

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 成果報告書

2025年3月12日

株式会社データグリッド

# 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	..... 32
2-3. 本実証後の展望	..... 37

# 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

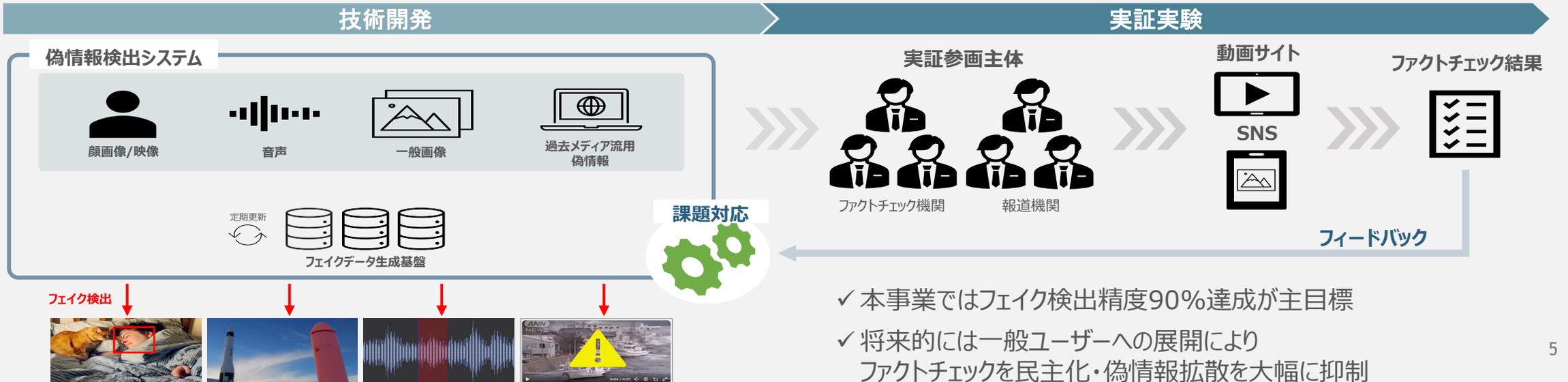
- 2-1. 開発した技術・ツールの詳細 ..... 18
- 2-2. 社会実装のための実証の結果 ..... 32
- 2-3. 本実証後の展望 ..... 37

# 1. 実証事業の概要

## 1-1 実証概要のサマリ

### 実証件名: 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

<b>実施体制</b> <small>(下線: 代表機関)</small>	株式会社データグリッド、大学共同利用機関法人 情報・システム研究機構 国立情報学研究所、国立大学法人名古屋大学、一般社団法人セーファーインターネット協会、特定非営利活動法人ファクトチェック・イニシアティブ、株式会社日本経済新聞社	<b>対象とする偽・誤情報</b>	<ul style="list-style-type: none"> <li>➢ 情報の対象: 人物、物体、風景、過去の情報を流用した偽情報</li> <li>➢ 情報の形態: 画像、映像、音声、画像とテキストの複合コンテンツ</li> </ul>
<b>対策技術</b>	<ul style="list-style-type: none"> <li>➢ 技術1: 顔画像・映像メディアを対象としたフェイクデータ生成基盤及びディープフェイク検出技術</li> <li>➢ 技術2: 音声メディアを対象としたフェイクデータ生成基盤及びディープフェイク検出技術</li> <li>➢ 技術3: 一般画像メディアを対象としたフェイクデータ生成基盤及びディープフェイク検出技術</li> <li>➢ 技術4: 過去メディアを流用した偽情報に対する真偽判別支援技術</li> </ul>	<b>目標</b>	<ul style="list-style-type: none"> <li>➢ 技術1: フェイク顔映像・画像で90%以上の検出精度を達成</li> <li>➢ 技術2: フェイク日本人音声で90%以上の検出精度を達成</li> <li>➢ 技術3: フェイク一般画像で90%以上の検出精度を達成</li> <li>➢ 技術4: 検索エンジンと連携した時間や場所などの関連度から偽情報か否かを判別できる機能を構築</li> </ul>
<b>実証概要</b>	<ul style="list-style-type: none"> <li>➢ 【課題】顔画像・映像、音声、一般画像を対象としたディープフェイクや過去メディアを流用した偽情報に対する技術的対策では、最新の生成AI技術を含めた国内外のディープフェイクの脅威に対して継続的に対応することが困難であるという課題が存在している。</li> <li>➢ 【提案する取組】最新の生成AI技術によるフェイクデータ生成基盤及びディープフェイク検出技術に加え、検索エンジンと連携した過去メディアの検証技術を活用したシステムを開発し、主要なファクトチェック機関や報道機関との実証実験を実施する。</li> <li>➢ 【目標】中長期的にはファクトチェック機関だけでなく、広く一般ユーザーにもサービスを開放することで、誰もが簡単にSNS等の情報に対してファクトチェックを実施するとともに、一般ユーザーから提供された偽情報と疑われる情報を収集し、偽情報の発生状況をタイムリーに把握可能にする。また、提供された偽情報から半自動的に拡張されるフェイクデータ生成基盤によりモデルを強化し、新たなディープフェイクに対して継続的に対応することで、偽情報の氾濫を極限まで抑制された社会の実現に貢献する。</li> </ul>		



- ✓ 本事業ではフェイク検出精度90%達成が主目標
- ✓ 将来的には一般ユーザーへの展開によりファクトチェックを民主化・偽情報拡散を大幅に抑制

## 1. 実証事業の概要

### 1-2 事業の目的

#### 事業の目的

本事業では、日々発展し続ける多様な生成AI技術や国内のディープフェイク対策で重要となる日本人・日本語にも対応したファクトチェックサービスの開発・実証を行うことで、ファクトチェック機関や報道機関におけるディープフェイクの検証作業の高度化・効率化を実現し、偽・誤情報対策技術の社会実装を推進する。

ディープフェイク対策技術の確立には、ディープフェイクの「生成」と「検知」の両面から取り組むことが必要不可欠である。生成技術に関しては、本提案の代表機関である株式会社データグリッド（以下、「データグリッド」）が中心に取り組む。データグリッドは、2017年の創業以来、生成AIに特化して研究開発に取り組んでいる企業であり、独自の大規模データセットの整備とアルゴリズムの改良により、2018年には高品質な日本人顔画像の生成、2019年にはデジタルヒューマンの全身画像生成に世界でいち早く成功している。

こうした生成AIの社会実装に取り組んできた中で、データグリッドはディープフェイクの台頭に危機感を抱いており、この問題に対処すべく、本事業では、ディープフェイク検出の分野で最も重要なフェイク顔画像の検出を世界で初めて提案し、Deepfake detectionと呼ばれる新たな研究分野を創生するとともに国内初の実用化に成功した国立情報学研究所の越前・山岸及び名古屋大学の安田と連携することで、生成と検知の両面から最新のディープフェイクに対抗する技術開発に取り組む。さらに複数のファクトチェック機関や報道機関との実証実験を通して、実行主体が実際のファクトチェック業務で効果的に利用できるようなサービスを開発する。

#### 偽・誤情報への技術的対策の課題の状況、今後目指す姿

##### <課題の状況>

次々と開発される最新の生成AI技術による多様なディープフェイクに対して、既存のディープフェイク検知ツールは対応できておらず、さらに一般画像への対応も限定的である。また、既存のディープフェイク検知ツールのほとんどは海外製であり、それらに搭載されるディープフェイク音声検出モデルは英語音声もしくは中国語音声で学習と評価がなされており、日本語音声のディープフェイク検出に適していない。

このような状況を打開するために、最新の生成AI技術を含めた国内外のディープフェイクの脅威に対して、継続的に対抗できる技術的対策の確立が求められている。

##### <今後目指す姿>

中長期的には偽・誤情報は更に多く発生すると想定されるため、ファクトチェック機関や報道機関によるファクトチェックだけでは対処が追いつかない可能性が高い。そこで、偽情報検出サービスを広く一般に開放し、誰もが簡単にSNS等の偽情報と疑われる情報に対するファクトチェックを実施できるようにする。そして、一般ユーザーから提供された偽情報と疑われる情報を収集し、偽情報の発生状況をタイムリーに把握し、偽情報の拡散を抑制することに貢献する。また、提供された偽情報から半自動的に拡張されるフェイクデータ生成基盤によりディープフェイク検知モデルを強化し、新たなディープフェイクに対して継続的に対応することで、偽情報の氾濫を極限まで抑制された社会の実現に貢献する。

