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

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
成果報告書 簡易版

2026/3/19

技11_Sakana AI株式会社

1-1. 開発・実証のサマリ

アプローチする課題・目指す姿	• 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。 • 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。		
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	実施体制（下線：技術開発主体）	Sakana AI株式会社
対象とするモーダル種	文章、画像、音声、動画		

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るべき対応策の立案
- 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。