



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

目次

- 1. 本事業の背景・目的 P. 3
- 2. 本事業における実施事項 P. 10

目次

1. 本事業の背景・目的
2. 本事業における実施事項

目次

1. 本事業の背景・目的

1-1. 本事業の背景・目的

1-2. 本事業の実施事項の全体像

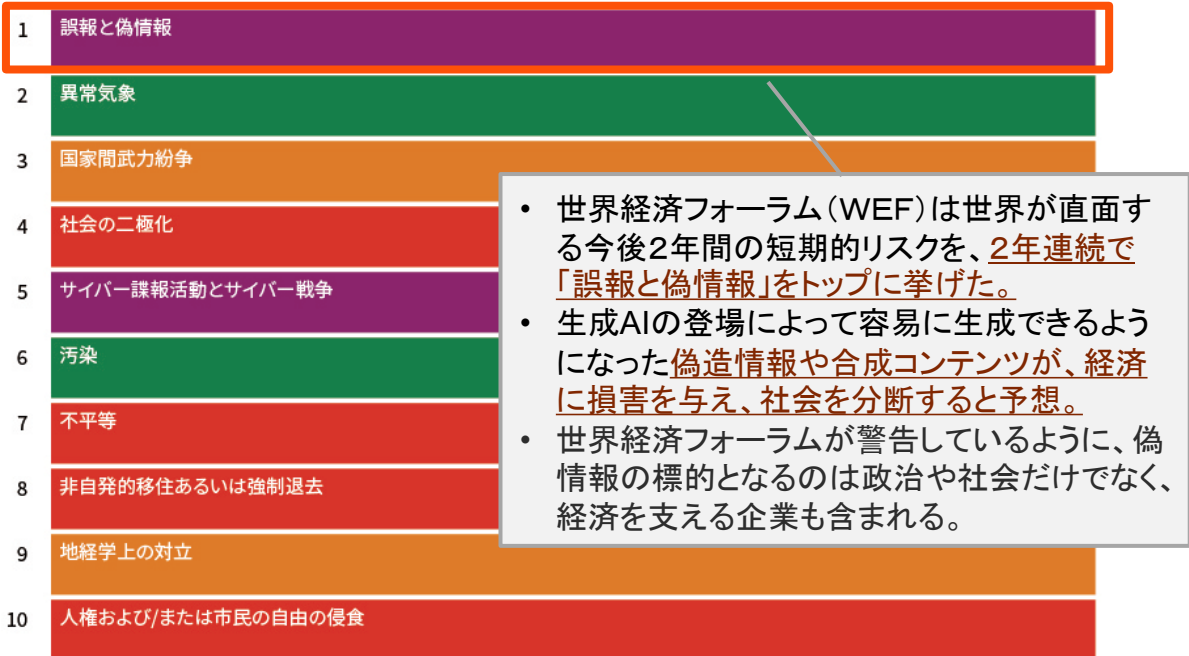
2. 本事業における実施事項

① 本事業の背景

- ✓ 生成AIをはじめとする人工知能を用いた意図的な偽・誤情報(文章、画像、音声、動画等)は、世界各国で政治や外交のみならず社会に大きな影響を与えている。そのため、現代の生活・ビジネスにおいて不可欠なインターネット上を安心・安全な空間にするために、**偽・誤情報対策技術の開発・実証および社会実装が急務である。**

グローバルリスクへの指摘

短期(今後2年間)



リスク分類 ● 経済 ● 環境 ● 地政学 ● 社会 ● テクノロジー

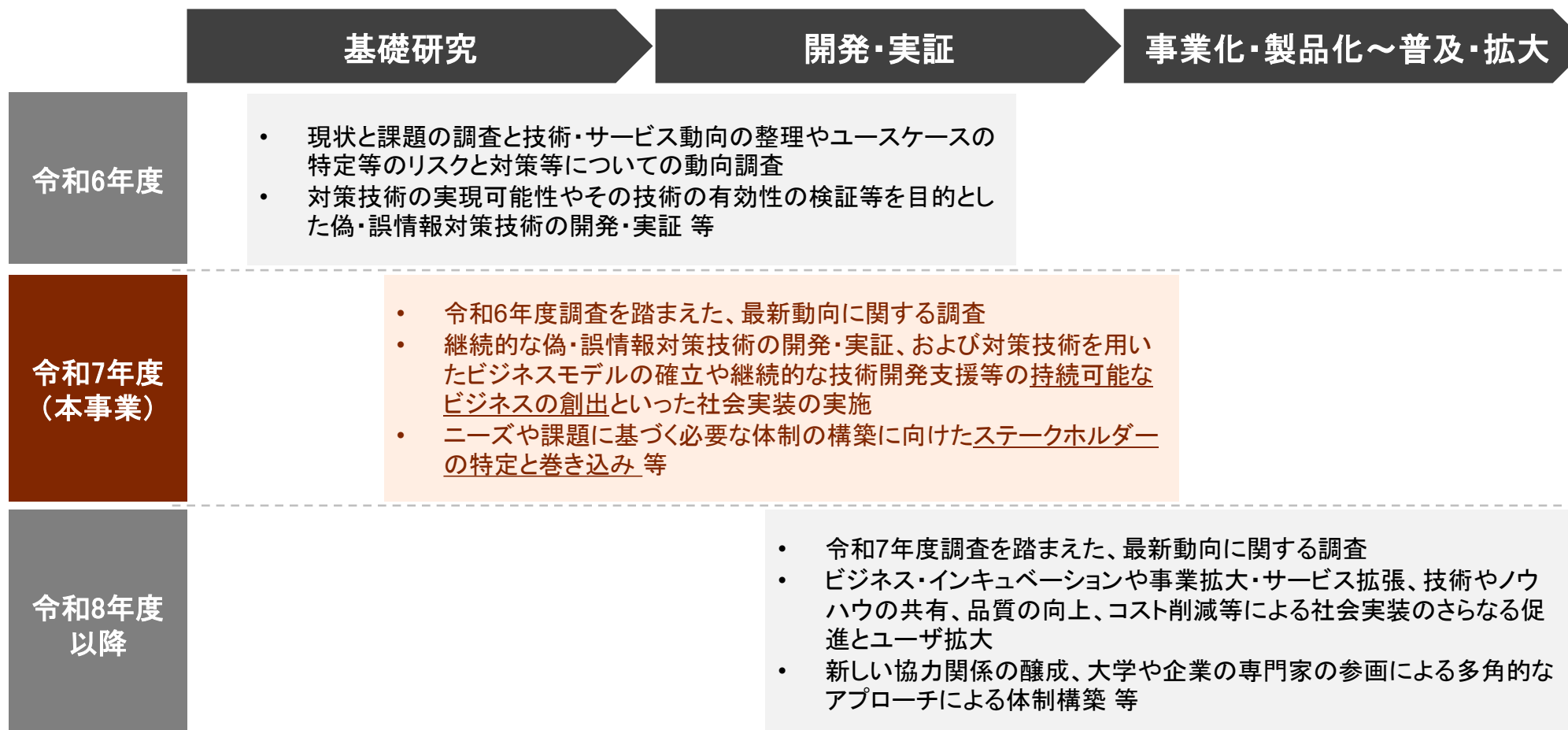
出典: グローバルリスクの短期的な重要度ランキング(世界経済フォーラム「グローバルリスク報告書2025年版」)
<https://www.weforum.org/publications/global-risks-report-2025/>

偽・誤情報関する被害事例

| カテゴリ | 内容例 |
|---------|--|
| なりすまし動画 | <ul style="list-style-type: none"> 政治家のなりすまし動画によって誤ったメッセージが発信されれば、世論操作のみならず、国家を揺るがすような重大な事態を招く可能性もあり、社会的、経済的なダメージは計り知れない。 特定の人物の評判を落とす、社会的地位を失わせるといった目的で、生成AIを悪用し、映像を捏造するようなケースも危惧される。 |
| デマ情報 | <ul style="list-style-type: none"> 生成AIによって作られた「架空の災害映像」「架空の衛星写真」などが災害時にSNSなどで拡散され、デマによって大きな混乱を引き起こすリスクがある。 |
| 詐欺 | <ul style="list-style-type: none"> 芸能人や有名人の映像を使ってディープフェイクを作成し、詐欺広告を出したり、フィッシングサイトや詐欺サイトに誘導させたりするようなケースもみられる。 さらに、ディープフェイク音声やビデオによる詐欺もみられ、手口はより巧妙化している。 |
| 不正認証 | <ul style="list-style-type: none"> ディープフェイクによって、顔認証や声紋認証などのセキュリティシステムの突破を目論む事例が存在する。 あらゆるサービスへ不正にログインできるようになるため、サイバー攻撃が容易になる。 |

② 本事業の目的

- ✓ 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・拡散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に本事業を推進している。昨年度の技術検証等に加え、今年度は対策技術の具体的な社会実装に向けての方策や必要な体制構築等について検討した。



目次

1. 本事業の背景・目的

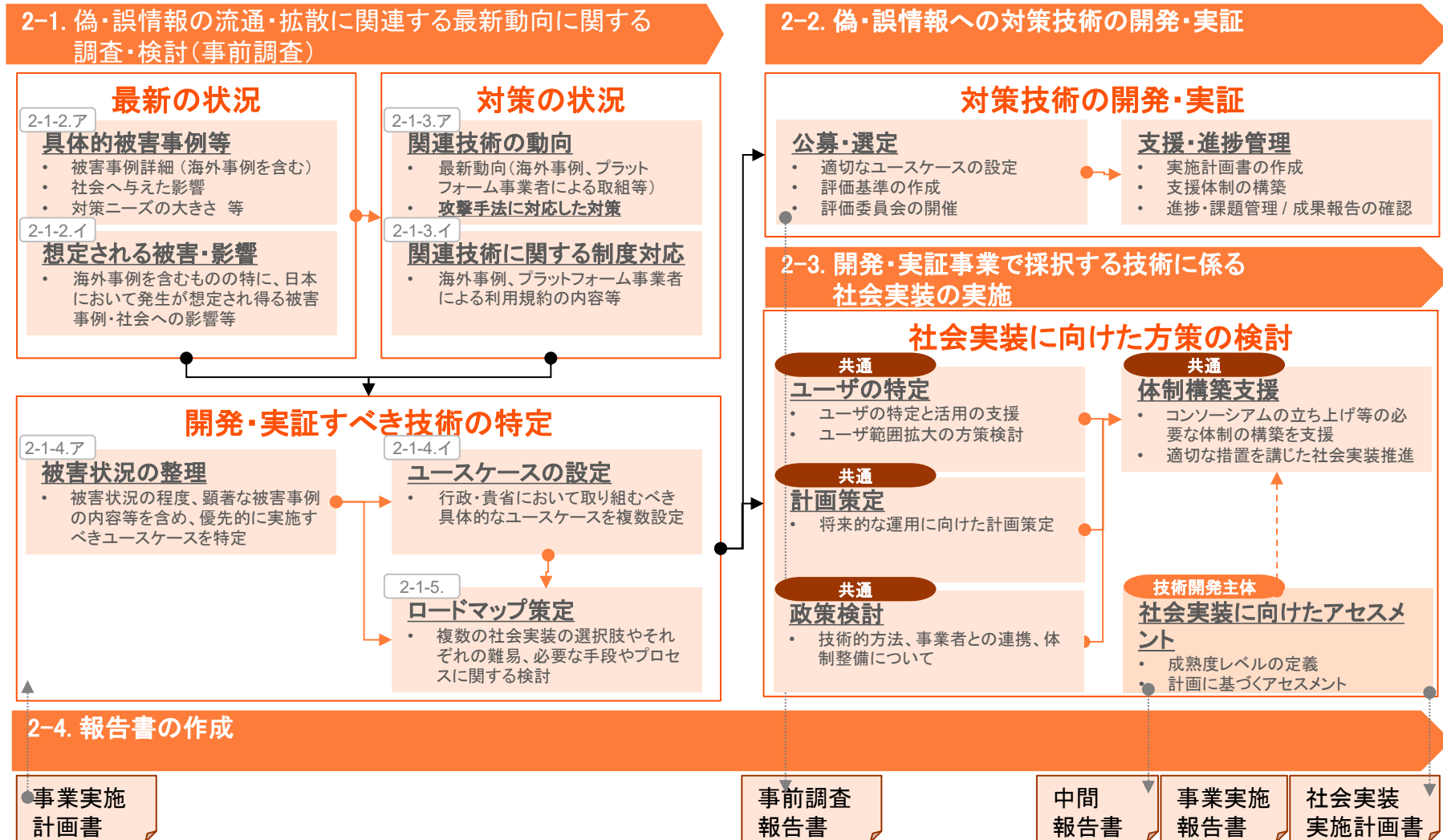
1-1. 本事業の背景・目的

1-2. 本事業の実施事項の全体像

2. 本事業における実施事項

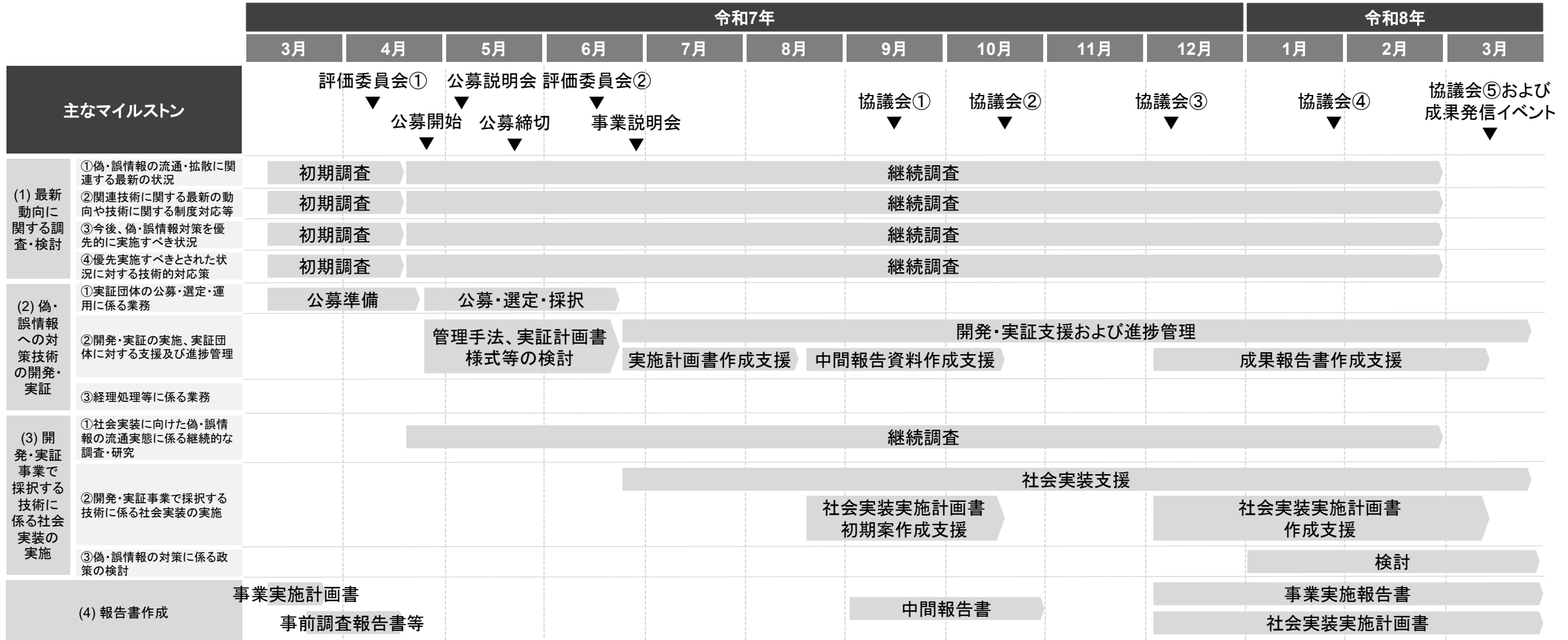
① 本事業の実施事項の全体像

✓ 本事業内容の全体像は以下の通り。直近5年のディープフェイクなどの被害事例を踏まえ、偽・誤情報対策技術の開発と実証及び社会実装に向けた取組を実施した。



② 本事業の実施スケジュール

✓ 本事業の実施スケジュールは以下の通り。



目次

1. 本事業の背景・目的
2. 本事業における実施事項

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

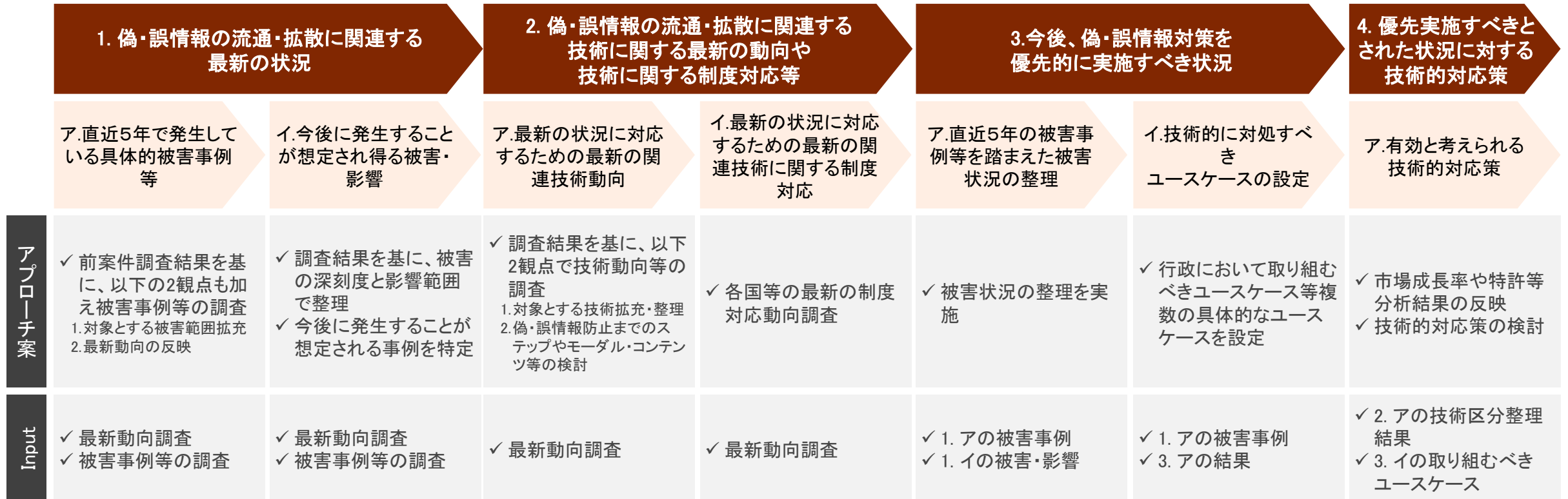
2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

① 事前調査の全体像

✓ 偽・誤情報の流通・拡散に関連する最新動向に関する調査として、以下のステップで調査を進める。

1.被害事例等の調査、2.技術動向等の調査を踏まえ、3.技術的に対処すべきユースケースを設定し、4.優先実施すべきとされる技術的対応策を策定する。



目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

① 本調査にて得られた被害事例一覧

✓ 直近5年で発生している、偽・誤情報による被害事例として、以下事例が得られた。

事例タイトル(偽・誤情報全般に係るもの)

- ① ドナルド・トランプ Jr.の偽音声
- ② “容疑者は不法移民” 偽情報が英で暴動に
- ③ プーチン大統領が降伏を宣言したという動画
- ④ “地震予知” 偽情報 スпамで荒れたSNS
- ⑤ 天皇陛下や上皇陛下がエプスタイン氏と写った偽画像
- ⑥ キング牧師がトランプを支持したという音声は偽
- ⑦ トランプ大統領が地名区の変更を望む偽音声
- ⑧ 静岡県の水害巡りフェイク画像が拡散
- ⑨ センリス刑事裁判所が偽ニュース動画の作者を名誉毀損で有罪判決
- ⑩ 台湾選挙に関するアメリカ合衆国の議員のディープフェイク
- ⑪ トランプ大統領とアルゼンチン、イタリア、ハンガリー首脳の偽記念写真
- ⑫ 兵庫県斎藤知事に関する偽画像、百条委が否定
- ⑬ 菅元首相がワクチン接種を誤魔化したと誤情報
- ⑭ 国会乱闘中における捏造画像
- ⑮ 米大統領選、ハリス陣営が集会画像を生成?
- ⑯ 有名人かたる偽の広告“放置で被害” メタを全国5地裁に提訴
- ⑰ クリステア・フリーランドのディープフェイク動画とトロント・スターの偽記事
- ⑱ 巨大な剣が発見されたという偽画像
- ⑲ GTA 5のゲームプレイ映像であり、ブラジルの飛行機事故ではない
- ⑳ 金閣寺の金箔を剥がしたという偽情報

事例タイトル(主にディープフェイクに係るもの)

- ① 中国で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 米国防総省近くで爆発偽画像をSNS上で拡散
- ④ 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 架空の制度を承認をする米大統領のフェイク動画
- 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
- ⑦ 「オバマ氏が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 マウイ島の山火事が人為的な災害と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の偽情報が多発
- ⑫ 大手食品会社に対し「ワクチン入りの食品を販売予定」との偽画像が拡散
- ⑬ 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め詐欺事件
- ⑮ 国際イベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映
- ⑯ 著名シンガーのわいせつ画像がソーシャルメディアで拡散
- ⑰ 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
- ⑱ 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
- ⑲ フェイク音声による振り込め詐欺事件
- ⑳ 著名人の顔でDFを作成できるアプリへの訴訟
- ㉑ ディープフェイク技術で別人になりすまし、面接を受験

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関連する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関連する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-1-2.イ 今後の発生が想定され得る被害・影響

① 今後に発生することが想定され得る被害・影響

✓ 過去事例を鑑みると、**国民の安全を脅かす「国民の扇動」は対策ニーズが大きいと想定される。**

調査事例(偽・誤情報全般に係るもの)

対策ニーズ大

- ① ドナルド・トランプJr.の偽音声
- ② “容疑者は不法移民”偽情報が英で暴動に
- ③ ブーテン大統領が降伏を宣言したという動画
- ④ “地震予知”偽情報 スпамで荒れたSNS

対策ニーズ中

- ⑤ 天皇陛下や上皇陛下がエプスタイン氏と写った偽画像
- ⑥ キング牧師がトランプを支持したという音声は偽
- ⑦ トランプ大統領が地名区の変更を望む偽音声
- ⑧ 静岡県の水害巡りフェイク画像が拡散
- ⑨ センリス刑事裁判所が偽ニュース動画の作者を名誉毀損で有罪判決
- ⑩ 台湾選挙に関するアメリカ合衆国の議員のディープフェイク
- ⑪ トランプ大統領とアルゼンチン、イタリア、ハンガリー首脳の写真偽記念写真

対策ニーズ小

- ⑫ 兵庫県斎藤知事に関する偽画像、百条委が否定
- ⑬ 菅元首相がワクチン接種を誤魔化したと誤情報
- ⑭ 国会乱闘中における捏造画像
- ⑮ 米大統領選、ハリス陣営が集会画像を生成?
- ⑯ 有名人かたる偽の広告“放置で被害”メタを全国5地裁に提訴
- ⑰ クリステア・フリーランドのディープフェイク動画とトロント・スターの偽記事
- ⑱ 巨大な剣が発見されたという偽画像
- ⑲ GTA 5のゲームプレイ映像であり、ブラジルの飛行機事故ではない
- ⑳ 金閣寺の金箔を剥がしたという偽情報

調査事例(主にディープフェイクに係るもの)

対策ニーズ大

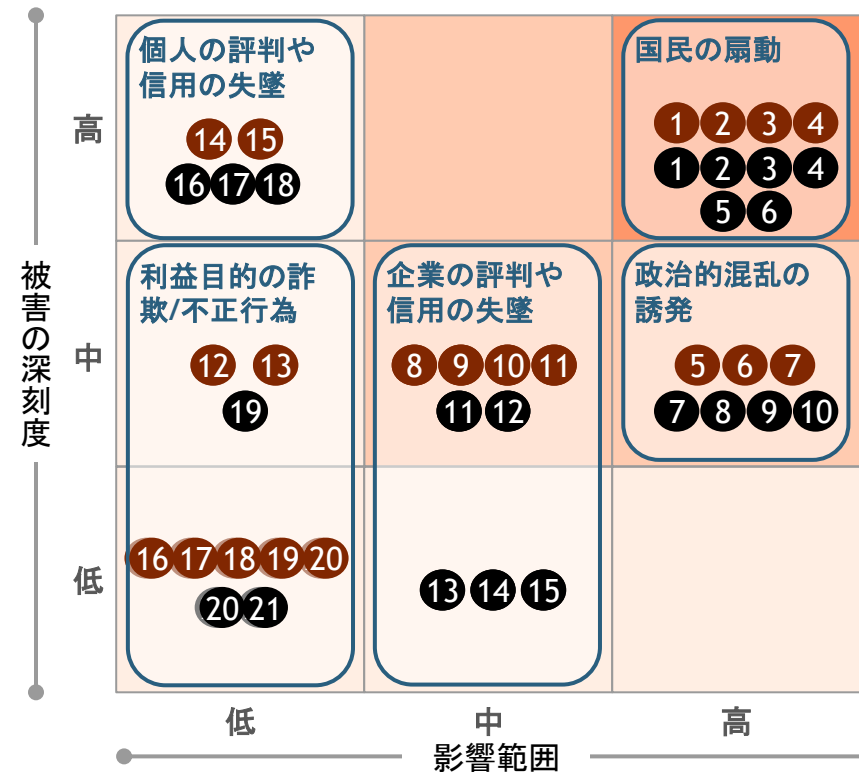
- ① 中国で列車の衝突に関する偽ニュースを拡散
- ② 日本政府の処理水に関する誤情報が記者会見のフェイク画像と共に拡散
- ③ 米国防総省近くで爆発偽画像をSNS上で拡散
- ④ 米国のモデルの声を学習し、イスラエル支持のスピーチを生成し拡散
- ⑤ 架空の制度を承認をする米大統領のフェイク動画
- ⑥ 能登半島地震の際に、架空の地名での救助要請などのデマ情報が拡散

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 マウイ島の山火事が人為的な災害と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の偽情報が多発
- 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

対策ニーズ小

- ⑬ 大手食品会社に対し「ワクチン入りの食品を販売予定」との偽画像が拡散
- 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
- ⑮ 国際イベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映
- ⑯ 著名シンガーのわいせつ画像がソーシャルメディアで拡散
- ⑰ 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
- ⑱ 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
- ⑲ フェイク音声による振り込め詐欺事件
- ⑳ 著名人の顔でDFを作成できるアプリへの訴訟
- ㉑ ディープフェイク技術で別人になりすまし、面接を受験



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: 公衆衛生又は安全、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

1. <https://www.cisa.gov/news-events/news/cisa-national-cyber-incident-scoring-system-ncis>16

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関連する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関連する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

② 偽・誤情報の流通・拡散に関連する技術区分(1/2)

✓ 偽・誤情報の流通・拡散に関連する最新の技術動向の調査を踏まえ、偽・誤情報への対応技術を整理・分類した。

| 偽・誤情報への対応技術 | | | 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・ 改ざん検知技術 | コンテンツ単独 による検知 | 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 <ul style="list-style-type: none"> 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)の処理の痕跡など 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、まばたきのパターンなど、生体信号の欠如 | — | ● | ● | ● |
| | | 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 <ul style="list-style-type: none"> データドリブンでのディープフェイクの検知 | — | ● | ● | ● |
| | 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 <ul style="list-style-type: none"> 疑義言説収集等 | ● | ● | ● | ● |
| 真正性保証 ・信頼性判断 支援技術 | 真正性保証 | 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 <ul style="list-style-type: none"> デジタル証明書 オリジネータープロフィール等 | ● | ● | ● | ● |
| | 信頼性判断・ 改ざん検知 | 電子透かし | コンテンツに対して人間の知覚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 | ● | ● | ● | ● |
| 情報流通状況の 可視化・分析 技術 | トレンド可視化 | マクロ分析(ソーシャルリスニング) | テキストマイニングや時系列分析クラスタリングなどの手法を用いて、情報流通空間全体のトレンドを可視化する技術 | ● | ● | ● | ● |
| | 偽・誤情報の目的・ 対象の予測 | マイクロ分析(ソーシャルモニタリング・アトリビューション) | 特定のトピックや個別ユーザー・アカウントの発信内容を把握し、拡散状況や影響を分析し偽情報が拡散しそうなトピックや、情報操作の相手になりそうな集団・個人を前もって予測する技術 | ● | ● | ● | ● |

② 偽・誤情報の流通・拡散に関連する技術区分(2/2)

✓ 偽・誤情報の流通・拡散に関連する最新の技術動向の調査を踏まえ、偽・誤情報への対応技術を整理・分類した。

| 偽・誤情報への対応技術 | | 概要 | 技術が適用可能なコンテンツ | | | | | |
|------------------------------|-----------|---------|---|--|----|----|---|---|
| | | | 文章 | 画像 | 音声 | 動画 | | |
| 情報の拡散防止 情報の拡散防止・無効化技術 | 情報の拡散防止 | 情報のタグ付け | 改ざんリスクの高いコンテンツに注釈やタグ付けを行い、情報の受け取り手に対する注意喚起を行う仕組み | ● | ● | ● | ● | |
| | | 情報の拡散制限 | 疑わしいと判断されたコンテンツについて拡散範囲の制限を行う仕組み | ● | ● | ● | ● | |
| | 偽・誤情報の無効化 | 情報の復元 | 予め(人間の顔など)映像や画像の特徴に関する情報を、その外側の部分に埋め込んでおくことで、ディープフェイク攻撃を受けた際、復元モデルを用いることで、周辺に埋め込まれたオリジナルの情報を基に復元できる技術 | — | ● | — | ● | |
| | | 訂正情報の発信 | プレバンキング(事前介入) | 偽・誤情報が拡散する前に、その手法や特徴を学ばせたり、「…のようなデマが流れる可能性があるが、それは間違いである」旨の注意喚起を実施する技術 | ● | ● | ● | ● |
| | | | デバンキング(事後修正) | 流通した偽・誤情報に触れた人々に対し、間違っている箇所を提示するなどして正しい情報を効果的に提供する対応技術 | ● | ● | ● | ● |

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関連する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関連する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

② 海外における偽・誤情報に関する制度・規約(1/3)

| 地域 | 制度等名称 | 詳細 |
|----|--|--|
| 米国 | Disinformation Governance Board (国土安全保障省(DHS)) | <ul style="list-style-type: none"> 2022年4月27日に国土安全保障省(DHS)が「Disinformation Governance Board (偽情報ガバナンス委員会)」の創設を発表。トップとなる事務局長には、非政府組織のフェローであるNina Jankowicz (ニーナ・ヤンコヴィッチ)氏が着任し、委員会は、<u>偽情報の脅威に対処するためのベストプラクティスの収集を行う</u>ことを目的とする。 委員会の創設に対しては、ソーシャルメディア上で炎上し、メディアが批判を行った。ジョージ・オーウェルの小説「1984」のThe Ministry of Truth(真理省)になぞらえてメディアも国民を監視する目的との批判が行われ、DHSは委員会の検討を中止、2022年8月に活動を停止した。 |
| | 競争及び、技術プラットフォームの説明責任の強化 | <ul style="list-style-type: none"> 「通信品位法 第230条」により、技術プラットフォームは第三者が発信する情報について原則として責任を負わず、また、有害なコンテンツに対する削除等の対応に関して責任を負っていない」との指摘があった。同法における免責規定について、<u>プラットフォームに偽・誤情報の流通について責任を負わせる方向での議論</u>が実施。 2022年9月8日にホワイトハウスは専門家や実務家とのリスニングセッションを開催し、「競争及び、技術プラットフォームの説明責任の強化」を公表した。大統領がけん引し、超党派の議員グループの活動となった。 |
| | NY State Assembly Bill 2023-A4217 ※ニューヨーク州 | <ul style="list-style-type: none"> 不正な意図でディープフェイクを作成することを犯罪行為とする法案であり、Steven Otis議員(民主党)によって提出。 意図的に誤解を招くフェイク画像又はフェイクオーディオを作成した者に罰金や懲役刑を課すことや<u>ディープフェイク作成に対する刑事罰に加えて、被害者に対して損害賠償を支払う</u>ことが規定されている。 |

② 海外における偽・誤情報に関する制度・規約(1/3)

| 地域 | 制度等名称 | 詳細 |
|------|---|--|
| 米国 | Bill H.72 | <ul style="list-style-type: none"> 州政府に影響を与えるディープフェイクの拡散等に対応するため、タスクフォースを設置し、ディープフェイクの影響評価やディープフェイク防止技術のベストプラクティス等について評価を行うためのもの。同タスクフォースは、2024年1月1日までにディープフェイク技術の潜在的な用途、リスク等について報告した。 作業部会の報告書には、デジタルコンテンツ偽造及びディープフェイクの定義の修正に関する推奨事項と、デジタルコンテンツ偽造技術及びディープフェイクの展開によって影響を受ける可能性がある他のコードセクションの修正に関する推奨事項が含まれる。 |
| 欧州連合 | DSA法 (Digital Services Act) NY State Assembly Bill 2023-A4217 ※ニューヨーク州 | <ul style="list-style-type: none"> <u>偽情報や操作的な悪用行為、未成年者への悪影響など、システミックリスクが社会と民主主義に及ぼしうる負の影響も考慮すべき領域</u>であるとし、これには、意図的に不正確な、あるいは誤解を招くような情報、又は経済的利益を得る目的で作成されたボットや偽アカウントを使用するなど、偽情報を含む情報の増幅を目的とした協調的な操作が含まれ、これらは<u>特に未成年者などサービスの受け手である弱者にとって有害であると定義</u>のうえ、以下のように規定。 <ul style="list-style-type: none"> 前文2項: 社会的リスクとして、違法コンテンツとオンライン上の偽情報を併記 前文9項: DSAの目的として、<u>オンラインでの違法コンテンツの流布と、偽情報やその他のコンテンツの流布が引き起こす可能性のある社会的リスクへの対処を明記</u> 前文69項: 広告ターゲティングによる悪影響として、偽情報キャンペーンへの加担の可能性を明記 前文83項: VLOPやVLOSEのシステミックリスクが公衆衛生に関連する組織的な偽情報キャンペーンからも発生する可能性があることを明記 前文84項: VLOPやVLOSEの提供者はシステミックリスクを評価する際、<u>違法ではないが特定されたシステムリスクに寄与する偽情報などの、誤解や欺瞞的なコンテンツを増幅するためのどのようにサービスが利用されるか</u>について、特に注意を払うべきと明記 前文88項: システミックリスクが偽情報キャンペーンに関連する場合には、VLOPやVLOSEの提供者は意識向上の活動も検討する必要があることを明記 前文95項: オンライン広告のもたらす可能性のあるリスクの例として、違法な広告等と併記して、偽情報を明記 前文104項: 偽情報に対して、違法コンテンツとは別に、<u>自主規制によって検討されるべき特定の分野</u>であると明記 前文106項: DSAがEUで確立されている自主規制の基礎になりうるとした上で、その具体例として<u>偽情報に関する行動規範について言及</u> |