## 1. 実証事業の概要

### 1-3 開発する技術の概要 (1/4)

#### 開発技術

#### 技術1: 顔画像・映像メディアを対象としたデータ生成基盤及びディープフェイク検出技術

Stable Diffusionを始めとした最新の拡散モデルの多様化・進化によりフェイク顔映像・画像による脅威が深刻化している。NIIが開発した「SYNTHETIQ VISION」はすでに最新の動画生成AIであるOpenAI社のSoraによる生成動画に対するディープフェイク検出に成功しているが、多種多様なフェイク顔画像・映像メディアに対して検出技術を高度化するために、データグリッドのデジタルヒューマン生成技術と連携したフェイクデータ生成基盤を活用し、多様な利用シナリオにおける個々の課題を克服する汎用的な技術を確立する。さらに、検知AIの真偽判定の際に、その判断根拠を単にヒートマップを出力するだけでなく、自然言語で出力する方法を開発する。

#### 本事業全体における当該技術の役割

ディープフェイクの中でも顔画像・映像は形態が多様かつ被害が深刻であるため、当該技術である顔画像・映像メディアのデータ生成基盤及びディープフェイク検出技術の構築は、本事業全体の中核をなす。

## 1. 実証事業の概要

### 1-3 開発する技術の概要 (2/4)

#### 開発技術

#### 技術2: 音声メディアを対象としたデータ生成基盤及びディープフェイク検出技術

英語圏および中国語圏を中心に、人の音声に極めて近いディープフェイク音声の悪用が広がっているが、英語圏でボイスクローンやディープフェイク作成の際に頻繁に利用されるElevenLab等は既に日本語にも対応しており、日本語のディープフェイクが今後悪用される可能性は極めて高い。しかし、日本語音声に対応したディープフェイク検出ツールは存在しておらず、この技術の確立が喫緊の課題である。本提案では、これまでの知見を最大限に活かし、英語・中国語のフェイクデータ生成基盤に加えて、日本語のディープフェイクデータ生成基盤も構築し、日本語にも対応したディープフェイク音声検出モデルも開発する。

#### 本事業全体における当該技術の役割

ディープフェイクの中でも音声メディアは人間の認識に大きな影響を与えるため、当該技術である音声メディアのデータ生成基盤及びディープフェイク検出技術の構築は、本事業全体の重要な要素となる。特に日本語に対応したディープフェイク音声検知ツールの確立は、日本国内でのディープフェイク被害の防止に直結し、社会的な信頼性を向上させる役割を果たす。

## 1. 実証事業の概要

### 1-3 開発する技術の概要 (3/4)

#### 開発技術

#### 技術3: 一般画像メディアを対象としたデータ生成基盤及びディープフェイク検出技術

フェイク顔映像を対象としたSYNTHETIQ VISIONの枠組みをさらに発展させ、物体や風景等の一般画像に対する検出技術を開発する。現在、プロンプト（文章）を入力することで画像を生成するツールも数種類存在し、生成された画像のリアルさも高まっていることから、生成された一般画像が悪用され始め、真偽判定のニーズも非常に高まっている。

「SYNTHETIQ VISION」は顔特有の特徴を判定に利用していないため、検知モデルを学習するデータベースを用意し、偽顔検知モデルと同様の最適化アルゴリズムを実行することにより、一般画像に対する真偽判定も可能になる。検知モデルの学習は、“GenImage”データセット及びフェイクデータ生成基盤を活用する。モデル学習完了後は、自然災害に便乗した偽画像・動画の検出等について評価を行う。

#### 本事業全体における当該技術の役割

一般画像メディアにおけるディープフェイク検出技術の開発は、顔画像・映像に留まらず広範なメディアに対応するための重要なステップである。これにより、様々なシナリオにおけるフェイク画像の検出が可能となり、自然災害などの非常時における偽情報の拡散を防止し、社会の安全と情報の信頼性を確保するための基盤となる。

## 1. 実証事業の概要

### 1-3 開発する技術の概要 (4/4)

#### 開発技術

#### 技術4: 過去メディアを流用した偽情報に対する真偽判別支援技術

本年1月の能登半島地震において、東日本大震災の映像等をSNSや動画サイトから取得し、能登半島地震の映像のように見せかける偽情報が拡散されるなど、過去メディアを流用した偽情報はその手軽さからもこれまでに散見されている。

このような偽情報に対するファクトチェック作業では、現状検索エンジンを用いて過去のソース記事・動画を調査し、人手で照合する必要があるため、ファクトチェック作業の負担となっている。そこで、偽情報と疑われるコンテンツに対して検索エンジン（Google画像検索等）と連携して、検索結果の画像や記事の日付・内容を分析・スコアリングすることで、ファクトチェッカーが確認すべきソースを自動的に絞り込み、ファクトチェック作業を大幅に効率化する真偽判別支援技術の開発を行う。

#### 本事業全体における当該技術の役割

過去メディアを流用した偽情報はその拡散速度と範囲が広いいため、真偽判別支援技術の開発は本事業において重要な役割を果たす。この技術により、ファクトチェック作業が効率化され、迅速かつ正確な情報の提供が可能となり、社会全体の情報リテラシーの向上と偽情報の拡散防止に寄与する。

## 1. 実証事業の概要

### 1-4 社会実装のための実証の詳細 (1/2)

#### 社会実装のゴール

中長期的にはファクトチェック機関だけでなく、広く一般ユーザーにもサービスを開放することで、誰もが簡単にSNS等の情報に対してファクトチェックを実施するとともに、一般ユーザーから提供された偽情報と疑われる情報を収集し、偽情報の発生状況をタイムリーに把握可能にする。また、提供された偽情報から半自動的に拡張されるフェイクデータ生成基盤によりモデルを強化し、新たなディープフェイクに対して継続的に対応することで、偽情報の氾濫を極限まで抑制された社会の実現に貢献する。

#### 社会実装に向け、必要な検討事項

**【提案する取組】** 最新の生成AI技術によるフェイクデータ生成基盤及びディープフェイク検出技術に加え、検索エンジンと連携した過去メディアの検証技術を活用したシステムを開発し、主要なファクトチェック機関や報道機関との実証実験を実施する。

#### 【対策技術と目標】

- **技術1:** 顔画像・映像メディアを対象としたフェイクデータ生成基盤及びディープフェイク検出技術（本事業で実証予定）  
目標：フェイク顔映像・画像で90%以上の検出精度を達成
- **技術2:** 音声メディアを対象としたフェイクデータ生成基盤及びディープフェイク検出技術（本事業で実証予定）  
目標：フェイク日本人音声で90%以上の検出精度を達成
- **技術3:** 一般画像メディアを対象としたフェイクデータ生成基盤及びディープフェイク検出技術（本事業で実証予定）  
目標：フェイク一般画像で90%以上の検出精度を達成
- **技術4:** 過去メディアを流用した偽情報に対する真偽判別支援技術（本事業で実証予定）
  - 目標：検索エンジンと連携した時間や場所などの関連度から偽情報か否かを判別できる機能を構築

## 1. 実証事業の概要

### 1-4 社会実装のための実証の詳細 (2/2)

#### 実証内容

##### 1. 意見交換会の実施

2～3回程度の打ち合わせを通じて主に右記想定ゴールの①、②、③について協議し、テスト利用時に提供するシステム仕様を整理する。

##### 2. テスト利用の実施

上記の意見交換会での内容を踏まえて開発したシステムを実証参画主体に提供し、実際のファクトチェック業務を想定してテスト利用を実施する。検証対象とするディープフェイクコンテンツは世の中に出回るディープフェイクや公開ツールなどで作成したディープフェイクを対象とする。

※ テスト用のシステムでは主に顔画像・映像、一般画像のディープフェイクに対応

(音声のディープフェイク判定についてはスケジュールを前倒して開発できた場合に搭載)

