

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・誤情報およびデバンキング情報拡散の
シミュレーション技術の開発・実証
成果報告書 概要版**

2026/3/19

技06_株式会社コンステラセキュリティジャパン

目次

1. 開発・実証における対策技術の開発

1. 開発技術によりアプローチする課題・目指す姿
2. 技術開発の取組・成果

2. 開発・実証における社会実装に向けた取組

1. 社会実装に係る取組・成果
2. 社会実装時のビジネスモデル等
3. 技術開発及び社会実装にあたっての課題・展望
4. 事業の拡大に向けた中長期的な計画

目次

1. 開発・実証における対策技術の開発
 1. 開発技術によりアプローチする課題・目指す姿
 2. 技術開発の取組・成果

2. 開発・実証における社会実装に向けた取組
 1. 社会実装に係る取組・成果
 2. 社会実装時のビジネスモデル等
 3. 技術開発及び社会実装にあたっての課題・展望
 4. 事業の拡大に向けた中長期的な計画

1-1. 開発技術によりアプローチする課題・目指す姿

開発技術によりアプローチする課題

現代の情報環境においては、偽・誤情報の発生頻度が増加しており、大きな脅威となっている。このような状況下において、対策に用い得るリソースは逼迫しており、優先的に対処すべき偽・誤情報の評価と迅速な対応が不可欠となっている。

課題	課題の詳細
①生成AIによる脅威の増大と対策リソースの不足	生成AIにより偽・誤情報の流布が容易になるにつれ、分析・対策リソースが逼迫しており、対処すべき情報の優先順位を効率的に評価することが重要である
②国際的な情報共有の潮流と国内における連携の課題	西側諸国では偽・誤情報対策の情報共有が標準化されつつある一方、日本では国内でも共有が標準化されておらず、組織間での情報共有が困難である
③日本市場・偽情報対策に特化した分析ツールの不在	多数のSNS分析ツールは海外市場や別用途を主眼としており、日本語等特定言語の処理及び理解精度や偽・誤情報に特化した分析が不十分である
④変化する情報環境と既存ソリューションの限界	SNSや分析技術が不規則に追加・更新される中、柔軟に機能を追加でき、急速に変化する情報環境に対応可能なツールが必要である

上記課題を踏まえ目指す姿・ゴール

本開発技術は、インターネット上に拡散される情報のリスクを定量的に評価し、影響が大きいと見込まれる偽・誤情報を早期に検知・分析し、偽・誤情報の打ち消しを支援する仕組みを提供する。

	本開発技術の機能	提供する価値
①	日本語環境でSNSの情報がどのように広がるかを再現し、 <u>シミュレーション</u> を通じて偽・誤情報やそのファクトチェックの広がりを予測する	拡散の可能性を元に対応の必要性を判断し、対応のためのリソースの効果的な分配を可能とする
②	脅威情報の共有・管理基盤として欧米で広く利用されている <u>OpenCTI</u> 上で、脅威情報を記述するための国際標準である <u>STIX</u> に基づき情報を整理することで、他ツールとの連携および偽・誤情報の評価基準の統一を可能にする	国内外の組織間での円滑な情報共有を実現する
③	日本語ならではの <u>特徴や表現を理解したうえでSNS上の情報を分析できる仕組みを整備する</u>	日本語の特性を踏まえることで、文意をより正確に理解した分析が可能となり、実務に活用可能な信頼性の高い分析結果を得られる
④	<u>データの取得・加工・表示方法をあらかじめ統一した設計とすることで、新たなデータ源や分析手法の柔軟な追加が可能である</u>	SNSや分析技術の追加・更新に容易に対応可能な構造であり、最新の情報環境に合わせて運用できる

1-2. 技術開発の取組・成果

開発技術に係る取組・成果

取組の概要

SNS上の偽・誤情報を分析するための統合プラットフォームを開発した。本プラットフォームは、脅威情報管理プラットフォームであるOpenCTIをベースとし、脅威情報を記述するための国際標準であるSTIXに基づき、データを標準化して管理する。本開発技術は、データの統合を担うコネクタ、データに対して加工を行うエンリッチメント機構、シミュレーション機能を含む高度な分析を提供する分析機構の三つの主要要素から構成される。これらの処理を単一のシステム上で一元的に実行できる点を特徴とする。

