

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**

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

1-1. 開発・実証のサマリ

<p>アプローチする課題・目指す姿</p>	<p>近年、ソーシャルメディアや生成AIの発展により偽・誤情報が急速に増加している。一方、国内では標準化された仕組みが存在せず、組織を跨いだ知見の共有が十分に行われていないほか、既存の商用プラットフォームでは日本語・偽・誤情報に特化した分析精度や技術更新に課題がある。本開発技術は、偽・誤情報及びデバッキング情報におけるナラティブ分布の可視化や拡散予測を通じて、効果的な発信ノードやネットワーククラスターを特定することにより、限られた人的資源の中でも効率的な対策立案を可能とする。加えて、OpenCTIとの相互運用性を確保し、日本の脅威環境に即した統合的な偽・誤情報対策基盤の構築を目指す。</p>		
<p>技術区分</p>	<p>情報流通状況の可視化・分析技術、情報の拡散防止・無効化技術</p>	<p>実施体制 (下線：技術開発主体)</p>	<p>(株) コンステラセキュリティジャパン</p>
<p>対象とするモデル種</p>	<p>文章、画像、動画</p>		

技術開発の取組・成果

- 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OpenCTIを基盤とし、国際標準に準拠した偽・誤情報分析プラットフォームを開発した
- 複数のプラットフォームに対応したデータ格納基盤を整備し、取得データを脅威情報を記述するための国際標準であるSTIX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を可能とした
- 集積されたデータに対して自動で分析結果や評価を付与するエンリッチメント機構を実装し、偽・誤情報の早期検知を支援する機能群を実装した
- 日本語環境におけるSNS上の情報拡散をモデル化し、同モデルを用いて偽・誤情報およ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機能を実装した。これにより、偽・誤情報の影響度や対応の有効性を予測し、効果的な対策立案およびリソースの最適配分を支援する

社会実装に係る取組・成果

- ステークホルダーへ開発段階ごとにシステム紹介及びヒアリングを実施し、実務的観点から本開発技術の有効性を検証するとともにシステム機能に関する意見を収集した
- 社内アナリストが開発技術を用いて偽・誤情報を分析し、その結果を当社の洞察及び推奨事項とともに月次レポートとして顧客・取引先に配布した。これにより、社会全体における偽・誤情報対策への意識啓発を行うとともに、本開発技術の有用性を示した
- 令和7年富士調査研究会同へ出展し、本開発技術の紹介及び意見交換を行った
- 顧客企業に対して本開発技術を用いたサービスのPoCを実施し、来年度への契約獲得に向けた提案を行った

技術開発及び社会実装にあたっての課題・展望

- 外部基盤の仕様に起因する制約を緩和するため、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
- 公開情報に基づく分析には、アカウント属性等の推定精度に限界があるため、社会学・心理学的知見や関連情報を統合することで推定精度の向上を図る
- 製品化に向けてスケーラビリティやユーザインターフェース等の非機能要件の定義を進める
- 市場拡大に向けて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
- 実運用を想定したエンドユーザへの教育やサポート体制を確立するとともにユースケースを整理する
- 持続的な運用体制を確保するため運用・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

代表者コメント



株式会社コンステラセキュリティジャパン 代表取締役社長 宮村信男

偽・誤情報及びデバッキング情報の拡散は、安全保障の観点からも極めて重要な課題である。本事業は挑戦的なテーマであったが、想定していた主要機能は概ね計画どおり実現できたと認識している。今後は、処理性能やスケーラビリティ、運用上の安全性を一層高め、より幅広い実環境で活用可能な技術として成長させたい。

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

インターネット上の偽・誤情報等への対策技術に関して、諸組織が現在抱える課題は、大きく分けて以下の4点である。

①生成AIによる脅威の増大と
対策リソースの非対称性

②国際的な情報共有の潮流と
国内における連携不足

③日本市場・偽情報対策に特化した
分析ツールの不在

④変化する情報環境と既存ソリューションの限界

以下では、各課題の詳細について記載する。

①生成AIによる脅威の増大と対策リソースの非対称性

必要とされるリソースの非対称性

防御側

生成AIを使用する攻撃側

偽・誤情報の大量・高精度分析と的確な対応の必要性



多数の偽・誤情報のうち一つでも拡散されれば
攻撃が成立する構造

対応すべき対象の選別と資源配分の最適化が必要

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

②国際的な情報共有の潮流と国内における連携不足

偽・誤情報対策のための情報共有の現状

西側諸国

日本

権威主義国家による西側諸国を対象とした協働的な情報操作活動の拡大に対抗するため、偽・誤情報共有基盤を整備し、情報共有を促進



組織間の情報共有が不足しており、横断的なインテリジェンス強化が進展せず

日本における情報共有の課題

背景

偽・誤情報等の対策のために提供されるツール群や対策役務の出力フォーマットの未標準化・未正規化

情報共有形式

レポート形式の情報共有

CSV形式の情報共有

課題

膨大となるレポート全てへの対応の蓋然性や知識の集積における属人性

利用システムの差異に基づく自動的な書き出しや読み込みの困難

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

③ 日本市場・偽情報対策に特化した分析ツールの不在

背景

海外市場向けツールへの依存

マーケティングを主用途とした分析ツールへの依存

課題

日本語文脈・感情・含意理解の不足

政治的分析・プロファイリング・ボット検知機能が低精度

④ 変化する情報環境と既存ソリューションの限界

背景

商用ツールへの依存

モリシック設計によるツール開発

課題

対応プラットフォームや分析機能の制限
データ統合の柔軟性不足

新たな媒体や新規分析機能への対応の遅滞

ソーシャルメディアや関連技術の頻繁な仕様変更に対応可能な偽・誤情報分析ツールの必要性

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

本事業では、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・誤情報の早期検知・分析・無効化を支援する機能群を提供する。

①生成AIによる脅威の増大と対策リソースの非対称性へのアプローチ

取組

- システムに収集され分析された偽・誤情報事案について、それらがどの程度、どの対象に拡散するかの予測を、拡散されている情報、その拡散源及び周辺のアカунトの特性及びネットワーク、過去の類似事案の実績等に基づく機械学習により提供する
- 偽・誤情報等を打ち消す情報の有効的な注入方法及び注入先についてシミュレーションを実施する



想定される顧客への提供価値

- 情報拡散が乏しい、ないし衆目に触れない程度の偽・誤情報等については分析・対応によってあえて世間の耳目を集める事を避ける事で使用者の労力を保持する
- 重要事案の対応への注力を可能にすることで、リソース分配の最適化を実現する

②国際的な情報共有の潮流と国内における連携不足へのアプローチ

取組

- OpenCTI (P.41参照)におけるソーシャルメディア・その他インターネット関連のデータや、偽・誤情報等において必要となる各種分析をSTIX (P.41参照)形式で定義する
- モジュール方式により既存のツール等との親和性を担保する
- データ・一次分析及び分析官により評価を標準化する



