制作・流通のワークフローにおける脅威への対策として、NHKにおける来歴情報提示技術の研究開発事例を2つ紹介する。これらは2025年5月29日~6月1日に開催されたNHK放送技術研究所の一般公開イベント「技研公開2025」で展示された。

◆C2PA対応動画視聴プレーヤー

C2PAを前提とするエコシステム全体や要素ごとの課題抽出、ユーザーが実際にコンテンツの信頼性を判断するために重要な要素の調査・検証を目的に、CAI(Content Authenticity Initiative)が提供するC2PA準拠のライブラリーを用いて、来歴情報をMPEG-DASH形式のストリーミング用動画コンテンツに付与するシステム及びコンテンツを再生しながら来歴情報をリアルタイムで検証・提示する動画視聴プレーヤーを試作した。本プレーヤーは、NHKで開発したMPEG-DASH動画視聴プレーヤー basjoo.js に対し、来歴情報の検証・提示処理を追加したものであり、コンテンツの信頼性を判断する要素を分かりやすく提供することを目的に、以下に示す4つの機能を組み込んだ。なお、本プレーヤーはPCだけでなくスマートTV上でも動作可能である。

・来歴情報の有無の視覚的提示

来歴情報が付与されたコンテンツを再生する場合、C2PAの公式アイコンであるcrアイコンを画面右上に提示することで、来歴情報があることをユーザーに示す(図4(a))。来歴情報が付与されていないコンテンツを再生する場合、crアイコンを提示しないことで、来歴情報がないコンテンツであることをユーザーが一目で分かるようにした。

・シークバーによる検証結果の提示

動画コンテンツが改ざんされていないかどうかの検証結果はシークバーで確認できる(図4(b))。試作したプレーヤーはMPEG-DASHのセグメントごとに改ざん検知が可能であり、検証に成功したセグメントは青、改ざんが検知されたセグメントは赤、未再生のセグメントは灰色で表示することで、検証結果をユーザーに分かりやすく提示する。また、再生前であっても、改ざんされたセグメントをプレーヤーが読み込んだ際は、事前に改ざんの通知を行う機能も備えており、ユーザーは落ち着いて適切な対処ができる。

・来歴情報の提示

ユーザーの要求（crアイコンのクリックやリモコンのボタン押下）により、現在再生しているコンテンツの来歴情報を提示することができる（図4 (a)）。検証に成功した来歴情報を実際に確認することで、コンテンツの出どころと制作過程を把握でき、コンテンツの信頼のための判断材料として活用できる。

なお、本プレーヤーでは、再生中のコンテンツに含まれるすべての来歴情報のうち、アクティブマニフェストと呼ばれる最新の来歴情報のみを提示することとした。例えば、「撮影」・「編集」・「配信」の3つの来歴情報が含まれるコンテンツを再生した場合、「配信」の来歴情報のみを提示する。これは、ユーザー評価実験を行った結果、制作過程の詳細よりも、どの事業者が責任をもって配信したコンテンツか分かることが重要である、という意見が多かったためである。

・改ざんの検知と注意喚起

セグメントの改ざんを検知した場合、文字とcrアイコンを用いてユーザーに通知する（図4 (b)）。再生中のコンテンツ

が改ざんされている場合、偽・誤情報の可能性をユーザーに提示することで、コンテンツを注意深く視聴してもらう、または、視聴をやめてもらうなどの行動を喚起することができる。

◆来歴情報を検証・提示するWebブラウザ

C2PA対応動画視聴プレーヤーをWebアプリとして実装する場合、偽の来歴情報や検証結果を提示するアプリが流通する可能性がある。そこで、オープンソースのWebブラウザであるChromiumを改修し、来歴情報の検証・提示が可能なWebブラウザを試作した。第三者がWebブラウザに介入することは困難なため、提示される来歴情報と検証結果の信頼性をより高めることができる。本ブラウザでは、Webサイトに来歴情報が付与されたコンテンツが含まれている場合、アドレスバーの右側にcrアイコンを表示する（図5 (a)）とともに、crアイコンをクリックすると来歴情報が表示される（図5 (b)）。コンテンツの改ざんを検知した場合、文字とcrアイコンを用いてユーザーに通知する（図5 (c)）。