##### 3. フィードバック会の実施

1～2回程度の打ち合わせを通じて、テスト利用のフィードバックを含めて主に②、③、④について協議し、次年度以降に本格的なシステム提供及び業務利用に資するシステム要件を整理する。

#### 実証パートナー

- 一般社団法人セーフインターネット協会 日本ファクトチェックセンター
- 特定非営利活動法人ファクトチェック・イニシアティブ
- 株式会社日本経済新聞社

#### 想定ゴール

##### ① 対応すべきディープフェイク種別の優先順位付け

世の中で問題になっているディープフェイク等について議論し、優先的に対応すべきディープフェイクの種類や生成AIツールを整理することで、効率的にディープフェイク検出技術の対応範囲を拡張する。

##### ② ユーザビリティの明確化

直感的に操作しやすいユーザーインターフェース等について、システムの開発前からテスト利用にかけて議論することで、ファクトチェック業務での高度化・効率化に資するシステム構成を明らかにする。

##### ③ 利用プロセスの確立

ファクトチェック業務のオペレーションの中でのディープフェイク検出ツールの位置付けや利用方法を議論することで、実際のファクトチェック業務に如何に役立てていくのか、その適切なプロセスを確立する。

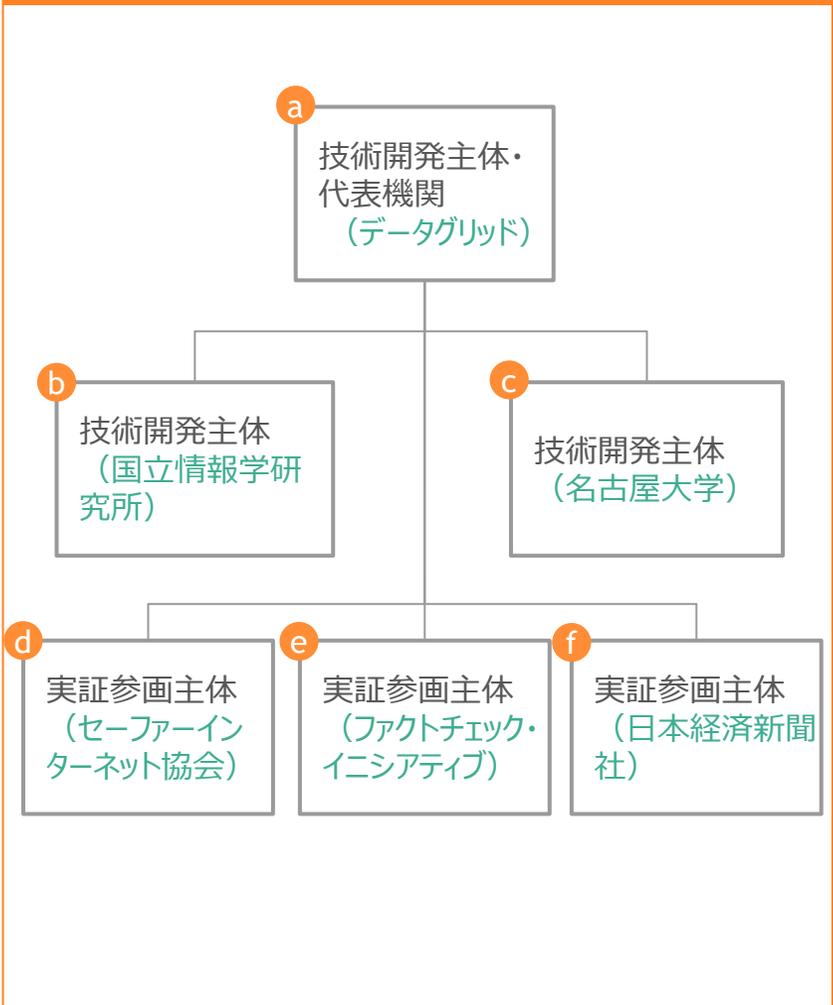
##### ④ 連携プロセスの確立

システムの判定結果に対してユーザーフィードバックの取り込みや、ユーザーから未対応のディープフェイク情報をシステムに取り込むといった、システムの運用プロセスにユーザーが介在することにより、継続的な品質改善を実現する具体的な仕組みを確立する。

# 1. 実証事業の概要

## 1-5 実証の実施体制

実施体制図



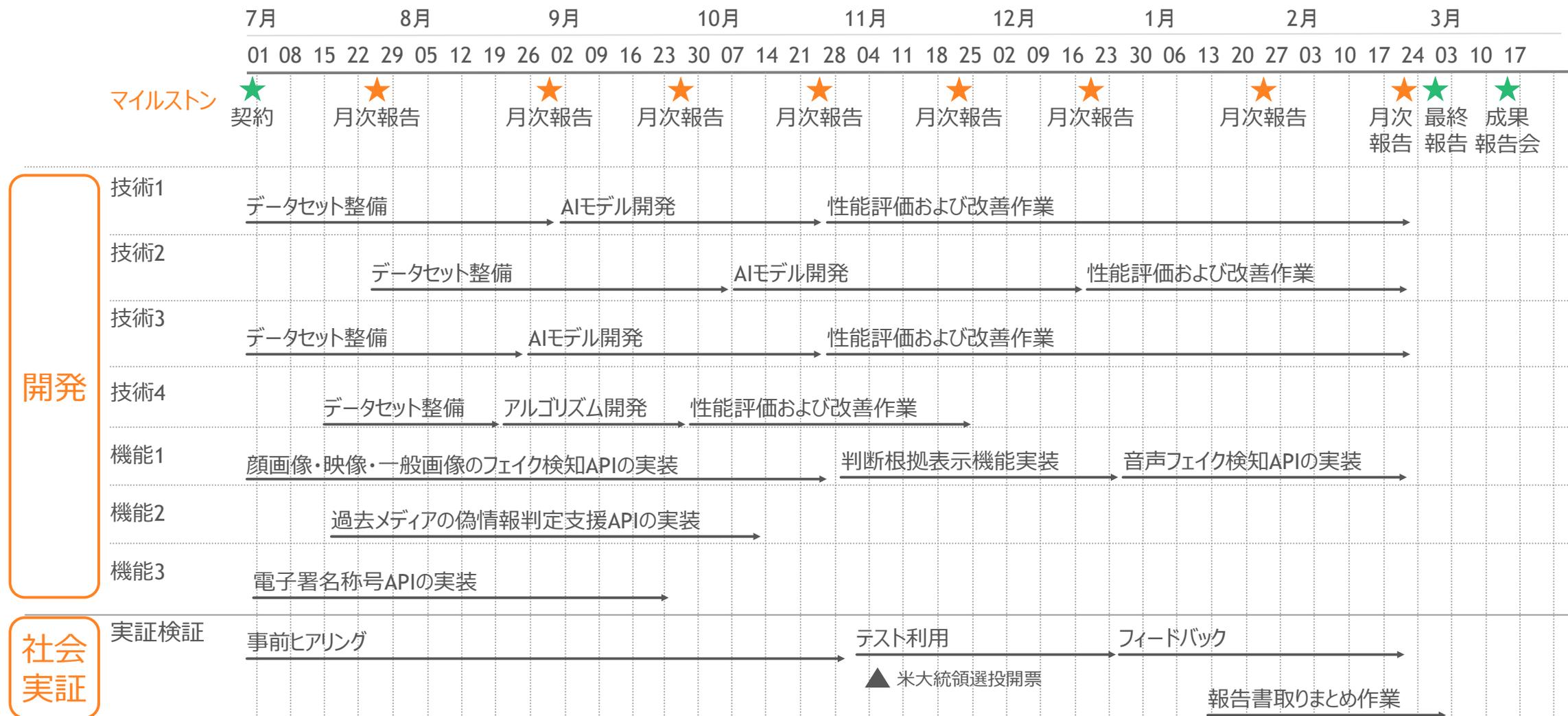
### 団体名

### 役割及び責任

a 株式会社データグリッド (技術開発主体・代表機関)	本事業全般の管理・統括業務、フェイクデータ生成基盤の開発、ディープフェイク検出モデルのチューニング、システム開発
b 大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 (技術開発主体)	SYNTHETIQ VISIONを中心とするディープフェイク検出技術の開発
c 国立大学法人名古屋大学 (技術開発主体)	SYNTHETIQ VISIONを中心とするディープフェイク検出技術の開発
d 一般社団法人セーフアーインターネット協会 (実証参画主体)	ディープフェイク検出サービスのテスト利用及びフィードバックの提供
e 特定非営利活動法人ファクトチェック・イニシアティブ (実証参画主体)	ディープフェイク検出サービスのテスト利用及びフィードバックの提供
f 株式会社日本経済新聞社 (実証参画主体)	ディープフェイク検出サービスのテスト利用及びフィードバックの提供