② 海外における偽・誤情報に関する制度・規約(1/3)

| 地域 | 制度等名称 | 詳細 |
|---------|---|---|
| 英国 | Online Safety Act 2023 | <ul style="list-style-type: none">2023年10月に発効。虚偽であると知っている情報を受信者に心理的又は身体的危害を与えることを意図してインターネット上で送信した者に、6か月の禁錮刑を科す内容が含まれている。特に、相手に苦痛、不安や屈辱等を与える加害意図や、自分が性的満足を得ようとする意図があったと立証されれば、最高刑が懲役2年となる。 |
| オーストラリア | Disinformation and Misinformation Code (偽情報及び誤情報コード) | <ul style="list-style-type: none">誤情報と偽情報は、オーストラリア国民の安全と福祉、そして民主主義、社会、経済に脅威をもたらす。政府は、通信法改正法案(誤報及びディスインフォメーションとの闘い)2024を進めるつもりはないとした。 |
| 韓国 | ディープフェイク性犯罪防止法 | <ul style="list-style-type: none">2024年9月、韓国の国会はディープフェイクの技術を悪用した性的な偽の画像などが通信アプリで拡散され深刻な社会問題となっていることを受けて、所持したり、視聴したりした場合についても処罰の対象にするさらなる対策強化の法案を可決した。法案では「ディープフェイク」で作成された偽の画像や動画を所持、保存、それに視聴するなどした場合も処罰の対象となり、3年以下の懲役又は3000万ウォン以下、日本円にして320万円余りの罰金を科すとしている。また、偽の画像や動画を作成するなどした場合については、現行の5年以下の懲役から7年以下に引き上げるとした。 |

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

- 2. 本事業における実施事項
 - 2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)
 - 2-1-4.ア 直近5年の被害事例等を踏まえた被害状況の整理

① 優先的に実施すべきとする技術的対応策導出の考え方

✓ 「過去の被害事例をもとに整理した対処すべきユースケース」と、「日本・グローバルの調査結果等から注力すべき技術」等の観点から、有効と考えられる技術的対応策を特定する。

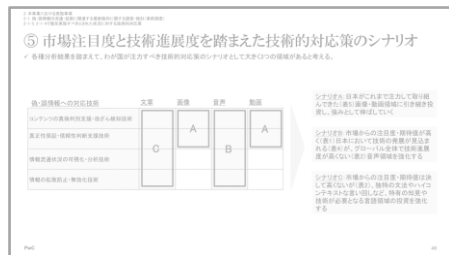
直近5年の偽・誤情報を踏まえ、対処すべきユースケースは何か

過去5年にて発生している被害事例をもとに、被害の深刻度及び影響範囲より対処すべきユースケースを特定。



日本・グローバルから注力すべき技術は何か

グローバルの投資・特許データ等をもとに「市場からの注目度」や日本・グローバルの特許数から注力すべき技術を導出。



有効と考えられる技術的対応策の特定

ユースケースと注力すべき技術より、技術的対応策を特定

⑥ ユースケースを踏まえた技術的対応策のシナリオの優先度

- ✓ 深刻度が大きく影響範囲の広い事例を、流布された情報の種類(モーダル)で分類したところ、特に画像・動画を使用するケースが多かった。
- ✓ 前項の技術的対応策のシナリオのうち、シナリオAを優先的に、次いでシナリオB、シナリオCに取り組みべきと考えられる。

| 事例分析結果 | 深刻度・影響範囲 | 代表的なユースケース | 事例数 | | | |
|-----------------|----------|---|-----|----|----|----|
| | | | 文章 | 画像 | 音声 | 動画 |
| | 高・高 | <ul style="list-style-type: none"> ✓ 災害時に、生成AIを用いて被災地の画像を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした。 ✓ 戦時下や暴動発生時において、生成AIにて動画を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした。 | 2 | 5 | 3 | 7 |
| | 中・高 | <ul style="list-style-type: none"> ✓ 選挙時に、候補者に関するSNSの偽文章を生成し、投票権を有する国民の興味を引き拡散を促すことで、特定地域の市民の支持率・政治情勢に影響を与えた。 ✓ 首相や大統領の偽音声を生じ、SNSで拡散することで、国民の支持率や政治的な混乱を促した。 | 4 | 6 | 3 | 4 |
| | 中・中 | | | | | |
| 合計 | | | 6 | 11 | 6 | 11 |
| 対応する技術的対応策のシナリオ | | | C | A | B | A |

より広範囲に影響を及ぼす、深刻な被害を防止するためには、シナリオAが特に有効と考えられる。

② 被害状況の程度、顕著な被害事例の大きさの評価

実施内容

- ア 直近5年の被害事例等を踏まえた被害状況の整理
 - ①の調査結果により判明した偽・誤情報の被害事例等を踏まえ、被害状況の程度、顕著な被害事例の内容等を含め、今後に偽・誤情報対策を優先的に実施すべきユースケースが明らかになるよう、直近5年の被害事例等を踏まえた被害状況を分析・整理する。
- イ 技術的に対処すべきユースケースの設定
 - ③アの分析・整理結果を踏まえ、行政において取り組むべきユースケース、その中で総務省が取り組むべきユースケースが分かるよう、複数の具体的なユースケースを設定する。

タスク/アウトプットイメージ

- ア 被害状況の程度、顕著な被害事例の大きさの評価
 - ①において整理した被害事例等において、被害の大きさを評価する



- イ 技術的に対処すべきユースケースの設定
 - 前項で特定した被害状況をもとに、行政において取り組むべきユースケース、その中で優先的に取り組むべきユースケースを整理する

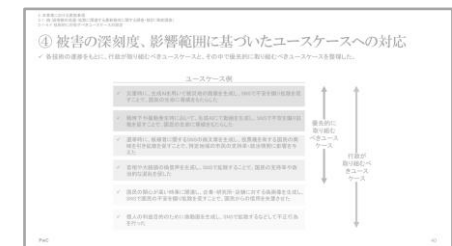
ユースケース案

各事例において、技術的に対処すべきと考える、被害の深刻度、影響範囲を踏まえ、ユースケースを導出する



ユースケースの整理

左記より得られたユースケースをもとに、行政が取り組むべきユースケースと、その中で優先的に取り組むべきユースケースを整理する



2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-1-4.ア 直近5年の被害事例等を踏まえた被害状況の整理

③ 直近5年の被害事例等を踏まえた被害状況の整理

✓ 被害の深刻度・影響度を鑑みると、**国民の扇動や政治的混乱、企業の評判の失墜を促す事例が発生しうると想定。**

調査事例(偽・誤情報全般に係るもの)

対策ニーズ大

- ① ドナルド・トランプJr.の偽音声
- ② “容疑者は不法移民”偽情報が英で暴動に
- ③ ブーチン大統領が降伏を宣言したという動画
- ④ “地震予知”偽情報 スпамで荒れたSNS

対策ニーズ中

- ⑤ 天皇陛下や上皇陛下がエプスタイン氏と写った偽画像
- ⑥ キング牧師がトランプを支持したという音声は偽
- ⑦ トランプ大統領が地名区の変更を望む偽音声
- ⑧ 静岡県の水害巡りフェイク画像が拡散
- ⑨ センリス刑事裁判所が偽ニュース動画の作者を名誉毀損で有罪判決
- ⑩ 台湾選挙に関するアメリカ合衆国の議員のディープフェイク
- ⑪ トランプ大統領とアルゼンチン、イタリア、ハンガリー首脳の偽記念写真

対策ニーズ小

- ⑫ 兵庫県斎藤知事に関する偽画像、百条委が否定
- ⑬ 菅元首相がワクチン接種を誤魔化したと誤情報
- ⑭ 国会乱闘中における捏造画像
- ⑮ 米大統領選、ハリス陣営が集会画像を生成?
- ⑯ 有名人かたる偽の広告“放置で被害”メタを全国5地裁に提訴
- ⑰ クリстяア・フリーランドのディープフェイク動画とトロント・スターの偽記事
- ⑱ 巨大な剣が発見されたという偽画像
- ⑲ GTA 5のゲームプレイ映像であり、ブラジルの飛行機事故ではない
- ⑳ 金閣寺の金箔を剥がしたという偽情報

調査事例(主にディープフェイクに係るもの)

対策ニーズ大

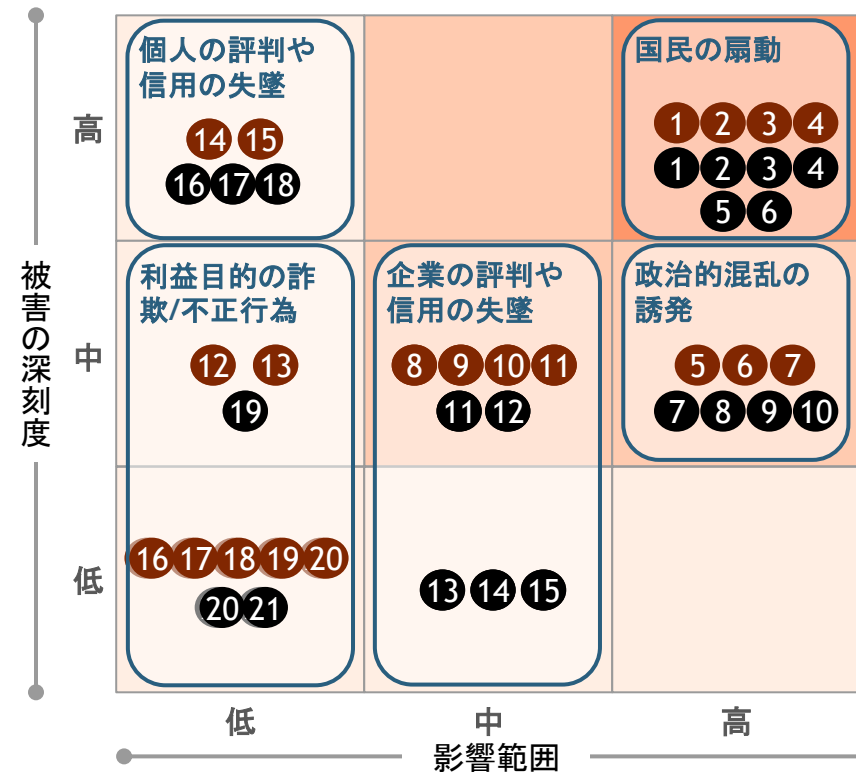
- ① 中国で列車の衝突に関する偽ニュースを拡散
- ② 日本政府の処理水に関する誤情報が記者会見のフェイク画像と共に拡散
- ③ 米国防総省近くで爆発偽画像をSNS上で拡散
- ④ 米国のモデルの声を学習し、イスラエル支持のスピーチを生成し拡散
- ⑤ 架空の制度を承認をする米大統領のフェイク動画
- ⑥ 能登半島地震の際に、架空の地名での救助要請などのデマ情報が拡散

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 マウイ島の山火事が人為的な災害と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の偽情報が多発
- 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

対策ニーズ小

- ⑬ 大手食品会社に対し「ワクチン入りの食品を販売予定」との偽画像が拡散
- 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
- ⑮ 国際イベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映
- ⑯ 著名シンガーのわいせつ画像がソーシャルメディアで拡散
- ⑰ 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
- ⑱ 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
- ⑲ フェイク音声による振り込め詐欺事件
- ⑳ 著名人の顔でDFを作成できるアプリへの訴訟
- ㉑ ディープフェイク技術で別人になりすまし、面接を受験



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: 公衆衛生又は安全、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

1. <https://www.cisa.gov/news-events/news/cisa-national-cyber-incident-scoring-system-ncis29>

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関連する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関連する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

① 今後、偽・誤情報対策を優先的に実施すべき状況

実施内容

- ア 直近5年の被害事例等を踏まえた被害状況の整理
 - ①の調査結果により判明した偽・誤情報の被害事例等を踏まえ、被害状況の程度、顕著な被害事例の内容等を含め、今後に偽・誤情報対策を優先的に実施すべきユースケースが明らかになるよう、直近5年の被害事例等を踏まえた被害状況を分析・整理する。
- イ 技術的に対処すべきユースケースの設定
 - ③アの分析・整理結果を踏まえ、行政において取り組むべきユースケース、その中で総務省が取り組むべきユースケースが分かるよう、複数の具体的なユースケースを設定する。

タスク/アウトプットイメージ

- ア 被害状況の程度、顕著な被害事例の大きさの評価
 - ・ ①において整理した被害事例等において、被害の大きさを評価する



- イ 技術的に対処すべきユースケースの設定
 - ・ 前項で特定した被害状況をもとに、行政において取り組むべきユースケース、その中で優先的に取り組むべきユースケースを整理する

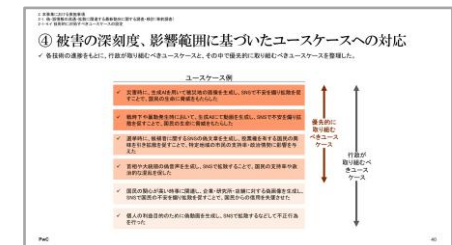
ユースケース案

各事例において、技術的に対処すべきと考える、被害の深刻度、影響範囲を踏まえ、ユースケースを導出する



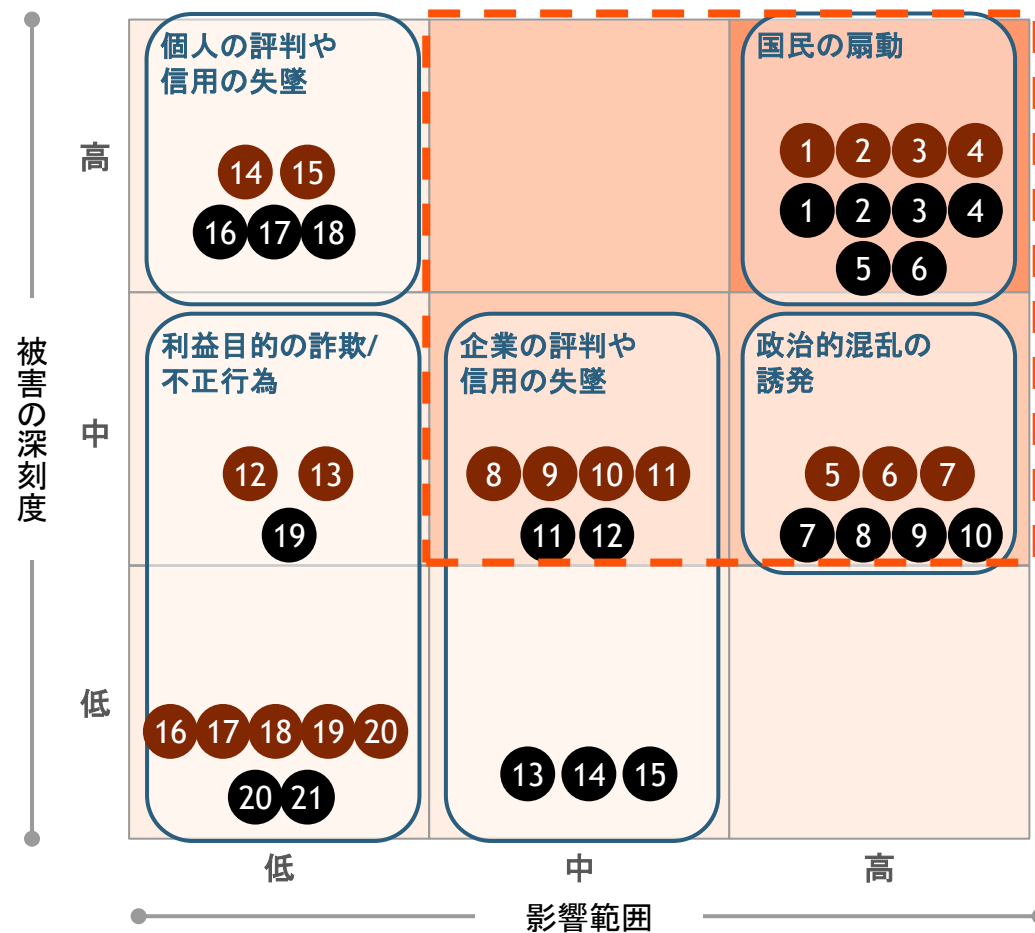
ユースケースの整理

左記より得られたユースケースをもとに、行政が取り組むべきユースケースと、その中で優先的に取り組むべきユースケースを整理する



② 被害の深刻度、影響範囲に基づいた本事業の優先度

✓ 「被害の深刻度」及び「影響範囲」が高い領域を主に貴省が取り組むべきと考える。



③ 被害の深刻度、影響範囲に基づいたユースケースの導出

✓ 被害の深刻度、影響範囲に基づき、マッピングした各事例より対処すべきユースケースを導出した。

調査事例(偽・誤情報全般に係るもの)

対策ニーズ大

- ① ドナルド・トランプJr.の偽音声
- ② “容疑者は不法移民” 偽情報が英で暴動に
- ③ ブーテン大統領が降伏を宣言したという動画
- ④ “地震予知” 偽情報 スпамで荒れたSNS

対策ニーズ中

- ⑤ 天皇陛下や上皇陛下がエプスタイン氏と写った偽画像
- ⑥ キング牧師がトランプを支持したという音声は偽
- ⑦ トランプ大統領が地名区の変更を望む偽音声
- ⑧ 静岡県の水害巡りフェイク画像が拡散
- ⑨ センリス刑事裁判所が偽ニュース動画の作者を名誉毀損で有罪判決
- ⑩ 台湾選挙に関するアメリカ合衆国の議員のディープフェイク
- ⑪ トランプ大統領とアルゼンチン、イタリア、ハンガリー首脳の写真偽記念写真

対策ニーズ小

- ⑫ 兵庫県斎藤知事に関する偽画像、百条委が否定
- ⑬ 菅元首相がワクチン接種を誤魔化したと誤情報
- ⑭ 国会乱闘中における捏造画像
- ⑮ 米大統領選、ハリス陣営が集会画像を生成?
- ⑯ 有名人かたる偽の広告“放置で被害”メタを全国5地裁に提訴
- ⑰ クリステア・フリーランドのディープフェイク動画とトロント・スターの偽記事
- ⑱ 巨大な剣が発見されたという偽画像
- ⑲ GTA 5のゲームプレイ映像であり、ブラジルの飛行機事故ではない
- ⑳ 金閣寺の金箔を剥がしたという偽情報

調査事例(主にディープフェイクに係るもの)

対策ニーズ大

- ① 中国で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 米国防総省近くで爆発偽画像をSNS上で拡散
- ④ 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 架空の制度を承認をする米大統領のフェイク動画
- 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 マウイ島の山火事が人為的な災害と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の偽情報が多発
- 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

対策ニーズ小

- ⑬ 大手食品会社に対し「ワクチン入りの食品を販売予定」との偽画像が拡散
- 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
- ⑮ 国際イベントにて、テスラの商品の欠陥を示唆するフェイク動画が放映
- ⑯ 著名シンガーのわいせつ画像がソーシャルメディアで拡散
- ⑰ 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
- ⑱ 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
- ⑲ フェイク音声による振り込め詐欺事件
- ⑳ 著名人の顔でDFを作成できるアプリへの訴訟
- ㉑ ディープフェイク技術で別人になりすまし、面接を受験

ユースケース例

- ✓ 災害時に、生成AIを用いて被災地の画像を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした
- ✓ 戦時下や暴動発生時において、生成AIにて動画を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした
- ✓ 選挙時に、候補者に関するSNSの偽文章を生成し、投票権を有する国民の興味を引き拡散を促すことで、特定地域の市民の支持率・政治情勢に影響を与えた
- ✓ 首相や大統領の偽音声を生成し、SNSで拡散することで、国民の支持率や政治的な混乱を促した
- ✓ 国民の関心が高い時事に関連し、企業・研究所・店舗に対する偽画像を生成し、SNSで国民の不安を煽り拡散を促すことで、国民からの信用を失墜させた
- ✓ 個人の利益目的のために偽動画を生成し、SNSで拡散するなどして不正行為を行った

④ 被害の深刻度、影響範囲に基づいたユースケースへの対応

✓ 各技術の進捗をもとに、行政が取り組むべきユースケースと、その中で優先的に取り組むべきユースケースを整理した。

ユースケース例

✓ 災害時に、生成AIを用いて被災地の画像を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした

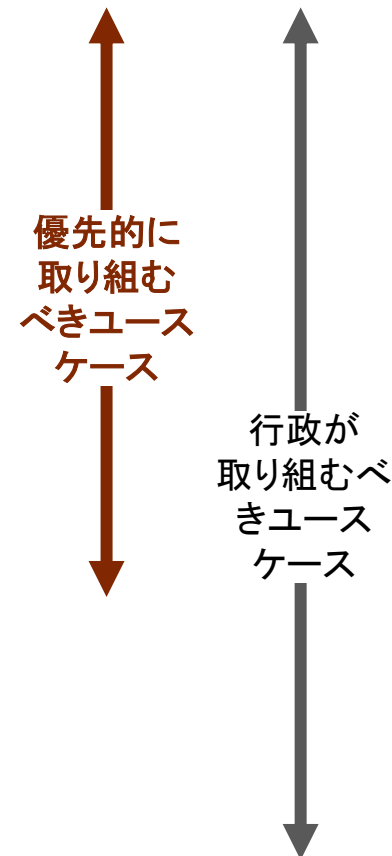
✓ 戦時下や暴動発生時において、生成AIにて動画を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした

✓ 選挙時に、候補者に関するSNSの偽文章を生成し、投票権を有する国民の興味を引き拡散を促すことで、特定地域の市民の支持率・政治情勢に影響を与えた

✓ 首相や大統領の偽音声を生成し、SNSで拡散することで、国民の支持率や政治的な混乱を促した

✓ 国民の関心が高い時事に関連し、企業・研究所・店舗に対する偽画像を生成し、SNSで国民の不安を煽り拡散を促すことで、国民からの信用を失墜させた

✓ 個人の利益目的のために偽動画を生成し、SNSで拡散するなどして不正行為を行った



目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-1-1. 事前調査の全体像

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 ア 直近5年で発生している具体的被害事例等

2-1-2. 偽・誤情報の流通・拡散に関連する最新の状況 イ 今後の発生が想定され得る被害・影響

2-1-3. 偽・誤情報の流通・拡散に関連する技術 ア 最新の関連技術動向

2-1-3. 偽・誤情報の流通・拡散に関連する技術 イ 技術に関する制度対応等、最新の関連技術に関する制度対応

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 ア 直近5年の被害事例等を踏まえた被害状況の整理

2-1-4. 今後、偽・誤情報対策を優先的に実施すべき状況 イ 技術的に対処すべきユースケースの設定

2-1-5. 2-1-4で優先実施すべきとされた状況に対する技術的対応策

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

① 優先的に実施すべきとする技術的対応策導出の考え方

✓ 「過去の被害事例をもとに整理した対処すべきユースケース」と、「日本・グローバルの調査結果等から注力すべき技術」等の観点から、有効と考えられる技術的対応策を特定する。

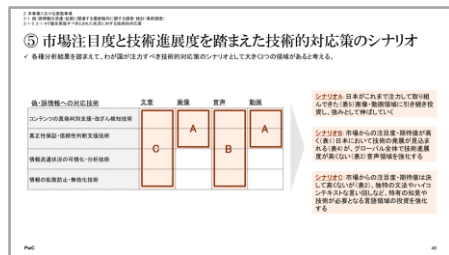
直近5年の偽・誤情報を踏まえ、対処すべきユースケースは何か

過去5年にて発生している被害事例をもとに、被害の深刻度及び影響範囲より対処すべきユースケースを特定。



日本・グローバルから注力すべき技術は何か

グローバルの投資・特許データ等をもとに「市場からの注目度」や日本・グローバルの特許数から注力すべき技術を導出。



有効と考えられる技術的対応策の特定

ユースケースと注力すべき技術より、技術的対応策を特定

⑥ ユースケースを踏まえた技術的対応策のシナリオの優先度

- ✓ 深刻度が大きく影響範囲の広い事例を、流布された情報の種別(モーダル)で分類したところ、特に画像・動画を使用するケースが多かった。
- ✓ 前項の技術的対応策のシナリオのうち、シナリオAを優先的に、次いでシナリオB、シナリオCに取り組みむべきと考えられる。

| 事例分析結果 | 深刻度・影響範囲 | 代表的なユースケース | 事例数 | | | |
|-------------|----------|---|-----|----|----|----|
| | | | 文章 | 画像 | 音声 | 動画 |
| [事例分析結果の図表] | 高・高 | <ul style="list-style-type: none"> ✓ 災害時に、生成AIを用いて被災地の画像を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした。 ✓ 戦時下や暴動発生時において、生成AIにて動画を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした。 | 2 | 5 | 3 | 7 |
| | 中・高 | <ul style="list-style-type: none"> ✓ 選挙時に、候補者に關するSNSの偽文章を生成し、投票権を有する国民の興味を引き拡散を促すことで、特定地域の市民の支持率・政治情勢に影響を与えた。 | 4 | 6 | 3 | 4 |
| | 中・中 | <ul style="list-style-type: none"> ✓ 首相や大統領の偽音声を生成し、SNSで拡散することで、国民の支持率や政治的な混乱を促した。 | | | | |
| 合計 | | | 6 | 11 | 6 | 11 |

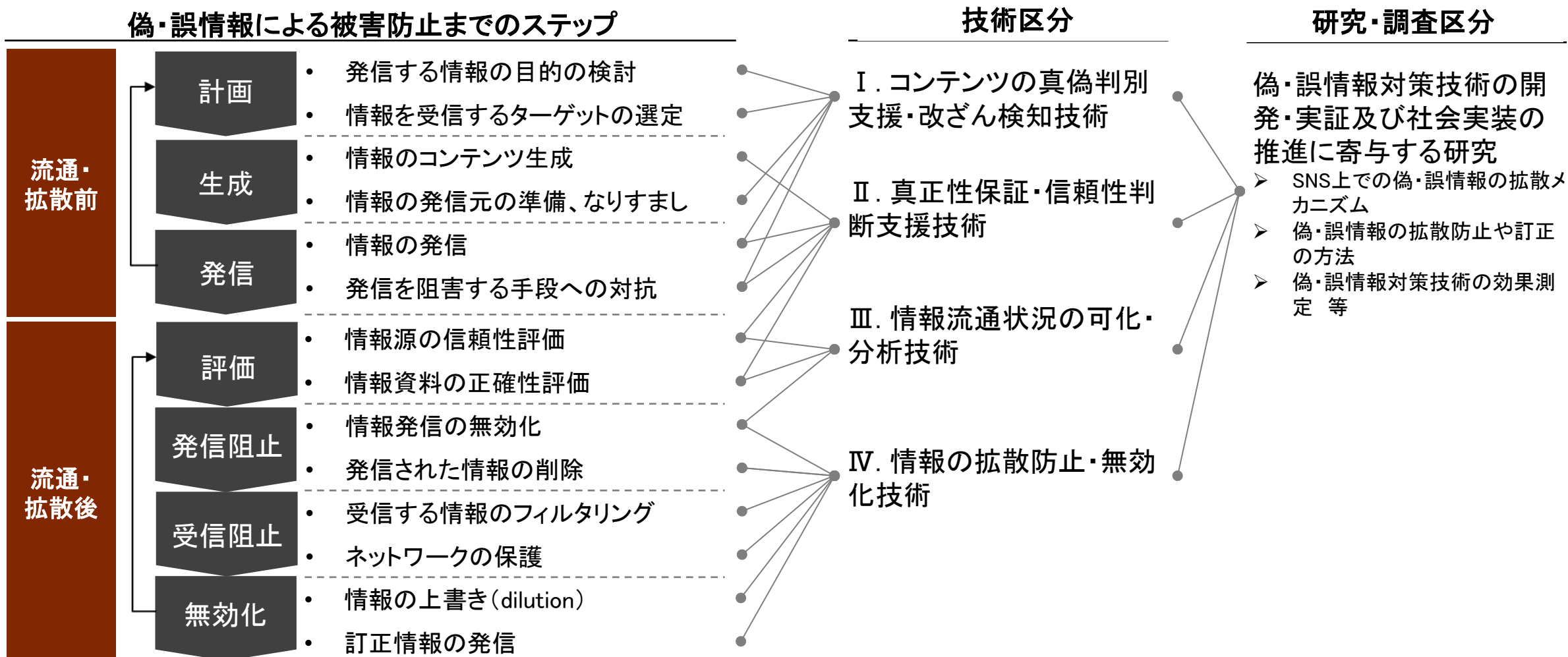
対応する技術的対応策のシナリオ

| | | | |
|---|---|---|---|
| C | A | B | A |
|---|---|---|---|

より広範囲に影響を及ぼす、深刻な被害を防止するためには、シナリオAが特に有効と考えられる。

② 偽・誤情報の流通・拡散に関する区分

- ✓ 本調査で整理した技術区分は、**偽・誤情報の流通・拡散から防止までのアプローチと対応**している。
- ✓ 全てのステップを網羅的にカバーする必要はなく、いずれかに対応する対策技術を適用することで、偽・誤情報による被害を抑止することが可能。
- ✓ 並行して偽・誤情報対策関連の調査研究を行うことが、技術開発・実証を推進させ、効果的な偽・誤情報対策につながる。



③ 偽・誤情報対策技術における各モーダルの定義

✓ 本パートにおいては、情報の表現形式であるモーダル種「文章」「画像」「音声」「動画」を以下の通り定義している。

| | 定義 | ユースケース例 |
|----|--|--|
| 文章 | 自然言語や記号によって表現された情報 文字列の集合として表現され、人間の言語に基づいて意味を持つ | ネット上の記事やニュース報道、様々な言説のファクトチェック ニュース記事とそれに対する反応を分析することで「疑義言説」を自動的に検知し、誤りの可能性を予測し、スコアリングする |
| 画像 | 視覚的な情報を静止画として表現したもの 2次元空間におけるピクセルの集合で、色情報を含む | 画像コンテンツの改ざん検知 コンテンツに対して人間の知覚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 |
| 音声 | 空間内の音の波形を時間軸に沿って記録したデータ 人間の耳に聞こえる範囲の空気の振動をデジタル化したもの | 詐欺電話の真偽判定 音響特徴量の分析や、音声認証・声紋分析などによって「フェイク(合成)であるか否か」「本人であるか否か」を判定・識別する |
| 動画 | 連続する画像と音声を含むメディア 時間軸に沿って変化する視覚情報を表現 | ソーシャルメディア上の動画コンテンツの真偽判定 動画内での時系列での不自然さを分析し、生成AIの生成物か否かの判別を行う |

※「音声」と「動画」との違い

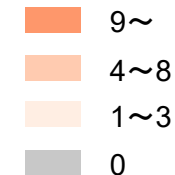
「音声」も「動画」も音を含んだモーダルだが、「音声」は詐欺電話やチャットボットのようなリアルタイムコミュニケーションの中でフェイクかどうかの判断が求められるケースがある点が特徴

④ 日本における偽・誤情報対応ソリューションの現状

- ✓ ピックアップした偽・誤情報関連技術の開発実施企業・研究機関が現状カバーしている技術分類をヒートマップ化。
- ✓ 画像・動画の真偽判別技術や真正性保証・信頼性判断技術を中心に研究開発・実用化が進んでいる。

日本における対応技術×ソリューション数

| 偽・誤情報への対応技術 | 文章 | 画像 | 音声 | 動画 |
|----------------------|------|-------|------|-------|
| コンテンツの真偽判別支援・改ざん検知技術 | 5(1) | 12(3) | 6(2) | 10(4) |
| 真正性保証・信頼性判断支援技術 | 4(0) | 9(3) | 6(2) | 8(3) |
| 情報流通状況の可視化・分析技術 | 7(6) | 2(0) | 0(0) | 0(0) |
| 情報の拡散防止・無効化技術 | 2(0) | 4(0) | 1(0) | 4(0) |



※()の数字は実用化されているソリューション数

画像・動画の真偽判別技術、真正性保証・信頼性判断ソリューションが多く出ている

⑤ 市場注目度と技術進展度を踏まえた技術的対応策のシナリオ

✓ 各種分析結果を踏まえて、わが国が注力すべき技術的対応策のシナリオとして大きく3つの領域があると考える。

| 偽・誤情報への対応技術 | 文章 | 画像 | 音声 | 動画 |
|----------------------|----|----|----|----|
| コンテンツの真偽判別支援・改ざん検知技術 | | A | | A |
| 真正性保証・信頼性判断支援技術 | | | B | |
| 情報流通状況の可視化・分析技術 | C | | | |
| 情報の拡散防止・無効化技術 | | | | |

シナリオA: 日本がこれまで注力して取り組んできた(表5)画像・動画領域に引き続き投資し、強みとして伸ばしていく

シナリオB: 市場からの注目度・期待値が高く(表1)日本において技術の発展が見込まれる(表4)が、グローバル全体で技術進展度が高くない(表2)音声領域を強化する

シナリオC: 市場からの注目度・期待値は決して高くないが(表2)、独特の文法やハイコンテキストな言い回しなど、特有の知見や技術が必要となる言語領域の投資を強化する

