

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 簡易版

2026/3/19

技08_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

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.
対象とするモデル種	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 情報摂取行動の可視化:** 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。