本取組において開発するプラットフォーム



取組の成果

統合プラットフォーム

内部に取り込んだデータやその分析及び評価結果を表示するダッシュボード機能を実装した

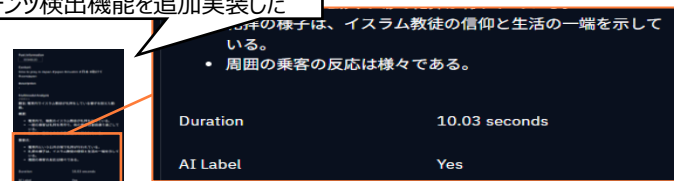
プラットフォーム別投稿数やエンゲージメントの推移などデータ全体の傾向を可視化



OpenCTI上で表示される分析結果のダッシュボード

機能単位で構成された設計を採用しており、柔軟な機能の追加や置換が可能である

実際に、ステークホルダーからのフィードバックを踏まえ、AI生成コンテンツ検出機能を追加実装した



プラットフォームの操作画面上で表示されるAI生成コンテンツ検出の結果例

データ収集・構造化

機能概要

- X、YouTube、TikTok、ニュース記事等複数の情報源からのデータを格納
- 各プラットフォームに対応したコネクタを実装し、リアルタイム/準リアルタイム分析に対応
- 生データを自動でSTIX 2.1形式へ変換し、構造化

ユーザへの提供価値

構造の異なる複数情報源のデータを統一形式へ自動変換することで、単一プラットフォーム上でのデータの一元管理が可能となる

1-2. 技術開発の取組・成果

開発技術に係る取組・成果：エンリッチメント機構

エンリッチメント機構とは、OpenCTI上に集積されたデータに対し、自動で分析結果や評価を付与し、データの価値を向上させるための機能群である。具体的には以下のようなデータ加工と分析を提供する。

モジュール	機能概要	ユーザへの提供価値
センチメント・エモーション分析	特定の対象に対し、各投稿が抱くセンチメントやエモーションを確信度スコアとともに判定	特定の対象に対して投稿者の抱く感情や意図的に喚起される感情を把握することで、偽・誤情報の危険性や影響の評価に活用できる
SNSアカウント真正性分析	SNSアカウントが人間によって運用されているか、自動投稿や拡散を行うボットによって運用されているかを、確信度スコアとともに判定	人工的に引き起こされた情報拡散を識別するための指標を提供する
埋め込み表現の取得	投稿テキストの持つ意味を数値化	キーワードや表記揺れに左右されることなく、意味的に近い投稿群を抽出できるため、話題のまとめりや関連投稿の発見が容易となる
エンティティ検知	人物・組織・場所・イベント等テキスト内に登場する対象を複数言語で自動抽出	拡散する情報の中から特定の対象に関連する投稿を正確に抽出できるようになる
モーダル横断メディア理解	画像・動画形式のデータの意味を解析し、テキスト形式データと一元的に分析できるようデータ処理	テキストに加え、画像や動画を含めた統合的な分析が可能となり、偽・誤情報を含む画像・動画の検知を効率化できる
アカウントプロファイリング	偽・誤情報分析において、拡散に関与するアカウントの特性を直近1か月の投稿とプロフィール情報から分析	偽・誤情報の拡散に関与し得るアカウントの特徴を、統合された情報に基づき自動的かつ短時間で判断できる
AIコンテンツ検出※	SNS投稿に添付された画像・動画を対象としてAI生成コンテンツであるか否かとその根拠を自動判定	生成AIによる偽情報を検出するための判断材料を提供する

※ステークホルダーからのフィードバックを踏まえ追加実装

1-2. 技術開発の取組・成果

開発技術に係る取組・成果：分析機構①

トピック分析・ナラティブ検知・主張抽出及び拡散傾向の分析・可視化

本機能により、特定のデータ群から主要なトピックやナラティブ、主張を抽出し、「何が語られているか」を特定できる。さらに、それらが時間の経過とともにどのように伝播していくかを可視化することが可能である。