想定される顧客への提供価値

- 円滑なインテリジェンス共有を実現する

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

③ 日本市場・偽情報対策に特化した分析ツールの不在へのアプローチ

取組

- 日本語話者かつ偽・誤情報等の分析において実務経験・知見を有する者を開発に参与させ、大規模言語モデルによるテキスト分析のベースラインやトレーニングデータセットにそれら担当者による評価を追加する
- 実務者視点で必要な分析を用意し、モジュール形式により追加の要望にも柔軟に対応する

想定される顧客への提供価値

- 日本において偽・誤情報等の効果的な対策を行うためのシステムを構築する

④ 変化する情報環境と既存ソリューションの限界

取組

- モジュール式のアーキテクチャを採用し、情報の収集・加工・分析の各領域について入出力形式を標準化する

想定される顧客への提供価値

- 変化するSNS環境に対し、新規プラットフォームやニッチSNSを含む収集対象を迅速に追加できる
- 設計上、各機能においてモデルの追加や置換を柔軟に行うことができ、顧客の要望に応じたモデルの選択や、技術発展に伴う最新のモデル、他組織が開発したモデル等の平易な統合を可能とする

以上のように、本開発技術では、限られた人的資源の中で効果的な情報分析・共有・検証・対応を実現する。欧米で主流となっているOpenCTIとの相互運用性を担保しつつ、日本語を含む国内の脅威環境に即した実効的な偽・誤情報対策を可能にする統合的な運用基盤の構築を目指す。

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：全体的な分析の流れ

分析段階	詳細	得られる知見
① 入力	ユースケースに応じて選定したキーワード等を含む検索クエリから抽出されたデータを入力として用意する	—
② コネクター・OpenCTIによるデータ統合及びエンリッチメント機構によるデータ加工	用意されたデータを統合・構造化し、正規化した上で、各種エンリッチメント機能（センチメント・エモーション分析、アカウント真正性分析によるボット検知、投稿文章に含まれる分析対象主体の抽出等）の結果を付与し、可視化する	<ul style="list-style-type: none"> ・ 拡散された情報の性質及び拡散者の特徴の把握 ・ 調査対象及び組織化された拡散の兆候の特定
③ 分析機構を用いた拡散状況の把握と対策支援	投稿等に含まれたナラティブを抽出して拡散の起点や経路を分析し、影響され得るオーディエンスの反応を予測する。さらに、拡散シミュレーションを通じて、発信クラスター間の拡散の流れと規模を可視化し、偽・誤情報への対抗情報の発信を支援する	<ul style="list-style-type: none"> ・ 偽・誤情報/対抗情報の拡散経路および影響規模 ・ 各種対策の内容、発信タイミング、発信チャネルの最適化
④ 期待される効果	偽・誤情報の早期検知と拡散構造・影響範囲の把握を通じて、効果的な対抗情報発信や対策の立案を支援する	

2-3. 開発技術により対処可能なユースケース

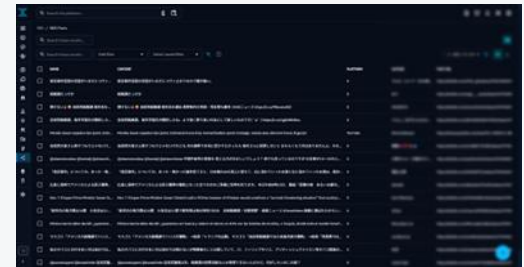
開発技術により対処可能なユースケース①：選挙における偽・誤情報拡散の検知・対策

① 入力

選挙関連キーワード（政党名等）・期間等を含む検索クエリから抽出されたデータ

② コネクター・ OpenCTIによる データ統合 及び エンリッチメント機 構による データ加工

- 選定したSNS等から抽出されたデータを統合・構造化し、時系列で表示する
- センチメント・エモーション分析により、候補者・政策等に関する感情傾向を把握し、社会的反応や異常な反応の兆候を捉える
- アカウント真正性分析により、ボット等の組織的拡散の兆候を識別する
- アカウントプロファイリングを通じて、特定の主張を増幅するアカウント群を把握し、選挙について協調的・組織的な拡散構造を明らかにする
- AI生成コンテンツ検出機能を用いて、候補者関連の画像・動画を分析し、ディープフェイク等の不正コンテンツを検知する



OpenCTI上で表示される投稿の一覧



OpenCTI上のダッシュボードで表示される
エンリッチメント結果

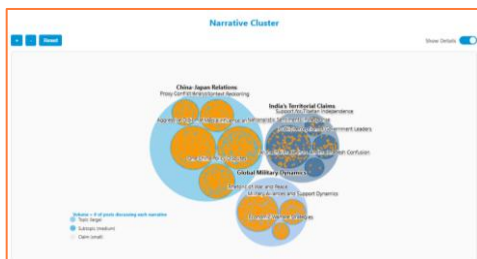
2-3. 開発技術により対処可能なユースケース

開発技術により対処可能なユースケース①：選挙における偽・誤情報拡散の検知・対策

③ 分析機構を用いた拡散状況の把握と対策支援

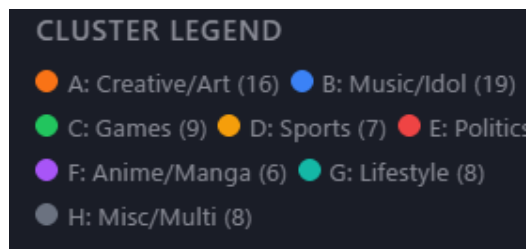
- 選挙に関する特定の偽・誤情報やナラティブについて、拡散、経路、速度、規模を可視化・分析
- 特定の有権者層が影響を受けやすい偽・誤情報を評価
- 選挙管理委員会、ファクトチェック機関、報道機関等が発信すべきデバンキング・プレバンキング情報について、内容・タイミング・発信チャネルの最適化を拡散シミュレーションにより支援

ナラティブ抽出



選挙に関する議論のトピック及び拡散経路を把握

オーディエンスモデリング



影響を受けやすい有権者層を特定

拡散シミュレーション



デバンキング・プレバンキング情報について、内容・タイミング・発信チャネルの最適化

④ 期待される効果

早期検知・把握

選挙妨害を目的とした偽・誤情報キャンペーンの早期発見と影響範囲の迅速な把握

構造理解

偽・誤情報を拡散する主要なアクターやネットワークの特定と、その活動パターンの解明

戦略的対処

客観的なデータに基づく効果的な偽・誤情報対策戦略の立案と、限られたリソースの最適配分

2-3. 開発技術により対処可能なユースケース

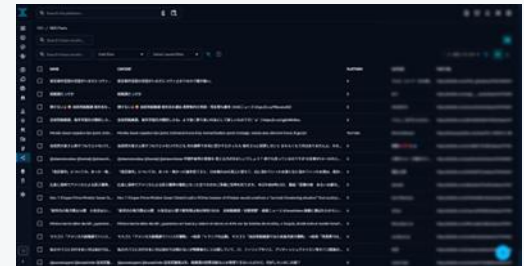
開発技術により対処可能なユースケース②：BtoCメーカーのネット上での風評被害の検知・対処

