

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

# 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

## 成果報告書 概要版

2026/3/19

技08\_株式会社データグリッド

# 目次

## 1. 開発・実証における対策技術の開発

1. 開発技術によりアプローチする課題・目指す姿
2. 技術開発の取組・成果

## 2. 開発・実証における社会実装に向けた取組

1. 社会実装に係る取組・成果
2. 社会実装時のビジネスモデル等
3. 技術開発及び社会実装にあたっての課題・展望
4. 事業の拡大に向けた中長期的な計画

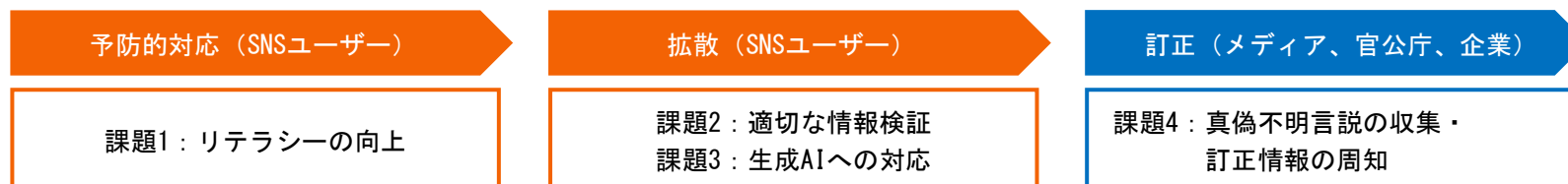
# 目次

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

# 1-1. 開発技術によりアプローチする課題・目指す姿

## 開発技術によりアプローチする課題

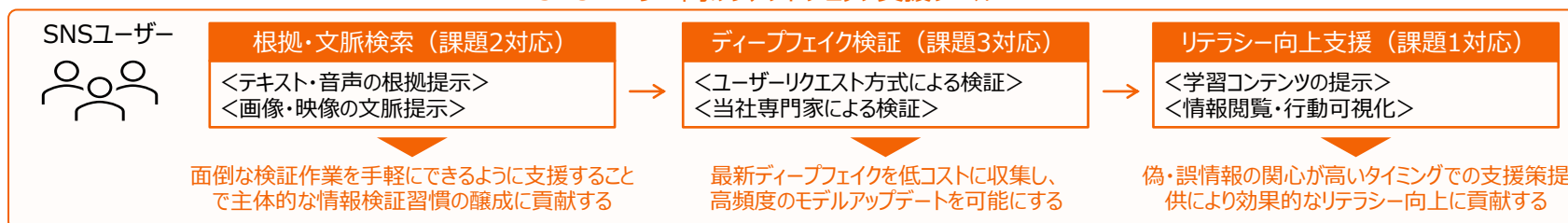
- 本事業では、SNSユーザーへの直接的な技術的支援を中核に据えつつ、メディア、官公庁、企業、研究機関などが連携し、訂正情報の発信やリテラシー向上プログラムの提供といった多角的な役割を担う社会全体での対策を推進する。この方針のもと、偽・誤情報の「予防的対応」、「拡散」、「拡散後の「訂正」の各段階における課題に対処する技術・ツールを開発・実証することで、健全なデジタル空間の実現に貢献する。



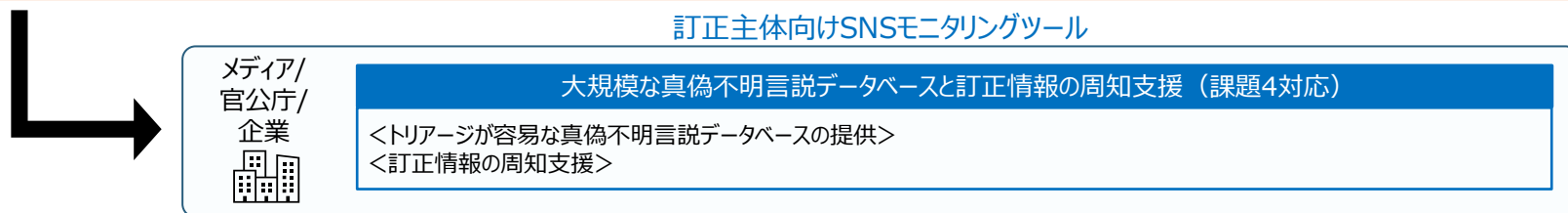
## 上記課題を踏まえ目指す姿・ゴール

- 前述した4つの課題解決に貢献する『SNSユーザー向けファクトチェック支援ツール』及び『訂正主体向けSNSモニタリングツール』を開発し、その有効性やユースケースを検証・調査する。

