令和6年度 生成AI に起因するインターネット上の偽・誤情報等への対策技術に係る調査の請負 (実証事業)

多様なメディアにおける最新のディープフェイクに追従した偽・誤情報検出技術の開発・実証 成果報告書 概要版

2025年3月12日

株式会社データグリッド

実証番号1 株式会社データグリッド: R5補正 成果概要

主たる実証の成果

＜開発した技術・ツールの詳細＞
本実証では主として次の技術1~4を開発した。
技術1.顔画像・映像ディープフェイク検知技術: 生成AIで合成された顔画像・映像を97%以上の精度で検知できる技術
技術2. 音声ディープフェイク検知技術: 生成AIで合成された日本語音声を98%以上の精度で検知できる技術
技術3. 一般画像ディープフェイク検知技術: 生成AIで合成された一般画像（特に災害画像）を90%以上の精度で検知できる技術
技術4. 過去メディアを流用した偽情報に対する真偽判別支援技術：
対象データに関してリバースイメージサーチを実施することで過去に同じ画像がないか効率的に照合する技術

＜社会実装のための実証の結果＞
世の中で実際に問題になっているディープフェイクの整理や海外製のディープフェイク検出ツールでは対応が困難なディープフェイクの系統について検証した。
また、ファクトチェック業務の中で直感的に操作しやすいユーザーインターフェース等について、システムの開発前からテスト利用にかけて議論し、フィードバックをいただくことで、実業務の効率化に資するディープフェイク検知アプリの機能・構成を明らかにし、本事業で開発したアプリに落とし込むことに成功した。

今後の課題・展望等

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

＜本実証後の展望＞
2025年度 SNS上の偽・誤情報対策サービスの開発・実証：
ディープフェイク対策を含む、実行性のある偽・誤情報対策サービスのコアコンセプトを実装・実証する。
2026年度 サービス普及に向けた本格展開：
官公庁での利活用や対応SNSの拡充等により、サービス普及に向けた本格展開を開始する。
2027年度 ビジネスモデルの確立：
収益性を確保したビジネスモデルを導入し、持続的な事業運営フェーズに移行する。