

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**

2026/3/19

技05_株式会社Classroom Adventure

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

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】

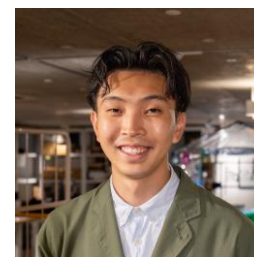
本技術は、生成体験を通じて判断行動へ転移させる設計である一方、教材内で生成された文章・画像が教材外へ持ち出されることによる誤認・拡散や二次利用のリスクへの対応が中核課題である。また生成AIの安全機能には、迂回表現・文脈依存・多言語等による抜け道が残っているため、教育上必要な生成体験を成立させつつ危険領域へ逸脱しない教材特有の安全に使うための制御設計が継続課題となる。さらに実環境では、偽・誤情報が短尺動画・音声なりすまし等へ移行しつつあり、教材側のマルチモーダル対応が不可欠である。一方で、ログ取得や運用統制はプライバシー・社内規程・審査対応と接続するため、導入先が説明可能なデータ範囲・権限・保存・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

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。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

2-1. 開発技術によりアプローチする課題

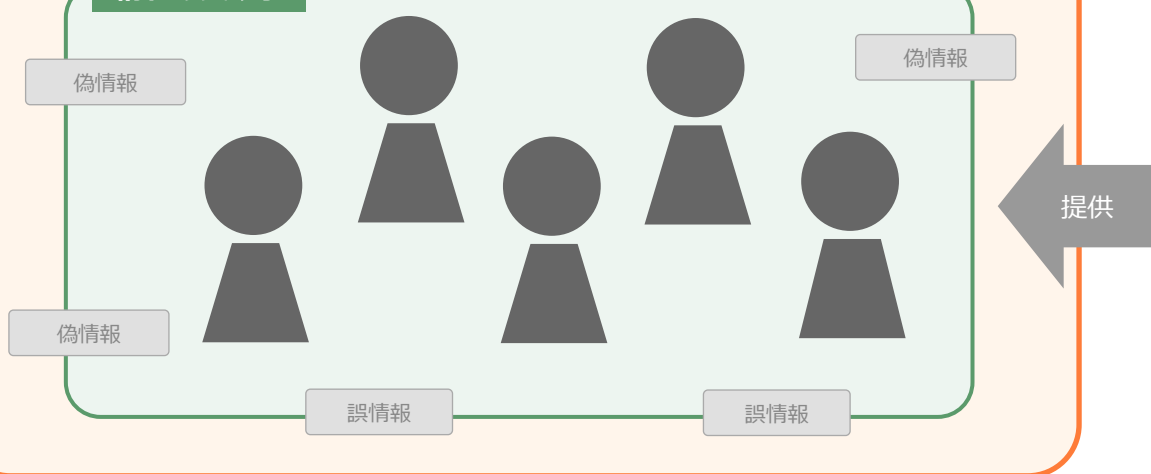
開発技術によりアプローチする課題

本事業が取り組む課題は、インターネット上の偽・誤情報をめぐる環境変化を踏まえ、主に次の三点に整理される。

社会環境

生成AIの普及により偽・誤情報が見分けにくくなり、大量流通する環境

情報の受け手



既存教材・対策



プレバンキング教材等

提供

生成AIの発展による偽・誤情報の変化

生成AIの普及により、偽・誤情報が高精度化・低コスト化し、短時間で大量に作成・流通し得る状況が生まれている

情報の受け手に求められるリテラシー

個別の訂正情報の提示のみでは限界があり、情報の受け手に実践的なリテラシーが従来以上に求められている

現状のプレバンキング教材が抱える課題

従来のプレバンキング教材は有効だが、生成AI時代の偽・誤情報の多様性や学習効果の持続に課題が残っている

2-1. 開発技術によりアプローチする課題

開発技術によりアプローチする課題

本事業が取り組む課題の詳細について下記に記載する。

生成AIの発展による偽・誤情報の変化

近年の生成AIの発展と普及により、偽・誤情報は「専門的な能力を持つ一部の者が作るもの」から、「一般の利用者でも短時間で作成できるもの」へと変化している。文章、画像等の生成が容易になり、見た目の自然さや説得力が増したことで、受け手が直感的に真偽を判別することは一層困難となっている。

また、生成AIの特徴は作成の容易さととどまらず、短時間で大量に作成できる点にある。従来であれば、一定の手間やコストを要した文章作成や画像素材の用意が、同一テーマであっても表現や構成を変えながら連続的に作成できるようになった。

その結果として、同じ出来事や話題をめぐる、異なる言い回しや異なる画像を用いた投稿が同時多発的に流通しやすくなり、受け手が接触する偽・誤情報の総量が増加している。

このような環境では、個別の偽・誤情報に対して訂正情報を提示する対応だけでは、発生と拡散の速度に追いつきにくい。したがって、偽・誤情報の流通・拡散を前提に、受け手自身が日常的に「疑う」「確かめる」「拡散しない」といった判断行動をとれる状態を整える必要がある。

情報の受け手に求められるリテラシー

偽・誤情報が巧妙化し、流通量が増えるほど、受け手側には「正しい情報を知っているか」だけでなく、「不確かな情報に出会った際に、拡散せず、立ち止まり、確かめる」ための判断力と行動が求められる。とりわけ、SNS等を通じて情報を受け取る場面では、短時間で内容を理解したつもりになり、感情や先入観に基づいて共有することが起こりやすい。

したがって、出所や根拠の確認、複数の情報源との照会、第三者視点から見た評価の参照といった検証行動を、日常的な習慣として身につける必要がある。

一方で、現実には、ファクトチェック機関等が訂正記事を作成しているものの、作成できる数には限りがある。生成AIにより大量に生じる偽・誤情報を、個別のファクトチェックのみで全てカバーすることは困難である。したがって、社会全体としては、受け手一人ひとりが偽・誤情報に対して耐性を持ち、誤って信じない、誤って広めないための知識を獲得することが不可欠である。この点が、教育的対策を必要とする大きな理由である。

また、教育現場や地域の啓発現場においては、指導者側が十分な教材や時間を確保しにくく、結果として断片的な知識提供にとどまりやすいという課題がある。受け手の行動変容まで到達させるには、知識の説明だけでなく、実践的な学習体験が必要である。

2-1. 開発技術によりアプローチする課題

開発技術によりアプローチする課題

現状のプレバンキング教材が抱える課題

プレバンキングは、偽・誤情報に接触する前に典型的な手口や心理的誘導を学ばせ、耐性を高めるという点で有効な枠組みである。一方で、生成AI時代の偽・誤情報は、作成の自由度と表現の多様性が高く、受け手が遭遇する具体例が状況に応じて変化する。このため、固定的な例示や受動的な解説に依存する教材では、「なぜ信じてしまうのか」「どの部分が印象操作されているのか」「どのように作られ、どのように広がるのか」といった構造理解に到達しにくいという限界がある。

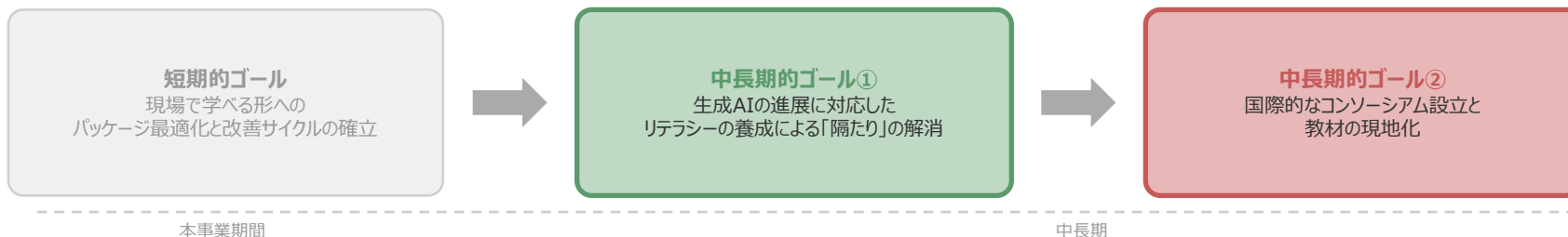
また、現状の学習方法として、事例を動画等で視聴し理解する形式が中心になりやすく、学習者自身が試行錯誤しながら「どこを操作すればもっともらしく見えるのか」「何が判断を誤らせる要因になるのか」を体験的に掴む実践機会が少ない。結果として、知識としては理解しても、実際の情報環境で瞬時に疑い、確認に立ち返る行動へ接続しにくいという課題が残る。

加えて、学習効果の持続という観点でも課題がある。単発の学習は理解を得やすい反面、時間の経過とともに注意が薄れ、行動として定着しにくい。偽・誤情報への対処は、知識ではなく反射的な行動として求められる場面が多く、反復による習慣化が重要となる。しかし、従来の教材は継続利用や再学習を前提とした設計になっていない場合が多く、受講後の効果を維持する仕組みが弱い。

したがって、3つの課題を解決するためには、受け手が「作り手の視点」も含めて偽・誤情報の構造を理解でき、かつ短時間で反復可能な形で学習を継続できる技術開発が求められる。

2-2. 開発技術により目指す姿・ゴール

開発技術を通して目指す姿・ゴール



中長期的ゴール①：生成AIの進展に対応したリテラシーの養成による「隔たり」の解消

生成AIの進展に伴い高度化・多様化する偽・誤情報に対して、単に知識として注意点を理解することにとどまらず、実際の情報環境に近い状況で「疑う」「確かめる」「拡散しない」といった判断行動をとれる実践的な耐性を継続的に育成することを目標とする。

生成AIは今後も性能向上と社会実装が進むことが見込まれ、その結果、技術進歩の速度と利用者のリテラシーの水準との間に隔たりが生じやすい。

本事業は、この隔たりを縮小するため、受け手が作り手の視点も踏まえて偽・誤情報の構造を理解し、未知の事例に対しても応用可能な判断力を獲得できる教育技術を整備する。これにより、技術進歩に対して利用者側の能力が遅れて追従するのではなく、変化に対応できる基礎的な耐性を社会に広く形成することを中長期の中心目標とする。

中長期的ゴール②：国際的なコンソーシアム設立と教材の現地化

国際的なコンソーシアムを設立し、東アジア、東南アジアを中心に教材の現地化を進めることを掲げる。偽・誤情報の課題は国境を越えて共通しつつも、言語、文化、社会制度、情報環境によって具体的な流通形態や影響が異なる。

そのため、単一の教材をそのまま展開するのではなく、地域ごとの実態に即した形で内容・事例・表現を調整し、現地の教育機関・関係団体と連携して運用可能な形に整えることが不可欠である。

本事業で開発する教材・サンドボックスは、偽・誤情報の典型的な手口を構造として学ぶ点と、生成AIを用いた生成によって実践的理解を得る点に特徴がある。これらは各地域においても応用可能な枠組みであるため、コンソーシアムにより共同で改善・評価・現地化を進め、継続的にアップデートされる国際展開の基盤を構築することを中長期の到達目標とする。

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケースの全体像

開発技術のユースケース

(1) 教育機関向け

① 探究学習での活用
(対面・オンライン)

② 探究学習での活用
(Eラーニング)

(2) 法人向け

① 新入社員研修
情報セキュリティ研修

② 専門職向け研修

③ メディア企業向け
編集体制強化

(3) 公的機関・自治体向け

① イベント・市民講座
啓発

② 自治体職員研修

各ユースケースにおいて「提供形態」と「期待される効果」を整理し、以下のスライドで詳述する。

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

(1) 教育機関向け

(2) 法人向け

(3) 公的機関・自治体向け

(1) ニーズ

教育機関における中心的ニーズは、学習者が日常的に接触する情報環境が複雑化する中で、探究学習等において大量の情報を扱う際に「誤った情報を正しい前提として取り込み、誤った価値観や判断基準を身につけてしまう」リスクを低減することにある。特に、探究学習を開始する段階では、情報収集・整理・発信の基盤が形成される時期であり、この段階で情報の見極め方を体験的に学ぶ意義が大きい。

また、学習指導上、情報科目（情報Ⅰ）において情報モラルやメディアリテラシーの観点から情報を見極める力を養うことが求められる一方、授業内では知識としての理解にとどまりやすく、実際の情報環境に近い状況で判断を繰り返す機会が不足しやすい。このため、情報Ⅰで学ぶ内容を「復習教材」として再体験させ、行動として定着させる教材需要が存在する。

以上のニーズに対し、本事業の対策技術は、偽・誤情報の典型パターンを体系的に学ぶことに加え、仮想環境でのAI生成を通じて「作り手の視点」も踏まえた構造理解を促すことで、学習者が未知の事例に遭遇しても応用可能な判断行動を獲得できる点に特徴がある。

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

(1) 教育機関向け

(2) 法人向け

(3) 公的機関・自治体向け

① 探究学習での活用

ユースケース（提供形態）

対面（出張授業形式）：

私立・国公立の中学校、高等学校等において、弊社スタッフが現地に赴き、ワークショップ形式の授業を実施する。探究学習を開始する学年を主対象とし、情報収集の基礎姿勢として「出所確認」「根拠点検」「拡散抑制」の判断行動を体験的に学ばせる。

オンライン（遠隔ワークショップ形式）：

Zoom等を用いて、地理的制約を超えて実施する。対面と同様に、個人作業、グループでの検討、発表、振り返りを組み合わせ、学習者が「もっともらしさ」に対して立ち止まる経験を得られるよう設計する。

期待される効果（教育効果）

- ・ 探究学習における情報収集の質の向上
- ・ 情報の信頼性や意図を多角的に検討する力の向上
- ・ SNSや検索で遭遇する偽・誤情報に対する判断行動の獲得
- ・ 受け身の注意喚起ではなく、体験に基づく心理的耐性の形成
- ・ 探究学習の教育的価値の担保（誤った価値観・判断基準の固定化を回避）

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

(1) 教育機関向け

(2) 法人向け

(3) 公的機関・自治体向け

② 探究学習での活用

ユースケース（提供形態）

Eラーニング（教材提供・自走利用）：

学校等に対し、弊社プログラム管理システムのライセンスを提供し、授業内または家庭学習で実施できる形で導入する。教員の運用負担を抑えつつ、学年進行に合わせて反復的に学べるよう、短時間モジュールで構成する。

探究学習開始前の導入教材としての利用に加え、探究学習の途中段階での学習教材としても活用可能とする。

さらに、情報 I で扱う情報モラル、メディアリテラシー、情報の信頼性に関する内容を、知識の確認にとどめず、実際の判断として反復させる「復習教材」として位置づける。情報 I で学んだ観点を、サンドボックス上の具体的な題材で適用させることで、授業内容の定着と行動化を図る。

期待される効果（教育効果）

- 探究学習に必要な情報リテラシーの標準的習得
- 情報 I の学習内容の定着
- 反復学習による習慣化
- 教員側の準備負担の軽減と、実施機会の拡大

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

(1) 教育機関向け

(2) 法人向け

(3) 公的機関・自治体向け

(2) ニーズ

法人領域では、生成AIの普及に伴い、外部からの攻撃（詐欺や情報盗難等）だけでなく、社内外で「生成AI由来の不確かな情報を誤って信じる」「それを基に発信・意思決定する」ことが企業価値の毀損につながり得る点が顕在化している。特に、社外発信を担う部門や、情報の正確性が事業に直結する部門では、偽・誤情報の混入や誤発信が信用問題に直結する。

したがって、単なる注意喚起ではなく、実務に近い形で判断行動を訓練できる研修ニーズが存在する。

また、情報セキュリティやコンプライアンスの観点からも、生成AIにより巧妙化した攻撃への耐性が求められる。従業員一人ひとりが不審点に気づき、確認し、適切な相談経路につなげる行動が組織的リスク低減の要となる。

① 新入社員研修・情報セキュリティ研修での活用

ユースケース（提供形態）

新入社員研修、情報セキュリティ研修、コンプライアンス研修の一部として導入し、偽・誤情報と生成AI活用の注意点を、体験型で学ばせる。受け手としての判断に加え、仮想環境でのAI生成を通じて構造を理解させ、日常業務での予防行動につなげる。

期待される効果

- 従業員の生成AI・SNSリスクの具体的理解
- 不審情報に対する確認・相談行動の定着
- 情報漏えい・誤送金・誤対応等の組織的リスクの低減
- 危機管理体制の実効性向上

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

(1) 教育機関向け

(2) 法人向け

(3) 公的機関・自治体向け

② 専門職向け研修での活用

ユースケース（提供形態）

広報、法務、マーケティング等の部門別研修として導入し、偽・誤情報への対応、発信物の点検、生成AI利用に伴うリスク管理を専門的に扱う。業務上の判断が企業価値に直結する部門に対し、ケースに基づく判断訓練として位置づける。