⑥ ユースケースを踏まえた技術的対応策のシナリオの優先度

- ✓ 深刻度が大きく影響範囲の広い事例を、流布された情報の種別(モーダル)で分類したところ、**特に画像・動画を使用するケースが多かった。**
- ✓ 前項の技術的対応策のシナリオのうち、シナリオAを優先的に、次いでシナリオB、シナリオCに取り組むべきと考えられる。

| 事例分析結果 | 深刻度・ 影響範囲 | 代表的なユースケース | 事例数 | | | |
|-----------------|--------------|---|-----|----|----|----|
| | | | 文章 | 画像 | 音声 | 動画 |
| | 高・高 | <ul style="list-style-type: none"> ✓ 災害時に、生成AIを用いて被災地の画像を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした ✓ 戦時下や暴動発生時において、生成AIにて動画を生成し、SNSで不安を煽り拡散を促すことで、国民の生命に脅威をもたらした | 2 | 5 | 3 | 7 |
| | 中・高 | <ul style="list-style-type: none"> ✓ 選挙時に、候補者に関するSNSの偽文章を生成し、投票権を有する国民の興味を引き拡散を促すことで、特定地域の市民の支持率・政治情勢に影響を与えた | 4 | 6 | 3 | 4 |
| | 中・中 | <ul style="list-style-type: none"> ✓ 首相や大統領の偽音声を生成し、SNSで拡散することで、国民の支持率や政治的な混乱を促した | 4 | 6 | 3 | 4 |
| 合計 | | | 6 | 11 | 6 | 11 |
| 対応する技術的対応策のシナリオ | | | C | A | B | A |

より広範囲に影響を及ぼす、深刻な被害を防止するためには、シナリオAが特に有効と考えられる。

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-2. 偽・誤情報への対策技術の開発・実証

2-2-1. 実証団体の公募・選定・運用に係る業務

2-2-2. 開発・実証の実施、実証団体に対する支援及び進捗管理

2-2-3. 経理処理等に係る業務

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

① 公募内容

- ✓ 偽・誤情報等への対策技術の開発を行う技術開発区分に加え、今年度は偽・誤情報等への対策技術に係る研究を行う研究・調査区分を設け公募を行った。具体的には、偽・誤情報の流通・拡散に関連する最新動向に関する調査結果を踏まえ、技術開発区分では、5の対策技術を対象技術とし、研究・調査区分はそれらの対策技術に係る研究を対象とした。
- ✓ **技術開発区分の採択数は12-15件程度、研究・調査区分は5件程度とし、採択合計額は15億円以内とした。**

公募対象とした技術開発、研究・調査内容

- 今年度の開発・実証事業の公募では、偽・誤情報等への対策技術の開発を行う技術開発区分に加え、偽・誤情報等への対策技術に係る研究を行う研究・調査区分を設け、公募を行った。

【技術開発区分】

1. コンテンツの真偽判別支援技術
2. 真正性保証・信頼性判断支援・改ざん検知技術
3. 情報流通状況の可視化・分析技術
4. 偽・誤情報の拡散を防止する技術
5. 偽・誤情報を無効化する技術

【研究・調査区分】

1. 偽・誤情報対策技術に係る研究
(研究テーマ例)
 - SNS上での偽・誤情報の拡散メカニズムに関する定量的・心理学的な分析
 - 偽・誤情報の拡散防止や訂正の方法等に関する定量的・心理学的分析
 - 偽・誤情報対策技術の効果測定に関する定量的・心理学的分析 等

採択予定数や採択上限金額、スケジュール

- 技術開発区分の採択数は12-15件程度、研究・調査区分は5件程度、採択合計額は15億円以内とし、各区分での上限金額を以下のとおり設定した。

【技術開発区分】

- 技術開発区分は1件あたり上限を5億円(税込)。

【研究・調査区分】

- 研究・調査区分は1件あたり上限を1千万円(税込)。

※ 技術開発主体、研究・調査主体が提出する資金計画書の妥当性等も踏まえて、支弁する金額を決定することとした。

- 開発・実証スケジュールは以下のとおり。

スケジュール

令和7年4月～5月頃 公募
令和7年6月頃 外部有識者による評価及び採択案件の発表
令和8年2月頃 成果報告書案等の提出
令和8年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。

② 実施要領

✓ 公募にあたり、事業目的や公募内容、業務内容等をまとめた実施要領を作成した。

実施要領※¹ (抜粋)

総務省

インターネット上の偽・誤情報等への
対策技術の開発・実証事業

実施要領

【応募受付期間】
令和7年4月22日(火)～同年5月26日(月)正午

【事務局】
PwC コンサルティング 合同会社

1. 事業の目的
 多くの利用者が情報を収集・閲覧するプラットフォームサービス等のインターネット上では、偽・誤情報の流通・拡散に伴うリスクが急速に拡大しています。特に、災害時の情報収集や、民主主義の根幹である選挙等において、SNS（ソーシャルネットワークサービス）等を利用した情報発信・収集が活発化する中で、AIを利用した精緻・巧妙な偽・誤情報（画像、映像、音声等）の流通・拡散に伴うリスクをはじめ、偽・誤情報の流通・拡散に伴うリスクが増大しています。偽・誤情報の流通・拡散により、利用者が多様な偽情報を正確に理解し適切な判断を下すことが困難になり、利用者が安心・信頼してデジタルサービスを利用することができなくなる可能性があります。

これらの状況を踏まえ、インターネット上の偽・誤情報の流通・拡散リスクへの対策に資するため、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ます。

また、偽・誤情報の対策技術の開発・実証及び社会実装の推進を目的とした研究を実施し、より迅速に利用者が安心・信頼してデジタルサービスを活用することを目的としています。

以上の目的等のため、総務省では、インターネット上の偽・誤情報への対策技術の開発・実証を実施することとし、必要となる調査研究の事務局としてPwC コンサルティング 合同会社（以下「事務局」という。）を選定しました。このたび、総務省では、インターネット上の偽・誤情報への対策技術の開発・実証の実施を希望する団体を一般に広く募ることとしました。

2. 公募内容等
(1) 公募内容
 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（以下「本開発・実証」という。）

本開発・実証事業の対象とする技術開発区分と研究・調査区分としては具体的には以下を想定しています。技術開発区分の詳細は別添1を参照してください。

【技術開発区分】

- コンテンツの真偽別支援技術
- 真正性保証、信頼性判断支援、改ざん検知技術
- 情報流通状況の可視化・分析技術
- 偽・誤情報の拡散を防止する技術
- 偽・誤情報を無効化する技術

【研究・調査区分】

- 偽・誤情報対策技術に係る研究
 - SNS 上での偽・誤情報の拡散メカニズムに関する定量的・心理学的な分析
 - 偽・誤情報の拡散防止や訂正の方法等に関する定量的・心理学的分析
 - 偽・誤情報対策技術の効果測定に関する定量的・心理学的分析 等

提案事業が複数の技術開発区分に該当する場合には、提案内容における主たる技術開発区分を1つのみ選択ください。

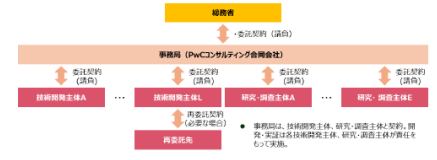
なお、提案事業が上記の区分のいずれに該当するかが不明な場合には、事前に事務局に相談、確認の上、応募してください。

(2) 業務委託（請負）費
 技術開発区分は1件あたり上限を5億円（税込）とします。
 研究・調査区分は1件あたり上限を1千万円（税込）とします。

※技術開発主体、研究・調査主体が提出する資金計画書の妥当性等も踏まえ、支弁する金額を決定します。

(3) 採択予定数、採択合計額
 技術開発区分の採択数は12-15件程度、研究・調査区分は5件程度、採択合計額は15億円以内を予定しています。

(4) 契約等の流れ
 事務局（PwC コンサルティング 合同会社）は、総務省の指示に基づき、技術開発主体、研究・調査主体と委託（請負）契約を締結します。



(5) 委託期間
 契約締結日から令和8年3月19日（木）まで

3. 業務内容
 本業務は「補助」ではなく、「委託」に基づくものです。業務委託費は調査研究の業務委託契約に基づく対価の性格を有する経費であって、補助金のような助成的性格のものとは異なりますので、この点、十分にご留意ください。

業務事項は技術開発区分と研究・調査区分で異なるため、提案する区分に応じた業務事項を確認してください。

(1) 業務事項【技術開発区分】
 採択を受けた技術開発主体は、①～⑤までの事項を実施してください。

① インターネット上の偽・誤情報等への対策技術の開発
 インターネット上の偽・誤情報等の対策技術について適切な開発手法を検討し、それを実行すること。

※ なお、企画提案書においては、5（1）に示す評価の観点を十分に踏まえ、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・誤情報等への対策技術の有効性等に関する検証及び調査
 インターネット上の偽・誤情報等への対策技術の開発に当たり、その有効性を定量的に検証するとともに、社会実証に向けた課題の解消に関する調査・検討等を行うこと。

なお、定量的な評価に当たっては、開発技術を検証するための指標だけでなく、社会実証に向けた効果検証を意識した指標を事務局と協議して設定した上で、適切な検証手法を検討すること。

※ 企画提案書において、5（1）に示す評価の観点を十分に踏まえ、具体的な実施内容を提案すること。

③ 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に向けた取組
 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に向けて、導入先となるユーザーや関係するステークホルダーへのヒアリング、ビジネスモデルの検討等を行い、適宜必要となる調査を実施すること。

※ なお、ビジネスモデルの検討に当たっては、国内での導入だけでなくグローバルでの展開や国際標準等の取組も含まれていることが望ましい。

※ 企画提案書において、5（1）に示す評価の観点を十分に踏まえ、具体的な実施内容を提案すること。

④ 成果報告書及び社会実装実施計画書の作成
 成果報告書については、上記①～③の実施内容や成果等について、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。詳細については採択後に指示する事項に従うこと。

社会実装実施計画については、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。詳細については採択後に指示する事項に従うこと。

⑤ 普及啓発活動への協力
 本開発・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、本開発・実証期間の終了後も含め、総務省が実施する本開発・実証成果の普及啓発活動に当たって、本開発・実証内容に関する資料提供等の協力を行うこと。

(2) 業務事項【研究・調査区分】
 採択を受けた研究・調査主体は、①～④までの事項を実施してください。

① インターネット上の偽・誤情報等への対策技術に係る研究の実施
 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、加速化させるために、研究を行うこと。

※ なお、企画提案書においては、5（1）に示す評価の観点を十分に踏まえ、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・誤情報等への有効性等に関する分析・調査
 インターネット上の偽・誤情報等に存する先進的な取組について、その有効性を定量的に分析するとともに、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、5（1）に示す評価の観点を十分に踏まえ、具体的な実施内容を提案すること。

③ 成果報告書の作成
 成果報告書については、上記①、②の実施内容や成果等について、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。詳細については採択後に指示する事項に従うこと。

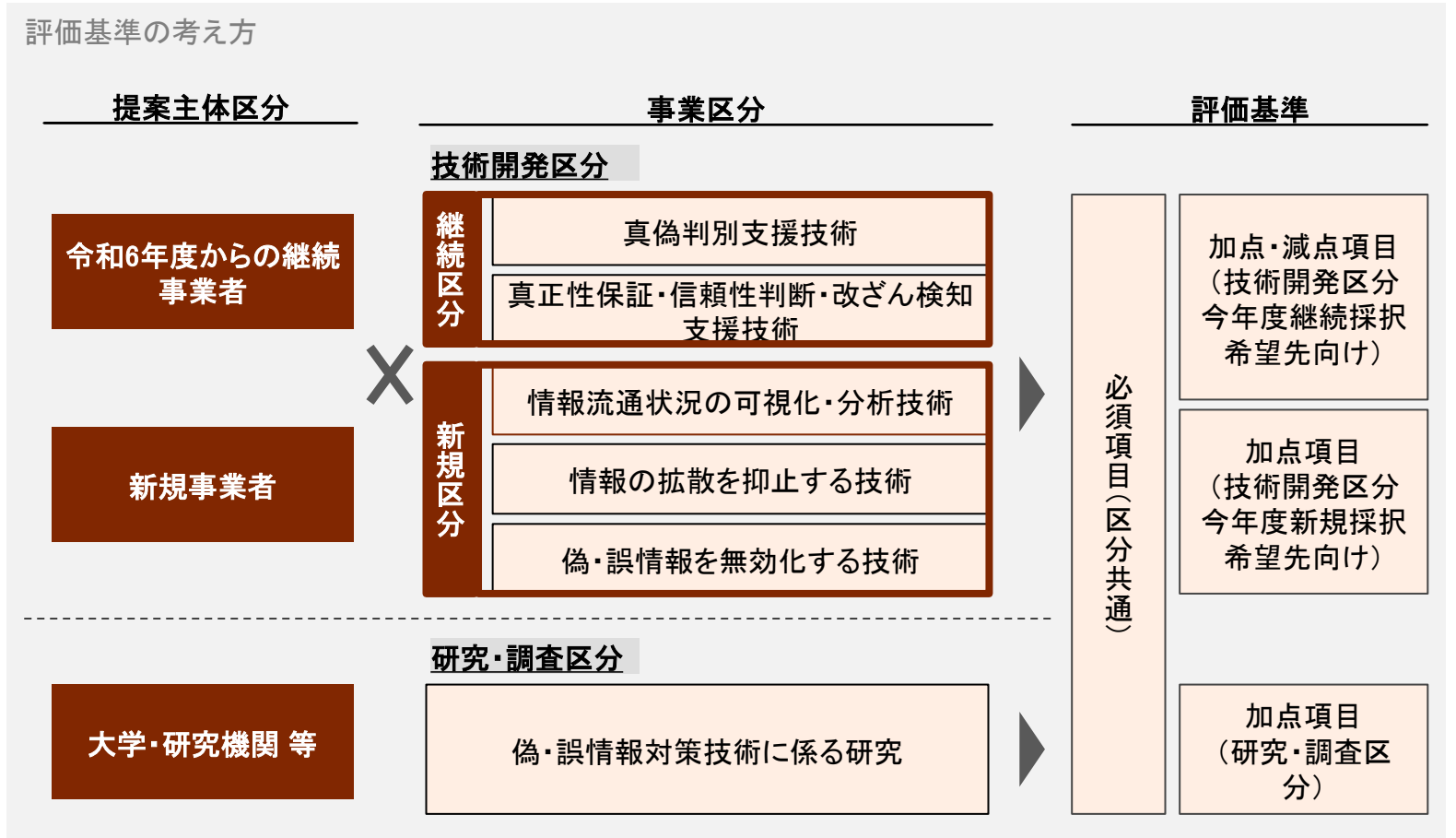
④ 普及啓発活動への協力
 本開発・実証成果について、メディア対応やイベント開催、学会参加等を通じて、積極的に普及啓発活動に取り組むこと。また、本開発・実証期間の終了後も含め、総務省が実施する

※1 実施要領は下記のリンク先で公開
https://www.soumu.go.jp/main_content/001006475.pdf

③ 評価基準の考え方

- ✓ 総合的に評価を行うため、有識者3名に評価委員の委嘱を行い、評価委員の意見を踏まえ、技術開発区分、研究・調査区分それぞれに評価基準を設けた。
- ✓ また、昨年度採択を受けた実証団体については、昨年度の成果を加味するため、今年度新規採択希望先向けの評価基準とは別に基準を設けた。

| # | 評価委員 | ご所属 |
|---|---------|---|
| 1 | 笹原 和俊 様 | 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 |
| 2 | 山口 真一 様 | 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 |
| 3 | 小林 宏彰 様 | 株式会社東京大学 エッジキャピタルパートナーズ パートナー |



④ 評価基準（必須項目）

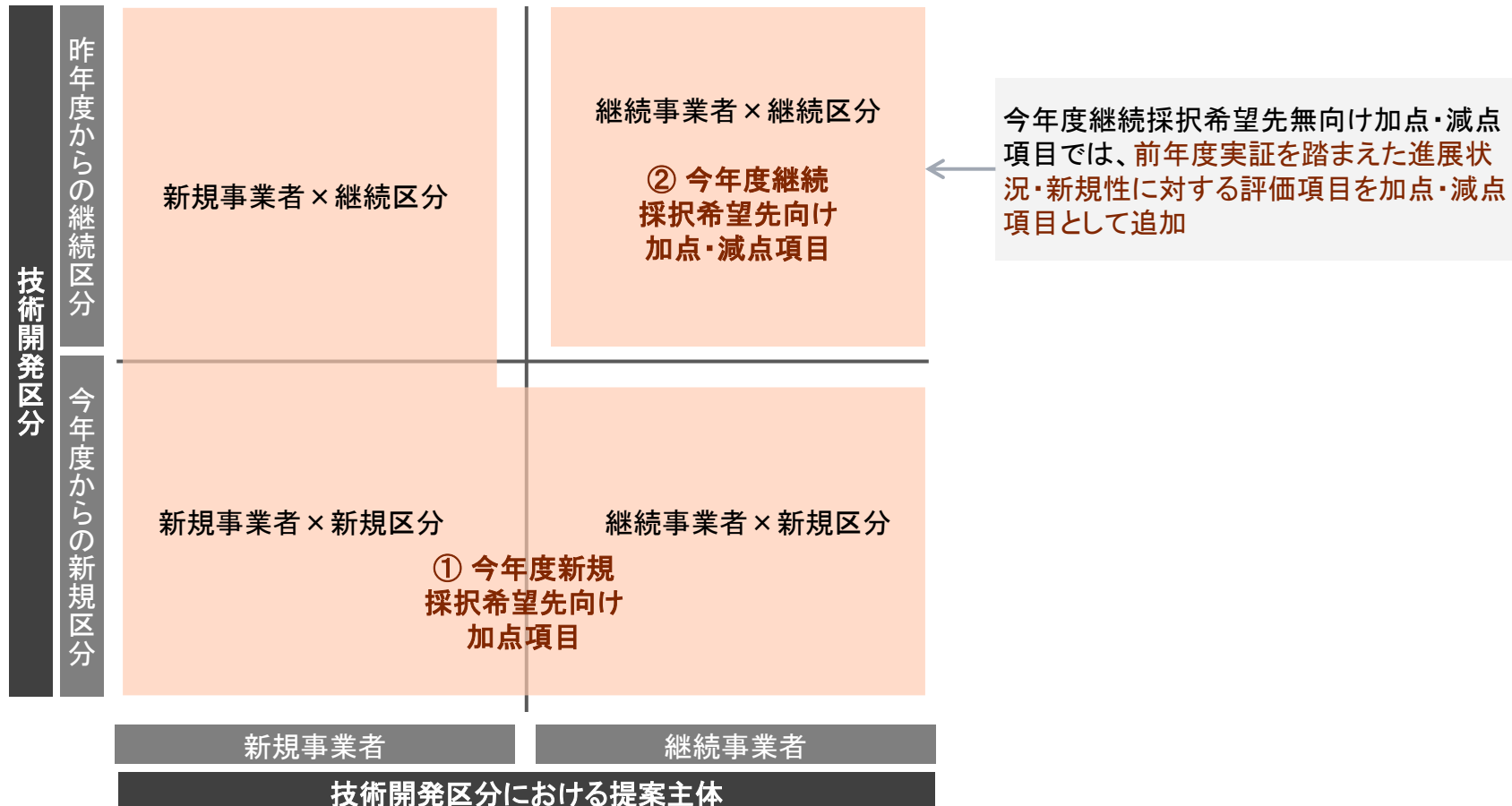
- ✓ 技術開発区分、研究・調査区分共通の必須項目としては、業務事項で定めた内容の網羅性、実施体制の事業構築能力、実施スケジュールの妥当性、情報セキュリティ対策の適切さを評価できる仕組みとした。
- ✓ なお、必須項目の要素が満たされていない事業者は審査対象としての基準を満たさないものとし、加点項目の評価にかかわらず採択候補先としないこととした。

| 分類 | 評価基準 |
|----------------------------|----------------|
| 必須項目 【技術開発区分、研究・調査区分共通】 | 業務事項で定めた内容の網羅性 |
| | 実施体制の事業遂行能力 |
| | 実施スケジュールの妥当性 |
| | 情報セキュリティ対策の適切さ |

- 必須項目の要素を満たさない事業者は審査対象外とし、加点項目の評価に関わらず、採択候補先としないこととした。

⑤ 技術開発区分の加点・減点項目の考え方

- ✓ 技術開発区分は昨年度から継続して募集する区分と今年度から新たに募集する区分の2つに分けられ、技術開発区分における提案主体も今年度から新規に採択を希望する事業者と、昨年度採択を受けており今年度も継続して採択を希望する事業者の2つに分けられる。
- ✓ このため、技術開発区分の評価にあたっては、① 今年度新規採択希望先向け評価基準と② 今年度継続採択希望先向け評価基準の2つを設けた。



⑥ 加点項目(技術開発区分 今年度新規採択希望先向け)

✓ 今年度新規採択を希望する事業者向けにはKPIとして定める観点、技術の進化・可用性、社会実装に向けた進展を加点項目として設けた。

| カテゴリー | | 観点 | 具体的な評価基準 |
|-------|-------------|----------------------------|--|
| 加点項目 | KPIとして定める観点 | 技術面でのKPIの観点 | <ul style="list-style-type: none"> 開発・実証の対象となる技術の達成度合いが定量的に評価できるようなKPIの観点が適切に定められていること(偽・誤情報の検知率、削減率等) |
| | | ビジネス面でのKPIの観点 | <ul style="list-style-type: none"> 社会実装を進める上でのビジネス視点・ユーザー視点での達成度合いが定量的に評価できるようなKPIの観点が適切に定められていること |
| | 技術の進化・可用性 | 期待成果の明確さ | <ul style="list-style-type: none"> 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分ができるか |
| | | 技術の横展開の可能性 | <ul style="list-style-type: none"> 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 |
| | | 技術が対処可能なユースケース/偽・誤情報のカバレッジ | <ul style="list-style-type: none"> 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 |
| | | 社会実装に向けた課題解決 | <ul style="list-style-type: none"> 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 |
| | 社会実装に向けた進展 | 社会実装に向けた体制 | <ul style="list-style-type: none"> 技術の実用化や利用の主体が開発・実証の体制に含まれるか 上記による開発・実証後の円滑な社会実装が見込まれるか |
| | | 社会実装に向けた拡がり | <ul style="list-style-type: none"> 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 開発・実証成果が、幅広い主体の利用の後押しになるか |
| | | ビジネスモデルの検討状況 | <ul style="list-style-type: none"> 社会実装を進める上での、ターゲット市場や顧客価値の明確化、ターゲット・ステークホルダー分析等が具体的に示しているか 費用対効果/事業採算性の検討等が適切に行えているか |
| | | | |

⑦ 加点・減点項目(技術開発区分 今年度継続採択希望先向け)

✓ 前年度採択を受けており今年度も継続採択を希望する事業者向けには、今年度新規採択希望先向けの加点項目に加え、前年度開発・実証を踏まえた進展状況及び新規性を加点・減点項目として設けた。

| カテゴリー | | 観点 | 具体的な評価基準 |
|---------|------------------------|----------------------------|--|
| 加点項目 | KPIとして定める観点 | 技術面でのKPIの観点 | ・ 開発・実証の対象となる技術の達成度合いが定量的に評価できるようなKPIの観点が適切に定められていること(偽・誤情報の検知率、削減率等) |
| | | ビジネス面でのKPIの観点 | ・ 社会実装を進める上でのビジネス視点・ユーザー視点での達成度合いが定量的に評価できるようなKPIの観点が適切に定められていること |
| | 技術の進化・可用性 | 期待成果の明確さ | ・ 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分ができるか |
| | | 技術の横展開の可能性 | ・ 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 |
| | | 技術が対処可能なユースケース/偽・誤情報のカバレッジ | ・ 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 |
| | | 社会実装に向けた進展 | ・ 社会実装に向けた課題に定める提案内容になっているか ・ 技術の実用化や利用の主体が開発・実証の体制に含まれるか ・ 上記による開発・実証後の円滑な社会実装が見込まれるか |
| | 社会実装に向けた進展 | 社会実装に向けた体制 | ・ 社会実装に向けた課題に定める提案内容になっているか ・ 技術の実用化や利用の主体が開発・実証の体制に含まれるか ・ 上記による開発・実証後の円滑な社会実装が見込まれるか |
| | | 社会実装に向けた体制 | ・ 技術の実用化や利用の主体が開発・実証の体制に含まれるか ・ 上記による開発・実証後の円滑な社会実装が見込まれるか |
| | | 社会実装に向けた拡がり | ・ 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 ・ 開発・実証成果が、幅広い主体の利用の後押しになるか |
| | | ビジネスモデルの検討状況 | ・ 社会実装を進める上での、ターゲット市場や顧客価値の明確化、ターゲット・ステークホルダー分析等が具体的に示しているか ・ 費用対効果/事業採算性の検討等が適切に行えているか |
| 加点・減点項目 | 前年度開発・実証を踏まえた進展状況及び新規性 | 前年度開発・実証の成果及び進展状況 | ・ 前年度開発・実証における成果品質(特許やプロダクト等) ・ 前年度成果を踏まえた現時点での新たな進展 |
| | | 前年度開発・実証を踏まえた新規性 | ・ 前年度開発・実証を踏まえた課題の具体的な改善案が示されているか ・ 今年度提案における新たな提案要素や独自性の有無 |

⑧ 加点項目(研究・調査区分)

✓ 研究・調査区分向けには、専門知識・適格性、研究内容、研究経験及び実績を加点項目として設けた。

| カテゴリー | | 観点 | 具体的な評価基準 |
|-------|----------|---------------------|--|
| 加点項目 | 専門知識・適格性 | 研究内容に関する専門性の有無、適格性 | <ul style="list-style-type: none"> 偽・誤情報への対策について、当該分野に関する社会動向、先行研究、制度等を理解するための学術的及び専門的知見を有しているか 偽・誤情報の動向に関する深い知見を有し、より効果的な調査手法を提案・実施する能力を有していること |
| | 研究内容 | 研究内容の妥当性、独創性 | <ul style="list-style-type: none"> 本実施要領記載の業務事項について全て提案されていること 研究・調査の抽出・分析手法が妥当であること(分析手法に事業成果を高めるための工夫があれば加点する) 研究において、独創性・先進性があること |
| | 研究経験及び実績 | 研究・調査主体の研究業務の経験及び実績 | <ul style="list-style-type: none"> 提案内容の研究業務に寄与する技術に関する研究・調査を実施した経験及び実績があるか 上記以外にインターネット上の偽・誤情報等への対策に寄与する技術に関する研究・調査を実施した経験及び実績があるか ※業務の実績内容により加点する。 |

- 2. 本事業における実施事項
 - 2-2. 偽・誤情報への対策技術の開発・実証
 - 2-2-1. 実証団体の公募・選定・運用に係る業務

⑨ 応募書類・記載例(1/2)

✓ 応募書類として、① 企画提案書等、② 企画提案書概要版、③ 業務の実施体制及び管理体制、④ 業務従事者名簿、⑤ 事業スケジュール、⑥ 資金計画書を作成するとともに、提案者の作成の助けとなるよう、記載例も作成した。

① 企画提案書等※1

① 企画提案書等※1

インターネット上の偽・誤情報等への対策技術の開発・実証事業 企画提案書

※本事業の実施体制及び管理体制あり。記載内容は申請書類に提出し審査対象とする。 提出日：令和7年 月 日

② 企画提案書概要版※1

研究の背景・目的や提案する取組及び期待される効果等について、図等を用いて自由に記載。

③ 業務の実施体制及び管理体制※1

本事業実施に係る情報取扱表

※記載内容は申請書類に提出し審査対象とする。

④ 業務従事者名簿※1

本事業実施に係る業務担当者名簿

※記載内容は申請書類に提出し審査対象とする。

② 企画提案書概要版※1

② 企画提案書概要版※1

技術的区分向け 実証件名

実証体制 (公募・競争型実証) 対社 (社(株)、D社(株)) 技術的区分

対象技術 > 実証で活用する対象技術について記載。

KPI > 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。※背景・課題(技術・社会実装における課題等)を記載。
> ○○(対策技術)を活用し、○○○○を克服。※課題解決のために提案する取組の概要。
> ○○○○の実現を図る。
※150-300文字程度で、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。

研究・調査区分向け 研究件名

実証体制 (公募・競争型実証) 対社(研究開発) 対社(社(株)、D社(株)) 対象とする偽・誤情報

研究概要 > ○○(対象とする偽・誤情報)には、○○○○という課題が存在。※背景・課題(技術・社会実装における課題等)を記載。
> ○○○○という研究を実施。※課題解決のために提案する取組の概要。
> ○○○○の実現を図る。
※150-300文字程度で、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。

研究の背景・目的や提案する取組及び期待される効果等について、図等を用いて自由に記載。

研究概要を表す図・イラスト等を記載

③ 業務の実施体制及び管理体制※1

③ 業務の実施体制及び管理体制※1

実証・研究件名 業務の実施体制及び管理体制

※以下は記載内容の例または説明であり、記載の内容が直まらるように記載すること。

業務の実施体制及び管理体制を示すこと。

情報管理責任者

情報管理責任者・経理責任者

業務従事者

※記載内容は申請書類に提出し審査対象とする。

※以下は記載内容の例または説明であり、記載の内容が直まらるように記載すること。

情報管理責任者

情報管理責任者・経理責任者

業務従事者

※記載内容は申請書類に提出し審査対象とする。

※1 応募書類は下記のリンク先で公開
https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/taisakugijutsu_fy2025.html

- 2. 本事業における実施事項
 - 2-2. 偽・誤情報への対策技術の開発・実証
 - 2-2-1. 実証団体の公募・選定・運用に係る業務

⑩ 公募用ウェブページ

✓ 公募にあたり、総務省ウェブページに本開発・実証事業の概要や公募概要、公募結果、担当窓口に関するウェブページを作成した。

インターネット上の違法・有害情報に対する対応(情報流通プラットフォーム対処法)

- ▶ [インターネット上の違法・有害情報に対する対応\(情報流通プラットフォーム対処法\)](#)
- ▶ [インターネット上の違法・有害情報に関してお困りの方へ](#)
- ▶ [プロバイダ\(サイト管理者等\)の方へ](#)
- ▶ [プロバイダ責任制限法Q&A](#)
- ▶ [インターネット上のフェイクニュースや偽情報への対策](#)
- ▶ [インターネット上の偽・誤情報等への対策技術の開発・実証事業](#)
- ▶ [インターネット上の誹謗中傷への対策](#)

インターネット上の偽・誤情報等への対策技術の開発・実証事業(令和7年度)

1 事業の概要

総務省は、生成AIに関連する偽・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・拡散に対応するため、「インターネット上の偽・誤情報等への対策技術の開発・実証事業」を通じ、対策技術の開発・実証及び社会実装を推進することとしています。

本件は、事業者や研究機関等を対象に、事務局(PwCコンサルティング合同会社)を通じてインターネット上の偽・誤情報等への対策技術の開発・実証事業の公募を行うものです。

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・誤情報等への対策技術の開発・実証事業

(2) 提案主体
事業者、研究機関等

(3) 全体の流れとスケジュール
公募は令和7年4月22日(水)～6月28日(月)12:00に実施します。
採択案件は令和7年6月下旬頃に公表予定です。

公募
令和7年4月22日～
5月26日

評価
令和7年5月下旬～
6月中旬

採択案件発表
令和7年6月下旬頃

実施
令和7年6月下旬～
令和8年2月頃

成果報告会
令和8年3月頃

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・誤情報等への対策技術の開発・実証事業」実施要領](#)

(4) 応募方法
下記リンクから様式をダウンロードして企画提案書等を作成し、実施要領に記載された提出先で一斉を提出してください。

[「インターネット上の偽・誤情報等への対策技術の開発・実証事業」応募資料一式](#)

(5) 公募説明会の実施
開発・実証の概要や公募に関する説明会をオンラインにて開催いたします。参加を希望の方は以下の申込用フォームからお申し込みください。詳細は実施要領を参照してください。

① 開催日時
令和7年5月1日(木)13:00～14:00

② 参加申込期限
令和7年4月30日(水)14:00

③ 参加申込方法
以下の事務局による申請フォームより必須項目をご入力ください。お申し込みいただいたアドレス宛に説明会前日までには採用のURLを送付いたします。
[公募説明会申込用フォーム](#)

(6) 選定手続
外部有識者による評価の結果に基づいて総務省が選定します。

3 公募結果

(1) 決定内容
公募の結果、技術開発主体に21者(22件)、研究・調査主体に6者(6件)の合計27者(28件)の提案があったところ、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、以下のとおり、技術開発主体14者、研究・調査主体6者の事業の採択を決定しました。

【技術開発主体】

| No | 技術開発主体 | 事業概要 |
|----|---------------------------|---|
| 1 | 一般社団法人 コード・フォー・ジャパン | SNSにおける偽情報・真偽不明情報の市民参加型可視化・分析技術の開発・実証 |
| 2 | 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・実証 |
| 3 | エヌ・ティ・ティ・コミュニケーションズ株式会社 | 情報の真正性を可視化するO2PA技術を活用した偽・誤情報対策技術の開発・実証 |
| 4 | Originator Profile 技術研究組合 | 発信者の真正性を担保する「Originator Profile」技術の開発・社会実装 |
| 5 | 株式会社Classroom Adventure | 偽・誤情報サンドボックスを活用した実践的ゲーム型プレバンキング技術の開発・実証 |
| 6 | 株式会社 コンテラセキュリティジャパン | 偽・誤情報およびデバンキング情報拡散のシミュレーション技術の開発・実証 |
| 7 | 株式会社TDW Lab | デジタル情報空間における多層的意味解析と拡散ダイミクス解明プラットフォームの開発・実証 |
| 8 | 株式会社データグリッド | SNSユーザー支援を中核とした偽・誤情報対策技術の開発・実証 |
| 9 | 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・誤情報対策技術の開発・実証 |
| 10 | SEARCHLIGHT株式会社 | アトリーミング動画コンテンツの真偽検証支援ツールの開発・実証 |
| 11 | Sakana AI株式会社 | 画像・動画を中心としたSNS上の投稿の真偽判定システムの開発・実証 |
| 12 | サン電子株式会社 | 多元統合型偽・誤情報検出技術の開発・実証 |
| 13 | NAELAS株式会社 | 電話音声フェイク検知および自治体向け偽・誤情報総合対策技術の開発・実証 |
| 14 | 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 |

