

令和5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

事業実施報告書

2025年3月31日

Agenda

背景・目的、実施内容、スケジュール

(1) 生成AIを始めとする技術の進展に伴うインターネット上の偽・誤情報によるリスクに関する調査・検討

(2) ディープフェイク等のインターネット上の偽・誤情報への対策技術の開発・実証

(3) 報告書作成等

その他 検討事項

Agenda

➤ 背景・目的、実施内容、スケジュール

(1) 生成AIを始めとする技術の進展に伴うインターネット上の偽・誤情報によるリスクに関する調査・検討

(2) ディープフェイク等のインターネット上の偽・誤情報への対策技術の開発・実証

(3) 報告書作成等

その他 検討事項

背景・目的

背景

近年、対話型言語モデル「Chat GPT」や、テキストを入力すると画像を生成する「プロンプト型画像生成AI」、テキストから作曲を行うAI等、**多様な用途での生成AIが登場**。生成AIやディープフェイク技術により、**偽画像・動画を誰でも容易に作成できるようになり**、偽のニュース動画や偽の災害画像など、**人の目では真贋を見分けることが困難な情報**に、国民がSNS等を通じて日常的に触れる機会が増加しており、**トラブルを生じさせている**。

政府内でも、AI戦略チームやAI戦略会議において、我が国における生成AIをめぐる課題について検討しており、2023年5月26日付で公表された「AIに関する暫定的な論点整理」では、偽情報等が社会を不安定化・混乱させるリスクが高まっており、**ディープフェイクを検知する技術やディープフェイクが流通しない仕組みの開発・普及等の奨励が望ましい**と取りまとめられている。

また、令和6年能登半島地震に関して、事実に基づかない不確実な情報や、明らかに偽と思われる情報、真偽を直ちに判別することが困難な情報などがSNS上等で存在するなどの課題が指摘されており、2024年1月25日に公表された被災者の生活と生業支援のためのパッケージにおいて、「切れ目のない被災者支援」の一つとして「被災地におけるインターネット上の偽情報・誤情報対策を行う。」と位置付けられたところである。

目的

生成AIの急進展に伴い顕在化しているリスクへの対策に資するため、以下を通じ、**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の社会実装を推進**する

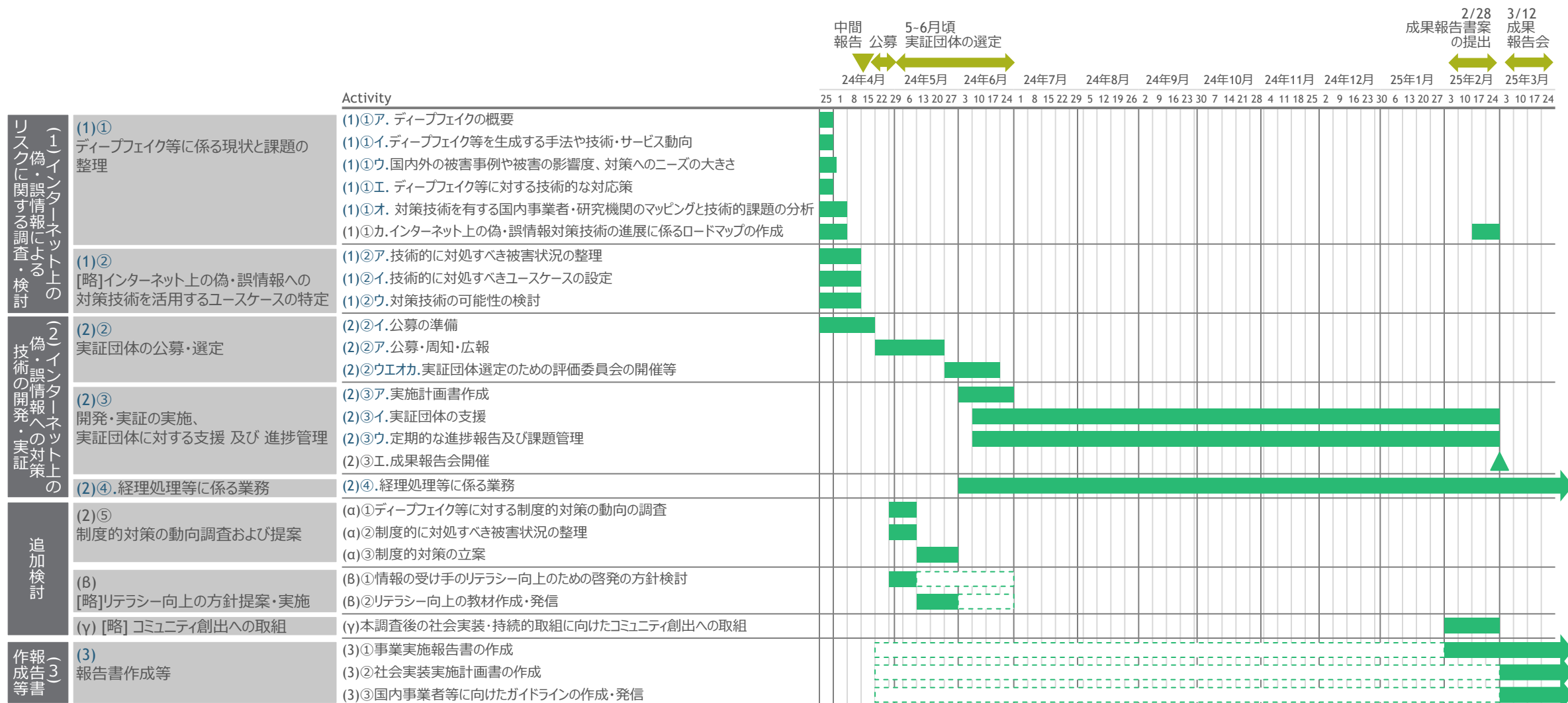
- 調査・検討を通じた以下の整理
 - ディープフェイク等のインターネット上の偽・誤情報の現状と課題
 - 対策すべきユースケース
 - 対策技術
- 対策技術の開発・実証並びに社会実装に向けた実施計画書の策定
 - 情報コンテンツが、生成AIにより生成された可能性が高いかどうか判別する真偽判別支援技術
 - 情報コンテンツや発信者の信頼性等を確保する真正性保証技術・信頼性判断支援技術 等

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の社会実装を推進するために必要な包括的な内容を実施

実施内容



実施スケジュール



Agenda

背景・目的、実施内容、スケジュール



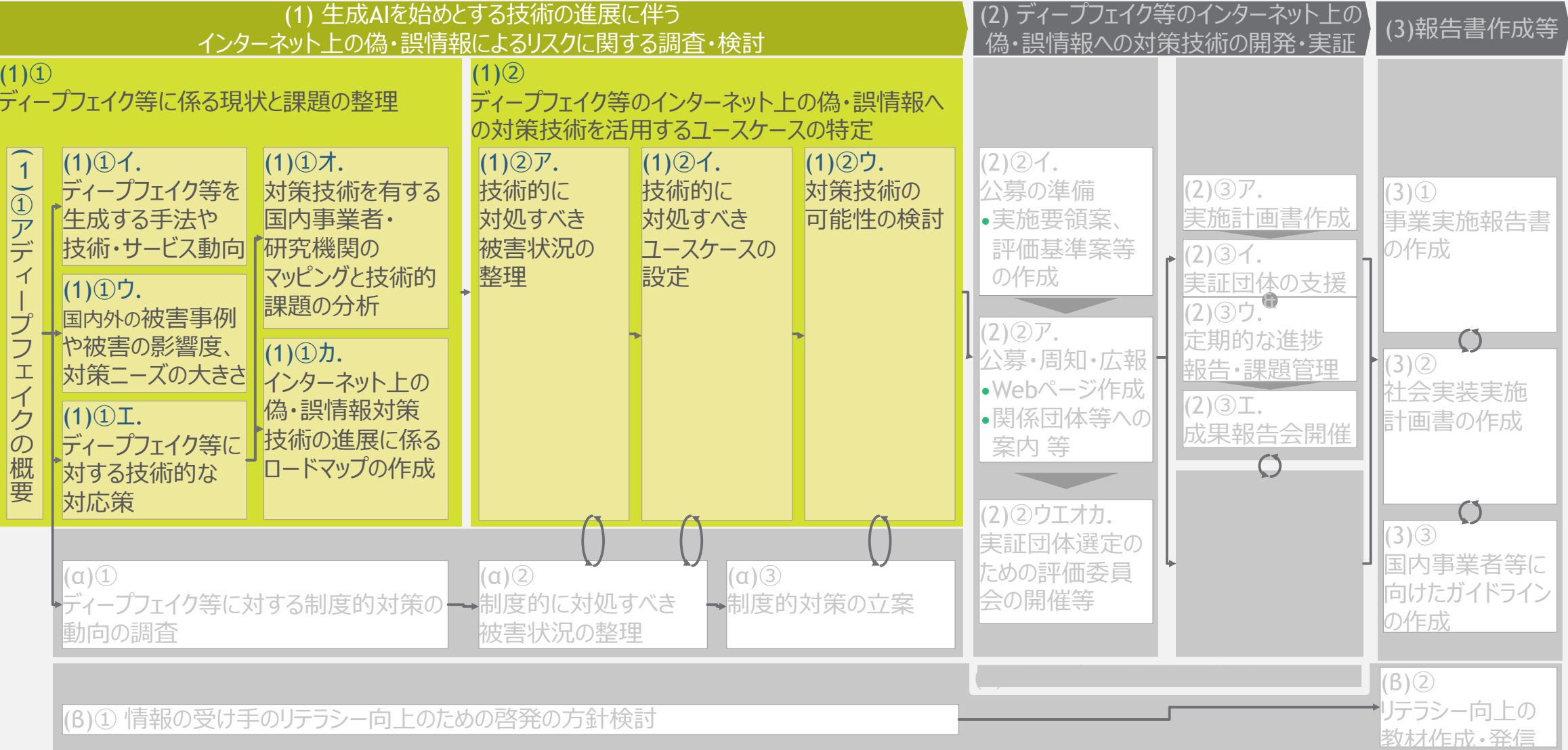
(1) 生成AIを始めとする技術の進展に伴うインターネット上の偽・誤情報によるリスクに関する調査・検討

(2) ディープフェイク等のインターネット上の偽・誤情報への対策技術の開発・実証

(3) 報告書作成等

その他 検討事項

実施内容





ディープフェイクは、AIで生成された、人間の顔/声/行動や風景、物体をリアルに模倣した偽のメディアコンテンツ。特に2022~23年頃から一般ユーザーにも容易に生成可能になった

ディープフェイクの概要[1/2] - 定義と進化の歴史

I. ディープフェイクの定義

II. チープフェイクの定義

III. 両者の比較

定義

ディープフェイクは、人工知能(AI)を用いて生成された、人間の顔、声、行動や風景、物体をリアルに模倣した偽のメディアコンテンツ。
狭義の本来の定義と、広義の現在一般に用いられている定義がある

狭義のディープフェイク (本来の定義)

- 本来は、人物の顔を別の人物の顔に置き換える技術
- この定義では、技術の使用目的は、元のビデオや画像に登場する人物の顔を、別の人物の顔に非常に高いリアリズムで置き換えることに焦点を当てている

広義のディープフェイク (現在、一般に用いられている定義)

- 顔の置き換えに限らず、声の模倣、動作や表情の複製、さらには特定の人物が存在しないシナリオを生成する技術も含む
- 動画・画像だけでなく、音声・テキストコンテンツも含む
 - 音声合成により、特定の人物の架空の発現をリアルに再現
 - 表情や動作をリアルタイムで別の人物にマッピング 等

本調査研究では、より広範な被害を防止・抑制するねらいから、広義のディープフェイクを対象とする

進化の歴史

過去約10年で技術進化が進んできたが、特に2022~23年頃から一般ユーザーにも容易に生成可能になった

前史:

- 映像の顔の動きを別の音声にあわせて描写し、映画吹き替えに応用することを目的としたVideo Rewrite Projectが1997年に発表された

黎明期:

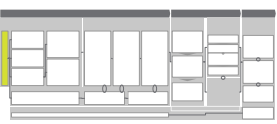
- ディープフェイクに用いられる技術 (例:敵対的生成ネットワーク/GANやオートエンコーダなど)の研究が2014年頃から公表されはじめた
- 2017年頃から、オンラインプラットフォームで、ディープフェイクが共有され始めた

発展期:

- 2018年から2020年にかけて、ディープフェイク技術は顕著に発展
- この時期には、より高品質でリアルなディープフェイクコンテンツを短時間で生成できるツールが登場

普及期:

- 2021年以降、特に生成AIが飛躍的に進化・普及した2022年から2023年にかけて、専門知識がなくても簡単にリアルなディープフェイクを作成できるツール/サービスが多数登場



ディープフェイクには、適切な使用が為されるケースもあるが、インターネット上での偽・誤情報(フェイクニュース等)の流布・拡散などの悪用による負の影響がある

ディープフェイクの概要[2/2] - 利活用状況と社会への影響

I. ディープフェイクの定義

II. チープフェイクの定義

III. 両者の比較

利活用状況

適切な使用も行われているが、悪用されるケースが拡大しつつある

適切な使用

- エンターテインメント分野でのコンテンツ制作:
映画やビデオゲームで俳優の顔を別の俳優の顔に置き換える、若返らせる又は歳を取らせるなど、リアルな特殊効果を生成する
- ヘルスケア分野での発声機能の再現:
疾患等により発声機能を失った患者に対して、予め本人の声を記録し、AIに学習させておくことで、本人の声でテキストを読み上げる 等

悪用

- 偽・誤情報(フェイクニュース)の流布:
悪意ある人物や企業への攻撃・政治的プロパガンダ・社会の扇動や、悪意が限定的ないたずらなどの目的で、偽情報を作成・流布する
- なりすましによる詐欺:
特定人物になりすました音声や映像により、ターゲットになる人物を欺くことで、資金を不正に振り込ませる、機密情報を聞き出す など

社会への影響

本調査で扱うインターネット上の偽・誤情報では、下記影響が生じている

一般個人への影響:

- リアルな偽情報による混乱と不信感の増大
- 個人のイメージの損害

著名人への影響

- 不適切発言やフェイクポルノによる名誉やキャリアの毀損
- 無許可での広告等への使用

企業への影響:

- ブランドイメージの損害による信頼性の低下
- 偽情報に基づく株価の不正な操作

政府・公共機関への影響:

- 選挙への干渉や政治的不安の引き起こし
- 偽の政治的声明やニュースによる公衆の誤解

等



(参考)フェイク画像を悪ふざけで投稿し、騙された閲覧者がSNSで拡散する事案が発生 生成された情報の悪用の具体事例

I. ディープフェイクの定義

II. チープフェイクの定義

III. 両者の比較

2022年9月、台風15号による水害被害が発生している静岡県の画像がSNS上で拡散

投稿者は「ドローンで撮影された静岡県の水害」と称し画像を投稿したが、実は画像生成AIで作成したフェイク画像と公表

- 濁流の流れや一部の建物に不自然な部分があり、投稿に対して「画像生成AIが作成した偽物ではないか？」など疑問の声が上がっていた
- 投稿者は問題の画像がフェイクだと認め、画像生成AI「Stable Diffusion」で作成したことを表明
- "Flood Damage Shizuoka"の簡単なキーワードで生成

投稿した理由について、" 大した目的はない。そもそも安易にこれが広まると想定していなかった"とコメント



生成AIの進化に伴い、フェイク画像が簡単により精度高く作れるようになり、
社会が混乱する恐れあり



チープフェイクは、映像・画像等の編集ソフトウェアで制作される偽のメディアコンテンツ。制作に映像等の編集技術を要するが、痕跡等で偽物と認識され易く、社会への影響は限られる

チープフェイクの概要

I. ディープフェイクの定義

II. チープフェイクの定義

III. 両者の比較

定義

チープフェイクは、高度なAI技術を用いず、映像・画像等の編集ソフトウェアで制作される偽のメディアコンテンツ

- 例: 速度変更、カット、シーンの再編成などにより、誤解を招くように編集された映像
- 例: サイズ・角度・色調等を調整した合成により、顔を置き換えた写真 等

特徴

制作のために、画像や映像の編集技術が必要。特に、説得力のあるものを制作するためには高いスキルが求められる。

一方で、編集の痕跡が残ったり、リアリティが欠けたりすることで、受け手が偽物と認識し易いケースが多い

社会的影響

一般個人/著名人/企業/政府等への影響が生じている

一般個人への影響:

- 個人のイメージの損害

著名人への影響

- スキャンダルの捏造による名誉の毀損

企業への影響:

- 偽広告などによるブランドイメージの毀損

政府・公共機関への影響:

- 選挙への干渉や政治的不安の引き起こし
- 偽の政治的声明やニュースによる公衆の誤解

等



チープフェイクよりディープフェイクの方が、一般個人にも作成が容易かつ説得性が高いため、社会への影響が大きくなることが懸念され、かつ、技術的対策の有効性が高い

ディープフェイクとチープフェイクの比較

I. ディープフェイクの定義

II. チープフェイクの定義

III. 両者の比較

ディープフェイク

作成に
利用できる
ツール

ディープフェイク生成用ツール

- AI(オートエンコーダ、GAN等)による、映像・画像等の元データからの自動生成



作成の
容易さ

より容易

- ツールの登場により、特別な技術が不要
- 数秒~数時間で自動生成



偽情報の
説得性

より高い

- リアル・高精度で人には見分けにくい(かつ、技術進化で急速に向上中)



社会への
影響

より大きくなるリスクがある

- 上記により、説得性の高い偽情報が、多数の人により大量に生成・流布される可能性がある



技術的対策
の有効性

より高い

- 映像・音声にアーチファクト等の特徴が現れるため、その分析により偽情報が判別し得る



チープフェイク

映像・画像などの編集ツール

- 手作業による、映像の速度変更、カット、シーンの再編成や、画像の合成 など



より難しい

- 映像・画像などの編集技術が必要
- 長い時間と手間をかけて編集

より低い

- 編集の痕跡や、コンテキストと照らした不自然さで、比較的、人にも見分け易い



より限定される

- 上記により、作成できる人も、その人が作成できる量も、説得性も、比較的に限られる

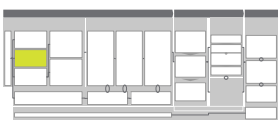
より低い

- 元の映像・音声は実物のため、機械的な真偽判別は困難(元情報との照合や背景分析等が必要)



ディープフェイク等生成する手法には、対象と形態により異なる技術がある

対象	形態	主な生成技術	概要
人	画像	フェイス・シンセシス (顔の合成)	非実在または合成された人物の顔を生成する
		フェイス・モーフィング (顔の変換)	二つ以上の顔画像を融合させて、新しい顔を生成する
	映像	フェイス・スワップ (顔の交換)	ある画像/映像から取り出した人物の顔を、別の画像/映像にある人物の顔と置き換える
		フェイシャル・リエナクトメント (顔の再現)	ある人物（ソース）の表情や動きを別の人物の顔に移し替える
		フェイス・マニピュレーション (顔の操作)	人物の顔の外見を変更または操作する
	音声	テキスト・トゥ・スピーチ (テキストの音声化)	書かれたテキストを合成音声に変換する
		ボイス・コンバージョン (音声変換)	ある人物の音声を別の人物の音声特性に変換する
		スピーチ・モーフィング (音声合成)	二つ以上の音声サンプルを組み合わせて、新しい音声を生成する
人以外 (風景、物体等)	画像	Cycle GAN	ペアになっていない画像を用いて、一方のドメイン(例:平常時の風景)の画像を他方のドメイン(例:災害時の風景)の画像に変換する
	文章	事前学習+ファインチューニング	人間のように自然な言語で様々なトピックに関する文章を生成する



(1)①ウ. 国内外におけるディープフェイク等による被害事例や被害の影響度合い、対策へのニーズの大きさ

ディープフェイクは様々な媒体を経由して、社会・企業・個人それぞれに被害をもたらす

ディープフェイクの被害の全体像と今日までの被害事例

被害状況		ディープフェイクの種類						未確認（但し、将来的に被害・影響は発生する可能性）	
		文章生成	画像生成		音声生成		動画生成		
			ヒトを生成	モノ・コト(風景等)を生成	ヒトを生成	モノ・コト(風景等)を生成	ヒトを生成	モノ・コト(風景等)を生成	
社会への攻撃	国民の扇動	①中国でChatGPTを使って列車の衝突に関する偽ニュースを拡散	②日本政府の処理水に関する誤情報が記者会見のフェイク画像と共に拡散	③アメリカ、ワシントンの国防総省近くで大規模な爆発という説明と共に偽画像をSNS上で拡散	④パレスチナ出身の父を持つ米国のモデルの声を学習し、イスラエル支持のスピーチを生成し拡散	未確認	⑤米大統領のフェイク動画を使い、非公式なプログラムへのサインアップを奨励	⑥能登半島地震の際に、架空の地名での救助要請などのデマ情報が拡散	
	政治の混乱の誘発	⑦「オバマ氏がヒラリー・クリントンの支持を撤回」という誤情報をラジオで拡散	未確認	⑧マウイ島の山火事の際に人為的な災害であることを示唆するデマ画像が拡散	⑨米大統領に似た声のフェイク音声で予備選挙に介入	未確認	⑩岸田首相が実際には発言していない内容のフェイク動画がSNSにて拡散	未確認	
企業への攻撃	企業の評判や信頼の失墜	⑪飲食店等の店舗に対しコロナ陽性者が確認されたというデマ情報が多発 ⑫AIが作成したフィッシングメールによるサイバー攻撃	未確認	⑬大手食品会社に対して「ワクチン入りの食品を販売予定」というデマ画像が拡散	⑭フェイク音声での取引先のなりすましによる振り込め詐欺事件	未確認	未確認	⑮国際的なイベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映	
個人への攻撃	個の評判や信用の失墜	⑯米国著名司会者をChatGPTで調べたところ自身が横領を働いたとする虚偽で有害な情報がでっち上げられたと主張し、Open AIを提訴	⑰著名シンガーのわいせつなフェイク画像がソーシャルメディアで拡散	未確認	未確認	未確認	⑱日本の女性芸能人150名のフェイクボロボ動画が違法に公開 ⑲知人の喫煙や飲酒をしているディープフェイク動画を嫌がらせ目的で拡散	未確認	
	利益目的の詐欺/不正行為	⑳仮想通貨取引所において偽造された本人確認書類で本人確認を突破 - パスポート、マイナンバーカード等	㉑銀行口座開設時のAIで生成した顔の画像で顔認証を突破	未確認	㉒アーティストの声を模して生成AI作成の楽曲を拡散 ㉓銀行口座開設時のAIで生成した声の音声で声帯認証を突破 ㉔フェイク音声による振り込め詐欺事件	未確認	㉕著名人の顔でディープフェイクを作成できるアプリに対しての集団訴訟 ㉖ディープフェイク技術で別人になりすまし、オンライン面接を受験	未確認	