① 入力

メーカー関連キーワード（ブランド名、商品名等）を基に抽出されたデータ

② コネクター・ OpenCTIによる データ統合 及び エンリッチメント機 構による データ加工

- 主要SNS等から収集された投稿、コメント、ニュース記事等のデータをOpenCTI上に正規化した形式で表示する
- センチメント・エモーション分析によりメーカーやブランドに対するネガティブキャンペーン及び風評被害の兆候を検知する
- エンティティ検知により、批判の対象となっている製品・人物等の主体を早期に特定し、批判内容を整理する
- アカウント真正性分析により、ボットやなりすましアカウントの関与を把握し、風評被害が組織的に増幅されている可能性を見極める
- AI生成コンテンツ検出機能を用いて、AIによって生成・加工された画像及び動画を検知し、拡散コンテンツの信頼性を判定する



OpenCTI上で表示される投稿の一覧



OpenCTI上のダッシュボードで表示される
エンリッチメント結果

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②：BtoCメーカーのネット上での風評被害の検知・対処

③ 分析機構を用いた拡散状況の把握と対策支援

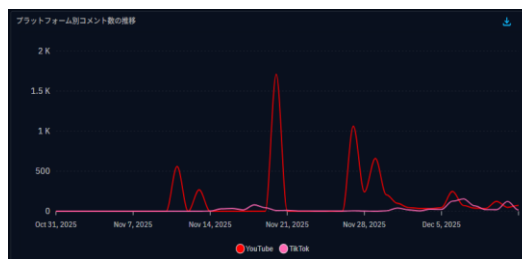
- ネットワーク分析及び時系列分析により、複数のSNSで展開されるネガティブ情報の起点や拡散経路を特定する
- 拡散傾向の分析機能を用いて、プラットフォームを横断した拡散状況を可視化し、到達範囲や拡散速度、拡散の増幅要因を把握する
- 拡散シミュレーションにより、企業による公式声明等の対策ごとに沈静化効果と再燃リスクを比較する

ネットワーク分析



OpenCTI上の可視化で表示される分析

プラットフォームを横断した拡散



プラットフォームごとの情報拡散の可視化

拡散シミュレーション



対抗情報発信の最適化

④ 期待される効果

早期検知・把握

ブランド・商品に関するネガティブ投稿や偽・誤情報の兆候の早期検知により、炎上・風評被害の拡大リスクを低減

構造理解

批判対象の製品・人物・キャンペーンを抽出して批判言論を整理し、なりすましアカウントの関与を可視化し、拡散構造を把握

戦略的対処

公式声明等の対応策を比較し、鎮静効果と見積もった上で、優先すべき対象と推奨事項を提示し、迅速な信頼回復を支援

2-3. 開発技術により対処可能なユースケース

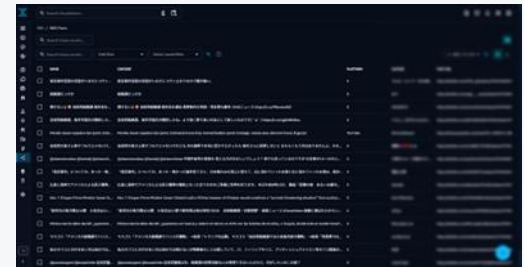
開発技術により対処可能なユースケース③：特定のトピックにおける海外勢力による干渉の検知・対処

① 入力

事前に定義したトピック、既知のFIMIキャンペーン関連キーワードを基に抽出されたデータ

② コネクター・ OpenCTIによる データ統合 及び エンリッチメント機 構による データ加工

- 国内外のオープンソース情報とサイバー脅威情報を一元化し、サイバー・認知等の複数ドメインに跨る、海外勢力による干渉の兆候の把握と分析を効率化する
- アカウント真正性分析やSNSアカウントプロファイリングを通じて、海外勢力による情報操作を示唆する可能性のあるアカウント及びボットの活動を識別する
- モーダル横断分析により、加工された画像や映像の利用の検知し、情報操作や印象誘導の可能性を見極める



OpenCTI上で表示される投稿の一覧



OpenCTI上のダッシュボードで表示される
エンリッチメント結果

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③：特定のトピックにおける海外勢力による干渉の検知・対処

③ 分析機構を用いた拡散状況の把握と対策支援

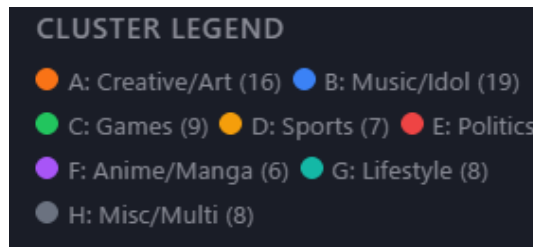
- ネットワーク分析及びナラティブの時系列追跡により、複数のプラットフォームを横断して展開される可能性のあるFIMIキャンペーンの全体像の明確化
- オーディエンスモデリング等を通じて特定の主張や情報を増幅するアカウント群とターゲットコミュニティを特定し、その影響を評価
- 拡散シミュレーションを用いて、FIMIキャンペーンの到達範囲や影響を予測し、対抗情報発信戦略の有効性を事前に検討

ナラティブ抽出・追跡

政権影響下の市場操作	高市政権による解散情報を利用した株式市場の動向と投資機会
選挙相場の形成	解散前後の市場の動きや市況への期待とそのリスク
長期的視点の欠如	選挙の頻度が増すことで政策が短期志向になる懸念
短期的なバラマキ政策	選挙前の支持率上昇を狙った即効性のある政策提案の増加
有権者の受け止め方	短期的な政策や選挙の運営が有権者の判断に与える影響

特定のトピックに関するナラティブ一覧

ターゲットコミュニティの特定



情報を増幅するアカウント群を特定

拡散シミュレーション



FIMIキャンペーンの到達範囲や影響を予測する

④ 期待される効果

早期検知・把握

海外勢力による情報操作・干渉活動の早期検知と帰属推定支援能力の向上

構造理解

FIMIの戦術、標的オーディエンス、主要ナラティブに関する理解の深化

戦略的対処

外国からの不当な影響工作へのより迅速かつ効果的な対処（関係省庁間の情報共有、プラットフォーム事業者への協力要請、注意喚起等）への貢献

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

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ー、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。