期待される効果

- ・ 誤発信の抑止
- ・ 組織全体の対応力向上

③ メディア企業向け研修・編集体制強化での活用

ユースケース（提供形態）

編集部門、制作部門、SNS運用部門等に対し、生成AI由来の偽・誤情報の典型パターンと検証の観点を、体験型で学ぶ研修として導入する。

期待される効果

- ・ 誤報・誤引用の抑止と訂正対応コストの低減
- ・ 企業信頼性の維持・向上

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

(1) 教育機関向け

(2) 法人向け

(3) 公的機関・自治体向け

(3) ニーズ

災害・行政手続・地域課題に関する情報がSNS等で流通する中、偽・誤情報が住民の不安増幅や混乱を招き、問い合わせ増・訂正広報・対応工数など行政コストを増大させる。詐欺・なりすまし等の被害も住民生活に直結する。

一方で、講演型啓発は受け身になりやすく、学習後の行動変容につながりにくい。

したがって、住民が「出所確認／根拠の確認／拡散抑制」を体験的に反復できる、継続運用可能な啓発・研修手段が求められる（常設展示・講座・地域連携などに組み込みやすい形）。

① 住民向け啓発（図書館・公民館・生涯学習・防災イベント等）

ユースケース（提供形態）

短時間モジュール（体験→振り返り）を常設展示やイベント枠で提供。地域課題（災害・医療・行政情報など）に近い題材で、判断行動を反復させる。

期待される効果

- ・ 偽・誤情報の拡散抑制
- ・ 詐欺・なりすまし被害の予防
- ・ 家族・世代間の対話促進
- ・ 地域のデジタル安全性向上

2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース

(1) 教育機関向け

(2) 法人向け

(3) 公的機関・自治体向け



② 自治体職員・指定管理者向け研修（広報・窓口・防災・危機管理）

ユースケース（提供形態）

部門別研修として導入し、住民対応で起きる典型ケース（誤情報への問い合わせ、訂正広報、SNS運用）をもとに、判断・対応フローを訓練。

期待される効果

- 問い合わせ・訂正対応コストの低減
- 広報品質の向上
- 危機対応の標準化と属人性低下

③ 学校・地域連携（中高・地域団体との共同実施）

ユースケース（提供形態）

学校の探究／情報教育と、自治体の啓発施策を接続し、地域課題テーマで共同開催（学校実施＋地域公開回など）。

期待される効果

- 若年層への浸透力向上
- 地域への波及（家庭・地域コミュニティ）
- 継続的な啓発基盤の形成

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

3-1. 技術開発の全体像

技術開発に係る取組・成果の全体像

本事業における対策技術は、「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」を成立させるために必要となる技術群を統合したものである。単に偽・誤情報の知識を提示するのではなく、受講者が生成・改変の過程を自ら操作し、誤認が生じる条件や心理的誘導の働き方を体験的に把握したうえで、実際の情報環境で再現可能な確認行動へ接続する一連の学習過程を、同一の仕組みとして実装した点に特徴がある。

本技術は、学習効果と現場運用の両立を図るため、次の通りに整理して構成している。各領域は独立した要素ではなく、学習体験の一貫性を担保するために相互に連動する設計としている。

技術1：偽・誤情報サンドボックス

偽・誤情報が「なぜもっともらしく見えるのか」「どのように心理的誘導が組み込まれるのか」を体験的に理解するための演習環境である。加えて、偽・誤情報サンドボックスはモーダル横断（テキスト・画像、将来的には音声・動画）を前提として設計し、フェーズ1では文書と画像を対象に開発する。これにより、文章の誤情報だけでなく、視覚情報を伴う誤情報の生成・検証も含め、現実の情報環境に近い学習条件を確保する。

技術2：学習設計

学習設計は、変化する偽・誤情報の手口に対し、学習を一過性で終わらせず、反復によって確認行動を定着させるための設計領域である。偽・誤情報の典型パターンを学習単位として整理し、受講者が理解から実践、振り返りまでを一貫して行えるように学習導線を標準化している。さらに、新たな手口の出現に対応できるよう、学習単位の追加・更新が可能な拡張性を前提として整理している。

技術3・4：運用/管理システム、UI/UX設計

授業・研修の現場で「実施できる」状態を担保し、学習体験の品質を安定させるための設計領域である。学習者側の体験導線だけでなく、実施者側が受講状況を把握し、成果を整理し、再学習へつなげる運用が成立することを目的として整備している。

3-1. 技術開発の全体像

技術開発に係る取組・成果の全体像

本技術は、「偽・誤情報サンドボックス」「学習設計」「運用／管理システム」を一体として設計することで、既存のプレバンキング技術と比較して優位性を持つ。

① 安全な偽・誤情報サンドボックスによる生成体験

受講者が偽・誤情報を生成・改変する体験を通じ、その構造や心理的誘導の仕組みを理解する設計。生成内容の制御・不適切出力の抑止・生成過程の記録等の安全管理を組み込んだ教育専用環境であり、「生成の仕組み」を体験的に理解させる点が本技術の中核。

② 事例に基づく学習内容の体系化と迅速な追加設計

実在する偽・誤情報事例を基に特性や拡散構造を整理し、「事例の提示→生成体験→解説→テスト」の共通フォーマットで学習を構成。新たな手法にも同構造で迅速に追加可能な拡張性を備える。

③ 継続的学習設計

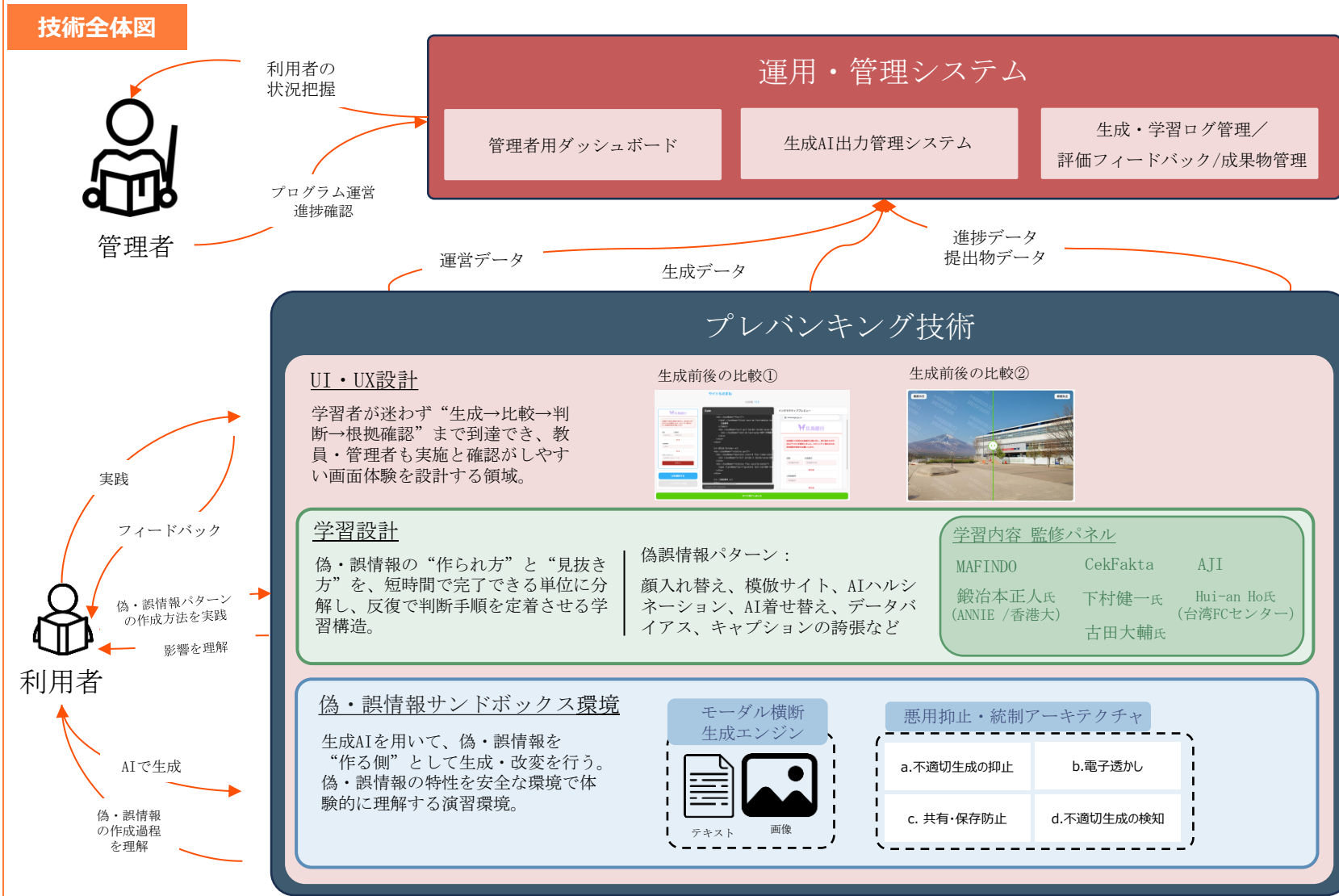
運用／管理システムにより学習者の進捗やスコアを管理者が把握可能（学習ログの記録）。進捗遅れの早期把握・フォローに加え、ゲーミフィケーション技術で「続けたい」学習導線を実現し、反復・再学習を支援。

他技術との比較

項目	既存プレバンキング技術	海外研究型ゲーム	本技術
学習形式	動画視聴・事例解説	ゲーム形式（限定シナリオ）	生成制御型サンドボックス
偽・誤情報の生成体験	なし	一部簡易シミュレーション	実際に生成AIを操作
安全管理	明確な制御設計なし	教育ゲーム範囲	出力制御・ログ記録・持出抑止
学習内容の構造	個別事例中心	ストーリー中心	事例に基づく体系化設計
新規事例への対応	更新ごとに再設計	シナリオ追加が必要	共通フォーマットで迅速追加
学習の継続支援	単発型	単発型	進捗把握＋フォロー設計
ログ取得	限定的	限定的	学習・生成履歴を記録
社会実装前提	教材	研究用途中心	教育現場運用を前提

3-1. 技術開発の全体像

技術開発に係る取組・成果の全体像



3-2. 技術開発の個別詳細

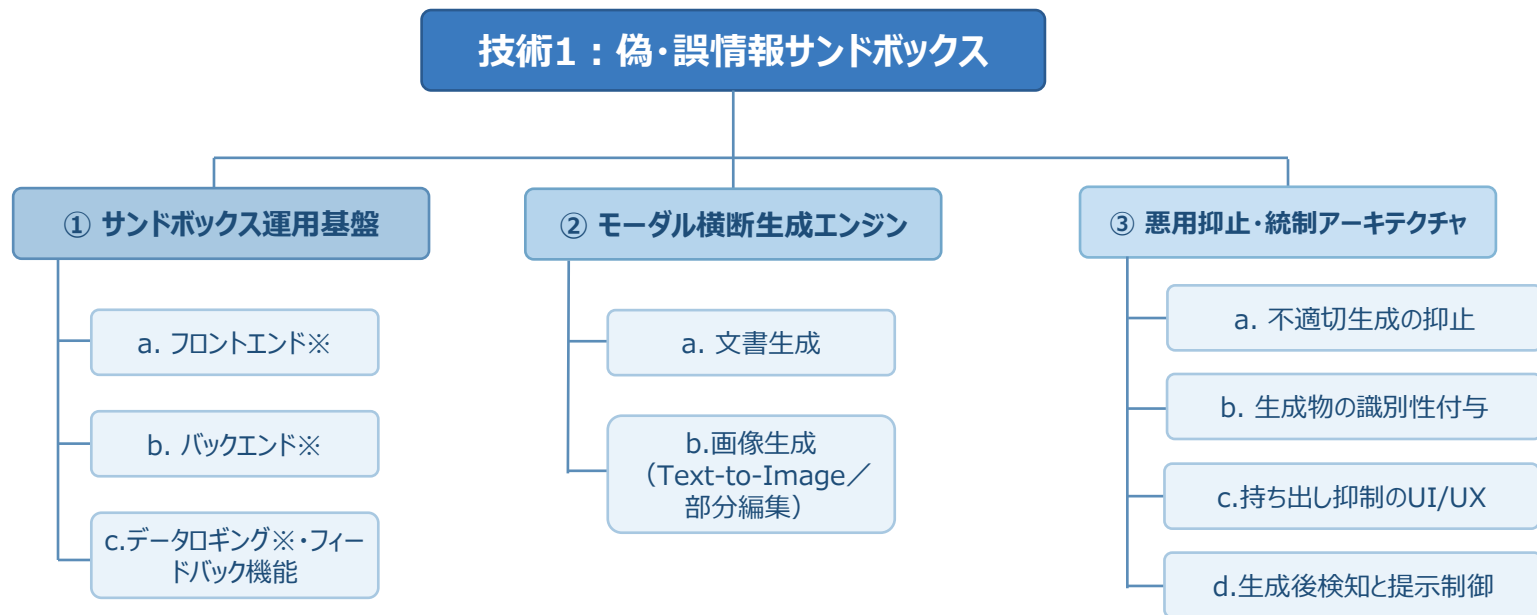
技術1：偽・誤情報サンドボックスについて

目的と設計方針

偽・誤情報サンドボックスは、偽・誤情報が「なぜもっともらしく見えるのか」「どのように心理的誘導が組み込まれるのか」を体験的に理解するための演習環境である。

モーダル横断（テキスト・画像、将来的には音声・動画）を前提として設計し、フェーズ1では文書と画像を対象に開発する。これにより、文章の誤情報だけでなく、視覚情報を伴う誤情報の生成・検証も含め、現実の情報環境に近い学習条件を確保する。

加えて、本事業の偽・誤情報サンドボックスは、教育用途としての実践性と運用性を両立させるため、主として次の三つの技術要素により構成する。



※用語：フロントエンド＝受講者が操作する画面／バックエンド＝裏側で判定・制御する仕組み／データロギング＝操作や結果を記録し、後から確認できる機能

3-2. 技術開発の個別詳細

技術1：偽・誤情報サンドボックス - サンドボックス運用基盤

① 運用基盤

② 生成エンジン

③ 悪用抑止

サンドボックス運用基盤は、学習者が「生成して終わり」にならないように、生成・比較・振り返りまでを一続きの導線として成立させる実装基盤である。同時に、生成AIの呼び出しやアクセスを統制して授業や研修でも安定稼働させ、操作ログや生成結果、フィードバックを回収することで、学習効果の可視化と継続的な改善に接続する役割を担う。主として次の三つの要素により構成する。

a. フロントエンド※

※フロントエンド = 受講者が操作する画面

受講者が直感的に操作できるよう、Webアプリとしての操作画面を整備する。プロンプト入力、トーン選択、出力モード選択（テキスト・画像）等を一画面内で扱える構成とし、学習者が試行錯誤を重ねやすい導線を確認する。加えて、本教材の学習効果を高めるため、受講者が作成した偽・誤情報の「どこが、どのように変わったか」を把握しやすいUI設計を行う。

具体的には、画像改変の場合に元画像と生成後画像を並置し差分を比較できる表示、あるいは受講者が意図した改変内容（生成しなかった状態）と実際の生成結果を照合し、意図と結果のずれを認知できる仕組みを設ける。

左：参照画像
右：参照画像を参考に模倣した生成サイト



画像改変前・後を中央のバーを用いて確認可能



3-2. 技術開発の個別詳細

技術1：偽・誤情報サンドボックス - サンドボックス運用基盤

① 運用基盤

② 生成エンジン

③ 悪用抑止

b. バックエンド※

※用語：バックエンド=裏側で判定・制御する仕組み/ API=外部システム連携機能

生成AIとの通信・制御はAPI※で処理し、クラウド上の生成AI基盤と連携する。
教育用途としての安全性を前提に、生成リクエストに対して必要な制御を付与し、模擬偽・誤情報の生成が学習に必要な範囲に収まるよう設計。

あわせて、処理の安定性確保、アクセス制御、流量制限等の基本対策を講じ、授業・研修等の実施環境で支障なく運用できる基盤とする。

c. データロギング※・フィードバック機能

※データロギング=操作や結果を記録し後から確認できる機能

学習効果の可視化および継続的改善のため、利用者の操作情報を自動的に記録し、教材改善、安全管理、学習効果分析に活用する。具体的には、利用者が入力したプロンプト、生成された出力（テキスト・画像）、システムが表示した警告等を、データ分析基盤およびオブジェクトストレージに保存する。