(1)①ウ. 国内外におけるディープフェイク等による被害事例や被害の影響度合い、対策へのニーズの大きさ

広範に国民の安全を脅かす国民の扇動は対策ニーズが大きい

ディープフェイク被害の対策ニーズの初期分析

分析の前提

分析対象

国内外のディープフェイク被害事例の類型で掲載した事例26例

分析方法

対策のニーズを被害の深刻度と許容範囲で評価

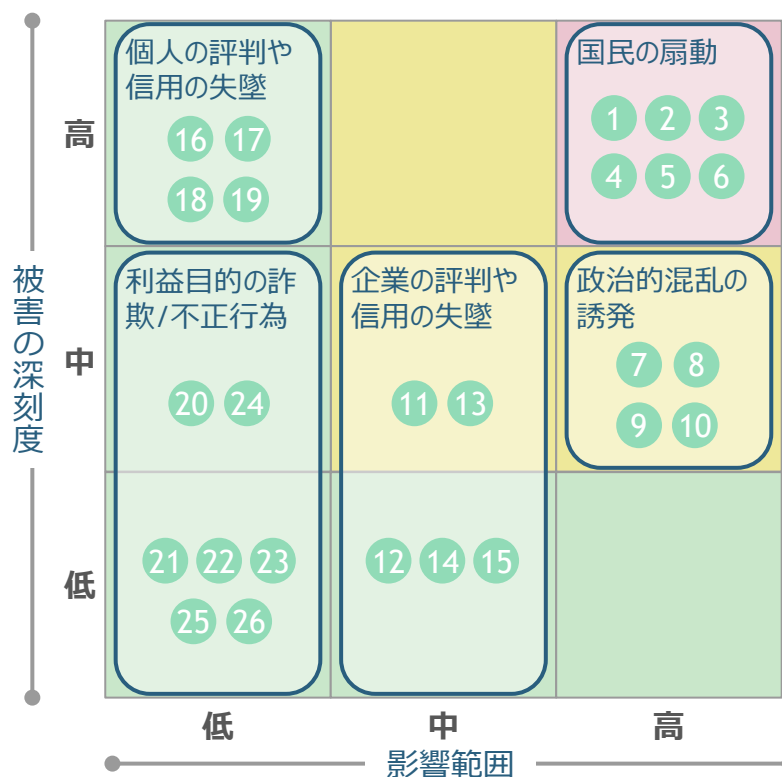
被害の深刻度

- 米国のサイバーインシデント深刻度¹を基準に三段階で評価
 - 高:** 国家の重要なインフラサービスの提供、国家政府の安定、国民の生命に差し迫った脅威をもたらす可能性がある
 - 中:** 公衆衛生または安全、国家安全保障、外交関係、市民の自由、国民の安全に影響を与える可能性がある
 - 低:** 上記に当てはまらない

影響範囲

- 被害発生時の影響範囲を三段階で評価
 - 高:** 日本国/世界全体に影響
 - 中:** 特定地域/産業に影響
 - 低:** 個人,限定的なコミュニティに影響

分析結果



対策ニーズ大

- 中国でChatGPTを使って列車の衝突に関する偽ニュースを拡散
- 日本政府の処理水に関する誤情報が記者会見のフェイク画像と共に拡散
- アメリカ、ワシントンの国防総省近くで大規模な爆発という説明と共に偽画像をSNS上で拡散
- パレスチナ出身の父を持つ米国のモデルの声を学習し、イスラエル支持のスピーチを生成し拡散
- 米大統領のフェイク動画を使い、非公式なプログラムへのサインアップを奨励
- 能登半島地震の際に、架空の地名での救助要請などのデマ情報が拡散

対策ニーズ中

- 「オバマ氏がヒラリー・クリントンの支持を撤回」という誤情報をラジオで拡散
- ハワイ島の山火事の際に人為的な災害であることを示唆するデマ画像が拡散
- 米大統領に似た声のフェイク音声で予備選挙に介入
- 岸田首相が実際には発言していない内容のフェイク動画がSNSにて拡散
- 飲食店等の店舗に対しコロナ陽性者が確認されたというデマ情報が多発
- 大手食品会社に対して「ワクチン入りの食品を販売予定」というデマ画像が拡散

対策ニーズ小

- AIが作成したフィッシングメールによるサイバー攻撃
- フェイク音声での取引先のなりすましによる振り込み詐欺事件
- 国際的なイベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映
- 米国著名司会者をChatGPTで調べたところ自身が横領を働いたとする虚偽で有害な情報がでっち上げられたと主張し、Open AIを提訴
- 著名シンガーのわいせつなフェイク画像がソーシャルメディアで拡散
- 日本の女性芸能人150名のフェイクポルノ動画が違法に公開
- 知人の喫煙や飲酒をしているディープフェイク動画を嫌がらせ目的で拡散
- 仮想通貨取引所において偽造された本人確認書類で本人確認を突破
 - パスポート、マイナンバーカード等
- 銀行口座開設時のAIで生成した顔の画像で顔認証を突破
- アーティストの声を模して生成AI作成の楽曲を拡散
- 銀行口座開設時のAIで生成した声の音声で声帯認証を突破
- フェイク音声による振り込み詐欺事件
- 著名人の顔でディープフェイクを作成できるアプリに対する集団訴訟
- ディープフェイク技術で別人になりすまし、オンライン面接を受験

1. <https://www.cisa.gov/news-events/news/cisa-national-cyber-incident-scoring-system-nciss>



(1)①エ. ディープフェイク等に対する
技術的な対応策

ディープフェイク対策は配布されたコンテンツの真贋判定と配布するコンテンツの真正性や信頼性を保証する技術が存在

ディープフェイクへの対応技術		技術概要	技術が適用可能なコンテンツ							
			文章	画像	音声	動画				
コンテンツの 真偽判別 支援技術	コンテンツ 単独による 検知	1 直接的 真偽判別	コンテンツ内の情報の位置関係 (空間的な歪み) と時系列で見たときの歪みを解析し、 ディープフェイクを見破る技術 - 空間的な歪み - 照明や背景の不自然さ - 生体的な特徴 (顔の輪郭や構造)の不自然さ - 敵対的生成ネットワーク (GAN) の処理の痕跡 - 時系列で見たときの不自然さ - 異常なジェスチャーや行動 - 隣接するフレーム間での生物学的な特徴の不一致 - まばたきのパターンなど、生体信号の欠如 - フレーム間のジッターなどの歪み - スピーチデータから推定された声帯データに基づく音声的な歪み				—	○	○	○
		2 間接的 真偽判別	蓄積したデータを基にコンテンツを分類し、ディープフェイクを見破る技術 - データドリブンでのディープフェイクの検知 - CNN/DNNによる分類				—	○	○	○
		3 複合的なコンテンツによる 真偽判別	インターネット上の情報等、複数のコンテンツを組み合わせたディープフェイクの検出技術 疑義言説収集 等				○	○	○	○
コンテンツの 真正性 保証・ 信頼性判断 支援技術	真正性 保証	4 電子署名	電子的なコンテンツに署名を付与することで、誰が何を作成したか真正性を証明する技術 デジタル証明書 等				○	○	○	○
	信頼性 判断支援	5 電子透かし (知覚困難型)	コンテンツに対して人間の近くでは判断できない形で任意の情報を埋め込み、不正利用や 改ざんを防止・検知することでコンテンツの信頼性を担保する技術				○	○	○	○



(1)①オ. 対策技術を有する国内事業者・研究機関のマッピングと技術的課題の分析

ディープフェイク対策技術はそれぞれの技術固有の課題や制約があり、社会実装に向けては単一の技術だけを使うのではなく複数の技術を組み合わせる必要

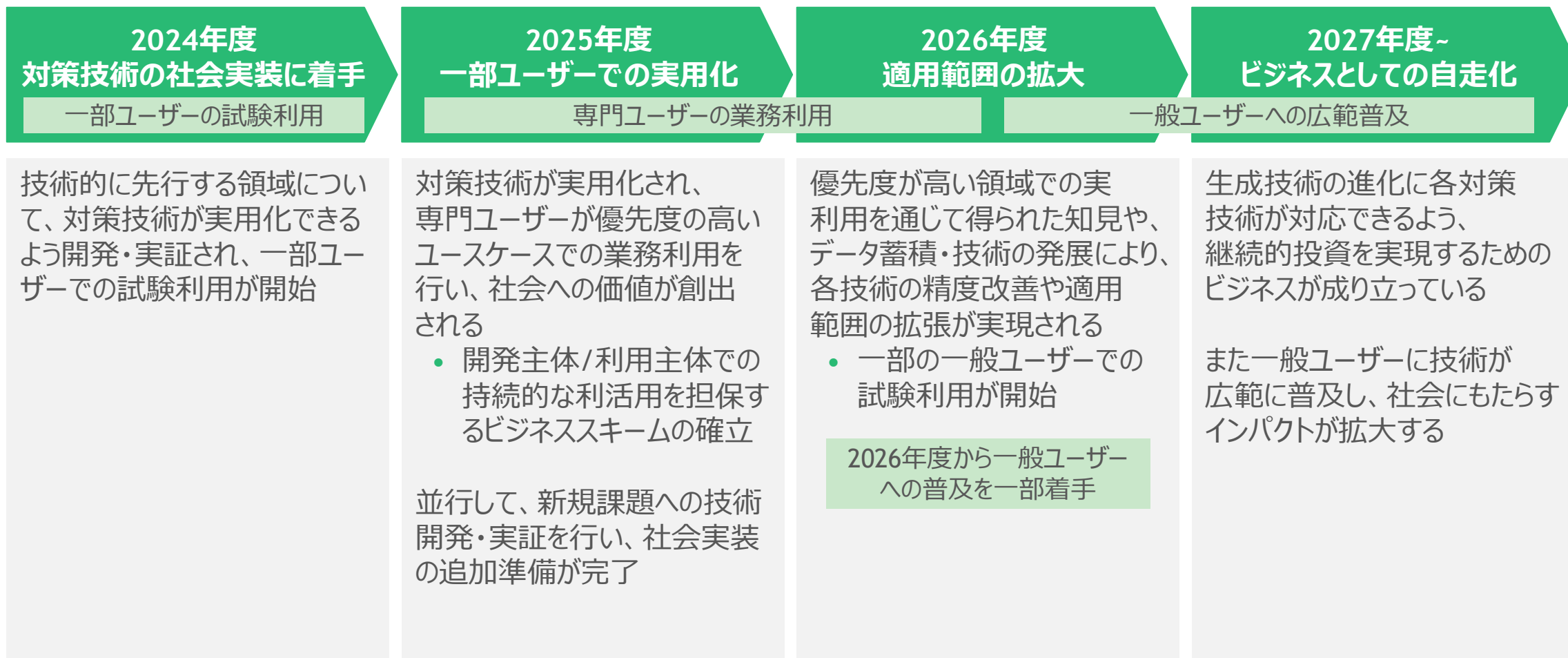
技術の研究開発の現状と今後の方向性

ディープフェイクへの対応技術			現在実現できていること	課題・制約	今後の研究開発の方向性
コンテンツの真偽判別支援技術	コンテンツ単独による検知	直接的真偽判別	特定の状況下であれば90%以上の精度で人物のディープフェイクを検知可能	- ロバスト性が低い - 画質(ノイズ, 解像度)に検出精度が左右されやすい - 特徴量が特異な対象(精神病患者の表情等)が検知できないケースが報告されている	ロバスト性の向上
		間接的真偽判別	特定の状況下であれば90%以上の精度で人物のディープフェイクを検知可能	- 学習データにない未知のシナリオ/精製方法の検知精度が悪い - 説明可能性(なぜその答えを出したのかを説明できる能力の高さ)と精度にトレードオフがある	データ量を増やすことによる未知のシナリオの潰しこみ
	複合的なコンテンツによる真偽判別		特定の状況下で80%以上の感度でフェイクとして要検証なニュースを検知可能	検知精度が低い	検知精度の向上
コンテンツの真正性保証・信頼性判断支援技術	真正性保証	電子署名	コンテンツへの電子署名の付与	- データ量が増えて処理の負荷が大きい - 量子コンピュータ等により従来の暗号化技術では改ざんのリスクがある	- データ量の圧縮 - より強固な改ざん防止技術の開発 - ブロックチェーン等
	信頼性判断支援	電子透かし(知覚困難型)	- 一定のコンテンツの加工に耐えうる電子透かしの付与 - トリミング、非可逆圧縮、色の変更等	- 電子透かし方式の公開が難しく、普及させづらい - 電子透かしの破壊のリスクがある - 特にテキスト向けの電子透かしは容易に破壊されやすい	堅牢性の向上 ※アルゴリズム秘匿と標準化の両立は現状解ナシ



(1)①カ.インターネット上の
偽・誤情報対策技術の進展に係る
ロードマップの作成

25年度は、対策技術の社会実装を加速させ、喫緊の課題となるユースケースで成果創出。
段階的に対策技術の適用範囲を拡大し、中期的にビジネスとしての自走化を目指す





(1)①カ.インターネット上の
偽・誤情報対策技術の進展に係る
ロードマップの作成

ユースケースの中でも、市場性を見込みやすい領域から社会実装が拡大する見立て。 優先度高の領域の中では、「詐欺」・「企業評判」の領域から立ち上がり社会実装を拡大 社会実装(ユースケース)観点のロードマップ

			2024年度 対策技術の社会実装に着手	2025年度 一部ユーザーでの実用化	2026年度 適用範囲の拡大	2027年度~ ビジネスとしての自走化
目標			先行領域：一部ユーザーでの試験利用 新規領域：研究が進捗	優先度が高い領域で 専門ユーザーが業務利用	適用範囲が拡大し精度が向上。 専門ユーザーの業務利用範囲 拡大／一般ユーザーの試用開始	継続的な技術開発・投資に 向けてビジネスモデルが確立
情報空間	優先度：高	災害	必要な要素技術について、 社会実装に資する形で研究が 行われ、実証実験が進む	利用主体を含むコンソーシアムが 組成され、一部ユーザーの利用 が進む	感度の高い / 親和性の高い 利用主体(自治体等)での活用 が広がる	先行事例の確立に伴い、幅広い自治体での活用が進み、専門ユーザーの業務利用が拡大
		詐欺 企業評判		技術を有する企業が、顧客接点を有する企業と組み、一部ユーザーの利用が進む 企業向けの多様な攻撃シナリオへの対処	顧客接点を有する企業との拡販により、広い事業者等で専門ユーザーが業務利用を行う 個人向けのチープフェイクからディープフェイクへの対処	企業ユーザーが拡大し、ビジネスとして確立される / 広いユーザーに広範に普及する
	優先度：中	平時扇動 政治 個人評判 肖像権	要素技術について、社会実装に資する形で研究が行われる	幅広いユースケースをカバーする対策技術の活用において、 実証実験が行われる	幅広いユースケースをカバーする対策技術の活用において、 利用ユーザーが段階的に拡大	
(参考) 非情報空間	eKYC 損害保険		技術開発と同時進行で、既存サービス (例：eKYCサービス、損保申請、等)に対策技術が組み込まれ始める	活用の広がりに伴い、データが蓄積され、更に技術精度が高まる。 加えて、近接領域での活用が拡大する		



(1)①カ.インターネット上の偽・誤情報対策技術の進展に係る
ロードマップの作成

PFerが取り組む「拡散防止」や、既に商用化している「流通情報の健全性把握」が先行。 次いで、「発信情報の真偽判別」や「発信者・発信情報の真正性担保」が段階的に発展 対策技術観点のロードマップ

		2024年度 対策技術の社会実装に着手	2025年度 一部ユーザーでの実用化	2026年度 適用範囲の拡大	2027年度~ ビジネスとしての自走化
目標		先行領域：一部ユーザーでの試験利用 新規領域：研究が進捗	優先度が高い領域で 専門ユーザーが業務利用	適用範囲が拡大し精度が向上。 専門ユーザーの業務利用範囲 拡大/一般ユーザーの試用開始	継続的な技術開発・投資に 向けてビジネスモデルが確立
流通情報の健全性把握	情報流通状況の可視化	ソーシャルリスクニングは商用レベルで活用。但し、ディープフェイク対策とは必ずしも連動せず	ソーシャルリスクニングは商用レベルで活用。一部のユースケースでは、検知との連動を実現	検知と連動するユースケースを拡大	
二重・誤情報の作成防止	ファイル防御・改ざん修復	基礎研究が進む	基礎研究・要素技術開発を実施。 一部優先領域で、試験利用着手	一部優先領域での専門ユーザーによる、業務利用が開始される	他ユースケースでの業務利用を拡大
発信情報の真偽判別	生成・改変有無の検知別	要素技術開発と技術検証、 一部ユーザーでの試験利用着手	一部優先領域での専門ユーザーによる、業務利用が開始される ・ 音声の多様なシナリオへ対処	他のユースケースでの活用が拡張 また検知対象のモーダルも拡大	一般ユーザーに広範に普及し、 ビジネスとして自走化
	情報収集・データベース提供	要素技術開発と技術検証が進む	検知技術と連携して、精度向上の 技術検証と試験利用を実施	一部優先領域での専門ユーザーによる、業務利用が開始される	検知技術とソーシャルリスクニングツールと連携して業務利用を拡大
発信者・発信情報の真正性担保	情報発信者の認証付与・確認 発信情報の認証付与・確認	横展開を見据えた技術開発と、 一部ユーザーでの試験利用が進む 要素技術開発と技術検証、 一部ユーザーでの試験利用着手	一部優先領域での専門ユーザーによる、業務利用を推進 (発信情報は対象モーダルの技術開発を拡大) ・ 利用者向けの簡便な 検証基盤の構築	他のユースケースでの活用が拡張	一般ユーザーに広範に普及し、 ビジネスとして自走化
拡散防止	拡散者へのアラート・投稿制限	生成AIの普及に伴い、PFは自社のSNSサービス等に対策技術を組み込み	生成側の進化や各国政府からの規制に合わせて、継続的に安全対策の技術開発・実装を強化		



(1)②ア.技術的に対処すべき
被害状況の整理

(再掲) 広範に国民の安全を脅かす国民の扇動は対策ニーズが大きい

ディープフェイク被害の対策ニーズの初期分析

分析の前提

分析対象

国内外のディープフェイク被害事例の類型で掲載した事例26例

分析方法

対策のニーズを被害の深刻度と許容範囲で評価

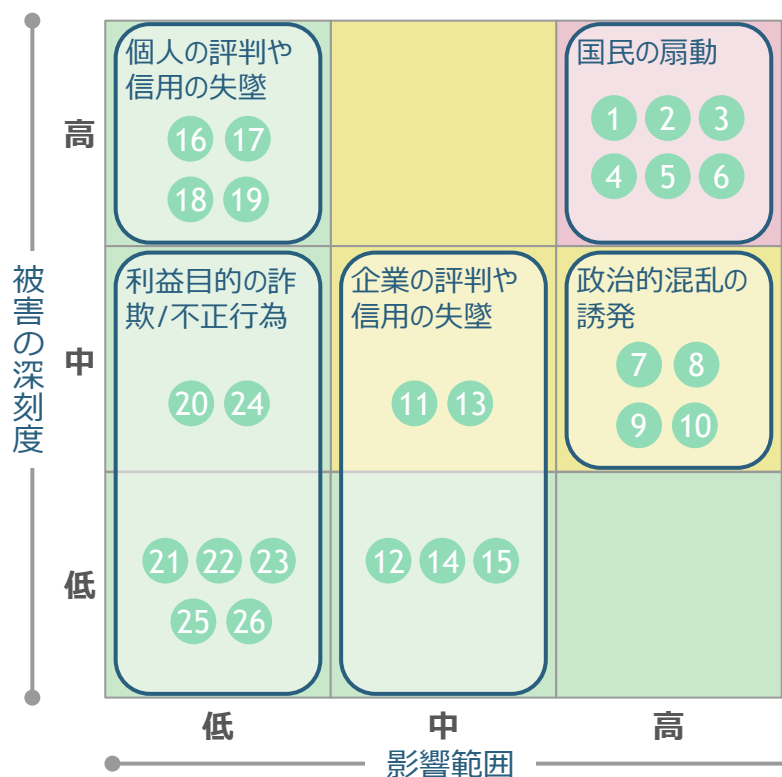
被害の深刻度

- 米国のサイバーインシデント深刻度¹を基準に三段階で評価
 - 高:** 国家の重要なインフラサービスの提供、国家政府の安定、国民の生命に差し迫った脅威をもたらす可能性がある
 - 中:** 公衆衛生または安全、国家安全保障、外交関係、市民の自由、国民の安全に影響を与える可能性がある
 - 低:** 上記に当てはまらない

影響範囲

- 被害発生時の影響範囲を三段階で評価
 - 高:** 日本国/世界全体に影響
 - 中:** 特定地域/産業に影響
 - 低:** 個人,限定的なコミュニティに影響

分析結果



対策ニーズ大

- ① 中国でChatGPTを使って列車の衝突に関する偽ニュースを拡散
- ② 日本政府の処理水に関する誤情報が記者会見のフェイク画像と共に拡散
- ③ アメリカ、ワシントンの国防総省近くで大規模な爆発という説明と共に偽画像をSNS上で拡散
- ④ パレスチナ出身の父を持つ米国のモデルの声を学習し、イスラエル支持のスピーチを生成し拡散
- ⑤ 米大統領のフェイク動画を使い、非公式なプログラムへのサインアップを奨励
- ⑥ 能登半島地震の際に、架空の地名での救助要請などのデマ情報が拡散

対策ニーズ中

- ⑦ 「オバマ氏がヒラリー・クリントンの支持を撤回」という誤情報をラジオで拡散
- ⑧ マウイ島の山火事の際に人為的な災害であることを示唆するデマ画像が拡散
- ⑨ 米大統領に似た声のフェイク音声で予備選挙に介入
- ⑩ 岸田首相が実際には発言していない内容のフェイク動画がSNSにて拡散
- ⑪ 飲食店等の店舗に対しコロナ陽性者が確認されたというデマ情報が多発
- ⑬ 大手食品会社に対して「ワクチン入りの食品を販売予定」というデマ画像が拡散

対策ニーズ小

- ⑫ AIが作成したフィッシングメールによるサイバー攻撃
- ⑭ フェイク音声での取引先のなりすましによる振り込み詐欺事件
- ⑮ 国際的なイベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映
- ⑯ 米国著名司会者をChatGPTで調べたところ自身が横領を働いたとする虚偽で有害な情報がでっち上げられたと主張し、Open AIを提訴
- ⑰ 著名シンガーのわいせつなフェイク画像がソーシャルメディアで拡散
- ⑱ 日本の女性芸能人150名のフェイクポルノ動画が違法に公開
- ⑲ 知人の喫煙や飲酒をしているディープフェイク動画を嫌がらせ目的で拡散
- ⑳ 仮想通貨取引所において偽造された本人確認書類で本人確認を突破
 - パスポート、マイナンバーカード等
- ㉑ 銀行口座開設時のAIで生成した顔の画像で顔認証を突破
- ㉒ アーティストの声を模して生成AI作成の楽曲を拡散
- ㉓ 銀行口座開設時のAIで生成した声の音声で声帯認証を突破
- ㉔ フェイク音声による振り込み詐欺事件
- ㉕ 著名人の顔でディープフェイクを作成できるアプリに対しての集団訴訟
- ㉖ ディープフェイク技術で別人になりすまし、オンライン面接を受験

1. <https://www.cisa.gov/news-events/news/cisa-national-cyber-incident-scoring-system-nciss>