### SNSユーザー向けファクトチェック支援ツール



### 訂正主体向けSNSモニタリングツール



## 1-2. 技術開発の取組・成果

### ツールの基盤となる要素技術の実施目的・概要

- SNSユーザー向けツールおよび訂正主体向けツールで活用する5つの要素技術の開発を実施

#### <要素技術 1. 画像・映像・音声のディープフェイク検出の運用・高度化技術>

- 生成 AI の急速な発展に対応するために、最新の生成AIで作られたDFデータセット（学習用データ）を構築し、それをを用いて昨年度開発したDF検出エンジンをファインチューニング
- 生成 AI の急速な発展に対応するために、機械学習技術の開発から運用、改善までの一連のフローを効率化する MLOps（Machine Learning Operations）パイプラインを構築

#### <要素技術 2. Deep Research 技術による信頼性評価付き根拠情報検索技術>

- SNS 上のテキスト、動画、音声の真偽不明言説の検証に不可欠である信頼性の高い複数の根拠情報に容易にアクセスできる技術を開発

#### <要素技術 3. 画像・動画に対する出所及び文脈の検証技術>

- SNS 上の画像や動画コンテンツの初出情報及び過去にどのような文脈で利用されてきたかを特定・追跡し、検証に必要な情報を提示する技術を開発

#### <要素技術 4. 情報接種・行動の分析技術>

- SNS ユーザーのタイムラインで流れてくる情報の偏りの分析や、「いいね」や「リポスト」などのエンゲージメント行動をした情報に偽・誤情報が含まれていたかなど、自らの情報環境をわかりやすく振り返る機会を提供する技術を開発

#### <要素技術 5. 真偽不明言説のリスクアセスメント・拡散予測技術>

- 真偽不明言説に対して、リスク評価技術によるトリアージ支援と拡散予測技術による早期の対抗策実施を支援する技術を開発

## 1-2. 技術開発の取組・成果

### ツールの基盤となる要素技術の実施概要

- 各要素技術の開発・検証を実施し、目標を超える精度を達成

要素技術	成果	KPI		
		評価指標	目標	結果
1	最新の生成AIを含む包括的なディープフェイクデータセットを構築 昨年度開発したディープフェイク検出エンジンを最新のDFデータセットの一部を用いてファインチューニング（特定の用途に最適化する調整）し、高い検出精度を実現。 最新の生成 AI 技術によって作られたディープフェイクに対しても短期間の再学習で自動的に検知できる MLOps パイプラインを構築。	ディープフェイク検出精度	90%	93.1%
2	SNS ユーザーが真偽不明言説を検証する際に、信頼できる支持・反証両面の根拠情報への容易なアクセスを実現し、検証作業を大幅に効率化した。さらに、ユーザー自身が根拠情報に基づき主体的に真偽を判断する支援になることが期待される。	正しい根拠を提示する割合（適合率）	85%	92.9%
3	画像や動画に対するリバースサーチ（類似画像・動画検索）と VLM（Vision Languageモデル; 視覚情報と言語情報の両方を使ったモデル）による視覚的理解を基にした検索技術により、本来の文脈で画像や動画が使われているかを迅速に検証することが可能になった。	SNS 上で実際に流通した改変・流用画像/動画に対し、正しい初出情報・利用文脈を提示する割合（適合率）	80%	90%

## 1-2. 技術開発の取組・成果

### ツールの基盤となる要素技術の実施概要

- 各要素技術の開発・検証を実施し、目標を超える精度を達成

要素技術	成果	KPI		
		評価指標	目標	結果
4	SNS ユーザーに対して、情報接種の偏りやエンゲージメント行動における真偽不明言説の拡散履歴などの可視化により、情動的健康を意識するきっかけと行動変容に繋がる機会を提供することができた。	ラベリングデータで情報接種分析の結果に関して QWK ※多クラス分類の評価指標の 1 つで、混同行列に重みを付けて評価する指標	0.30	0.62
5	訂正主体において、真偽不明言説に対し、対応すべき優先度を即時かつ的確に判断するための高度なリスク評価技術と実用的なトリアージシステムを確立し、高リスク事案の特定率向上や検知時間改善を実現。	ツールが算出したリスクスコアQWK ※多クラス分類の評価指標の 1 つで、混同行列に重みを付けて評価する指標	0.30	0.84