【研究・調査主体】

| No | 技術開発主体 | 事業概要 |
|----|-----------------|--|
| 1 | 株式会社新領域安全保障研究所 | グローバル・メタアナリシスと国内実証による対策技術の有効性の研究・調査 |
| 2 | 中央大学 | 偽・誤情報の判別に知識および認知的熟慮が及ぼす影響に関する研究・調査 |
| 3 | 東京大学大学院 情報学環 | 生成AI時代における偽情報流通と認知特性の解明に関する研究・調査 |
| 4 | 東京大学大学院工学系研究科 | 偽・誤情報の拡散を抑制するためのSNSにおけるシェア行動プロセス可視化と信頼性を評価する表示の研究・調査 |
| 5 | 名古屋工業大学 | 誤情報対策における訂正のリスク・心理的メカニズムの研究・調査 |
| 6 | 日本エンタープライズ株式会社 | 制作者意図と異なる作品解釈に係る情報の急激な増加の影響等に関する研究・調査 |

4 担当窓口

技術開発主体を取りまとめる事務局(PwCコンサルティング合同会社)
E-mail: jp_cons_kobo_niseqciyocho_atmark_pwc.com
迷惑メール対策のため、「@」を「_atmark_」と表示しております。送信の際には「@」に変更してください。

⑪ 公募説明会

✓ 公募に関する説明や提案予定者からの質疑を受け付けるために、公募期間中に公募説明会を開催した。

(5) 公募説明会の実施

開発・実証の概要や公募に関する説明会をオンラインにて開催いたします。参加をご希望の方は以下の申込用フォームからお申し込みください。詳細は実施要領を参照してください。

<1> 開催日時

令和7年5月1日(木) 13:00～14:00

<2> 参加申込期限

令和7年4月30日(水) 14:00

<3> 参加申込方法

以下の事務局による申請フォームより必須項目をご入力ください。お申し込まいただいたアドレス宛に説明会前日までに視聴用のURLを送付いたします。

[公募説明会申込用フォーム](#)

⑫ 評価委員会

✓ 公募締切後、提案内容の評価を行うために、評価委員3名からなる評価委員会を開催。評価委員により評価が行われ、とりまとめを行った。

| # | 評価委員 | ご所属 |
|---|---------|---|
| 1 | 笹原 和俊 様 | 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 |
| 2 | 山口 真一 様 | 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 |
| 3 | 小林 宏彰 様 | 株式会社東京大学 エッジキャピタルパートナーズ パートナー |

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・誤情報等への対策技術の開発・実証事業

(2)提案主体

事業者、研究機関等

(3)全体の流れとスケジュール

公募は令和7年4月22日(火)～5月26日(月)12:00に実施します。

採択案件は令和7年6月下旬に公表予定です。



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・誤情報等への対策技術の開発・実証事業」実施要領](#)

⑬ 採択先一覧(技術開発区分)

✓ 評価委員会において、**技術開発主体の採択先として14事業者が選定**された。

凡例: 該当する技術区分※1

- ① コンテンツの真偽判別支援技術、改ざん検知技術
- ② 真正性保証・信頼性判断支援技術
- ③ 情報流通状況の可視化・分析技術
- ④ 情報の拡散防止・無効化技術

| # | 技術開発主体名 | 開発・実証件名 | 該当する技術区分※1 | | | |
|----|---------------------------|--|------------|---|---|---|
| | | | ① | ② | ③ | ④ |
| 1 | 一般社団法人コード・フォー・ジャパン | SNSにおける偽情報・真偽不明情報の市民参加型可視化・分析技術の開発・実証 | | | ○ | |
| 2 | 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・実証 | ○ | ○ | | |
| 3 | NTTドコモビジネス株式会社 | 情報の真正性を可視化するC2PA技術を活用した偽・誤情報対策の開発・実証 | ○ | ○ | | |
| 4 | Originator Profile 技術研究組合 | 発信者の真正性を担保する「Originator Profile」技術の開発・実証 | ○ | ○ | | |
| 5 | 株式会社Classroom Adventure | 偽・誤情報サンドボックスを活用した実践的ゲーム型プレバンキング技術の開発・実証 | | | | ○ |
| 6 | 株式会社コンステラセキュリティジャパン | 偽・誤情報及びデバンキング情報拡散のシミュレーション技術の開発・実証 | | | ○ | ○ |
| 7 | 株式会社TDAI Lab | デジタル情報空間における多層的意味解析と拡散ダイナミクス解明プラットフォームの開発・実証 | | | ○ | |
| 8 | 株式会社データグリッド | SNSユーザー支援を中核とした偽・誤情報対策の開発・実証 | ○ | | | |
| 9 | 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・誤情報対策技術の開発・実証 | | ○ | | ○ |
| 10 | SEARCHLIGHT株式会社 | ストリーミング動画コンテンツの真偽検証支援ツールの開発・実証 | ○ | | | |
| 11 | Sakana AI株式会社 | 画像・動画を中心としたSNS上の投稿の真偽判定システムの開発・実証 | ○ | | | |
| 12 | サン電子株式会社 | 多元統合型偽・誤情報検出技術の開発・実証 | | ○ | | |
| 13 | NABLAS株式会社 | 電話音声フェイク検知及び自治体向け偽・誤情報総合対策の開発・実証 | ○ | ○ | | |
| 14 | 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 | ○ | | | |

※1 採択先の選定を行う過程で、技術区分を4区分に再整理した。表中の該当する技術区分は各社の取組内容を踏まえた整理である。

⑭ 採択先一覧(研究・調査区分)

✓ 評価委員会において、研究・調査主体の採択先として6事業者が選定された。

| # | 研究・調査主体名 | 研究・調査件名 |
|---|----------------|--|
| 1 | 株式会社新領域安全保障研究所 | グローバル・メタアナリシスと国内実証による対策技術の有効性の研究・調査 |
| 2 | 中央大学 | 偽・誤情報の判別に知識及び認知的熟慮が及ぼす影響に関する研究・調査 |
| 3 | 東京大学大学院情報学環 | 生成AI時代における偽誤情報流通と認知特性の解明に関する研究・調査 |
| 4 | 東京大学大学院工学系研究科 | 偽・誤情報の拡散を抑制するためのSNSにおけるシェア行動プロセス可視化と信頼性を評価する表示の研究・調査 |
| 5 | 名古屋工業大学 | 誤情報対策における訂正のリスク:心理的メカニズムの研究・調査 |
| 6 | 日本エンタープライズ株式会社 | 制作者意図と異なる作品解釈に係る情報の急激な増加の影響等に関する研究・調査 |

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-2. 偽・誤情報への対策技術の開発・実証

2-2-1. 実証団体の公募・選定・運用に係る業務

2-2-2. 開発・実証の実施、実証団体に対する支援及び進捗管理

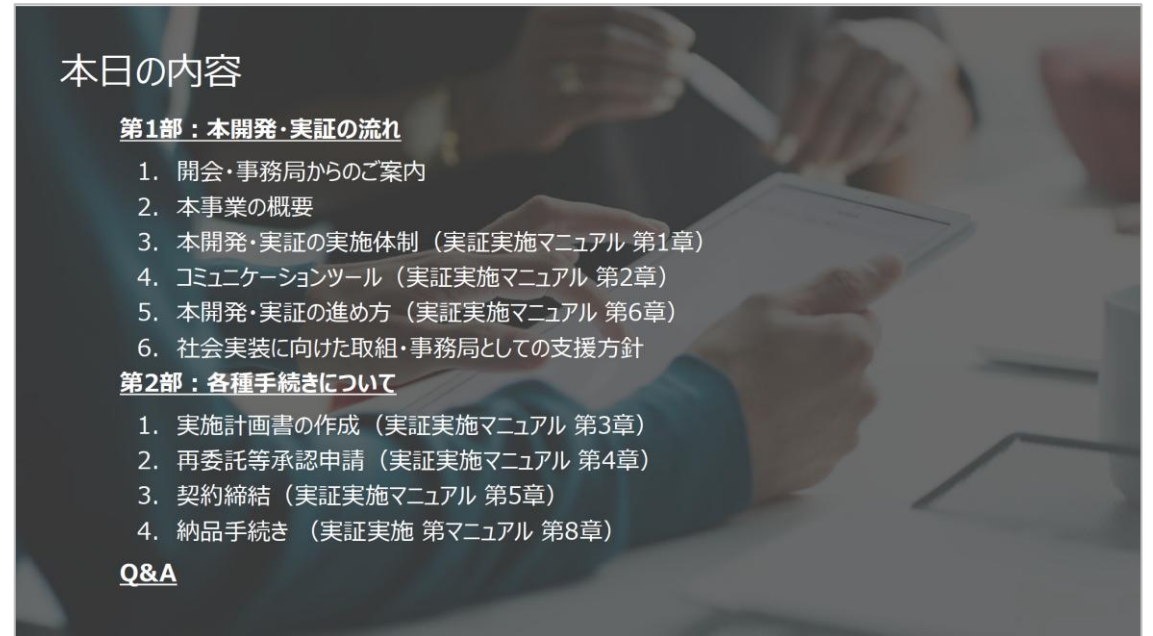
2-2-3. 経理処理等に係る業務

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

② 事業説明会

- ✓ 本事業に参画する実証団体に対して、本開発・実証の進め方に関する疑問点・不明点を解消し、円滑な事業推進を図ることを目的として、事業説明会を開催した。



③ 本事業における各評価指標

- ✓ 実施計画書に定めたKPI及びTRL/ARL等の評価指標の運用方針を策定のうえ、進捗管理を実施した。
- ✓ 技術開発主体に対しては、KGIの進捗状況について、初期(8月)、中間(11月)、最終(2月)の各段階において定期的なアセスメントを実施し、進捗状況の可視化を行った。

| | 実証団体ごとの進捗状況 | 実証団体ごとのKPI | 実証団体ごとのKGI |
|----------------------|---|---|--|
| 評価指標 | 事業スケジュールを踏まえた遅延・課題有無 | 実施計画書記載のKPI | TRL/ARLスコア (及び3×3のマトリクス) |
| 実証団体/事務局間での作成資料 | 定期報告書、課題管理表、遅延管理表 | 定期打合せ資料・議事録 | TRL/ARL評価シート |
| 実証団体/事務局間でのコミュニケーション | <ul style="list-style-type: none"> 月次打合せにて確認 | <ul style="list-style-type: none"> 10~12月の月次打合せにて、達成見込みと現時点での取組状況についてヒアリング | <ul style="list-style-type: none"> 10月の月次打合せにて、達成見込みをヒアリング 11月~12月に中間アセスメントを実施 2月に最終アセスメントを実施 |

④ 社会実装の進捗状況を評価するKGI(TRL/ARL)の考え方

- ✓ 社会実装とは、本事業にて開発・実証された技術が、公共機関や企業等の現場において実際に導入されていること(PoCでの導入事例を含む)と定義した。
- ✓ 本評価では、技術の完成度と社会・市場での活用に向けた環境・体制整備の状況という2つの観点から、“技術成熟度指標(TRL)”及び“採用成熟度指標(ARL)”を設定した。そして、両指標を縦軸・横軸とした下記のマトリクスにより、取組の進捗状況を可視化し、社会実装に向けた目標の設定と個別支援に活用した。

社会実装の進捗状況を評価する指標

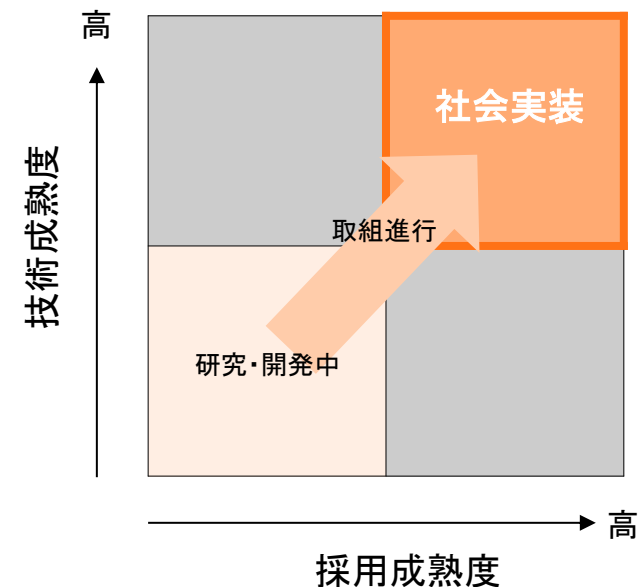
技術成熟度(TRL)

- 実用化に向けた技術の成熟度を評価する指標
- 技術そのものの完成度を評価する開発・統合・検証・安全性といった観点に加え、実用化に必要な市場性や事業化といった観点も含めて評価する

採用成熟度(ARL)

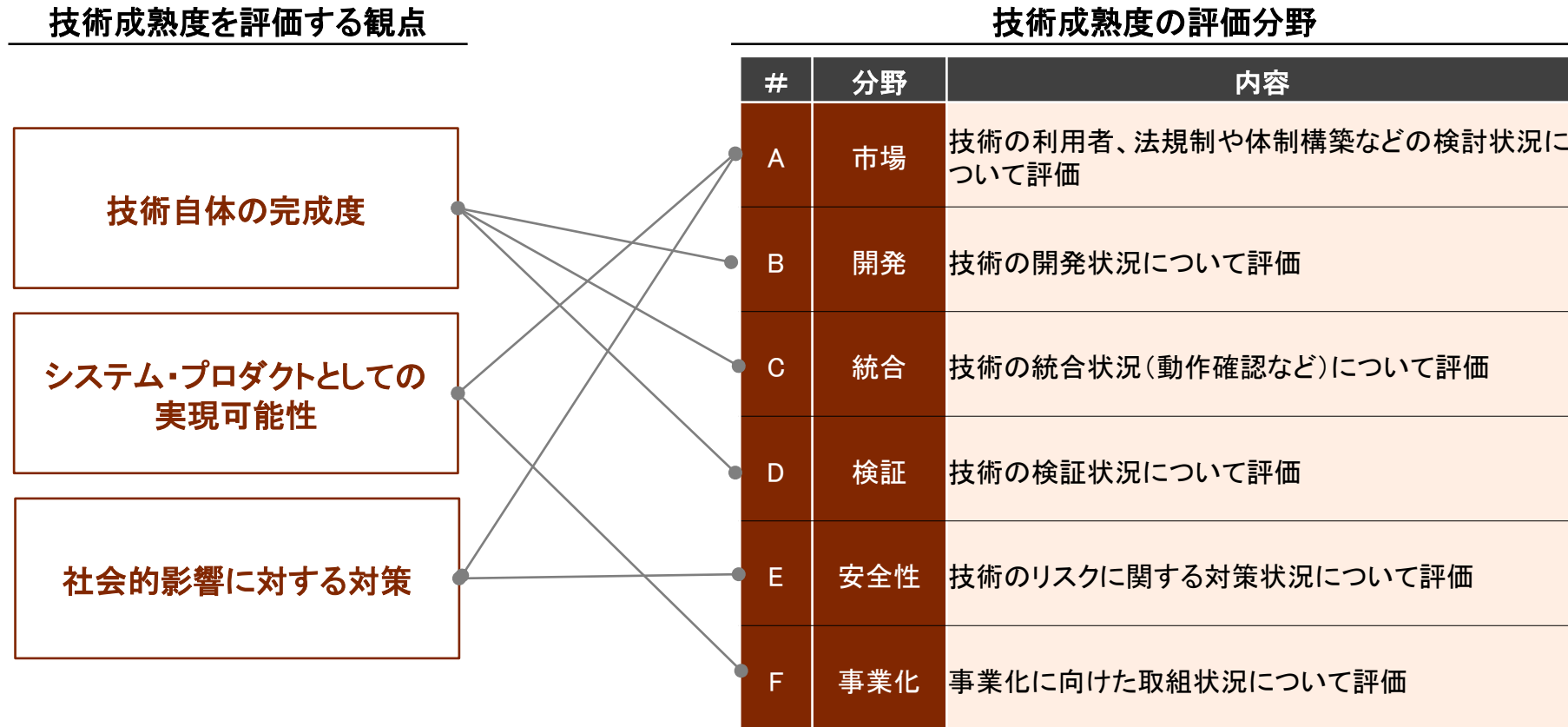
- ある技術が社会や市場にどの程度受け入れられているかを評価する指標
- 価値提案、市場での受け入れ状況、導入に必要なリソースの成熟度、社会的な許容性といった観点から、その技術の導入・活用に向けた環境・体制の整備状況を評価する

社会実装の進捗状況マトリクス



⑤ TRLとは

- ✓ TRL (Technology Readiness Level)とは技術開発の実用化に向けて、技術成熟度を測る指標である。
- ✓ 技術が“成熟している”と判断するためには、単に技術そのものが完成しているだけでは不十分であり、**システムやプロダクトとしての実現可能性や社会的影響に対する対策という観点から、総合的に評価する必要**がある。
- ✓ これらの観点を踏まえ、「A.市場」、「B.開発」、「C.統合」、「D.検証」、「E.安全性」、「F.事業化」の6つの評価分野に分類したものである。



⑥ ARLとは

- ✓ TRL (Technology Readiness Level) だけでは、**技術実証後の市場スケールにおける「社会受容性(理解や支持を得られているか)」や「競争力」の評価が困難**なため、社会実装を促進する目的でアメリカのエネルギー省 (DOE) がAdoption Readiness Level (ARL、開発技術の採用成熟度レベル) を採用。
- ✓ 評価においては、4分野「A.価値提案」、「B.市場受容性」、「C.リソース成熟度」、「D.社会許容性」にて評価。これを細分化し、本プロジェクト用にカスタマイズした計13の次元で評価するCommercial Adoption Readiness Assessment Tool (CARAT) を用いてリスクの程度を明らかにし、評価レベル「高」、「中」、「低」にて評価を行い、評価レベルが高いほど社会から支持が得られており、当該技術が継続的に利用されている状態を示す。

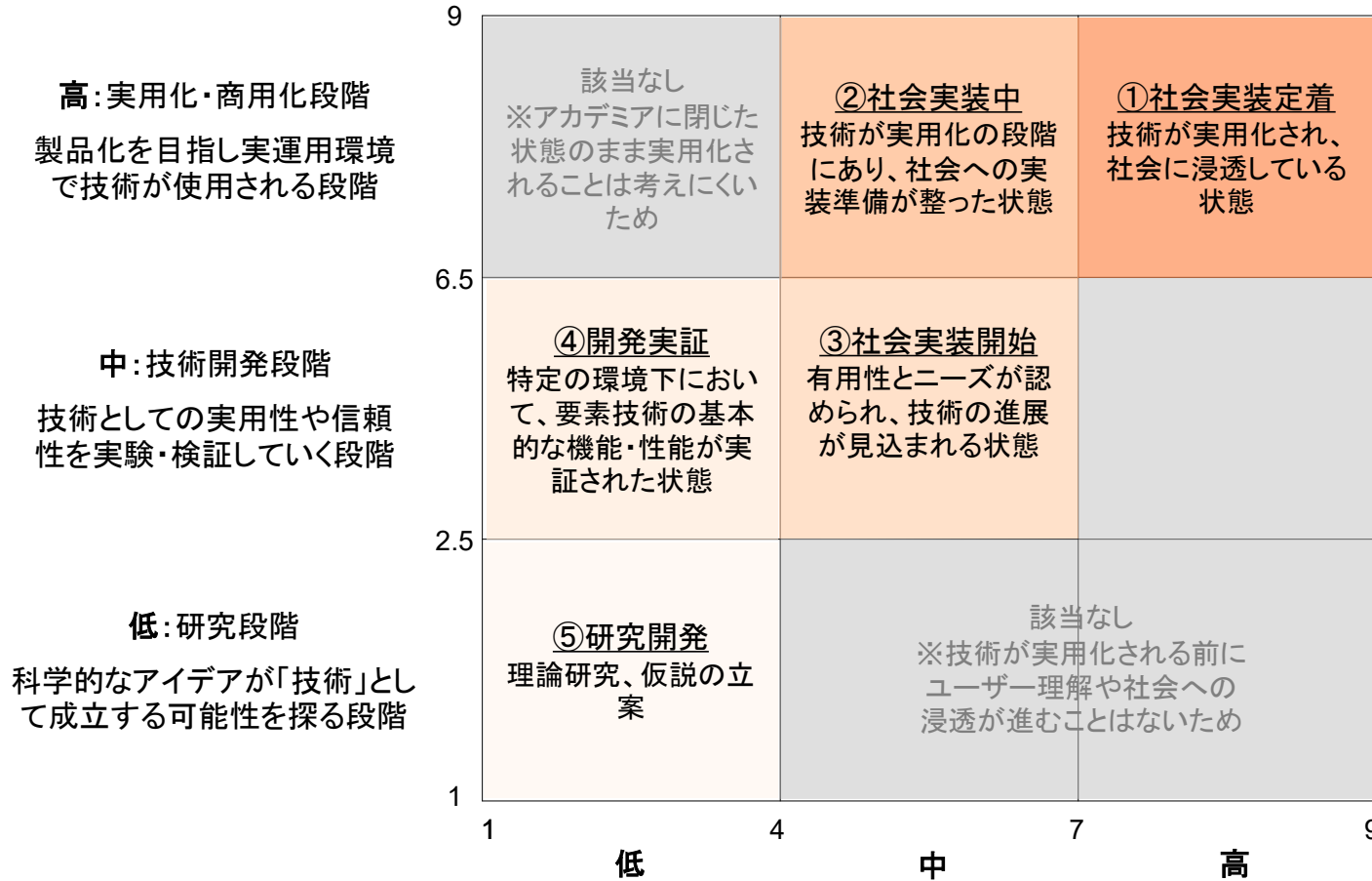
評価項目

| | | |
|-----------|--|----------------------------|
| A.価値提案 | 市場の需要を満たすために、顧客が支払ってもよいと考える価格で、新しいテクノロジーが市場で求められる機能を満たす能力の評価 | A-1.コスト競争力 |
| | | A-2.機能性 |
| | | A-3.ユーザーフレンドリー性 |
| B.市場受容性 | 競合他社、顧客、その他バリューチェーンプレーヤーを含む既存のプレーヤーによってもたらされるターゲット市場の需要特性とリスクの評価 | B-1.需要の確実性 |
| | | B-2.市場規模 |
| | | B-3.バリューチェーンに関するリスク |
| C.リソース成熟度 | 技術ソリューションを生み出すために必要な資材調達への妨げとなるリスクの評価 | C-1. 追加的資本投資の必要性 |
| | | C-2. プロジェクト方法論の成熟性 |
| | | C-3. システムインフラの成熟性 |
| | | C-4. 知識・知見を有する人材の調達に関するリスク |
| | | C-5. 保守・運用等の人的資本 |
| D.社会許容性 | 理解・支持を阻害し、技術の導入を妨げる可能性のある国や自治体による社会的からのリスクの評価 | D-1. 法令規制対応の必要性 |
| | | D-2. 地域社会の受容性 |

⑦ TRL×ARLマトリックスのエリア定義

✓ TRLとARLの指標を踏まえて、技術の状態を以下のようにマップ化した。

Technology Readiness Level (TRL):
 技術成熟度レベル



研究者だけに技術の存在や有用性が認知された状態

初期のユースケースに関わるユーザーに理解され、有用性が認められた状態

広く社会で受容され継続的に生産・利用されている状態

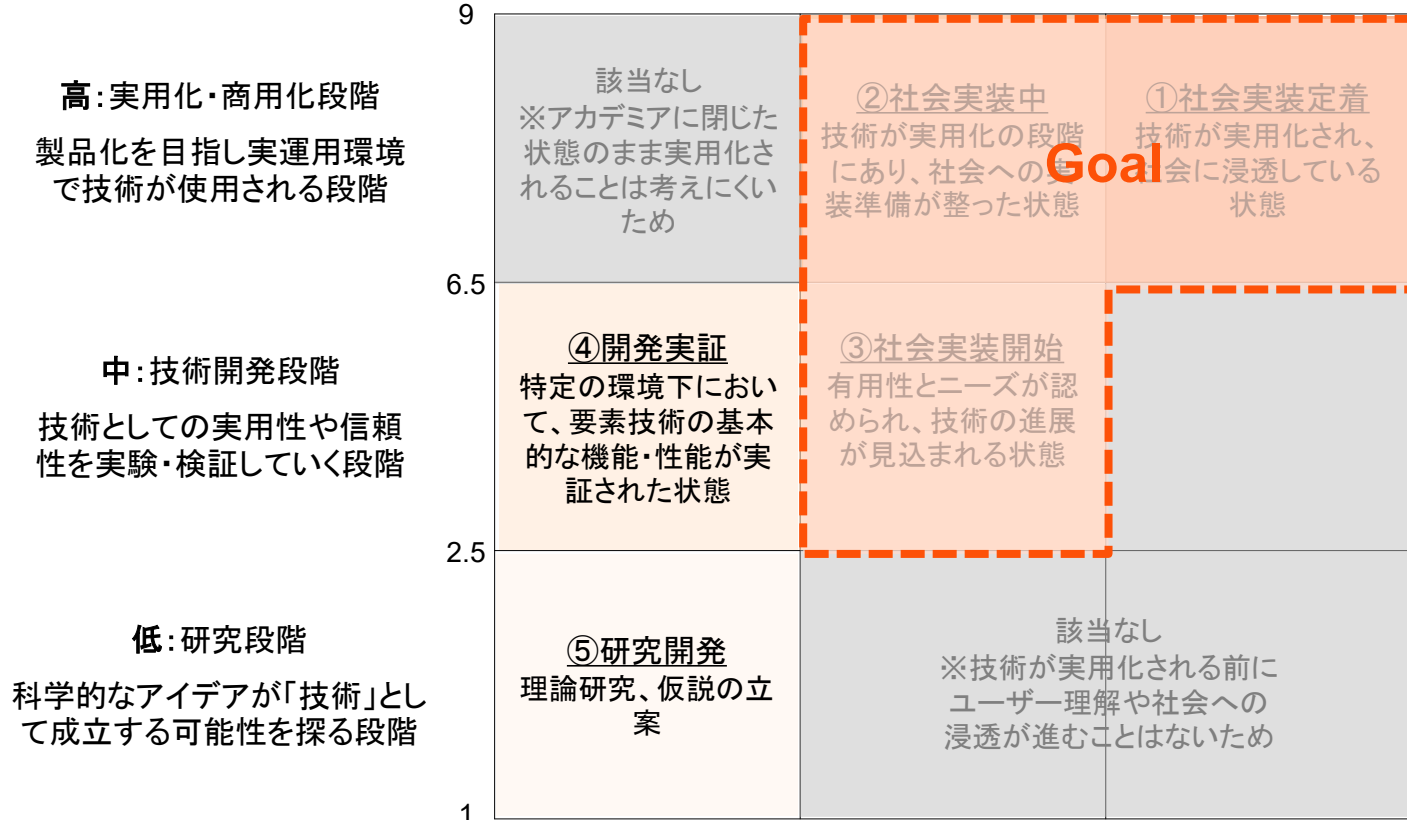
Adoption Readiness Level (ARL):

採用成熟度レベル

⑧ 本事業を通じた今年度のゴール

✓ 「①社会実装定着」を目標にしつつ、**各実証団体の実証開始時点の開発状況や技術特性に応じて最低限「②社会実装中」「③社会実装開始」の状態に到達するよう支援を行った。**

Technology Readiness Level (TRL):
 技術成熟度レベル



Adoption Readiness Level (ARL):
 採用成熟度レベル

低 研究者だけに技術の存在や有用性が認知された状態
 中 初期のユースケースに関わるユーザーに理解され、有用性が認められた状態
 高 広く社会で受容され継続的に生産・利用されている状態

- 2. 本事業における実施事項
 - 2-2. 偽・誤情報への対策技術の開発・実証
 - 2-2-2. 開発・実証の実施、実証団体に対する支援及び進捗管理

⑨ 成果発信イベントの企画・運営

- ✓ 3/16(月)に東京・大手町サンケイプラザにて成果発信イベントを開催した。
- ✓ 周知・広報においては、端的にイベント内容が伝わり、また、視覚的に興味を持ってもらえるようなリーフレットを作成した。
- ✓ 当日コンテンツ以外にも、付加価値となる情報を来場者へ提供できるよう、**本事業での調査結果の要点をパンフレットに纏め、来場者に配布した。**

イベントリーフレット

2026年 3/16 総務省 主催
13:00~17:00 (最終入場16:30)
会場: 大手町サンケイプラザ

参加 無料

ご関心をお持ちの企業や団体、地方自治体の皆さまへ、お待ちしております！

- コンテンツの真偽を見分けるお手伝いをする
- 情報の流通状況を見える化する
- コンテンツの出所と改ざん有無を保証する
- 情報の拡散を抑制・停止する

アプリなど 偽・誤情報対策の最新技術が一挙に集結！

開会あいさつ 13:00 ~ 13:05 総務省
基調講演 13:05 ~ 13:30 笹原和俊 教授 (東京科学大学)
『インターネット上の偽・誤情報対策技術の現在地とこれから』
出展団体(下記)による **ポスターやプロトタイプ等の展示**を実施！

ポスター 展示セッション 及び 出展者による 講演

参加登録フォームURL
<https://entry.smtkg.jp/public/seminar/view/5571>

お問い合わせ先は、「インターネット上の偽・誤情報等への対策技術の開発・実証事業 成果発信イベント 運営事務局」
saika2025_visitor_nasogycho@newsbase.co.jp

当日パンフレット

令和7年度
インターネット上の偽・誤情報等への対策技術の開発・実証事業 成果発信イベント
～偽・誤情報のリスクへ立ち向かう技術の最前線～

日時 2026年3月16日(月) 13:00~17:00
会場 大手町サンケイプラザ4階ホール
主催 総務省

発生した事象
虚偽の救助要請
拡散を促進した要因
リスクと対応技術

市民の混乱
自治体・報道機関が迅速に正確な情報を発信できず、危機感が増し市民の不安が増幅する

展示ポスター

事業者名 (正式名称)
請負契約件名

1 タイトル
原稿入力カスベス

2 タイトル
原稿入力カスベス

3 タイトル
原稿入力カスベス

4 タイトル
原稿入力カスベス

事業者ロゴ QR

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-2. 偽・誤情報への対策技術の開発・実証

2-2-1. 実証団体の公募・選定・運用に係る業務

2-2-2. 開発・実証の実施、実証団体に対する支援及び進捗管理

2-2-3. 経理処理等に係る業務

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

① 経理処理等に係る業務の概要

- ✓ 公的資金の執行という観点から、開発・実証費の支出について、それが適正であるかを検証することが必要である。このことから、実証団体から証憑を提出いただき、経理処理等に係る業務を実施した。
- ✓ 具体的には開発・実証開始時に経理処理説明会を実施するとともに、経理検査として、10月末に中間経理検査、3月下旬に確定経理検査を実施した。

経理処理の必要性

- 公的資金の執行という観点から、開発・実証費の支出について、それが適正であるかを検証することが必要である。
- このことから、事務局にて適正な経理処理の方法を整備し、適正な考え方に則り、実証団体が経費を支出しているかの確認を行った。

経理処理等に係る業務の概要

経理処理説明会の
開催
(6月末頃)

- 技術開発主体、研究・調査主体の経理処理に係る担当者向けに、本開発・実証に係る経理処理の進め方及び留意点を説明する経理処理説明会を開催。

中間経理検査の
実施
(10月末頃)

- 開発・実証開始日から10月中旬までに発生した経費の状況を把握し、適切な経理処理が行われているかを確認するために、中間経理検査を実施。

最終経理検査の
実施
(3月下旬頃)

- 開発・実証開始日から終了日までに発生した経費が適正に処理されているかを確認するために、最終経理検査を実施。

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

① 社会実装に向けた事務局の支援内容例

✓ 技術開発主体の意向、開発・実証状況及び本事業において目指すゴール設定に応じて、下記の支援内容について協議を行い、必要な支援を実施した。

| 主な対応事項例 | | 支援内容例 | 主な対応事項例 | 支援内容例 | |
|--------------|---------|-------------------------|---------------|---------------|--|
| 事業化戦略の策定・深掘り | 事業計画の策定 | ビジネスモデルの検討 | マーケティング | 調査・分析 | <ul style="list-style-type: none"> 市場構造及び競合動向を分析し、技術に関する規制や市場展開を後押しする施策を調査する 想定導入企業の課題やニーズを抽出し、対応方針に関する仮説を構築する |
| | | ビジネスモデルの具体化、事業展開のイメージ整理 | | 戦略設計 | <ul style="list-style-type: none"> 導入候補企業及びパートナー企業とのマッチングを実施し、初期接点の形成や提案候補先のリストアップを行う パートナー企業との協業体制の構築方針を策定し、ロードマップやKPIを作成する |
| | | 事業計画書の作成・事業計画書の内容精緻化 | | 市場展開 | <ul style="list-style-type: none"> 導入候補企業への提案活動を実施し、提案資料の作成、打合せへの同席等を行う パイロット導入やフィードバックの収集を通じて、実運用開始に向けた施策を作成する |
| | 営業活動試行 | ビジネスプラン持ち込み先候補出し | 社会実装に向けた枠組み作成 | 目的整理 | <ul style="list-style-type: none"> 標準化の目的・ビジョンを明確化し、社会・市場における役割や価値提供の方向性を整理する 実現すべき成果(信頼性・安全性等)を定義し、目的に応じた取組の基本方針を策定する |
| | | アポ取り | | 標準化 | <ul style="list-style-type: none"> ユースケースの作成や技術コアの特定を行い、標準化の形態(デファクトスタンダード含む)を選定する 産・官・学との連携体制を構築し、研究会等を組成する |
| | | 商談同席／事業説明など | | 標準化への対応活動 | <ul style="list-style-type: none"> 新規作業項目提案のドラフトを作成し、標準化団体関係者へのヒアリングを実施する |
| | | | | コンソーシアムの設立・普及 | <ul style="list-style-type: none"> コンソーシアムの目的・目標を定め、参加候補となる企業・大学等の調査及びマッチングを行う ガバナンス体制や運営ルール等の設計、想定リスクの洗い出しや対応方針の整理を行う |
| | | | | 設立準備 | <ul style="list-style-type: none"> 設立手続きの推進及び進捗管理を行い、役割分担や組織体制を構築する 各ワーキンググループの活動計画を策定する |
| | | | | 設立手続き | |

② 社会実装の共通支援内容 - ユースケース整理(1/5)