(1)②ア.技術的に対処すべき
被害状況の整理

対策ニーズの大きい国民の扇動を目的としたディープフェイクのうち、特に人物を生成AIの生成対象とし、画像や音声、動画による被害内容は技術的な対策の有効性が高い

技術的な対策の有効性の評価の考え方

技術的な対策の有効性は事例内で複雑性(生成AIが生成する対象)と情報量(生成AIのアウトプット形態)で評価可能

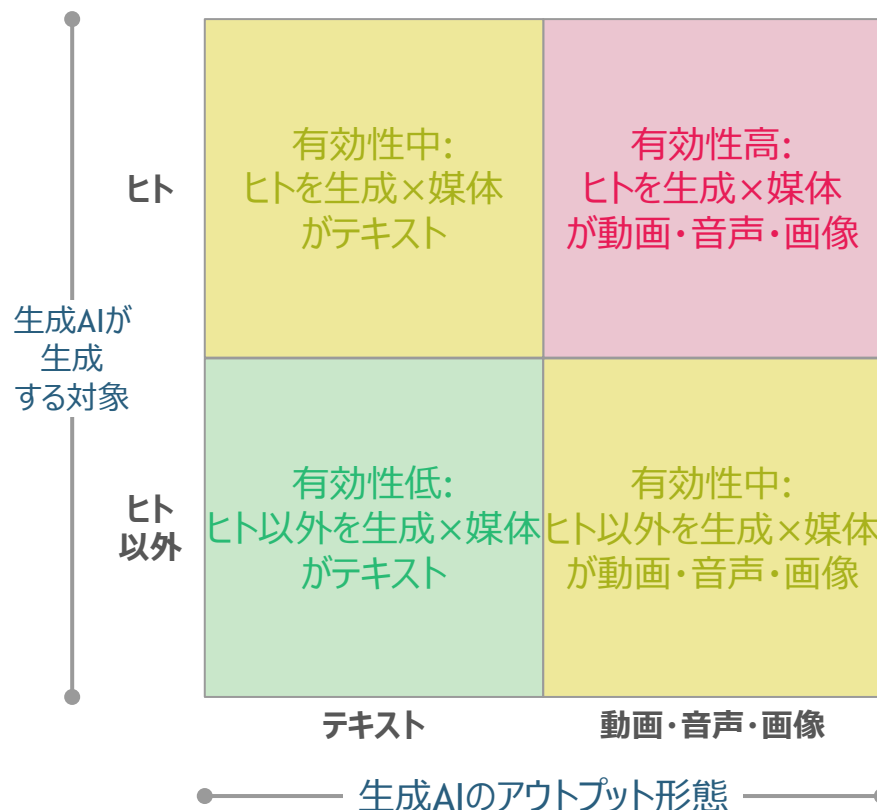
複雑性:生成AIが生成する対象

- 生成AIの生成対象が特徴量が明確なヒトの場合はヒト以外(風景等)を生成するよりもより高精度に検知可能なため

情報量:生成AIのアウトプット形態

テキストか、動画・音声・画像か

- 生成AIが生成するアウトプットの形態は情報がより多く含まれる動画・音声・画像の方がテキストよりも相対的に検知精度は高くなるため



国民の扇動を目的とした偽・誤情報のうち
対策技術の開発・実証に取り組むべき被害状況

優先度高

- 国民の扇動を目的とし、動画・音声・画像を媒体としてヒトを生成する偽・誤情報
 - 日本政府の処理水に関する誤情報が記者会見のフェイク画像と共に拡散 など

優先度中

- 国民の扇動を目的とし、動画・音声・画像を媒体として風景を生成する偽・誤情報
 - アメリカ、ワシントンの国防総省近くで大規模な爆発という説明と共に偽画像をSNS上で拡散 など
- 国民の扇動を目的とし、テキストを媒体としてヒトになりすます偽・誤情報

優先度低

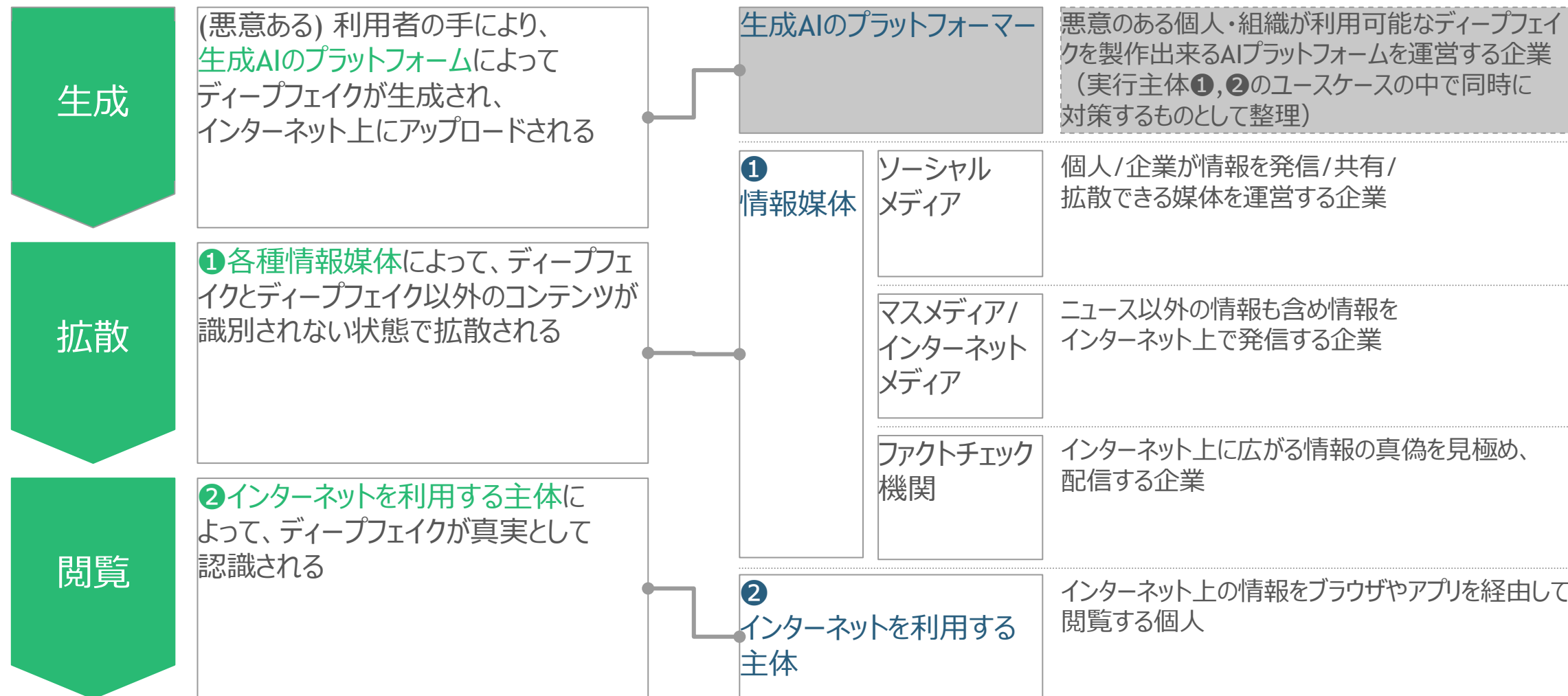
- 国民の扇動を目的とし、テキストを媒体として風景等ヒト以外に言及する記事などを生成する偽・誤情報
 - 中国でChatGPTを使って列車の衝突に関する偽ニュースを拡散 など



ディープフェイクが流布するまでのプロセスを踏まえ、対策の実行主体を洗い出し

ディープフェイクが流布するプロセス

対策の実行主体





技術的に対処すべき被害状況と対策実行の主体を基にユースケースを設定

技術的に対処すべき被害状況

偽・誤情報の流布の目的

- 国民の扇動
 - 日本政府の処理水に関する誤情報が音声と記者会見の画像/動画と共に拡散
 - マウイ島の山火事の際に人為的な災害であることを示唆するデマ画像が拡散など

生成AIが生成する対象

- ヒト (または風景)

情報媒体

- 動画、画像、音声

対策実行の主体

情報媒体

ソーシャル
メディア

ファクトチェック
機関

インターネットを利用する主体

ユースケース

Ⅱ ソーシャルメディアが、自社の提供するプラットフォーム上を流通する人物及び風景・街並みに関する画像・動画等のコンテンツのうち、自社のポリシーに反する投稿（著名人を騙ったコンテンツ等）を削除する用途で、生成AI によるものかについて確認を行う場面

Ⅰ ファクトチェック機関が、ソーシャルメディア上の投稿の人物及び風景・街並みに関する画像・動画等のコンテンツが、生成AI により生成されたものかについて確認を行う場面

Ⅲ インターネットを利用する主体が、インターネット上の情報媒体を閲覧する際に、情報コンテンツの作成者や配信サイト運営者、広告主といった発信者情報や情報コンテンツの信頼性の確認を行う場面



前項で特定したシーンについて、対策技術を技術的に実装する場合の方法を幅出し

ユースケース案

ソーシャルメディア運営企業が、
自社の提供するプラットフォーム上を流通する人物及び風景・街並みに
関する画像・映像等のコンテンツのうち、自社のポリシーに反する投稿を
削除する用途で、
生成AI によるものかについて確認を行う場面



真偽判定支援技術

① ソーシャルメディアのプラットフォームに真偽判別
検知機能を組み込み、投稿された情報コンテンツ
について生成AIか否かの真偽判別を行う

真正性保証・信頼性判断支援技術

① ソーシャルメディアのプラットフォームに電子透かしや電子証明書の読み込み機能を組み込み、生成AIか否かも含めコンテンツの信頼性を確認する

マスメディア/インターネットメディアが、
ソーシャルメディア上の投稿の人物及び風景・街並みに関する画像・映像等のコンテンツのうち、真偽の不透明な記事を削除/訂正する用途で、
生成AI により生成されたものかについて確認を行う場面



② 各マスメディア/インターネットメディアが保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて生成AIか否かの真偽判別を行う

② 各マスメディア/インターネットメディアが保有するシステムに電子透かしや電子証明書の読み込み機能を組み込み、生成AIか否かも含めコンテンツの信頼性を確認する

ファクトチェック機関が、
ソーシャルメディア上の投稿の人物及び風景・街並みに関する画像・映像等のコンテンツが、
生成AI により生成されたものかについて確認を行う場面



③ ファクトチェック機関が保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて生成AIか否かの真偽判別を行う

③ ファクトチェック機関が保有するシステムに電子透かしや電子証明書の読み込み機能を組み込み、生成AIか否かも含めコンテンツの信頼性を確認する

インターネットを利用する主体が、
インターネット上の情報媒体を閲覧する際に、情報コンテンツの作成者や
配信サイト運営者、広告主といった発信者情報や情報コンテンツの
信頼性の確認を行う場面



④ 個人の端末のアプリ・ブラウザに真偽判別検知機能を組み込み、端末にダウンロードされた情報コンテンツについて真偽判別を行い、信頼性を確認する

④ 個人の端末のアプリ・ブラウザに電子透かしや電子証明書の読み込み機能を組み込み、生成AIか否かも含めコンテンツの信頼性を確認する



何れのシーンも、真偽判別技術または真正性保証・信頼性判断技術で技術的な対策は可能だが、シーンにより技術的な実装の実現性が低いものもあると考えられる

(再掲) 対策技術を使用する場合に想定される技術的な実装方法の幅出し

ソーシャルメディア	①ソーシャルメディアのプラットフォームに真偽判別検知機能を組み込み、投稿された情報コンテンツについて生成AIか否かの真偽判別を行う 真偽判定	➤ 実現性あり ・ 投稿される情報コンテンツの真偽判別は誤・偽情報の流布防止に効果的であり、プラットフォームのシステムリソースは潤沢であるため実装上のハードルも低い
	①ソーシャルメディア上のプラットフォームに電子透かしや電子証明書の読み込み機能を組み込み、電子透かしや電子証明書を通じて生成AIか否かの真偽判別を行う 真正性保証・信頼性判断	➤ 実現性あり ・ 投稿される情報コンテンツの真正性・信頼性の明示は誤・偽情報の流布防止に効果的であり、実装上のハードルも低い
マスメディア/インターネットメディア	②各マスメディア/インターネットメディアが保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて生成AIか否かの真偽判別を行う 真偽判定	➤ 実現性あり ・ マスメディア/インターネットメディアの真偽判別技術導入はインターネットを利用する主体への誤・偽情報の伝達を防ぐことができ、プラットフォームも企業で構築しうる
	②各マスメディア/インターネットメディアが保有するシステムに電子透かしや電子証明書の読み込み機能を組み込み、電子透かしや電子証明書を通じて生成AIか否かの真偽判別を行う 真正性保証・信頼性判断	➤ 実現性が低い (インパクトが小さい) ・ マスメディア/インターネットメディアは情報を発信する立場で真正性・信頼性が明示されない情報コンテンツが取材対象であるため、実装してもインパクトが見込みづらい
ファクトチェック機関	③ファクトチェック機関が保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて生成AIか否かの真偽判別を行う 真偽判定	➤ 実現性あり ・ ファクトチェック機関での真偽判別技術導入はインターネットを利用する主体への誤・偽情報の伝達を防ぐことができ、プラットフォームも機関で構築しうる
	③ファクトチェック機関が保有するシステムに電子透かしや電子証明書の読み込み機能を組み込み、生成AIか否かも含めコンテンツの信頼性を確認する 真正性保証・信頼性判断	➤ 実現性あり ・ 真正性保証・信頼性の検証は端末・プラットフォーム側に高性能を要求しないため、実現可能性は高い
インターネットを利用する主体	④個人の端末のアプリ・ブラウザに真偽判別検知機能を組み込み、端末にダウンロードされた情報コンテンツについて真偽判別を行い、信頼性を確認する 真偽判定	➤ 実現性が低い ・ 真偽判別技術を導入できる性能水準を個人が保有する端末で満たすことは現時点困難であると想定される
	④個人の端末のアプリ・ブラウザに電子透かしや電子証明書の読み込み機能を組み込み、生成AIか否かも含めコンテンツの信頼性を確認する 真正性保証・信頼性判断	➤ 実現性あり ・ 真正性保証・信頼性の検証は端末・プラットフォーム側に高性能を要求しないため、実現可能性は高い

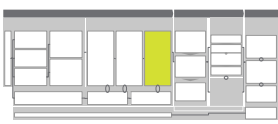


開発・実証は、ユースケース(利用主体)と対策技術の2つの切り口が考えられる

各ユースケースにおける対策技術の詳細化

			ソーシャルメディア	マスメディア/インターネットメディア	ファクトチェック機関	インターネットを利用する主体
			ユースケース単位の進め方			現時点実現困難と想定
コンテンツの真偽判別支援技術	コンテンツ単独による検知	直接的真偽判別	ソーシャルメディアのプラットフォームにコンテンツ単独での真偽判別検知機能を組み込み、投稿されたコンテンツについて真偽判別を行う	各マスメディア/インターネットメディアが保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて真偽判別を行う	各機関のシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて真偽判別を行う	
		間接的真偽判別				
複合的なコンテンツによる真偽判別			ソーシャルメディアのプラットフォームに複合的なコンテンツによる真偽判別機能を組み込み、投稿されたコンテンツについて真偽判別を行う	各メディアが保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて真偽判別を行う	各メディアが保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて真偽判別を行う	
コンテンツの真正性保証・信頼性判断支援技術	真正性保証	電子署名	ソーシャルメディアのプラットフォームに電子署名の読み込み機能を組み込み、投稿されたコンテンツについて真正性を示す	現時点実現困難と想定	各メディアが保有するシステムに電子署名の読み込み機能を組み込み、収集したコンテンツについて真正性を示す	個人が保有する端末に電子署名の読み込み機能を組み込み、収集したコンテンツについて真正性を示す
	信頼性判断支援	電子透かし(知覚困難型)	ソーシャルメディアのプラットフォームに電子透かしの読み込み機能を組み込み、投稿されたコンテンツについて信頼性を示す		各メディアが保有するシステムに電子透かしの読み込み機能を組み込み、投稿されたコンテンツについて信頼性を示す	個人が保有する端末に電子透かしの読み込み機能を組み込み、投稿されたコンテンツについて信頼性を示す

27



(参考) なお、ユースケースを切り口に開発・実証を行う場合、1つのユースケースでも複数の対応技術が利用し得るため、開発・実証テーマの数が膨大になる可能性がある

ユースケース	技術的な対策方針			想定される技術的な実装方法	技術的な課題
	対策技術				
Ⅱ ソーシャルメディアが、自社の提供するプラットフォーム上を流通する人物及び風景・街並みに関する画像・動画等のコンテンツのうち、自社のポリシーに反する投稿（著名人を騙ったコンテンツ等）を削除する用途で、生成AI によるものかについて確認を行う場面	真偽判別支援技術	コンテンツ単独による検知	直接的 真偽判別	ソーシャルメディアのプラットフォームに真偽判別検知機能を組み込み、投稿された情報コンテンツについて生成AIか否かの真偽判別を行う	● ロバスト性が低い
			間接的 真偽判別		● 未知のシナリオの検知精度が低い ● 説明可能性と精度にトレードオフがある
	真正性保証・信頼性判断支援技術	複合的なコンテンツによる真偽判別		プラットフォームに複合的なコンテンツによる真偽判別を組み込み、真偽判別を行う	● 検知精度が低い
		真正性保証	電子署名		● 量子コンピュータ等により従来の暗号化技術では改ざんのリスクがある
Ⅰ ファクトチェック機関が、ソーシャルメディア上の投稿の人物及び風景・街並みに関する画像・動画等のコンテンツが、生成AI により生成されたものかについて確認を行う場面	真偽判別支援技術	コンテンツ単独による検知	直接的 真偽判別	ファクトチェック機関が保有するシステムに真偽判別検知機能を組み込み、収集した情報コンテンツについて生成AIか否かの真偽判別を行う	● 依然として電子透かしの破壊のリスクがある ● 電子透かし方式の公開が難しく、普及させづらい
			間接的 真偽判別		● ロバスト性が低い
	真正性保証・信頼性判断支援技術	複合的なコンテンツによる真偽判別		各機関のシステムに電子署名の読み込み機能を組み込み、真正性を確認する	● 未知のシナリオの検知精度が低い ● 説明可能性と精度にトレードオフがある
		真正性保証	電子署名		● 検知精度が低い
Ⅲ インターネットを利用する主体が、インターネット上の情報媒体を閲覧する際に、情報コンテンツの作成者や配信サイト運営者、広告主といった発信者情報や情報コンテンツの信頼性の確認を行う場面	真偽判別支援技術	コンテンツ単独による検知	直接的 真偽判別	個人保有の端末に電子署名の読み込み機能を組み込み、真正性を確認する	● 未知のシナリオの検知精度が低い ● 説明可能性と精度にトレードオフがある
			間接的 真偽判別		● 未知のシナリオの検知精度が低い ● 説明可能性と精度にトレードオフがある
	真正性保証・信頼性判断支援技術	複合的なコンテンツによる真偽判別		個人保有の端末に電子署名の読み込み機能を組み込み、真正性を確認する	● 量子コンピュータ等により従来の暗号化技術では改ざんのリスクがある
		真正性保証	電子署名		● 依然として電子透かしの破壊のリスクがある ● 電子透かし方式の公開が難しく、普及させづらい

Agenda

背景・目的、実施内容、スケジュール

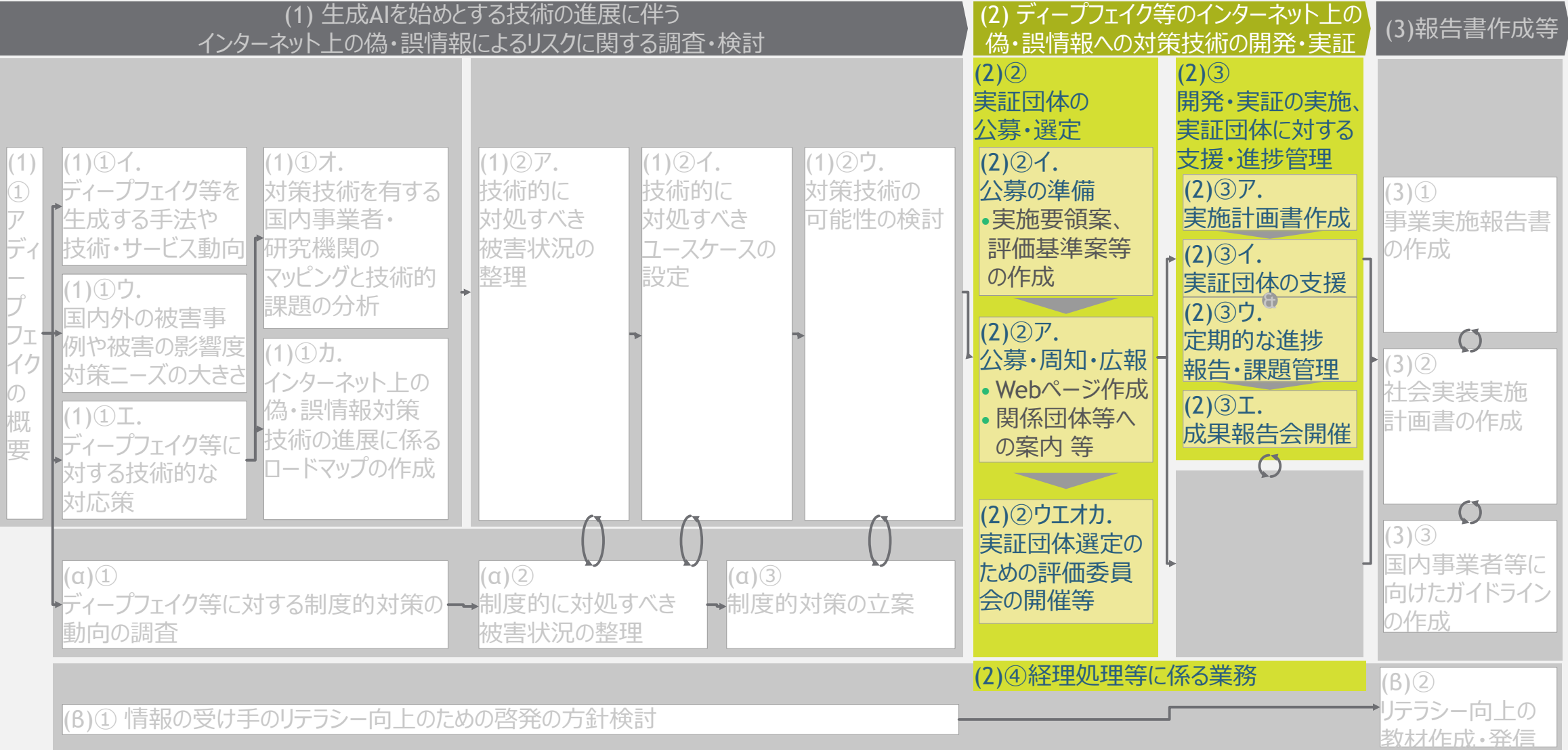
(1) 生成AIを始めとする技術の進展に伴うインターネット上の偽・誤情報によるリスクに関する調査・検討

➤ (2) ディープフェイク等のインターネット上の偽・誤情報への対策技術の開発・実証

(3) 報告書作成等

その他 検討事項

実施内容





(2)① 業務の前提

ディープフェイク等のインターネット上の偽・誤情報への対策技術の開発・実証の業務の前提は、以下のとおり

ア 実施主体

開発・実証の実施主体は、請負者が公募し、外部有識者等で構成する評価委員会の開催を経て選定する事業者等（以下「実証団体」という。）とする。
実証団体は、請負者が3．(1)②イにより整理したユースケースを基に、インターネット上の偽・誤情報が問題となる適切なユースケースを設定した上で、対策技術を開発・実証する。

イ 対象となる技術

3.(1)②により特定したユースケース及び対策技術を対象とする。

なお、3 (1)②のユースケース例における、インターネット上で流通する偽・誤情報への対策技術の開発・実証を想定しているが、これらのユースケースに限定せず、更なるユースケースの提案を行い、主管課と協議の上、設定する。

ウ 実施件数・事業規模

少なくとも4件程度の実証の実施を想定する。

また、1件当たりの事業規模は8千万円～2億円程度を目安とし、総額で4.5億円(税込)を上限とし、残額が生じた場合は、他項目に流用せず精算する。
見積もりの作成に当たっては、上記の上限額を見込んだ額を計上する。

なお、公募・選定に際して、評価基準に基づいて、提案内容・規模の評価を実施する予定。

エ スケジュール

実証の実施スケジュールは以下のとおり。

- 令和6年4月～5月頃 公募
- 令和6年5月～6月頃 実証団体の選定、決定通知
- 令和7年2月頃 成果報告書案の提出
- 令和7年3月頃 最終報告会の開催

最終的なスケジュールは、契約締結後、主管課と請負者で協議の上で決定するものとする。また、状況に応じてスケジュールを変更する可能性がある。

なお、募集状況によっては2次公募、3次公募を実施する場合がある。



実施要領 (R5補正 使用資料)



実施要領 (1/6)

