

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

**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
成果報告書 簡易版**

2026/3/19

技05_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

<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>対象とするモデル種</p>	<p>文章、画像</p>	<p>実施体制 (下線：技術開発主体)</p>	<p>株式会社Classroom Adventure</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】

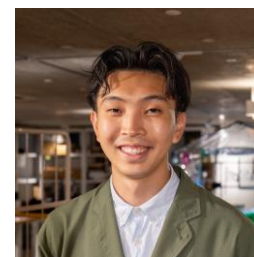
本技術は、生成体験を通じて判断行動へ転移させる設計である一方、教材内で生成された文章・画像が教材外へ持ち出されることによる誤認・拡散や二次利用のリスクへの対応が中核課題である。また生成AIの安全機能には、迂回表現・文脈依存・多言語等による抜け道が残っているため、教育上必要な生成体験を成立させつつ危険領域へ逸脱しない教材特有の安全に使うための制御設計が継続課題となる。さらに実環境では、偽・誤情報が短尺動画・音声なりすまし等へ移行しつつあり、教材側のマルチモーダル対応が不可欠である。一方で、ログ取得や運用統制はプライバシー・社内規程・審査対応と接続するため、導入先が説明可能なデータ範囲・権限・保存・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み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、私たちにあ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。ゲームフィク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。