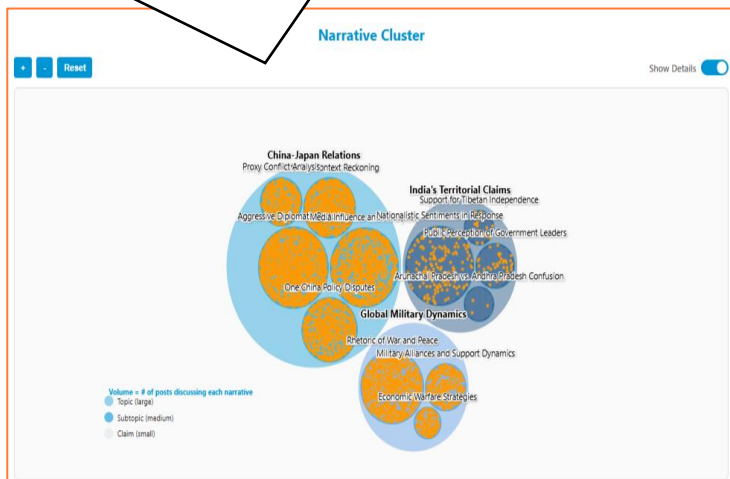
機能

- 特定のデータ群を対象にトピック・ナラティブ・主張を階層的（トピック > ナラティブ > 主張）に自動抽出する
- 抽出したトピック・ナラティブ・主張の伝播の推移や規模を分析し、クラスタ形式で可視化する

ユーザへの提供価値

- 特定の期間やプラットフォームにおける議論のトピックや具体的な主張を構造的に把握し、それらを中長期的に比較することが可能となる
- 偽・誤情報拡散状況や拡散の起点の理解に活用できる

最も大きな円はトピックのまとまりを示し、その内部にナラティブおよび主張が階層的に可視化されている。また、時間の経過に伴う伝播状況が色の変化で表される



外部の可視化機能により表示されるナラティブクラスター

左図のナラティブクラスターをOpenCTI上で可視化した結果。中央にはトピックが配置されており、そこから線で接続された先に当該トピックに含まれるナラティブが示されている。さらに、そのナラティブから線で接続された先に主張が位置付けられている



OpenCTI上で可視化されるトピック・ナラティブ分析結果例

1-2. 技術開発の取組・成果

開発技術に係る取組・成果：分析機構②

オーディエンス・モデリング及び拡散シミュレーション

本機能によりSNS上における偽・誤情報及びファクトチェック情報の拡散を予測・シミュレーションすることで、効果的な偽・誤情報対策立案を支援する。

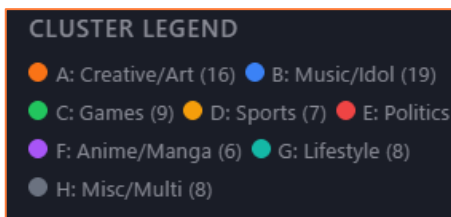
機能

- SNS環境におけるオーディエンスを興味・関心に基づいてグルーピングし、その上で各グループがどのような情報に影響を受けやすいかをモデル化する
- 情報の発信元となるクラスタ、発信メッセージ、何時間先までを何時間ごとに、何段階先の拡散までシミュレーションするかなどユーザが設定した条件に基づいた拡散シミュレーションを実行できる
- 設定した発信メッセージが、時間の経過とともにどのオーディエンスにどの程度受容されるか、どのような経路・速度で拡散するかを予測できる
- 偽・誤情報の拡散のみならず、その対応として発信されるファクトチェック情報がどのように拡散し、どの程度の抑制効果を持ち得るかを評価できる

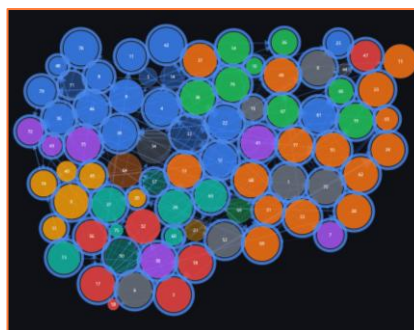
ユーザへの提供価値

- 情報拡散シミュレーションにより、各偽・誤情報の影響度を予測し、優先的に対応すべき対象を効率的に判断できる
- 偽・誤情報への対抗措置として、ファクトチェック情報を「いつ・どのように・どの主体から」発信すべきかの判断材料を提供する
- 特定の偽・誤情報が拡散している状況下において、ファクトチェック情報が有効に機能するか、あるいは逆に偽・誤情報の露出を拡大させる可能性がないかを予測し、適切な対応判断に資する