1. 実証事業の概要

1-5 実施計画 (スケジュール)



## 1. 実証事業の概要

### 1-5 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

- 技術1のデータセット整備：多様な生成手法・個別シナリオによる顔画像・映像メディア学習・評価データセット用のデータ整備を開始
- 技術3のデータセット整備：GenImage・フェイクデータ生成基盤による学習・評価データセット整備を開始
- 技術4のデータセット整備：過去のインターネット上の偽情報データの収集を開始
- 機能1の設計：画像フェイク検知APIの設計を開始
- 機能2の設計：過去メディアの偽情報判定支援APIの設計を開始
- 機能3の設計：電子署名照合APIを設計を開始

～8/26週  
(夏季休暇想定で5w)

- 技術1,3のデータセット整備・AIモデル開発：顔映像と一般画像のデータセット整備を完了し、AIモデル開発を開始
- 技術2のデータセット整備：ElevenLabを含む10種類以上の合成音声の学習・評価データセットの整備を開始
- 技術4のアルゴリズム開発：Google画像検索でデータを収集し、データの関連度等を計算するアルゴリズムの実装を開始
- 機能1の設計・実装：画像フェイク検知APIの設計完了、実装開始
- 機能3の設計・実装：電子署名称号APIの設計を完了し、実装を開始

～9/23週

- 技術4のAIモデル開発：Google画像検索でデータを収集し、データの関連度等を計算するアルゴリズムの実装を完了、性能評価開始
- 機能2の実装：過去メディアの偽情報判定支援APIの設計を完了し、実装開始
- 機能3の実装：電子署名照合APIの実装完了
- 実証検証：現場での課題感等を検証パートナーに事前ヒアリングを実施

～10/21週

- 技術1,3のAIモデル開発：顔映像と一般画像に対する複数の検知アルゴリズムの開発完了
- 技術2のAIモデル開発：日本人音声に対する複数の検知アルゴリズムの開発を開始
- 技術4の性能評価と改善作業：アルゴリズムを改善し、正しく類似データや関連度を表示できるか確認
- 機能1の実装：画像フェイク検知APIの実装完了
- 機能2の実装：過去メディアの偽情報判定支援APIの実装完了

## 1. 実証事業の概要

### 1-5 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

- 技術1,3の性能評価と改善作業：アルゴリズムを改善し、フェイク顔映像・画像、一般画像を対象に精度80%を達成
- 技術4の性能評価と改善作業：アルゴリズムを改善し、正しく類似データを表示できるか確認
- 機能1の設計：判断根拠表示機能の設計を開始
- 実証検証：各機能に関してパートナーとテスト利用を実施

～12/16週

- 技術1,3の性能評価と改善作業：アルゴリズムを改善し、フェイク顔映像・画像、一般画像を対象に精度85%を達成
- 技術2のAIモデル開発：日本人音声に対する複数の検知アルゴリズムの開発を完了
- 機能1の設計・実装：判断根拠表示機能の設計を完了、実装を開始
- 実証検証：各機能に関してパートナーとテスト利用を実施

～1/20週

(年未年始想定で5w)

- 技術1,3の性能評価と改善作業：アルゴリズムを改善し、フェイク顔映像・画像、一般画像を対象に精度90%を達成
- 技術2の性能評価と改善作業：アルゴリズムを改善し、合成音声データセットを対象に精度80%を達成
- 機能1の設計・実装：判断根拠表示機能の実装を完了し、音声フェイク検知APIの実装を開始
- 実証検証：検証パートナーと使用感や性能等に関するフィードバック会の実施

～2/17週

- 技術2の性能評価と改善作業：アルゴリズムを改善し、合成音声データセットを対象に精度90%を達成
- 機能1の実装：音声フェイク検知APIの実装を完了
- 実証検証：検証パートナーと使用感や性能等に関するフィードバック会の実施

最終報告

- 技術1-4の検証結果を報告
- 機能1-3について開発結果を報告
- 実証検証の結果を報告
- 本事業全体の総括を報告

# 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 ..... 32
- 2-3. 本実証後の展望 ..... 37

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (1/14)

#### 開発技術の詳細内容

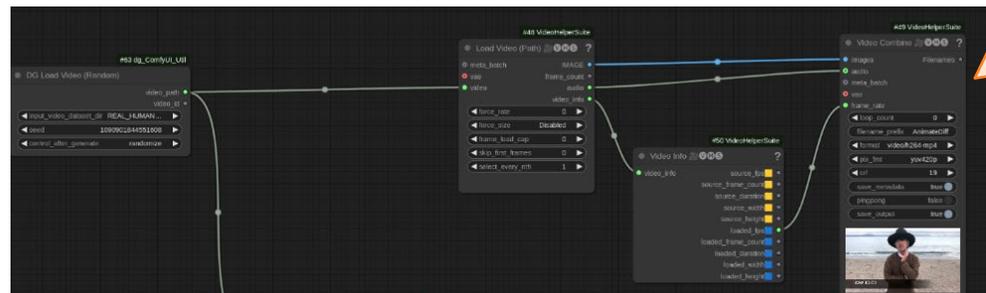
##### 技術1：顔画像・映像メディアを対象としたディープフェイク検出技術

- **目的：顔画像・映像メディアを対象としたディープフェイク検出モデルの高精度化**
  - フェイクメディア検知技術を実社会へ導入する事を前提とした場合、公開後に提案される新しいフェイクメディア生成手法への迅速な追従が必要不可欠であり、これは従来の機械学習の仮定と大きく異なるため、機械学習を用いた検出モデルに関する戦略の高度化が必要
- **技術内容：**
  - **大量リアルデータで学習された基盤モデルの活用**：リアル側のメディアデータを可能な限り大量に学習した基盤モデルを活用することで、フェイク側データの密度分布は不明でも、リアル側メディアの分布をモデリングすることでリアル側の境界を正確に定義する事が可能になる。本事業では、12億枚の画像で事前学習され、メタ社により公開されたDINO(v2)を活用し、多種多様な画像生成・変換アルゴリズムによって合成されたフェイク顔画像を97%以上で検知する技術を構築した
  - **フェイクデータ生成基盤による大規模フェイクデータを学習したモデルを構築**：デジタルヒューマン生成技術と連携したフェイクデータ生成基盤を活用することで、多様な生成手法によるフェイク画像・動画を作成し、学習データに活用することでより頑健なモデルを構築できる。本事業ではFaceID (Stable Diffusion) やghost (GAN) 等を含む9種類の生成手法で生成した顔画像・映像データセットを学習した検出モデルを構築した

作成したフェイク画像データの例



デジタルヒューマン生成技術と連携したフェイクデータ生成基盤



フェイスワップや人物画像生成、リップシンクなど人物関連の多種多様な合成データをフレキシブルに生成可能な生成基盤

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (2/14)

#### R5補正における技術開発の成果

##### 技術1：顔画像・映像メディアを対象としたディープフェイク検出技術

- **何を実現できる技術なのか**
  - 入力画像もしくは映像中の人の顔を分析し、Stable diffusionなどの画像生成AIツールや顔を入れ替えるツールなどで生成もしくは加工された顔画像であるかを映像中のフレームごとに自動分析することができる
  - 分析結果に基づいて、画像もしくは映像中のどの時刻のどの顔が合成・加工された顔であるかを自動判断することができる
- **誰が使用可能な水準なのか**
  - 民間IT企業が自社内サーバに上記ディープフェイク検知プログラムを導入した後、他のプログラムから上記検知機能呼び出すことが可能
  - **OEM製品を作ることに興味がある民間企業がディープフェイク検知機能を持つ製品開発に利用可能**
- **その他特徴・精度など**
  - ディープフェイク検知のベンチマークは様々な生成アルゴリズムにより生成・加工された顔画像を含む独自のテストセット①、②で実施
    - テストセット①：過去から現在に至るまでの多様な生成手法
    - テストセット②：現在流通している主要な生成手法
  - 上記独自テストセット①において生体認証の標準指標にて、**フェイク顔画像の検出精度97%を達成**
    - ここでは検出精度を「100%－等価誤り率(EER)」で定義
  - 主要な商用の生成AIツール（Midjourney, AKOOL, Magic Hourなど）で構築した独自テストセット②で機械学習でよく用いられるACC指標として**98%以上の検出精度を達成**
  - フェイクの判断根拠として、生成手法を自然言語で出力することができる（機能1 参照）