また、利用後のアンケート導線に加え、システム内フィードバック機能を提供し、利用者が感じた難易度、理解度、操作性、改善要望等を収集する。これらの定量・定性情報を統合して分析し、教材内容、UI設計、生成制御、安全対策等の改善に反映することで、ユーザー体験の継続的改善を図る。

3-2. 技術開発の個別詳細

技術1：偽・誤情報サンドボックス - モーダル横断生成エンジン

① 運用基盤

② 生成エンジン

③ 悪用抑止

テキストと画像の両方で「もっともらしく見える偽情報」を短時間で生成できる状態を作り、偽・誤情報が成立する仕組みを体験として学ばせる中核機能である。生成前後の差分を意識させることで、どこがどう欺きに寄与したのかを理解させ、最終的に発信源・根拠・関連情報といった確認ポイントに沿って見抜く演習素材を提供する役割を担う。

a. 文書生成

※API=外部システム連携機能

文書生成には、クラウド上の大規模言語モデルをAPI※で利用する。短文の偽・誤情報生成にとどまらず、報告書や論文等の長文資料を読み込ませ、その内容に基づく要約文の作成など、資料体裁の整った文章を生成できる構成とする。

また、文章生成に加えて、簡易なプログラムコード等の形式で情報を出力できるようにし、外形の完成度が高い生成物が短時間で作成され得ることを学習対象として扱う。

これらの体験を通じて、体裁の整合性や表現の自然さが内容の正しさを保証するものではないこと、ならびに受け手側において根拠や関連情報の確認等が不可欠であることを理解させる設計とする。

レポート作成の例

生成回数: 1/3

パン・全粒穀物

「全粒穀物パン」が命を救う

「パンを食べると長生きする」——これは78万人以上のデータに基づく、科学的に証明された事実である。感覚や伝説ではなく、数字が真実を語っている。

米国心臓協会の権威ある学術誌「Circulation」に2016年に掲載されたメタ解析は、全粒穀物摂取と死亡リスクの関係について、これまで最も包括的な証拠を提示した。米、英国、スカンジナビア諸国などで実施された14件の研究結果を統合し、総計76,078人の参加者、97,867件の死亡を解析した結果である。

結果は明確だった。全粒穀物（特に全粒粉パンなど）摂取量が最も多いグループは、最も少ないグループと比較して、全死因死亡リスクが16%低下し、心血管疾患による死亡が18%低下、がんによる死亡が2%低下していた。さらに用量反応解析では、全粒穀物を1日16g増やすごとに全死因死亡リスクが7%、心血管死亡が9%、がん死亡が5%ずつ低下するといふ、きれいな直線関係が確認された。全粒粉パン1枚（約40g）は、全粒粉100gに相当する。この関係を基

パンの種類	全粒粉含有率	繊維含有率	糖質含有率	脂質含有率	水分含有率
全粒粉パン	100%	10g	45g	10g	35g
全粒粉入りパン	80%	8g	48g	12g	32g
全粒粉入りロールパン	60%	6g	50g	14g	30g
全粒粉入り食パン	40%	4g	52g	16g	28g
全粒粉入りロールケーキ	20%	2g	54g	18g	26g

すべて完了しました

コード生成の例

生成回数: 1/3

投資スクール

たった2日で +80万円

「投資タカヒロ」は、今すぐ投資スクールをオープンしました！この特別な機会に、人生を変えよう保証です。貴重！時を、逃してください。メールでもOK、takahiro@ipencil.com 頂けるよあなたに大きなお礼を申し上げます！

```

Code
<div class="wrapper">
  <div class="text">
    <p>たった2日で +80万円</p>
  </div>
  <div class="caption">
    <p>「投資タカヒロ」は、今すぐ投資スクールをオープンしました！この特別な機会に、人生を変えよう保証です。貴重！時を、逃してください。メールでもOK、takahiro@ipencil.com 頂けるよあなたに大きなお礼を申し上げます！</p>
  </div>
</div>
    
```

UIを選択する

ウェブサイトに追加

すべて完了しました

3-2. 技術開発の個別詳細

技術1：偽・誤情報サンドボックス - モーダル横断生成エンジン

① 運用基盤

② 生成エンジン

③ 悪用抑止

b.画像生成（Text-to-Image／部分編集）

※API=外部システム連携機能

画像生成には、画像生成モデルをAPIで利用する。受講者が「ゼロからの画像生成」と「既存画像の改変」の双方を体験できるよう、次の二機能を中核として構成する。

ゼロからの画像生成：

テキスト指示により新たな画像を生成し、「現実には存在しない風景」「架空の製品」等を作成する体験を通じて、視覚情報が付与されることで説得力が高まる構造を理解させる。

既存画像の改変（インペインティング）：

画像内の特定範囲を指定し、指示に従って当該部分を再描画する。受講者は、提供されたサンプル画像または自分で用意した画像をアップロードし、インターフェース上のブラシツールで編集したい箇所を選択（マスク範囲設定）する。変更内容をテキストで指示すると、光源や質感に自然に馴染む形で該当部分が再生成され、人物や物体の入れ替え、背景の変更等が実現される。

顔画像の入れ替えの例



生成結果



画像の部分編集の例



3-2. 技術開発の個別詳細

技術1：偽・誤情報サンドボックス - 悪用抑止・統制アーキテクチャ

① 運用基盤

② 生成エンジン

③ 悪用抑止

a. 不適切生成の抑止

生成AI側の安全機能を有効化したうえで、教材目的に照らして不適切性が高い語彙・文脈を抑止する独自フィルタを重ね、入力（プロンプト）および生成時の逸脱を抑制する。

b. 生成物の識別性付与

生成されたテキスト・画像が外部に持ち出され、二次利用等により悪用されるリスクを低減するため、生成物の識別性を担保する措置を講じる。具体的には、生成物に対し、当該教材環境で生成されたことが識別可能となる電子透かしを付与する。

a. 不適切生成抑止フィルタ

服装の説明（英語推奨）

水着

詳しく説明するほど良い結果が得られます

🚫 禁止アイテムを含むプロンプトは自動的に編集されます

▶ 高度な設定

服を変更する (残り3回)

禁止キーワードが含まれています: "水着"

b. 電子透かし



3-2. 技術開発の個別詳細

技術1：偽・誤情報サンドボックス - 悪用抑止・統制アーキテクチャ

① 運用基盤

② 生成エンジン

③ 悪用抑止

c. 持ち出し抑制のUI/UX

保存・転載を前提としない体験設計とし、安易な保存動機を生みにくい表示・導線を採用するなど、ブラウザ上での共有・保存が容易にならないUI抑止を組み合わせ、行動レベルでの拡散を抑える。

d. 生成後検知と提示制御

生成後の出力について、画像理解等の検知処理を通じて不適切性を判定し、疑いがある場合は表示制限・差し戻しを行うなど、生成後段でも安全性を担保する。

c. コピーガード機能



d. 生成後検知機能



3-2. 技術開発の個別詳細

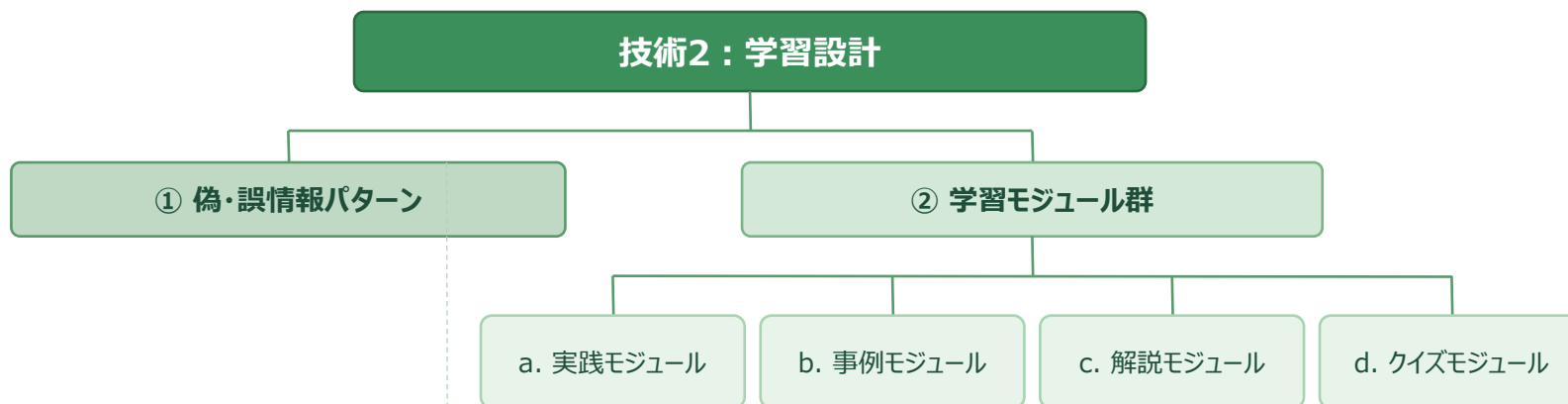
技術2：学習設計について

概要

偽・誤情報の手口は日々変化しており、特定のパターンだけを教える方法ではすぐに陳腐化してしまうという課題がある。本技術では、この課題に対応するため、どのパターンが追加・更新されても学習体験全体の整合性が保たれる「更新可能な学習アーキテクチャ」を設計した。具体的には、偽・誤情報の各手口を独立した学習単位（モジュール）として管理し、それぞれに共通のモジュール列（実践→事例→解説→クイズ）を適用する構成とした。この設計により、新しい手口が出現した際にも既存の学習フレームワークを壊すことなく、コンテンツを柔軟に拡張できる。

この構造が持つ3つの役割

- ① **拡張性**：
新たな偽・誤情報の手口が出現した場合でも、新しいパターンを追加するだけで対応でき、全体の学習体験の一貫性を損なわずに拡張が可能
- ② **再現性**：
学習の導線を共通フレームで固定しているため、学校・企業・自治体など異なる運用現場でも、実施者によるブレを最小限に抑えた展開ができる
- ③ **転移性**：
判断の根拠を言語化させるプロセスを組み込むことで、学習したパターン以外の偽・誤情報に遭遇した場合にも応用できる批判的思考力を育成する



学習の継続と反復を動機づけるため、ゲーミフィケーションの仕組みを導入した。パターンごとの進行状況や完了度を可視化し、学習者が自分の到達点を常に把握できるようにしている。スコアは単なる正答率ではなく、確認ポイントの達成度や判断根拠の明確さといった質的な指標にも紐づけることで、弱点を認識した上で再挑戦する学習の循環を生み出す。また、接種理論※に基づき、獲得した耐性が時間とともに減衰することを踏まえ、一定期間後に再学習を促すブースターショット※の仕組みを組み込むことで、学習効果の長期的な持続を図っている。

※接種理論（Inoculation Theory）：誤情報の弱い形態にあらかじめ触れさせ、心理的抵抗力を形成する理論

※ブースターショット（Booster Shot）：効果の減衰を防ぐため、一定期間後に再度学習機会を設ける再介入手法

3-2. 技術開発の個別詳細

技術2：学習設計 – 偽・誤情報パターン

① 偽・誤情報パターン

② 学習モジュール群

偽・誤情報は「何を学ぶか」が曖昧だと、受講者にも現場にも学習の焦点が残りにくい。このため本教材では、専門家との意見交換を踏まえて、教育的に扱うべき“典型手口”を「偽・誤情報パターン」として学習単位に整理した。あわせて、生成AIの発展により手口が増え続ける前提に立ち、パターンは固定せず、社会で新たな作成パターンが出現するのに比例して順次追加できる拡張型として設計している。取り扱うパターンは下記の通りである。

偽・誤情報パターン

※偽・誤情報パターンの名称は仮称です。

01 Face Swap	顔の差し替えなどで、実在人物が「言っていない・していない」言動をしたように見せる手口を学ぶ。どこが怪しいか、何を確認すべきかを、比較と検証の手順として身につける。
02 ものまねサイト	本物そっくりの見た目と利用者をだますサイトの仕組みを学ぶ。URLや運営主体など「見た目以外で確認するポイント」を押さえ、誤入力や被害を防ぐ判断を身につける。
03 演出家	同一の事実素材でも、キャプションやBGM、フィルターによって受け取り方が変わることを学ぶと同時に、情報が「どう演出されているか」を見抜き、判断が誘導されない読み方を身につける。
04 嘘つきAI	生成AIが、もっともらしい文章・引用・数値を提示し得ることで、間違った情報や架空の情報が生成されてしまうことを理解する。
05 Inpainting	画像の一部分だけを自然に差し替えて、事実をねじ曲げる手口を学ぶ。元画像との比較や改変の疑い点を押さえ、拡散前に確認する手順を身につける。
06 データバイアス	数字やグラフがあっても、条件の置き方で結論が変わることを学ぶ。母集団、期間、比較条件など「先に見るべき前提」を押さえ、数字にだまされない見方を身につける。
07 動画の切り抜き	短い切り抜きで、発言や行動の意味を反転させる手口を学ぶ。前後の文脈や元動画を確認し、早合点しない判断の仕方を身につける。
08 論文作成	生成AIを用いて、どんな主張にももっともらしいデータをつけていかにも正当な意見だと表現できてしまう危険性と、関連情報や真逆の主張データを見る習慣を身につける。
09 偽キャプション	画像や事実素材に対してそれらしいキャプションを付すことで、受け手の理解が容易に歪むことを理解する。
10 AI装着替え	人物の服装・属性情報を改変することで、当事者像や社会的評価を誤誘導できることを理解する。

3-2. 技術開発の個別詳細

技術2：学習設計 - 学習モジュール群

① 偽・誤情報パターン

② 学習モジュール群

偽・誤情報の「典型パターン」を学習単位として整理し、各パターンについて、学習者が理解から実践、振り返りまでを一貫して行えるように構成した。各パターンは、原則として次のモジュール群で構成している。

実践モジュール

偽・誤情報サンドボックスを用い、当該パターンに対応する生成・改変の過程を体験させる。生成条件や編集要素を操作し、生成結果の差分を観察することで、「もっともらしさ」が生じる要因を理解させる。

実践モジュール例①

ハルシネーション体験

1 ステップ1
会話

あなた
御所達の戦いについて詳しく教えてください

AI
御所達の戦いについてですね。
御所達の戦いは、1586年（天正14年）に、九州を平定しようとした豊臣秀吉と、それに抵抗する島津氏との間で行われた戦いの一部です。
具体的には、秀吉の九州征伐の最中、島津氏の拠点のひとつであった**御所浦島（現在の熊本県上天草市御所浦町）**において、豊臣軍が島津軍を破った戦いです。
戦いの背景:
1586年、豊臣秀吉は九州統一を目指し、大規模な軍事行動（九州征伐）を開始しました。
島津氏はこの九州統一の最大の障害であり、各地で激しく抵抗していました。
御所浦島は、豊後（現在の Oita）や日向（現在の Miyazaki）方面から九州本土への進軍ルート、あるいは島津氏の重要な拠点として重要な役割を果たしていました。

このステップは完了しました

次へ

実践モジュール例②

AI caption

AIフェイクキャプション

画像をアップロードすると、AIが2種類の偽キャプションを生成します

残り実行回数: 4/5

同じ画像から生成された2種類の偽キャプション
AIは画像から正反対の印象を与える投稿を作成できます

真上・批判タイプ 美化・感動タイプ

匿名ユーザー @anonymous_user · 2分前

これは、文化祭の出展地を受けた後に、発表のために再び正門前に集まって記念撮影をした元在校生グループの画像。

これだけ間接的にして選挙になっても、また学校の前でヘラヘラできる神経が本当に理解できない、先生たちが可哀想すぎる。



30 59 2.2k

匿名ユーザー @anonymous_user · 10分前

これは、経済的な事情で塾に通えない子供たちのために、週末を返上して無償で学習支援を行っている学生ボランティアグループが、活動を終えて充実した笑顔を見せている一瞬。

自分の選びたい時間を削ってまで「子供たちの未来の選択肢を広げたい」と活動を続ける彼らの軽妙な思いに、胸が熱くなりました。見返りを求めず誰かのために行動できる彼らのような若者がいることが、何よりの希望です。



41 300 1.1k

実践モジュール例③

プリセットを選択 加工後の動画

動画
感動的な背景音楽

効果
明るく楽しい雰囲気

面白い
知られたい印象

字幕を編集

テキスト 1
知れば消される

テキスト 2
日本のヤバい話

日本のヤバい話

加工後の動画

知れば消される
日本のヤバい話

一時停止

3-2. 技術開発の個別詳細

技術2：学習設計 - 学習モジュール群

① 偽・誤情報パターン

② 学習モジュール群

偽・誤情報の「典型パターン」を学習単位として整理し、各パターンについて、学習者が理解から実践、振り返りまでを一貫して行えるように構成した。各パターンは、原則として次のモジュール群で構成している。

事例モジュール

現実に拡散した事例を題材として、学習者が偽・誤情報の成立と影響を具体的に把握できるよう、次の観点で整理して提示する。

拡散した言説、考えられる発信者の意図、状況、もしみんなが信じたら？、ファクトチェックの結果。これにより、単なる正誤判断にとどまらず、なぜ拡散するのか、何が社会的リスクとなるのかを理解させる。

事例モジュール例①

02. ものまねサイト 事例1.