OpenCTIを中核とする情報統合基盤

本プラットフォームは、OpenCTIを中核とする情報統合基盤として、STIX 2.1に基づく統一的なデータ管理により、SNS投稿・アカウント情報・ニュース記事などの構造化データを一元的に蓄積・管理している。あわせて、カスタムSTIX 2.1拡張（SnsPost、SnsAccount、Article）を実装することで、ドメイン固有要件に対応しつつ標準準拠を維持し、国際標準に基づくデータ共有基盤を整備した

コネクター

X、YouTube、TikTok、ニュース記事などの各情報源からのデータをSTIX 2.1形式へ自動変換して取り込むコネクターを実装した。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

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を加工し、自動で分析結果や評価を付与することで、データの価値を向上させるための機能群である。エンリッチメントモジュールを実装し、センチメント・エモーション分析、エンティティ検知、アカウント真正性分析、埋め込み生成、アカウントプロファイリング、モーダル横断のメディア理解などを付加情報として付与できるようにした。加えて、分析需要を踏まえてAI生成コンテンツ検出機能も追加し、画像・動画のAI生成可能性を判定できるようにした

分析機構

分析機構として、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・可視化に加え、オーディエンスモデリング、拡散シミュレーション機能を実装し、偽・誤情報の拡散やオーディエンス反応を分析・予測するためのシミュレーション基盤を整備した

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

前述の各構成要素について、単一の処理に統合するのではなく、SNS投稿データの統合から分析結果の出力までが相互に連携・協調する形で、一貫して動作することを確認した。あわせて、STIX 2.1 準拠のデータ管理とカスタム拡張を前提とした運用を想定し、ユースケース適用を通じて実運用上の有効性と拡張性を検証した。

統合プラットフォームのデプロイを完了

OpenCTI、データ統合コネクタ、エンリッチメント機構、分析機構の4コンポーネントを統合したパイプラインとして稼働を確認した。機能ごとにモジュールが独立しており、要件に応じて組み合わせ可能な設計を採用することで、将来要件も見据えた拡張性と保守性を確保した

データ管理基盤

STIX 2.1に基づく統一管理を実現し、カスタムSTIX 2.1拡張（SnsPost、SnsAccount、Article）を実装した。標準準拠を維持しながらドメイン固有要件に対応できる拡張メカニズムを採用したことで、国際標準に沿ったデータ共有・連携を前提とした運用に対応可能な基盤を整備した

データ格納基盤

X、YouTube、TikTok、ニュース記事及びウェブサイトからのデータに対応し、データをSTIX 2.1形式へ自動変換して取り込む仕組みを実装した。その結果、データ取得網羅性、STIX 2.1への変換精度ともに実際の運用で利用できる水準を確保した

エンリッチメント機構及び分析機構

エンリッチメント6モジュールと分析4モジュールを開発し、統合を完了した。全エンリッチメントモジュールがベースラインKPIまたはターゲットKPIを達成し、各モジュールは独立に動作しつつ運用要件に応じて組み合わせる構成とした

実運用適用性の検証

3つのユースケースを用いて実運用適用性を実証し、本プラットフォームが実用的な分析基盤として機能することを確認した。統合パイプラインの稼働確認やSTIX 2.1準拠の管理・拡張、KPI達成、及びユースケース実証を通じて、実運用を見据えた分析基盤としての有効性を立証した

3-1. 技術開発の全体像

技術開発がユーザに与える価値

本開発技術を利用することにより最終的に得られる情報は**偽・誤情報及びファクトチェック情報の拡散シミュレーション結果**である。まず、シミュレーションのための前処理として、偽・誤情報の早期検知に有用な各種分析を実行する。次に、それらの分析結果を基に拡散シミュレーションを実行し、偽・誤情報拡散への対策立案を支援する。

情報拡散状況の把握

得られる情報	不自然な投稿の増加や、注目を集めた影響力のある投稿の特定	拡散される情報に含まれるトピックやナラティブの把握	投稿が特定の対象に対して示す感情の判定	人工的に操作されたボットアカウントの検出	AIによって生成された偽・誤情報の検出	偽・誤情報を拡散するアカウントの特徴把握
--------	------------------------------	---------------------------	---------------------	----------------------	---------------------	----------------------

拡散シミュレーション

活用方法	偽・誤情報の拡散を予測、シミュレーション	偽・誤情報の拡散と、これを打ち消すためのファクトチェック情報の拡散をシミュレーションし、その効果を比較	複数のファクトチェック情報の拡散をシミュレーションし、その効果を比較
得られる情報	日本のSNS環境における、特定の情報の拡散速度、拡散範囲、影響度、および影響を受け得るオーディエンスの予測結果		
得られる示唆	<ul style="list-style-type: none"> 各偽・誤情報に対して対応の必要性を判断 優先的に対処すべき偽・誤情報の判断 	<ul style="list-style-type: none"> ファクトチェック情報の有効性を評価 ファクトチェック情報の拡散によりかえって偽・誤情報の露出を拡大させてしまうリスクを評価 	<ul style="list-style-type: none"> ファクトチェック情報をいつ、どのように、どの主体から情報を発信した場合が最適かを判断

これらの仕組みにより、偽・誤情報の早期検知にとどまらず、その拡散や対抗するファクトチェック情報の拡散予測・シミュレーションを行い、対策立案を支援する。従来の多くのツールが偽・誤情報の検知にとどまっていたのに対し、本システムでは対策の検討まで支援することで、対策の実行を促進する