下記リンク先で公開

https://www.soumu.go.jp/main_content/000944632.pdf

インターネット上の偽・誤情報対策技術の開発・実証事業 実施要領

1 事業の目的

近年、対話型言語モデル「Chat GPT」や、テキストを入力すると画像を生成する「プロンプト型画像生成 AI」、テキストから作曲を行う AI 等、多様な用途での生成 AI が登場している。生成 AI やディープフェイク技術により、偽画像・動画を誰でも容易に作成できるようになり、偽のニュース動画や偽の災害画像、詐欺サイト等、人の目では真偽を見分けることが困難な情報に、国民が SNS 等を通じて日常的に触れ得る機会が増加している。

政府内でも、AI 戦略チームや AI 戦略会議において、我が国における生成 AI をめぐる課題について検討しており、2023 年 5 月 26 日付で公表された「AI に関する暫定的な論点整理」では、偽情報等が社会を不安定化・混乱させるリスクが高まっており、ディープフェイクを検知する技術やディープフェイクが流通しない仕組みの開発・普及等の奨励が望ましいと取りまとめられている。

また、令和 6 年能登半島地震に関して、事実に基づかない不確実な情報や、明らかに偽と思われる情報、真偽を直ちに判別することが困難な情報等が SNS 上で存在する等の課題が指摘されており、2024 年 1 月 25 日に公表された被災者の生活と生業支援のためのパッケージにおいて、「切れ目のない被災者支援」の一つとして「被災地におけるインターネット上の偽情報・誤情報対策を行う。」と位置付けられたところである。

本事業は、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（以下、単に「インターネット上の偽・誤情報」という。）の流通リスクに対応するため、対策技術の開発・実証を実施し、社会実装を推進することを目的としている。

2 事業の概要

（1）対象事業

インターネット上の偽・誤情報への対策技術の開発・実証事業。

対策技術は以下を想定する。

- 1 SNS 等のインターネット上に流通する情報コンテンツが、生成 AI により生成された可能性が高いかどうか判別する真偽判別支援技術
 - 1-1 ファクトチェック機関やメディア等の各主体のシステムに、一つの判別対象に対して、画像又は動画といった単一の情報コンテンツを読み込み、その情報コンテンツが生成 AI により生成されたか否かを判別する機能を組み込み、真偽判別を支援する方法を検討
 - 1-2 ファクトチェック機関やメディア等の各主体のシステムに、一つの判別対象に対して、画像や動画等に加えて、インターネット上の情報等も含めた複数の情報を組み合わせ、その情報コ

ンテンツが生成 AI により生成されたか否かを判別する機能を組み込み、真偽判別を支援する方法を検討

2 情報コンテンツや発信者の信頼性等を確保する真正性保証技術・信頼性判断支援技術

- 2-1 メディア、公的機関（国・自治体）、プラットフォーム事業者等の各主体のプラットフォーム、システム又は端末に電子署名の付与・読み込み機能を組み込み、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及び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
- 2-2 メディア、公的機関（国・自治体）、プラットフォーム事業者等の各主体のプラットフォーム、システム又は端末に電子透かしの付与・読み込み機能を組み込み、各作成者が情報コンテンツの作成時に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を検討

※上記以外のインターネット上の偽・誤情報への対策技術については、事前に技術開発主体を取りまとめる管理団体（以下、単に「管理団体」という。）に相談、確認の上応募すること。

なお、技術的に対処すべきユースケース例として、以下を想定する。

＜ユースケース（例）＞

- I ファクトチェック機関が、ソーシャルメディア上の投稿の人物及び風景・街並みに関する画像・動画等のコンテンツが、生成 AI により生成されたものかについて確認を行う場面
- II ソーシャルメディアが、自社の提供するプラットフォーム上に流通する人物及び風景・街並みに関する画像・動画等のコンテンツのうち、自社のポリシーに反する投稿（著名人を騙ったコンテンツ等）を削除する用途で、生成 AI によるものかについて確認を行う場面
- III インターネットを利用する主体が、インターネット上の情報媒体を閲覧する際に、情報コンテンツの作成者や配信サイト運営者、広告主といった発信者情報や情報コンテンツの信頼性の確認を行う場面

また、今回の公募・採択は、後述の「9 採択候補先の選定等」に記載があるように、採択候補先の選定後に、外部有識者のコメント等を踏まえて、管理団体より提案内容の調整を打診する場合がある。

（2）実施概要

- ・ 総務省が契約する管理団体（ポストン コンサルティング グループ合同会社）と技術開発主体との間で契約を締結し、管理団体は、技術開発主体に対する支援及び進捗管理を行う。



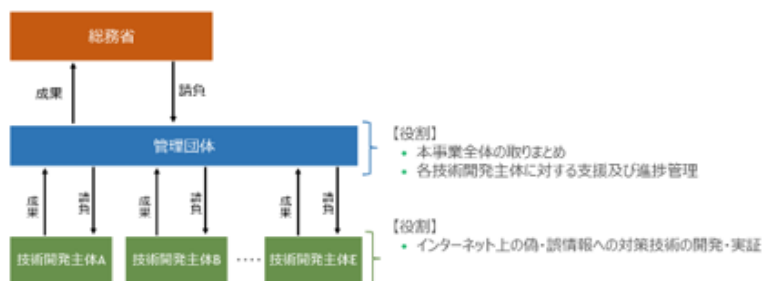
実施要領 (2/6)

下記リンク先で公開

https://www.soumu.go.jp/main_content/000944632.pdf

- 事業費規模の目安は最大 1.5 億円程度とする。
- 活用する技術の種類や費用対効果等も踏まえて提案内容の評価を実施する。
- また、技術開発主体が提出する支出計画書の妥当性等も踏まえて、総務省と管理団体との間で協議の上、支弁する金額を決定する。

＜参考＞開発・実証事業の全体像イメージ



(3) スケジュール

令和 6 年 5 月 20 日 (月) まで 企画提案書提出
 5 月下旬～6 月上旬 外部有識者による評価等
 6 月上旬 採択案件 (技術開発主体) の発表、管理団体との契約締結、事業開始

令和 7 年 2 月頃 成果報告書案の提出
 3 月頃 成果報告会の開催

※採択候補先の選定状況等によって多少前後する場合がある。

3 応募要件

(1) 提案主体

インターネット上の偽・誤情報への対策技術の開発や社会実装に取り組む事業者、研究機関等。
 上記の者で構成するコンソーシアムを組成する場合には、事業の取りまとめを行う代表機関を定め、当該代表機関は、本実施要領に定める一義的な責任を負うものとする。

(2) 対象経費

原則として、電子計算機等の物品の購入費用は対象外とするが、以下に該当する物品のリース経費、消耗品やリースで調達できない物品の購入経費等は対象経費として認める。

実証終了後における購入物品の取扱いについては、管理団体と協議の上、技術開発主体において適切に管理・活用すること。

当該契約に係る事業遂行後、対応する成果物が全て含まれた納品物 (成果報告書及び証憑書類等) の提出を求め、期限までに実施計画書どおり事業が実施されていること等、納品物の検収を経た上で、事業に要した経費を支出する。

その他詳細については、採択後に管理団体から案内する「経理処理マニュアル」等に従うこと。

支援対象経費	備考
電子計算機等実証に必要な物品のリース経費 (機器のサブスクリプション型サービスを含みます)	実証期間内に発生した経費のみが対象
取得単価が税込 10 万円未満 又は 使用可能期間が 1 年未満の物品の購入経費	「使用可能期間が 1 年未満」とは、一般的に消耗性のものと認識され、かつ、平均的な使用状況等からみて、その使用可能期間が 1 年未満であると認められるものをいう。
リース等で調達できないネットワーク/ソリューション機器の購入経費	リース等で調達できない理由及び継続利用の計画を記した理由書 (様式任意) の提出が必須
役務費	実証環境の構築やアプリケーション開発等の実証、国際標準化活動に係る人件費 等
その他	実証に係る通信運搬費・光熱費・旅費等

支援対象外経費

- 電子計算機等の物品の購入経費 (「対象経費」に該当するものを除く)
- 実証目的の遂行に必要と認められない経費及び目的遂行に必要であっても一般的に合理的と認められる範囲を超える経費 等

4 実施事項

技術開発主体は、以下の (1) から (5) までの事項を実施すること。

(1) インターネット上の偽・誤情報への対策技術の開発



実施要領 (3/6)

下記リンク先で公開

https://www.soumu.go.jp/main_content/000944632.pdf

真偽判別支援技術・真正性保証技術・信頼性判断支援技術等の、インターネット上の偽・誤情報への先進的な対策技術について、社会実装に向けた有効性を増進させるために、開発・実証を行うこと。
企画提案書において、9（3）に示す評価の観点を中心に踏まえ、具体的な実施内容を提案すること。

(2) インターネット上の偽・誤情報への技術的対策の有効性等に関する検証

真偽判別支援技術・真正性保証技術・信頼性判断支援技術等の、インターネット上の偽・誤情報への技術的対策を図る先進的な取組について、その有効性を定量的に検証すると共に、社会実装に向けた課題の解消に関する調査・検討等を行うこと。
企画提案書において、9（3）に示す評価の観点を中心に踏まえ、具体的な実施内容を提案すること。

(3) 成果報告書の作成

上記（1）及び（2）の実施内容や成果等について、管理団体が指示する報告様式に沿って、特段の専門知識を有することなく容易に理解できる表現で文書化し、報告書を作成すること。詳細については「11 納入成果物」を参照するほか、採択後に管理団体が指示する事項に従うこと。

(4) 社会実装実施計画書の提出

本事業終了後の社会実装に向けた計画として、社会実装実施計画を策定すること。事業化に向けた実効的な取組計画、事業化活動、体制、資金、及び社会実装する際の製品、サービス又はシステム要件を記述すること。本社会実装実施計画書に基づく取組状況や実績については、本事業が終了した翌年度以降も、主管課が必要に応じてフォローアップを行い、随時報告を求めることがある。

(5) 普及啓発活動への協力

開発・実証成果について、メディア対応やイベント開催、学会参加等を通じて、積極的に普及啓発活動に取り組むこと。また、開発・実証期間の終了後も含め、総務省が実施する開発・実証成果の普及啓発活動に当たって、開発・実証内容に関する資料提供等の協力を行うこと。

5 実施体制

本事業の実施に当たっては、実証参画主体との協力関係を含め、4に定める実施事項を確実に効果的に遂行できる体制を構築すること。また、必要に応じて、真偽判別支援技術・真正性保証技術・信頼性判断

支援技術等のインターネット上の偽・誤情報への先進的な対策技術の開発に取り組む専門家等の協力を得られる体制を構築すること。

また、実施体制内部の契約関係や協力関係、役割分担を確認できる実施体制図を企画提案書に記載すること。また、実証参画主体との協力関係等を示す資料を提示可能である場合には企画提案書に添付すること。

なお、技術開発主体の構成員は対外的に公表することを前提とする。

6 事業実施スケジュール

開発・実証内容の特性、対象技術の開発期間、検証項目等を踏まえて、効果的に事業を実施するために必要な期間を確保すること。

企画提案書において、各工程（準備、技術開発、各種検証、報告書作成等）の実施内容の詳細及びスケジュールを記載すること。

7 情報セキュリティ対策

外部委託先を含め、必要な情報セキュリティ対策を講ずること。

また、本事業で使用する設備・機器やシステム等については、「IT 調達に係る国の物品等又は役務の調達方針及び調達手続に関する申合せ」（令和5年4月一部改正）等に留意し、サプライチェーンリスク対応を含む十分なサイバーセキュリティ対策を講ずること。

特に、クラウドサービスの利用等、外部のネットワークへの接続やデータ伝送を伴う場合、個人情報の管理を含め、情報の流通経路全般にわたるセキュリティが適切に確保されるよう、情報の流通経路全般を見渡した形で、必要なセキュリティ対策等を実施すること。

8 提案手続

(1) 提出書類

□ 様式 1：企画提案書

□ 様式 2：実施体制説明書

□ 様式 3：事業スケジュール

□ 様式 4：資金計画書

□ 様式 5：企画提案書概要

※ 上記のほか、提案内容を補足する資料があれば、A4判（様式自由）10ページ程度で添付すること。

※ 提出された書類の返却はしない。また、採択された案件の企画提案書の概要（主に様式5）について、総務省ホームページ等で公開する場合がある。



実施要領 (4/6)

下記リンク先で公開

https://www.soumu.go.jp/main_content/000944632.pdf

(2) 提出期限

令和 6 年 5 月 20 日 (月) 午後 5 時まで (必着)

(3) 提出先

本事業の管理団体のボストン コンサルティング グループ合同会社に対して、(1)の提出書類(電子媒体)を電子メールで提出すること(郵送、持ち込みは不可)。

連絡先は以下のとおり。

TOKGigajouhoutsaku@bcg.com

※ メール の 件 名 は「インターネット上の偽・誤情報対策技術の開発・実証事業(企業・団体名)」とすること。

※ 提出書類のサイズが 10MB を超える場合には、事前に上記連絡先に相談し、提出方法について指示を受けること。

9 採択候補先の選定等

(1) 選定方法

外部有識者で構成する評価委員会において審査を行った後、その結果に基づき総務省と協議の上、管理団体が採択候補先を選定する。

評価は書面審査及び必要に応じてヒアリングを実施することにより行う。ヒアリングの実施が必要な場合又は追加の資料提出を求める場合等には、管理団体から連絡する。

(2) 提案内容の採択と修正

今回の公募・採択は、上記の採択候補先の選定後に、外部有識者のコメント等を踏まえて、管理団体から採択候補先へ提案内容の修正等を打診し(例:「アイデアは良いが、△△は直して欲しい」、「規模を縮小または拡大して欲しい」等)、内容・費用についての交渉を進め、最終的に事業内容に合意することをもって最終的な委託契約が成立し、提案が採択されるものとする。その過程において、調整未了により、提案が採択されず、委託契約に至らない可能性があることも留意すること。

2(1)で定めた 4 つの対策技術において、偏りが出ないように総務省・管理団体に配慮しながら採択先の決定を行う場合がある。

なお、管理団体から提案内容の打診や採択の通知等を行う時期については案件によって前後する。最終的な採択結果は、全ての契約締結を終えた後に纏めて公開・通知する。不採択の理由については公開しないので留意すること。

(3) 評価の観点

採択候補先の選定に当たっては、以下の観点から総合的に評価を行う。これらの観点を十分に踏まえて、できる限り具体的かつ網羅的に企画提案書を作成すること。

なお、今後、評価の観点に変更が生じた場合には、総務省ホームページにおいて公表する。

本事業の目的を踏まえ、特に重視する要素として以下が挙げられる。

- ✓ 対策技術の開発だけでなく、社会実装に向けた進展が見込まれるか
- ✓ 社会実装を後押しするような具体的な計画が立てられているか
例)
 - インターネット上の偽・誤情報対策の実行主体となるファクトチェック団体等に対して、技術開発後の一定の期間、無償もしくは安価な価格で開発した技術が提供され、取組を加速させられる。
 - インターネット上の偽・誤情報対策の実行主体となるファクトチェック団体等が、国民がインターネット上の情報の真偽判別を行う際の支援を実施する場合も含め、公共の利益に資する活動を行なう場合に、開発した技術が無償もしくは安価な価格で提供し、取組を加速させられる。

必須項目として、実施要領で定めた全ての項目に対して具体的な提案が示されているか。

- 1 実施事項で定めた内容の網羅性
- 2 実施体制の事業遂行能力
- 3 実施スケジュールの妥当性
- 4 情報セキュリティ対策の適切さ

加算項目として、各項目の提案内容に具体性や独創性があるか。

- 5 技術の期待成果の大きさと実現可能性
 - 5.1 期待成果の明確さ・技術の新しさ
 - ✓ 本開発・実証による期待成果と実現可能性が明確か
 - ✓ 従前の研究開発等と明確に区分ができるか
 - 5.2 技術の横展開の可能性
 - ✓ 本開発・実証の成果として得られる技術等が、提案主体だけでなく、幅広い主体に利用可能なものか
 - 5.3 技術が対処可能なユースケース/偽・誤情報のカバレッジと期待効果の大きさ
 - ✓ 本開発・実証の対象のユースケース/偽・誤情報に対処可能か
 - ✓ 他のユースケース/偽・誤情報も汎くカバーできるか
- 6 社会実装への期待効果の大きさと実現可能性



実施要領 (5/6)

下記リンク先で公開

https://www.soumu.go.jp/main_content/000944632.pdf

6.1 社会実装の意義・期待効果の大きさ

- ✓ 社会実装により課題解決が可能な提案内容になっているか

6.2 社会実装に向けた体制

- ✓ 技術の実用化や利用の主体が実証の体制に含まれるか
- ✓ 上記による本事業終了後の円滑な社会実装が見込まれるか

6.3 社会実装の拡張可能性

- ✓ 対象技術が、幅広い主体において、アクセス可能かつユーザビリティの観点でスムーズに導入可能か
- ✓ 対象技術が、特に公的な機関において、費用面で導入しやすいか
- ✓ 対象技術が、国際標準化を見据え、具体的に実行できるか

7 費用対効果

- ✓ 費用対効果を高めるための検討や工夫がされているか
- ✓ 実装・横展開も見据え、費用対効果が検討されているか

10 採択後及び開発・実証期間中の流れ

(1) 事業説明会

採択が決定した技術開発主体は、管理団体が開催する事業説明会に必ず出席すること（開催日時及び方法は別途指示する）。

(2) 実施計画書の作成

技術開発主体は、採択決定後、事業説明会等において管理団体が示す作成要領に従って、成果（アウトカム）目標、開発・実証内容、経費、スケジュール、再委託内容等、企画提案書の内容についてより詳細に記載した実施計画書を作成し、採択決定後 2 週間以内に管理団体に提出すること。

実施計画書の内容は、管理団体によるレビュー及び総務省の承認を経て確定するものとする。また、実施計画に変更が生じた場合は、速やかに変更箇所を反映した実施計画書を改めて管理団体に提出し、再度総務省の承認を得ること。

(3) 実証期間中の進捗管理

- ・ 技術開発主体は、管理団体が別途指定する成果物の納入期日までの間、管理団体の指示に従って、進捗報告書及び課題管理表を作成し、定期的に（少なくとも週 1 回程度）報告すること。報告の頻度については、進捗状況等を踏まえて見直す場合がある。
 - ◇ 進捗報告書の記載内容例：当月の作業内容、遅延状況、経費支出等
 - ◇ 課題管理表の記載内容例：課題内容、対応者、対応方針、対応結果等

- ・ 報告内容や課題への対応状況を踏まえて、管理団体が会議（原則オンライン）の開催を求めた場合、技術開発主体は当該会議に出席し、管理団体の指示に従って状況説明等を行うこと。
- ・ 技術開発主体のプロジェクトリーダーは、管理団体や総務省から進捗状況や実証内容に関する確認等があった場合には、迅速に技術開発主体内で確認の上で報告すること。
- ・ 管理団体が技術開発主体に対して実証の効果を高める助言等を行った場合には、当該助言等に従って技術開発主体は適切に対応すること。

(4) 成果報告

技術開発主体は、管理団体が事業全体の成果を取りまとめる際に必要となる情報提供等について協力すること。

技術開発主体は、管理団体の指示に従って成果報告会に参加するとともに、資料作成等の事前準備について協力すること。成果報告会では、技術開発主体から、KPI 達成状況、実装に向けた見通し、今後の課題等について説明すること。

11 納入成果物

技術開発主体は、実証の成果等について、管理団体が指示する報告様式に沿って、以下の（1）及び（2）の資料を作成し、管理団体が別途指定する納入期日までに技術開発主体内の了解を得て取りまとめること。

成果報告書の添付資料及び個人情報等を除き、原則として公開することを前提として作成すること。

(1) 成果報告書

実証の実施内容及びその成果等について、特段の専門知識を有することなく容易に理解できる表現で文書化すること。Microsoft Word/Excel/PowerPoint を使用して、A4・30～100 ページ程度（添付資料を含まない）で作成すること。

(2) 成果報告書 概要版

主たる実証の成果及び今後の課題等について、Microsoft PowerPoint を使用して、A4・1 ページで作成すること。

12 契約手続

(1) 本開発・実証における契約に係る基本的条件

技術開発主体の代表機関は、採択決定後、管理団体の指示に従い、本実施要領の内容に即して管理団体が作成する仕様書に基づいて契約を締結し、当該契約に係る一義的な責任を負うものとする。



実施要領 (6/6)

下記リンク先で公開

https://www.soumu.go.jp/main_content/000944632.pdf

仕様書は、原則として全技術開発主体で共通のものとし、実施内容の詳細は実施計画書で定めることとする。また、契約条件等について変更を求めることは認められない。

技術開発主体の代表機関と管理団体との契約は、総務省と管理団体との請負業務の再委託に当たるため、採択決定後、管理団体から総務省に対して再委託の申請を実施する。契約手続は当該申請について総務省が承認した後、速やかに進めるものとする。

(2) 再委託について

技術開発主体の代表機関は、技術開発主体の構成員に限らず、実証に関する業務の一部を他の企業・団体等へ再委託する場合、全ての再委託先について、委託契約等を締結する前に、管理団体に再委託等承認申請を行い、承認を得る必要がある。

管理団体及び総務省によって再委託等が承認される前に委託契約を締結した場合、当該委託契約に係る費用は実証の対象経費として一切認められないため、注意すること。

13 知的財産権等

- 1 請負者（管理団体、技術開発主体及び技術開発主体からの再委託先をいう。以下同じ。）は、本契約に関して総務省が開示した情報（公知の情報を除く。以下同じ。）及び契約履行過程で生じた納入成果物に関する情報を本契約の目的以外に使用又は第三者に開示若しくは漏洩してはならないものとし、そのために必要な措置を講じること。
なお、当該情報を本契約以外の目的に使用又は第三者に開示する必要がある場合は、事前に主管課に承認を得ること。
- 2 本事業履行過程で生じた納入成果物に関し、著作権法（昭和 45 年法律第 48 号）第 27 条及び 28 条に定める権利を含む全ての著作権及びノウハウ（営業秘密）は総務省に帰属し、総務省が独占的に使用するものとする。ただし、請負者は、本契約履行過程で生じた納入成果物に関し、著作権又はノウハウ（営業秘密）を自ら使用又は第三者をして使用させる場合は、総務省と別途協議するものとする。なお、請負者は総務省に対し、一切の著作人格権を行使しないこととし、また、第三者をして行使させないものとする。
- 3 請負者は、納入成果物に第三者が権利を有する著作物（以下「既存著作物」という。）が含まれている場合は、総務省が特に指示した場合を除き、当該著作物の使用に必要な費用の負担及び使用承諾契約に係る一切の手続を行うこと。この場合、請負者は当該契約等の内容について事前に主管課の承認を得ることとし、総務省は既存著作物について当該許諾条件の範囲内で使用するものとする。
- 4 請負者は、本仕様書に基づく作業に関し、第三者との間に知的財産権に係る権利侵害の紛争等が生じた場合は、当該紛争等の原因が専ら総務省の責めに帰す場合を除き、請負者の責任及び負担

において一切を処理すること。この場合、総務省に係る紛争等の事実を知ったときは、請負者に通知し、必要な範囲で訴訟上の防御を請負者に委ねる等の協力措置を講じるものとする。

- 5 本契約の履行に当たり、第三者の産業財産権又はノウハウ（営業秘密）を実施又は使用するとき、請負者はその実施又は使用に対する一切の責任を負うものとする。
- 6 本実証事業の実施中に知的財産権が発生した場合は、産業技術力強化法（平成 12 年法律第 44 号）に基づき、以下の一定の条件の下、受託者側に帰属させることを可能とする。