技術区分 I : コンテンツの真偽判別支援技術・改ざん検知技術

| 立場 | 想定されるリスク・課題(例) | 本技術でできること(例) | 関連実証団体 |
|------|--|--|--|
| 公的機関 | 災害時の偽・誤情報による誤判断 災害対応や行政処分、選挙管理において、インターネット上の不正確情報を前提に判断してしまい、対応の正当性が損なわれる恐れ。 | 拡散中の動画等が実際の被害状況を撮影したものを検証することにより、改ざんされた情報に基づく誤った初動判断を防止できる。 | <ul style="list-style-type: none"> エヴィクサー NTTドコモビジネス データグリット SEARCHLIGHT Sakana AI サン電子 NABLAS・NTT東日本 日本電気 |
| | 公的発信のなりすまし・誤認 行政機関や職員を装った情報が拡散し、住民が正規の公式情報と誤認して行動してしまう恐れ。 | 行政発信を装った投稿の真贋確認を支援することにより、なりすましを識別し、住民が正規の公式情報か否かを判断できる。 | |
| | 選挙関連情報による投票行動への影響 偽情報の拡散により、住民の不安や混乱が増幅し、社会秩序や公共安全に影響が及ぶ恐れ。 | 選挙関連のコンテンツが生成AIによって作成されたものか、意図的に改ざんされていないか確認することにより、投票判断に影響を与える偽・誤情報の拡散を抑制できる。 | |
| | 証拠の信頼性が争点化するリスク 行政処分や紛争対応でWeb・SNS上の画像・動画・音声を用いた際、それが生成AIによって作られたものである恐れ。 | Web・SMS上のコンテンツについて、生成・改ざんされていないか確認することにより、行政判断や紛争対応における証拠の信頼性を担保できる。 | |
| 企業 | ブランド毀損につながる虚偽投稿 自社製品の不具合や不祥事を装った動画等がSNSで拡散し、事実と異なる内容により信用が低下する恐れ。 | 投稿内容が実際の出来事に基づくものか、過去素材の流用や改ざんはないかを確認することにより、虚偽情報によるブランド価値の低下を防止できる。 | |
| | 誤報・誤公表リスク SNSや外部影響者から入手した画像等を自社のコンテンツに用いた結果、生成・改ざん素材だった場合に信用低下の恐れ。 | 公表前に使用する画像等の真正性を確認することにより、生成物や改ざん素材の誤使用による信用低下を防止できる。 | |
| | なりすましによる不正取引 経営者や取引先担当者を装ったオンライン会議映像・音声により、誤った意思決定や送金を行う恐れ。 | オンライン映像や音声本人によるものかを確認することにより、経営者や取引先担当者になりすました詐欺や不正送金を防止できる。 | |
| 一般市民 | 偽・誤情報に気づかぬまま拡散 SNS等で見た情報を事実と誤認し、拡散してしまう恐れ。 | SNS投稿内容が事実に基づくものかを確認することにより、誤った情報を拡散してしまうリスクを低減できる。 | |
| | なりすまし詐欺被害・偽サイトへの誘導 広告やSNS投稿から偽のECサイトや投資詐欺に誘導されてしまう恐れ。 | 広告画像やWeb表示が正規の事業者によるものか確認することにより、偽ECサイトや投資詐欺への誘導を回避できる。 | |

② 社会実装の共通支援内容 - ユースケース整理(2/5)

技術区分Ⅱ: 真正性保証・信頼性判断支援

| 立場 | 想定されるリスク・課題(例) | 本技術でできること(例) | 関連実証団体 |
|------|---|---|--|
| 公的機関 | 公式情報の誤認・混乱 Webサイト上やSNSで発信した災害情報・制度案内が、なりすましや改変された投稿と区別できず、住民がどれを信じればよいか分からなくなる恐れ。 | 行政が発信した正規の情報であることを示す真正性情報を付与・表示し、住民が一目で公式情報と判断できる。 | <ul style="list-style-type: none"> エヴィクサー NTTドコモビジネス OP組合 関西テレビソフトウェア サン電子 NABLAS・NTT東日本 |
| | 公的見解の切り取り・誤引用 注意喚起コンテンツが一部のみ切り取られて拡散され、行政の意図と異なる理解が広がる恐れ。 | 発信主体・コンテンツに真正性情報を付与することで、住民の誤認を予防できる。 | |
| | 選挙関連情報の正当性の主張困難 選挙期間中に、公式情報と偽情報が混在し、本物の情報であるにもかかわらず、虚偽の情報であると拡散される恐れ。 | 正式な機関が発信した正規情報であることを示す真正性情報を付与することで、住民の誤認を予防できる。 | |
| 企業 | 公式情報と非公式情報の区別困難 自社発信のニュースリリースや製品情報が第三者の改変投稿と見分けられず、ブランド毀損につながる恐れ。 | 自社が正式に発信したコンテンツであることを示す真正性情報を付与し、受け手に公式性を可視することで誤認を予防できる。 | |
| | 取引連絡において正当性の立証困難 オンラインでの見積・契約連絡が、なりすましと疑われ取引が滞る恐れ。 | 正規取引担当者・正規プロセスによる連絡であることを示す真正性情報を付与することで、相手方を確認できる。 | |
| | 生成AIコンテンツとの混同 自社が制作した画像・動画が生成AIによるものと誤解され、企業イメージや信用が低下する恐れ。 | 人手で制作された正規コンテンツであることを示す真正性情報を付与することで、生成AIコンテンツと誤解されることを予防できる。 | |
| 一般市民 | 信頼できる情報か判断困難 SNS・Webで見た情報について、どれが正規情報か分からず、不安や誤解が生じる恐れ。 | 発信元で作成履歴が表示されることで、信頼できる情報化を直感的に判断できる。 | |
| | なりすまし詐欺被害への不安 企業や著名人を装った情報が多く、本物かどうか見分けられない恐れ。 | 正規の発信元であることを示す真正性情報を付与・表示することで誤認を予防できる。 | |

② 社会実装の共通支援内容 - ユースケース整理(3/5)

技術区分Ⅲ: 情報流通状況の可視化・分析技術(1/2)

| 立場 | 想定されるリスク・課題(例) | 本技術でできること(例) | 関連実証団体 |
|------|--|---|---|
| 公的機関 | <p>偽・誤情報拡散による初動対応の遅れ 住民からの電話・窓口相談や現場対応の報告により、初めてSNS等で虚偽の被災情報や誤解を招く投稿が拡散していることに気づき、注意喚起や関係機関連携等の初動対応が遅れる恐れ。</p> | <p>SNS上で流通する投稿をリアルタイムで収集し、時系列分析により、災害関連の真偽不明投稿の立ち上がりを検知できる。</p> | <ul style="list-style-type: none"> • コードフォーージャパン • コンステラセキュリティジャパン • TDAI Lab • Sakana AI |
| | <p>拡散規模・影響範囲の把握困難 新たな施策・制度の開始後、SNS等で賛否が集まっていることは把握できても、それが、一部コミュニティ内で目立っているだけなのか、広範囲に拡散しているのかが見えにくく、全体像を捉えにくい可能性。</p> | <p>投稿の拡散ネットワーク(リポスト・引用・リンク関係)を可視化し、拡散の中心となるコミュニティ数や到達範囲の広がりを把握することで、「一部コミュニティでの話か、広域に波及しているか」の判断を支援できる。</p> | |
| | <p>拡散起点の特定困難 災害・感染症・選挙等に関する偽・誤情報が拡散している際、それがどこを起点として拡散しているのか、特定の意図を持つアカウント群による一斉拡散が起きているのか等を判断するのが難しい恐れ。</p> | <p>初出と見られる投稿候補を抽出し、早期の拡散経路(誰が誰を増幅したか)を辿れる形で可視化することで、拡散の起点や初期増幅要因の把握を支援できる。</p> | |
| | <p>偽・誤情報に関する継続監視の運用負担 災害・感染症・選挙等の複数テーマを監視する中で、夜間・休日に急増する投稿、同文面の大量投稿等まで追い切れず、結果として対応すべき火種への気づきが遅れる、または担当者の経験に依存して抜け漏れが生じる恐れ。</p> | <p>自動モニタリングツールとして活用することで、夜間・休日に急増する投稿等の情報を自動で収集し、担当者の張り付き負担を軽減できる。</p> | |

② 社会実装の共通支援内容 - ユースケース整理(4/5)

技術区分Ⅲ: 情報流通状況の可視化・分析技術(2/2)

| 立場 | 想定されるリスク・課題(例) | 本技術でできること(例) | 関連実証団体 |
|------|--|---|---|
| 企業 | <p>自社関連偽・誤情報による初動対応の遅れ 顧客からの問い合わせや営業現場からの報告(例:「SNSで危険だと見た」)をきっかけに初めて自社・自社製品に関する偽・誤情報が拡散していることに気づき、事実確認や対外対応の初動が遅れる恐れ。</p> | <p>ブランド名・製品名・役員名等に紐づく投稿を収集し、ネガティブ反応や誤解を招きやすい表現の急増を検知して、問い合わせ・現場報告より前の段階で火種を把握できる。</p> | <ul style="list-style-type: none"> • コードフォーージャパン • コンステラセキュリティジャパン • TDAI Lab • Sakana AI |
| | <p>組織的拡散か否かの見極め困難 初出投稿や初期拡散アカウント、同文面の大量投稿等の兆候を把握しにくく、意図的な攻撃や組織的行動の可能性を見誤る恐れ。</p> | <p>初期拡散アカウント群の投稿頻度等を可視化することで、意図的な攻撃や不自然な増幅の兆候の把握を支援できる。</p> | |
| | <p>訂正・注意喚起に向けた根拠整理が困難 会見やSNS等で訂正・注意喚起を出すべきか検討する際に、論点が複数並行して生じ、拡散状況も時々刻々と変化するため、どの論点を優先して説明すべきか、どのような表現が誤解の拡大につながりにくいかに、判断根拠を整理しにくい恐れ。</p> | <p>投稿を論点別に自動整理し、論点ごとの拡散量・拡散速度・誤解が生じやすい表現のパターンを並べて提示することで、「どの論点を優先して説明するか」の根拠整理を支援できる。</p> | |
| 一般市民 | <p>急拡散時の状況把握困難 災害時等に情報が短時間で大量に流通する局面では、断片的な情報や誤った情報が先行して拡散しやすく、状況の理解が難しくなる可能性。</p> | <p>大量投稿を内容の近さで自動グルーピングし、「どの話題がどれくらい増えているか」を数個の主要論点に圧縮して表示することで、状況理解を支援できる。</p> | |

② 社会実装の共通支援内容 - ユースケース整理(5/5)

技術区分Ⅳ: 偽・誤情報の拡散を防止・無効化する技術

| 立場 | 想定されるリスク・課題(例) | 本技術でできること(例) | 関連実証団体 |
|------|---|---|---|
| 公的機関 | 災害・制度変更等に関する偽・誤情報の拡散 災害対応や制度改正に関する偽・誤情報がSNS等で拡散し、住民の誤解や問い合わせ集中を招く恐れ。 | SNS上で急増している関連投稿に対し、投稿閲覧時点で注意喚起を表示させることで、利用者に慎重な判断を促すことができる。 | <ul style="list-style-type: none"> Classroom Adventure コンステラ セキュリティジャパン 関西テレビソフトウェア Sakana AI |
| | 行政発表の切り取り・誤解拡散 公的発表が一部だけ切り取られ、本来の意図と異なる理解広まる恐れ。 | 公的機関の投稿を関連投稿として優先表示させることで、正確な情報への接触機会を高めることができる。 | |
| | 偽・誤情報拡散後の是正困難 一度広まった偽・誤情報が訂正されず、住民の理解が固定化する恐れ。 | 偽・誤情報を含む投稿を閲覧した利用者に対して、訂正情報を発信することで、誤った理解を是正することができる。 | |
| | 職員等による誤拡散 SNS等で拡散した偽・誤情報を基に誤った対応をする恐れ。 | 偽・誤情報対策に関する学習コンテンツを提供・活用することで、偽・誤情報を適切に判断できる。 | |
| 企業 | 製品・サービスに関する誤認 製品仕様や効果に関する偽・誤情報が拡散し、消費者対応・苦情対応が増加する恐れ。 | 注意喚起を表示させるとともに、公式ページへのリンクを表示させることで、利用者が正確な情報にアクセスできる。 | |
| | 風評・炎上による事業影響 虚偽や誤解に基づく投稿がSNS等で急速に拡散し、ブランド価値が毀損される恐れ。 | 過去事例や想定シナリオを基に、偽・誤情報がどのように広がるかを事前に試算し、効果的な対応策を検討できる。 | |
| | 広告・広報表現の切り取り拡散 広告等が部分的に切り取られ、意図しない解釈が拡散する恐れ。 | 切り取られる前の元広告へのリンクを表示させることで正式な情報にアクセスできる。 | |
| | 偽・誤情報拡散後の是正困難 自社製品に関する誤った情報が訂正されず、顧客の理解が固定化する恐れ。 | 誤っている箇所を明示し、正しい情報を段階的に提示することで、偽・誤情報を事後的に是正することができる。 | |
| | 従業員等による誤拡散 SNS等で拡散した偽・誤情報を基に誤った対応をする恐れ。 | 偽・誤情報の手口や真偽検証の方法に関する学習コンテンツを提供・活用することで、偽・誤情報を適切に判断できる。 | |
| 一般市民 | 偽・誤情報に気づかぬまま拡散 SNS等で見た情報を事実と誤認し、拡散してしまう恐れ。 | 共有ボタンを押下時に「未確認情報の可能性」を明示することで、利用者が共有前に事実確認を行う機会を設けることができる。 | |

③ 開発・実証から得られた社会実装に向けた政策的示唆

- ✓ 本年度の開発・実証を通じて、偽・誤情報対策の技術的有効性は一定程度確認された。一方で、当初想定していたユーザ層であっても導入には至らないケースが見られ、**社会実装に向けては、市場形成に加え、利用主体・利用場面の明確化や具体的な活用イメージの提示が課題**であることが明らかになった。
- ✓ このため、需要創出とあわせて、利用主体・利用場面の明確化や活用イメージの具体化を通じた**市場形成の促進が重要**である。

開発・実証から得られた知見

現状認識

- 本事業開始時点で、想定ユーザ・ユースケースを整理した
- 開発実証を通じて技術的有効性は一定確認された
- 一方で、当初想定していたユーザ層でも、実際の購入・導入まで至らないケースが確認された

社会実装上の課題

- インターネット上の偽・誤情報に関する被害は拡大する一方で、導入・予算化・業務への組み込みとの間にギャップが存在する
- 社会実装の初期フェーズにおいては、当該ギャップの解消が課題であり、特に以下の点が重要である
 - ✓ **市場形成の促進**（課題を有する主体において、本対策技術の需要が自然に生じる状態の形成）
 - ✓ **利用主体・利用場面の明確化**（誰が、どの業界で・どの業務において利用するのかを具体化し、社会側から見える状態の構築）

次年度以降に取り組むべき事項

市場形成の促進

- 偽・誤情報対策技術の導入が期待される分野・利用主体を整理し、重点領域における需要創出を図る
- 開発・実証成果や被害事例を活用し、導入効果を可視化することで、社会的受容及びユーザ導入を促進する。
- 初期導入を後押しし、市場の立ち上がりを支援する

利用主体・利用場面の明確化

- 利用主体ごとの課題や利用目的を整理し、**技術を活用対象(主体・業務内容)を明確化**する
- 業界・業務単位での利用場面を具体化し、**導入検討に資する形で提示**する
- ユーザの業務フローや運用体制への組み込みを前提とした活用モデルを整備するなど、社会実装の可能性を高める

④ 技術区分ごとの社会実装に向けたポイント

| | | 技術区分Ⅰ コンテンツの真偽判別支援技術 | 技術区分Ⅱ 真正性保証・信頼性判断支援・ 改ざん検知支援技術 | 技術区分Ⅲ 情報流通状況の可視化・分析技術 | 技術区分Ⅳ 情報の拡散を防止・無効化する技術 |
|--------------|---------|--|--|---|---|
| 社会実装に向けた課題 | 技術開発 | <ul style="list-style-type: none"> 動画生成AI等の急速な進化により新たな手口のディープフェイクが次出しており、検出モデルの精度劣化リスクが常態化 ディープフェイク検知や言説疑義を検証・確認するためのインターフェースが、情報検証の習慣のない一部のユーザーには理解困難 | <ul style="list-style-type: none"> 一定の市場普及率が求められるにも関わらず、C2PAやOPを検証・確認するためのインターフェースが一般ユーザ視点では理解しづらい | <ul style="list-style-type: none"> 表示される内容や出力される結果の信頼性の確保に向け、分析精度の向上や可視化における誤認防止に取り組む必要がある 利用者の業務特性に応じて求められる分析の粒度や速度が異なることから、多様な業務要件への柔軟な対応が必要 | <ul style="list-style-type: none"> 動画生成AI等の急速な進化により新たな手口のディープフェイクが次出しており、拡散対策対象の特定・優先度付けが難しい |
| | 事業化・製品化 | <ul style="list-style-type: none"> 一定の市場普及率を確保するためには各業界にアーリーアダプターを見つける必要があるが、実証団体のみではユーザとなり得る事業者を見つけれない | <ul style="list-style-type: none"> 現時点でC2PAやOPは普及しておらず、導入効果を一般ユーザ視点では理解しづらい 一定の市場普及率を確保するには早期にアーリーアダプターを見つける必要があるが、C2PAやOPに関しては専門性が高く、実証団体のみでユーザとなる事業者を確保が困難 | <ul style="list-style-type: none"> 単独での社会実装に加え、既存のSNS炎上対策等のデジタルリスク対策サービスの一機能として実装していくことが想定される 多様な業種・業務における具体的なユースケースを創出し、付加価値を明確化するなど、ビジネスとして成立させるための具体的な方策の検討が必要 | <ul style="list-style-type: none"> 教育機関・法人・公的機関・一般ユーザーそれぞれに一定の需要はあるが、運用面の負担から当該技術の本格導入や継続的な取り組みに二の足を踏む事業者が多い |
| | 国際標準 | <ul style="list-style-type: none"> ディープフェイク技術は高度化・多様化しており、真偽判別技術も様々な方法が存在する。国際標準化により、真偽判別の評価指標や検証手法、データフォーマットなどの共通基準が確立されることで、技術の信頼性や比較可能性向上が期待できる | <ul style="list-style-type: none"> C2PAやOPのようなコンテンツに対して真正性を付与する技術の場合、その他多数のユーザが同一の仕組みを利用しない限り、その効果は限定的である。 | <ul style="list-style-type: none"> 現時点で、ITU、ISO、IEEEなどの主要団体において、情報流通状況の可視化・分析に特化したWG(作業部会)やSG(研究課題)が設置されておらず、どの標準化団体が主導すべきかすら不明確な状況 | <ul style="list-style-type: none"> 現時点で、ITU、ISO、IEEEなどの主要団体において、情報の拡散防止・抑止技術に特化したWG(作業部会)やSG(研究課題)が設置されておらず、どの標準化団体が主導すべきかすら不明確な状況 |
| | 市場構築 | <ul style="list-style-type: none"> 一般ユーザの中には、情報の真偽を検証することの必要性について理解に乏しい層が存在し、導入に向けたインセンティブが無い場合がある | <ul style="list-style-type: none"> そもそも一般ユーザ視点では、情報の真正性や信頼性を保証することの必要性について理解に乏しく、導入に向けたインセンティブが無い場合がある。(写真撮影されたデータの改ざんリスクを考慮できていない) | <ul style="list-style-type: none"> 当該技術単独での市場が現時点では確立されていない。市場構築に向けては、当該技術を導入することによるメリットを明確化し、潜在的な需要を顕在化させ、市場としての規模や持続性を確保していくことが必要 | <ul style="list-style-type: none"> 当該技術単独での市場が現時点では確立されていない。情報の拡散防止・無効化の技術単体では、事業者の予算確保が難しい |
| 社会実装に向けたポイント | | <ul style="list-style-type: none"> 一般利用者が有料で利用するほどのインセンティブが高まっておらず、利用動機の創出や価値訴求の強化がマネタイズの実現に向けて重要 | <ul style="list-style-type: none"> 一定の市場普及を経ることによりはじめて効力を発揮することなどを踏まえて、個社のサービス導入だけでなく、国際標準や市場構築など面での導入に向けた施策を実施することが重要 | <ul style="list-style-type: none"> 技術面での信頼性の確保や多様な業務要件への柔軟な対応に加え、ビジネスとしての成立させるための他の技術領域との連携したサービス提供による市場の形成といった多面的な課題に対応していくことが重要 | <ul style="list-style-type: none"> 情報の拡散を防止・抑止する技術の社会実装に向けては、当該技術単位での普及だけでなく、他の技術領域を組み合わせたパッケージ開発や市場訴求が重要 |

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

2-4-1. 作成資料概要

2-4-2. 実証団体成果

① 中間報告書 - 事務局作成

✓ 中間報告書は以下内容で構成した。



目次

| | |
|--|------|
| 1. 偽・誤情報の流通・拡散に関連する最新動向(海外事例を含む。)に関する調査・検討 | p3 |
| 1. 偽・誤情報の流通・拡散に関連する最新動向(海外事例を含む。)に関する調査・検討(* 事前調査報告書内容の再掲) | |
| 2. 調査事例の詳細化・掘り下げ | |
| 3. 令和7年度参議院議員選挙における偽・誤情報事例 | |
| 2. 偽・誤情報に関連する最新生成AI技術 | p168 |
| 1. 最新技術モデルと、昨年・過去のモデルを比較した技術発展 | |
| 2. 生成AIによる災害等の事例 | |
| 3. 生成AIの見分けに関するレポート | |
| 3. マルチモーダルな偽誤情報 | p192 |
| 1. マルチモーダルな偽誤情報の定義・事例 | |
| 2. (参考)マルチモーダルAI | |
| 4. 各実証団体の進捗状況 | p214 |

2

② 成果報告書 中間報告版 - 技術開発主体作成

✓ 技術開発主体の成果報告書 中間報告版を以下内容で取りまとめた。

令和7年度 インターネット上の偽・誤情報等への対策技術の開発・実証事業

**インターネット上の偽・誤情報等への対策技術の開発・実証事業
(開発・実証件名)**

成果報告書 中間報告版

2025/MM/DD
技XX_ (実証団体名)

技術開発主体 研究・調査主体

技XX_ (実証団体名)
1. 成果報告書 中間報告版
(開発・実証件名)

| | | | |
|--------------------------------------|---|----------------------------|--|
| 開発予定の 対価形態 および 課料 | SNS中提示版などでは、誤情報が動画表現や皮肉、質問の形式などで巧みに投稿されるケースが多く、自動検出の困難である。 特に○○、△△といった点に誤情報が多く、誤検知や見逃しが発生しやすい。このため、誤情報の特徴を的確に捉える精度を■技術の開発が求められている。 | 課題を 克服する 目標 | SNS上で拡散される誤情報表現、動画表現を含む偽・誤情報も、文脈を踏まえて高精度に検出・分類できる技術を開発し、様々なケースに対応可能にする。 これにより、アカウントチーム業務を一貫ユーザーが対応可能・軽減できる健全な情報流通環境の構築を目指す。 |
| 技術 開発区分 | コンテンツの真偽判別支援技術、改ざん検知技術 | 実施体制 (所属) | A(株)、(株)B、(一社)C |
| 対象とする モデル種 | 文章、画像、音声、動画 | | |

| | |
|---|---|
| 開発・実証を行う対策技術の概要 | 対策技術の 開発・検証 |
| <ul style="list-style-type: none"> 本開発・実証では○○という技術の開発を進める。対策技術の全体像は下記となる。 ○○には、xxx、△△という機能を実装する予定である。 具体的には○○というデータを用いて、xxxという機能で■といった分析を行い、分析結果を用いることで○○ということが可能になる。 類似技術として○○や△△といった技術が挙げられる。類似技術と比較して、△△という点やxxという点が強みである。 | <ul style="list-style-type: none"> 対策技術の開発においては、○○、△△、■といった機能の開発を行うとともに、△△、××といった手法を組み込むことで、判定精度向上を目指す。 開発した対策技術の検証においては、自社内において、KPIとして設定している○○率や△△スコアの指標を検証する。 また、○○といった観点から○○という取組を実施する。 |
| 社会的実証に 対応する | <ul style="list-style-type: none"> ○○機関や△△団体へのヒアリングを通じ、連携可能性や導入時の懸念点を把握するとともに懸念点解消に向けて○○に取り組む。 また、連携可能なビジネスモデル構築するために、○○といった取組を行う。加えて国際標準化を担っていくための取組として、××の推進もする。 |
| 開発・実証 における主要KPI | <ul style="list-style-type: none"> 技術面でのKPIでは、○○という指標を用いる。 ビジネス面でのKPIでは、○○という指標を用いる。 |
| 社会実証によって 期待される効果・期待 される効果 | <ul style="list-style-type: none"> 本技術の実証により、SNS上で急速に拡散される偽・誤情報を早期に検出、拡散の未然防止につながる可能性がある。これにより、○○、△△が期待される。 また、○○の作業負担が△△%軽減し、偽・誤情報対応の迅速化と質的向上に貢献する。 利用基盤としても、○○といった点が促進され、結果として■、○○といった形成に寄与することが期待される。 |

開発・実証を行う技術の全体像
(画像やイラストを用いた説明)

Template

技術開発主体 研究・調査主体

技XX_ (実証団体名)
1. 成果報告書 中間報告版
本開発・実証のスケジュールと進捗状況

| 主な実施事項 | 令和7年 | | | | | | 令和8年 | | |
|--------|------|----|-----|-----|-----|----|------|----|--|
| | 8月 | 9月 | 10月 | 11月 | 12月 | 1月 | 2月 | 3月 | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |

| | |
|---------------------------|--|
| 現在までの 進捗状況 | <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX |
| 本事業における 課題やリスク | <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX |
| 実証に係る 今後の予定 | <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX |

③ 社会実装実施計画書 初期案 - 技術開発主体作成

✓ 各技術開発主体の社会実装実施計画書初期案を以下内容で取りまとめた。

令和7年度 インターネット上の偽・誤情報等への対策技術の開発・実証事業

インターネット上の偽・誤情報等への対策技術の開発・実証事業
 (開発・実証件名)

社会実装実施計画書 初期案

2025/MM/DD
 技XX_ (実証団体名)

技XX_ (実証団体名)
 2. 社会実装実施計画書 初期案
 (開発・実証件名)

社会実装時のビジネスモデル案

ユーザー-導入先の詳細とそのポイント
 (以下記載例)

- 具体的なターゲットユーザー-導入先としては、Xデビア社XX部やXデビア社XX部を想定。
- これらのユーザーが抱える課題(ペインポイント)としては、xxxや△△が挙げられ…

ユーザー-導入先への提供形態
 (以下記載例)

- xxxや△△という課題に対する解決方法として、本技術を用いた○○ツールを～の形式により提供する。
- 具体的には、Webブラウザの拡張機能として本技術を組み込み…

技XX_ (実証団体名)
 2. 社会実装実施計画書 初期案

事業化に向けた実効的な取組計画や展望

今年度を実施する社会実装の取組計画・事業活動

(以下記載例)

- 今年度は、○○機関や△△団体へのヒアリングを通じ、連携可能性や導入時の懸念点を把握するとともに懸念点解消に向けて○○に取り組み、
- また、持続可能なビジネスモデルを構築するために、～といった取組を行う。
- 加えて国際標準化を促していくための取組として、××の推進を図る。

事業の拡大に向けた中長期的なスケジュール・取組計画

(以下記載例)

- xxx年度以降、○○や△△を達成するため、～といった取組を…
- また、取組の高度化・広域化を進めていくため、本技術の事業化に向けた今後の深化の観点としては、○○や△△の観点が重要となる。
- の観点においては、～といった工夫が必要であり…

今年度の社会実装に向けた課題

(以下記載例)

- 本技術の事業化に向けた課題としては、xxx、△△といった課題が挙げられる。
- 当該課題に対する対策としては、○○頃～～を実施する想定であり…

有識者の皆さまへの相談事項

- XXXXXXXXXXXXXXXXXXXX
- XXXXXXXXXXXXXXXXXXXX

④ 成果報告書 中間報告版 - 研究・調査主体作成

✓ 研究・調査主体の成果報告書 中間報告版を以下内容で取りまとめた。

令和7年度 インターネット上の偽・誤情報等への対策技術の開発・実証事業

インターネット上の偽・誤情報等への対策技術の開発・実証事業
 (研究・調査件名)

成果報告書 中間報告版

2025/MM/DD
研XX_ (実証団体名)

[詳細調査主体] [研究・調査主体]

研XX_ (実証団体名)
1. 成果報告書 中間報告版
(研究・調査件名)

| | | | |
|-----------------------|--|--------------------------|--|
| 研究・調査内容に追加する事項 | 偽・誤情報の流通・拡散の防止として、○や△の手法が広く用いられているが、当該手法では～～という限界があり、十分な偽・誤情報対策として機能しない可能性がある。このため、偽・誤情報の特徴を的確に捉える精度な■の解明が求められている。 | 課題を研めたい目標 | xxxに関する新たな仕組みを解明することにより、～～によって、偽・誤情報対策の迅速化と質的向上を目指す。 |
| 研究・調査区分 | 偽・誤情報対策技術に係る研究 | 実施体制 (代表者、連絡先) | A(姓)、(株)B、(一社)C |

| | |
|--|---|
| 研究・調査の概要 | 研究の実績 |
| <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX | <ul style="list-style-type: none"> ○◎名程度の実験参加者を、xxx、△△のグループに分け、それぞれ～への回答を行った。 その際、～～という(ターン)A、～～という(ターン)Bを用意し、■の条件について比較を行うことで… |
| 研究の進捗状況 | <ul style="list-style-type: none"> ○◎名程度の実験参加者を、xxx、△△のグループに分け、それぞれ～への回答を行った。 ～～への回答内容をグループごとに比較することで、■の有効性を検証し… |
| 研究・調査によって期待される効果 | <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX |

本研究・調査の全体像
(画像やイラストを用いて説明)

Template

[詳細調査主体] [研究・調査主体]

研XX_ (実証団体名)
1. 成果報告書 中間報告版
本研究・調査のスケジュールと進捗状況

| 主な実施事項 | 令和7年 | | | | | | 令和8年 | | |
|--------|------|----|-----|-----|-----|----|------|----|--|
| | 8月 | 9月 | 10月 | 11月 | 12月 | 1月 | 2月 | 3月 | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |
| XX | → | → | | | | | | | |

| | |
|----------------------|--|
| 現在までの進捗状況 | <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX |
| 本事業における課題やリスク | <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX |
| 実証に係る今後の予定 | <ul style="list-style-type: none"> XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX |

⑤ 事業実施報告書 - 事務局作成

✓ 事業実施計画書は以下内容で構成した。



⑥ 成果報告書 - 技術開発主体作成

✓ 技術開発主体の成果報告書を以下内容で取りまとめた。

成果報告書

令和7年度 インターネット上の偽・誤情報等への対策技術の開発・実証事業

(開発・実証件名を記入ください。)

成果報告書

2026/1/26
 技XX_実証団体名

成果報告書 概要版

令和7年度 インターネット上の偽・誤情報等への対策技術の開発・実証事業

(開発・実証件名を記入ください。)

成果報告書 概要版

2026/1/26
 技XX_実証団体名

成果報告書 簡易版

令和7年度 インターネット上の偽・誤情報等への対策技術の開発・実証事業

(開発・実証件名を記入ください。)

成果報告書 簡易版

2026/1/26
 技XX_実証団体名

Template

技XX_実証団体名 成果報告書

目次

- 開発・実証のサマリ
 - 開発・実証のサマリ
- 開発・実証の背景・目的
 - 開発技術によりアプローチする課題
 - 開発技術により目指す姿・ゴール
 - 開発技術により対処可能なユースケース
- 開発・実証における「対策技術の開発」
 - 技術開発の全体像
 - 技術開発の個別詳細
- 開発・実証における「対策技術の有効性等に関する検証及び調査」
 - 検証及び調査の全体像
 - 検証及び調査の個別詳細
- 開発・実証における「対策技術の社会実装に向けた取組」
 - 社会実装に向けた取組の全体像
 - 社会実装に向けた取組の個別詳細
- 開発・実証における「普及啓発活動への協力」
 - 普及啓発活動の全体像
 - 普及啓発活動の個別詳細
- 開発・実証の課題・展望
 - 技術開発及び社会実装における課題・展望
- 開発・実証の実施体制等
 - 実施体制及び役割分担
 - 全体スケジュール

技XX_実証団体名 成果報告書 概要版

目次

- 開発・実証における対策技術の開発
 - 開発技術によりアプローチする課題・目指す姿
 - 技術開発の取組・成果
- 開発・実証における社会実装に向けた取組
 - 社会実装に係る取組・成果
 - 社会実装時のビジネスモデル等
 - 技術開発及び社会実装にあたっての課題・展望
 - 事業の拡大に向けた中長期的な計画

技XX_実証団体名 成果報告書 簡易版

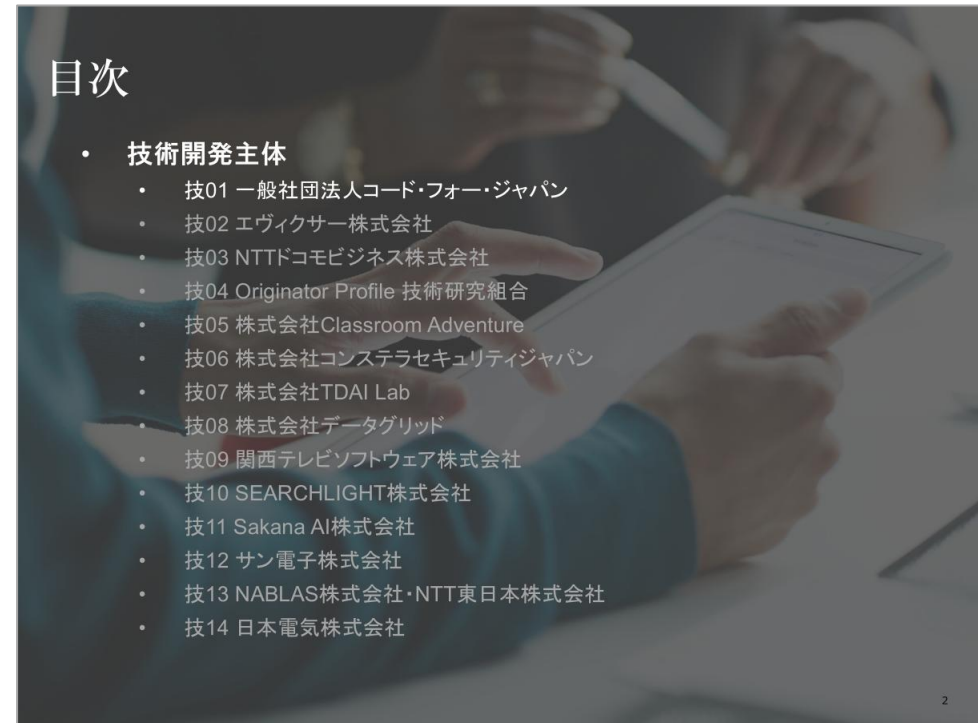
(開発・実証件名を記入ください。)