■図4. C2PA対応動画視聴プレーヤーの画面例

◆映像管理システム

ニュース制作の現場では、視聴者による投稿映像や通信社から伝送される映像など、外部からの膨大な数の映像素材に対し、放送・配信に利用できるかどうかの判断をすべて人手でチェックしなければならない。そこで、人手でチェックする前に、C2PAの来歴情報を用いて映像素材を機械的に分類することで、制作現場の負担を軽減する映像管理システムをNHKメディアイノベーションセンター、報道局と連携して試作した。図6に映像素材の分類ワークフロー、図7に映像管理システムの画面例を示す。外部からの映像素材が本システムに入力されると、C2PAの来歴情報が付与さ



■図5. 来歴情報を検証・提示するWebブラウザの画面例



れているかどうかの判別を行う。来歴情報が付与されていない場合は「来歴情報なし」に分類される。一方、映像素材に来歴情報が付与されている場合はその検証を行い、検証に成功した場合は以下の分類を行う。すなわち、カメラで撮影したという来歴のみがある場合は「来歴情報あり（カメラ撮影）」、生成AIで作成／加工したという来歴がある場合は「来歴情報あり（生成AI使用）」、上記以外の場合は「来歴情報あり（その他）」に分類する。そして、来歴情報の検証に失敗した場合は「来歴情報あり（改ざん検出）」に分類する。本システムの管理画面では、映像素材に含まれる来歴情報を元にすべての映像素材が上記5つのいずれかに分類された状態で表示される。ニュースの映像として利用できるか否かの判断は従来どおり人手で行う必要があるが、例えば、「来歴情報あり（カメラ撮影）」に分類された映像を優先的にチェックするなどの方法により、制作現場の負担を軽減することができると考えられる。

なお、図6のワークフローは外部からの映像素材がC2PAに準拠したカメラや生成AIソフトウェアを用いて制作されることを前提として設計されている。例えば、ソニーはC2PA



■図6. 映像素材の分類ワークフロー

の技術規格に対応したミラーレス一眼カメラを開発しており、撮影時にリアルタイムで静止画に署名付き来歴情報を付与することが可能である。現在は静止画のみ対応しているが、今後の動画への対応が期待される。Adobeは、C2PAに対応した動画編集ソフトPremiere Proを開発しており、動画コンテンツの編集過程で生成AIを活用した場合、生成AIを使用したことを表す来歴情報がコンテンツに付与される。また、生成AIツール「Adobe Firefly」を用いて生成されたコンテンツには、生成AIを使用したことを表す来歴情報が付与される。このように、撮影・編集などコンテンツ制作のワークフローごとにC2PAの来歴情報をコンテンツに付与できる環境が整いつつある。

5. まとめ

フェイクニュースをはじめとする偽情報・誤情報への対策として、コンテンツの信頼性を確保するための来歴情報提示技術とNHKにおける研究開発事例について紹介した。YouTubeでは、アップロードされた動画がカメラで撮影され、生成AIなどによる改変がされていないことをユーザーに示す機能を提供しており、また、FacebookやInstagramでは、AIで作成されたコンテンツを表示する際に「Made with AI」のラベル付けを行うなど、プラットフォーム側もC2PAの来歴情報を用いた偽情報・誤情報対策を進めている。今後は、国内外の放送局と連携しながら、来歴情報提示技術の実用化に向けて取り組んでいくとともに、来歴情報がコンテンツに付与されることが当たり前となり、コンテンツを安心して視聴できる情報空間の実現に向け、啓発普及活動にも取り組んでいく。



■図7. 映像管理システムの画面例