【条件（遵守項目）】

- 本実証事業に係る成果（実証事業の実施により新たに発見ないし生み出されたもの全てをいい、知的財産権に関するもの、ノウハウに関するもの等全てを含む。）が得られた場合には、遅滞なく、総務省にその旨を報告すること。
- 総務省が公共の利益のために特に必要があるとしてその理由を明らかにして求める場合には、無償で当該知的財産権を利用する権利を国に許諾すること。
- 当該知的財産権を相当期間活用していないと認められ、かつ、当該知的財産権を相当期間活用していないことについて正当な理由が認められない場合において、総務省が当該知的財産権の活用を促進するため特に必要があるとしてその理由を明らかにして求めるときは、当該知的財産権を利用する権利を第三者に許諾すること。
- 第三者に当該知的財産権の移転又は当該知的財産権についての専用実施権若しくは専用利用権の設定その他日本国内において排他的に実施する権利の設定若しくは移転の承諾をするときは、一部の場合を除き、あらかじめ総務省の承認を受けること。
- 上記のほか、必要と認められる事項がある場合には、契約書等において別途定める。

14 その他

本事業の実施については、本実施要領のほか、今後新たに取り決めるべき事項が生じた場合に総務省が定める事項によるものとする。

総務省が新たに定める事項については、総務省ホームページで公開するものとする。

(https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/taisakujiutsu.html)

15 本事業に関する問合せ先

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

E-mail: TOKGigojouhoutaisaku@bcg.com



評価基準 (R5補正使用資料)



評価基準 - 実施要領より抜粋

採択候補先の選定に当たっては、以下の観点から総合的に評価を行う。これらの観点を十分に踏まえて、できる限り具体的かつ網羅的に企画提案書を作成すること。

なお、今後、評価の観点に変更が生じた場合には、総務省ホームページにおいて公表する。

本事業の目的を踏まえ、特に重視する要素として以下が挙げられる。

- ✓ 対策技術の開発だけでなく、社会実装に向けた進展が見込まれるか
- ✓ 社会実装を後押しするような具体的な計画が立てられているか 例)
- インターネット上の偽・誤情報対策の実行主体となるファクトチェック団体等に対して、技術開発後の一定の期間、無償もしくは安価な価格で開発した技術が提供され、取組を加速させられる。
- インターネット上の偽・誤情報対策の実行主体となるファクトチェック団体等が、国民がインターネット上の情報の真偽判別を行う際の支援を実施する場合も含め、公共の利益に資する活動を行なう場合に、開発した技術が無償もしくは安価な価格で提供し、取組を加速させられる。

評価委員

No.	評価委員	所属
1	山口 真一	国際大学グローバル・コミュニケーション・センター 准教授
2	笹原 和俊	東京科学大学 環境・社会理工学院 イノベーション科学系 教授
3	川村 正樹	山口大学 大学院創成科学研究科 教授

必須項目として、実施要領で定めた全ての項目に対して具体的な提案が示されているか。

- 1 実施事項で定めた内容の網羅性
- 2 実施体制の事業遂行能力
- 3 実施スケジュールの妥当性
- 4 情報セキュリティ対策の適切さ

加点項目として、各項目の提案内容に具体性や独創性があるか。

5 技術の期待成果の大きさと実現可能性

5.1 期待成果の明確さ・技術の新しさ

- ✓ 本開発・実証による期待成果と実現可能性が明確か
- ✓ 従前の研究開発等と明確に区分ができるか

5.2 技術の横展開の可能性

- ✓ 本開発・実証の成果として得られる技術等が、提案主体だけでなく、幅広い主体に利用可能なものか

5.3 技術が対処可能なユースケース/偽・誤情報のカバレッジと期待効果の大きさ

- ✓ 本開発・実証の対象のユースケース/偽・誤情報に対処可能か
- ✓ 他のユースケース/偽・誤情報も汎くカバーできるか

6 社会実装への期待効果の大きさと実現可能性

6.1 社会実装の意義・期待効果の大きさ

- ✓ 社会実装により課題解決が可能な提案内容になっているか

6.2 社会実装に向けた体制

- ✓ 技術の実用化や利用の主体が実証の体制に含まれるか
- ✓ 上記による本事業終了後の円滑な社会実装が見込まれるか

6.3 社会実装の拡張可能性

- ✓ 対象技術が、幅広い主体において、アクセス可能かつユーザビリティの観点でスムーズに導入可能か
- ✓ 対象技術が、特に公的な機関において、費用面で導入しやすいか
- ✓ 対象技術が、国際標準化を見据え、具体的に実行できるか

7 費用対効果

- ✓ 費用対効果を高めるための検討や工夫がされているか
- ✓ 実装・横展開も見据え、費用対効果が検討されているか



応募書類案・記載例 (R5補正 使用資料)



インターネット上の偽・誤情報対策技術の開発・実証事業 企画提案書

様式1: 企画提案書

提出日：令和 6 年 月 日

提案者	技術開発主体名	技術開発主体（代表機関）となる事業者・研究機関等の名称を記載。		
	法人番号			
	代表者名	技術開発主体（代表機関）となる事業者・研究機関等の代表者の氏名・役職を記載。 （事業者代表など）		
	技術開発主体 の属性 ※技術開発主体が民間団 体の 場合のみ記載	下記のいずれかに該当する場合はチェックを入れること。（□ クリックでチェックが済みます） ☐ 課税事業者（納税義務者）に該当する民間事業者 ☐ 免税事業者に該当する民間事業者		
	プロジェクトリーダー （所属・役職・ 氏名）	電話番号		
		E-mail アドレス		
技術開発主体 構成員	技術開発主体の構成員として参画する団体を全て記載。			
実証の概要	実証件名			
	対象技術	真偽判別支援技術／真正性保証技術・信頼性判断支援技術の別 及び その具体的な技術の名称等を記載。 【例：真正性保証技術 e シール】		
	実証概要・ イメージ図	<実証概要> 簡潔に記載すること。 インターネット上の偽・誤情報の技術開発・社会実装に向けて抱える課題、提案事業に期待される 効果、実証における目標や中長期的な成果（アウトカム）目標を必ず含めること。 提案事業で活用する対策技術によってどのように偽・誤情報への技術的対策が実現されるのか、 分かりやすく（論理的・定量的な指標などを用いて）記載すること。 <事業全体のイメージ図> 事業の全体概要を説明するイメージ図を添付すること。 ※ 事業概要イメージ図（様式 5）を添付してもよい。		
	費用	〇〇千円（ <u>税込み</u> ） 本事業における対象経費を <u>税込み</u> で記載すること。		



(2)②イ.公募の準備

様式1: 企画提案書

事業の詳細	
1. 事業の目的 2. 実施計画 (1) 対象技術 ※ 事業の対象技術の概要を記載すること (2) 対象技術による偽・誤情報への技術的対策の課題の状況、今後目指す姿 (3) 実用化を目指す技術的対策の概要 ※ どのような対策技術を使って、どのような機能・サービスを実現するのかなどについて具体的に記載すること ※ サービス等の対象となる偽・誤情報の種類・形態や生じ得る被害の想定等についても記載すること 3. 「評価の観点」に関する事項 ※ 実施要領 9 (2) を十分に踏まえて記載すること 1 技術の進化・可用性 1.1 期待成果の明確さ <記載ポイント> ✓ 本開発・実証による期待成果が明確か ✓ 従前の研究開発等と明確に区別ができるか 1.2 技術の横展開の可能性 <記載ポイント> ✓ 本開発・実証の成果として得られる技術等が、提案主体だけでなく、幅広い主体に利用可能なものか 1.3 技術が対処可能なユースケース/偽・誤情報のカバレッジ <記載ポイント> ✓ 本開発・実証の対象のユースケースに対処可能か ✓ 他のユースケース(偽・誤情報)も汎くカバーできるか 2 社会実装に向けた進展 2.1 社会実装に向けた課題解決 <記載ポイント> ✓ 社会実装に向けた課題に応える提案内容になっているか 2.2 社会実装に向けた体制	<記載ポイント> ✓ 技術の実用化や利用の主体が実証の体制に含まれるか ✓ 上記による本事業終了後の円滑な社会実装が見込まれるか 2.3 社会実装に向けた拡がり <記載ポイント> ✓ 対象技術が、幅広い主体が利用可能なものか ✓ 対象技術が、国際標準化を見据え、具体的に実行できるか ✓ 実証成果が、幅広い主体の利用の後押しになるか 3 費用対効果 <記載ポイント> ✓ 費用対効果を高めるための検討や工夫がされているか ✓ 実装・横展開も見据え、費用対効果が検討されているか 5. その他 ※ 事業内容について補足すべき点があれば記載すること

注) 「事業の詳細」は 8 ページ程度に収めること。

注) 必要に応じて、補足資料 (10 ページ以内) を添付すること。



様式2: 実施体制説明書

[様式 2]

実施体制説明書

1. 実施体制

図等を用いて実施体制を分かりやすく記入すること。

提案者のみならず、実証事業の実施に関わる者について全て本様式に役割、責任を明記すること。

(例)

△△大学〇〇系研究科◇◇研究室

本事業全般の管理・統括業務（実施主体）、技術開発、技術提供

□□株式会社

社会実装に係る実証用データの提供

××株式会社

技術の実証用の実用に係るシステム開発

2. 各主体の役割

NO	名称	役割及び責任 上記組織図に対応した主体別に役割を明確に記入すること
1	△△大学〇〇系 研究科◇◇研究室	・事業計画の立案、報告書の作成をはじめとする事業全般の管理・統括業務 ・〇〇技術の開発・提供
2	□□株式会社	・協議会への参加 ・社会実装に係る実証用データの提供
3	××株式会社	・協議会への参加 ・技術の実証用の実用に係るシステム開発

3. プロジェクトリーダー（実施責任者）

氏名（ふりがな）	勤務先及び職歴概要

4. 連絡担当者

No	名称	連絡先 ※ 所属、役職、氏名、住所、電話番号、メールアドレスを記入。 ※ プロジェクトリーダーと同一でも良い。複数名記載しても良い。
1	△△大学〇〇系 研究科◇◇研究室	〇〇（所属） （役職） 総務 太郎（そうむ たろう） 〒000-0000 〇〇県〇〇市〇〇1-1-1 電話：00-0000-0000 メール：xxxx@xxxxxxxxxxxx
2	□□株式会社	〇〇部〇〇課 （役職） 総務 次郎（そうむ じろう） （役職） 総務 三郎（そうむ さぶろう） 〒000-0000 〇〇県〇〇市〇〇1-1-1 電話：00-0000-0000 メール：xxxx@xxxxxxxxxxxx, xxxx@xxxxxxxxxxxx
3	××株式会社	〇〇部門〇〇担当 （役職） 総務 花子（そうむ はなこ） 〒000-0000 〇〇県〇〇市〇〇1-1-1 電話：00-0000-0000 メール：xxxx@xxxxxxxxxxxx

注）枚数制限なし。

様式3:
事業スケジュール

事業スケジュール

単位：百万円

事業内容	令和6年 3月	4月	5月	6月	7月	12月	令和7年 1月	2月	3月	事業費
実証事業手続等 スケジュール(案)										
1.再委託申請の完了・ 契約締結(見込み)				★						
2.事業期間										
3.成果報告書案提出					★					
4.成果報告書提出								★		
5.最終報告会									★	
(例)										
ア) 実証										
1.××の調査				→						*****
2.システムの設計				→	→					*****
3.システムの構築					→	→				*****
イ) システムの試験稼働								→	→	*****
ウ) 成果の検証・分析								→	→	*****
					★		★		★	*****
合計										*****

注) 技術開発、システム構築、実証用データ整備、実証、報告書作成などの作業工程をできる限り細分化し、具体的な作業内容も含めて記載すること。具体的な作業内容については、必ずしも上記のスケジュール表内に記載する必要はない。

注) 事業費は消費税込みの金額を記入すること。

注) 事業スケジュールについて、事前に留意する点がある場合は明記すること。

注) 枚数制限なし。



(2)②イ.公募の準備

【様式4】

様式4: 資金計画書

【実証件名を記載】

適宜、行の追加・削除、行・列の高さ
や幅の変更、セル結合して構わない

資金計画書

赤字は全て例示であり、記載の際は赤字を
削除し、黒字にして記載

※金額はすべて消費税込み（10%）
設備・備品費等のリース料金については一ヶ月当たりのリース料金を単価として表記

大項目	項目	費用区分				
1 直接経費						#REF!
(1) 実証環境の構築						20,600,000
① 技術開発関係	研究開発作業費	人件費	2人×100時間×	@	12,000円 =	2,400,000
	クラウドサーバー利用費	物品リース、レンタル費、サーバ利用費	1式×3ヶ月×	@	600,000円 =	1,800,000
		物品・ソフトウェア等購入費（税込10万円未満/消耗品）	1式×	@	1,000,000円 =	2,000,000
② 実証用データ整備関係	システム費	物品リース、レンタル費、サーバ利用費	5台×90日×	@	10,000円 =	4,500,000
	開発費	不動産賃貸費	90日×	@	100,000円 =	9,000,000
③ その他	印刷	物品リース、レンタル費、サーバ利用費			600,000円 =	600,000
	旅費	旅費・交通費・宿泊費			300,000円 =	300,000
(2) 検証作業などの実施						1,100,000
① …に関する検証	検討作業費	人件費	2人×50時間×	@	11,000円 =	1,100,000
② …に関する調査・検討	検討作業費	人件費	2人×50時間×	@	11,000円 =	1,100,000
(3) 成果報告書の作成						2,560,000
①	報告書作成作業費	人件費	4人×40時間×	@	16,000円 =	2,560,000
2 間接経費（一般管理費のみ）						100,000
3 合計					（直接経費+間接経費）	税込 #REF!

【資金計画書記載の注意点】

- ・記載にあたっては、原則として費用項目ごとにその内訳（費用区分、単価、数量等）を示すこと。
- ・費用のうち「ネットワーク/リユース機器等購入費」については、リースなどで調達できないもののみ実証対象経費として認める。
なお、上記の購入に当たってはリースなどで調達できない理由及び継続利用の計画を記した理由書（様式任意）を提出し、事前に承認を得ること。
- ・赤字は全て例示であり、記載の際は赤字を削除し、黒字にして記載すること。
- ・記載にあたっては、適宜、行の追加・削除、行・列の高さや幅の変更、セル結合して構わない。
- ・間接経費は管理等に必要経費として使用する経費であり、販売費を除く一般管理費を対象とする。

**様式5:
企画提案書概要**

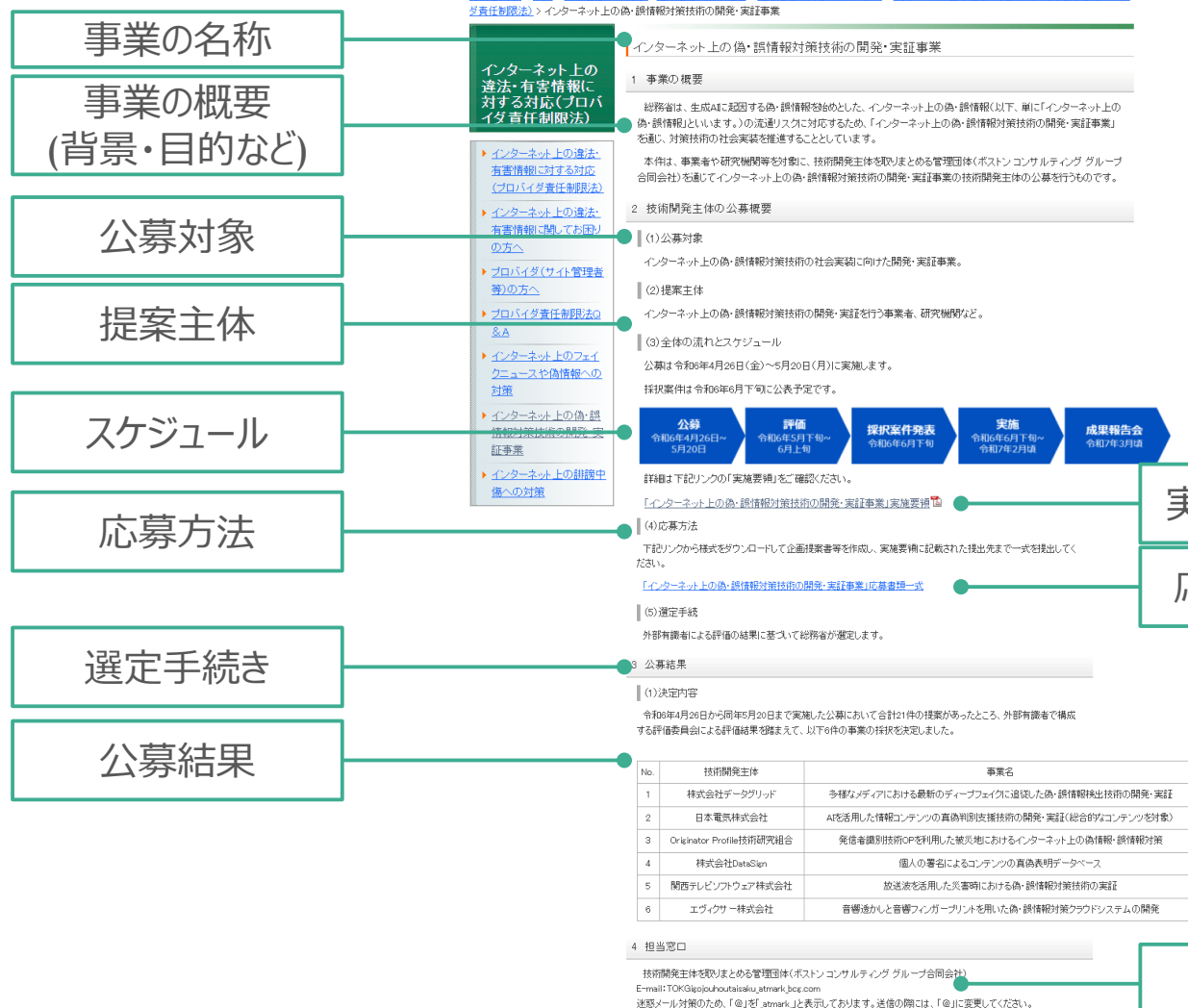
1		実証件名		様式 5
実施体制 (下線：代表機関)	A大学X研究室、B社(株)、C社(株)		対象とする偽・誤情報	➤ 対象とする偽・誤情報の種類を記載 － 情報の対象：人、物体、風景 など － 情報の形態：画像、映像、音声、テキスト など ➤ 中長期的な成果（アウトカム）目標を記載
対策技術	➤ 実証で活用する対策技術について記載		目標	➤ 実証における目標を記載 ➤ 中長期的な成果（アウトカム）目標を記載
実証概要	➤ ○○(対象とする偽・誤情報)への技術的対策には、○○○○という課題が存在。 ➤ ○○(対策技術)を活用し、○○○○を実施。 ➤ ○○○○の実現を図る。 ※150～300字程度で、原則としてフォーマットに記載されている文章を用い、○○○○の箇所に提案内容を記載すること。 ※背景・課題（技術・社会実装における課題など）を記載 ※課題解決のために提案する取組みの概要 ※実証における目標・中長期的な成果(アウトカム)目標を記載 ＜技術開発・社会実装に向けた課題の概要、提案する取組み及び期待される効果などの概要＞ ＜ 事業概要を表す図・イラスト等を記載			



公募用ウェブページの構成要素

R5補正 使用ウェブページ (下記リンク先で公開)

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



実施要領(PDF)のダウンロードリンク

応募様式一式のダウンロードリンク

管理団体への問合せ先



評価委員を組成の上で、評価委員会を実施

2 技術開発主体の公募概要

(1) 公募対象

インターネット上の偽・誤情報対策技術の社会実装に向けた開発・実証事業。

(2) 提案主体

インターネット上の偽・誤情報対策技術の開発・実証を行う事業者、研究機関など。

(3) 全体の流れとスケジュール

公募は令和6年4月26日(金)～5月20日(月)に実施します。

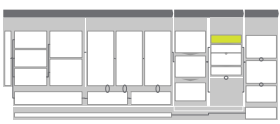
採択案件は令和6年6月下旬に公表予定です。

評価委員会の実施



詳細は下記リンクの「実施要領」をご確認ください。

[「インターネット上の偽・誤情報対策技術の開発・実証事業」実施要領](#) 



採択事業者は3月までの実施計画を作成し、計画に基づき事業を実施

実施計画書の目次

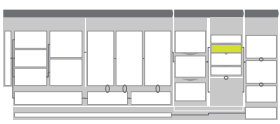
目次

1. 実証事業の概要

1-1. 実証概要のサマリ	 XX
1-2. 事業の目的	 XX
1-3. 開発する技術の詳細	 XX
1-4. 社会実装のための実証の詳細	 XX
1-5. 本実証後の展望	 XX

2. 計画の詳細

2-1. 実証の実施体制	 XX
2-2. 実証スケジュール	 XX
2-3. 支出計画書	 XX



開発・実証の効果を高め、インターネット上の偽・誤情報対策の社会実装を推進するため、
実証団体の状況に応じて、専門的な見地から助言等の支援を実施できる体制を構築

管理団体の支援体制

関係事業者を監督・管理する体制

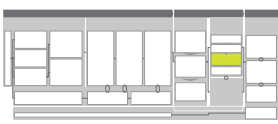
ディープフェイク対策への専門的知見に対する体制

複数の実証事業を同時並行で支援するための体制

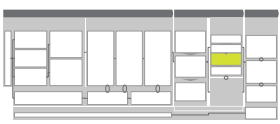


外部有識者の協力体制

ディープフェイク対策に知見のある
学識経験者の協力体制



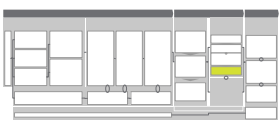
進捗報告・課題管理用の資料フォーマット (R5補正 使用資料)



進捗状況サマリーフォーマット

- 予定通り進捗
- 課題発生し、対応策検討中・立案済
- 課題発生し、一部方針転換要

No.	代表機関	事業名	進捗状況		備考
			前回	今回	
1	データグリッド	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	(各月の概況を記載)		
2	NEC	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			
3	OP	発信者識別技術OPを利用した被災地におけるインターネット上の偽情報・誤情報対策			
4	DataSign	個人の署名によるコンテンツの真偽表明データベース			
5	関西テレビソフトウェア	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			
6	エヴィクサー	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の開発			



成果報告会の開催概要

- 実施形式
 - 会場：総務省会議室
 - 日時：3/12(水) 16:00-18:00
 - ハイブリッド形式
- 参加者
 - 総務省関係者
 - 管理団体
 - 実証事業者（6社）
- アジェンダ
 - 冒頭挨拶
 - 事業概要の説明
 - 事業者からの成果報告
 - 締めの挨拶
- 事業者からの成果報告の発表内容
 - ①実証事業での取組概要
 - ②事業の成果
 - 目標・KPI達成状況
 - 開発した技術・ツールの詳細（開発した技術やツールの特長・精度など）
 - 社会実装のための実証の結果（実証結果として得られた効果や、発見した課題等）
 - ③今後に向けた課題
 - ④社会実装に向けた今後の展望



経理処理マニュアルを整備の上、経理処理を実施

目次

第1章：本実証における経理処理業務の概要	P.2
第2章：基本情報登録	P.5
第3章：支出計画書	P.8
第4章：対象経費及び提出方法	P.9
別紙・別表	P.20