お問い合わせ
 問合せ先
 技術開発
 技術開発の取組・成果
 社会実装に係る取組・成果
 技術開発及び社会実装にあたっての課題・展望

本文
 ○○社 XX部 課長
 XXX(氏名)

⑧ 社会実装実施計画書 - 事務局作成

✓ 社会実装実施計画書は以下内容で構成した。



⑨ 社会実装実施計画書 - 技術開発主体作成

✓ 技術開発主体の社会実装実施計画書を以下内容で取りまとめた。

令和7年度 インターネット上の偽・誤情報等への対策技術の開発・実証事業

(開発・実証件名を記入ください。)

社会実装実施計画書

2026/1/26
技XX_実証団体名

技XX_実証団体名

社会実装実施計画書

目次

1. 開発技術の社会実装の概要
 1. 社会実装時のビジネスモデル
 2. ユーザ・導入先とそのペインポイント
 3. サービスのシステム構成・提供形態等
2. 開発技術の社会実装に向けた取組
 1. 事業の拡大にあたっての課題・展望
 2. 事業の拡大に向けた中長期的な計画
3. 開発技術の社会実装に向けた実施体制等
 1. 事業化に向けた実施体制
 2. 事業化に向けた収支計画

目次

1. 本事業の背景・目的

2. 本事業における実施事項

2-1. 偽・誤情報の流通・拡散に関連する最新動向に関する調査・検討(事前調査)

2-2. 偽・誤情報への対策技術の開発・実証

2-3. 開発・実証事業で採択する技術に係る社会実装の実施

2-4. 報告書作成等

2-4-1. 作成資料概要

2-4-2. 実証団体成果

① 技術開発主体における主たる成果等サマリ (1/4)

| | 主たる技術開発の成果 | 今後に向けた主な課題・展望 |
|--------------------------------------|--|---|
| 技01_ 一般社団法人 コードフォージャパン | <ul style="list-style-type: none"> ・ X(旧Twitter)が提供するコミュニティノート(投稿に対してユーザーが背景情報を作成・評価)を活用し、国内の偽・誤情報に関する実態を可視化・分析する基盤を開発 ・ 利用者ニーズ把握のため、多様なステークホルダーと関係性を構築 | <ul style="list-style-type: none"> ・ 現在取り扱っているデータ量に限りがあることに加え、データを取得後のデータ処理に時間と費用を要する ・ データ処理の効率化及び他組織が使用できるインターフェイスを開発することで広範囲での活用を企図 |
| 技02_ エヴィクサー株式会社 | <ul style="list-style-type: none"> ・ 音響透かし・フィンガープリントを活用した真偽判定システム判定の精度向上に加え、レポート機能や文字起こし機能を拡張 ・ 新たな活用方法として、音響透かし技術を用いたアプリ不要の参加型イベント、電話網を通じた音声情報のリアルタイム検知機能の実証を実施 | <ul style="list-style-type: none"> ・ 自律的な普及・定着に向けて、手対応を伴う運用負荷、多様なニーズへの対応、インフラ要件の高さによる電話網等への参画障壁の高さが課題 ・ システムのモジュール化(機能の部品化)を進めることで柔軟な提供形態を構築予定 |
| 技03_ NTTドコモビジネス 株式会社 | <ul style="list-style-type: none"> ・ デジタルコンテンツの生成元や編集履歴を証明するC2PA技術を応用し、撮影した画像の位置情報や撮影時刻の改ざん有無を検証するシステムを開発。 ・ 実証実験では、放送・メディア事業者と協力し、SNSコンテンツの真偽検証に要するコストや時間をどの程度削減できるかを検証 | <ul style="list-style-type: none"> ・ スマホ端末搭載に向けたメーカー側の動機付けが不十分な点、SNSやC2PA非対応ツールを経由すると真正性保証が困難になる点が課題 ・ 国際標準化団体への必須要件化の働きかけ、SNSと連携した来歴情報の保持、一般的な画像編集アプリへの対応拡大を企図 |
| 技04_ Originator Profile 技術研究組合 | <ul style="list-style-type: none"> ・ Webサイトの作成主体・発信主体や所属する業界団体等の情報を、第三者機関が確認し、改ざん不可能な形で付与し、受け手がサイトや記事などのコンテンツ上で作成者・発信者を確認できるようになる仕組み(Originator Profile)を構築 ・ 実証実験では、複数企業・自治体と協力し、自社サイトやサイト埋め込み型広告におけるOriginator Profileの機能検証を実施 | <ul style="list-style-type: none"> ・ 今年度はブラウザ拡張機能を開発したが、一般ユーザーにとってブラウザの拡張機能の利用はハードルが高く、一般利用者向けの利便性向上への対応が課題 ・ Originator Profileのマネタイズと自律的な運営を見据え、OP付与の広告取引を商取引に組み込んでいくための手法やビジネスモデルの精緻化を企図 |

① 技術開発主体における主たる成果等サマリ(2/4)

| | 主たる技術開発の成果 | 今後に向けた主な課題・展望 |
|-------------------------------------|---|---|
| 技05_ 株式会社 Classroom Adventure | <ul style="list-style-type: none"> 受け手が未知の事例でも「疑い、確かめ、拡散しない」行動を取れるよう、偽・誤情報生成の疑似体験を中心とした技術を開発 使用ユーザに対する偽・誤情報の真偽判定スコア向上・満足度調査の結果より、学習効果を定量的に検証 | <ul style="list-style-type: none"> 仮想環境で生成された偽・誤情報の外部流通リスクへの対策や、教育目的を逸脱しない安全な設計、テキスト・画像に限らず、多様な形式への対応が必要 受講者の管理システム者側開発、学習段階に応じて段階的に利用範囲を拡張する設計、実環境条件での学習機能拡充を予定 |
| 技06_ 株式会社 コンステラセキュリティ ジャパン | <ul style="list-style-type: none"> OpenCTI基盤・STIX 2.1準拠の偽・誤情報分析プラットフォームを構築し、複数のプラットフォームからのデータ自動変換・一元管理基盤を整備 AI分析モジュールによるセンチメント分析・エンティティ検知・AI生成コンテンツ検知・ナラティブ検知・拡散シミュレーション等の多角的な分析・可視化機能を実装 | <ul style="list-style-type: none"> 大量データの取り込み時の処理待ちの発生やOpenCTIの可視化カスタマイズの制約による柔軟性の制限、分析手法上の推論の限界、想定顧客への訴求力の強化が課題 処理方式の改善や処理基盤の強化、API連携を含むデータソース多様化、既存サービスとの差別化ポイントの明確化等に取り組む予定 |
| 技07_ 株式会社TDAI Lab | <ul style="list-style-type: none"> SNS上の言説の論理構造や表現を深層的に解釈し、ナラティブ分類や論点整理等により、短時間で全体把握可能な情報拡散分析基盤を構築 ベンチマークを用いた評価では、既存研究を上回る性能を実現。ユーザーのニーズに合わせ、事前データあり/なし拡散分析、ファクトチェック分析機能の3つの機能を実装 | <ul style="list-style-type: none"> 生成AI出力結果を根拠情報として用いる際の信頼性懸念やユーザーの業務特性に合わせた異なる分析観点・粒度への対応が課題 まずは判断補助用途での導入を進め、実績を蓄積しつつ、用途別に機能をパッケージ化し利用者属性に応じたカスタマイズを段階的に提供予定 |
| 技08_ 株式会社 データグリッド | <ul style="list-style-type: none"> SNSユーザ及び訂正主体向けツールを開発。生成AIの進化に対応可能な基盤を構築し、画像・動画・音声のディープフェイク検出精度を向上 SNSユーザ向けには、情報接種の偏りや真偽不明言説の拡散履歴を可視化し、情報的健康への気付きと行動変容の機会を提供する機能を開発。訂正主体向けには、高度なリスク評価とトリアージにより、対応優先度を迅速・適切に判断し、高リスク事案の特定率向上と検知時間短縮する機能を開発 | <ul style="list-style-type: none"> 生成AI技術の急速な進化への追従、検出対象メディアの拡大、多言語対応の高度化に加え、SNSユーザーの獲得・定着、訂正情報の到達率向上、訂正主体によるツール活用の拡大、法規制・倫理面への対応が課題 ショート動画・ライブ配信への対応、ゲーミフィケーション等によるUX設計の高度化、訂正情報の流通チャネル強化、パートナー企業との協業による導入促進等に取り組む予定 |

① 技術開発主体における主たる成果等サマリ(3/4)

| | 主たる技術開発の成果 | 今後に向けた主な課題・展望 |
|----------------------|---|---|
| 技09_ 関西テレビソフトウェア株式会社 | <ul style="list-style-type: none"> ブロックチェーンデータを、放送局内設備を経由して有線で受信機に配信する実証を実施 受信した情報を基に、災害時における拡散画像の改ざん有無を検知できるアプリケーションと、偽・誤情報を位置情報とともに提示することができるアプリケーションを開発 | <ul style="list-style-type: none"> 社会実装に向けては、実電波を用いたフィールド実証が必要となるため受信機側の実装設計をはじめ技術面・運用面に課題 平常時活用、災害時のデータ管理者の範囲拡大を含め、技術面・運用面の課題を段階的に整理予定 |
| 技10_ SEARCHLIGHT株式会社 | <ul style="list-style-type: none"> 動画コンテンツの内容をAIが解析・構造化することで、「全編視聴」や「目視確認」に依存していた確認作業を自動化・効率化するシステムを構築 急速に拡散する動画に対する報道機関などが迅速に事実確認できるモニタリングツール、動画コンテンツの法令適合性検証時の支援ツールとして実証を実施 | <ul style="list-style-type: none"> 日本国内の情報空間が外資系プラットフォームに依存しており、正確な情報流通状況の把握や対策が困難なことが課題 本技術の機能を拡充するとともに、他のパートナー企業にも技術提供して活用範囲と収益を広げ、持続可能な国内プラットフォーム市場の創出を企図 |
| 技11_ Sakana AI株式会社 | <ul style="list-style-type: none"> X(旧Twitter)上の真偽判定システムに加え、X上の論調とそれを構成する投稿群を偽情報スコアで可視化し対策優先度を判断する技術を開発 AIが偽情報への対応策立案を支援する、次アクション提案とカウンター発信の効果検証を行うシミュレーション技術を確立 | <ul style="list-style-type: none"> 言論空間の主戦場がX(旧Twitter)から動画プラットフォームへと移行しており、他プラットフォームへの横展開にあたり機能差分の調整が課題 利用者のニーズ(多様な分析観点・リアルタイム性)を効率的に満たす共通基盤とカスタマイズの境界を明確に定義・標準化し製品化を推進 |
| 技12_ サン電子株式会社 | <ul style="list-style-type: none"> 複数の検知技術を統合し、異なる形式の判定結果を統合可能な形式に正規化し三段階判定を実装する統合判定ロジックを構築 判定結果の評価・検証を効率的に行うため、統合ダッシュボードや指標の自動算出・可視化に関する機能要件および非機能要件を整理 | <ul style="list-style-type: none"> 統合判定では重み付け調整により過検知／見逃しを切り替えられる一方、統合性能とユーザーニーズのバランスを取る判断基準が未整備であり、悪意ある編集や文脈操作を伴うコンテンツは単純な真偽判定ではリスク評価が難しいことが課題 統合判定ロジックの評価軸を整理し、ユースケース別に方針を選択可能な設計とするとともに、コンテキスト判別を高度化したリスク評価システムの構築に向けて検証を進める予定 |

① 技術開発主体における主たる成果等サマリ(4/4)

| | 主たる技術開発の成果 | 今後に向けた主な課題・展望 |
|--------------------|--|--|
| 技13_ NABLAS株式会社 | <ul style="list-style-type: none">電話・通話環境にフェイク音声の自動検知が組み込まれ、利用者が意識せずともなりすまし・詐欺から守られるシステムを構築自治体向けに、SNS情報から多面的にフェイク検出・真偽判定可能なアプリケーションにて発信者情報の信頼性担保技術を開発 | <ul style="list-style-type: none">最新のフェイク音声は高度化で検出が困難となり、単一検出モデルでは対応が困難。自治体では、体制未整備によりツールを有効活用できていない状況多様なケースに対応する検出モデルを構築し、最適モデルを自動選択できるシステム開発と、ユーザにおける検知段階から対応方針策定までを支援するサービス構築に取り組む予定 |
| 技14_ 日本電気株式会社 | <ul style="list-style-type: none">複合的なコンテンツの真偽判別支援技術を開発。主に総合真偽判別の高度化、様々な利用ユーザの意図に応じた分析の実現と SNS 等の分析対象情報の検知機能追加を実施 | <ul style="list-style-type: none">専門用語への対応、想定より長いレスポンスタイム、生成AIで加工・作成されたフェイク画像の判別精度不足が課題ユーザの利用頻度が高いと想定される領域を強化したソリューション開発(専門用語集対応含む)、偽誤情報判別ジョブの並列化、生成AIによる画像フェイク検知の精度向上に取り組む予定 |

② 研究・調査主体における主たる成果等サマリ(1/2)

| | 主たる研究・調査の成果 | 今後に向けた主な課題・展望 |
|--------------------------------|---|--|
| 研01_ 株式会社 新領域安全保障 研究所 | <ul style="list-style-type: none"> 英語圏偏重の既存研究を相対化し、日本固有の文脈に即した対策を立案するため、世界的な誤情報の研究動向を明らかにした その上で、日本の文脈を再現可能なLLMエージェントを用いた安全な検証環境の構築を実施し、コストの低い再現実験の手法を提示 | <ul style="list-style-type: none"> 非英語圏やローカルプラットフォームを含む網羅的な研究動向の分析へと発展させ、地理や言語、媒体の盲点を継続的に分析 比較用の人間データを整備し、シミュレーションの予測精度を評価・補正できる枠組みを確立し、また、より現実に近いダイナミクスを組み込むことが必要 |
| 研02_ 中央大学 | <ul style="list-style-type: none"> ウェブ調査により、偽・誤情報への接触状況と、それを規定する社会的属性や個人的属性、政治意識との関連を検討 サーベイ実験により、認知反射テストへの取り組みが熟慮を促すかどうか、また、それが偽・誤情報の判断を促進するかどうかを解明 | <ul style="list-style-type: none"> どうすれば熟慮が促進されるのかについても今後検討が必要 また、社会への不安や疎外感を煽る情報への注意喚起に加えて、信頼できる情報源をどう提示するか、その情報源に対する信頼をどう獲得するかという観点での検討が必要 |
| 研03_ 東京大学大学院 情報学環 | <ul style="list-style-type: none"> 生成AIを活用した偽・誤情報がデジタル空間でどのように生成・流通しているかを大規模データ分析により解明し、現在及び将来の脅威を体系化 先行研究を踏まえ設計したオンライン実験を行い、生成AI偽・誤情報を受容する日本固有の特徴を他国と比較しながら解明 | <ul style="list-style-type: none"> 実際の拡散構造やアルゴリズム影響は未解明であり、データアクセスと透明性の確保が必要 加えて、どの介入設計が判断精度を上げるのかについての検証や、行動レベルでの追跡、偽・誤情報に脆弱な層や介入効果がある層などに分けての設計、実環境での持続効果等に対する検証等が必要 |
| 研04_ 東京大学大学院 工学系研究科 | <ul style="list-style-type: none"> Webブラウザ上で動作する拡張機能を開発し、SNSにおける投稿者の過去のコミュニティノート付与履歴等を可視化するツールを実装 投稿者の信頼性情報を可視化することによって、情報の受け手であるユーザーの判断プロセスへ介入し、SNS利用者の自律的な判断を促すメカニズムを解明 | <ul style="list-style-type: none"> 本研究の結果を統計的に確かなものとするためには、より多くの対象者を用いた大規模な実証が必要 ブラウザ拡張機能に限らず、SNSプラットフォーム自体が標準機能としてこのような信頼性可視化システムを組み込むことを期待 |

② 研究・調査主体における主たる成果等サマリ(2/2)

| | 主たる研究・調査の成果 | 今後に向けた主な課題・展望 |
|--------------------|--|--|
| 研05_ 名古屋工業大学 | <ul style="list-style-type: none">誤情報対策のひとつである訂正行為に内在するリスクとして、誤情報そのものに触れていなくても、訂正情報に触れることで、新たに誤情報の影響を生み出しうる可能性に着目し、その心理的メカニズム等を明らかにした撤回表現に代替説明を付加することで、誤情報撤回効果が有意に低下することを確認 | <ul style="list-style-type: none">今後の研究において、誤情報撤回効果を生み出す詳細な認知メカニズムの解明や、因果推論以外の文脈や日本語環境以外での検証等を検討上記の課題を通して、誤情報撤回効果の理解をさらに深化させるとともに、訂正という主要な誤情報対策を、より効果的で副作用の少ないものへの発展を期待 |
| 技06_ 日本エンタープライズ | <ul style="list-style-type: none">SNS等において制作者の意図と異なる作品解釈に基づく批判が投稿され、急速に拡散するといった炎上事例について、X(旧Twitter)上のポストを対象に、投稿データを収集・分析専門家ヒアリング等を通して、炎上事案の構造やその推移に影響を与える要因を解明 | <ul style="list-style-type: none">炎上事案について、より多くの事案の調査を行い、さまざまな炎上パターンや鎮静化要因を解析して対策の精緻化を図る炎上事案に対する予防、早期探知と有効な介入を可能にするため、コンテンツごとの炎上構造を踏まえた早期探知・警告ツールの開発や、二次創作品等に関するガイドライン整備を推進 |

SNSにおける偽情報・真偽不明情報の市民参加型可視化・分析技術の開発・実証

| | | | |
|----------------|--|---------------------|--------------------|
| アプローチする課題・目指す姿 | <ul style="list-style-type: none"> アプローチする課題：2023年のX API有料化（月額約75万円～）に伴い、APIを無償で使うことを前提としていた国内の偽情報研究や対策ツール開発が停滞している社会的障壁の解決 目指す姿：無償公開されているコミュニティノートのデータを活用し、高額なデータ購入に頼らずとも市民・報道・研究者が協力して偽情報を可視化・分析できる、持続可能な情報流通環境の構築 | | |
| 技術区分 | 情報流通状況の可視化・分析技術 | 実施体制 (下線：技術開発主体) | 一般社団法人コード・フォー・ジャパン |
| 対象とするモデル種 | 文章 | | |

技術開発の取組・成果

【取組み】

- ダッシュボードの開発：コミュニティノートのデータを一元化し、真偽不明情報の全体像や詳細情報を可視化するダッシュボードを専門家向けにリリース
- データパイプライン構築：無償データ活用の「シングル」を含む3つの処理経路を開発。低コスト運用と、選挙や災害時の速報性を両立するインフラ基盤を確立

【成果】

- 限定公開と初期ユーザーの獲得：研究者や報道関係者、開発者、関係企業等に限定してツールを公開。授業や報道の調査等で実際に使いたいという評価をいただいた
- 技術検証とKPI目標の達成：可視化までのリードタイムを目標の2日以内に対して平均31時間30分でできることを検証

社会実装に係る取組・成果

【取組み】

- 連携ネットワークの構築：報道機関や研究者等の国内外のステークホルダーとの重層的な協力体制の構築を推進
- 寄付モデルの策定：多額の収入がなくても持続的に事業を継続するため、法人・個人の寄付により運営費を賄う独自のビジネスモデルを策定

【成果】

- 連携の成果：読売・西日本新聞の参議院選挙に関連する記事掲載、社会情報学会と計算社会科学会での研究報告、カナダ大使館との共催イベントなど既にツールを用いた分析を社会に広く発信（KPI6件に対して実績7件）
- ビジネスプランの策定：システム設計の工夫により運用費を200万まで抑制。起業への営業試行を通じて寄付型モデルの実現性を実証

技術開発及び社会実装にあたっての課題・展望

【課題】

- ユーザーからの要望：実際に利用したユーザーからさらに必要な具体的な機能に関する要望やUI・ユーザビリティの改善についての要望があった
- 継続的な運営：大規模のデータの処理を効率的に行うことで運用費用を抑制するとともに、運用費用を賄うための継続的な資金を得る必要がある

【展望】

- 追加機能などの開発：ユーザーから要望のあったナラティブの具体的な内容が分かる機能やトレンドを把握できる機能の開発、UIやユーザビリティの改善などを行い、一般向けにリリースする
- 成果と寄付の循環による持続的運用の確立：報道・研究等での社会実装成果を出して寄付獲得へ繋げ、得られた資金を分析支援等へ再投資する「持続的なエコシステム」を確立する

代表者コメント



一般社団法人コード・フォー・ジャパン 副代表理事 陣内一樹

本技術はこれまで少人数で開発してきましたが、本事業により開発を加速し、報道・研究者向けリリースを実現できました。さらに、国内外の連携も進み、可視化するための基盤ができたことは大きな成果です。今後は持続的な仕組みを整え、市民主体で偽・誤情報に向き合うこの技術をさらに成長させていきます。

音響透かしと音響フィンガープリントを用いた偽・誤情報対策クラウドシステムの開発・実証

| | | | |
|-----------------------|--|--|---|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> 音声生成AIの急進歩により真偽判別が難しく、導入インセンティブの弱さから社会実装も進みにくい。しかし詐欺電話対策などには即応性が不可欠だ。そのため、高精度かつリアルタイムな検知技術の開発と、普及を促進する仕組み作りが強く求められている。 生成AIや詐欺電話など巧妙化する偽情報に対し、音響透かし・フィンガープリントを活用した高精度かつ即時的な真偽判定技術を確立。放送局や通信・プラットフォーム事業者が、信頼ある情報を容易に発信・判定できる情報流通環境の実現を目指す。 | | |
| <p>技術区分</p> | <p>真正性保証・信頼性判断支援・改ざん検知技術（信頼性判断・改ざん検知／電子透かし）</p> | <p>実施体制 <small>（下線：技術開発主体）</small></p> | <p>エヴィクサー株式会社、TMI総合法律事務所（再委託先）、一般社団法人CiP協議会（再委託先）、株式会社IP Bridge（再委託先）、大坪太郎（再々委託先）</p> |
| <p>対象とするモデル種</p> | <p>音声、動画</p> | | |

技術開発の取組・成果

- ・EAF2024（Evixar Audio Forensicsの略称で、ここでは2024年度に本実証事業で開発したソリューション全体を指す）の精度向上と開発：AI判定エンジンの精度向上と対応モデルを拡充。併せてUI/UXやレポート機能の改善も実施。
- ・音響透かし・フィンガープリント活用のキャンペーン構築：登録素材でのキャンペーン機能実装により導入インセンティブを付与。Webベースの検出連携で手軽な実施を実現。
- ・リアルタイム透かし埋め込みと検出：ストリーミングやIP電話網での利用を想定し、既存の仕組みをストリーム処理可能なりリアルタイム機能として実装。

社会実装に係る取組・成果

- 「シーズ起点」と「多分野での同時並行PoC」を戦略の柱とし、以下の実証を通じて技術の有効性とビジネスモデルを検証した。
- ・行政：京都府
 - ・ODM提供：シヤチハタ株式会社
 - ・イベント：聴覚障害者のための世界規模の総合スポーツ競技大会、Bリーグ「アルバルク東京」のホームゲーム（トヨタ・モビリティ基金「Mobility for ALL」プロジェクトの実証実験）
数千人規模の同時アクセス下での安定稼働を実証し、ユーザ参加型の新たな価値提供を実現した。
 - ・連携模索：台湾ChungHwa Telecom社、株式会社radiko

技術開発及び社会実装にあたっての課題・展望

技術開発の観点：

- ・課題：AI生成技術の進化速度は速く、新たなモデルに常に対応し続けるための、継続的なアルゴリズム更新が不可欠である。
- ・対応：引き続き多様なAI音声のデータセットを拡充し、判定エンジンの学習と精度向上を推進。新モデルへの迅速な対応プロセスを構築する。

社会実装の観点：

- ・課題：偽情報対策という「守り」の価値に加え、導入企業の投資に見合う「攻め」の付加価値（事業メリット）の提示が不可欠である。また、業界ごとに異なるニーズに合わせたビジネスモデルの具体化も課題である。
- ・対応：PoCの成果を基に、キャンペーン機能による収益向上など、費用対効果を明確に訴求する。SaaSやODMといった、業界の特性に合わせた柔軟な提供形態と料金体系を確立し、本格導入を推進する。

代表者コメント



エヴィクサー株式会社
取締役COO
鈴木久晴

本事業を通じて、音響透かしと音響フィンガープリントを組み合わせた偽情報に対抗する技術を前進し、行政・放送・通信からエンターテインメントまで多様な現場でその有効性を実証できたことは、大きな成果です。
今後はパートナーとの連携をさらに強化し、本技術が安心・安全なデジタル社会のインフラとなるよう、その普及と発展に努めてまいります。

情報の真正性を可視化するC2PA技術を活用した偽・誤情報対策の開発・実証

| | | | |
|-----------------------|--|-----------------------------|--------------------------------|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> ・ 昨今、高度に加工された画像や動画を判別するには多くの人的コストを要している点に課題があり、報道機関を始めとするファクトチェック実施する事業者が存在する。 ・ 撮影の瞬間からデバイス情報・日時・位置情報等のメタデータを含む電子署名を付与し、改ざんの有無を確認できる形で記録する仕組みに加え、記録された情報を可視化することで撮影されたコンテンツの真正性を誰もが簡単に証明・検証できる社会基盤の構築を目指す。 | | |
| <p>技術区分</p> | <p>真正性保証・信頼性判断支援・改ざん検知技術</p> | <p>実施体制 (下線：技術開発主体)</p> | <p>NTTドコモビジネス(株)、(株)NTTドコモ</p> |
| <p>対象とするモダリティ</p> | <p>画像、動画</p> | | |

技術開発の取組・成果

- ・ 以下3点の技術開発を実施
 - ①コンテンツ取得時のメタデータの真実性チェック技術
 - ②標準仕様「C2PA(Coalition for Content Provenance and Authenticity)」に準拠した形式でのコンテンツへの情報付与
 - ③署名後データのファクトチェック支援技術
- ・ 開発した技術を用いて実際にファクトチェックを実施している協力会社との実証実験を実施
 - ・ 偽・誤情報の検知率 85%以上を達成
(選挙シナリオ：96.5% 災害シナリオ86.4%)
 - ・ ファクトチェックの稼働削減率 15%以上を達成
(選挙シナリオ：73.8% 災害シナリオ：31.1%)

社会実装に係る取組・成果

- ・ 想定されるユーザへのヒアリングの実施
 - ・ 開発技術の有効性、活用方法について確認
→画像・動画の真偽判定を含む業務における報道機関だけでなく
その他業界においても、活用シーンがあること確認
- ・ 社会実装されるにあたっての懸念を各業界について確認
→C2PAで取得した署名データの取り扱いや技術普及時の業界全体の開発動向や標準化に向けた動きについての意見の収集

技術開発及び社会実装にあたっての課題・展望

- ・ 【技術開発の観点における課題・展望】
 - ・ ファクトチェック時に使用する情報の信頼性の向上をさせる必要がある
 - ・ 開発ツールで完結する機能の拡張が必要
 - ・ 今後、実際のファクトチェック作業に適したUI/UXの実現をファクトチェック事業者とのヒアリング等も実施しながら実現可能か検討
- ・ 【社会実装の観点における課題・展望】
 - ・ 開発技術の普及に向けて情報の取り扱い方法や関連する各種業界との連携を行う必要がある
 - ・ 継続的な開発を行うとともに、各種業界への開発技術の有効性や普及に向けた調整を実施していく

代表者コメント



NTTドコモビジネス社
ソリューションサービス部
主査 角田 有希

本実証実験の結果、実際のファクトチェック作業においても、開発技術が作業負荷の削減や真偽判定結果の支援に大きく貢献できることが明らかとなった。一方で、社会実装に向けては依然として多くの課題が存在することも判明した。今後は、開発技術のさらなる有効化に向けた改修を進めるとともに、各種業界との連携や情報発信を積極的に実施していくことを検討している。

情報の真正性を可視化するC2PA技術を活用した偽・誤情報対策の開発・実証

| | | | |
|-----------------------|--|-----------------------------|--------------------------------|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> ・ 昨今、高度に加工された画像や動画を判別するには多くの人的コストを要している点に課題があり、報道機関を始めとするファクトチェック実施する事業者が存在する。 ・ 撮影の瞬間からデバイス情報・日時・位置情報等のメタデータを含む電子署名を付与し、改ざんの有無を確認できる形で記録する仕組みに加え、記録された情報を可視化することで撮影されたコンテンツの真正性を誰もが簡単に証明・検証できる社会基盤の構築を目指す。 | | |
| <p>技術区分</p> | <p>真正性保証・信頼性判断支援・改ざん検知技術</p> | <p>実施体制 (下線：技術開発主体)</p> | <p>NTTドコモビジネス(株)、(株)NTTドコモ</p> |
| <p>対象とするモダリティ</p> | <p>画像、動画</p> | | |

技術開発の取組・成果

- ・ 以下3点の技術開発を実施
 - ①コンテンツ取得時のメタデータの真実性チェック技術
 - ②標準仕様「C2PA(Coalition for Content Provenance and Authenticity)」に準拠した形式でのコンテンツへの情報付与
 - ③署名後データのファクトチェック支援技術
- ・ 開発した技術を用いて実際にファクトチェックを実施している協力会社との実証実験を実施
 - ・ 偽・誤情報の検知率 85%以上を達成
(選挙シナリオ：96.5% 災害シナリオ86.4%)
 - ・ ファクトチェックの稼働削減率 15%以上を達成
(選挙シナリオ：73.8% 災害シナリオ：31.1%)

社会実装に係る取組・成果

- ・ 想定されるユーザへのヒアリングの実施
 - ・ 開発技術の有効性、活用方法について確認
→画像・動画の真偽判定を含む業務における報道機関だけでなく、その他業界においても、活用シーンがあること確認
- ・ 社会実装されるにあたっての懸念を各業界について確認
→C2PAで取得した署名データの取り扱いや技術普及時の業界全体の開発動向や標準化に向けた動きについての意見の収集

技術開発及び社会実装にあたっての課題・展望

- ・ 【技術開発の観点における課題・展望】
 - ・ ファクトチェック時に使用する情報の信頼性の向上をさせる必要がある
 - ・ 開発ツールで完結する機能の拡張が必要
 - ・ 今後、実際のファクトチェック作業に適したUI/UXの実現をファクトチェック事業者とのヒアリング等も実施しながら実現可能か検討
- ・ 【社会実装の観点における課題・展望】
 - ・ 開発技術の普及に向けて情報の取り扱い方法や関連する各種業界との連携を行う必要がある
 - ・ 継続的な開発を行うとともに、各種業界への開発技術の有効性や普及に向けた調整を実施していく

代表者コメント



NTTドコモビジネス社
ソリューションサービス部
主査 角田 有希

本実証実験の結果、実際のファクトチェック作業においても、開発技術が作業負荷の削減や真偽判定結果の支援に大きく貢献できることが明らかとなった。一方で、社会実装に向けては依然として多くの課題が存在することも判明した。今後は、開発技術のさらなる有効化に向けた改修を進めるとともに、各種業界との連携や情報発信を積極的に実施していくことを検討している。

技04_Originator Profile技術研究組合 「Originator Profile」の開発と社会実装

| | | | |
|-----------------------|--|----------------------------|--|
| アプローチする課題・目指す姿 | <ul style="list-style-type: none"> インターネットでは、情報に責任を持つ組織が発した情報と、無責任に発信された情報を見分ける方法がない。不確かな情報に接触したユーザーの一部は、その情報を真実と誤認し詐欺的行為の被害にあふ、災害時に命に係わる偽情報が社会に混乱を招く、といった事態も発生しており、情報の発信者を正確に確認できる技術の開発が求められている。 発信する情報に責任を持つ様々な主体がWeb上で情報発信する際に、サイトやコンテンツにOPがつくことで、ユーザーが「どのような発信者による情報か」「コンテンツが発信者以外の第三者によって改ざんされていないか」を識別でき、ユーザーによる「偽・誤情報かどうか」の判断が容易になる。これによって、インターネット上で起きるアテンションエコノミーの抑制や、詐欺的行為による被害防止などを通じ、安心して利用できるインターネット環境の構築を目指す。 | | |
| 技術区分 | 真正性保証・信頼性判断支援・改ざん検知技術 | 実施体制 (下線：技術開発主体) | Originator Profile技術研究組合 WebDINO Japan、三菱総合研究所、電通総研、大日本印刷、他・OP技術研究組合組合員各社など |
| 対象とするモダリティ | 文章、画像 | | |

技術開発の取組・成果

- 本実証でメディアや自治体がOP技術とそのサイトに導入する際に必要となる**OPレジストリ、CAサーバなどの開発が完了**した。これにより、OP技術を実装したサイトや、記事などのコンテンツを、OPが付いた状態で公開することが可能となった。
- 広告にOPを付けて流通させる**基礎的なシステムが完成**した。
- 上記のサイト・コンテンツ（記事・広告）を閲覧しながら、ブラウザの拡張機能を利用すると、それぞれの情報発信者を確認し、途中で改ざんがなされていないことも確認ができるようになった。
 - OPレジストリ: OPレジストリ(Originator Profileレジストリ)は、Originator Profile (OP)を登録するサーバです。新しくOriginator Profileを発行する場合は、発行に必要な情報（企業であれば登記情報など存在確認ができる情報、ドメイン名等）とともにOPレジストリに登録します。
 - CAサーバ: CAサーバ(Content Attestation サーバ)とは、発信者がコンテンツに対して署名しCAという形式でOPを付与する操作を行うサーバです。例えば、Webコンテンツ管理システム(CMS)等から、CAサーバを呼び出すことでOPを付与することができます。