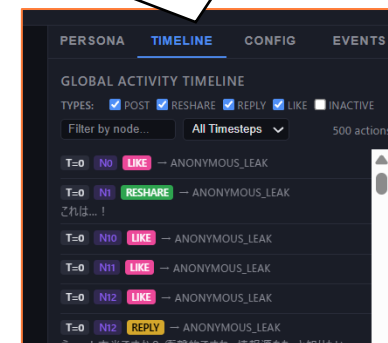
拡散された情報に対するオーディエンスの反応を予測しタイムライン形式で表示



カテゴリによって色分けしたグループ間の情報拡散経路を矢印で可視化



プラットフォームの操作画面上で表示される拡散シミュレーション



目次

1. 開発・実証における対策技術の開発
 1. 開発技術によりアプローチする課題・目指す姿
 2. 技術開発の取組・成果

2. 開発・実証における社会実装に向けた取組
 1. 社会実装に係る取組・成果
 2. 社会実装時のビジネスモデル等
 3. 技術開発及び社会実装にあたっての課題・展望
 4. 事業の拡大に向けた中長期的な計画

2-1. 社会実装に係る取組・成果

社会実装に係る取組・成果

1. ステークホルダーへのシステム紹介とヒアリングの実施

開発の主要な進捗ごとにステークホルダーに対してシステム紹介及びヒアリングを行い、各段階における機能や想定する分析等に関するレビューや意見を収集した。

主な取組成果

総ヒアリング件数：25件

協力組織（公的機関・報道機関）：17組織

主なフィードバック

全体的な評価

開発技術の有用性については概ね好評を得た

主なニーズ

実務に耐える分析精度の確保が求められた

主なリスク・懸念点

データの誤解釈やシステム誤用のリスクが指摘され、当社サポート体制の整備が求められた

2. 月次レポートの発信

関係組織に対して本開発技術を用いた偽・誤情報の調査・分析成果をまとめたレポートを配信した。

主な取組成果

毎月平均170人にレポートを配布することで一定数の関係組織に対し継続的なエンゲージメントを図ることができた

偽・誤情報の分析事例を、当社の洞察や推奨事項とともに共有することで、偽・誤情報対策に対する社会的な意識向上及び啓発を図ることができた

2-1. 社会実装に係る取組・成果

社会実装に係る取組・成果

令和7年富士調査研究会同への出展

マネージドサービスの試験導入

概要

- 令和7年富士調査研究会同へ出展し、官側関係者等に対し開発技術の初期紹介を行った

- 開発技術を用いて対象企業に関するSNS上の言説分析を実施し、その分析結果をレポートとして同社に提供した

主な取組成果

- 官側関係者等と技術に関する情報共有・意見交換を行い、有用なフィードバックを収集することができた

- 開発技術を用いた分析結果をレポートとして提供するマネージドサービスを試験的に導入した
- 同社に対し、来年度に向けて、現在マネージドサービスの実運用を提案中である

主なフィードバック

- 対応プラットフォームの拡充や、ポット検知・ファクトチェック等の機能強化に対するニーズが示された
- 実務上、報告用途を想定し、数値指標を重視した分析が求められた