Agenda

背景・目的、実施内容、スケジュール

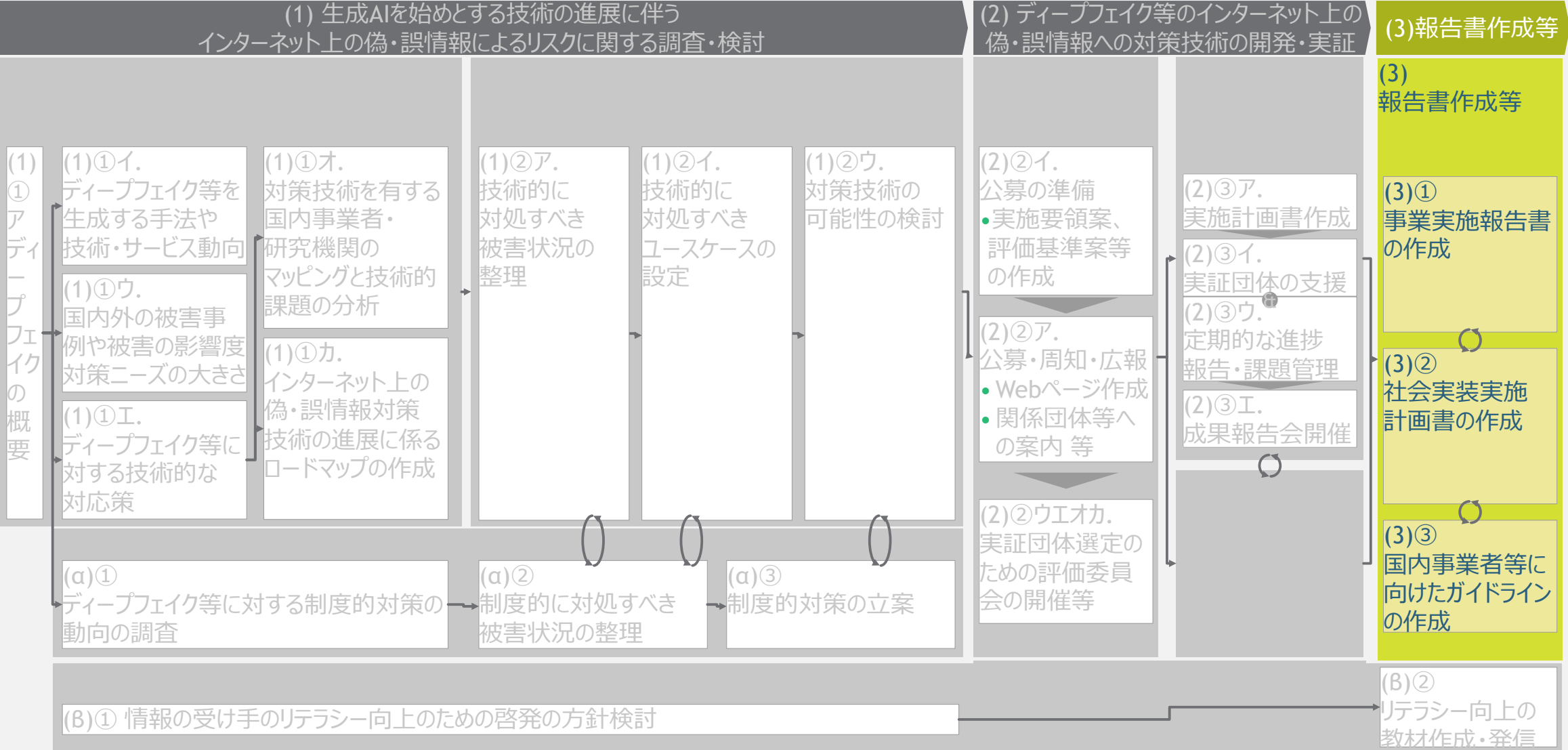
(1) 生成AIを始めとする技術の進展に伴うインターネット上の偽・誤情報によるリスクに関する調査・検討

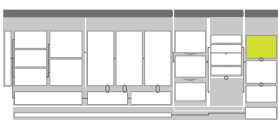
(2) ディープフェイク等のインターネット上の偽・誤情報への対策技術の開発・実証

➤ (3) 報告書作成等

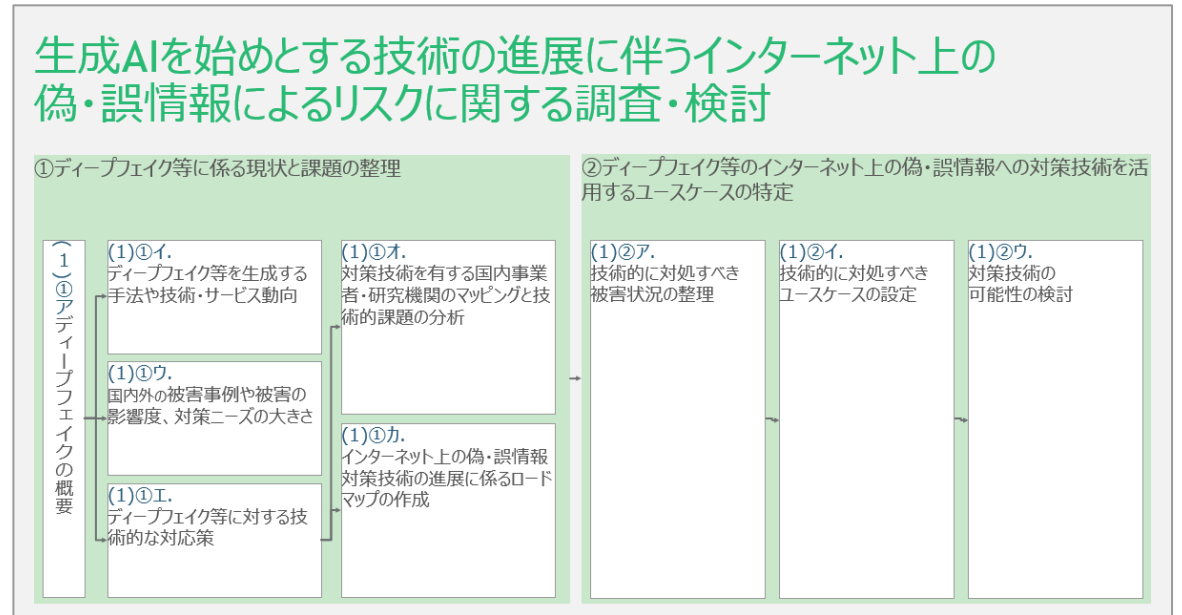
その他 検討事項

実施内容





中間報告書は以下の内容で構成





各実証団体の成果報告書を取りまとめて整理

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

2

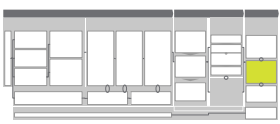
目次

1. 実証事業の概要

- 1-1. 実証概要のサマリ
- 1-2. 事業の目的
- 1-3. 開発する技術の概要
- 1-4. 社会実装のための実証の詳細
- 1-5. 実施計画の詳細

2. 実証事業の成果

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



各実証団体の報告内容を取りまとめた社会実装実施計画書を作成

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

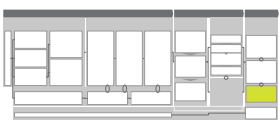
インターネット上の偽・誤情報技術の開発・実証事業 管理団体
ボストン・コンサルティング・グループ合同会社

0

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
- 1-3. 体制
- 1-4. 資金
- 1-5. 社会実装する際の製品、サービス又はシステム要件



開発・実証事業の概要・成果を取りまとめた資料を作成



インターネット上の偽・誤情報 対策技術の開発・実証事業

総務省 情報流通行政局 情報流通振興課

令和7年4月



MIC

総務省

Ministry of Internal Affairs
and Communications

インターネット上の偽・誤情報 対策技術の開発・実証事業

総務省 情報流通行政局 情報流通振興課

令和7年4月

インターネット上の偽・誤情報対策技術の開発・実証事業 概要

背景・目的



近年、多様な生成AIやディープフェイク技術により偽画像・動画を誰でも容易に作成できるようになり、偽のニュース動画・災害画像、詐欺サイト等、人の目では真偽を見分けることが困難な情報に、国民がSNS等を通じて日常的に触れ得る機会が増加

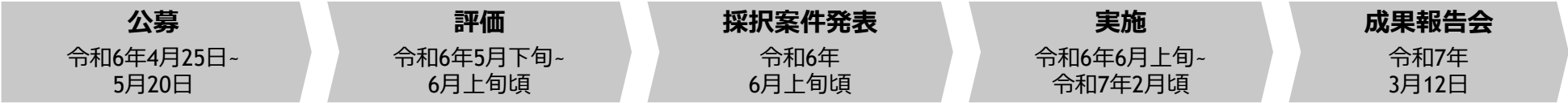
- 令和6年能登半島地震に際しても、事実に基づかない不確実な情報や、明らかに偽と思われる情報、真偽を直ちに判別することが困難な情報等がSNS上等で存在する等の課題が指摘

本事業は、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、対策技術の開発・実証を実施し、社会実装を推進することを目的とする

全体の流れとスケジュール



令和6年4月~5月に公募を実施し6月上旬に採択案件を発表、その後各事業者にて技術開発・社会実装の実証を実施いただき、令和7年3月に成果報告会を実施



採択事業者/事業名



公募にて合計21件の提案があり、外部有識者で構成する評価委員会による評価結果を踏まえて、以下6件の事業の採択を決定

No.	技術開発主体	事業名
1	株式会社データグリッド	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
2	日本電気株式会社	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 (総合的なコンテンツを対象)
3	Originator Profile 技術研究組合	発信者識別技術OPを利用した被災地におけるインターネット上の偽情報・誤情報対策
4	株式会社DataSign	個人の署名によるコンテンツの真偽表明データベース
5	関西テレビソフトウェア株式会社	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
6	エヴィクサー株式会社	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

主要な技術の開発・検証を完了し、関係者との協業と社会実装に着手

インターネット上の偽・誤情報対策技術の開発・実証事業: R5補正成果

本事業の 目的

生成AI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、対策技術の開発・実証を実施し、社会実装を推進する

- 社会的安定性の確保に向けた、情報信頼性の向上のための持続可能な技術基盤の構築

R5補正にて 見込む 到達点

対策技術が実用化できるよう開発・実証され、社会実装に着手する

a 対策技術の開発・実証

- 真偽判別、真正性保証/信頼性判断支援の各技術領域における幅広い技術開発・技術検証

b 社会実装への着手

- 画像/映像の検知、真正性保証/信頼性判断支援は、想定される利用主体等と試用等の実証実験を実施

R5補正 成果

主要な技術の開発・実証を完了し、関係者との協業と社会実装に着手

技術開発の成果



社会実装の進展



株式会社データグリッド

偽の映像・画像や音声を見分ける技術を開発し、事実確認担当者が使用可能な水準の精度に到達

日本電気株式会社

検索情報を組合せて情報内容が本物か見分ける技術を開発し、事実確認担当者の調査に要する時間を削減

Originator Profile
技術研究組合

情報の発信者や、改ざんが行われていないことを見極める技術を開発

株式会社DataSign

情報の発信者が自分の発言や文章が本物だと証明できる技術を開発

関西テレビソフトウェア
株式会社

インターネットが使えない状況で、災害時の支援作業者を判断できる技術を有線で動く形で構築

エヴィクサー株式会社

放送コンテンツに音響透かしを埋め込み、改ざん有無を確認できる技術を開発

ファクトチェック機関 (FIJ/SIA) やメディア (日経新聞) の事実確認担当者のテスト利用を通じてフィードバックを確認し、機能改善を実施

ファクトチェック機関 (JFC) の事実確認担当者のテスト利用を通じてフィードバックを確認し、レポート機能や参照情報等必要な機能や使い勝手を改善

自治体がWebサイトで公式発表した情報を市民が確認できる機能の実装・動作検証を実施

- 鳥取県の公式HP 等

利用主体となる著名人やジャーナリストからのフィードバックを取り入れながら、機能改善を実施

一般市民が判断するためのアプリの操作性をテスト利用により改善

ケーブルテレビ局等 (ゼクサTV) で透かし入り放送を流す実証を実施

個別技術の作りこみ、利用主体・実装体制の具体化により社会実装を加速

インターネット上の偽・誤情報対策技術の開発・実証事業: R5補正採択技術の今後の展望

モノ
(技術)



R5補正成果と深化の観点

個別技術の開発を完了し、
今後は各ユースケースごとの作りこみが必要

- ・ 「影響範囲」「被害の深刻度」「市場性」の観点より災害・詐欺・企業の評判を優先領域と想定
- ・ 個別の文脈に応じた社会実装には、専用のデータセットと詳細な調整が必要

現時点の生成技術に対する技術開発を完了し、
今後は技術進歩への対応が必要

- ・ 日進月歩の生成AI技術に対して、常に既存技術の向上が必要

今後の社会実装の方向性

個別技術のユースケースへの落とし込み

- ・ 社会実装に際した要求仕様の具体化
- ・ 検知精度の向上
- ・ 運用手順の定義・マニュアル化

個別技術の高度化

- ・ 生成側の進化への追加対応
- ・ 音声/テキストのモーダル拡充 等

ヒト
(利用主体・
実装体制)



初期的な利用主体の特定が完了し、
今後は新たな展開先の獲得が必要

- ・ 中長期的な利用主体の特定・合意形成が必要

実装の立上げに必要なステークホルダーの巻き込みを開始、
今後は体制・役割の明確化が必要

- ・ 対象の業界や企業・組織の仮説に基づき、特定・合意形成が必要

対象ユーザーの拡大

- ・ ユースケースごとに自治体・報道機関・地域住民など利用主体の拡大先を具体化

座組の明確化

- ・ 協業先の特定・役割分担の具体化
- ・ 利用主体を巻き込むコンソーシアムの構築 等

災害・詐欺・企業への風評など喫緊の課題となっているユースケースでの成果創出を目指す

各実証団体の主たる実証の成果及び今後の課題・展望サマリ

	主たる実証の成果 (技術開発の成果 及び 社会実装の進展)	今後の課題・展望
株式会社 データグリッド	- 偽の映像・画像や音声を見分ける技術を開発し、事実確認担当者等が使用可能な水準の精度に到達 - ファクトチェック機関 (FIJ/SIA) やメディア (日経新聞) の事実確認担当者のテスト利用を通じてフィードバックを確認し、機能改善を実施	- 技術開発では技術進展への対応、テキスト等の情報形態への対応を予定 - 社会実装では偽・誤情報対応の一通貫プロセスの支援や対象ユーザーの一般のSNSユーザーや自治体・官公庁等への拡大を企図
日本電気 株式会社	- 検索情報を組合せて情報内容が本物か見分ける技術を開発し、事実確認担当者の調査に要する時間を削減 - ファクトチェック機関 (JFC) の事実確認担当者のテスト利用を通じてフィードバックを確認し、レポート機能や参照情報等必要な機能や使い勝手を改善	- 技術開発では個別技術の高度化・SNS情報等との連携、総合的な偽・誤情報対策ソリューションの構築を企図 - 社会実装では初期顧客となるメディア (放送局、新聞社) へのアプローチ・組み先の開拓が必要
Originator Profile 技術研究組合	- 情報の発信者や、改ざんが行われていないことを見極める技術を開発 - 自治体がWebサイトで公式発表した情報を市民が確認できる機能の実装・動作検証を実施 - 鳥取県の公式HPで実装を開始、加えて石川県/金沢市やNHK等の組合員の報道機関と協業を合意、CMS提供企業とも連携を開始	- 技術開発ではWeb動向への対応と安定稼働化を検討 - 社会実装ではユースケース拡大/広告サプライチェーン企業との合意形成、及び国際標準化の推進を予定
株式会社 DataSign	- 情報の発信者が自分の発言や文章が本物だと証明できる技術を開発 - 利用主体となる著名人やジャーナリストからのフィードバックを取り入れながら、機能改善を実施	- 技術開発では各種デバイスへの対応やUI/UXの改善を予定 - 社会実装ではメディアや関係団体との連携・広報強化や各プラットフォームへの組み込み・オンラインサービスとの連携による技術標準化を予定
関西テレビ ソフトウェア 株式会社	- インターネットが使えない状況で、災害時の支援作業者を判断できる技術を有線で動く形で構築 - 一般市民が判断するためのアプリの操作性をテスト利用により改善	- 技術開発では無線技術向けの追加開発及び対象モダルの拡充 - 社会実装では利用主体との協力関係の構築が必要
エヴィクサー 株式会社	- 放送コンテンツに音響透かし等を埋め込み、改ざん有無の確認技術を開発 - ケーブルテレビ局等 (ゼクサTV) で透かし入り放送を流す実証を実施	- 技術開発では対象データの形態拡張及び特許出願を検討予定 - 社会実装では組み先及び顧客の開拓と、国際標準化を企図

主たる実証の成果と今後の課題・展望

実証番号1: 株式会社データグリッド

代表者コメント



株式会社データグリッド
代表取締役CEO

岡田 侑貴 様

“ 本テーマ・事業に 関連するコメント

生成AIの急速な進展の一方で、ディープフェイクによる偽・誤情報が拡散し、情報の信頼性が揺らいでいます。当社は独自の判定技術により、ディープフェイクを見破り、安心安全なインターネット空間に貢献します。

主たる実証の成果 (技術開発の成果 及び 社会実装の進展)



- 偽の映像・画像や音声を見分ける技術を開発し、事実確認担当者等が使用可能な水準の精度に到達
- ファクトチェック機関 (FIJ/SIA) やメディア (日経新聞) の事実確認担当者のテスト利用を通じてフィードバックを確認し、機能改善を実施

開発した技術・ツールの詳細

本実証では主として次の技術1-4を開発した。

- 技術1. 顔画像・映像ディープフェイク検知技術:
生成AIで合成された顔画像・映像を97%以上の精度で検知できる技術
- 技術2. 音声ディープフェイク検知技術:
生成AIで合成された日本語音声で98%以上の精度で検知できる技術
- 技術3. 一般画像ディープフェイク検知技術:
生成AIで合成された一般画像 (特に災害画像) を90%以上の精度で検知できる技術
- 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:
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

社会実装のための実証の結果

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。また、ファクトチェック業務の中で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し、フィードバックをいただくことで、実業務の効率化に資するディープフェイク検知アプリの機能・構成を明らかにし、本事業で開発したアプリに落とし込むことに成功した。

今後の課題・展望



- 技術開発では技術進展への対応、テキスト等の情報形態への対応を予定
- 社会実装では偽・誤情報対応の一気通貫プロセスの支援や対象ユーザーの一般のSNSユーザーや自治体・官公庁等への拡大を企図

社会実装のための実証で得られた課題

- 課題1: 生成AI技術の急速な発展に追従するディープフェイク検知AIの運用技術の確立
- 課題2: ディープフェイク以外の偽・誤情報手法に対する真偽判定技術の確立
- 課題3: メディアにおけるSNSファクトチェックのさらなる活性化による偽・誤情報の影響抑止
- 課題4: SNSユーザーによる主体的なファクトチェックによる偽・誤情報の拡散抑制

本実証後の展望

- 2025年度 SNS上の偽・誤情報対策サービスの開発・実証:
 - ディープフェイク対策を含む、実行性のある偽・誤情報対策サービスのコアコンセプトを実装・実証する。
- 2026年度 サービス普及に向けた本格展開:
 - 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。
- 2027年度 ビジネスモデルの確立:
 - 収益性を確保したビジネスモデルを導入し、持続的な事業運営フェーズに移行する。

主たる実証の成果と今後の課題・展望

実証番号2: 日本電気株式会社

代表者コメント



日本電気株式会社
クロスインダストリービジネス
ユニット インフラDX事業
部門 メディア統括部
プロフェッショナル

上田 宏一 様

“ 本テーマ・事業に 関連するコメント

NECは先進テクノロジーと今まで積み上げてきた知見を活用し、非常に難しい本問題の解決に取り組んでいました。今後も総合的な偽誤情報の分析/検知プラットフォームの実現により社会実装を加速し、社会課題の解決を目指します。

主たる実証の成果 (技術開発の成果 及び 社会実装の進展)



- 検索情報を組合せて情報内容が本物か見分ける技術を開発し、事実確認担当者の調査に要する時間を削減
- ファクトチェック機関 (JFC) の事実確認担当者のテスト利用を通じてフィードバックを確認し、レポート機能や参照情報等必要な機能や使い勝手を改善

今回の実証で予定していた機能の実装が完了したこと、本事業で開発したものがファクトチェック機関や放送局での作業効率化につながることへの確認ができた。

開発した技術・ツールの詳細

- テキスト/映像/音声から複合的に真偽を判定し、内容抽出と情報収集による分析が可能
- フェイク判定精度は、日本ファクトチェックセンターのレポートと比較して85%相当
- 社会実装を加速するため、OSS主体の構成
最新のAPI/サービスに適応可能な拡張性を持ち、導入しやすい設計

社会実装のための実証の結果 (ファクトチェック機関)

- 業務時間の短縮に寄与し、特に証拠出展情報の収集を2時間程度から5〜10分に短縮できる
- 分析レポートの作成時間は削減可能だが、具体的な時間については示しづらい。ただし、生成されたレポートの組み立て方を参考にすることで効率化が図れる

社会実装に向けた要件・課題の洗い出し (放送局)

- 業務フロー図を作成し、東京3局、大阪1局の偽・誤情報分析を行う部署にヒアリングを実施
- 信頼性の低い情報を使用する際の検証やクロスチェックに使用することで放送局の場合でもファクトチェック作業効率化ができることを確認

今後の課題・展望



- 技術開発では個別技術の高度化・SNS情報等との連携、総合的な偽・誤情報対策ソリューションの構築を企画
- 社会実装では初期顧客となるメディア (放送局、新聞社) へのアプローチ・組み先の開拓が必要

今回の実証を通して、社会実装することへの有用性の確認、課題の洗い出しが実施できたため、翌年度以降に社会実装を加速するため、技術の洗練だけでなくプラットフォーム化が必要と判断し、社会実装計画の具体化を行った。

社会実装に向けた課題

- 出展情報の取得方法のさらなる工夫
(二次から一次、クローズド情報の活用)
- 速報用途での対応
(証拠が少ないため、風景照合や過去画像照合等の機能追加)
- 求めている機能の違い
 - ファクトチェック機関、マスメディア: 真偽分析
 - その他企業: 真偽判定

上記の困難な課題の解決と社会実装を加速していくため、2024年度実証中の他案件成果物 (NEC以外) 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を構築する計画を提案する

上記についても今年度実施した、総合分析による真偽判別支援技術で培ったノウハウが主となる予定である

主たる実証の成果と今後の課題・展望

実証番号3: Originator Profile技術研究組合

代表者コメント



Originator Profile
技術研究組合 理事長
村井 純 様

“ 本テーマ・事業に
関連するコメント

「オリジネーター・プロフィール (OP)」は、Web の世界をより健全で透明性の高いものにするために日本から貢献することを目指し、発信者の検証を可能とする技術です。本実証によって、メディアと地方自治体の情報発信に関する技術開発と国際標準化活動が進展し、社会実装の道筋が見えました。今後も引き続き開発と普及に向けた取組を拡大させてまいります

主たる実証の成果 (技術開発の成果 及び 社会実装の進展)



- 情報の発信者や、改ざんが行われていないことを見極める技術を開発
- 自治体がWebサイトで公式発表した情報を市民が確認できる機能の実装・動作検証を実施
 - 鳥取県の公式HPで実装を開始、加えて石川県/金沢市やNHK等の組合員の報道機関と協業を合意、CMS提供企業とも連携を開始

開発した技術・ツールの詳細

- Originator Profile (OP) はWeb上での発信者のさまざまな情報を検証できる形で付与する技術である。デジタル署名技術を使っており、悪意ある第三者のなりすましを防ぐこともできる。OPは 情報発信ポリシーとそれを実行するガバナンスが整備され、それが遵守されている組織に発行されるものである。
- インターネットはさまざまな発信者が作ったコンテンツが混在した状態で提示されるが、OP は個々のコンテンツの範囲を特定、それぞれの発信者についての情報をデジタル署名とともに紐づけることができる。
- 従って、OPを使えば、市民が、災害時に混乱を目的としたり、インプレッション数を獲得するために発信される偽・誤情報コンテンツに接触した時に、単に当該サイトの運営者だけでなく、サイト内のページに記載されたコンテンツごとの発信者を確認し、それが責任を持った情報発信を行っている企業・団体等から発信された情報かを市民が判断できるようになる。その結果、市民自身が偽・誤情報の可能性を判定しやすくなる。