拡散した言説	あなたならどう判断？	状況
	<p>鹿児島県東部の被災地支援を名乗り【緊急】鹿児島県東部 支援活動にご支援をお願いします！などの文言とともに、目標金額や達成率、支援人数を表示して仮想通貨 USDT での寄付を呼びかける寄付サイトへのリンクがLINE のオープンチャットや各種 SNS などで共有されている。</p> <p>見た目やレイアウトはふるさと納税サイト「ふるさとチョイス(GO)」で公益社団法人が行っていた鹿児島県東部被災地のページに酷似しており、ロゴも勝手に使用されている。</p> <p>しかし、通貨単位は日本円ではなく仮想通貨 USDT になっており、寄付ボタンを押すと名前、メールアドレスなどの個人情報が表示され、その仮想通貨送金を促す仕掛けになっていた。</p>	<p>2024年1月1日の鹿児島県東部地震後</p> <p>考えられる発信者の意図</p> <ul style="list-style-type: none"> 寄付フォームを通して集めた氏名やメールアドレスなどの個人情報の悪用 送られた寄付金に対する金銭目的 <p>もしみんなが信じたら？</p> <p>本物とよく似たサイトの構成や実在団体のロゴを用いているため、正味の寄付先と勘違いして送金してしまう可能性が高い。その結果、本来、被災地に届くはずだった支援金が詐欺グループに流れたり、正味の寄付サイトや支援団体の信頼が損なわれるリスクがある。</p>

誤り 正味の寄付ページを確認したところ、検証対象のサイトとは URL やロゴが異なり、仮想通貨での寄付も受け付けていないことが分かった。

[詳しく見る \(外部サイト\)](#)

事例モジュール例②

01. なりすまし 事例2.

拡散した言説	あなたならどう判断？	状況
	<p>実業家の著名人になりました AI 生成とみられる偽の動画広告がYouTube などで拡散されています。</p> <p>動画の中ではそっくりの人物が視聴者に株式投資を勧誘し、「100倍もしない人、今すぐ耳を離してください」と呼びかけようとして「LINE を無料でお試し「受取」と送るだけで翌日振り上がりする銘柄がわかる」「毎週60%の利益が期待できる優良銘柄情報も入手できる」といった言葉はめがけられている。</p> <p>動画上部には会社名が表示され、著名人が関わる投資サービスと印象があるかのような印象を与えている内容となっています。</p>	<p>YouTube 広告上</p> <p>考えられる発信者の意図</p> <ul style="list-style-type: none"> 金銭目的での誘導 登録させた LINE や連絡先を通じた個人情報の取得 本人や関連サービスのブランド・イメージの低下 <p>もしみんなが信じたら？</p> <p>生活費や老後の貯金をつき込んでしまう人が流出し、返済できない借金を抱えて日常生活が立ちゆかなくなる人が生まれる</p>

誤り 一部不自然な日本語で話していたり、顔の動きと発言内容が一致していない場面が確認でき、AIによって生成された映像である可能性が高いことが分かる。また、著名人が運営するサービスの公式サイトや SNS で、これらの広告は自分や該当サービスとは一切関係のない完全な詐欺であると明白に注意喚起を行っている。

[詳しく見る \(外部サイト\)](#)

3-2. 技術開発の個別詳細

技術2：学習設計 - 学習モジュール群

① 偽・誤情報パターン

② 学習モジュール群

解説モジュール

当該パターンの特徴や誤認が生じやすい点を整理した上で、学習者が実際の情報環境で用いることができる対策行動として、次の三つの観点から学ばせる。

- 発信源を見る（誰が発信しているのか、公式情報か、なりすましや改変の可能性はないか等）
 - 根拠を見る（一次情報・データ・引用の実在性、条件や前提、検証可能性等）
 - 関連情報を見る（他媒体・他視点との突合、反対情報や追加情報の確認、時系列・文脈の確認等）
- これにより、判断の観点を「手順」として定着させ、再現可能な確認行動につなげる。

クイズモジュール

学習内容の理解度を確認し、どこを疑い、何を確認するかという判断の根拠を言語化させる。

クイズモジュール例①

拡散した言説



あなたならどう判断？

実業家の著名人になりました AI 生成とみられる偽の動画広告が YouTube などでも掲載されています。

動画の中ではそっくりの人物が視聴者に株式投資を勧誘し、「100%にしたい人、今すぐ耳を刺してください」と呼びかけたうえで「LINE を無料でお返し『変換』と表示されたら即座にブロックし、その後のメールも読まない」という旨のメッセージを送る。この動画は「毎週10%の利益が期待できる優良株情報も入手できる」といった現実にはありえない誇大宣伝を繰り返しながら LINE 登録を促しています。

動画上部には会社名が表示され、著名人が関わる投資サービスと関係があるかのような印象を与える内容となっています。

状況

YouTube広告上

考えられる発信者の意図

- ・金銭目的での詐欺
- ・登録させたLINE や連絡先を通じた個人情報取得
- ・本人や関連サービスのブランド・イメージの低下

もしみんなが信じたら？

?

誤り

一部不自然な日本語で話していたり、顔の動きや発言内容が一貫していない場面を確認でき、AI によって合成された映像である可能性が高いことが分かる。また、著名人本人の発言するサービスの公式サイトや SNS で、これらの広告は自分や該当サービスとは一切関係のない完全な詐欺であると明言し注意喚起を行っている。

もしみんなが信じたらどんなことが起こる？

クイズモジュール例②

レベルテスト

問題 1 / 1

10 ポイント

「なりすまし詐欺」の正しい説明は下記のうちどれか。

「なりすまし詐欺」の正しい説明は下記のうちどれか。

匿名で情報も発信すること全般

本人や顔面できる範囲に成り代わり、本物のように見せて人をだます行為

SNS での誹謗中傷全般

不正アクセスによるアカウント乗っ取りのみ

答えを確認

3-2. 技術開発の個別詳細

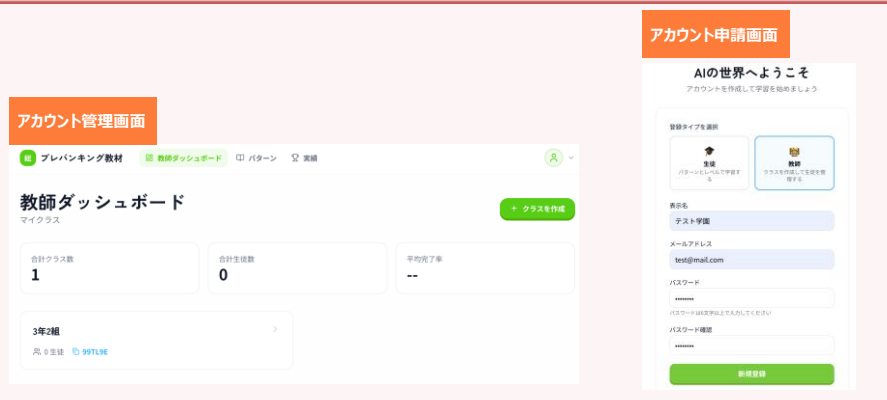
技術3 : 運用・管理システム

教材運用・管理システムは、授業・研修の実施者（教員、研修担当者等）が、受講者の受講状況を把握し、学習成果を整理したうえで次の指導・再学習に接続できるようにするための運用基盤として整備した。主な機能は以下のとおりである。

アカウント申請システム

教員や研修担当者が自らWeb上で申請を行うだけで、教材利用に必要なアカウントを即時発行できるシステムを構築した。

申請完了後にアクセスコードが自動発行され、受講者はそのコードを入力するだけで本教材へアクセスが可能となる。これにより、管理者への個別依頼や手動のアカウント発行作業が不要となり、導入から利用開始までのリードタイムを大幅に短縮できる。



受講者管理

受講者をクラス・グループ単位で登録・管理し、実施回（授業回・研修回）に紐づけて運用できるようにした。

これにより、現場での一斉実施や複数回の実施においても、対象者の取り違えや記録の分散を防止できる。



3-2. 技術開発の個別詳細

技術3：運用・管理システム

進捗把握

学習者がどのパターン・どのモジュールまで完了したかを可視化し、未完了者のフォローや再受講の促進に役立てられるようにした。

特に、事例→実践→解説→クイズの各ステップの到達状況を確認できる構造とし、学習者がどこで停滞しているかを把握可能とした。

進捗確認ページ



結果出力

授業・研修後に、実施者が成果を整理できるよう、進捗・到達度等の結果を出力できるようにした。

これにより、実施報告や受講者へのフィードバック、次回実施時の改善検討に必要な情報を、運用の中で確実に蓄積できる。

個人結果確認ページ



提出物確認ページ



3-2. 技術開発の個別詳細

技術4 : UI/UX設計

UI/UX設計は、学習者が「何をすべきか」「次にどこへ進むか」を迷わず理解し、短時間でも学習を完了できることを重視して整備した。特に、偽・誤情報対策は学習者の前提知識に差が出やすいため、操作負担を下げることを基本方針とした。

全体導線（ステップ進行）：

学習者が、実践→事例理解→解説→テストの順に自然に進行できるよう、画面遷移と表示情報を整理した。各ステップの目的と到達目標を明示し、学習の現在地が把握できるようにした。

実践モジュールの導線

偽・誤情報サンドボックスの操作は、学習者が試行錯誤しやすいように、入力・生成・比較・振り返りの流れを一画面内で理解できる構成とした。

生成結果については、元素材との比較や差分把握がしやすい表示を採用し、学習者が「どの操作が、どの印象差を生んだか」を捉えられるようにした。

テストモジュールの導線

学習者が判断根拠を整理しながら回答できるように設計し、回答後には学習内容に立ち返れるよう、解説モジュールへの接続を明確化した。

これにより、誤答を単なる失敗で終わらせず、再学習に直結させる。

解説モジュールの導線

解説は、学習者が実践で体験した内容を、確認行動として整理できるように設計した。

すなわち、発信源・根拠・関連情報という観点に沿って、何を見て、どう判断するかを明確にし、次のテストや次回学習に転用できる形で提示する。

以上のとおり、付帯要素として整備した教材運用・管理システムおよびUI/UX導線は、教材の継続運用と学習効果の定着を支える基盤であり、授業・研修の現場で無理なく実施できる状態を実現している。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

4-1. 検証及び調査の全体像

検証及び調査に係る取組・成果の全体像

本事業における有効性の検証及び調査は、「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」が、受講者の判断に与える影響を、感想の集約に留めず、教育現場で説明可能な形で把握することを目的として実施した。

検証の中心は、①受講直後に判断の質が改善しているか、②教材として継続利用に耐える体験価値が成立しているか、③一定期間後も学習した確認ポイントに基づく判断が維持されるかの三点を同時に確認する点に置いた。

KPI	詳細	目標	成果
学習効果	受講前後で同一条件の設問に回答させ、偽・誤情報を適切に疑える方向に信頼評価が変化するかを測定する構造とした。	80%以上の参加者が21%以上のスコア向上	参加者の約55%が約37%のスコア向上
ユーザー満足度	楽しさ・理解実感・日常での有用感を問うことで、単なる知識獲得ではなく、継続利用・普及に耐える体験価値が成立しているかを確認する構造とした。	「満足」以上が「85%以上」を占める	約98%の参加者が「満足」「とても満足」と回答
行動変容持続	受講から一定期間経過後に追跡調査を行い、学習した確認ポイント（発信源・根拠・関連情報）に沿って偽・誤情報を見破れるかを測定する構造とした。	「50%」の参加者が3ヶ月経過後もプレバンキング行動を実践	怪しい情報に接触した参加者のうち75%が確認行動を継続

※KPIの目標値は、先行研究や類似のゲーム型介入で報告される改善幅を踏まえつつ、授業時間内で実施可能であること、回答負担が過度にならないこと等、教育現場での運用条件と両立する水準として設定した。

4-2. 検証及び調査の個別詳細

実証の目的と検討事項

本事業における有効性の検証及び調査は、「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」が、受講者の判断に与える影響を、感想の集約に留めず、教育現場で説明可能な形で把握することを目的として実施した。

検証の中心は、

- ①偽・誤情報に対する判断力が改善しているか
- ②教材として継続利用に耐える体験価値が成立しているか
- ③一定期間後も学習した確認観点に基づく判断が維持されるかの三点を同時に確認する点に置いた。

このため、本検証では評価体系（KPI）を下記のとおりとした。

KPI	詳細	目標
学習効果	受講前後で同一条件の設問に回答させ、偽・誤情報を適切に疑える方向に信頼評価が変化するかを測定する構造とした。	80%以上の参加者が21%以上のスコア向上
ユーザー満足度	楽しさ・理解実感・日常での有用感を問うことで、単なる知識獲得ではなく、継続利用・普及に耐える体験価値が成立しているかを確認する構造とした。	「満足」以上が「85%以上」を占める
行動変容持続	受講から一定期間経過後に追跡調査を行い、学習した確認ポイント（発信源・根拠・関連情報）に沿って偽・誤情報を見破れるかを測定する構造とした。	「50%」の参加者が2ヶ月経過後もプレバンキング行動を実践

※KPIの目標値は、先行研究や類似のゲーム型介入で報告される改善幅を踏まえつつ、授業時間内で実施可能であること、回答負担が過度にならないこと等、教育現場での運用条件と両立する水準として設定した。

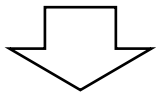
4-2. 検証及び調査の個別詳細

実証の流れ

評価体系を現場で成立させるため、検証は次の一連の流れで実施した。

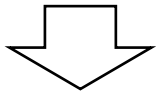
① 監修者との意見交換

教材が「体験として面白い」だけで終わらず、受講者の判断手順へ接続するよう、学習導線と説明の粒度を整えた。具体的には、体験→事例理解→対策理解という共通構造を各パートに定着させ、悪用リスクを前提とした倫理・安全上の配慮を運用要件として整理した。



② a版テストの実施

講前後の設問による学習効果測定と、体験後アンケートによる満足度・学習実感の把握を同時に行い、評価指標が実際に取得可能であること、および教材としての受容性と効果の方向性を確認した。



③ フィードバック

学習者の所感、実施者・監修者の改善提案、倫理・安全・対象者設定等の運用上の懸念を体系的に収集し、改善論点を「学習構造」「説明提示」「安全配慮」「体験導線」の軸で整理して改良に反映した。

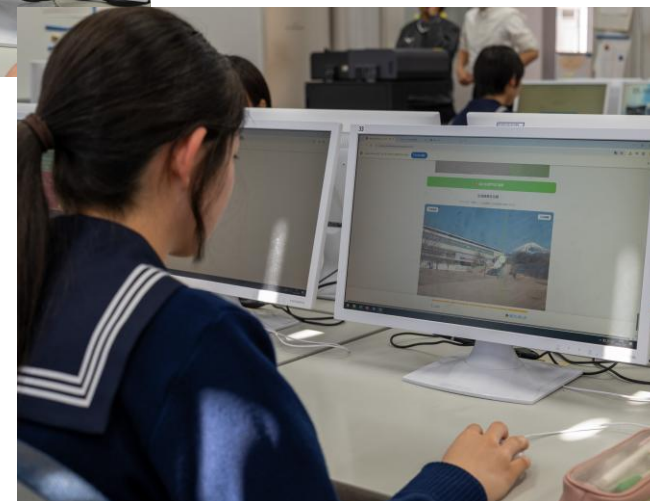
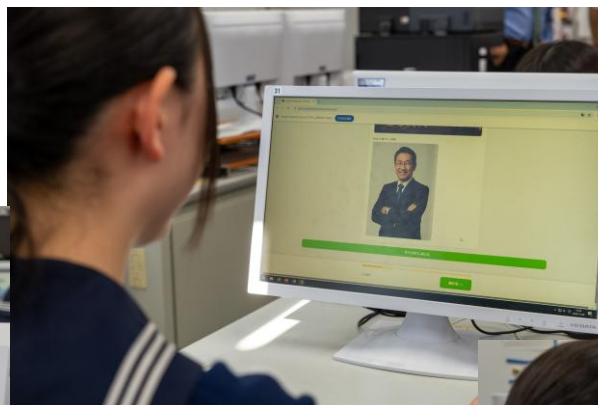
以上のとおり、本検証は、KPIの設定と測定、監修者による妥当性確認、a版テストによる効果と受容性の確認、フィードバック会による改善サイクルへの接続を一体の枠組みとして実施したものである。これにより、有効性の確認を単発の評価で終わらせず、教育現場での実装に耐える形で継続的に改良できる検証・調査として成立させた。