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (3/14)

#### 開発技術の詳細内容

##### 技術2：音声メディアを対象としたディープフェイク検出技術

- **目的：日本語の合成音声を対象としたディープフェイク検出モデルの構築と評価**
  - 画像や映像のディープフェイク検出モデルとは異なり、**合成音声のディープフェイク検出モデルは対象言語の影響を強く受け、言語毎のモデル構築が必要**
  - これまで構築されてきた合成音声のディープフェイク検出モデルは英語・中国語が中心であり、**日本語を適切に扱えるディープフェイク検出モデルは無い**
- **提案内容：日本語の合成音声を対象とし、実環境で利用可能なディープフェイク検出モデルを構築する**
  - 日本語の多種多様な合成音声を生成できる日本語音声フェイクデータ生成基盤を構築し、その基盤を使って日本語音声フェイクデータを自動生成することでディープフェイク検出モデル学習・評価用のデータベースを作成する
  - クリーン環境だけでなく、雑音、圧縮、反響など劣悪環境を想定し検出モデルを構築する
- **研究・実験内容：大量リアルデータにより事前学習された基盤モデルを活用するディープフェイク検出実験を実施**
  - 英語の合成音声の検出タスクにおいて優れた検出性能を実現しているSSL-ASSISTネットワークを採用
  - 画像・映像、および、英語音声におけるディープフェイク検出タスクに倣い、日本語音声においても、学習データに含める生成手法と、評価データに含める生成手法を分け、未知の日本語音声合成モデルを検知することを評価

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (4/14)

#### R5補正における技術開発の成果

##### 技術2：音声メディアを対象としたディープフェイク検出技術

- **何を実現できる技術なのか**
  - 入力日本語音声进行分析し、与えられた日本語音声ファイルが、テキスト音声合成ツールもしくは声質変換ツールなどで生成もしくは加工された合成音声であるか、それとも自然音声なのかを入力単位で分析することができる
- **誰が使用可能な水準なのか**
  - 現時点では主に研究者やAIエンジニアが利用可能な水準
    - 国立情報学研究所が開発しているSYNTHETIQ VISIONシステムにおいて日本語合成音声の検知モデルを利用するためには、上記検知モデルをSYNTHETIQ VISIONシステムにおいて利用するための開発作業が今後必要
    - また入力音声ファイル全体のディープフェイク分析ではなく、ある一定時間間隔毎のディープフェイク分析も行える様に検知モデルのアルゴリズムやネットワーク構造をさらに改良する必要あり
    - これらの開発・改良後に、SYNTHETIQ VISIONシステムを利用する民間IT企業が音声のディープフェイク検知を利用可能となる
- **その他特徴・精度など**
  - 日本語音声のディープフェイク検知ベンチマーク用のテストデータセットを様々な生成アルゴリズムにより合成・変換された日本語音声で構成
    - 上記テストデータセットは雑音、圧縮、反響など劣悪環境をシミュレート
  - 上記独自テストデータセットにおいて生体認証の標準指標で、**フェイク日本語音声の検出精度98%～99.8%を達成**
    - ここでは検出精度を「100%－等価誤り率(EER)」で定義
    - **この数値は劣悪環境において十分な検知性能を持っていることを示しており、顕著な成果と言える**

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (5/14)

#### 開発技術の詳細内容

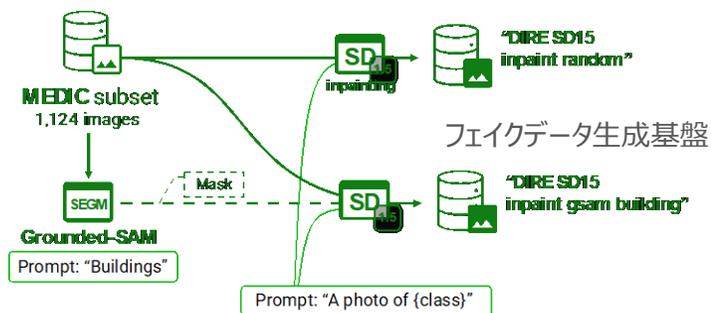
#### 技術3：一般画像メディアを対象としたディープフェイク検出技術

##### 開発した技術の内容

- 近年生成AIにより画像の一部を改変する手法が提案されており、このような状況を鑑み、生成AIにより画像に部分的操作を施した一般画像（災害画像）のディープフェイクデータセットを構築した
- 部分的に改変された一般画像のディープフェイクに対して、効果的に真贋判定を行う高度な検出・ローカライズ手法を開発し、構築したデータセットを用いた評価で、90.4%の検出精度を達成した

##### 一般画像（災害画像）のディープフェイクデータセットの構築

- 災害画像（リアル）のデータセット（MEDIC）の一部を生成AIにより改変するフェイクデータ生成基盤（下図）を構築し、災害状況を過度に演出した災害画像のフェイク画像を生成
- 生成したフェイク画像は視覚的違和感が少なく、見た目ではフェイクと判断するのは困難、従来の二値分類に基づくディープフェイク検出手法による検出も困難



災害状況を過度に演出したフェイク画像の生成

##### データセット情報

種類	画像数
リアル災害画像（MEDICから）	834
<b>フェイク災害画像</b>	<b>834</b>
合計	1,668



データセットのサンプル画像

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (6/14)

#### R5補正における技術開発の成果

#### 技術3：一般画像メディアを対象としたディープフェイク検出技術

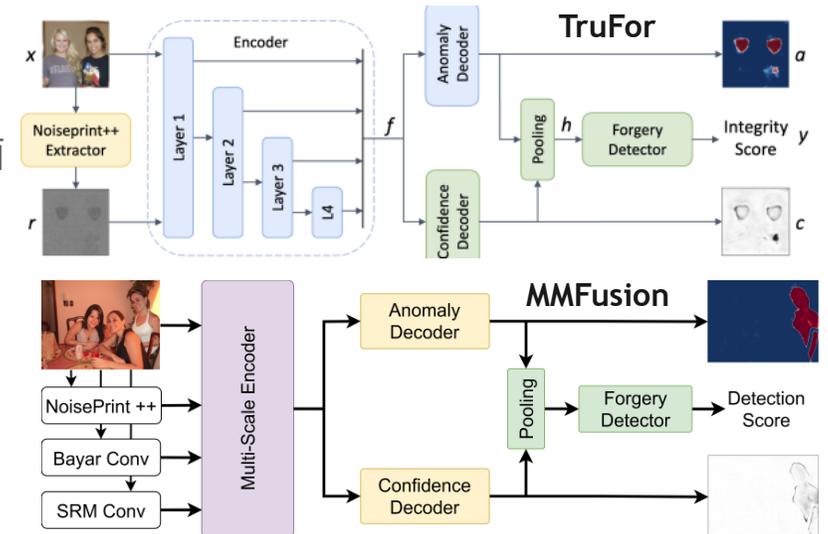
- 何を実現できる技術なのか
  - 部分操作された一般画像（特に本事業では災害画像を対象にした）の検出及び変更箇所の特定をすることができる
- 誰が使用可能な水準なのか
  - 現時点では主に研究者やAIエンジニアが利用可能な水準
- その他特徴・精度など
  - 災害画像のディープフェイクデータセットを用いて、最先端のディープフェイク検出・ローカライゼーションモデルであるTruForとMMFusionをファインチューンすることで、評価指標ACCで**90.4%の検出精度を達成**