3-1. 技術開発の全体像

本開発技術の優位性

海外ツールには本開発技術と類似するものがいくつか存在する。以下では、本開発技術とそれらの海外ツールを比較し、本開発技術の優位性を整理する。

本開発技術と他ツールとの類似機能

複数プラットフォームのデータを集約・統合

感情や投稿量等の拡散傾向を可視化

ナラティブの検出と分析

ナラティブやネットワーク構造に基づく偽情報及びアクター分析

アクター間の関係性の視覚的表示

本開発技術の優位性

拡散予測

類似技術

ネットワークの可視化や偽情報の量を追跡することに焦点を当てているが、それらの情報の拡散を予測する能力は限定的



本開発技術

ネットワーク構造やアカウントの性質に基づく情報の拡散予測を実行することが可能

統合や拡張の柔軟性

類似技術

独自のプラットフォームを通じて分析を提供しており、統合や拡張の柔軟性は限定的



本開発技術

OpenCTIとSTIX2.1に基づくモジュール形式のアーキテクチャにより、分析の拡張や他システムとのシームレスな統合が可能

日本語分析能力

類似技術

主に英語や国際市場向けに最適化されているため、日本語のナラティブや文化的背景に対する性能に留保が必要



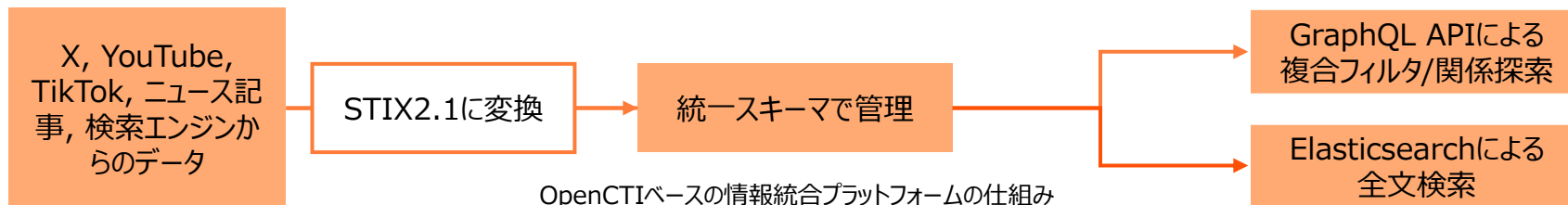
本開発技術

日本語話者の知見と日本における情報作戦の専門的知識に基づき、日本特有のナラティブや構文への対応が可能

3-2. 技術開発の個別詳細

OpenCTIベースの情報統合プラットフォーム

OpenCTIを中核とした情報統合基盤を構築した。SNS投稿・アカウント・ニュース記事・ウェブサイト等のオープンソースデータについて、抽出・標準化・管理・検索・関係性分析といった処理を一貫して提供する。OpenCTIを軸に、STIX2.1の拡張メカニズムを用いてSNS投稿やアカウントなどを表す独自のデータ形式であるカスタムドメインオブジェクトを実装した。

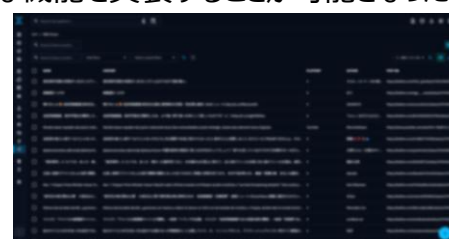


これにより、標準準拠を維持しつつドメイン要件を満たす統合基盤を実現し、投稿—アカウント関係、引用関係、拡散の時系列などをグラフ形式で可視化できるようにした。さらに、データの入出力やエンリッチメント等はコネクタ方式で実装した。これらの機能は本体と独立性の高い状態で連携させることで、システム本体を改修することなく、新規データソースの追加や分析手法の導入を継続的に実施できる構造とした。

プラットフォームは、オープンソースソフトウェアであるOpenCTIのデータモデル、GraphQLスキーマ、フロントエンド表示の一部を拡張することで実装した。STIX形式を使用することにより、改変範囲を最小限としつつ必要な機能を実装することが可能となった。



OpenCTI上で表示される分析結果のダッシュボード



OpenCTI上で表示されるSNS投稿の一覧

ユーザへの提供価値

OpenCTIを基盤とし、STIX形式で情報を構造化して統合することにより、他ツールとの連携や偽・誤情報の評価基準の統一を可能とした。統合したデータには、統計情報や後述する各種エンリッチメントおよび分析結果に基づいた可視化が行われる。また、データ、分析機能、可視化機能を追加しやすいアーキテクチャを採用することで、ユーザのニーズに基づいてSNSや分析技術の追加・更新に柔軟に対応できる構造とした

3-2. 技術開発の個別詳細

OpenCTI・STIX形式の概要・メリット

本プラットフォームは、OpenCTIを中核とする情報統合基盤として、偽・誤情報に関するあらゆる非構造化データをSTIX形式で一元管理する。以下ではOpenCTI・STIX形式の概要とメリットについて整理する。

OpenCTI

概要

フランスFiligran社開発のオープンソース脅威インテリジェンスプラットフォーム（TIP）であり、偽・誤情報に関する情報共有の用途でも欧州・米国でデファクトスタンダードとなりつつある

OpenCTI上での実装のメリット

各種のデータを構造化し格納する情報統合能力

FIMIを始めとする偽・誤情報分野における情報共有のデファクトスタンダードへの準拠

連携する諸外国や国際機関との間においても安全かつ標準化された方法でデータを共有することが可能

STIX形式

概要

STIX（Structured Threat Information eXpression、脅威情報構造化記述形式）は、サイバー脅威情報を共有するための標準化言語

STIX形式採用のメリット

オブジェクト間の関連付けにより、SNSにおける投稿やそれを行ったアカウント、またアカウント間等の関係性の統一的表現が可能となる

STIX 2.1形式を独自にカスタムすることで、STIX形式での情報共有の効率性を検証、さらに偽・誤情報の分析に有用な要素を実装し、将来的なSTIX形式への採択を目指すことも可能となる

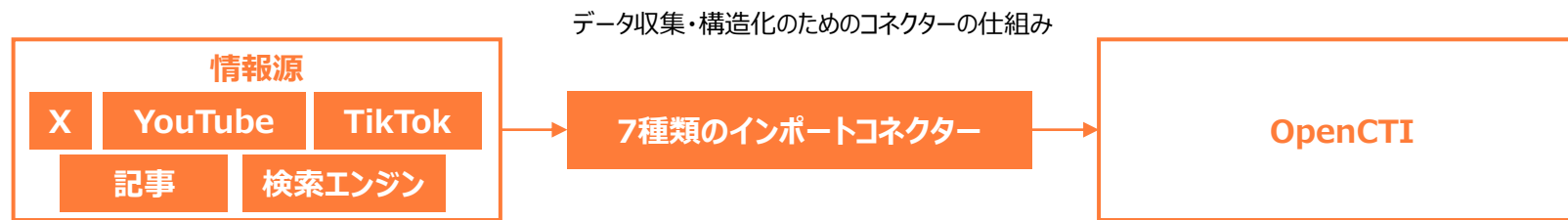
	SNS1	SNS2	SNS3
"author_handle"	"@watcher2025"	"truthmediajp"	"@deepinfo2025"
"text"	"@officialgov 某国が秘密兵器を使っているという噂が... #真実を暴け"	"【衝撃】世界の裏で何が起きているのか？ #マスメディアが報道しない真実を暴露。ソースはこちら→ https://unknownsource.org"	"テレビでは絶対に隠れない事実を暴露... #目を覚ませ #メディア操作"
"posted_time"	"2025-06-24T00:00:00Z"	"2025-06-20T00:00:00Z"	"2025-06-22T00:00:00Z"
"engagement"	195 (like:150, share:20, quote:15, comment:10)	6692 (like:5800, share:-, quote:-, comment:892)	15720 (like:12000, share:3050, quote:-, comment:670)
"hashtags"	"[#真実を暴け]"	"[#報道されない真実", "#暗諷論]"	"[#目を覚ませ", "#メディア操作]"
"mentions"	"[@officialgov]"	"[https://unknownsource.org]"	"[@truthagency]"
"media"	"[media/G13y9p/format=jpg]"	"[sns2.com/424829016/video]"	"[sns3.com/video/987zyx654]"