4-2. 検証及び調査の個別詳細

α版テストの概要

実施対象と実施方法

本α版テストは、協力機関である実践女子学園の中学3年生38名を対象に、教室に集合して一斉実施した。まず事前アンケートに回答してもらい、その後、偽・誤情報の典型パターンを題材にした学習プログラムを体験してもらったうえで、事後アンケートに回答してもらってプレ／ポスト形式で実施した。プログラムでは、5つの偽・誤情報パターンについて「実践（生成・改変の体験）→事例紹介→解説」の流れで学習した。



4-2. 検証及び調査の個別詳細

a版テストのアンケート設計

本a版テストのアンケート設計は、「ゲーム型プレバンキング介入の効果を、主観だけでなく定量で示す」ことを目的に、監修者からの助言と先行研究で用いられている測定枠組みを参照して決定した。具体的には、ゲーム介入による誤情報への耐性を扱った研究や、ゲーミフィケーションによる接種効果を扱う研究で一般的なプレテスト／ポストテストと追跡調査の構成を踏まえ、授業内で無理なく実施できる設問数と尺度に最適化した。

測定設計

KPI

測定手法

学習効果

受講前後で同一条件の設問に回答させ、提示された情報に対する信頼度（1～7）の変化として測定した。本物1件とフェイク／ミスリード3件（なりすまし、AI画像生成、ものまねサイトに対応）を提示し、フェイク情報は信頼度が下がるほど「適切に疑えている」と解釈し、本物情報は信頼度が高いほど望ましいとして扱った。

これらをもとに事前事後の変化量をスキルスコアとして算出し、参加者のうち80%以上が21%以上のスコア向上を達成することを目標に設定した。併せて、判断への自信度（1～7）も同時に取得し、誤った判断のまま自信だけが過剰に高まるような状態が生じていないかを点検する補助指標として用いた。

ユーザー満足度

体験直後アンケートで「楽しさ」「理解が深まった実感」「実生活で役立つ実感」を質問し、教材として継続利用・普及に耐える体験価値が成立しているかを測定した。

具体的には「プログラムは楽しめましたか」「フェイクニュース／偽・誤情報についての理解は深まりましたか」「実際のSNSやニュースを見るときに役立ちそうだと感じますか」の3項目で把握した。

行動変容持続

受講から2ヶ月経過後に追跡アンケートを実施し、学習した確認ポイント（発信源・根拠・関連情報）に沿った行動が実生活で実践されているかを測定する構造とした。

追跡アンケートでは、期間中に怪しい情報を見たか、見た場合にどの確認を行ったか、どのような場面で行動が起きたかを回答させ、1回でも確認行動を行ってれば「プレバンキング行動を実施」として扱う。

参考論文（一部）

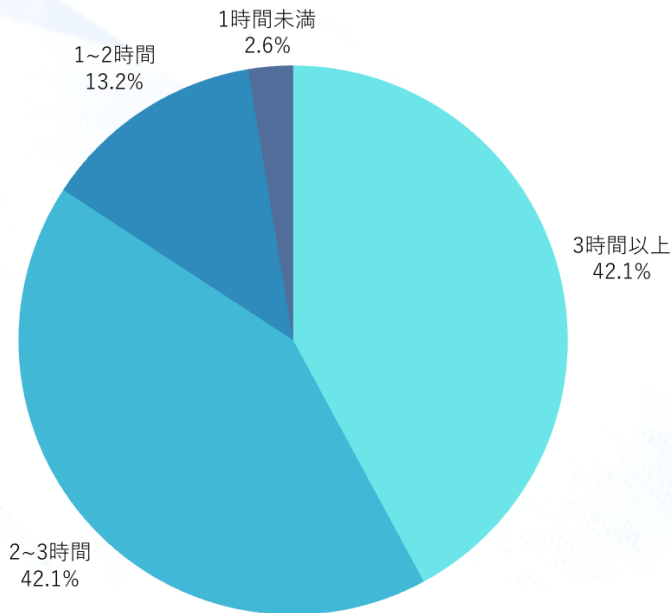
- ・「Fake news game confers psychological resistance against online misinformation (2019)」 : <https://www.nature.com/articles/s41599-019-0279-9>
- ・「Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News (2020)」 : <https://journalofcognition.org/articles/10.5334/joc.91>

4-2. 検証及び調査の個別詳細

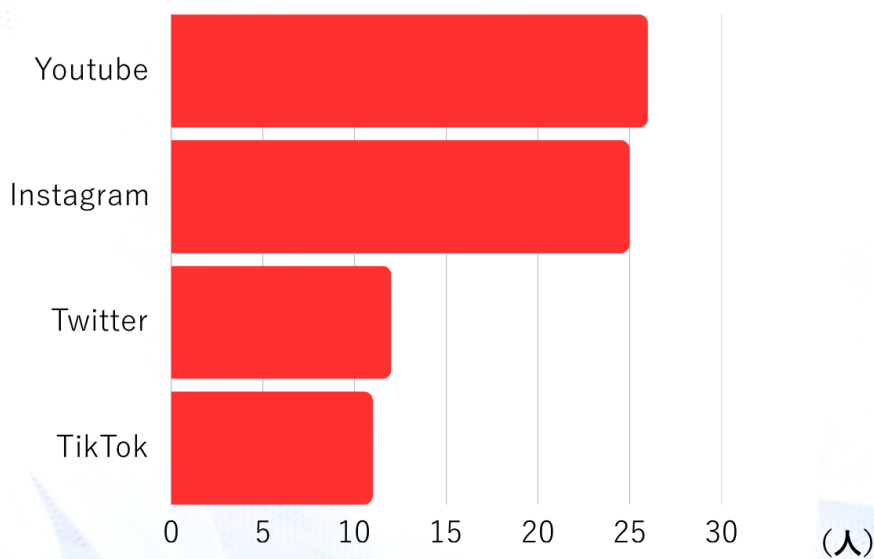
α版テスト参加者のSNS利用状況

生徒の約8割が「1～3時間／日」SNSを利用し、YouTube・Instagram が主要な情報源

SNS利用時間（1日あたり）



よく利用しているSNS（複数回答）



4-2. 検証及び調査の個別詳細

成果① 学習効果 – 55%の生徒が約37%のスコア向上

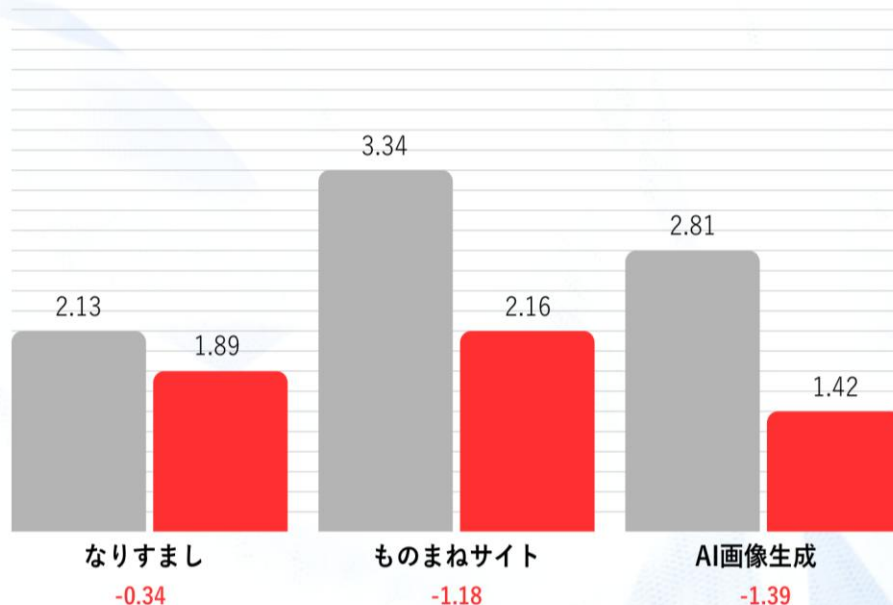
受講直後の識別能力の向上については、プログラム前後で提示した「本物1件+フェイク/ミスリード3件」に対する信頼度（1～7）の変化から、フェイク側を「本物だと思う傾向」が全体として減少したことを確認した。具体的には、なりすまし・ものまねサイト・AI画像生成のいずれも平均信頼度が低下しており、個人レベルでもフェイクに対する信頼スコアが向上した生徒が全体の約55%となった。また、向上した生徒に限ると、最大変化幅に対して平均37%分だけ改善していることが整理されている。

併せて自信度を測定し、判断の確かさも確認した。誤った方向への判断かつ自信度上昇（過信）は16.7%に留まり、補助指標として一定の妥当性が確認できた。

フェイクニュースを見抜く力が向上

信頼スコア 平均

● 事前 ● 体験後



なりすまし -1.85 (31%)

信頼スコアが向上した生徒人数：13名 (37.1%)

向上した生徒のスコア 変動：事前 3.31 → 事後 1.46

ものまねサイト -2.85 (48%)

信頼スコアが向上した生徒人数：20名 (57.1%)

向上した生徒のスコア 変動：事前 4.50 → 事後 1.65

AI画像生成 -1.92 (32%)

信頼スコアが向上した生徒人数：25名 (71.4%)

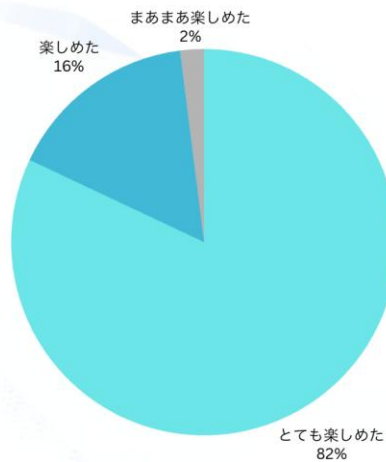
向上した生徒のスコア 変動：事前 3.28 → 事後 1.36

4-2. 検証及び調査の個別詳細

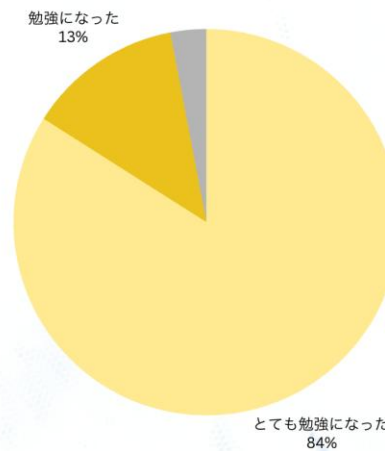
成果② ユーザー満足度-約98%

学習体験としての納得感・継続意欲の確保については、体験後アンケートで「楽しめた」「勉強になった」「今後も役立ちそう」がいずれも97%以上となり、サンドボックス型教材として体験価値が成立していることを確認した。

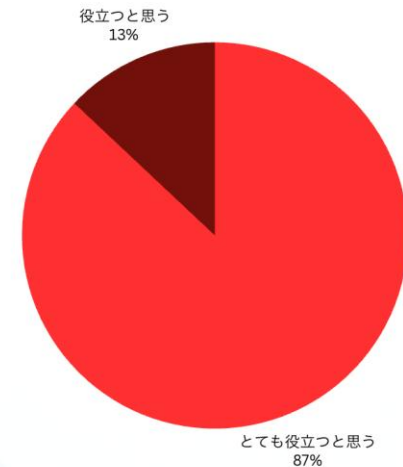
プログラムは楽しめましたか？



フェイクニュース／偽・誤情報についての理解は深まりましたか？



実際のSNSやニュースを見るときに役立ちそうと感じますか？



4-2. 検証及び調査の個別詳細

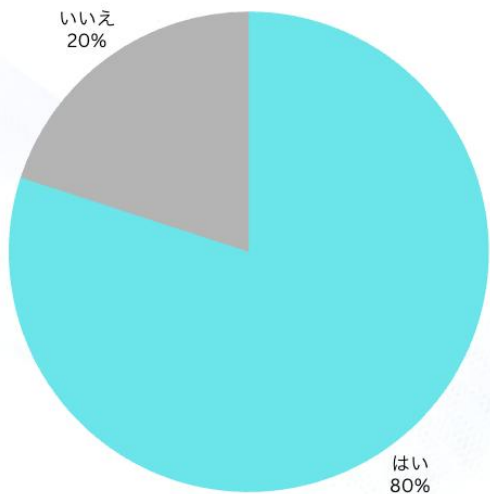
成果③ 行動変容持続

本追跡アンケートは、受講時に学習した判断基準が2ヶ月後も維持されているか、また実生活の情報環境の中で実際に活用されているかを測定することを目的として実施した。

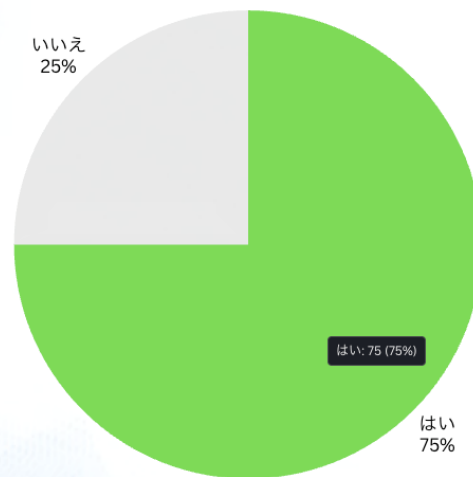
その結果、生徒は怪しい情報に触れた際に、「発信者やアカウント情報を確認する」「検索で他の報道や公式情報を探す」「画像検索で元画像を調べる」「すぐに共有せず一度立ち止まる」といった具体的な確認行動を一定割合で実践していることが確認された。

これにより、本プログラムで扱った確認ポイントが知識として理解されるだけでなく、日常において行動として定着している可能性が示唆された。

体験後に、SNSやニュース等で
怪しい情報を見たり受け取ったりしましたか



怪しい情報に対して、
学習した“確かめる行動”を一回でも行いましたか



4-2. 検証及び調査の個別詳細

成果④ 体験者の声

本実証では、信頼度スコアや満足度といった定量指標だけでは把握しきれない「どこが難しかったか」「何が腹落ちしたか」「どの場面で行動に移せそうか」といった学習体験の中身を補足するため、体験者の声（自由記述を含む定性的フィードバック）も併せて収集した。これは、数値の改善有無だけでなく、学習者がどの要素に納得し、どの説明や導線でつまづいたのかを具体的に把握し、教材のモジュール構成、説明の粒度、UIの分かりやすさ、安全配慮の伝え方などを次版開発で改善するための根拠として位置づけたものである。さらに、定量結果の解釈を補強し、改善点の優先順位づけ（どの課題が学習効果や満足度に影響しているか）に接続する目的も含め、継続的な改善サイクルを回すための入力データとして活用する設計とした。