#### ■ ローカライゼーション精度(MMFusion)：ファインチューニングによる精度向上

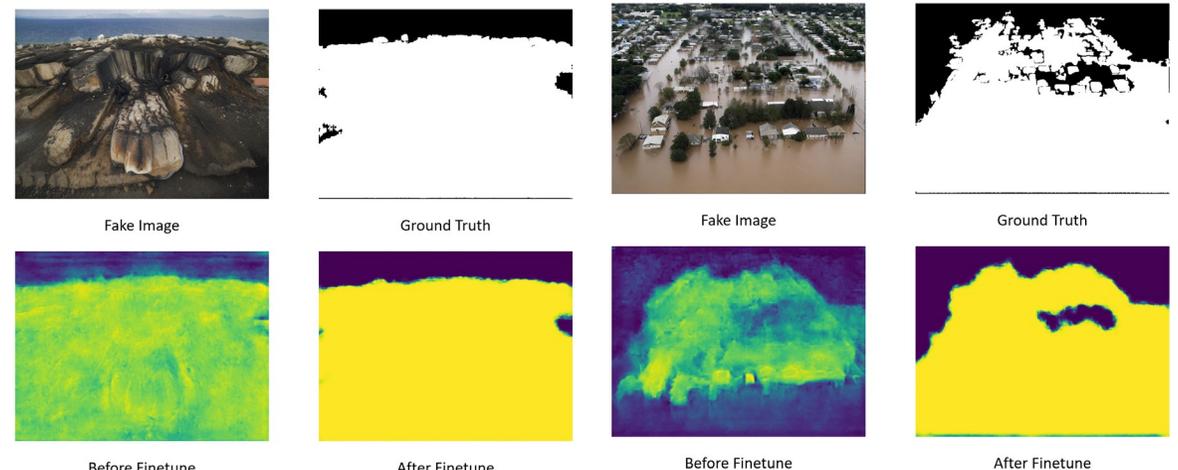
モデル	F1-best (%)	F1-fixed (%)
MMFusion (オリジナル)	72.75	53.18
<b>MMFusion (ファインチューニング)</b>	<b>85.97</b>	<b>76.78</b>

#### ■ 検出 (二値分類) 精度(TruFor, MMFusion)：ファインチューニングによる検出精度の向上

モデル	検出精度 (%)
TruFor (オリジナル)	57.78
MMFusion (オリジナル)	64.40
<b>MMFusion (ファインチューニング)</b>	<b>90.40</b>



TruForとMMFusionのアーキテクチャ



ファインチューニングによるローカライゼーション精度の向上 (MMFusion)

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (7/14)

#### 開発技術の詳細内容

#### 技術4：過去メディアを流用した偽情報に対する真偽判別支援技術

- 技術の背景・目的

- 本年1月の能登半島地震において、東日本大震災の映像等をSNSや動画サイトから取得し、能登半島地震の映像のように見せかける偽情報が拡散されるなど、過去メディアを流用した偽情報はその手軽さからもこれまでに散見されている
- このような偽情報に対するファクトチェック作業では、現状検索エンジンを用いて過去のソース記事・動画を調査し、人手で照合する必要があるため、ファクトチェック作業の負担となっている

- 技術の概要

- 対象データに関してリバースイメージサーチを実施することで過去に同じ画像がないか効率的に照合する技術

- 技術の詳細

- 偽情報と疑われるデータ（画像を想定）に検索エンジン(Google 画像検索等)と連携して、メタデータを取得し、分析することで、ファクトチェッカーが確認すべきソースを自動的に絞り込み、ファクトチェック作業を大幅に効率化する真偽判別支援技術である。具体的にはユーザーがアップロードしたデータに関してリバースイメージサーチの1つであるGoogle 画像検索と連携して、以下の情報を取得することで、ファクトチェック作業を効率化する
  - 画像の関連度
    - 画像の類似性を評価する画像構造の類似度を計算するSSIM (Structural Similarity) と学習済みのニューラルネットワークを用いて画像の類似性を判定するDISTS (Deep Image Structure and Texture Similarity) を組み合わせて計算
  - ページタイトル（記事の概要）

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (8/14)

#### R5補正における技術開発の成果

#### 技術4：過去メディアを流用した偽情報に対する真偽判別支援技術

- **何を実現できる技術なのか**
  - 偽情報と疑われるコンテンツに対してGoogle検索と連携して、検索結果の画像の関連度（画像の類似度）、ページタイトル、記事の更新日の一覧を表示し、ファクトチェッカーが優先的に確認すべき対象を明示することで、ファクトチェック作業を大幅に効率化する真偽判別を支援する技術
- **誰が使用可能な水準なのか**
  - ファクトチェック業務を行う事実確認担当者が使用可能
- **その他特徴・精度など**
  - 対象モーダル：画像を含むweb上の記事
  - 使用技術
    - リバースイメージサーチ
    - 画像の類似度算出
      - 類似度の評価方法に関しては、画像構造の類似度を計算するSSIM（Structural Similarity）と学習済みのニューラルネットワークを用いて画像の類似性を判定するDISTS（Deep Image Structure and Texture Similarity）を用いた

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (9/14)

#### 開発技術の詳細内容

#### 機能1：顔画像・映像、音声、一般画像を対象としたディープフェイクの検出・解析をする機能

#### ディープフェイク判別機能のUI

##### ● 機能の概要

- 対象の画像・映像・音声データを生成AIで生成されたものか判別する機能およびユーザーインターフェース

##### ● 機能の詳細

- ① ユーザーがアップロードした画像・映像・音声データを生成AIで生成されたものか判別
  - 顔画像・映像：画像・映像中の顔領域を検出し、その顔領域が生成AIで作成されたものかどうかを判別する
  - 音声：動画中の音声が生成AIで作成されたものかどうか判別する
  - 一般画像：顔を含まない画像（自然画像、災害画像、動物画像など）が生成AIで作成されたものかどうかを判別する
- ② 生成スコアを表示
  - データが、生成AIで作られたものがある可能性を0~100%で表示
    - 生成スコアが高ければ高いほど、生成AIで作られたデータである可能性が高いことを示す
- ③ 生成スコアの根拠（分析結果）を表示
  - どんな生成手法で生成されたか、どの部分が生成されているかなどを自然言語の形態で表示

b43379da-7b8f-449d-9665-b306ab66dc16.jpeg 容量：244.80KB アップロード日時：2024/11/



判定結果

**フェイク**

生成スコア

**99.99%**

分析結果

Stable Diffusionで生成された可能性が高い

C2PA情報

署名者名：なし

署名時刻：なし

画像検索判定

判定の詳細 →

類似の画像が見つかりませんでした。

Stable Diffusionで生成された人物画像を、生成AIで作成されたものと判定している

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (10/14)

#### R5補正における技術開発の成果

##### 機能1：顔画像・映像、音声、一般画像を対象としたディープフェイクの検出・解析をする機能

商用サービスで生成した顔データに対する検出精度

###### ● 何を実現できる機能なのか

- 顔画像・映像メディア、音声メディア、一般画像メディアを対象としディープフェイク検出することができる機能
- 検出結果に対する判断根拠等を出力する機能

###### ● 誰が使用可能な水準なのか

- ファクトチェック等専門的な業務を行う事実確認担当者が使用可能
- メディア企業など具体的な利用主体が使用可能

###### ● その他特徴・精度など

- **顔画像・映像メディア**：顔フェイクメディアの検出率に関して**精度98%**を確認（技術1を参照）
  - フェイク人物顔画像・生成用途で用いられることが多い商用の画像・動画生成サービスである「AKOOL」、「DeepSwapper」、「Magic Hour」などで実利用想定でのフェイク顔・映像の評価データセットを作成して上記精度を確認した
- **音声メディア**：日本語音声フェイクメディアの検出率に関して**精度90%**を確認（技術2を参照）
- **一般画像メディア**
  - 災害画像のフェイクメディアの検出率に関して**精度90%**を確認（技術3を参照）
  - 有名な画像生成サービスである「Midjourney」や商用のソフトウェアである「Adobe Firefly」、「Canva」などで災害画像以外の実利用想定でのフェイク一般画像の評価データセットでは**精度89%**を確認
  - 本実証のスコープ外ではあるが、最近X（旧twitter）等で特に投稿が多い、OpenAIの「Sora」、Googleの「Veo2」、Runwayの「Gen3」、Lumaの「DreamMachine」などの動画生成AIで作成した動画で評価したところフェイク動画の検出**精度87%**を確認し、フェイク一般画像向けの技術がフェイク一般動画に転用可能であることも確認した