社会実装に係る取組・成果

- 本実証で愛知県、秋田魁新報、朝日新聞、産経新聞、時事通信、ジャパントイムズ、小学館、中国新聞、中日新聞、電通、鳥取県、福島民友新聞、北海道新聞、北國新聞、毎日新聞、宮崎日日新聞、読売新聞などの**公開WebサイトにOP技術が導入され、OPが社会実装できるレベルに達したことが確認**できた。
- W3Cの年次総会TPAC2025において、**OPの技術仕様をグローバルに公開、発表**を行い、**国際標準化のために技術理解を求める場の確保に成功**した。
- サントリーウエルネス、資生堂ジャパン、パナソニック コネクトといった民間企業、愛知県、鳥取県といった自治体など、**広告の社会的責任を重視する広告主のOP付きデジタル広告を流通**させた。
- 上記広告は、国内メディアの公開用Webサイトに掲載され、広告取引においてOPを利用する基礎実証を成功**させることができた。

技術開発及び社会実装にあたっての課題・展望

- OPのWebサイトへの実装を浸透させるためには、**サイト運営者が簡単にOP技術を実装**できることが必要で、Webサイトが採用している**さまざまなCMSへの対応が課題**となる。本実証においては、最も普及しているCMSであるWordPress (WP) に**OP登録、CAデータ発行を自動化するプラグインを提供**したが、自治体やメディアにおいてWPを採用している割合が少ない。本実証中には、メディア企業の利用が多いCMS開発企業と協力関係を築いたが、今後、これらの取り組みをさらに加速させ、簡単に実装ができる状況を目指す。
 - ただし、**すべてのCMSでOP対応プラグインを開発することは非現実的**であるため、**コンテンツを配信する際にOPを第三者サーバから付与するSaaSサービスの開発に着手**し、一部事業者の検証系サイトへの実装テストを行なった。この仕組みを活用することで、より多くのWebサイトが比較的容易にOP実装することが可能となることから、利用方法の検討を進めていく。
 - 一般ユーザーが閲覧するWebサイト・コンテンツでOPを確認するための仕組みを検討・開発していく。現在、実証用に用意しているのはブラウザ拡張機能だが、インストールなどユーザーに能動的なアクションが必要になるため、OP実装されたサイトにアクセスした際に、**ユーザーが行動しなくても、ブラウザに発信者の情報が表示されるUI**を検討した。今後、ブラウザに同様の機能を搭載する道筋を計画していく。
 - OP付き広告取引について、本実証で**実環境での広告配信が成功**した。**OP有無による広告配信管理やトレーサビリティ確保の検討**も開始しており、今後、これらの実現に取り組んでいく。
- CMS: Content Management Systemの略でWebサーバでコンテンツを発行するシステム
 ● OP対応プラグイン: CMSが発行するWebページに、発信者とページの作成者を自動的に登録できるようにするシステムのこと。
 ● UI: User Interfaceの略、利用者がシステムを動かすために使う画面のデザインのこと

代表者コメント



Originator Profile
技術研究組合
理事長 村井 純

偽・誤情報・アドフランドなど事態そのものは急速に深刻化し、民主主義の脅威として社会の安全を脅かすことが一層懸念されています。本実証では、OP技術の開発がさらに進化し、その国際標準化の可能性も視野に入ってきました。

Web業界、コンテンツ業界、オンライン広告業界、ブラウザベンダーなどにも働きかけ、世界中の人々が安心して利用できる健全なインターネットの発展に貢献することを目指します。私たち OP-CIP の取り組みに、どうぞ注目ください。

偽・誤情報サンドボックスを活用した実践的ゲーム型プレバンキング技術の開発・実証

| | | | |
|-----------------------|--|-----------------------------|--------------------------------|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> 本事業が取り組む課題は三点である。第一に、偽・誤情報が本物と見分けが付きにくくなると同時に低コスト化し、短時間で大量に流通し得る点。第二に、訂正情報の提示だけでは限界があり、受け手に「疑い、確かめ、拡散しない」判断行動まで含む実践的リテラシーが求められる点。第三に、従来の受動型学習では構造理解と実践、さらに反復による定着が難しい点である。 目指す姿は、受け手が未知の事例でも「疑い、確かめ、拡散しない」判断行動を継続できる状態を広げることである。そのために、作り手視点を含む体験型学習と、短時間で反復可能な教材・運用パッケージを整備し、効果測定と改善サイクルを確立する。さらに、言語・文化・制度差に対応できる国際展開の基盤形成を進める。 | | |
| <p>技術区分</p> | <p>情報の拡散防止・無効化技術</p> | <p>実施体制 (下線：技術開発主体)</p> | <p>株式会社Classroom Adventure</p> |
| <p>対象とするモデル種</p> | <p>文章、画像</p> | | |

技術開発の取組・成果

【取組】

- 偽・誤情報サンドボックスの開発：受講者が偽・誤情報を「作る側」として仮想環境で生成し、誤認が生じる条件や心理的誘導の仕組みを体験的に理解したうえで、確認行動（立ち止まる・出所を確かめる・拡散を控える）へ転移させるサンドボックス環境を設計・実装。
- 反復型学習設計の構築：偽・誤情報の典型パターンを学習単位として整理し、事例提示→実践→解説→振り返りの一貫した学習導線を標準化。
- 安全対策と運用・管理システムの整備：教材内生成物の外部流通リスクに対し、電子透かし・ワードフィルタ・コピーガード・生成物解析AIによる不適切性チェック等の多層防御を実装。加えて、受講者管理・進捗把握・結果出力等の管理機能を備え、教員・研修担当者が現場で運用できる体制を整備した。

【成果】

事前・事後の同一条件設問により学習効果を測定した結果、参加者の約55%でスコアが約37%向上し、受講直後の判断の質の改善が確認された。さらに、体験の満足度・理解実感・有関感を問う調査では、約98%が「満足」以上と回答し、単発イベントに留まらず教材として継続利用に耐えうる体験価値の成立が示唆された。

※サンドボックス：実際のインターネット空間とは切り離された安全な検証環境であり、利用者が偽・誤情報や生成AIの出力を実験的に作成・検証できる閉鎖的な学習環境

社会実装に係る取組・成果

【社会実装の取組】

- 学校・自治体・企業等の導入現場を想定し、ニーズ調査・関係者ヒアリングを通じて、導入意思決定に影響する条件を整理した。
- 個人情報・画像取り扱い・生成物の外部持ち出し・不適切利用・権利侵害等の論点を洗い出し、同意設計・運用ルール・透かし／制限・ログの取り扱いを含む対応方針を明確化した。
- 導入前のフィルタリング等の確認や実施手順、実施後レポート等を含む導入パッケージの整備を進め、反復利用を前提とした運用の標準化に着手した。

【社会実装の成果】

導入は学習効果だけで決まらず、「授業内で完結し担当者負担が小さいこと」「安全対策を説明できること」「反復利用時の費用対効果が示せること」が意思決定の主要因である点を具体化できた。
また、個人情報・生成物取り扱いの主要論点を整理し、同意・ルール・制限・透かし・監査ログを組み合わせた実装・運用方針を導入先に提示可能な形に整備した。

技術開発及び社会実装にあたっての課題・展望

【今後の課題】

本技術は、生成体験を通じて判断行動へ転移させる設計である一方、教材内で生成された文章・画像が教材外へ持ち出されることによる誤認・拡散や二次利用のリスクへの対応が中核課題である。また生成AIの安全機能には、迂回表現・文脈依存・多言語等による抜け道が残りのため、教育上必要な生成体験を成立させつつ危険領域へ逸脱しない教材特有の安全に使うための制御設計が継続課題となる。さらに実環境では、偽・誤情報が短尺動画・音声なりすまし等へ移行しつつあり、教材側のマルチモーダル対応が不可欠である。一方で、ログ取得や運用統制はプライバシー・社内規程・審査対応と接続するため、導入先が説明可能なデータ範囲・権限・保存・削除の設計を明確化する必要がある。加えて国際展開では、各国の事例・制度・文化差に合わせて共同で更新し続ける体制づくりと、プレバンキングの反復実施を阻害しない運用設計が課題となる。

【展望】

生成物の外部流通リスクに対しては、可視な電子透かし中心から、機械可読な来歴情報＋不可視な電子透かし＋運用統制を組み合わせた多層設計へ段階的に移行し、加工後も検出可能性と説明可能性を高める。安全に使うための制御については、シナリオ制約と自由度の段階設計により「教育上必要な危うさ」を管理しつつ、生成前後の判定と監査ログ、管理者の例外運用を統合して安全性・統制・説明可能性を一体で担保する。

マルチモーダル対応は、音声・動画へ段階的拡張し、悪用リスクと運用負荷を踏まえた限定的導入から実環境への近接を図る。社会実装面では、セキュリティ／プライバシー設計と事故時対応フローを標準化した導入パッケージを整備し、導入審査で説明可能な状態を完成させる。国際展開は、生成・安全制御・学習ログ・管理機能を共通化し、各国に合わせて「共通基盤＋現地提供設計」で分業し、将来的には外部システム連携機能・開発部品化により、既存の学習管理システム等への組込みを可能にして導入スケールと継続利用を同時に実現する。

代表者コメント



株式会社Classroom Adventure
代表取締役 今井善太郎

本事業で開発・実証した偽・誤情報サンドボックスは、私たちにとって大きな学びになりました。学習者が強く引き付けられ、偽情報を「自分事」として捉え直すことで、出所確認・根拠点検・拡散抑制といったプレバンキング行動が、知識ではなく実践として身につけ始めていることを実感しています。ゲーミフィケーションによって受け身の学びから抜け出したことも大きく、生成AIの全面解放に慎重な学校現場でも「この形なら検討できる」と言っていただけなのは重要な成果でした。

一方で、情報環境の変化は想像以上に速く、私たち開発側にも同じスピードで更新し続ける体制が求められます。次年度以降は、画像だけでなく音声・動画を含むマルチモーダル対応を進め、現実のリスクにより近い形で学べる環境を整えていきます。同時に、「逆に偽情報の作り方を教えてしまうのではないか」という懸念には真正面から向き合い、透かし・検知・制約・監査といった多層的な対策を継続的に強化しながら、安全性と学びを両立させていきます。

偽・誤情報およびデバンキング情報拡散のシミュレーション技術の開発・実証

| | | | |
|-----------------------|---|-----------------------------|----------------------------|
| <p>アプローチする課題・目指す姿</p> | <p>近年、ソーシャルメディアや生成AIの発展により偽・誤情報が急速に増加している。一方、国内では標準化された仕組みが存在せず、組織を跨いだ知見の共有が十分に行われていないほか、既存の商用プラットフォームでは日本語・偽・誤情報に特化した分析精度や技術更新に課題がある。本開発技術は、偽・誤情報及びデバンキング情報におけるナラティブ分布の可視化や拡散予測を通じて、効果的な発信ノードやネットワーククラスターを特定することにより、限られた人的資源の中でも効率的な対策立案を可能とする。加えて、OpenCTIとの相互運用性を確保し、日本の脅威環境に即した統合的な偽・誤情報対策基盤の構築を目指す。</p> | | |
| <p>技術区分</p> | <p>情報流通状況の可視化・分析技術、情報の拡散防止・無効化技術</p> | <p>実施体制 (下線：技術開発主体)</p> | <p>(株) コンステラセキュリティジャパン</p> |
| <p>対象とするモダリティ</p> | <p>文章、画像、動画</p> | | |

技術開発の取組・成果

- 脅威情報の共有・管理基盤として欧米で広く利用されているOpenCTIを基盤とし、国際標準に準拠した偽・誤情報分析プラットフォームを開発した
- 複数のプラットフォームに対応したデータ格納基盤を整備し、取得データを脅威情報を記述するための国際標準であるSTIX形式へ自動変換することで、単一プラットフォーム上でのデータの一元管理を可能とした
- 集積されたデータに対して自動で分析結果や評価を付与するエンリッチメント機構を実装し、偽・誤情報の早期検知を支援する機能群を実装した
- 日本語環境におけるSNS上の情報拡散をモデル化し、同モデルを用いて偽・誤情報およびファクトチェック情報の拡散を予測・シミュレーションする機能を実装した。これにより、偽・誤情報の影響度や対応の有効性を予測し、効果的な対策立案およびリソースの最適配分を支援する

社会実装に係る取組・成果

- ステークホルダーへ開発段階ごとにシステム紹介及びヒアリングを実施し、実務的観点から本開発技術の有効性を検証するとともにシステム機能に関する意見を収集した
- 社内アナリストが開発技術を用いて偽・誤情報を分析し、その結果を当社の洞察及び推奨事項とともに月次レポートとして顧客・取引先に配布した。これにより、社会全体における偽・誤情報対策への意識啓発を行うとともに、本開発技術の有用性を示した
- 令和7年富士調査研究会同へ出展し、本開発技術の紹介及び意見交換を行った
- 顧客企業に対して本開発技術を用いたサービスのPoCを実施し、来年度への契約獲得に向けた提案を行った

技術開発及び社会実装にあたっての課題・展望

- 外部基盤の仕様に起因する制約を緩和するため、データ処理方式の改善や処理基盤の強化によりデータ処理性能を緩和するとともに、OpenCTI開発元との協議や外部ダッシュボードの活用、分析結果の表示画面のカスタマイズなどの代替手段を検討し可視化形式の柔軟性を補完する
- 公開情報に基づく分析には、アカウント属性等の推定精度に限界があるため、社会学・心理学的知見や関連情報を統合することで推定精度の向上を図る
- 製品化に向けてスケーラビリティやユーザーインターフェース等の非機能要件の定義を進める
- 市場拡大に向けて幅広い組織との連携を強化し、各種組織にとって適切な利用用途の検討を進める
- 実運用を想定したエンドユーザーへの教育やサポート体制を確立するとともにユースケースを整理する
- 持続的な運用体制を確保するため運用・保守体制を整備し、顧客からのフィードバックを踏まえ継続的な改善を図る

代表者コメント



株式会社コンステラセキュリティジャパン 代表取締役社長 宮村信男

偽・誤情報及びデバンキング情報の拡散は、安全保障の観点からも極めて重要な課題である。本事業は挑戦的なテーマであったが、想定していた主要機能は概ね計画どおり実現できたと認識している。今後は、処理性能やスケーラビリティ、運用上の安全性を一層高め、より幅広い実環境で活用可能な技術として成長させたい。

デジタル情報空間における多層的意味解析と拡散ダイナミクス解明プラットフォームの開発・実証

| | | | |
|-----------------------|--|-----------------------------|--------------------|
| <p>アプローチする課題・目指す姿</p> | <p>•SNS上では、論理的誤謬や印象操作を伴うミスリーディング情報が増加している一方、その構造や拡散メカニズムを俯瞰的・定量的に把握する技術は要素技術の研究レベルにとどまっており、かつ従来の真偽判別技術だけでは対応が困難である。 •本開発・実証では、大規模言語モデルによる深層の意味解釈と情報拡散分析を統合し、ミスリーディング情報の構築論理・影響を包括的に可視化する。これにより、災害時や選挙時などの重要局面において、偽・誤情報の早期検知と迅速な是正が機能する予防型の情報空間を実現し、健全な言論環境の維持を目指す。</p> | | |
| <p>技術区分</p> | <p>情報流通状況の可視化・分析</p> | <p>実施体制 (下線：技術開発主体)</p> | <p>(株)TDAI Lab</p> |
| <p>対象とするモーダル種</p> | <p>文章</p> | | |

技術開発の取組・成果

- 意味的解釈エンジンを中核とした情報拡散分析技術を開発し、言説のナラティブ分類、論点整理、可視化を自動で行う分析基盤を構築した。これにより、数千件規模のコメントデータに対しても短時間で全体像を把握可能とした。
- 意味的解釈エンジンでは、単純なキーワード集計ではなく、言説の意図や論理構造を捉える分類モジュールを実装した。実証用ベンチマークを用いた評価により、**既存研究を上回る性能**を達成した。
- 事前にSNSデータを準備せず、特定テーマに関する情報流通状況を把握するデータなし情報拡散分析機能を実装した。これにより、速報性が求められる調査初期段階において、ファクトチェックや深掘り対象となる言説の洗い出しを効率化できることを確認した。
- ウェブ上の公開情報を基に、特定言説の信頼性や根拠状況を整理するファクトチェック分析機能を開発した。分析処理は実務上十分な速度で実行可能であり、自治体等における偽・誤情報対応の初動支援として有効であることが示された。

社会実装に係る取組・成果

- 報道機関、行政機関、企業へのヒアリングを実施。これにより、実利用者が期待するユースケースや、導入時に懸念される点を把握し、得られたフィードバックをもとに、分析方針の調整、UIの改善、結果提示の可視化方法の拡張といった機能改修を行い、懸念解消と利用現場に即した最適化を進めた。
- 実務者ヒアリングを通じ、本技術による分析結果は最終判断を代替するものではなく、根拠情報や分類定義を明示した「判断補助情報」として提供することが社会実装上適切であることを整理し、今後の精度向上および説明性強化に向けた改善方針を得た。
- 社会的インパクトの大きい社会問題など4テーマについて本開発技術を利用した分析記事を**メディアへ寄稿**。
- サービスプロバイダーへの**API連携**を通じた自治体へのサービス提供体制を実現。

技術開発及び社会実装にあたっての課題・展望

- 透明性の確保
AIの分析結果をそのまま実務で利用するのではなく、根拠となった情報源や分類ルール、分析時期を明示した「判断材料」として提示する設計が不可欠。
- 現場に即したカスタマイズ
分析の切り口は組織の業務ごとに異なるため、実務者が自身の意図を柔軟に反映できる操作性の向上が求められる。
- 分析基盤の高度化
今後はテキストだけでなく、画像や動画も対象としたマルチモーダル分析へ拡張することを検討。

代表者コメント



株式会社TDAI Lab
代表取締役 福馬智生

本事業に採択され、大変光栄に存じます。本事業では、真偽の判断が困難な論理的誤謬やプロパガンダ技法に基づくSNS上のコメントについて、その展開や波及の状況を分析し、リスク等をレポート化するシステムを構築しました。今後は、偽・誤情報対策にとどまらず、認知戦を含むさまざまなシナリオへの対応や、各事業者との連携を推進していきたいと考えております。

SNSユーザー支援を中核とした偽・誤情報対策技術の開発・実証

| | | | |
|----------------|---|----------------------|--|
| アプローチする課題・目指す姿 | SNS上では偽・誤情報が巧妙に投稿・拡散される一方、ユーザーが自ら検証する習慣・ツールが不足し、ディープフェイクの急速な進化や訂正情報の流通不足が深刻化している。本事業では、SNSユーザーの主体的な情報検証を支援するファクトチェック支援アプリ「シラベル」と訂正主体向けSNSモニタリングツールを開発・実証し、社会全体で偽・誤情報に対処できる健全な情報流通環境の構築を目指す。 | | |
| 技術区分 | コンテンツの真偽判別支援技術 | 実施体制 (下線: 技術開発主体) | 株式会社データグリッド、株式会社AIBOS、AIBOS Uganda Co., Ltd. |
| 対象とするモジュール種 | 文章、画像、音声、動画 | | |

技術開発の取組・成果

- **技術1 ディープフェイク検出:** 最新の生成AIにより作成された偽の画像・動画・音声を自動判定し、検出精度 93% を達成 (目標90%)
- **技術2 根拠情報の自動検索:** SNS上の真偽不明な言説に対し信頼性の高い複数の情報源から根拠を自動提示し、適合率 92% を達成 (目標85%)
- **技術3 画像・動画の出所検証:** 類似画像・動画検索により初出元や過去の利用文脈をWeb上から特定・提示し、適合率 90% を達成 (目標80%)
- **技術4 情報撮取行動の可視化:** SNS上でのユーザーの情報閲覧傾向や偽情報への反応・拡散の履歴を分析して提示し、精度(QWK) 0.62 を達成
- **技術5 リスク評価・拡散予測:** SNS上の真偽不明言説の社会的影響度を定量的に評価・対応優先順位付けを支援し、精度(QWK) 0.84 を達成

実証の取組・成果

- **検証1 SNSユーザー向けアルファ版テスト:** 59名のモニターにアプリを試用いただき、ファクトチェック支援の有用性とリテラシー向上効果を検証。満足度は10段階中 7.81 と高評価を獲得し、利用前後のリテラシーテストでは平均 +3.3点の統計的に有意な向上を確認。
- **検証2 SNSユーザー向けベータ版テスト:** 上記アプリの改善版を一般公開し、普及ポテンシャルとユーザー評価を検証。ポジティブ評価が 76.3% と高い支持を得ており、リリース後1ヶ月で約200名のユーザーを獲得。ファクトチェックニーズを確認するとともに、認知度向上に向けたメディアでのPRが今後重要となると認識。
- **検証3 訂正主体向けテスト:** 報道機関にSNSモニタリングツールをテスト利用いただき、実務への適合性を検証。ユーザービリティ評価は10段階中 7 を獲得し、早期探知・拡散分析・取材効率化など実践的なユースケースを検証。

技術開発及び社会実装にあたっての課題・展望

- 技術開発面では、①生成AI技術の急速な進化への追従 (本年度構築したMLOpsパイプラインの自動化率をさらに向上させ、新種ディープフェイクへの対応リードタイムを短縮)、②分析コストの最適化 (軽量モデルの活用やキャッシュ機構の導入)、③検出対象メディアの拡大 (ショート動画やライブ配信への対応)、④多言語対応の高度化 (越境的な偽情報キャンペーンに対応するための多言語根拠検索の精度向上) が主要課題として明確化された。
- 社会実装面では、①SNSユーザーの獲得・定着 (ゲーミフィケーションなどにより日常の行動フローに溶け込むUX設計を実現)、②訂正情報の流通促進 (「シラベル」を訂正情報の流通チャネルとして活用し、訂正情報の到達率を向上)、③訂正主体のツール活用促進 (パートナー企業との協業による既存顧客基盤を活用した間接的な導入促進)、④法規制・倫理面への対応 (あくまで「検証支援」として最終判断はユーザーに委ねる設計思想の堅持) に取り組む。

代表者コメント



株式会社データグリッド
代表取締役
岡田 侑貴

SNS上の偽・誤情報に対抗する最大の鍵は、私たち一人一人の情報リテラシーにあると考えています。生成AIで巧妙な情報が溢れる時代だからこそ、情報の真偽を丁寧に見極める慎重かつ冷静な視点が欠かせません。当社は本事業の取組みを通じて、社会全体のリテラシー向上に貢献します。

放送波を活用した災害時における偽・誤情報対策技術の開発・実証

| | | | |
|-----------------------|--|------------------------------|---|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> 災害時には通信環境の途絶により情報不足に陥ることが多く、わずかな情報に避難住民や災害対策支援者が過敏に反応しやすくなる。そのような状況下においては文脈から切り離された情報が流入し、真偽を確認できない不安から誤解や憶測が連鎖的に広がり、被災地全体の混乱を助長するリスクが高まる。 災害時などインターネットが利用できない環境下においても、放送波とブロックチェーンを活用することにより偽・誤情報の拡散を抑制し、避難者や災害対策支援者が信頼できる情報に基づいて判断できる環境の構築を目指す。 | | |
| <p>技術区分</p> | <p>情報の拡散防止・無効化技術、真正性保証・信頼性判断支援技術</p> | <p>実施体制 (下線: 技術開発主体)</p> | <p>関西テレビソフトウェア(株)、関西テレビ放送(株)、(株)アトラクター、(株)ベリサーブ、(株)Opening Line</p> |
| <p>対象とするモジュール種</p> | <p>画像、動画、文章 (文章はメタデータ・位置情報として活用)</p> | | |

技術開発の取組・成果

- 前年度実証に引き続き、証明書発行サブシステムと証明書検証サブシステムを開発。実運用を視野に操作性や速度パフォーマンスを大幅に改善。また、両システム上で稼働する「防災コンパス」と「ファクト注釈コード (FAコード: Fact Annotation Code)」を実装。
- 防災コンパスでは、真正性が保証された情報をコンパス上にプロットし、災害時に拡散する偽・誤情報に惑わされない判断を支援。
- ファクト注釈コードでは、拡散された画像に対するファクトチェックの存在有無が確認できる、かつ軽微な変更が画像に加えられた場合でも検知可能な視認性の高いコードを考案。

社会実装に係る取組・成果

- 放送に重畳させたデータを、放送局内設備を經由して有線で受信機に配信。放送マスタ管理システムに対して想定される影響の有無および運用上の課題を確認。
- ブロックチェーンを活用した単方向で受信したデータの信頼性確認手法を体系化し、第三者検証企業によるセキュリティ評価を実施。
- ブロックチェーン活用時に必要となる暗号資産について、代払い事業者を介在させることで、ユーザーの主体性を維持しつつ、会計面および税制面での運用負担の低減が可能であることを確認。

技術開発及び社会実装にあたっての課題・展望

- 社会実装に向けて、実電波を用いたフィールド実証が不可欠となる。段階的な実電波技術検証の実施のために、技術的および運用的な課題の整理を進める。
- 放送データを活用するためには、受信機側におけるデータ蓄積から検証までを安定稼働させる実装設計が重要となる。普及を見据え、低コストでの量産体制に必要な最小限の機能に絞った設計を行う。
- ユーザーの導入条件や予算感に応じて、放送と通信の両方の経路から受信可能なハイブリッド構成の設計。また、災害時に限らず、平時活用を前提とした運用モデルの構築を目指す。

代表者コメント



関西テレビソフトウェア(株)
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

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、継続的な開発・実証を重ね、実用化を目指します。

ストリーミング動画コンテンツの真偽検証支援ツールの開発・実証

| | | | |
|-----------------------|---|-----------------------------|--------------------------------------|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> ストリーミング動画は、社会的・政治的な情報流通の主要なチャネルとなりつつあり、公共的な議論や有権者の意思形成にも多大な影響を及ぼしている。特に動画形式による主張・評論・告発といったコンテンツは、感情的訴求力が高く、視聴者の印象形成に大きな役割を果たしている。しかし、ストリーミング動画コンテンツは検証コストが相対的に高く、多くのユーザーが利用する情報空間の健全性確保に、真偽検証を支援するツールの開発提供が寄与する可能性がある。 ユーザーベースが大きく、しかしデータへのアクセスが難しいストリーミング動画コンテンツに含まれる真偽の把握が困難なプラットフォームについて、広く対応した真偽検証支援ツールを提供。これにより、上記サービスの利用者が真偽不明の情報に早期対応・判断できる健全な情報流通環境の構築を目指す。【デバンキング対象の拡張】 | | |
| <p>技術区分</p> | <p>コンテンツの真偽判別支援技術、改ざん検知技術</p> | <p>実施体制 (下線：技術開発主体)</p> | <p>SEARCHLIGHT株式会社、AIQVE ONE株式会社</p> |
| <p>対象とするモデル種</p> | <p>動画</p> | | |

技術開発の取組・成果

- ストリーミング動画コンテンツの解析を目的とし、「動画内容の構造化」と「根拠情報の提示」を行う真偽検証支援技術を開発しました。大規模言語モデル（Gemini 3 Flash等）を活用し、動画内の主張や検証のためのエビデンスを、論拠強度を加味して収集・自動照合するシステムを構築したほか、実在しないURLを排除する出典照合の高度化も行いました。
- 主要報道機関等での実証実験において、検証業務の負荷軽減と効率化に対する有効性を確認しました。
- また、上記技術の広告・ブランドセーフティ領域や企業の炎上監視サービスへの応用を行いました。

社会実装に係る取組・成果

- 「AIが真偽を断定する」のではなく人間の判断を支援するという思想に基づき、報道機関等と対話を行い検閲リスクを回避する合意形成を図りました。
- また、マルチテナント対応のSaaS基盤の構築や、安全性検証（Red Teaming）を通じ、組織導入に耐えうる信頼性を確保しました。
- 成果として、2度の国政選挙において主要報道機関での試験導入を果たし有効性を実証したほか、広告・ブランドセーフティ領域や企業の炎上監視サービスへとビジネス展開を進め、持続可能な社会インフラとしての基盤構築のきっかけとなりました。

技術開発及び社会実装にあたっての課題・展望

【課題】

技術面では、陰謀論など「存在しない事実」を見抜く技術の確立と、外資プラットフォームに依存しない冗長性の確保が課題です。社会実装面では、対策市場が未成熟なため持続可能なビジネスモデル構築が急務です。社会実装にあたりベーシックなニーズは確認できたものの、平時での利用定着や広範な潜在顧客への販売チャネル強化が求められています。

【展望・対応事項】

今後は「プロ支援から社会インフラ」を掲げ改善を進めます。技術面では、主張の論理的飛躍を可視化する「論拠強度検証」や、AIが反証を生成する「ゴーストファクトチェック」機能を実装し検証精度を向上させます。社会実装面では、広告主へアプローチしブランドセーフティ市場を開拓します。SaaS基盤を活かして一般企業へ横展開して収益基盤を確立するとともに、将来的には海外での利用を目指し、持続可能な対策エコシステムの構築を目指します。

代表者コメント



SEARCHLIGHT株式会社
代表取締役
瀬戸 亮

実証期間が短く、機能面でも社会実装面でも非常に難しい状況でしたが、参院選と衆院選という大きいイベントが発生したことにより潜在的ニーズが顕在化し、多くのフィードバックを賜ることで、とても実りのある開発・実証となりました。

画像・動画を中心としたSNS上の投稿の真偽判定システムの開発・実証

| | | | |
|-----------------------|---|-----------------------------|----------------------|
| <p>アプローチする課題・目指す姿</p> | <p>• Xを始めSNSは偽・誤情報の主戦場となっており、社会に大きな影響を与えている。この対策として情報の真偽検証が挙げられるが、現状は多角的な視点を要するマニュアル作業であり、膨大な工数と高度な専門性が求められるほか、昨今の生成AI技術の進歩によりその判定はより難しくなっている。膨大な偽・誤情報の中から、どの情報を対象に、いかに効率よく、標準的に、高度に検知するかが、偽・誤情報対策の鍵となっている。</p> <p>• 本実証事業では、SNS空間で拡散される①偽・誤情報・論調(ナラティブ)を可視化し、真偽判別すべき情報の見極めと、②その判別の実施、及び③対策案の立案を支援するシステムを開発する。これにより、プラットフォーム事業者や一般ユーザが早期対応・判断できる健全な情報流通環境の構築を目指す。</p> | | |
| <p>技術区分</p> | <p>コンテンツの真偽判別支援技術、改ざん検知技術</p> | <p>実施体制 (下線:技術開発主体)</p> | <p>Sakana AI株式会社</p> |
| <p>対象とするモダリティ</p> | <p>文章、画像、音声、動画</p> | | |

技術開発の取組・成果

1. X上のナラティブ（論調）とそれを構成するXの投稿及びその偽・誤情報スコアの可視化
 - ナラティブを一覧化し、対策を打つべき偽・誤情報の優先度付けを可能とした（**効率化**）
2. 偽・誤情報の真偽判定システムの構築
 - 動画と画像を含む実用的な真偽判定システムの構築と、それを評価するために実際のXの投稿から収集したベンチマークを整備し、平均84%の検知精度を実現（**効率化、高度化**）
3. 偽・誤情報対策実施者が処置すべき対応策の立案
 - AIによる次のアクションの提案ならびにカウンター発信の効果検証を可能とするシミュレーション技術を確立（**効率化、標準化**）

社会実装に係る取組・成果

- 最前線の実務者（メディア、ファクトチェック有識者団体）との連携により、現場ワークフローに即した真に実効性のあるシステム要件を確立
- 国家安全保障に係るインテリジェンスの専門家である**中曽根平和研究所 情報空間のリスク研究会**様によるレビューを経て、その有用性を確認
- 普及活動を通じた導入見込み先の拡大
- 市場ニーズの多様性を踏まえ、「共通基盤」と「カスタマイズ」を組み合わせたハイブリッドな提供モデルを策定し、ビジネスモデルの具現化を実現

技術開発及び社会実装にあたっての課題・展望

【X以外のプラットフォームへの展開】

- 本システムのコア機能は、X以外のプラットフォームへも横展開が可能なアーキテクチャとなっている。実装にあたっては、プラットフォームごとのAPI仕様やデータ制約に起因する機能差分を考慮し、導入効果とコストのバランスを精査した上で、顧客ニーズに合わせた最適な適用範囲を定義していく。

【導入先ごとのカスタマイズ】

- 事業開始当初の想定を超え、幅広い事業者から関心が寄せられた結果、適用領域が拡大した。今後は、多様なニーズを効率的に満たすための、「共通基盤」と「カスタマイズ」の境界を明確に定義・標準化した上で、スケーラブルなプロダクトとしての製品化・商用化を推進する。

代表者コメント



Sakana AI
事業開発本部長
谷口博基

生成AIにより偽・誤情報の脅威が深刻化する中、本事業では幅広い有識者と密に連携し、可視化・判定・対策を一気通貫で支援する技術を開発し、情報分析の専門家より高い評価を得ることができました。今後は、実証で得られた知見を基に「社会を守るインフラ」としての製品化を推進し、信頼できるデジタル空間の実現に貢献してまいります。

多元統合型偽・誤情報検出技術の開発・実証

| | | | |
|-----------------------|--|-----------------------------|------------------------------|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> 近年、偽・誤情報リスクが急拡大する中、従来の単一技術では誤検知や精度低下といった限界が生じている。そのため、複数の検知技術を統合し、総合的な信頼性判断を可能にする新たなアプローチが不可欠となっている。 複数の改ざん・偽造検知技術を統合したシステムを構築し、統合検知結果を活用した信頼性判断支援技術を確認する。さらに、これらを用いた信頼性判断支援サービスの社会実装に向けた基盤を構築することを目指す。 | | |
| <p>技術区分</p> | <p>コンテンツの真正性保証・信頼性判断支援技術</p> | <p>実施体制 (下線：技術開発主体)</p> | <p>サン電子株式会社、サイバーコマンド株式会社</p> |
| <p>対象とするモード種</p> | <p>画像、音声、動画</p> | | |

技術開発の取組・成果

- 複数のツールが出力する異なる形式の判定結果を統合可能な形式に正規化し三段階判定を実装する統合判定ロジックを構築した。
- 統合ダッシュボードを開発し、評価・可視化・パラメータ最適化を一元的に行える環境を整備した。
- 統合ロジックにより判定精度が向上し、過検知や見逃しの低減を実現し、単独のツール製品では達成できない高精度・高信頼度の判定を実現した。
- 可視化UIと自動評価機能により、判定根拠の透明性が向上し、パラメータ調整や実務運用が効率化され、社会実装に耐える運用ロジックが確立した。