印象に残っていること

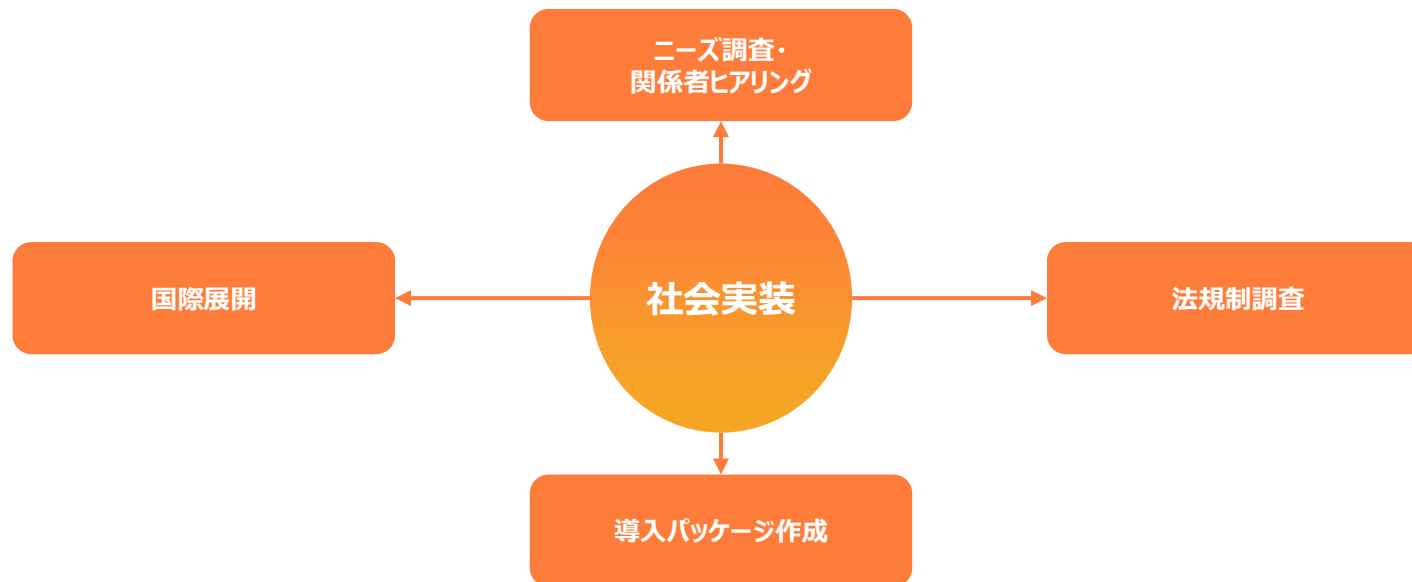
- ・光の当たり加減や影の感じを見てAIか判断できるというのは驚きました
- ・自分たちで実際にフェイク画像を作ったり、フェイク動画を作ることで、ほかの人がいかに簡単に人をだますことができるのかを知ることができたのが印象に残った。
- ・2つの画像を入れるだけで新しい画像が作れるのに驚きました。またAIは今まで正しいと思っていましたが間違っていた情報も教えられているのは怖いと思いました。
- ・SNSを使うときにあまり今まで気にしていなかったことを学ぶことができたので、今後気を付けながら使っていこうと思いました。
- ・自分の手でフェイクニュースを作ることでこういう風に作れるんだと驚きました。
- ・今まで単にスクロールしてただけのショートやインスタでも少しでも怪しいと思ったものは詳しく調べてみたいと思いました。またAIを使ったフェイクはだれでも簡単に作れると知って怖くなりました。
- ・よく顔をAIで変換してると出てるけどやり方も知れたことで前より、少しは見破れるようになったと思う
- ・リンクを見れば嘘か本当かわかるというのがすごく学びになった。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

5-1. 社会実装に向けた取組の全体像

社会実装に係る取組・成果の全体像



ニーズ調査・関係者ヒアリング

教育機関・法人・公的機関等へのヒアリングを通じ導入条件と障壁を抽出。

教育機関では、授業内完結・負担軽減が重視され、費用対効果で予算化の余地あり。

法人では短時間で業務接続できる設計への関心を確認。

法規制調査

導入の遅延要因となり得る個人情報・権利処理・不適切生成・周知表示等の論点を整理し、同意設計や運用ルールを含む対応方針を明確化。

導入パッケージ作成

開発したサンドボックス技術を、一般層でも受け入れやすい体験へ変換するため、ゲーム型の導入パッケージ（入口商品）を検討し、導入障壁を下げて本体導入へ接続する方針を整理。

国際展開

日本国内での社会実装に留まらず、東アジアや東南アジアを中心に世界へ展開可能な基盤づくりを進めるため、コンソーシアム設立を見据えたヒアリングとネットワーキングを検討・整理。

5-2. 社会実装に向けた取組の個別詳細

社会実装に向けた取り組みの目的と検討事項

本章は、社会実装を前提として設定した仮説が、実証および調査を通じてどの程度妥当であったかを整理し、導入障壁、提供形態、法務・運用条件、パッケージ化の方向性を明確化するためのパートである。

すなわち、単に「良い教材を作った」ことの記述ではなく、「市場に届け、継続導入される状態」を成立させるために何が必要かを、仮説→検証→示唆の順で整理する。

また、社会実装に向けて、学習効果指標とは別に、実際に導入が進むか、対価を伴う契約が成立するか、導入規模が実装・運用の前提として成立するかを測るためにビジネスKPIを設定した。

ビジネスKPI	詳細	目標
導入率	コンタクトを取った教育機関・企業・自治体のうち、サービス提供（無償含む）が実際に決定した割合。提案・協議記録と実施実績を照合し、「導入件数 ÷ 提案件数」により算出	2件
有償導入数	サービス提供が決定した案件のうち、対価を伴う契約に至った件数。契約書・請求書等により有償導入を確認	1件
行動変容持続	有償導入が成立した案件における、実際の受講人数の平均。参加記録、ログインデータ、授業・研修の出席者数から平均値を算出	10名

5-2. 社会実装に向けた取組の個別詳細

ニーズ調査・関係者ヒアリング - 教育機関向け

仮説検証

仮説A：探究学習での情報耐性不足

- 情報の扱い方を体系的に扱う教材が不足
- 教員の準備負担が継続運用の課題
- 「授業内完結」「短時間運用」が決め手

仮説B：実践型教材の不足

- 実践型教材が有効、授業時間内で運用可能な構造が必要
- 情報 I との接続・復習演習としての位置づけ
- 導入判断を後押しする要素として整理

得られた示唆

提供形態

- 出張授業：効果高いがコスト・日程制約あり
- オンライン：地理制約解消、操作支援が重要
- 教材提供：広域展開に適するがマニュアル整備が前提

価格・予算

- 学校の新規予算は限定的
- 反復・継続利用による費用対効果が説明できれば予算化可能
- 科目教材費等として一定範囲の予算化が見込まれる

導入条件

- 授業内完結 + 短時間運用が必須
- 教員側の準備負担の最小化
- 教材内容の妥当性に加え実施者が自立運用可能なこと

5-2. 社会実装に向けた取組の個別詳細

ニーズ調査・関係者ヒアリング

(1) 教育機関向け

(2) 法人向け

社会実装に向けたニーズ調査・関係者ヒアリングは、導入障壁の所在と、提供形態（対面・オンライン・教材提供）が成立する条件を明確化することを目的として実施した。特に、導入意思決定は学習効果だけでは決まらず、現場負担・予算化・説明可能性・安全性など複数要因の組合せで決まるため、仮説を置いたうえで論点を回収する設計とした。

実施方法に関しては、まずGoogleフォームを用いたアンケートとして実施し、回答企業のうち、追加のミーティングに応じた企業に対してヒアリングを行う二段階の設計とした。アンケート配布にあたっては、特にメディア企業を中心に、偽・誤情報対策に注力している実績等を参考にしつつ、関連性が見込まれる幅広い業界へ送付し、業種横断で導入条件や懸念点を回収できるようにした。

学校向けのヒアリングについては、既に提供している自社プログラムを通じて形成してきた教育機関との接点を活用し、現場の運用実態や予算・意思決定の観点を具体的に把握する形で実施した。

以下では、教育機関および法人を対象に実施したヒアリング調査について、事前に設定した仮説に照らして結果を整理し、そこから得られた示唆を取りまとめる。

仮説検証

仮説A：探究学習での情報耐性不足

探究学習の初期段階（第1学年等）では情報収集量が増える一方、偽・誤情報への耐性が十分でないため、探究の教育効果や成果物の信頼性が損なわれるリスクがある。

得られた示唆

探究における「情報の扱い方」を体系的に扱う教材の不足が指摘され、教員側の準備負担が大きいことが継続運用上の論点として現れた。

導入の決め手は、教材内容の妥当性に加え、「授業内で完結できること」「教員が短時間で運用できること」であるとの意見が見られた。

5-2. 社会実装に向けた取組の個別詳細

ニーズ調査・関係者ヒアリング

(1) 教育機関向け

(2) 法人向け

仮説検証

仮説B：実践型教材の不足

情報では情報モラルや情報の見極めが扱われるが、実社会の偽・誤情報や生成AIによる改変を"実践的に扱う教材"が不足しており、復習教材・演習教材への需要がある。

得られた示唆

知識中心の説明だけでは定着が弱いという認識が共有され、実践型の教材が有効である一方、授業時間制約の中で運用できる構造が必要であるとの意見が見られた。

あわせて、情報 I で扱う観点と接続し、授業の復習・演習として位置づけられることが、導入判断を後押しする要素として整理された。

仮説A・Bのヒアリングによって得られた他の示唆

提供形態

- 出張授業は効果が高いが、継続実施にはコストと日程制約がある。
- オンラインは地理制約を解消するが、操作支援や進行設計が重要となる。
- 教材提供（ライセンス）形式は、自治体・教育委員会等への広域展開に適する一方、実施者が自立運用できるマニュアル整備と事前研修が前提となる。

価格・予算

近年の物価高や家庭の経済状況の差を背景に、学校として大幅な新規予算を捻出することは難しいとの指摘があった。一方で、本教材が一回限りではなく反復利用・継続利用を前提とした構造であることから、学年を跨いだ活用や複数回実施による費用対効果が説明できる場合には、科目教材費等として一定範囲の予算化が可能ではないか、という示唆が得られた。

以上より、導入率を上げるには「導入形態の複線化」と「運用負担の最小化」に加え、学校の予算制約を踏まえた価格設計として、反復利用・継続利用による費用対効果が説明可能なパッケージ提示が重要論点であることが明確化された。

5-2. 社会実装に向けた取組の個別詳細

ニーズ調査・関係者ヒアリング – 法人向け

仮説検証

仮説A：AI誤用リスク

- 誤信・誤用を具体場面で体験する訓練が求められる
- 「自社業務に即した場面」「短時間実施」が条件
- 研修後に現場で使える手順への落とし込みを重視

仮説B：顧客教育の必要性

- 社内研修に限らず顧客向け啓発への関心
- 短時間で繰り返し使えるパッケージへのニーズ
- 判断観点を“手順”として定着させる設計が重要

仮説C：メディアの真偽判定力

- 内製困難な企業ほど外部教材ニーズが高い
- 「AI活用推進」ではなく検証プロセス強化として提示
- ブースター型トレーニングが受容されやすい

得られた示唆

提供形態

- ワークショップ：理解深いに対象人数に限界
- eラーニング：裾野拡大可能、理解の質担保が課題
- 教材提供：組織内展開容易、ガイドライン整備が前提

導入規模・体制

- 大規模企業：内製可能、最新技術キャッチアップも自社対応
- 中小規模企業：内製困難、外部教材ニーズが高い
- 規模・体制に合わせた運用負担の最小化が重要

導入条件

- 一般論ではなく自社業務に即した場面設計
- 短時間実施＋研修後に現場で使える手順
- 受け手のリテラシー差を踏まえた設計

5-2. 社会実装に向けた取組の個別詳細

ニーズ調査・関係者ヒアリング

(1) 教育機関向け

(2) 法人向け

仮説検証

仮説A：偽・誤情報リテラシーの企業研修ニーズ

偽・誤情報や生成AI由来の不正確情報を誤って信じることは、誤発信・コンプライアンス事故・意思決定ミスに直結し、企業価値やブランドを毀損し得る。加えて、全社的なAI活用が進む局面では「活用推進」と「誤用不安」が同時に存在し、従業員が「理解しないまま使う／不安で使えない」ことが生産性・品質双方のリスクとなる。

得られた示唆

生成AIの利活用を進めたい企業ほど、単なる注意喚起ではなく、業務の中で起こり得る誤信・誤用を具体場面で体験し、短時間で確認観点（発信源・根拠・関連情報）に立ち返れる訓練が求められる傾向が確認された。
導入条件としては「一般論ではなく自社業務に即した場面」「短時間で実施できること」「研修後に現場で使える手順に落ちること」が重視される。

仮説検証

仮説B：顧客教育を含めた対策の必要性

個人情報等を材料にした高度な詐欺・なりすましが増え、従来型の注意喚起だけでは不十分となっている。顧客が被害に遭うことは企業側の金銭的損失や信頼の毀損にもつながるため、社内研修だけでなく顧客教育まで含めた対策が必要である。

得られた示唆

研修対象を従業員に限定せず、顧客向けの啓発・教育に展開できるパッケージ（短時間で繰り返し使える教材、運用手順、説明用素材）への関心が見られた。
特に、受け手のリテラシー差が大きい前提で、判断観点を「手順」として定着させる設計が重要である。

5-2. 社会実装に向けた取組の個別詳細

ニーズ調査・関係者ヒアリング

(1) 教育機関向け

(2) 法人向け

仮説検証

仮説C：メディア現場での即時判定力ニーズ

日常的に巧妙な偽情報が流通する中、記者・編集部には短時間での真偽判定力が求められる。見抜けなかった場合、誤報・誹謗中傷の誘発や企業イメージの毀損に直結するため、忙しい現場でも数分で実施できる"見抜く力"のトレーニング教材へのニーズが見込まれる。

得られた示唆

大規模なメディア企業では、社内に研修機能・教材作成の体制を持ち、最新技術へのキャッチアップも含めて研修を内製できるケースがある一方、研修専門を置けない等の理由で、内製が難しい規模の企業ほど外部教材へのニーズが相対的に高い可能性が示唆された。

また、生成AIの業務利用は慎重に扱われる傾向があるため、メディア向けには「AI活用推進」ではなく、検証プロセスの強化と、短時間で反復できるブースター型トレーニングとして提示することが受容されやすい。

仮説A・B・Cのヒアリングによって得られた他の示唆

提供形態

- ・ ワークショップ形式は理解が深い一方、対象人数が増えるほど運用負担が課題となる。
- ・ eラーニング形式は裾野を広げられるが、理解の質を担保する仕組みが必要である。
- ・ 教材提供形式は、組織内で継続展開しやすい一方、導入時の研修・ガイドライン整備や、運用者が自走できる手順の提供が前提となる。

以上より、導入率を高めるには、提供形態の複線化と、対象組織の規模・体制（内製可否）に合わせた運用負担の最小化が重要論点である。

5-2. 社会実装に向けた取組の個別詳細

法規制調査

社会実装では、導入率・有償導入の障壁として、法務面の懸念が意思決定を遅らせる要因になり得る。このため、想定される論点を事前に洗い出し、対応方針を明確化することを目的に整理した。

懸念1：個人情報・プライバシー

メールアドレス等のアカウント情報や学習ログ等を取り扱う。パターンによっては受講者が持つ画像も取り扱う。

対応方針

- 取得データを「提供・運用・不正利用防止・学習効果可視化」に必要な最小限へ限定する。
- 画像は教材体験に必要な場合でも、システム内に保存しない設計を原則とし、保存が必要なケースは事前合意（保存期間・権限・匿名化/集計）を契約で取り決める。
- 未成年が対象となる場合、学校・保護者を含む同意設計（周知文、同意取得、代替教材の提示）をパッケージに含める。
- 運用面では、アクセス権限管理・ログ管理・保存期間・削除手順を「導入パッケージ」に標準同梱する。

懸念2：不適切利用・悪用

ユーザーが自分の画像をアップロードする形式は法的リスクが相対的に高く、未成年や悪意あるユーザーによる誤用リスクも想定される。

対応方針

- 利用規約・禁止事項（第三者の画像投入、誹謗中傷目的利用等）を明確化し、違反時の停止・通報フローを整備する。
- 出力物の再拡散・転用を抑制するため、画面上の透かし、ダウンロード制限、回数制限等を導入し、教育目的の範囲に閉じる。
- 「実在個人・著名人」を題材にしくても学べるよう、サンプル素材・架空人物素材を標準提供する。

懸念3：著作権・肖像権・名誉毀損等

体験型教材は、生成物が外部に持ち出された場合、第三者権利侵害や誹謗中傷誘発に接続し得る。

対応方針

- 教材内生成物は「学習目的の範囲でのみ利用」する設計・規約を明確化し、外部公開を前提にしない。
- 事例学習で扱う素材は権利クリアなものに限定し、引用・出典表示ルールをテンプレ化する。
- 重大リスク領域（第三者の顔、特定個人の名指し等）はプロンプト制御・UI上の注意喚起で抑止する。

5-2. 社会実装に向けた取組の個別詳細

導入パッケージ作成

本実証で開発した偽・誤情報サンドボックス技術は、画像生成AIおよびテキスト生成AIを統合し、悪用防止・統制アーキテクチャを備えた安全設計型の生成基盤である。今回開発した導入パッケージは、このサンドボックス技術を内包し、その生成機能および統制機構を活用した社会実装型コンテンツである。

すなわち、基盤技術を別用途に転用したものではなく、安全設計された生成環境を体験しやすい形に再構成した拡張モデルである。これにより、教育・研修現場のみならず、一般層を含む幅広い年齢層に対し、安全性を担保した形で偽・誤情報のリスクと向き合う体験機会を提供することを目指している。