- なし

展示会の様子

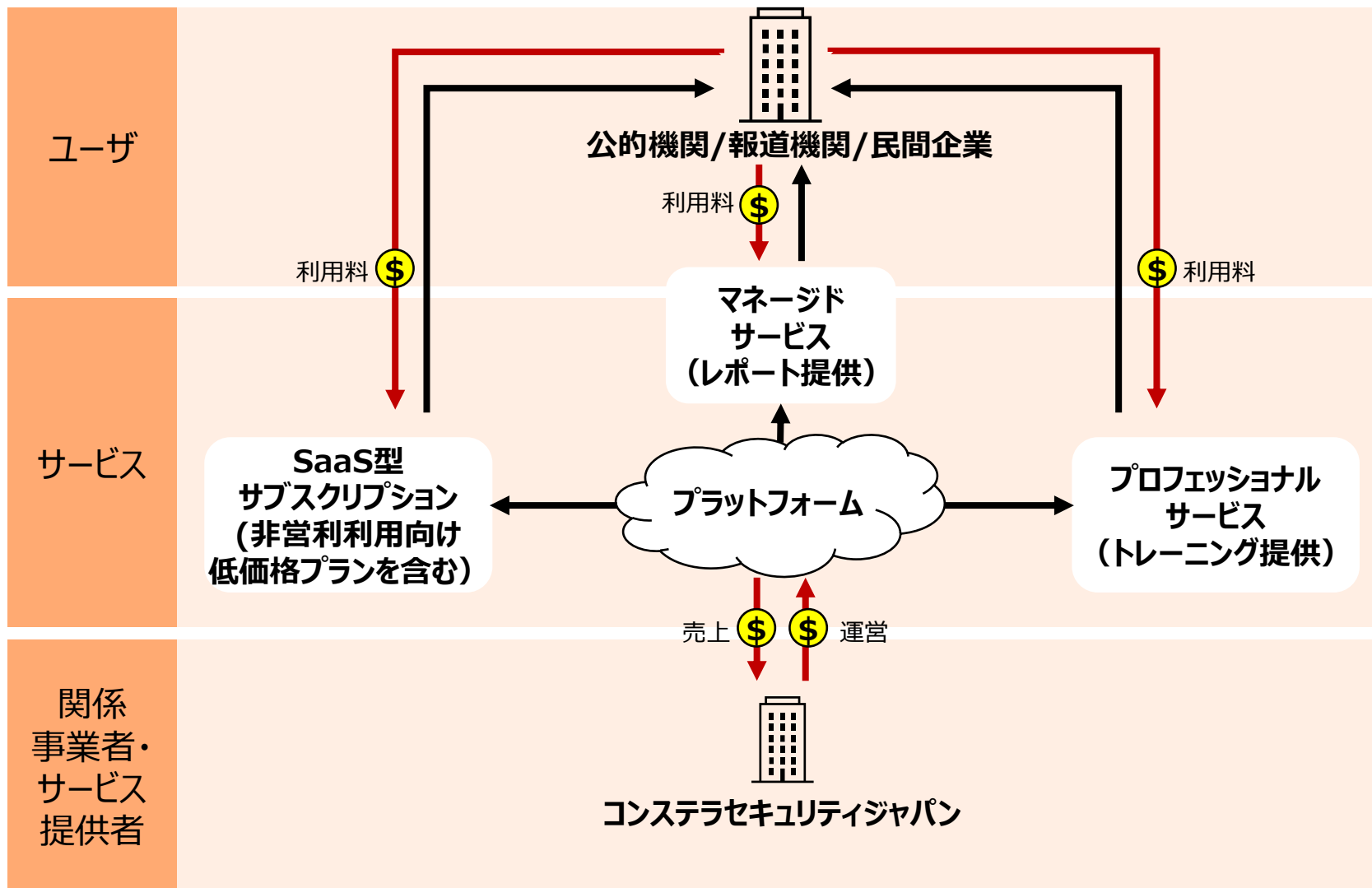
本開発技術の当時のプラットフォーム操作画面

写真・スクリーンショット



2-2. 社会実装時のビジネスモデル等

社会実装時のビジネスモデル



2-2. 社会実装時のビジネスモデル等

ユーザ・導入先の詳細とそのペインポイント

以下ではヒアリングの結果明らかとなった各ユーザ・導入先の主なペインポイントと各ペインポイントに対する開発技術の効果を整理した。

	ユーザ・導入先のペインポイント	本開発技術の効果
公的機関	既存ツールでは現状把握にとどまり対応施策の検討や実行へ展開することが難しい	偽・誤情報及びそれを打ち消す情報の拡散予測・シミュレーションにより、 <u>対策立案を支援する</u>
	多言語空間における偽・誤情報のモニタリングが必要とされているが十分な対応が困難である	システム構成と入出力形式の標準化により、必要に応じて対応言語を柔軟に追加可能である
	定量的表現が不足しており分析結果を実務上の報告資料として活用することが難しい	スコア判定を採用することで分析結果を数値化し、 <u>根拠として活用可能な形で出力できる</u>
報道機関	幅広いテーマに対するファクトチェックが求められる一方で十分な検証報道を実施することが難しい	拡散予測・シミュレーションにより優先的に対処すべき情報を効率的に評価できる
	偽・誤情報を報道することでかえって拡散を助長するリスクを適切に評価することが難しい	拡散予測・シミュレーションによりファクトチェック情報を拡散した際の効果を予測できる
	情報作戦においてその背後にいる主体までを分析し、根拠に基づいて報道することが求められている	アカウント特性の推定により偽・誤情報拡散に関与する <u>SNSアカウントの属性把握を支援する</u>
	既存ツールでは報道用途に耐えうるセンチメント分析精度の確保が困難な場合がある	主に日本語を対象としたセンチメント・エモーション分析により、特定の対象に対する感情を確信度付きで評価でき、 <u>信頼性の高い分析結果を提供する</u>
民間企業	風評被害がSNS上でどの程度拡散しているのかや対策の効果が不透明であるため、経済合理性の観点から対策の実行に踏み切りにくい	風評被害の拡散状況を把握し、対策の効果を予測することで、 <u>対策の有効性を判断できるようになり、経済合理性に基づいた意思決定を可能にする</u>