STIX形式のデータ構造正規化の模式図

3-2. 技術開発の個別詳細

データ統合・構造化のためのコネクタグループ

- 主要SNSプラットフォーム、ニュースサイト及び検索エンジンからのマルチプラットフォーム情報統合を目的として、インポートコネクタ群を開発・実装し、リアルタイム/準リアルタイムでデータを取込可能な統合基盤を整備した。
- X, YouTube, TikTok, ニュース記事、ウェブ検索結果等の主要なオープンソースデータについて、そのデータ抽出・変換のためのコネクタ機構をそれぞれ作成し、上記情報源に含まれる非構造化データをSTIX形式に変換することで、複数プラットフォームのデータを構造化し一元的に管理できる。



取組

- エラーへの対応を強化し、処理を自動でやり直す仕組みを導入することで、システムの動作を安定させた。また、データの確認方法を見直すことで、データの抜けや不一致を防いだ
- コネクタ方式の仕組みを採用することで、それぞれのプラットフォームの特徴に合わせた処理を実施可能とした。また、将来新しいプラットフォームを追加しやすい柔軟な構造とした

成果

- 複数のプラットフォームのデータを一元的に管理できるようにした。また、並列処理や接続管理、バッチ処理の効率化により、大量のデータを効率的に処理できるようにした
- 検証の結果、データ取得網羅性97.16%を達成し、実運用に耐える水準を確保した。STIX 2.1への変換精度は100%であり正確に標準形式へ変換できることを立証した

ユーザへの提供価値

結果として、複数プラットフォームに分散する情報を高い網羅性と信頼性を保持した状態で継続的に統合し、STIX2.1準拠の標準化データとして即時に扱えるようになったことで、後段の分析・可視化の立ち上げを速めるとともに、他ツールや他基盤との連携を前提とした運用が可能となった

3-2. 技術開発の個別詳細

エンリッチメント機構：センチメント・エモーション分析

- センチメント・エモーション分析は、エンリッチメント機構に含まれるモジュールの一つである。
- 本開発技術におけるセンチメント分析は、文章全体に対してではなく、特定のエンティティに対して投稿者が抱く感情がポジティブであるか、あるいはネガティブであるかといった感情傾向を自動的に判定するものである。
- エモーション分析では、分析対象となる投稿によって喚起される怒りや悲しみなどの多様なエモーションを把握することが可能である。
- 分析結果は確信度スコアとともに出力され、分析結果の信頼性の把握や、追加検証の必要性を判断する際の指標となる。これらの分析により、情報作戦の一環として行われ得る、人々の感情や心理に影響を与えようとする試みを検知することが可能である。

取組

- 精度と安定した運用の両方を実現するため、日本語特有の文字や句読点、文章の構造を考慮した前処理を設計した
- SNS特有の絵文字表現やURLの整理、句読点の扱いを調整することで、分析の精度を高めた

成果

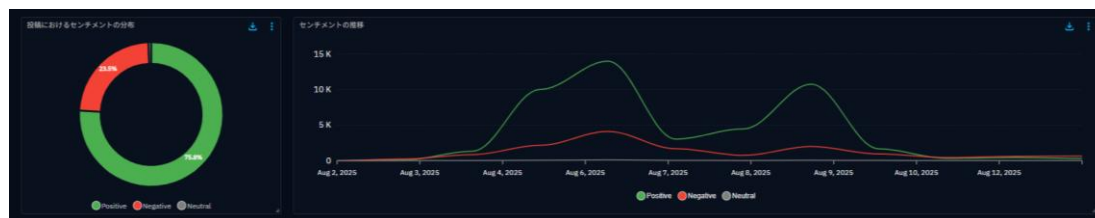
- 3値分類においてF1スコア (P55参照) 0.675を達成し実運用上必要な基準水準を満たした
- 日本語テキストに最適化された感情分析モデルをOpenCTIに統合し、投稿単位で一貫した分析結果を付与できるようになったことで、感情傾向を継続的に蓄積・検索・集計できる基盤が整備された

ユーザへの提供価値

- 日本語データにおける感情の定量化を実用精度で担保し、話題や事象に対する社会的反応の把握、偽・誤情報拡散時における反応傾向の分析などに活用可能となった
- 投稿に分析属性が付与されることで、データ整理・可視化に貢献し、異常な反応の兆候の早期検知が可能となる



UI上で表示されるセンチメント分析の結果例



ダッシュボード上で表示される投稿におけるセンチメントの推移グラフ例

3-2. 技術開発の個別詳細

エンリッチメント機構：SNSアカウント真正性分析

- SNSアカウントに対して実在する主体による運用の可能性、すなわち真正性を評価し、自動投稿や拡散を行うボットアカウントと人間の識別を支援するため、「SNSアカウント真正性分析」エンリッチメント機構を開発・実装した。アカウントの属性情報やメタデータを総合的に分析し、機械学習ベースの分類によって真正性を定量的に評価できるようにすることで、後段の拡散分析や調査の前処理を効率化することを目的とした。
- 本機構は、OpenCTIに取り込まれたSNSアカウントに対して自動的に推論を実行し、判定結果と確信度スコアをエンティティ属性として付与する。分析モデルを用い、1アカウントあたり23の特徴量を入力として、意味的・行動的指標を組み合わせた推論結果を出力する。

取組

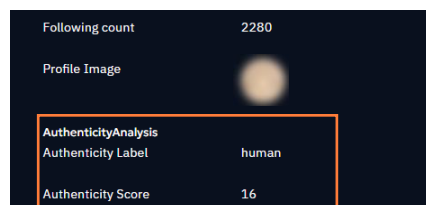
- 高精度化のため、多面的な属性を特徴量として設計し、アカウントの意味的な活動とその振る舞い等に関する指標を併用することで、アカウント特性を偏りなく捉える構成とした
- 学習済みモデルをプラットフォーム内で自動実行できる形に統合し、判定結果に確信度を付与することで、運用上の判断に活用可能とした

成果

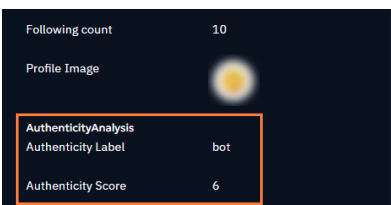
- 二値分類においてF1スコア0.9716を達成し、技術の潜在能力を最大限に発揮した場合の目標水準を満たした
- ボットアカウントと人間アカウントを高精度に区別できる自動エンリッチメント基盤が整備され、アカウント評価を継続的に蓄積・検索・活用することができる状態となった

ユーザへの提供価値

ボットである可能性が高いアカウントを効率的に発見できるようになり、判定の確信度に基づいて、すぐに対応すべきものと追加で確認すべきものを区別可能とした。これにより、初期対応の迅速化が期待できる。また、情報の拡散にどのような影響があるかの評価や、ボット群の特徴の分析を通じて、信頼性の低い主体による情報拡散の状況を把握できるようになった