主要な商用生成サービス		検出精度
汎用型	Midjourney	<b>99.1 %</b>
	Adobe Firefly	<b>97.5 %</b>
	Canva	<b>97.6 %</b>
顔特化	AKOOL	<b>98.7 %</b>
	Deep Swapper	<b>97.8 %</b>
	Magic Hour	<b>98.1 %</b>

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (11/14)

#### 開発技術の詳細内容

#### 機能2：過去メディアを流用した偽情報に対する真偽判別支援する機能

##### ● 機能の概要

- 対象データに関して過去に同じデータがないか効率的に検索し、ファクトチェック業務の支援をする機能およびユーザーインターフェース

##### ● 機能の詳細

- ユーザーがアップロードしたデータに関してGoogle 画像検索と連携して、
  - 画像の関連度
    - 画像の類似性を評価するSSIMやDISTSを組み合わせて計算
  - ページタイトル
  - ページURL
  - ページの最終更新日
- を表示することで、類似画像やその記事の最終更新日を、ユーザーは瞬時に確認することができる。（技術4を参照）例えば、過去に別媒体で使用されたデータが使いまわされて、偽誤情報のデータとして使われているか検証することができる

#### 類似画像検索機能のUI

##### 画像検索で判定

##### 確認対象コンテンツ



##### 関連するコンテンツ一覧

画像	関連度	ページタイトル	ページURL	最終更新日
	95.09%	「ドローンで...	<a href="https://ne...">https://ne...</a>	1970/01/01 09:00
	91.32%	あなたが見て...	<a href="https://ne...">https://ne...</a>	2022/10/28 18:30
	79.24%	生成AIで溢れ...	<a href="https://pr...">https://pr...</a>	1970/01/01 09:00
	33.58%	x.com	<a href="https://x...">https://x...</a>	1970/01/01 09:00
	33.58%	x.com	<a href="https://x...">https://x...</a>	1970/01/01 09:00
	30.53%	A I 使い「静...	<a href="https://w...">https://w...</a>	2022/09/27 20:16
	29.61%	静岡災害デマ...	<a href="https://sp...">https://sp...</a>	2022/09/28 00:00



## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (12/14)

#### R5補正における技術開発の成果

##### 機能2：過去メディアを流用した偽情報に対する真偽判別支援する機能

- **何を実現できる機能なのか**
  - 偽情報と疑われるコンテンツに対してGoogle検索と連携して、検索結果の画像の関連度（画像の類似度）、ページタイトル、記事の更新日の一覧を表示し、ファクトチェッカーが優先的に確認すべき対象を明らかにし、ファクトチェック作業を大幅に効率化する真偽判別を支援する機能
- **誰が使用可能な水準なのか**
  - ファクトチェック業務を行う事実確認担当者が使用可能
- **その他特徴・精度など**
  - 対象モーダル：画像を含むSNSの投稿やweb上の記事

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (13/14)

#### 開発技術の詳細内容

#### 機能3 : C2PA の規格に対応した電子署名照合機能

- 機能の概要

- 画像ファイル中のC2PA規格を読み取ることによって、コンテンツが誰によっていつ作成されたのかを識別する機能

- 機能の詳細

- 本機能は、C2PA規格で記録されている署名者名や署名時刻を読み取り、誰がどのタイミングで作ったコンテンツなのかを確認する機能で、これにより情報の出所やその信憑性を確かめることができる

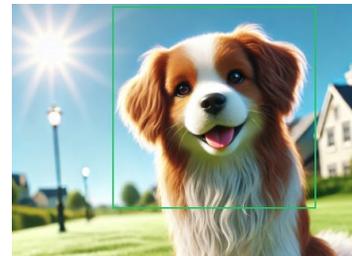
#### C2PAの読み取り例



##### C2PA情報

署名者名 : Adobe Inc.

署名時刻 : 2024/11/20 21:17



##### C2PA情報

署名者名 : OpenAI

署名時刻 : 2024/11/25 09:09

## 2. 実証事業の成果

### 2-1 開発した技術・ツールの詳細 (14/14)

#### R5補正における技術開発の成果

##### 機能3 : C2PA の規格に対応した電子署名照合機能

- **何を実現できる機能なのか**
  - C2PA規格で記録されている署名者名や署名時刻を読み取り、誰がどのタイミングで作ったコンテンツなのかを確認する機能で、これにより情報の出所やその信憑性を確かめることができる。これによりユーザーやファクトチェッカーが信頼性の高い情報かどうかを判断しやすくなり、ファクトチェック機能を効率化することができる
- **誰が使用可能な水準なのか**
  - ファクトチェック業務を行う事実確認担当者が使用可能
- **その他特徴・精度など**
  - 対象モジュール : C2PAを付与可能な画像や動画などのメディアデータ

## 2. 実証事業の成果

### 2-2 社会実装のための実証の結果 (1/5)

#### R5補正における社会実装の成果

##### ①対応すべきディープフェイク種別の整理：

世の中で実際に問題になっているディープフェイクの整理や海外製のディープフェイク検出ツールでは対応が困難なディープフェイクの系統について検証した

ディープフェイク種別の整理

手法	特徴	よく使われるシーン
全体生成	GANや拡散モデルを用い、画像・動画全体をゼロから生成。実在しない人物や風景をリアルに表現可能	偽ニュース映像（政治/災害/安全保障）、架空のSNSプロフィール作成など
インペインティング	映像・画像の一部をAIが自動で補完し、物体除去や追加	偽ニュース映像（政治/災害/安全保障）
リップシンク	音声に合わせて口元や顔の動きを同期させ、発話内容を変更	政治家や有名人の偽の発言動画
Faceswap	特定人物の顔を他人の映像・画像に自然に合成して置き換え	政治家や著名人の偽画像作成、フェイクポルノ
表情変換	映像・画像内の人物の表情を他の表情に変化させる	状況に合わない表情を作ることによる印象操作

海外製のディープフェイクツールでの性能検証（人物対象）

ツール	精度（Accuracy）	再現率（Precision）	適合率（Recall）
A	21%	21%	0%
B	58%	58%	38%
C	75%	83%	78%

海外製のディープフェイクツールでの性能検証（一般画像対象）

ツール	精度（Accuracy）	再現率（Precision）	適合率（Recall）
A	29%	29%	17%
B	44%	44%	33%
C	65%	71%	63%

※ Accuracyはすべての判定のうち、正しく判定できた割合

※ Precisionはフェイクと判定したうち、実際にフェイクだった割合

※ Recallは実際にフェイクだったもののうち、フェイクと判定できた割合

## 2. 実証事業の成果

### 2-2 社会実装のための実証の結果 (2/5)

#### R5補正における社会実装の成果

##### ② ユーザビリティの明確化：

直感的に操作しやすいユーザーインターフェースや根拠の提示等を含む判定結果の表示について、システムの開発前からテスト利用にかけて実証パートナーと議論することで、ファクトチェック業務を効率的に進めるためのユーザーインターフェースを構築した

トランプ前大統領を支持するテイラースウィフトの偽広告.jpeg 容量：73KB アップロード日時：20...