2-3. 技術開発及び社会実装にあたっての課題・展望

技術開発及び社会実装にあたっての今後の課題

関係者ヒアリングや市場動向・顧客ニーズの調査、社内検討を通じて明らかになった技術開発及び社会実装にあたっての課題を、今後の展望とともに以下に整理した。

技術開発における課題

- 開発過程において、開発技術はOpenCTI上に構築されているため、その仕様の影響により、データ処理性能や可視化形式の柔軟性に制約が生じ得ることを確認した
- 開発過程において、外部の大規模言語モデルの使用の影響により一部分析性能に制約が生じ得ることを確認した
- 拡散シミュレーション開発過程において、分析対象がアカウント自らが発信した公開情報に限定されることにより、アカウント特性の推定精度に制約が生じるという課題に直面した
- データ処理の停止や遅延を想定した検証の結果、対応手段の整備が不十分であり、復旧が長期化する可能性があることを確認した

社会実装における課題

- ステークホルダーへのヒアリングの結果、製品化に向けた具体的なニーズは把握できたが、導入に向けた製品化の詳細はまだ整理されていない
- 民間企業における偽・誤情報対策への関心は高まりつつあるものの、対応の優先度は依然として低く、導入の難易度が高いため、市場が公的機関に限定される可能性がある
- ステークホルダーから、シミュレーション結果が攻撃者視点でも有用になり得る点について懸念が示された
- 本開発技術を実際に事業化し、顧客提案へ移行するためには、監視・障害対応を含む持続可能な運用・保守体制の確立が不可欠である

上記課題を踏まえた今後の展望

技術開発における展望

- データ処理方式の改善や処理基盤の強化によりデータ処理性能を緩和するとともに、OpenCTI開発元との協議や外部ダッシュボードの活用、分析結果の表示画面のカスタマイズなどの代替手段を検討し可視化形式の柔軟性を補完する
- 検証済みデータを用いたモデル追加学習・微調整等により継続的に分析性能の向上を図る
- 社会的・心理学的知見や関連情報を統合することにより、推定精度の向上を図る
- 運用時を想定した保守・復旧体制を強化する

社会実装における展望

- 今後は分析の質に加え、必要なデータ量やアクセス人数など、製品化に必要な非機能要件の定義を進める必要がある
- 幅広い組織との連携を強化し、各種組織にとって適切な利用用途の検討を進める
- 特定のモジュールの提供を一部の顧客に限定するとともに、エンドユーザー向けの教育、ガイドライン、および管理体制を整備する
- 運用・保守体制を整備し、顧客からのフィードバックを踏まえ継続的な改善を図る

2-4. 事業の拡大に向けた中長期的な計画

事業の拡大に向けた中長期的な計画			
	フェーズ1 (2026年度)	フェーズ2 (2027-2028年度)	フェーズ3 (2029-2030年度)
公的機関	<ul style="list-style-type: none"> マネージドサービス プロフェッショナルサービス SaaS型サブスクリプション (パイロット運用) 	<ul style="list-style-type: none"> マネージドサービス プロフェッショナルサービス SaaS型サブスクリプション 	<ul style="list-style-type: none"> マネージドサービス プロフェッショナルサービス SaaS型サブスクリプション
報道機関	<ul style="list-style-type: none"> マネージドサービス (公益性の高い非営利連携を中心とした協力活動) 	<ul style="list-style-type: none"> マネージドサービス SaaS型サブスクリプション (パイロット運用) 	<ul style="list-style-type: none"> マネージドサービス プロフェッショナルサービス SaaS型サブスクリプション
民間企業	<ul style="list-style-type: none"> マネージドサービス プロフェッショナルサービス 	<ul style="list-style-type: none"> マネージドサービス SaaS型サブスクリプション (パイロット運用) 	<ul style="list-style-type: none"> マネージドサービス プロフェッショナルサービス SaaS型サブスクリプション
料金形態	<ul style="list-style-type: none"> 案件単位契約 	<ul style="list-style-type: none"> 案件単位契約 サブスクリプション 	<ul style="list-style-type: none"> 案件単位契約 サブスクリプション 非営利利用向け低価格プラン