ゲーム設計

導入パッケージでは、偽・誤情報の中でも特に直感的に理解されやすく、体験効果が高い「見た目の説得力」を題材に、“楽しいが怖い”体験を通じて判断の要点を残すゲームを設計対象とした。具体的には、参加者が写真を素材として扱い、システムが一部を改変したうえで、プレイヤー同士の対話と推理で真偽を見抜くパーティーゲーム型のアプリケーション（ゲーム仮名：ダウトAI）を導入パッケージの一類型として位置づけた。

想定する市場とターゲット

一般向け

イベント・市民講座・啓発

自治体・図書館・生涯学習等で実施される短時間イベントに参加する住民層。講義よりも体験型が適しやすく、家族・友人同士での参加が見込まれる。

学校向け

探究の入口・情報科の導入

探究学習や情報科の導入段階で「なぜ確認が必要か」を直感的に理解させたい層。短時間で盛り上がりを作りつつ、確認行動へ接続できる入口教材としての需要がある。

法人向け

全社員向け基礎研修・啓発

生成AIや偽・誤情報を「知っているが実態が腹落ちしていない」層に対し、体験を通じてリスク理解と判断基準を共有したい需要がある。誤認・誤発信が企業イメージに直結する組織ほど必要性が高い。

若年～現役世代

エンターテインメント・ゲーム市場

SNS利用頻度が高く、技術への好奇心と警戒が共存する層。体験の共有・拡散が起点になりやすく、普及導線を作りやすい。

5-2. 社会実装に向けた取組の個別詳細

導入パッケージ作成

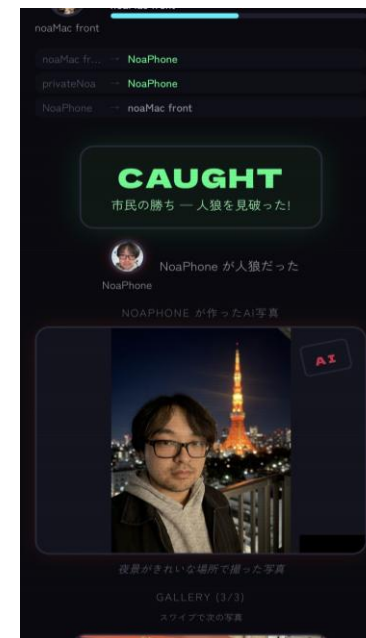
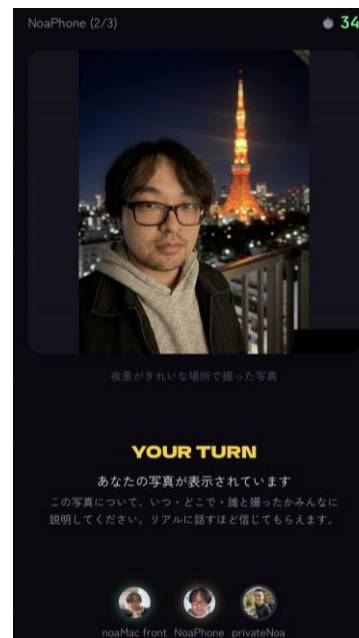
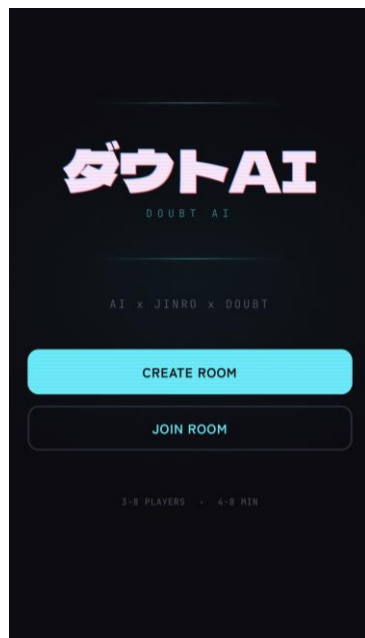
ゲーム概要

参加者は端末内の写真（本人の同意があるもの）を用いてゲームに参加し、システムが一部を改変した写真を混在させる。プレイヤーは順番に写真を提示し、その写真についての説明を行う。他プレイヤーは、違和感があれば「ダウト」を宣言し、改変の有無を確認する。この過程で、参加者は「見た目の自然さ」「説明の説得力」が真偽を担保しないこと、ならびに“疑うべき観点”が存在することを体験的に理解する。

ゲームの狙い

- ディープフェイクの危険性を「知識」ではなく「体験」として自分ごと化させる
- もっともらしいストーリー（嘘の物語）や雰囲気“真実の保証にならない”ことを腹落ちさせる
- 導入現場での受容性（盛り上がり・参加しやすさ）を確保し、学習パッケージ本体への接続を容易にする

※α版ゲーム画面



5-2. 社会実装に向けた取組の個別詳細

国際展開に向けて

偽・誤情報が社会にもたらす影響は、国内にとどまらず国境を越えて拡大している。そこで本実証では、東アジア、東南アジアを中心に世界展開可能な基盤づくりを進める。具体的には、ファクトチェック機関および関連団体と連携し、コンソーシアム設立を見据えたヒアリングとネットワーキングを実施した。

■ ヒアリング先

台湾ファクトチェック

CekFakta

MAFINDO

Aliansi Jurnalis
Independen

■ ヒアリングから見てきた4つの課題

① リソース不足

検証活動は実施できているが、教育活動に割く人的・資金的リソースが不足

② 言語・文化の差

翻訳のみでは不十分であり、国ごとの社会背景やセンシティブ領域に配慮した事例設計・表現調整が必要

③ AI法制度の差異

個人情報やAI利用の法制度が国ごとに異なり、安全設計とガイドライン整備が必要

④ 教育制度の違い

各国の教育課程や導入方針が異なり、統一的な展開は困難。現地との協働が不可欠

■ 今後の展開

STEP 1

共通基盤の整備
多言語対応・法制度調査



STEP 2

プレバンキング
コンソーシアム設立



STEP 3

各国パートナーとの
協働展開

※インドネシアや台湾をはじめとするファクトチェック団体へのヒアリングを通じ、国際展開に向けた課題と方向性を整理。今後、共通基盤を構築し、東アジアや東南アジアを含むファクトチェック機関とのコンソーシアムの設立を目指す。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

6-1. 普及啓発活動の全体像

普及啓発活動に係る取組・成果の全体像

本開発・実証期間において、以下の3つの普及啓発活動を実施し、認知拡大から導入検討への導線を構築した。

① プレスリリース配信 — 認知拡大と信頼獲得による導入相談の母数拡大

- 開発進捗に合わせた三段階配信（採択公表 → α版成果 → 正式リリース）で段階的に認知を拡大
- 教育系メディア（ICT教育ニュース、こどもとIT等）に転載され、教育ICT領域での認知拡大に寄与

② 自社主催イベントによる普及 — 体験を通じた「自分ごと化」と商談化

- 3/7（土）オンライン勉強会「フェイクニュースを"見破る×作ってみる"」を開催（Zoom、2時間）
- 教育関係者・企業・団体から計20名が参加し、「レイのブログ」および偽誤情報サンドボックスを体験
一部参加者から導入相談の打ち合わせ予約を獲得 → 普及導線としての成果を確認

③ 「レイのブログ」復習教材としての販促 — 既存導入先への継続利用・アップセル

- 既存導入校・問い合わせ校 約85校に対し、サンドボックス教材を復習・発展教材として導入促進
- 現在20校が探究学習・情報科目・長期休暇の教材等としての導入を前向きに検討中（来年度以降）

6-2. 普及啓発活動の個別詳細

プレスリリースの配信

本事業の取り組み内容（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）について、開発の狙い、社会的背景、実証の位置づけ、今後の社会実装方針を整理したプレスリリースを配信する。

狙いは、教育機関・自治体・法人の意思決定者に対して「何が新規性で、何が導入価値か」を短時間で理解できる情報パッケージを提供し、導入相談の母数を増やすことである。加えて、報道・業界媒体での露出を通じ、社会課題としての認知を底上げし、導入検討の心理的障壁（未知の取り組みへの不安）を下げる。

また、社会実装フェーズへの接続を強めるため、プレスリリースは開発進捗に合わせて三段階で配信する。第一段階は採択直後に、現在どのような実証プロジェクトに取り組むのかを公表し、協力機関の獲得やヒアリングの拡大につながる情報媒体として認知を形成する。第二段階はα版テスト完了時に、開発中の教材・技術が示した効果を定量データとして提示し、β版テストへの参加導線を作るとともに、プロダクトへの関心と期待を喚起する。第三段階は開発が完了したタイミングで、サンドボックス技術を用いたプロダクトの正式リリースとして発信し、具体的な導入検討層（教育機関・自治体・法人）に対する導入相談・契約獲得へ接続する。

総務省「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」に採択

「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」



総務省「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」における中間成果

生成AIを用いた「偽・誤情報サンドボックス」による能動的プレバンキングの実証結果—偽・情報への信頼スコアが全項目で低下、AIによる生成画像では7割超の生徒が見抜けるように—



メディア転載実績

なお、配信したプレスリリースは複数の教育系メディアにおいて転載・紹介され、教育ICT領域における認知拡大に寄与した。

主な掲載媒体は以下の通りである。

- ① ICT教育ニュース（2026年1月6日掲載）：中間成果発表に関する記事を転載 <https://ict-enews.net/2026/01/06classroom-adventure/>
- ② こどもとIT（2025年7月掲載）：総務省事業採択に関する記事を転載 <https://edu.watch.impress.co.jp/docs/news/2035078.html>

6-2. 普及啓発活動の個別詳細

自社主催イベントによる普及

3/7（土）にオンラインにて自社主催イベント「フェイクニュースを"見破る×作ってみる"オンライン勉強会」を開催。教育関係者・企業・団体から計20名が参加し、「レイのブログ」および「本実証事業で開発した偽誤情報サンドボックス」を用いた体験会を実施した。イベント後、一部の参加者から導入相談の打ち合わせ予約があり、普及導線としても成果を得た。

イベント概要

- ・ タイトル：フェイクニュースを「見破る×作ってみる」オンライン勉強会
- ・ 日時：2025/3/7（土）19:00～21:00 形式：Zoom（オンライン）
- ・ 対象：教育関係者／企業研修・広報担当／一般参加者
- ・ 参加者数：20名（教育関係者・企業・団体）

プログラム内容

- ・ 前半（約50分）：「レイのブログ」デモ体験 — 受け手の視点でフェイクニュースを見破る
- ・ 後半（約30分）：偽誤情報サンドボックス体験 — 作り手の視点で構造を理解する
- ・ 質疑応答（約30分）

成果

一部参加者が導入相談のための打ち合わせを予約

6-2. 普及啓発活動の個別詳細

自社教材「レイのブログ」の復習教材としての販促

本技術の役割

「レイのブログ」で獲得した基礎行動（疑う・調べる・判断する）を、生成AI時代の偽・誤情報環境（画像改変、もっともらしい文章生成等）に対応させる発展教材として提示し、既存導入先における継続利用・アップセルを狙うものである。新規導入と比較して意思決定・運用負担が相対的に小さく、導入率の改善に資する導線として位置づけた。



販促実績

約85校（「レイのブログ」導入校および問い合わせ校）に対し、復習教材（サンドボックス教材）としての導入促進を実施した。その結果、現在20校が、探究学習、情報科目、長期休暇の学習教材等としての導入を前向きに検討している。



- ※ 多くの学校では来年度予算での実施を想定しており、予算確定および年間指導計画の確定を待つ必要があるため、現時点では検討段階にある。
- ※ 新規導入と比較して意思決定・運用負担が相対的に小さく、導入率の改善に資する導線として位置づけた。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

7-1. 技術開発及び社会実装における課題・展望

技術開発にあたっての今後の課題

技術開発の課題

技術開発の展望

課題1：教材内生成物の外部流通・二次利用リスクへの対応

本技術は、受講者が偽・誤情報を生成する体験を通じて耐性を獲得する設計であるが、生成物が教材外へ流出した場合には、第三者の誤認・拡散、あるいは悪意ある二次利用につながる懸念がある。このため「教育効果を損なわずに、生成物の外部流通リスクを抑える」という両立が技術要件として重く、今年度の開発・実証を通じて、単一対策では十分でないことが明確になった。

今年度までに実装・運用している対策としては、

- (a) 生成画像に対してサーバーサイドで可視な電子透かしを付与し“当該ツール由来”を示す
 - (b) 生成AI側のセーフ機能に加えて独自のワードフィルタリング等を組み合わせ、センシティブな内容の生成を抑止する
 - (c) ブラウザ上で保存しづらくする抑止策（右クリック防止等）を入れる
 - (d) 生成物に対する解析AIでチェックし、不適切性を検知した場合に提示を制限する
- といった多層防御を採用している。

これらは、教材体験を成立させつつ悪用可能性を下げるための、現時点での実務的なバランスとして位置づけている。

しかし、今年度の議論と外部動向の調査から、「可視な電子透かしは消され得る」という現実が、対策設計上の構造課題として浮き彫りになった。生成AIや専用ツールによる“電子透かし除去”が継続的に高度化しており、可視透かしだけに依存した流出対策は時間とともに脆弱化する。加えて、研究としても電子透かし除去攻撃が提案され続けており、対策は「いたちごっこ」になり得ることが指摘されている。

課題2：生成AI安全機能の限界を踏まえた教材特有の安全に使うための制御設計の必要性

生成AIの安全機能は、不適切生成を抑止する上で重要な前提である。

一方で、実運用では、文脈依存で危険になる指示、迂回表現（スラング／伏せ字など）・多言語（英語/日本語/ローマ字を混ぜるなど）・文脈依存（「研究目的として」など、“設定”が付くと生成できるなど）を用いた抜け道等により、完全なブロックは困難である。

また、教育用途としては「危険な生成を完全に消す」ことが必ずしも最適ではなく、学習者が“なぜもっともらしく見えるのか”“どの表現が誘導になるのか”を理解するために、一定のリアリティや曖昧さを残した体験設計が必要となる。したがって本技術では、単に禁止を強化するのではなく、教育上必要な範囲の生成体験を成立させつつ、危険領域へ逸脱しないように制御する「教材特有の安全に使うための制御設計」が要件となる。

7-1. 技術開発及び社会実装における課題・展望

技術開発にあたっての今後の課題を踏まえた展望

技術開発の課題

技術開発の展望

①透かし・来歴情報・運用統制を組み合わせた多層的な悪用防止設計の強化

対応事項（改善案）は、可視透かし中心から「機械による読み取りが可読な来歴情報 + 不可視透かし + 運用統制」の組み合わせへ段階的に移行し、強度を上げることである。

具体的には、Google DeepMindのSynthIDのような不可視透かし／検出技術の活用可能性を検討し（テキスト・画像・音声・動画など複数モーダルを対象とする潮流を踏まえる）、生成物が加工されても検出可能な設計を取り入れる。

さらに、C2PA（Content Credentials）のようなコンテンツへの来歴情報の標準に沿って、生成・編集の事実を暗号学的に結びつけたメタデータとして付与し、外部流通時に「後から検証できる」状態を作る。

ただし、C2PA等のメタデータは、SNS等のプラットフォーム側で削除・非表示化される場合があり、技術を入れただけで社会全体に表示されるとは限らない。

したがって展望としては、

- (1)透かし・来歴を付ける
 - (2) それを読み取れる管理者・導入先の運用（研修内の確認、ログとの突合、持ち出し制限）を設計する
 - (3) 必要に応じて外部の検出・真正性確認の仕組みと併用する
- という三層（技術・運用・連携）でリスクを下げる方針が妥当である。

7-1. 技術開発及び社会実装における課題・展望

技術開発にあたっての今後の課題を踏まえた展望

技術開発の課題

技術開発の展望

② 自由度の段階設計と生成前後判定・監査機能による安全に使うための制御の高度化

今年度の開発・運用検討から、安全な制御設計の中核は以下の三点であることが明確になった。

① 自由度の設計

- 完全な自由生成ではなく、教材側でシナリオ・パターンを制約し、学習段階に応じて解放範囲を段階的に拡大
- 初期：条件選択による生成 → 中期：限定的な編集 → 上級：より自由な表現操作

② 判定の多重化

- 生成前：プロンプト解析により禁止カテゴリ・高リスク意図を事前制限
- 生成後：出力評価により危険性のある内容を制限・修正（両者の併用で迂回表現にも対応）
- 禁止語彙に依存せず、文脈・意図・対象を評価軸に含め多言語対応にも耐性を確保