UI上で表示されるアカウントの真正性分析の結果例



ダッシュボード上で表示されるアカウントの真正性の内訳グラフ例

3-2. 技術開発の個別詳細

エンリッチメント機構：埋め込み表現の取得

- OpenCTI内のデータの意味を数値として扱えるようにし、類似投稿の検索やグループ分けなどの分析を実行できるようにすることを目的として、投稿などの文章から「埋め込み表現」という数値データを作成し、OpenCTIのデータ属性として追加する「埋め込み表現の取得」エンリッチメント機構を開発・実装した。対象はOpenCTIに取り込まれた投稿や記事などの文章であり、単純な文字の一致だけに頼らず、意味の近さに基づいて関連性を判断でき、後段の分析に活用できる状態を目指した。
- 作成される埋め込みデータは384次元の数値ベクトルであり、保存する際に正規化処理を行い、コサイン類似度（文章などのデータを数値ベクトルに変換し、その向きの方の近さから内容の類似度を測定する手法）による比較に適した形式で保持する。これにより、投稿同士の類似度の計算や、その後の検索・グループ分けなどの処理にそのまま利用できるようにした

取組

- 実際の運用や将来の機能追加を考え、モデルの初期化から結果の反映までの処理を機構内で完結できるようにした
- また、保存するデータは、コサイン類似度による高速な計算ができるように、あらかじめ正規化したベクトル形式とした
- 評価では、検索結果の上位の質や、グループ分けの正確さと網羅性を重視し、実際の運用に耐えられる性能であることを確認する方針を採用した

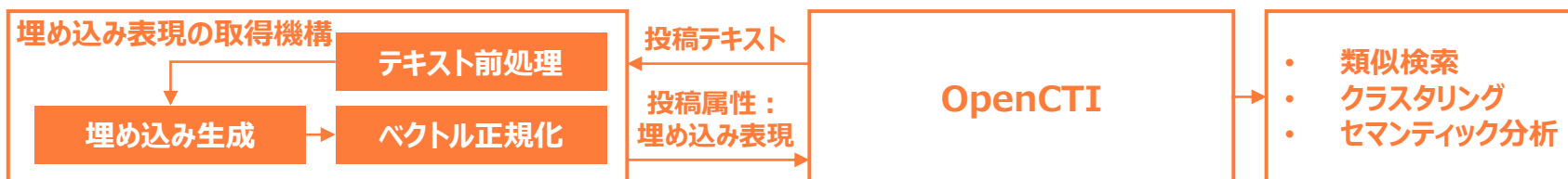
成果

- 埋め込み品質を検証した結果、nDCG@10 (P55参照) は0.4498、クラスタ品質はV-measure(avg) (P55参照) 0.5949を達成した。いずれも技術の潜在力を最大限発揮した場合の目標水準を満たした
- 埋め込み生成と正規化までを実装してOpenCTI上に保存可能としたことで、検索・クラスタリング機能の追加実装に必要な前提基盤を整備した

ユーザへの提供価値

ユーザはキーワードや表記揺れに左右されず、意味的に近い投稿群を抽出できるようになり、話題のまとめや関連投稿の発見が容易になる。埋め込みベクトルはOpenCTI上に蓄積されるため、将来的な類似検索・クラスタリング機能の実装や外部ツールでの分析に直接活用でき、偽情報・誤情報の拡散パターン分析における類似投稿探索とクラスタ発見の効率化への寄与が期待できる

埋め込み表現の取得



3-2. 技術開発の個別詳細

エンリッチメント機構：エンティティ検知

- データ内の文章から人物・組織・場所などのエンティティ・固有表現を自動抽出し、投稿・記事等の内容を構造化して扱えるようにすることを目的として、「エンティティ検知」エンリッチメント機構を開発・実装した。対象はOpenCTIに取り込まれたSNS投稿・ニュース記事・ウェブサイト等であり、複数言語に対応しつつ、人物、組織、場所、イベントなど多様な固有表現を抽出する。
- 固有表現を抽出するために、異なる3種類のモデルをOpenCTIに組み込み、初期化、文章の前処理、固有表現の抽出、結果の整理までを一連の処理として実装した。さらに、あらかじめ設定した優先順位に基づいて自動でモデルを選択し、使用できない場合は別のモデルに切り替えることで、処理を継続できる仕組みとした。

取組

- 実際の運用でも安定して使えるように、前処理ではデータの形式を整える正規化を行い、後処理では判定の確信度を示すスコアを計算するとともに、重複する結果を取り除いて分析結果の精度を高めた
- 単一モデルに依存しないマルチモデル方式とし、性能と安定した利用を両立させるためにモデルの自動選択と利用できない場合に別のモデルへ切り替える仕組みを設計に組み込んだ

成果

- ノイズを含むSNS投稿を対象とした評価においてF1スコア0.5764を達成し、実用水準として設定した基準値を満たした
- 整形済みテキストに限らず、URLや表記揺れ、不要文字が混在する投稿に対しても抽出を継続できるため、運用環境で固有表現を安定して付与できる基盤が整備された

ユーザへの提供価値

投稿・ニュース記事等の文章内の主体や対象を構造化して検索・集計できるようになり、話題に関与する人物・組織等の抽出を起点に調査を進めることが可能となる。偽・誤情報の拡散局面でも、関与し得る主体の特定や関連投稿の整理が効率化され、分析の初動と網羅性の向上に寄与する



UI上で表示される結果例



ダッシュボード上で表示される最も多く言及されたエンティティのランキング例

3-2. 技術開発の個別詳細

エンリッチメント機構：モーダル横断メディア理解

- SNS投稿に含まれる文章、画像、動画をまとめて分析し、文章だけ、画像だけでは分かりにくい文脈も含めて理解できるようにすることを目的として、「モーダル横断メディア理解」エンリッチメント機構を開発・実装した。対象はOpenCTIに取り込まれたSNS投稿であり、投稿本文に加えて添付された画像や動画も分析し、その内容をまとめて蓄積できるようにした。
- 文章・画像・動画をまとめて分析できるモデルをOpenCTIに組み込み、投稿ごとに自動で内容を分析することで、画像の分析結果、動画の分析結果、そしてそれらをまとめた分析結果を投稿の属性情報として追加する。画像については、画像内の文字を読み取る処理、写っている物の判別、場面の認識を行い、画像内の文字、主な物体、どのような場面かを出力する。動画については、時間の流れに沿って内容を分析し、重要な点を抽出して、内容の変化と重要なポイントを整理して出力する。さらに、投稿の文章と画像・動画の情報をまとめて解釈する処理を組み込み、文章と画像・動画がどのように関係しているかも含めて評価できるようにした。