社会実装のための実証の結果

- 実証実験には、自治体として鳥取県、OP技術研究組合 (OP CIP) 組合員の報道機関等が参加した。各主体がOP技術を検証系サイトのページ等に組み込むことに成功し、一部は引き続き組み込み作業を続けている。石川県、金沢市とは検証系サイトにOPを組み込むことの合意形成ができた。またOP実装したWebページのニュースプラットフォームなどへの再配信の技術仕様の開発にも成功した。さらに、OP技術を搭載したサイトとそのページにアクセスした際に当該サイトとページ内のコンテンツの発信者が誰かであるかと、コンテンツが第三者により改竄されていないことを確認できるブラウザ機能拡張の開発に成功した。
- 本実証で開発したOP技術が世界中で誰もが利用できるするためには、その国際標準化を実現することが必要である。そこで、国際標準化団体W3Cの技術者会議であるTPACに参加し、ブレイクアウトセッションでOP技術を紹介、W3C技術者と意見交換を行なった。
- 今後の国際標準化を進めるために、OP技術がさまざまなステークホルダーの意見を取り込みながら、透明性を持って開発されてきたことが確認できる文書シリーズとしてOP RFC (Request For Comment) の日本語・英語版を作成した。OP RFCは、今後も国際標準化に関わる様々なステークホルダーを巻き込みながら引き続き充実させていく。

今後の課題・展望



- 技術開発ではWeb動向への対応と安定稼働化を検討
- 社会実装ではユースケース拡大/広告サプライチェーン企業との合意形成、及び国際標準化の推進を予定

本実証後の展望

- 本実証で検証系サイトへのOP実装に成功した報道機関、自治体の通常Webサイト (本番系) へのOP実装を目指す。次に、2024年の実証で開発に着手したOP実装コンテンツの再配信技術を、OP CIP組合員のニュースアグリゲーターやプラットフォームとの実証に取りかかり、実装に必要なルールの検討に着手する。
- さらに発信元の識別をするブラウザ機能拡張アプリの改善と主要ブラウザ対応を行い、一般市民のOP拡張機能アプリ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具体的な提案活動を開始する。
- 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OP CIPの協業にも取り組み、より多くの自治体WebサイトへのOP実装を目指していく。

主たる実証の成果と今後の課題・展望

実証番号4: 株式会社DataSign

代表者コメント



株式会社DataSign
代表取締役

太田 祐一 様

“ 本テーマ・事業に 関連するコメント

この度「Boolcheck」という分散技術を用いたサービスを試験的にリリースすることができました。今後も特定の事業者に依存しない偽・誤情報対策の普及を目指し、技術開発および社会実装を推進していきます

主たる実証の成果 (技術開発の成果 及び 社会実装の進展)



- 情報の発信者が自分の発言や文章が本物だと証明できる技術を開発
- 利用主体となる著名人やジャーナリストからのフィードバックを取り入れながら、機能改善を実施

オンライン上のコンテンツの真偽について、著名人やジャーナリストが表明を投稿するシステムであるBoolcheckを開発した。投稿者の所属等の属性情報を証明する証明書 (Verifiable Credential) を組み合わせることで、身元を証明しながらコンテンツの真偽に言及することを実現している。なお、この証明書を管理するためのアプリケーションであるDIW (デジタルアイデンティティウォレット) についても併せて開発を行なった。

生活者は、本実証で開発したブラウザエクステンションを予めインストールしておくことで、オンライン上のコンテンツを閲覧している際にその真偽に関する投稿がBoolcheckに存在するか否かリアルタイムに確認することができ、偽誤情報の見極めに役立てることが可能となる。真偽に関する投稿が存在する場合は、閲覧するコンテンツの一部を遮るようにポップアップを表示することで、生活者に注意を促すことが可能である。このブラウザエクステンションは公開済みであり、容易に利用を開始することができる状態にある。

また、Boolcheckからデータを取得するためのAPIは一般に公開していることから、各種プラットフォームと連携し広告審査等に活用する可能性を備えている。

開発システムに対する実証実験では、前澤友作氏をはじめとする著名人やメディア関係者が参加し、実際にDIWやBoolcheckを利用して真偽情報の表明・確認を行うことで本システムの効果をより高めるための意見を得た。

今後の課題・展望



- 技術開発では各種デバイスへの対応やUI/UXの改善を予定
- 社会実装ではメディアや関係団体との連携・広報強化や各プラットフォームへの組み込み・オンラインサービスとの連携による技術標準化を予定

技術的観点

- コンテンツの永続保存とプライバシー保護の両立
 - オンラインコンテンツはURLで表現されるため、投稿時にスナップショットを取得し改ざん防止の上で永続保存が必要である。しかし、センシティブな内容を含む場合を考慮し、匿名化やアクセス制御、保存期間設定などのプライバシー対策も必須である。

非技術的観点

- 一般利用者の拡大と認知度向上
 - 対応ブラウザ・デバイスの拡充、利用ガイドの整備、広報活動強化により認知度向上を図る。利用者のフィードバックを反映し、ユーザビリティ向上と社会実装の促進を行う。
- マネタイズの検討と連携強化
 - 公開するBoolcheckのAPIを活用し、広告・情報伝送事業者と連携することで新たなビジネスモデルの構築が考えられる。関連組織との連携を通じた具体的なマネタイズ戦略で、持続可能な運用基盤の確立を目指す。
- 投稿者・認証組織の信頼性向上
 - 現状、投稿者の属性を証明するのみでありその投稿自体が正しいか否かに本システムは関与していない。最終的な偽誤情報の判断は生活者に委ねるモデルであることから、その判断材料として有益な検証可能な情報の拡大を進める必要がある。具体的には投稿者が投稿を行うだけの資格を有しているかや、その審査を行う主体の信頼性の表現についてである。

主たる実証の成果と今後の課題・展望

実証番号5: 関西テレビソフトウェア株式会社

代表者コメント



関西テレビソフトウェア
株式会社
ソリューションセンター
デジタルデザイングループ
チーフエキスパート
横島 裕明 様

“ 本テーマ・事業に 関連するコメント

災害時の偽・誤情報拡散は、
信頼できる情報配信による
早急な対策が必要です。
弊社では放送波を活用し、
通信困難な環境でも切れ目の
ない安心・安全な情報を届ける
ため、継続的な実証を重ね
実用化を目指します

主たる実証の成果 (技術開発の成果 及び 社会実装の進展)



- インターネットが使えない状況で、災害時の支援作業者を判断できる技術を有線で動く形で構築
- 一般市民が判断するためのアプリの操作性をテスト利用により改善

放送波を活用し、被災地にブロックチェーンと同期された環境の構築を実現。
コンパクトなデータサイズでもノイズ耐性のある偽誤情報の検証が可能に。

証明書発行サブシステム

- 耐改ざん性のあるブロックチェーンと同期したデータ構成を非ネット環境に構築し、災害時に通信キャリアが応急復旧するまでのおよそ2週間、検証可能な環境を維持。
- ブロックチェーンに偽・誤情報に関するデータが登録されてから3秒以内のIPDC受信を実現。

証明書検証サブシステム

- リアルタイムスキャン機能を備えたAndroidアプリケーションを開発し、誤検知率0.0368%の精度で事前に登録された偽・誤情報と同一性の高い画像 (60%以上で同一性ありと判断) を検知。

今後の課題・展望



- 技術開発では無線技術向けの追加開発及び対象モダルの拡充
- 社会実装では利用主体との協力関係の構築が必要

放送設備を使用した検証

- 今回は送信機と受信機を有線で結合しての実証であったが、放送局の放送設備 (待機系) を使用し、本放送への重畳に向けた実際に必要な手順を明確化する。

受信機開発

- 今回は製品化されているIPDC受信機を使用した。スマートフォンアプリケーションなどとシームレスにデータ連携できる受信機を開発する。また、実用化にむけて詳細なデータ設計を行い、偽・誤情報検証の効率化・高速化を図る。

モータル拡充

- 今回は画像・動画のみを対象とした偽・誤情報検証であったが、テキストも含めた総合的な偽・誤情報検証を行う。

主たる実証の成果と今後の課題・展望

実証番号6: エヴィクサー株式会社

代表者コメント



エヴィクサー株式会社
取締役執行役員COO

鈴木 久晴 様

“ 本テーマ・事業に 関連するコメント

偽・誤情報の拡散を防ぎ、信頼できる情報環境を守るために、音響技術を駆使した真正性保証システムを構築しました。誰もが安心して情報を受け取れる社会を実現するため、社会実装と普及に全力で挑みます！

主たる実証の成果 (技術開発の成果 及び 社会実装の進展)



- 放送コンテンツに音響透かし等を埋め込み、改ざん有無の確認技術を開発
- ケーブルテレビ局等 (ゼクサTV) で透かし入り放送を流す実証を実施

「音響透かしによる真正性保証・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放送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システムを構築した。上記の偽・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の高いシステムの設計、構築を行った。

再委託先のケーブルテレビ局にて、音響透かしを埋め込みと音響フィンガープリントのDB登録を行い、放送・配信を行った。一般視聴者環境のスマートフォンで音響透かしと音響フィンガープリントによる検出が可能かどうかの検証ができた。普及啓発に関して、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、関係者へのヒアリングを行った。

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。

今後の課題・展望



- 技術開発では対象データの形態拡張及び特許出願を検討予定
- 社会実装では組み先及び顧客の開拓と、国際標準化を企図
- 令和6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、プライシングモデルの実証実験
- リードユーザーの獲得
 - 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
 - 音楽団体
- 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ジネスモデルの検討
- 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、標準開発組織 (SDO) との連携検討、技術標準・ベストプラクティス開発に貢献
- 特許戦略の検討・実施
- 一般消費者が活用できるUI・UXを実現したサービス、アプリケーションの提供
- 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にも導入アプローチを進める予定



MIC

総務省

Ministry of Internal Affairs
and Communications

Agenda

背景・目的、実施内容、スケジュール

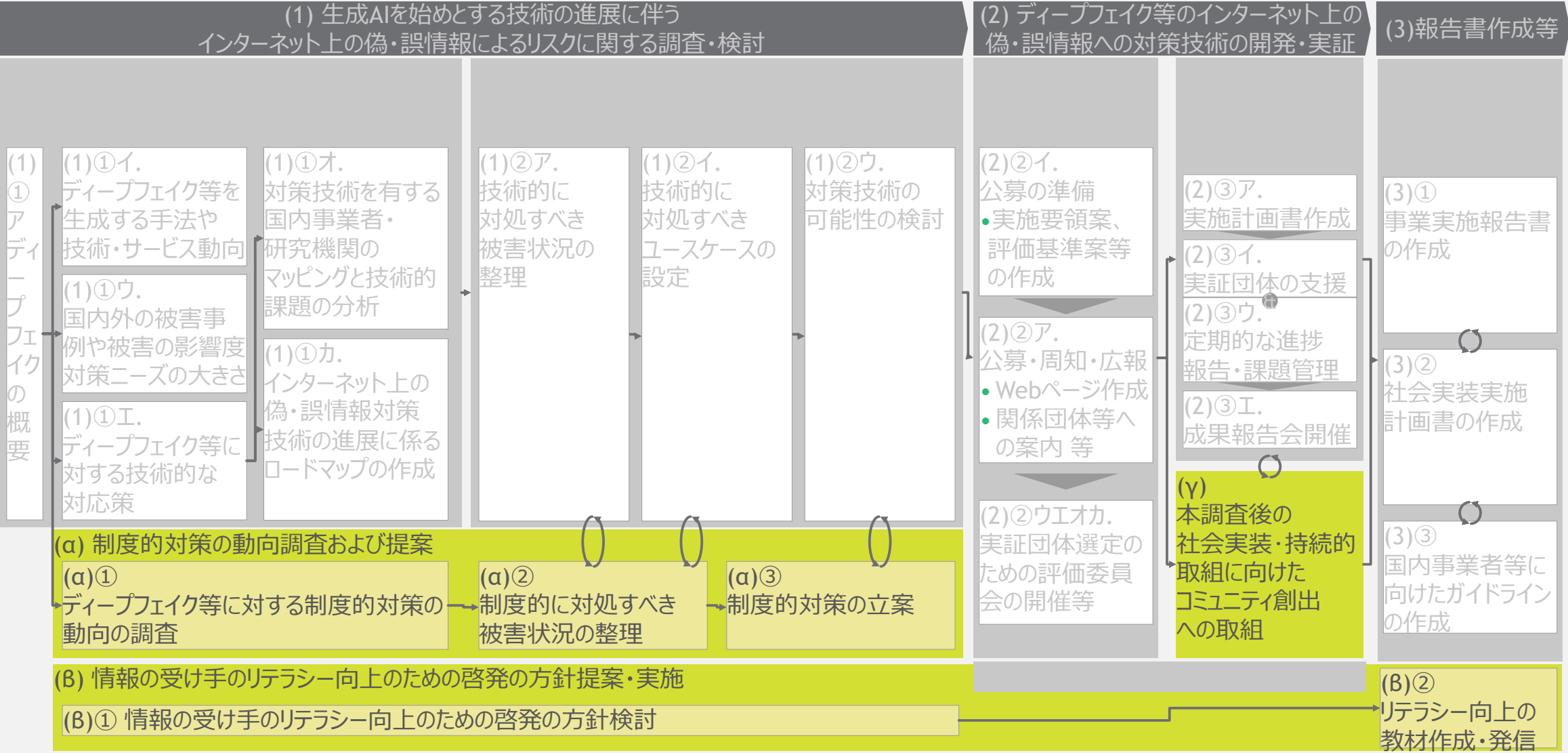
(1) 生成AIを始めとする技術の進展に伴うインターネット上の偽・誤情報によるリスクに関する調査・検討

(2) ディープフェイク等のインターネット上の偽・誤情報への対策技術の開発・実証

(3) 報告書作成等

➤ その他 検討事項

実施内容





各国はディープフェイク抑止のため規制・法整備・体制構築 等の制度的対策を推進 (1/3)

各国におけるディープフェイクへの制度的対策の例

下線: 2024年以降の最新動向

国名	規制対象	概要	対策内容
 アメリカ	AI開発事業者	大統領令により電子透かし義務化を含む規制	2023年10月、バイデン米大統領は、AIの安全性確保やプライバシー保護に関する規制を定めた大統領令¹に署名。<u>ディープフェイクコンテンツへの電子透かし義務化</u>の指針を商務省が策定すると発表
	政府・研究機関	ディープフェイクに関する研究支援を指示	2021年、全米科学財団 (NSF) と米国国立標準技術研究所 (NIST) に対し、<u>ディープフェイクを識別し、対策をとるための技術開発を支援することを命じる法律 (IOGAN法)</u> を成立
	AI開発事業者 SNS事業者 個人	総合的な偽・誤情報対策としてディープフェイク責任法案を審議中	2023年9月にディープフェイクへの総合的な対応方針として、<u>ディープフェイクの定義、体制構築、規則、罰則について示した法案</u>の審議が開始
	SNS事業者 個人	選挙妨害対策を目的に、州独自の法令を策定	- テキサス州とカリフォルニア州では2019年に<u>選挙の結果に影響を与えるディープフェイクを禁止</u>² - カリフォルニア州では2024年9月に、生成AIを用いて内容を改変した選挙広告に対して、<u>改変した事実の開示の義務付けや、ディープフェイクコンテンツを選挙管理委員会 等が意図的に拡散することを取り締まる法案に州知事が署名</u> - 同月、フェイスブックやX 等のオンラインプラットフォームに対し、選挙関連のディープフェイクの削除やラベル付けを義務化する法案に対して州知事が署名。2025年に施行予定 - さらに、AI開発事業者に対して、自社の生成AIによって作成・改変されたコンテンツに対し、その情報を付与する機能や、それらのコンテンツを検出可能なツール等を提供することを求める法案に署名。2026年1月に施行予定
	個人	肖像権侵害対策や選挙妨害対策を目的に、州独自の法令を策定	- カリフォルニア州は2023年に<u>同意なしのポルノディープフェイクを禁止</u> - ニューヨーク州、バージニア州では<u>不法な性的なディープフェイクに対し罰則</u>を設定
		AIを使用した非合意の性的コンテンツ作成の違法化	2024年7月に、米国上院で、AIを使用した非合意の性的コンテンツ作成を違法とする法案 (DEFIANCE法案) が可決 - 被害者は民事訴訟を提起し、損害賠償を請求可能
		同意なしに個人の声・顔・身体を再現したディープフェイク作成の違法化	2024年7月に、議員が上院議会に対し、<u>同意なしに個人の顔・声・身体 等を再現したディープフェイクが作成された場合、損害賠償請求を可能とする法案を提出</u>

1. "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" 2. カリフォルニア州では2023年1月までのサンセット法
Source: 公開情報をもとにBCG分析



(α)① ディープフェイク等に対する
制度的対策の動向の調査

各国はディープフェイク抑止のため規制・法整備・体制構築 等の制度的対策を推進 (2/3)

各国におけるディープフェイクへの制度的対策の例

下線: 2024年以降の最新動向

国名	規制対象	概要	対策内容
 EU	AI開発事業者 個人	人工知能の利用や規制をまとめた 欧州（EU）AI規制法が成立	2024年5月、ディープフェイクの脅威に対処するために包括的な規制(AI規制法案)として、ディープフェイクコンテンツをアップロードする場合にそのコンテンツが生成AIであることの明示を求めるAI規制法が成立 - 特定のコンテンツがディープフェイクかどうかを検証可能にするために電子透かしを提案
	SNS事業者	ソーシャルメディアにディープフェイクの 削除を義務付け	2022年6月、デジタルサービス法に基づき、SNS事業者を含むオンラインプラットフォームに対し、自社のプラットフォームからディープフェイクやその他の偽情報を削除・ブロックする方法の開示を求め、偽・誤情報対策の強化を要請 - 違反した場合は、最大でグローバルの売上高の6%の罰金が科される可能性あり
	SNS事業者	選挙に関するディープフェイクへの対 策を義務化	2024年3月、デジタスサービス法に基づき、オンラインプラットフォームに対して、ディープフェイク等のAIによって生成されたコンテンツのラベル付けや利用規約の制定、選挙期間中の選挙の安全性を脅かすコンテンツの収益化や拡散力の低減等を求めるガイドラインを発表 - 違反した場合は、最大でグローバルの売上高の6%の罰金が科される可能性あり
 英国	企業	オンライン安全法によりオンライン上 の違法行為の規制強化及び企業 への罰則	2023年10月26日に、英国オンライン安全法(Online Safety Act, 2023)が制定。オフコム(放送通信庁)を規制当局とし、違法又は子どもに有害なコンテンツや活動によるリスクを特定・軽減・管理する義務をオンラインサービスの提供者に課し、個人にとってより安全なオンラインサービスを提供
 韓国	個人	ディープフェイク犯罪への罰則を目的 に性暴力処罰法を改正	2020年に韓国は、性暴力犯罪の処罰などに関する特例法（性暴力処罰法）の改正により、性的なディープフェイク動画制作・頒布を犯罪として規定し、最高5年の懲役(営利目的の場合は最高7年)または最高5,000万ウォンの罰金を設定 2024年9月には法改正を行い、性的な偽画像の所持や閲覧まで処罰対象としたほか、制作・頒布への処罰を強化
	個人	ディープフェイクの政治利用を 禁止する改正公職選挙法を施行	2024年1月、改正公職選挙法が施行され、投開票日の90日前からは選挙運動を目的としたディープフェイクの制作などが禁止された
	個人	選挙におけるディープフェイクの 検知・削除要請部隊を設置	2024年の選挙において、中央選挙管理委員会は、ディープフェイクによる映像などを確認する専門の担当者をも70人配置し、違反投稿の削除要請や警察への告発対応を実施



各国はディープフェイク抑止のため規制・法整備・体制構築 等の制度的対策を推進 (3/3)

各国におけるディープフェイクへの制度的対策の例

下線: 2024年以降の最新動向

国名	規制対象	概要	対策内容
 シンガポール	政府・研究機関	国民教育の義務付け	デジタルメディア、リテラシー、サイバーセキュリティ、オンライン上の詐欺に関する国民の教育を義務付け
	政府・研究機関	オンライン上のディープフェイクを含む不正行為を防止する機関の設立	2024年3月、The Centre for Advanced Technologies in Online Safety(CATOS)を立ち上げ、 <u>ディープフェイクを含むオンライン空間での危害を監視・対処するための技術開発を推進</u> - <u>5年で5,000万シンガポールドル(約57.5億円)¹を投資したOnline Trust and Safety Research Programの一環。検知ツール、電子透かし、コンテンツ認証等の技術開発を目指す</u>
	SNS事業者	プラットフォームに対し、政府が命令を可能とする法案の可決	2023年7月、Online Criminal Harms Actを可決。オンライン上の詐欺に関連したアカウント・コンテンツから利用者を守るため、政府がオンラインプラットフォームに対して命令を下すことを認める
	SNS事業者 個人 個人	オンラインセーフティ法の施行 選挙におけるディープフェイク広告を禁止	2023年2月、オンラインセーフティ法を施行。個人、企業、インターネットサービスプロバイダーに悪意あるディープフェイクの削除、犯罪コンテンツの削除、アクセスのブロックなどを命令する事が可能に 2024年10月、選挙において、デジタルで生成・修正された、候補者が言っていない・していないことをしたかのように見せるデジタルコンテンツの公開を犯罪として扱う改正法案を可決
 中国	AI開発事業者 個人	ディープフェイク抑制のため「インターネット情報サービスディープフェイク管理規定」を策定	2023年1月、深層合成サービスプロバイダーとユーザーに対し、 <u>改ざんされたコンテンツには明示的にラベルを付け、個人の画像や音声を編集する場合、本人に通知することを義務付ける規制を施行</u>
 豪州	SNS事業者	プラットフォームへの規制を強化、プラットフォームへ広告責任を問う訴訟を実施	2022年3月、 <u>Meta Platforms社に対して、虚偽広告に対する取り組みが不十分だとして、プラットフォーム側の責任を問う訴訟を実施</u> 2023年8月、 <u>デジタルプラットフォームに対する新たな規制案を作成し、オーストラリア政府へ提言。最終的にオーストラリア政府と合意済</u>
	SNS事業者	SNS事業者への二重情報対策の要請	2024年9月、SNS事業者を対象に、二重情報対策を怠った場合、最大で売上高の5%に相当する罰金を科す法案を提出 (ただし、言論の自由との兼ね合い等により、12月に廃案)
	個人	性的なディープフェイクの作成・共有の違法化	2024年8月、非合意の性的なディープフェイク画像の作成者および共有者に、刑事罰を科す法案を可決

1. 1シンガポールドル=115円として計算 (2024年12月25日時点のレート)
Source: 公開情報をもとにBCG分析



(α)② 制度的に対処すべき
被害状況の整理

偽・誤情報による国内外の具体的被害事例(37事例)を7つの被害類型で整理

直近5年で発生している偽・誤情報による具体的被害事例（1/2）

未確認（但し、将来的に被害・影響は発生する可能性）

被害類型		ディープフェイクの種類							
		文章生成		画像生成		音声生成		動画生成	
				ヒトを生成	モノ・コト(風景等)を生成	ヒトを生成	モノ・コト(風景等)を生成	ヒトを生成	モノ・コト(風景等)を生成
社会への 攻撃	災害時の 扇動	②日向灘自身に係る 偽情報 ③能登半島自身に係る 偽情報	未確認	①ハリケーン・ヘレンでの ディープフェイク画像の 拡散 ⑤マウイ島の山火事の際 に人為的な災害であるこ とを示唆するデマ画像が 拡散	未確認	未確認	未確認	④能登半島地震の際に、 架空の地名での救助要 請などのデマ情報が拡散	
	平時の 扇動	②9中国でChatGPTを 使って列車の衝突に関す る偽ニュースを拡散	③0日本政府の処理水に 関する誤情報が記者會 見のフェイク画像と共に 拡散	③1アメリカ国防総省近く で大規模な爆発という説 明と共に偽画像をSNSで 拡散	未確認	未確認	未確認	未確認	
	政治的 混乱の 誘発 (選挙 含む)	⑧「オバマ氏がヒラリー・ クリントンの支持を撤回」 という誤情報をラジオで 拡散 ⑪米大統領選挙に係る 偽情報 ⑫韓国大統領選挙に 係る偽情報 ⑮ AIが作成したフィッ シングメールによるサイバ ー攻撃	未確認	未確認	⑥パレスチナ出身の父を 持つ米国のモデルの声を 学習し、イスラエル支持 のスピーチを生成し拡散 ⑨米大統領に似た声の フェイク音声で予備選挙 に介入 ⑬バイデン大統領の偽の 音声自動音声通話機 能により拡散、選挙妨害 に利用される	未確認	⑦米大統領のフェイク動 画を使い、非公式なプロ グラムへのサインアップを 奨励 ⑩岸田首相が実際には 発言していない内容の フェイク動画がSNSにて 拡散 ⑭インドの著名人が首相 を批判するディープフェ イク動画がインド国内の SNSに拡散	未確認	