生成スコアと同時に分析結果（フェイクの生成手法等）を表示できる

判定結果  
**フェイク**

生成スコア  
**95.70%**

分析結果  
Stable Diffusionで生成された可能性が高い

PA情報  
作者名：なし  
時刻：なし

画像検索判定  
判定の詳細 →  
類似の画像が1枚見つかりました。

#### 画像検索で判定

確認対象コンテンツ

関連画像のページタイトルとページURLを同時に表示し、ソースの概要と詳細に瞬時に確認できる

関連するコンテンツ一覧

画像 関連度	ページタイトル	ページURL	最終更新日
	Reddit - Dive into anything	<a href="https://www.reddit.com/r/Sw...">https://www.reddit.com/r/Sw...</a>	1970/01/01 09:00
	Ready for Nothing   John Str...	<a href="https://johnpstrohm.substac...">https://johnpstrohm.substac...</a>	1970/01/01 09:00
	Taylor Swift endorsing Kama...	<a href="https://www.thetimes.com/w...">https://www.thetimes.com/w...</a>	2024/09/11 18:40
	Trump in the Debate Sounde...	<a href="https://lawmoralityandethics...">https://lawmoralityandethics...</a>	1970/01/01 09:00

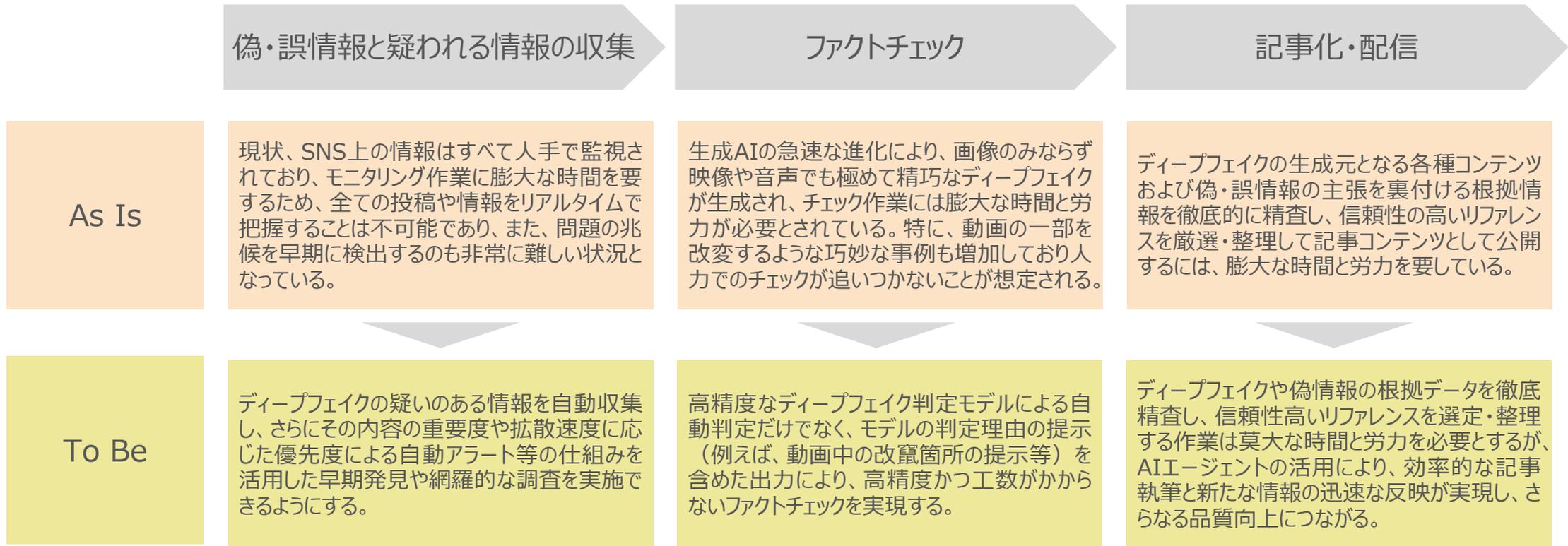
## 2. 実証事業の成果

### 2-2 社会実装のための実証の結果 (3/5)

#### R5補正における社会実装の成果

##### ③ 利用プロセスの確立：

ファクトチェック業務のオペレーションの中でのディープフェイク検出ツールの位置付けや利用方法を議論することで、実際のファクトチェック業務に如何に役立てていくのか、その適切なプロセスを明らかにした



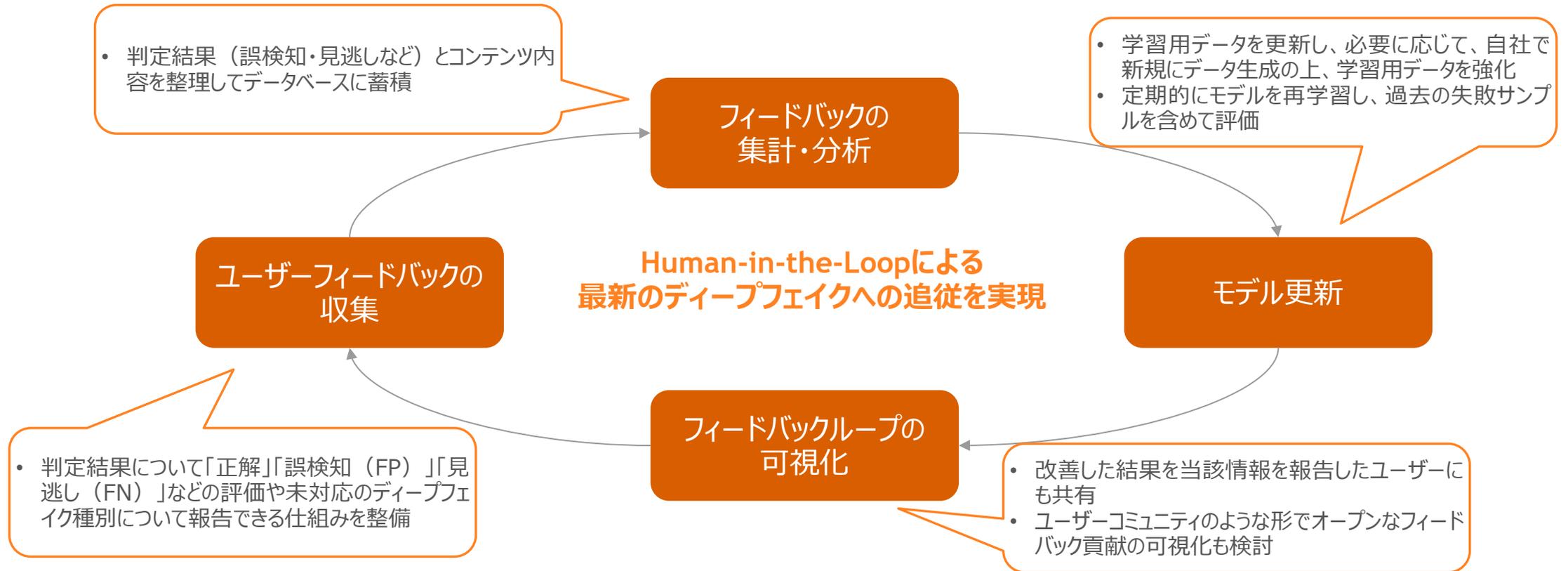
## 2. 実証事業の成果

### 2-2 社会実装のための実証の結果 (4/5)

#### R5補正における社会実装の成果

##### ④ 連携プロセスの確立：

システムの判定結果に対してユーザーフィードバックの取り込みや、ユーザーから未対応のディープフェイク情報をシステムに取り込むといった、システムの運用プロセスにユーザーが介在すること(Human-in-the-Loop)により、継続的な品質改善を実現する具体的な仕組みを明らかにした



## 2. 実証事業の成果

### 2-2 社会実装のための実証の結果 (5/5)

#### 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

## 2. 実証事業の成果

### 2-3 本実証後の展望

2025年度	2026年度	2027年度以降
<b>SNS上の 偽・誤情報対策サービスの開発・実証</b>	<b>サービス普及に向けた本格展開</b>	<b>ビジネスモデルの確立</b>
<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><b>メディア向けツールの開発・実証：</b></p> <ul style="list-style-type: none"><li>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</li></ul> <p><b>SNSユーザー向けツールの開発・実証：</b></p> <ul style="list-style-type: none"><li>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</li></ul> <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><b>システムの改良：</b></p> <ul style="list-style-type: none"><li>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</li><li>新たな偽・誤情報のパターンへの対応策の実装</li></ul> <p><b>偽・誤情報対策エコシステムの拡充：</b></p> <ul style="list-style-type: none"><li>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</li><li>発信者の真正性保証技術等との連携</li></ul> <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><b>収益モデルの確立：</b></p> <ul style="list-style-type: none"><li>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</li><li>海外市場への展開</li></ul> <p><b>継続的技術改善：</b></p> <ul style="list-style-type: none"><li>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</li><li>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</li></ul> <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>