# 1-2. 技術開発の取組・成果

## ツールの基盤となる要素技術の開発と成果

- SNSユーザー向けのファクトチェック支援アプリおよび訂正主体向けのSNSモニタリングツールを開発

### SNSユーザー向けファクトチェック支援アプリ



### 訂正主体向けSNSモニタリングツール



# 目次

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

## 2-1. 社会実装に係る取組・成果

### SNSユーザー向け「ファクトチェック支援アプリ」の実証成果

- SNSユーザー向けアプリのアルファ版・ベータ版テスト（59名）を通じ、情報リテラシーの向上とファクトチェック支援の有用性を確認。一方、更なる普及拡大に向けた認知度向上が課題として浮き彫りとなった。

項目	詳細
① 実証の概要 (目的・対象)	<ul style="list-style-type: none"> <li>• クローズドアルファ版（一般公開前の非公開環境下での検証）：主要機能の有用性評価および情報リテラシー向上効果の検証</li> <li>• パブリックベータ版（一般公開後の環境下での検証）：一般SNSユーザーへの普及ポテンシャルと実際の活用における実態調査</li> </ul>
② 定量的な成果 (KPIと達成度)	<ul style="list-style-type: none"> <li>• 定性評価での高い満足度： <ul style="list-style-type: none"> <li>➢ アルファ版満足度：10段階中7.81（KPI：7以上を達成）</li> <li>➢ ベータ版口コミ評価：<u>ポジティブ率76.3%</u>（KPI：70%以上を達成）</li> </ul> </li> <li>• 情報リテラシー向上効果の実証： <ul style="list-style-type: none"> <li>➢ アプリ利用前後で情報リテラシーテストの平均スコアが+3.3%と統計的に有意に向上</li> </ul> </li> </ul>
③ 抽出された課題と今後の展開	<ul style="list-style-type: none"> <li>• 利用者の成果実態：「<u>自力で探せないソースを提示してくれる</u>」「<u>検証作業を習慣化できる</u>」と高い評価を得た。</li> <li>• 社会実装に向けた課題：情報検証の習慣がない層へのアプローチには高いハードルがある。</li> <li>• 改善アプローチ：メディアと連携した戦略的PRにより認知度の向上を図るとともに、誰もが直観的に使えるようUI/UXの改善に継続的に取り組む。</li> </ul>