③ 管理者統制（例外・監査・説明）

- 教員・管理者が機能範囲を説明でき、入力・生成履歴を追跡可能な状態を標準化
- ログ確認機能、権限制御、逸脱兆候の検知・アラート等の統制機能を整備

→ 上記三点により、学習者の体験価値を維持しつつ、社会実装に耐える安全性・統制・説明可能性を一体的に担保する。

7-1. 技術開発及び社会実装における課題・展望

社会実装にあたっての今後の課題

社会実装の課題

社会実装の展望

課題1：偽・誤情報が動画・音声・複合へ移行しており、教材のマルチモーダル対応が必須になっている

今年度の開発・実証は、テキストと画像を中心に、典型パターン学習と仮想環境でのAI生成体験を組み合わせ、立ち止まる・出所を確かめる・拡散を抑えるという判断行動の獲得を狙った。

一方、実環境では短尺動画、音声、ライブ配信、切り抜き、ナレーション付き画像などへ接触形式が急速に移行している。さらに偽・誤情報は単一モーダルよりも「動画+テロップ」「画像+音声」「テキスト+画像+リンク」など複合で信ぴょう性を補強する傾向が強い。このため、テキスト・画像中心のままでは実環境との距離が開き、学習観点の転移が難しくなる懸念がある。今後は複合モーダル前提の学習設計へ拡張する必要がある。

課題2：プライバシーへの配慮

本技術は学習効果の可視化や運用管理のため、アカウント情報・利用状況・学習ログ・生成時の入力等を一定範囲で取り扱う設計である。一方、導入先（学校・企業・自治体）では、個人情報保護や内部規程により「学習ログが個人に紐づくこと」自体が導入障壁になり得る。

特に教育現場では未成年を含むため、説明責任・同意取得・保存期間・第三者提供の有無が厳密に問われ、企業では機密情報混入（受講者が業務情報をプロンプトに入力する等）もリスクになる。

課題3：国際展開において、各国の偽・誤情報特性に合わせて共同で作り続ける体制づくりが必要

国際化はUIや文章の多言語化だけでは不十分で、各国で異なる「典型的な誤情報テーマ」「信頼される情報源」「制度・プラットフォーム利用実態」「扱いが敏感な話題の境界」に合わせて、教材シナリオと評価指標を再設計する必要がある。あわせて、心理面の配慮も含め、現地知見を前提に共同編集できる体制（共同制作・共同評価・共同運用）をどう構築するかが課題となる。

課題4：継続利用（反復）を成立させる運用設計

プレバンキングは反復により効果が維持されやすい一方、学校・企業・自治体の現場では継続実施の工数が障壁になりやすい。社会実装では、教材品質だけでなく、短時間で完結するモジュール化、担当者が迷わない運用導線、実施後に効果が見えるレポート、定期的な更新による新鮮さが揃って初めて予算と工数が確保される。さらにマルチモーダル化は学習体験を強化する一方で、端末・回線・音出し・フィルタリングなど運用変数が増え、「運用の簡易性」とのトレードオフが生じる。

7-1. 技術開発及び社会実装における課題・展望

社会実装にあたっての今後の課題を踏まえた今後の展望

社会実装の課題

社会実装の展望

① マルチモーダル対応(動画・音声への段階的拡張)

今後は、現行のテキスト・画像に加え、音声や動画も対象に含め、実環境に近い学習条件へと拡張していく。その背景には、ニュースや情報接触の主戦場がソーシャルメディアや動画へ移行していることに加え、犯罪や詐欺においても音声・動画が悪用される場面が広がっていることがある。

このため、学習対象を実際の情報環境に近づけるうえで、音声・動画への対応は不可欠である。

ただし、音声生成や動画生成の機能を一度に広く提供することは、悪用リスクや運用負荷の観点から適切ではない。

そのため、社会実装にあたっては、段階的な拡張を基本方針とする。具体的には、初期段階では既存の音声・動画コンテンツを用いた検証や見抜きの学習モジュールを追加し、次段階で条件を限定した生成機能へと広げていく。

また、テキスト、画像、音声、動画を組み合わせた複合モーダルの教材についても、学習者が何を根拠に信頼性を判断すべきかを整理し、検証行動につなげる設計を進める。

② 国際化において既存海外パートナーと「共通基盤＋現地ラッパー」の分業で提供をスケール

国際展開においては、日本で作成した教材をそのまま翻訳して配布するのではなく、生成基盤や安全制御、学習ログ、管理者機能などの共通基盤を活用しつつ、各国パートナーが現地の事例、参照すべき信頼ソース、文化、制度、評価指標に応じて教材を最適化する形で提供する。

これにより、品質と安全性を一定水準で担保しながら、各地域の実情に即した導入を進めることができる。

そのためには、パートナー連携に関する契約設計や共同編集のワークフロー、各国での効果測定の方法を整備する必要がある。

加えて、ディープフェイクや偽情報への接触が心理面にも影響し得ることを踏まえ、刺激の強度やセンシティブなテーマの扱いも含めた学習設計ガイドラインを共通化し、国際展開に耐える運用基盤として整備していく。

参考

- Reuters Institute, Digital News Report 2025 (ソーシャル・動画経由のニュース接触増、若年層ほど顕著)
- Europol, EU-SOCTA 2025 (AI音声クローン、ライブ動画ディープフェイク等が詐欺・恐喝・なりすましを増幅)
- European Parliamentary Research Service (ディープフェイク詐欺・音声/動画被害の整理)

7-1. 技術開発及び社会実装における課題・展望

社会実装にあたっての今後の課題を踏まえた今後の展望

社会実装の課題

社会実装の展望

③ モーダルの柔軟性：to C向け展開と、ターゲット別ラッパー、API化

社会実装の拡大にあたっては、「教育機関・法人・自治体」向けの提供に加え、一般生活者が日常的に触れられるto C形態での展開も重要となる。

特に、若年層を中心に情報接触がソーシャルメディアや動画へ移行している中では、学習機会を学校や研修の場に閉じず、生活導線上で提供できる形が有効である。

そのため、現在の統合型モジュールに加え、短尺・単機能のモジュールを用途別に提供し、対象やリスクに応じて組み替えられる構造へと発展させる。また、対象別のラッパー設計を用意し、同じ根幹技術の上でUI、ストーリー、難易度、評価観点を調整する。

さらに中長期的には、中核機能をAPIやSDKとして外部提供し、e-learningシステムやLMS等への組み込みを可能にすることで、導入先の既存基盤の上でも活用できる形を目指す。

④ “安全だが学べる”ことを導入先が説明できる状態の整備

導入先の意思決定プロセス（学校の校内稟議、自治体の調達、企業のセキュリティ審査）において、「なぜ生成機能が必要なのか」「悪用対策は何をどこまでしているのか」「事故時の責任の所在はどうなるのか」を説明可能な資料・運用設計が求められる。

特に欧州では、合成コンテンツの表示・ラベリングの透明性義務を規制枠組みとして明確化する動きが進んでおり、将来的に国内外の制度・ガイドラインへの整合も競争力に直結する。

次年度以降は、導入先が安心して導入できるよう、

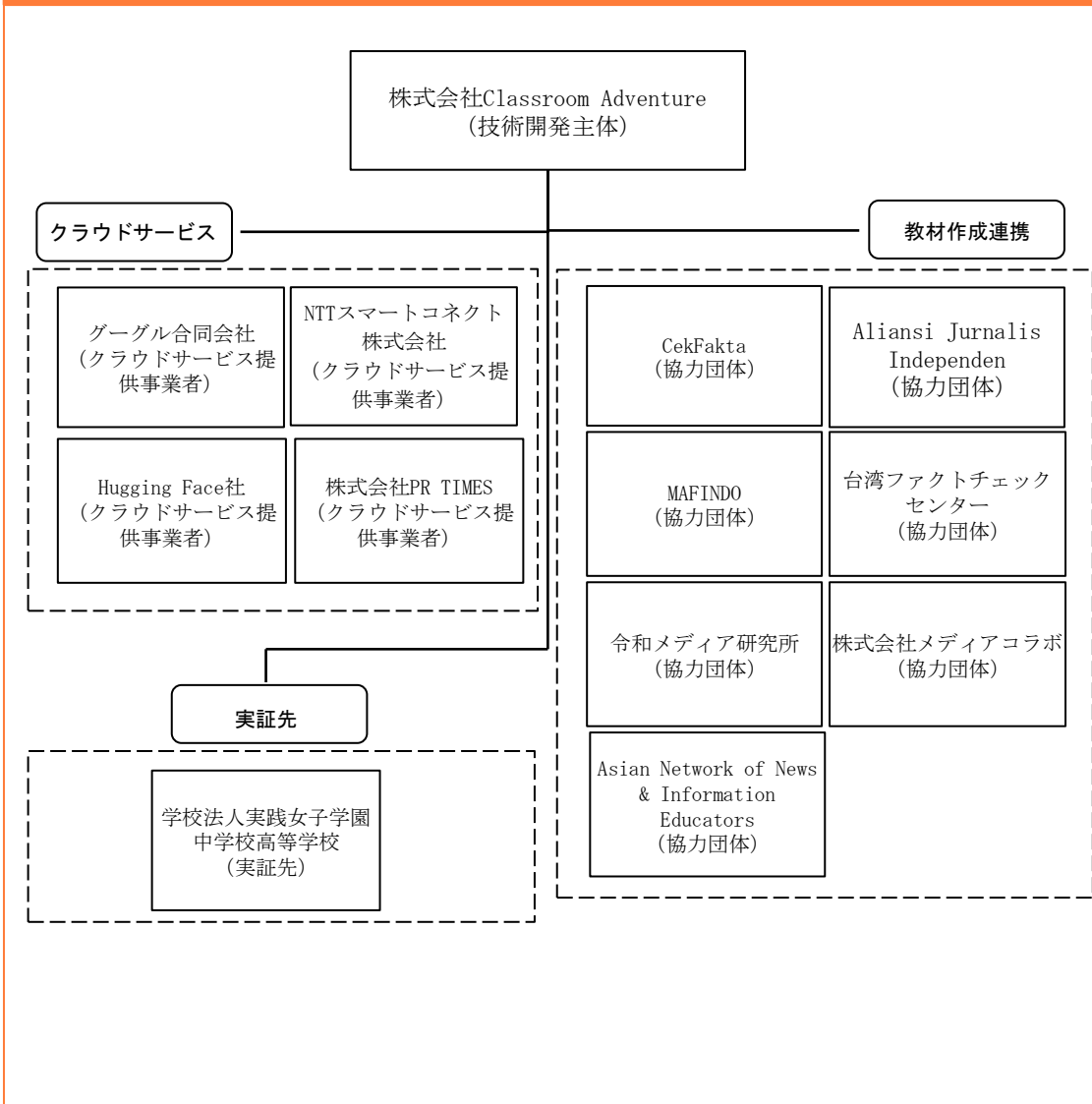
- (1) セキュリティ/プライバシー設計（データ範囲、権限、ログ、削除）を明文化した導入パッケージ
- (2) 教材内での安全教育（持ち出し禁止・二次利用禁止・倫理）を学習プロセスに組み込んだ設計
- (3) 事故・問い合わせ時の運用フロー（通報窓口、停止、ログ確認、再発防止）を標準化
- (4) の3点を整備し、社会実装に耐える設計の構築が必要である。

目次

1. 開発・実証のサマリ
 1. 開発・実証のサマリ
2. 開発・実証の背景・目的
 1. 開発技術によりアプローチする課題
 2. 開発技術により目指す姿・ゴール
 3. 開発技術により対処可能なユースケース
3. 開発・実証における「対策技術の開発」
 1. 技術開発の全体像
 2. 技術開発の個別詳細
4. 開発・実証における「対策技術の有効性等に関する検証及び調査」
 1. 検証及び調査の全体像
 2. 検証及び調査の個別詳細
5. 開発・実証における「対策技術の社会実装に向けた取組」
 1. 社会実装に向けた取組の全体像
 2. 社会実装に向けた取組の個別詳細
6. 開発・実証における「普及啓発活動への協力」
 1. 普及啓発活動の全体像
 2. 普及啓発活動の個別詳細
7. 開発・実証の課題・展望
 1. 技術開発及び社会実装における課題・展望
8. 開発・実証の実施体制等
 1. 実施体制及び役割分担
 2. 全体スケジュール

8-1. 実施体制及び役割分担

本事業の実施体制図



各団体の役割・業務範囲

【実証先】

学校法人実践女子学園（実証協力）：
 学習者参加の受け入れ、授業内での実施協力、アンケート実施協力、教育現場としてのフィードバック提供

【教材作成連携】

台湾ファクトチェックセンター（協力団体）：
 教材開発における内容の監修、国際展開連携
 CekFakta（協力団体）：
 教材開発における内容の監修、国際展開連携
 MAFINDO（協力団体）：
 教材開発における内容の監修、国際展開連携
 Aliansi Jurnal Independen（協力団体）：
 教材開発における内容の監修、国際展開連携
 令和メディア研究所：
 教材開発における内容の監修
 株式会社メディアコロバ：
 教材開発における内容の監修
 Asian Network of News & Information Educators：
 教材開発における内容の監修

【クラウドサービス提供事業者】

グーグル合同会社：
 Google Cloud Platformのサービス提供
 株式会社PR TIMES：
 プレスリリース発信プラットフォーム
 NTTスマートコネク株式会社：
 クラウド型UTMソフトウェアの提供
 Hugging Face社：
 Hugging Faceのサービス提供

8-2. 全体スケジュール

主な実施事項	令和7年						令和8年	
	8月	9月	10月	11月	12月	1月	2月	3月
(1) 対策技術の開発	→							
1. 教材の開発								
• 学習カリキュラムの構築 (フェーズ1・フェーズ2)	→							
• ゲームフィケーション要素の設計	→							
• 学習内容設計・アセット作成	→							
2. システム開発								
① 全体設計								
• デザインシステム作成	→							
• スタック作成	→							
② データベース構築								
• 「レッスン」データを整理・作成	→							
• 「テスト」データを整理・作成	→							
③ システム作成								
• ログイン・デバックシステム開発	→							
• 各階層フレーム設計・作成	→							
• モジュール作成(解説・クイズ・実践・テスト)	→							
• プロフィールページ・メインメニュー作成	→							
• 共有システム開発	→							
• 学習アセットのシステム反映	→							
• 監修の専門家の皆様からのフィードバック反映	→							

8-2. 全体スケジュール

主な実施事項	令和7年						令和8年	
	8月	9月	10月	11月	12月	1月	2月	3月
(3) 社会実装に向けた取り組み	→							
1. 社会実装に向けた設計								
・ 事務局支援のもと市場・競合分析		→						
・ 事務局支援のもとニーズ調査・関係者ヒアリング				→				
・ 事務局支援のもと法規制・政策調査		→						
・ 事務局支援のもとビジネスモデルの構築					→			
・ 事務局支援のもと事業戦略策定					→			
・ 事務局支援のもと連携先候補の洗い出しとアプローチ			→					
・ 導入パッケージ作成（導入先マニュアル等）					→			
・ 国際展開戦略検討								
・ 事務局支援のもと国際展開に向けたヒアリング						→		
2. 普及啓発活動に向けた準備								
・ 事務局支援のもと営業・広報戦略策定・実施				→				
・ 事務局支援のもと提案資料の作成		→						
・ プレスリリース記事作成（初期・中間・最終）	→				→		→	
・ 広報準備（チラシ・LP等）							→	→
・ 教育関係者向けイベント企画・準備							→	→
・ 事務局支援のもと次年度事業計画の策定								→

※実施予定

8-2. 全体スケジュール

主な実施事項	令和7年						令和8年	
	8月	9月	10月	11月	12月	1月	2月	3月
(4)成果報告書及び社会実装実施計画書の作成	→							
・ アウトライン策定	→							
・ 月次報告書作成	→ ※実施中							
・ 中間ドラフト執筆	→							
・ 中間成果整理		→						
・ 中間報告書作成			→					
・ 最終版報告書・社会実装計画書作成				→				
・ 実証実験最終データ・分析結果追記				→				
・ 品質チェック・主要パートナーによる内容確認							→	

8-2. 全体スケジュール

主な実施事項	令和7年						令和8年	
	8月	9月	10月	11月	12月	1月	2月	3月
(5)普及啓発活動への協力	→							
・ プレスリリース配信（初期報告・中間報告・最終報告）	→					→		→
・ SNS等情報発信							※実施予定 →	
・ 関係者向けイベント開催								→
・ 自社教材の復習教材として販促								→
・ DMでの販促								→