(α)② 制度的に対処すべき
被害状況の整理

偽・誤情報による国内外の具体的被害事例(37事例)を7つの被害類型で整理

直近5年で発生している偽・誤情報による具体的被害事例 (2/2)

未確認 (但し、将来的に被害・影響は発生する可能性)

被害類型		ディープフェイクの種類						
		文章生成	画像生成		音声生成		動画生成	
			ヒトを生成	モノ・コト(風景等)を生成	ヒトを生成	モノ・コト(風景等)を生成	ヒトを生成	モノ・コト(風景等)を生成
企業への 攻撃	企業の 評判や 信頼の 失墜	26 飲食店等の店舗に 対しコロナ陽性者が確認 されたというデマ情報が 多発	未確認	27 大手食品会社に対し 「ワクチン入りの食品を 販売予定」というデマ 画像が拡散	未確認	未確認	未確認	28 国際的なイベントにて、 テスラの商品の欠陥を 示唆するフェイク動画が 放映
	個人への 攻撃	個人の評 判や信用 の失墜	32 米国著名司会者を ChatGPTで調べたところ 自身が横領を働いたとす る虚偽で有害な情報が でっち上げられたと主張し、 Open AIを提訴	未確認	未確認	未確認	未確認	33 知人の喫煙や飲酒を しているディープフェイク 動画を嫌がらせ目的で 拡散
詐欺	詐欺	17 仮想通貨取引所にお いて偽造された本人 確認書類で本人確認を 突破 - パスポート、マイナ ンバーカード等 22 SNS上で著名人を装 い金銭を詐取24 偽CFOの指示に従い 送金	18 銀行口座開設時のAI で生成した顔の画像で 顔認証を突破	未確認	16 フェイク音声での取引 先のなりすましによる振り 込め詐欺29 銀行口座開設時のAI で生成した声の音声で 声帯認証を突破20 フェイク音声による振り 込め詐欺	未確認	21 ディープフェイク技術で 別人になりすまし、オンラ イン面接を受験23 生成AIを利用して有 名アイドルになりすまし、 ビデオ通話を通して600 万円詐取される被害25 イーロン・マスク氏を 装った投資詐欺	未確認
	利益目的 の肖像権 ・著作権 侵害	未確認	35 著名シンガーのわいせ つなフェイク画像がソー シャルメディアで拡散37 著名人の顔でディープ フェイクを作成できるアプリ に対する集団訴訟	未確認	34 アーティストの声を模し て生成AI作成の楽曲を 拡散	未確認	36 日本の女性芸能人 150名のフェイクボロボ 動画が違法に公開	未確認



(α)② 制度的に対処すべき
被害状況の整理

24年以降の直近では、「災害」、「詐欺」、「選挙」の被害事例が悪化・増加傾向にある 被害状況の最新動向

災害時の 扇動

ハリケーン・ヘレンでのディープフェイク画像の拡散 (2024年9月)

- 豪雨の被害にあって怯える子どもや、洪水の中救助活動を手伝う政治家のディープフェイク画像がSNSで拡散 (トランプ氏のフェイク画像は2日間で16万回シェア)

日向灘地震に係る偽情報 (2024年8月)

- 生成AIで作られたとみられる被害状況の偽画像や科学的根拠がない人工地震や地震予知とする偽情報がSNS上で拡散
- 注意喚起を装ったスパム投稿が相次ぎ、公的機関や公共放送からの転載画像をクリックするとアダルトサイトや通販サイトに誘導

能登半島地震に係る偽情報 (2024年1月)

- 生成AIで作られたとみられる被害状況の偽画像や科学的根拠がない人工地震や地震予知とする偽情報がSNS上で拡散
- 注意喚起を装ったスパム投稿が相次ぎ、公的機関や公共放送からの転載画像をクリックするとアダルトサイトや通販サイトに誘導

詐欺

イーロン・マスク氏を装った投資詐欺 (2024年11月)

- イーロン・マスク氏のディープフェイクによる詐欺は世界で120億ドルの被害が発生
- 騙されたユーザーが詐欺だとわかったあとに見直しても本物と信じられる品質のディープフェイクが拡散

SNS上で著名人を装い金銭を詐取 (2024年2月)

- 堀江貴文氏を装ったLINEアカウントに誘導され、アカウントで利用された音声を本人と思い込み、音声案内に従い2億2000万円の金額を振込

偽CFOの指示に従い送金 (2024年2月)

- ディープフェイクで作成された偽物の最高財務責任者 (CFO)のほか複数の同僚がビデオ会議に出席。会計担当者が本物だと信じ込み、偽CFOの音声指示に従い合計2,500万米ドルを送金

政治的 混乱の 誘発 (選挙 含む)

米大統領選挙に係る偽情報 (2024年11月)

- ハリス氏の偽の演説が、2億のフォロワーを持つイーロン・マスク氏により拡散
- 自動音声通話により、生成AIが利用され、バイデン氏に似せた声で「11月の大統領選に向けて投票を温存する」よう呼び掛け
- イランが有権者の間に分裂を起こすために、報道機関を装ったウェブサイトに掲載するためのフェイクニュースを生成

インドの著名人を装った選挙妨害 (2024年4月)

- インドの著名な俳優が、「モディ首相は選挙公約を守らず、深刻な経済問題に対処しなかった」と語り、国民会議派が選挙で使用するスローガンを発言したようにみせたフェイク動画がソーシャルメディア上で拡散

韓国大統領選挙に係る偽情報 (2024年3月)

- 過去の大統領の演説を編集した偽動画にて、「無能で腐敗した政府」と自らを批判する内容を捏造し、SNS上で拡散



(α)② 制度的に対処すべき
被害状況の整理

「影響範囲」x「被害の深刻度」の二軸で被害状況を分析した結果、災害、政治、詐欺、企業信用に係る被害事例の対策ニーズが大きい

偽・誤情報の被害状況の分析・整理

分析の前提

分析対象

直近5年で発生している、偽・誤情報による具体的被害事例の調査結果で得られた37事例

分析方法

対策のニーズを影響範囲と被害の深刻度で評価

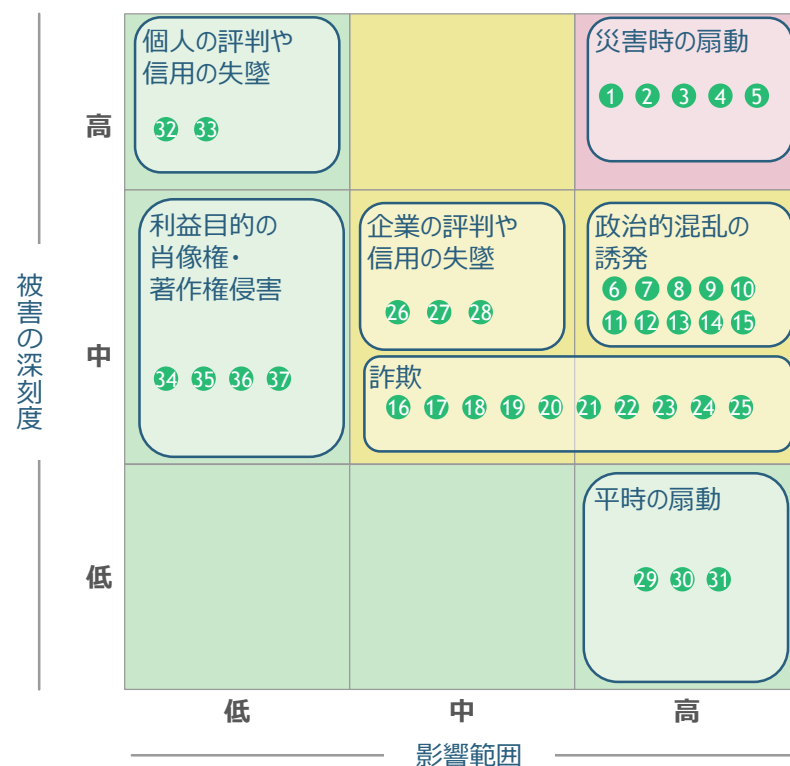
影響範囲

- 被害発生時の影響範囲を三段階で評価
 - 高: 日本国/世界全体に影響
 - 中: 特定地域/産業に影響
 - 低: 個人/限定的なコミュニティに影響

被害の深刻度

- 米国のサイバーインシデント深刻度¹を基準に三段階で評価
 - 高: 国家の重要なインフラサービスの提供、国家政府の安定、国民の生命に差し迫った脅威をもたらす可能性がある
 - 中: 公衆衛生または安全、国家安全保障、外交関係、市民の自由、国民の安全に影響を与える可能性がある
 - 低: 上記に当てはまらない

分析結果



対策ニーズ大

- ハリケーン・ヘレンでのディープフェイク画像の拡散
- 日向灘地震に係る偽情報
- 能登半島地震に係る偽情報
- 能登半島地震の際に、架空の地名での救助要請などのデマ情報が拡散
- マウイ島の山火事の際に人為的な災害であることを示唆するデマ画像が拡散

対策ニーズ中

- パレスチナ出身の父を持つ米国のモデルの声を学習し、イスラエル支持のスピーチを拡散
- 米大統領のフェイク動画を、非公式なプログラムへのサインアップを奨励
- 「オバマ氏がヒラリー・クリントンの支持を撤回」という誤情報をラジオで拡散
- 米大統領に似た声のフェイク音声で予備選挙に介入
- 岸田首相が実際には発言していない内容のフェイク動画がSNSにて拡散
- 米大統領選挙に係る偽情報
- 韓国大統領選挙に係る偽情報
- バイデン大統領の偽の音声で自動音声通話機能により拡散、選挙妨害に利用される
- インドの著名人が首相を批判するディープフェイク動画がインド国内のSNSに拡散
- AIが作成したフィッシングメールによるサイバー攻撃
- フェイク音声での取引先のみならず、振り込み詐欺事件
- 仮想通貨取引所において偽造された本人確認書類で本人確認を突破
- パスポート、マイナンバーカード等
- 銀行口座開設時のAIで生成した顔の画像で顔認証を突破
- 銀行口座開設時のAIで生成した声の音声で声帯認証を突破
- フェイク音声による振り込み詐欺事件
- ディープフェイク技術で別人になりすまし、オンライン面接を受験
- SNS上で著名人を装い金銭を詐取
- 生成AIを利用して有名アイドルになりすまし、ビデオ通話を通して600万円搾取される被害
- 偽CFOの指示に従い送金
- イーロン・マスク氏を装った投資詐欺
- 飲食店等の店舗に対しコロナ陽性者が確認されたというデマ情報が多発
- 大手食品会社に対して「ワクチン入りの食品を販売予定」というデマ画像が拡散
- 国際的なイベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映

対策ニーズ小

- 中国でChatGPTを使って列車の衝突に関する偽ニュースを拡散
- 日本政府の処理水に関する誤情報が記者会見のフェイク画像と共に拡散
- アメリカ、ワシントンの国防総省近くで大規模な爆発という説明と共に偽画像をSNSで拡散
- 米国著名司会者をChatGPTで調べたところ自身が横領を働いたとする虚偽で有害な情報が出たことと主張し、Open AIを提訴
- 知人の喫煙や飲酒をしているディープフェイク動画を嫌がらせ目的で拡散
- アーティストの声を模して生成AI作成の楽曲を拡散
- 著名シンガーのわいせつなフェイク画像がソーシャルメディアで拡散
- 日本の女性芸能人150名のフェイクポルノ動画が違法に公開
- 著名人の顔でディープフェイクを作成できるアプリに対する集団訴訟

1. <https://www.cisa.gov/news-events/news/cisa-national-cyber-incident-scoring-system-nciss>



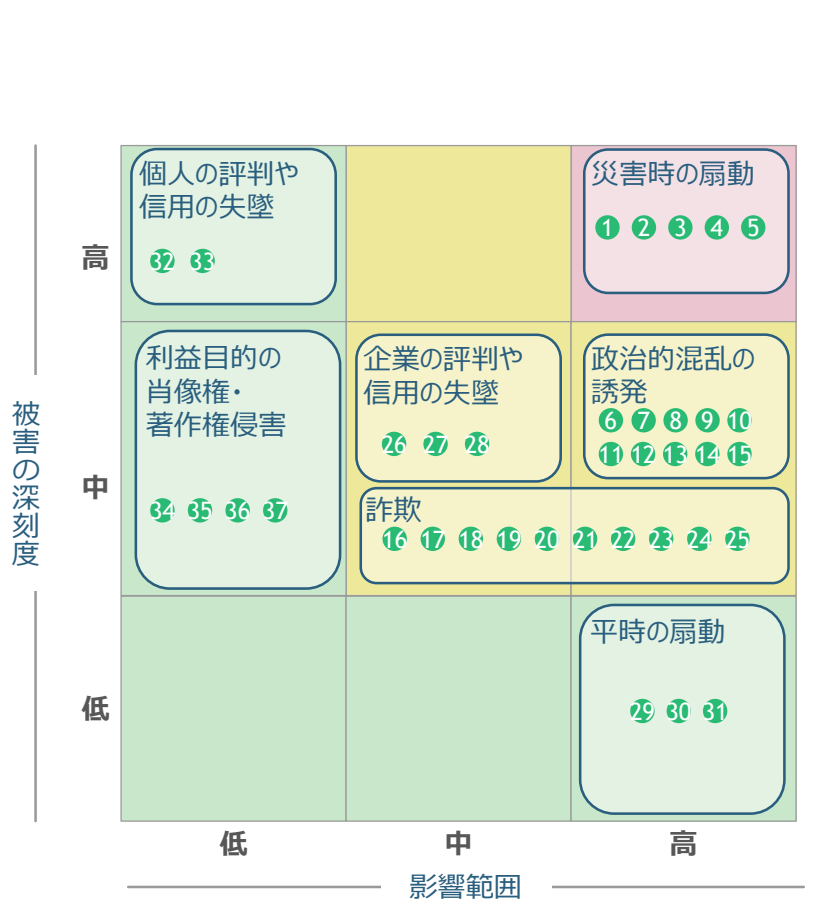
偽・誤情報の流通・拡散に関する技術的対策の類型

流通・拡散防止の対策		ITソリューション・サービス	関連技術	主たる内容
全体を把握する	流通情報の健全性把握	情報流通状況の可視化	ソーシャルリスニング	SNS等で流通する情報を可視化し、影響度を測定する
まず作らせない	二重・誤情報の作成防止	ファイル防御・改ざん修復	透かし	真正な情報が、ディープフェイク等を用いて改変された場合には、その痕跡が残るようにすることで、情報のみだりな改変を防ぐ
作られても 見つける	発信情報の真偽判別	生成・改変有無の検知	検知	コンテンツ単体、もしくは複合的なコンテンツを用いて、情報の真偽判断を行う
		情報収集・データベース提供	検知 データベース構築	SNS投稿等の発信情報を分析して偽・誤情報を検知・収集し、検知精度の向上に役立てる
そもそも正しい ものであることを 担保する	発信者・発信情報の真正性 担保	情報発信者の認証付与・ 確認	電子署名	インターネット上の情報に、発信者の情報を紐づけ、信頼性を担保する
		発信情報の認証付与・確認	電子署名 透かし	各種コンテンツに目に見えない形で、著作権者・使用許諾先などの各種情報を埋め込む
広がらない ように防ぐ	拡散防止	拡散者へのアラート・投稿 制限	検知 透かし データベース構築	SNS等で、ディープフェイクを検知し、拡散者アラート提示や投稿制限を行う



技術的対策と合わせ、制度的対策により技術提供者やプラットフォーム事業者への規制や、生成者・発信者等への刑罰の明確化などの制度的対策も必要と考えられる

偽・誤情報による被害に対する制度的対策 (1/2)



例)
国レベルあるいは企業・個人に関する偽・誤情報の拡散

例)
肖像権/著作権を侵害したコンテンツの作成による利益享受

例)
なりすましでの詐欺・不正行為による利益享受

制度的対策 (例)

- 偽・誤情報の検知の体制構築
- 技術提供者への規制 (生成時の電子すかしの義務化等)
- PFerへの流布防止策の導入 (特定のコンテンツへの警告表示、発信者の情報提供等)
- 生成者・発信者・拡散者への刑罰

- 技術提供者への規制(有名人は作成不可、性的コンテンツ作成禁止 等)
- 生成者・発信者・購入者への刑罰

- 技術提供者への規制(著名人は生成不可等)
- 生成者・詐欺犯罪者への刑罰



制度的対策としては、「技術提供者への規制」「メディア/PFerへの規制」「生成者・発信者等への規制」「検知・被害対応等が可能な体制の構築」が挙げられる

偽・誤情報による被害に対する制度的対策 (2/2)

規制・刑罰の整備

体制構築

①技術提供主体への規制

②生成者・発信者への規制

③メディア/PFerへの規制

④政府機関による 検知・被害対応体制構築

例)
国レベルあるいは
企業・個人に関する
偽・誤情報の拡散

- 生成物へのラベリング等の義務付け
- 生成物の限定
 - 著名人は作成不可
 - ポルノ作成禁止 等

例)
肖像権/著作権を
侵害したコンテンツの
作成による利益享受

- 生成物へのラベリング等の義務付け
- 生成AIツールの特定用途(政治工作等)での利用禁止・刑事罰の整備

- 流布防止策の導入義務化
 - コンテンツへの警告表示・削除への対応
 - 必要に応じた、発信者情報の提供協力義務化
- 偽・誤情報検知への協力要請
 - ディープフェイクの検出(検知ソフトウェア導入など)の義務付け
 - 情報の真正性の検出の義務付け
 - メディアに対する電子透かしの義務化 等

- (特に国・政治に関連する)ディープフェイク被害の検知・対応を行う組織体制の構築
 - 特に選挙等における偽・誤情報拡散を検知
 - 偽・誤情報投稿について、メディア/PFer/警察等と連携した対応の実施

例)
なりすましでの
詐欺・不正行為による
利益享受

- ディープフェイクでのなりすまし犯罪の定義明確化、刑事罰の整備

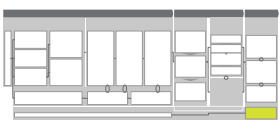
- 詐欺広告の防止等のための広告事前調査をメディア/PFerに義務付け



ハルシネーションの考慮、教養がある方からのスモールスタート、警鐘広告の活用、 既存リテラシーコンテンツの集約・更新の観点を要考慮

情報の受け手のリテラシー向上のための啓発の方針検討

1. **AIは今後も間違えるもの**だと認識したうえで、新規技術を利用することが大切、そのためにも利用者側のリテラシーの向上が必要
2. **最初はリテラシーの教養がある方**(メディアやファクトチェック機関)が使用、ある程度のリテラシーに関する教養を満たし、**社会の関心が高まってから一般公開**するなど長い道のりが必要
3. 動画コンテンツの広告として**フェイク動画の警鐘をならす広告**を挿入することで、リテラシーを高められる可能性はあり
4. 様々な企業が実施している**既存のリテラシーコンテンツをまとめアップデート**していく観点も重要



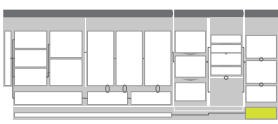
本調査結果を踏まえ、「生成AIリテラシー教材」へのディープフェイク関連の追加コンテンツ案 を作成

リテラシー向上の教材の作成

本調査を受けた追加項目

3 生成AI活用にあって注意すべきポイントの中で、偽・誤情報観点の資料が存在している。
そのため、その内容に追加する形で、以下コンテンツを新規作成

- ディープフェイクの基礎知識
 - ディープフェイクの概要
 - ディープフェイク技術の進化の歴史
 - チープフェイクとの違い
- ディープフェイクによる被害
 - ディープフェイクの被害の具体事例
- ディープフェイクへの対処の仕方
 - ディープフェイク対処のための技術の紹介
 - SNS事業者の取組紹介
 - 犯罪者にならないための心構え



作成した追加コンテンツの発信方法としては、官公庁からの発信・教育機関や企業における学習への導入・メディアを活用した周知など、複数手法の組み合わせが必要

リテラシー向上の教材の発信

所属組織からの呼びかけ/参加義務を伴う講義の実施:

学校・大学や職場等ターゲットが所属している組織に協力を要請し、組織経由で導入コンテンツを周知 (受動的に学習する人に対しては、学校や職場等所属組織からの強制力が必要)

官公庁からの案内:

関連する官公庁関連のウェブサイトやSNSを通して導入コンテンツを周知し、コンテンツの信頼感を担保

ニュースサイトやトラディショナルメディアへの掲載:

主にはデジタルリテラシーが高くない人が日々の情報収集に使用している新聞、テレビや、インターネットのニュースサイト等に導入コンテンツに関するリリースを行い、ニュースや記事で取り上げてもらい周知 (貴省の広報ご担当とも連携)

偽・誤情報流布媒体となりうるSNS等での、広告やコンテンツ提供:

コンテンツをもとに短い動画を作成し、SNSでの広告やおすすめコンテンツとして表示し、実際に偽・誤情報に触れる可能性の高い層に周知 (SNS事業者との連携を想定)



(γ)本調査後の社会実装・持続的
取組に向けたコミュニティ創出への
取組

持続的取組に向けてはコンソーシアム等の方法でコミュニティ形成をすることが有効

コンソーシアムの種類			
技術開発・実証		標準化・ガイドライン策定	
目的		情報共有・普及啓発	
具体例		情報共有・普及啓発	
▼		▼	
本件の趣旨に 照らした際の評価		技術開発主体と顧客候補の両者 を招聘し、前者のアピールや、後者 のニーズ共有などが可能であり、 マッチングに適している	

次ページに詳細案



(γ)本調査後の社会実装・持続的
取組に向けたコミュニティ創出への
取組

情報共有・普及啓発のコンソーシアムの例として、 被害類型ごとに、技術開発主体と顧客候補が集まり、情報共有を行うことを想定

詐欺		災害時の扇動		企業の評判の毀損	
参加者	技術開発主体		● 被害類型ごとに適用可能な技術を開発した 技術開発主体		
	顧客候補	エンドユーザー	● 特にリスクを感じている 企業 (大手中心)	● 地方公共団体	● 被害を受けやすい大手/著名企業
		技術の組込み先	● 企業向けセキュリティサービス 提供元 ● メールセキュリティ 提供元 等	● 総合防災情報システム 提供元 ● リアルタイム危機情報収集サービス 提供元 等	● ソーシャルリスニングツール 提供元 ● ブランド保護ソフトウェア 提供元 ● PR会社 等
アジェンダ (例)			● 事務局による発表： 国内外の最新動向 の共有 ● 被害の発生状況/事例、技術的対策の実施状況/事例 等 ● 技術開発主体による発表 ：開発している 対策技術に関する発表 - 関連する対策技術の一般的な概況、自社の技術のアピール (概要や優位性など) 等 ● 顧客候補による対策状況や課題意識 に関する発表 ● 技術開発主体と顧客候補のコネクションづくりを進める 交流会		



[bcg.com](https://www.bcg.com)