## 2-1. 社会実装に係る取組・成果

### 訂正主体向け「SNSモニタリングツール」の実証成果

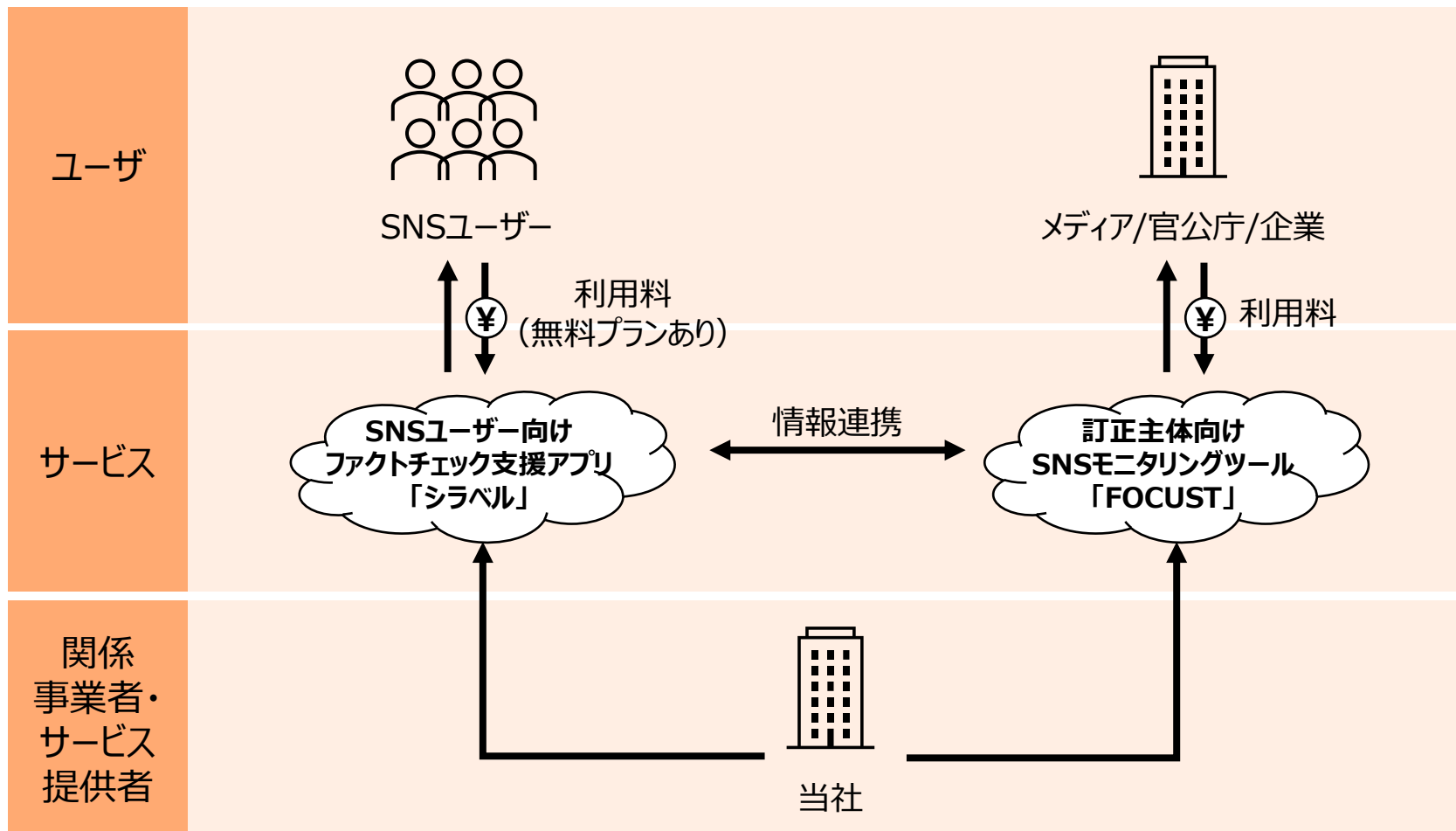
- 報道機関（2社）におけるテスト検証を通じ、偽・誤情報探知の迅速化と裏付け取材の効率化を実現に寄与できることを確認。社会実装に向けた有力な複数のユースケースを確立した。

項目	詳細
① 実証の概要 (目的・対象)	<ul style="list-style-type: none"> <li>対象：報道機関</li> <li>内容：実際の偽・誤情報モニタリングや検証業務でテスト利用。定期的なヒアリングを通じてユーザービリティやユースケースの検証を実施した</li> </ul>
② 定量的な成果 (KPIと達成度)	<ul style="list-style-type: none"> <li>ユーザービリティ評価：10段階中7と実務で利用可能であることを確認。</li> <li>主要なユースケースの確立（KPI：3つ以上の明確化を達成）： <ul style="list-style-type: none"> <li>➤ 検証対象の早期探知と優先順位付け</li> <li>➤ 拡散層の分析</li> <li>➤ 裏付け取材の時短など</li> </ul> </li> </ul>
③ 抽出された課題と今後の展開	<ul style="list-style-type: none"> <li>社会実装への手応え：人力リサーチをAIで代替でき、検証報道のスピード・質・中立性を大きく高める有効なソリューションとなることを確認。</li> <li>今後の展望：メディア・官公庁等の専門的な訂正主体向けSaaSとして正式版の提供を開始し、各組織のワークフローに適した形式で導入を推し進める。</li> </ul>

## 2-2. 社会実装時のビジネスモデル等

### 社会実装時のビジネスモデル

- SNSユーザー向け「ファクトチェック支援ツール」と訂正主体向け「SNSモニタリングツール」を提供し、相互に情報連携させるビジネスモデルである。一般ユーザーとメディア・官公庁・企業の両者から利用料を得る収益構造となっており、双方のアプローチを通じて情報空間の信頼性向上を図るエコシステムを構築する。