社会実装に係る取組・成果

- 市場ニーズの深掘りと社会実装モデルの策定を行った。
- 有識者・担当者ヒアリング（報道、プラットフォーム、タレントマネジメント、ブランド保護、行政機関、教育機関、ツール提供会社等）を実施し、実務における具体的なペインポイント（人件費、風評被害、ブランド毀損、法務コスト）を抽出した。
- ヒアリング対象者の視覚的な理解を促進し、課題やニーズを引き出す為、当初の計画にはなかったモックを作成した。
- モックを用いたヒアリングを実施した結果、根拠を示した判定支援の必要性を確認した。
- 市場ニーズを把握し、事業化における具体的な経済価値の特定、売上具体化に向けた具体性の高いビジネスモデルの構築を策定した。

技術開発及び社会実装にあたっての課題・展望

- 統合判定においては、重み付けの調整により過検知低減や見逃し低減といった判定特性の切り替えが可能である一方、統合性能の最適化とユーザニーズをどの程度反映させるべきかについて、明確な判断基準の必要性を感じた。
- 悪意をもって編集・加工が施されたコンテンツについては、元の著作物や発信者の意図とは異なる印象を受け手に与える可能性が高く、社会的影響が大きい。このようなケースでは、コンテンツ自体が真実であっても、文脈の切り取りや意図的な演出によって誤解を招くおそれがあり、単純な真偽判定のみでは社会的リスクを十分に評価できないことが大きな課題であることを認識した。
- 統合結果の性能指標とユーザニーズ（過検知許容度、見逃し許容度）を結び付ける評価軸を整理し、ユースケース別に適切な統合方針を選択できる設計を検討する。既に前処理や多段階評価は実施しており、今後は、コンテキスト判別の高度化により、単なる真偽判定を超えた、実効性の高いリスク評価システムの構築を目指していく。

代表者コメント



サン電子株式会社
グローバルDI事業部
事業部長 須藤慎二

本プロジェクトでは、複数の検知技術を統合し三段階で判定する独自ロジックを確立。実証を通じ、精度の向上に加え「判定根拠の可視化」が実務に不可欠であると確認しました。今後は単なる真偽判定を超え、文脈や意図まで考慮した高度なリスク評価システムの構築を目指し、誰もが情報を安心して信じられる社会の実現に貢献してまいります。

電話音声フェイク検知および自治体向け偽・誤情報総合対策の開発・実証

| | | |
|----------------|--|---|
| アプローチする課題・目指す姿 | <p>[アプローチする課題] 電話音声フェイク検知</p> <ul style="list-style-type: none"> 高精度な音声生成AIの急速な進化により、人間の耳では本物と区別できないなりすまし音声が増加し、電話詐欺被害が拡大している。 <p>自治体向け偽・誤情報総合対策</p> <ul style="list-style-type: none"> SNS上でフェイクニュースやディープフェイク投稿が急増し、特に災害等の有事において、虚偽情報と真に重要な情報の判別が困難になっている。 <p>[目指す姿]</p> <ul style="list-style-type: none"> フェイク検知・真正性担保・発信者証明が社会インフラとして普及し、デジタルコミュニケーション全体の信頼性が確保される社会 | |
| | <p>電話音声フェイク検知</p> <ul style="list-style-type: none"> 電話・通話環境にフェイク音声の自動検知が組み込まれ、利用者が意識せずともなりすまし・詐欺から守られるシステムの構築 <p>自治体向け偽・誤情報総合対策</p> <ul style="list-style-type: none"> 自治体・公的機関がSNS上の情報の真偽を即座に判断でき、市民が有事でも信頼できる情報にアクセスできる情報環境の構築 | |
| 技術区分 | コンテンツの真偽判別支援技術、改ざん検知技術 真正性保証・信頼性判断支援 | <p>実施体制 (下線：技術開発主体)</p> <p>NABLAS株式会社 再委託先：NTT東日本株式会社(2社間でコンソーシアムを組成) 再々委託先：ageet株式会社, NTTテクノクロス株式会社 再々々委託先：株式会社HBA</p> |
| 対象とするモダリティ | 文章、画像、音声、動画 | |

| | |
|--|---|
| <p>技術開発の取組・成果</p> | <p>社会実装に係る取組・成果</p> |
| <p>[電話音声フェイク検知]</p> <ul style="list-style-type: none"> 実際の電話環境でも精度を維持モデルを開発し、フェイク音声を検知を実現 最新生成モデルに対し迅速に検知性能を評価し改善できる体制を整備 電話サービス組込ヘリアルタイム処理可能な軽量かつ高精度検知エンジン設計 <p>[自治体向け偽・誤情報総合対策]</p> <ul style="list-style-type: none"> 総合的フェイク検知技術および電子透かしおよびDID/VC機能の新規開発 システムに組み込みにより、「検知」と「証明」を統合した、自治体が実用可能な総合対策システムを構築 | <p>[電話音声フェイク検知]</p> <ul style="list-style-type: none"> NTT東日本との共同開発により電話着信時のフェイク音声リアルタイム判定アプリを開発、ひかり電話・携帯電話・IP電話の実網で検証を実施 検知アプリにより、利用者がリアルタイムで判定結果を確認可能にした <p>[自治体向け偽・誤情報総合対策]</p> <ul style="list-style-type: none"> 伊那市と2回の虚偽投稿の発生を想定した実証を実施 実務フローに即した形で想定通り証明書（VC）が埋込まれることを確認し、SNSで拡散後でも真正性が把握できることを確認から実用レベルへ改善 |

技術開発及び社会実装にあたっての課題・展望

[電話音声フェイク検知]

- 音声処理の全環境を網羅的な対応は一つのモデルでは非常に難易度が高く、最新モデルの検知の難易度が急速に上昇している。
- 利用環境に応じ最適なモデルを自動選択し、環境に左右されず、最新のAIにも対応可能な検知機能をシステムを構築し、あらゆる音声コミュニケーションの場へ機能を拡張していく。

[自治体向け偽・誤情報総合対策]

- 多様な課題を担当する人員や業務フローが未整備、ないしは発段階のため、業務が停滞することもあり、危機や課題に直面し初めて対策を講じる形になり後手の対応となっている場合が多い。
- 偽・誤情報対策体制ごと代行可能なサービス・インフラを提供することで、自治体の負担なく対策可能なサービスを提供していく。

代表者コメント



NABLAS社 取締役
鈴木都生

電話・自治体向けどちらも実環境・実ニーズを想定した技術開発と実証実験を行いました。

様々な課題や改善点も出てきておりますが、実環境で提供できる価値も確認することができ、社会実装と社会課題解決に向けて大きな一歩になったと感じています。

AIを活用した情報コンテンツの真偽判別支援システムの開発・実証

| | | | |
|-----------------------|--|--|--|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> 複合的なコンテンツの組み合わせによる偽・誤情報への技術的対策については、マスメディア企業や自治体、その他企業において個別に対策・検討が進められているものの、手動分析に多くの工数を要し、業務負荷が増大しているという課題がある。 本開発では、AIを活用した偽・誤情報分析支援システムを用い、昨年度開発したシステムに対するユーザヒアリング等から得られた技術的課題を改善することで、①総合的な真偽判定の高度化を実現する。あわせて、SNS等多様な利用環境やユースケースに対応可能な機能拡張を行い、②マスメディア企業や自治体、その他企業・団体との連携による社会実装を目指す。 | | |
| <p>技術区分</p> | <p>①コンテンツの真偽判別支援・改ざん検知技術</p> | <p>実施体制 <small>(下線：技術開発主体)</small></p> | <p>日本電気(株)、東京大学、NECソリューションイノベータ(株)、日本アルゴリズム(株)</p> |
| <p>対象とするモジュール種</p> | <p>文章、画像、音声、動画</p> | | |

技術開発の取組・成果

- コンテンツの真偽判別支援技術のうち、複合的なコンテンツを真偽判別する技術について開発した。令和6年度で開発した真偽判別支援技術をベースと、実証事業の中で明確になった課題を解決するための機能強化及び機能の追加。具体的な機能強化及び機能の追加として大きく以下をおこなった。
 - 総合真偽判別の高度化
 - 様々な利用ユーザの意図に応じた分析の実現とSNS等の分析対象情報の検知機能追加
- 技術開発におけるKPIとしては、令和6年度の開発技術で偽・誤情報の真偽判別に失敗する情報コンテンツに対して80%以上判別可能となったことを確認しKPIを達成した。

社会実装に係る取組・成果

- ビジネス視点として、本実証により社会実装の利用主体の拡大について定量的に評価するKPIを設定した。具体的には20社以上のマスメディア(放送局・通信社・新聞社)と2つ以上の自治体でのユースケースに沿って開発したシステムの実証・ヒアリングを行うことを目指した。また、展示会への出展を2件以上(例：マスメディア向けInterBEE、他業界向け展示会等)、学会での論文採択・発表1件以上(例：AI系学会のAAAI等)を行い、社会実装の拡大に向けたユーザ探索を積極的に進めることを目指すこともKPIとして設定していた。
- 上記において69社のマスメディア(放送局・通信社・新聞社)と3つの自治体へシステムの実証・ヒアリングを実施。展示会への出展を3件、会での論文採択を1件実施したためKPIは達成となった。

技術開発及び社会実装にあたっての課題・展望

- 今年度予定の開発項目を組み込み、技術開発上のKPIとしていた誤判別率の向上は達成できたものの、ユーザの試用により、以下のような課題が確認された。
 - 専門用語に対応できない場合がある
 - レスポンスタイムが想定よりも長い
 - ユーザが望むユースケースである「生成AIによって加工・作成されたフェイク画像の検知」については正確な判別をおこなうことが難しい場合がある
- 上記を踏まえ、本事業の展望としては以下が考えられる。
 - ユーザの利用頻度が高いと想定される領域を強化したソリューション開発(専門用語集対応含む)
 - 偽誤情報判別ジョブの並列化
 - 生成AIによる画像フェイク検知の精度向上

代表者コメント



日本電気株式会社
メディアプラットフォーム
統括部
井上有二グループ長

昨年度の成果を踏まえ、本年度は技術面・社会実装面ともに活動の幅を広げることができた。特に、実証を通して多様な利用シーンでのニーズを把握し、それに応じた機能強化を進められたことは大きな収穫である。

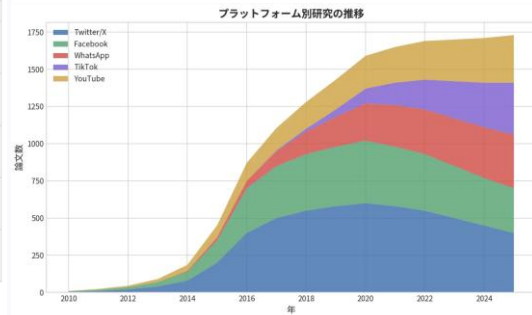
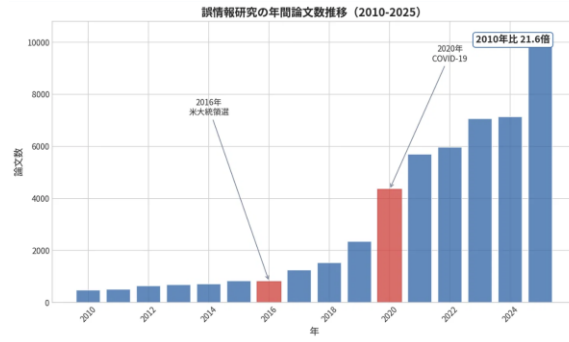
また、メディアや自治体との連携、展示会や学会での発信を通じて、社会実装に向けた基盤を着実に築くことができた。

グローバル・メタアナリシスと国内実証による対策技術の有効性の研究・調査

| | | | |
|---------------------------------|--|-----------------------------------|-----------------------|
| <p>アプローチ する課題 ・目指す姿</p> | <p>・ 英語圏偏重の既存研究を相対化し、日本固有の文脈に即した対策を立案するため世界的な誤情報研究動向を明らかにする。その上で、日本の文脈を再現可能なLLMエージェントを用いた「シリコンサンプリング」による安全な検証環境の構築を実施し、コストの低い再現実験の手法を提示する。</p> | | |
| <p>研究・ 調査区分</p> | <p>偽・誤情報対策技術に係る研究</p> | <p>実施体制 (下線：研究・ 調査主体)</p> | <p>株式会社新領域安全保障研究所</p> |

研究および有効性等に関する検証の取組・成果

- ・ 誤情報研究動向の分析：OpenAlexのAPIを用い、誤情報研究が過去15年で爆発的に拡大し、選挙・パンデミック・生成AI等の社会的危機が研究を牽引する「イベント駆動型」であることを整理。また、研究対象の地理・言語・プラットフォーム偏りを可視化し、誤情報研究における欧米の「政治」への関心に対しグローバルサウスでは「健康」への関心が高いことなどを示した。
- ・ 国内実証（シリコンサンプリングの提示・比較検証）：WVS（世界価値観調査）日本データを用い、LLMエージェントによる社会シミュレーションを構築。個別ペルソナ法／集約統計法、GPT系／Gemini系、複数介入条件を組み合わせ比較した。モデル・手法・入力情報によって推定が揺れること、介入によっては平均では効いて見えても特定層で逆効果を取り得ることを示唆。



研究・調査にあたっての課題・展望

- ・ 研究動向分析はOpenAlex + タイトル中心の抽出のため、非英語圏・ローカル語／領域固有語の取りこぼしがあり得る。→多言語・多概念での研究ウォッチへ拡張。
- ・ シリコンサンプリングは有望だが、現時点では妥当性（現実をどれだけ当てているか）検証のベンチマークが不足。→比較用の人間データ整備と評価枠組みの確立が必要。
- ・ ペルソナ設計（どの情報を与えるか）や、ネットワーク効果・反復曝露・画像／動画など現実環境要因の取り込みが今後の課題。→より実環境に近いシミュレーションへ改善し、日本での誤情報対策に資するエビデンス創出へつなげる。

代表者コメント



新領域安全保障研究所
代表取締役 齋藤孝道
(明治大学
理工学部教授)

世界的な研究動向を理解することは重要ですが、一方海外で成功したとされる対策手法を無批判に日本へ導入することは、予期せぬ副作用を生むリスクを伴います。今回は世界の研究動向を追いかけながらも、海外の知見についてどのように日本に適用するか低コストで検証する手法の提示を目指しました。

偽・誤情報の判別に知識および認知的熟慮が及ぼす影響

| | | | |
|-----------------------|--|------------------------------|-------------|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> 政治に関する偽・誤情報は、政局を混乱させ、民主主義への信頼を損なわせるという深刻な問題を引き起こす。 偽情報に惑わされてしまう反応の一部は、確証バイアスや認知的流暢性などの認知バイアスによって説明できる。 本研究では、(1) ウェブ調査により偽情報への接触状況および接触に関連する要因を検討すると同時に、(2) サーベイ実験で認知的反射テストを熟慮の操作として用いることで、「熟慮」が偽情報の正誤判断に及ぼす影響について検討を行う。 | <p>実施体制 (下線：研究・調査主体)</p> | <p>中央大学</p> |
| <p>研究・調査区分</p> | <p>偽・誤情報対策技術に係る研究</p> | | |

研究および有効性等に関する検証の取組・成果

本研究では、ウェブ調査及び（ウェブモニタ調査形式を用いた）サーベイ実験によって、以下の知見を得た。

【ウェブ調査】

- ①若く、ネットを情報源としている人ほど偽・誤情報に接しやすい。一方、テレビニュース接触は偽・誤情報判断に正の相関を示しており、注意喚起の効果もあったと推測される。
- ②社会への不安や政治的疎外意識が高い人ほど偽・誤情報に接しやすく、また信じやすい。
- ③批判的思考態度や「情報裏どり傾向」は、偽・誤情報への接触に正の効果を示していた。とくに批判的思考態度は、偽・誤情報の受容（本当かもしれないと思うこと）にも正の効果を持っていた。情報を深く処理することで、見たことを記憶しているということかもしれないが、「いろいろな考え方を知ろう」「裏付けを探そう」という努力が、場合によっては偽・誤情報への接触を促進してしまう可能性もある。

【サーベイ実験 1 および 2】

- ①認知反射テストの正解率は各問 2～4 割程度と低い水準であった。このことは、とくにネット上では「熟慮せずに反射的に反応する」ユーザが多いことを示唆している。
- ②認知反射テストへの取り組みとフィードバックを受けたグループは、「認知反射テストに取り組んだがフィードバックがない」「認知反射テストを受けていない」グループに比べて、「見たことがある、と判断する事実ニュースの数」が有意に多かった。ただし全体的に平均値は小さく、偽情報の判断には効果がなかった。
- ③「偽・誤情報を正しく判定した数」認知反射テストとFBを受けた回答者では、政治知識が多いほど「事実ではないと疑う偽・誤情報」の数が多くなっていた。以上の結果は、「考える作業」が、一定の条件で、情報の正誤判断を促進する可能性を示唆している。

研究・調査にあたっての課題・展望

- 本研究の知見として、特に重要なものは以下の 4 点である。
- ①社会への不安や政治的疎外意識が高い人ほど偽・誤情報に接しやすく、また信じやすい。
- ②批判的思考態度の強い、また情報の裏どりを重視する人ほど、偽・誤情報への接触が多い。（ただし、「見たことを記憶している」ことを意味している可能性もある。）
- ③認知反射テストの正解率は各問 2～4 割程度と低い水準で、ネット上では「熟慮せずに反射的に反応する」ユーザが多いことを示唆している。認知反射テストの正解数が多いほど提示情報を疑う傾向にあった。
- ④認知反射テストへの取り組みとフィードバックを受けたグループは、「認知反射テストに取り組んだがフィードバックがない」「認知反射テストを受けていない」グループに比べて、「見たことがある、と判断する事実ニュースの数」が有意に多かった。
- 以上の結果から、一般の有権者・ネットユーザが、「考える」面倒さを省きがちなこと、必ずしも事実への動機づけが高くないことを踏まえた情報発信が必要と考えられる。

代表者コメント



中央大学文学部教授
安野智子

本研究では、認知反射テストへの取り組みが「熟慮」を促すかどうか、またそれが偽・誤情報の判断を促進するかどうかについて検討しました。結果としては「事実判断」への効果、および、偽情報判断については微妙な交互作用がみられたのみでしたが、「考える作業」が一定の条件で正誤判断を促進するということが示唆されたと考えています。

生成AI時代における偽誤情報流通と認知特性の解明に関する研究・調査

| | | | |
|---------------------------------|---|------------------------------------|-------------------------|
| <p>アプローチ する課題 ・目指す姿</p> | <ul style="list-style-type: none"> 生成AI技術の急速な発展によって人間の判断を欺くコンテンツが容易に作成されるようになっている。膨大な情報の中で、すべての情報を正確に収集・理解・吟味することが不可能であり、不本意な意思決定をしてしまう場面が増加。情報の受け手である人間の認知特性自体が偽情報に対する脆弱性となっている。 そこで現在及び将来の偽誤情報脅威の体系化と日本固有の生成AI偽誤情報受容性の解明を図る | | |
| <p>研究・ 調査区分</p> | <p>偽・誤情報対策技術に係る研究</p> | <p>実施体制 (下線: 研究・ 調査主体)</p> | <p>東京大学大学院情報学環澁谷研究室</p> |

研究および有効性等に関する検証の取組・成果

【研究項目1】

- 偽誤情報関連キーワードで収集したX投稿（2025年10-11月、336万件、リツイートを除く）では、政治政党批判やメディア批判に関するトピックが最も多く、ついで、偽誤情報関連の言説を批判否定するもの、排外主義的な主張に関するトピックが多い
- 偽誤情報関連投稿で用いられている手法としては信用失墜型（Discrediting）が多い → 発信元を攻撃で正しい情報さえも受け付けない土壌が形成される懸念
- 偽誤情報対策は単なる真偽判定の問題ではない。情報をどのように受容し、どのように評価し、どのような行動へと接続するのかという認知過程そのものを対象化する設計が不可欠

【研究項目2】

- 日米独仏印の5カ国でオンライン実験を実施（N=5,443）し、強い警告（メッセージ型）と控えめな警告（ラベル型）の効果を検証
- 動画の正確性評価（7段階評価）に加え、ユーザーの迷い（不確実性）を測定し、個人属性や地域、AIの認識、性格、価値観などで制御の上比較
- 現在主流となっている控えめな生成AIラベルは、ユーザーの認知的警戒心を高める効果は限定的、誤情報受容抑制の十分条件とはなり得ない可能性が示唆
- 技術的対策に加え、ユーザー自身の認知バイアスへ働きかける教育が不可欠
- 現状把握や将来の脅威の理解を深めるためにも、SNS プラットフォームにおけるデータ公開・透明性を求める必要性

研究・調査にあたっての課題・展望

- 今後は、①プラットフォームとの連携によるデータアクセスと透明性の確保、②単なる注意喚起ではなく識別能力を高める介入設計の検証、③「どちらとも言えない」という不確実性の機能と拡散行動への影響の追跡、④利用者の異質性を前提とした適応的設計、⑤実環境での長期的・持続的効果の検証が求められる
- 誤情報を一律に抑制することではなく、生成AI時代に即した利用者の判断過程そのものを支えるインターフェース設計も重要

代表者コメント



東京大学
大学院情報学環
准教授
澁谷遊野

日々変化する情報空間を的確に捉え、日本の文脈に即した情報流通構造をより深く理解していく必要があると考えます。偽誤情報への対応には特効薬はなく多層的・多面的なアプローチが不可欠です。今後は、プラットフォームデータ透明性の向上を求めるとともに、異質性を前提とした介入設計、さらに実環境における長期的効果の検証を進めていきます。

偽・誤情報の拡散を抑制するためのSNSにおける シェア行動プロセス可視化と信頼性を評価する表示の検討

| | | | |
|---------------------------------|--|-----------------------------------|--------------------------------------|
| <p>アプローチ する課題 ・目指す姿</p> | <ul style="list-style-type: none"> 偽・誤情報の流通・拡散の防止として、コンテンツモデレーションの手法が用いられているが、表現の自由との兼ね合いで透明性確保が難しいという限界があり、十分な偽・誤情報対策として機能しない可能性がある。このため、偽・誤情報の特徴を的確に捉える精緻な拡散行動メカニズムの解明が求められている。 信頼性情報の可視化によって、利用者の自律的な判断を促すメカニズムを解明することにより、偽・誤情報対策の迅速化と質的向上を目指す。 | | |
| <p>研究・ 調査区分</p> | <p>偽・誤情報対策技術に係る研究</p> | <p>実施体制 (下線：研究・ 調査主体)</p> | <p>東京大学 鳥海研究室、(株)Lightblue、(株)電通</p> |

研究および有効性等に関する検証の取組・成果

- Chrome拡張機能の開発**: Webブラウザ上で動作する拡張機能を開発し、Webページの構造解析（DOM監視）により投稿者情報を特定および、X（旧Twitter）上で、投稿者の過去のコミュニティノート（CN）付与履歴を可視化するツールを実装した。
- 実証実験による効果検証**: 募集した一般ユーザー38名を対象に、実際のSNS環境を模したシステムを用いて比較実験を実施。介入群（ツール使用）は対照群（ツール使用なし）に比べ、全体のリポスト数を約59%、CN付与投稿へのリポストを約73%抑制した。
- 長期的な行動変容**: 追跡調査により、実験直後は高い抑制効果が見られるものの、2週間経過後には「慣れ」により効果が減衰する傾向（習慣化）を確認した。これらが報告書の主張と整合することを検証済み。

| 指標 | 対照群 (ツール使用なし) | 介入群 (ツール使用あり) | 抑制率 |
|--------------|------------------|------------------|-------|
| 全体リポスト | 7.37回 | 3.00回 | 59.3% |
| CN付与投稿へのリポスト | 2.74回 | 0.74回 | 73.1% |

研究・調査にあたっての課題・展望

- 大規模実証の必要性**: 本研究では中程度の実用的な効果量（Cohen's $d \approx 0.5$ ）を確認したが、統計的有意差の確立にはより大きなサンプルサイズが必要である。
- 習慣化への対策**: 長期利用における効果減衰を防ぐため、UIの動的変更や、信頼性スコアに応じた介入強度の調整など、持続的な効果維持のための機能改善が求められる。

代表者コメント



東京大学大学院工学系研究科教授
鳥海不二夫

本研究では過去の信頼性の低い投稿の可視化が情報拡散の判断にどのような影響を与えるのかを明らかにした。投稿者が過去にコミュニティノートが付与された事実の可視化が、情報拡散抑制に効果的であることが示された。当研究の結果は偽誤情報対策に大きな貢献が見込まれる。

誤情報対策における訂正のリスク：心理的メカニズムの研究・調査

| | | | |
|-----------------------|---|------------------------------|-----------------------|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> 訂正を行なった後に偽誤情報の影響が残るという現象には、従来の「誤情報持続効果」だけでは説明できない可能性がある。 撤回（-P）に含まれる誤情報（P）が後続の意思決定に影響を及ぼすという「誤情報撤回効果」仮説を厳密な実験手法によって検証する。 本仮説および関連する認知メカニズムを解明することにより、訂正リスクを可視化するとともに、効果的な訂正手法の実現を目指す。 | | |
| <p>研究・調査区分</p> | <p>偽・誤情報対策技術に係る研究</p> | <p>実施体制 (下線：研究・調査主体)</p> | <p>名古屋工業大学、神戸女子大学</p> |

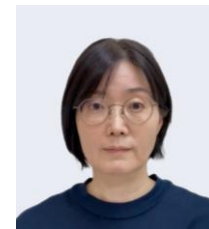
研究および有効性等に関する検証の取組・成果

- 本研究では、インターネット上に拡散した誤情報への主要な対策である「訂正」に内在するリスクに着目し、その心理的メカニズムを実証的に検証した。従来研究では、撤回前に提示された誤情報の影響が、訂正後も残存する現象として誤情報持続効果が報告されてきた。本研究は、訂正のリスク、すなわち、誤情報そのものに触れていなくても、訂正情報に触れることで、新たに誤情報の影響を生み出しうる可能性に焦点を当てた点に特徴がある。
- 研究1および研究2では、大規模サンプル（4実験、 $N \approx 7,400$ ）を用いた実験により、「撤回（-P）のみに接触した場合でも、後続の意思決定において誤情報（P）の影響が生じる」という「誤情報撤回効果」仮説を検証した。
- その結果、撤回は誤情報の影響を一定程度低減する一方で、誤情報持続効果と誤情報撤回効果の両方が一貫して確認された。さらに、これらの影響は1週間後の遅延測定においても維持されており、誤情報撤回効果が一過性ではなく持続的な現象であることが示された。
- また、記憶成績との関連分析および認知的負荷を操作した研究2の結果から、誤情報撤回効果は単純な記憶保持の強さや課題疲労によって説明されるものではなく、出来事の原因を理解しようとする因果推論過程に内在する認知的特性に基づく現象であることが示唆された。これらの結果は、撤回表現の中で言及された原因情報が、原因に対する認知的ポイド（空隙）を埋める形で利用されてしまうという、本研究が想定した認知メカニズムと整合的である。
- 研究3では、誤情報撤回効果を緩和する介入方法として、撤回表現に代替説明を付加する方法の有効性を検証した。その結果、代替説明を付加することで、誤情報撤回効果が有意に低下することが示された。これらの知見は、訂正そのものが新たな誤情報の影響を生み出しうるというリスクを可視化するとともに、より効果的で副作用の少ない訂正手法の開発に向けた科学的根拠を提供するものである。

研究・調査にあたっての課題・展望

- 本研究により、撤回という訂正形式に内在するリスクとして誤情報撤回効果の存在とその持続性が明らかとなったが、今後の研究に向けたいくつかの課題が残されている。
- 第一に、誤情報撤回効果を生み出す詳細な認知メカニズムの解明である。
- 第二に、因果推論以外の文脈においても誤情報撤回効果が生じるのかという点である。
- 第三に、日本語環境以外における再現性の検証が挙げられる。撤回表現の語順や否定形式は言語によって異なるため、異なる言語・文化圏における再現研究を通じて、本効果の普遍性と文化依存性を検討する必要がある。
- 最後に、代替説明を常に提示できない状況における対策の検討が課題として残されている。
- これらの課題に取り組むことで、誤情報撤回効果に関する理解をさらに深化させるとともに、訂正という主要な偽誤情報対策を、より効果的で副作用の少ないものへと発展させることが期待される。

代表者コメント



名古屋工業大学
基礎類 教授
(田中優子)

偽誤情報対策におけるデバンプとしての訂正の効果を妨げる要因とその認知プロセスに関心を持って研究をしています。

本研究では、訂正のみから誤情報の影響が新たに生まれる可能性というこれまで見過ごされてきた現象を検証しました。

認知的特徴を理解することで、よりよい偽誤情報対策につなげることができたらと思います。

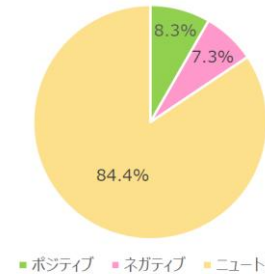
制作者意図と異なる作品解釈に係る情報の急激な増加の影響等に関する研究・調査

| | | | |
|-----------------------|--|------------------------------|--|
| <p>アプローチする課題・目指す姿</p> | <ul style="list-style-type: none"> ・ 制作者意図とは異なる作品解釈に係る情報の急激な増加により、健全な情報流通が変容し、これが表現の萎縮に繋がりがねないという課題が存在する。 ・ このため、偽・誤情報の拡散事例や拡散の背景から誤解情報の拡散抑止として取りうる効果的対策の解明が求められている。 ・ 誤解情報の拡散メカニズムを解明することにより、デバンキング（debunking）情報の発信による拡散抑止といった対策を明らかにすることを目指す。 | | |
| <p>研究・調査区分</p> | <p>偽・誤情報対策技術に係る研究</p> | <p>実施体制 (下線：研究・調査主体)</p> | <p>日本エンタープライズ(株)、(株)インフォシティ、株式会社エルテス</p> |

研究および有効性等に関する検証の取組・成果

- ・ 本研究調査では、作品・作者に対する根拠のない偽・誤情報的な批判の拡散（炎上）からかかる批判を真実であることを受け入れたかのごとき表現の不本意な変更へと繋がった事例の調査と、拡散の背景（予兆・拡散者の意図）や進行プロセスを類型化し、かかる偽・誤情報的な批判の拡散防止方法を考察する研究を実施した。
- ・ 具体的には、2段階での炎上事案調査（情報収集）を行った。リアルタイム捕捉を含め6事案の推移を検討し、うち2事案については専門ツールを用いてSNSポストデータの収集と分析を行い、批判発生から鎮静化までの推移や、状況に影響を与えた要因について検討を行った。
- ・ これらの調査・分析により、炎上事案の構造やその推移に影響を与える要因について、以下の知見が得られた。
- ・ 投稿全体の中でネガティブ（批判）投稿は一部に過ぎないが、炎上初期には制作者への攻撃的な批判が集中する。批判が作品の一方的解釈を持つ一部グループからに限られていても、制作者はその状況を認識することが難しいため、極めて強いプレッシャーとなり、謝罪等につながることもある。
- ・ 炎上拡大後、説得力あるポジティブ／ニュートラル投稿が大量にリポストされることにより、炎上が鎮静化に向かうプロセスがあることが確認された。説得力ある投稿では、批判が合理的な内容でないことを示す客観的な根拠や視点を示すことが必要になる。
- ・ 制作者側が事前に公開したガイドラインや、行き過ぎた初期批判投稿の記録などが根拠として有効であることが確認された。
- ・ 専門家ヒアリングでは、本調査結果は少数の作品批判が大きなインパクトを持つ構図が実証的に示され、業界への示唆、特に海外向けサービス等での示唆が得られると評価された。また、今後進めるべき対策として、継続的な炎上事案調査、制作者側のポリシーやガイドラインの公開、過激な投稿を抑制するプレバンキング（警告）機能のSNSへの実装などが挙げられた。

収集投稿の 카테고리構成比（例）



研究・調査にあたっての課題・展望

- ・ まず、多くの炎上事案を調査・分析しその知見を制作者等と共有することが必要である。併せて、技術的な諸課題に関する研究・調査の実施が望まれる。効果的なプレバンキング・デバンキング情報提供に関する研究、ガイドラインの在り方の検討、持続可能な対策推進体制の検討などが必要である。
- ・ 批判拡散防止のためには、炎上事案発生を早期に把握し批判の実態を記録することが重要である。コンテンツ独特の炎上構造を踏まえた早期探知ツール、警告ツールの開発・実装が求められる。
- ・ 根拠のない批判を抑制するため、作品表現や二次創作品に関するガイドラインを整備することが望ましい。個々の作品表現を縛るのではなく、考え方の枠組みや標準形を示す一般的ガイドラインをまず整備することが考えられる。
- ・ 中期的には、利害関係者を中心に対策推進ユニットを組織して作品や制作者に対する炎上のモニタリングを行い、炎上発生時にはSNSプラットフォームと連携して制作者・ユーザー双方に対しプレバンキング／デバンキング情報提供等を行って早期鎮静化を図る体制を整えることが望まれる。

代表者コメント



日本エンタープライズ(株)
取締役
杉山 浩一

意図しない偽・誤情報の発生メカニズムや、その情報の信憑性を確認するまでもなく、拡散される事であたかも多数派意見＝正しいかのようなインパクトを与えてしまう。情報の拡散スピードが早い昨今において、このインパクトが制作者に影響を与えないよう早期探知ツール、ガイドライン等の発信につながるべく研究調査の必要性を感じている

© 2026 PricewaterhouseCoopers Japan LLC. All rights reserved.

PwC refers to the PwC network member firms and/or their specified subsidiaries in Japan, and may sometimes refer to the PwC network. Each of such firms and subsidiaries is a separate legal entity. Please see www.pwc.com/structure for further details.

This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors.

本報告書は、総務省との間で締結された2025年3月5日付の請負契約書に基づき、PwCコンサルティング合同会社が作成したものです。PwCコンサルティング合同会社は、本報告書に関連して、総務省以外の第三者に対して、如何なる義務や責任も負いません。なお、PwCコンサルティング合同会社は、本報告書の日付後に発生した事象について、追加で報告をなし又は本報告書に反映させる責任を負うものではありません。