## 2-2. 社会実装時のビジネスモデル等

### ユーザ・導入先の詳細とそのペインポイント

- 両ツールは情報連携により、ユーザーの検証データが訂正主体の収集基盤となり、訂正主体の発信情報がユーザーに届くという相互補完の関係にある。

#### SNSファクトチェック支援ツール

##### ① 真偽不明言説を発見

ニーズ：自分で調べるのは手間がかかる  
 解決策：手軽な検証ツールを提供



##### ② ツールで検証

ニーズ：判定結果の根拠が見えない  
 解決策：信頼できる複数の根拠情報を提示



##### ③ いいね・拡散等判断

ニーズ：拡散可否の判断材料がほしい  
 解決策：検証結果に基づく拡散可否の判断支援



##### ④ 振り返り

ニーズ：リテラシーの自己評価手段がない  
 解決策：情報閲覧傾向を可視化し自己評価を支援

#### SNSモニタリングツール

##### ① 偽・誤情報の収集機能

ニーズ：人手で集めるのは手間がかかる  
 解決策：SNSユーザーの検証情報を活用



##### ② トリアージ

ニーズ：大量の情報から優先順位を判断したい  
 解決策：社会的影響度と拡散速度からリスク判定



##### ③ 検証

ニーズ：専門的な検証環境がない  
 解決策：ディープフェイク検出・根拠検索・画像の出所特定による多角的検証



##### ④ 訂正情報発信

ニーズ：正しい情報を偽情報に接したユーザーに届けたい  
 解決策：SNSファクトチェック支援ツールでの表示



## 2-3. 技術開発及び社会実装にあたっての課題・展望

### 技術開発及び社会実装にあたっての今後の課題

- 本年度の実証を通じ、技術面・社会実装面において以下の課題が明確化された。
  - ▶ **技術的な課題（精度維持とコスト）**：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、既存モデルの精度劣化リスクが顕在化している。また、LLMを活用した分析コストの最適化もスケールに向けた課題である。
  - ▶ **社会実装の課題（利用定着と情報流通）**：一般ユーザーへのベータ版提供を通じ、「真偽不明の情報を日常的に検証する習慣」が定着していない層へのアプローチが難しいことが判明した。なぜなら、情報の真偽を確かめる動機が薄い受容層において、情報を検証することへの価値が見出されていないためであると考えられる。また、訂正主体が発信する正しい情報が、誤情報に触れたユーザーに十分に届かないという情報の非対称性も存在している。

### 上記課題を踏まえた今後の展望

- 上記の課題を解決し、当社の技術を社会インフラとして定着させるべく、以下の展望を描く。
  - ▶ **継続的な技術アップデート（MLOps基盤の拡充）**：本事業で構築した「収集・監視→テスト→データ生成」のパイプラインをさらに自動化し、低コストかつ迅速に新種の偽情報モデルに対応し続ける体制を確立する。
  - ▶ **日常的な利用習慣の形成（UI体験とメディア連携）**：各メディアとの連携による積極的なPRを通じて認知度を高め、日常的な利用習慣の促進を推進する。また、検証行為そのものにベネフィット（インセンティブやゲーミフィケーション等）を付加する設計の検討を推進する。
  - ▶ **エコシステムとしての普及（SaaS・API展開）**：報道機関等に向けた業務支援SaaSとしての普及を進めつつ、将来的にはパートナー企業にAPIを提供し、あらゆるプラットフォーム上で判定機能が稼働するエコシステムの構築を目指す。

## 2-4. 事業の拡大に向けた中長期的な計画

### 事業の拡大に向けた中長期的な計画

- SNSユーザーを対象としたユーザー基盤を梃子にしてSNS上の偽・誤情報を流通状況を可視化するとともに、法人向けのソリューションにより事業の収益化を図る。

#### フェーズ1（2026年度） 基盤構築期

- **目標：**正式版プロダクト完成と初期顧客の獲得
- **実施事項：**
  - ✓ 「シラベル」「FOCUST」正式版のローンチ
  - ✓ パートナー企業との連携推進
  - ✓ 報道機関や官公庁への導入営業の実施

#### フェーズ2（2027年度） 市場拡大期

- **目標：**対応SNSの強化と収益化の推進
- **実施事項：**
  - ✓ 動画SNSへの対応および多言語対応
  - ✓ API/OEM提供など拡販パートナーを開拓
  - ✓ 一般企業向けへの販売パートナーを通じた展開

#### フェーズ3（2028年度） エコシステム確立期

- **目標：**偽・誤情報対策の社会インフラを確立
- **実施事項：**
  - ✓ SNSプラットフォームとの連携強化
  - ✓ C2PA等の国際標準への準拠と技術統合
  - ✓ 多言語版を活用した東アジアを中心とした海外市場の開拓