取組

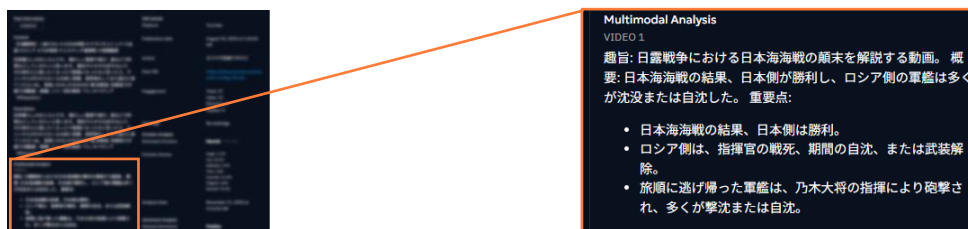
- 運用の安定性と処理効率を両立するため、画像・動画は特性に応じて処理系を分けつつ、最終的にLLMで統合解釈する構成とした
- 動画長に応じたフレームレート調整により処理時間と精度のバランスを確保した

成果

- テキスト生成品質では、BLEUスコア(P55参照) 0.387を達成した。物体・シーン認識では、mAP (P55参照) 0.772を達成し、いずれも実運用水準を満たした
- プラットフォームごとに異なる処理フローを確立し、様々な動画データを安定して解析できる基盤を整備した

ユーザへの提供価値

テキストだけでは判断できない画像・動画の内容を含めて投稿を包括的に把握でき、偽・誤情報に関与し得るメディアを効率的に抽出できるようになった。投稿本文とメディアの整合性や示唆内容を横断的に確認できるため、投稿の含意の理解を迅速化しつつ、見落としや誤解釈のリスクを低減する効果が期待できる



UI上で表示されるモーダル横断メディア理解の結果例

3-2. 技術開発の個別詳細

エンリッチメント機構：SNSアカウントプロファイリング

- SNSアカウントの属性把握を効率化することを目的に、OpenCTI上のSNSアカウントを対象として、「SNSアカウントプロファイリング」エンリッチメント機構を開発・実装した。本モジュールは直近1か月の投稿とプロフィール情報を併用し、設定されたカテゴリに基づく特性の推定を行う。アカウントの投稿を元にアカウント属性の推定を実行することで、偽・誤情報の拡散に関連すると想定される意味的・振舞い特徴への合致をカテゴリ別に出力する。
- 本機能により、単一の投稿のみではなく、アカウント情報を含む包括的な情報に基づいて、偽・誤情報拡散に関連した特性の把握が可能となる。また、カテゴリ別に一定の基準に基づいたアカウント間の比較や整理も可能となる。

取組

- カテゴリ・ラベルについては先行研究や自社における分析事例を元に定義し、カテゴリ別推定により出力の粒度や妥当性、一貫性を確保しつつ、客観的な評価を担保した

成果

- 専門家評価1.0を達成し、専門家評価とシステム評価との高い整合性を確認した
- 投稿内容とプロフィールの双方を統合して推定する方式により、アカウント特性を実用的な精度で把握できる状態を実現し、分析結果をOpenCTI上の属性として継続的に蓄積・検索可能とした

ユーザへの提供価値

ユーザーはアカウントの行動特性を短時間で把握でき、偽・誤情報に関する調査対象の選別や優先度付けを迅速に行うことができる。手作業での読解や整理を減少させ、ユーザーごとの振れ幅を抑制し、スケーラブルな分析を可能とする。



UI上で表示されるアカウントプロファイリング分析の結果例

3-2. 技術開発の個別詳細

エンリッチメント機構：AI生成コンテンツ検出

- SNS投稿に含まれる画像・動画がAI生成コンテンツかを自動判定し、偽・誤情報評価に活用するため、モーダル横断メディア理解機構を応用し「AI生成コンテンツ検出」エンリッチメントを実装した。OpenCTIに取り込まれた投稿に添付された画像や動画を対象に、OCR（画像や写真の中にある文字を読み取り、テキストデータに変換する技術）による文字の読み取りや画像・動画の分析を行い、その結果からAI生成の可能性を判定して付与できるようにした。
- 画像と動画それぞれに適した評価基準を設け、重み付きのスコアと判断の基準値を設定したうえで、AI生成の可能性を判定するモデルを構築した。結果は、AI生成フラグおよび3段階の確信度に加え、総合スコアと観点別スコアを記録することで、判定の根拠や妥当性を明確化し、さらなる分析への展開を容易にする設計とした。

取組

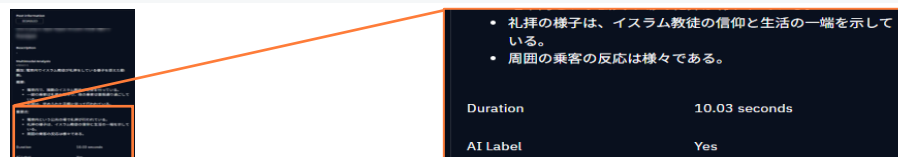
- 実装上は、画像・動画で評価観点と重み付けを分けた構造化評価とし、総合スコアに基づく閾値判定で運用可能な判断基準として整備した
- 評価では、ウォーターマークなどの出所を示す表示の有無、画像の不自然な部分や整合性、物理的に不自然な表現、編集の不自然さなどを組み合わせて判断する。また、判定結果とあわせて、その判断の根拠となる情報も保存する。
- モーダル横断メディア理解の機構に改変を加えることで、費用への影響を最小限にしつつ、プロジェクト計画予定外の追加機能を実装することができた

成果

- AI生成コンテンツ検出をモーダル横断メディア理解機構に統合し、画像・動画を対象とした自動判定と結果保存を実現した
- 検出精度は、F1スコア0.7890を達成した。特にロゴを含むコンテンツに限定した場合はF1スコア0.9947に達し、実用上十分な判定性能を確認した

ユーザへの提供価値

ユーザは投稿に含まれる画像・動画のAI生成可能性を自動で判定でき、AI生成コンテンツが拡散に与える影響を定量的に把握することが容易となる。動画内容を併せて確認できるため、内容に内在するリスクや判定の妥当性を検証しながら、優先対応や追加調査の判断を効率化できる



UI上で表示されるAI生成コンテンツ検出の結果例

3-2. 技術開発の個別詳細

分析機構：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

- OpenCTI内のデータから主要トピック、ナラティブ、主張を自動抽出し、拡散のネットワーク分析と時系列分析までを一体化して把握することを目的に、「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」システムを開発・実装し、文章の内容の解釈と拡散挙動を同じ枠組みで追跡する事を可能とした。
- トピックモデリング/クラスタリングにより主要トピックとナラティブを階層的（トピック > ナラティブ > 主張）に生成し、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。分析結果はOpenCTI上の可視化に加えて、独自開発の外部レポートビューアでも探索できる形で出力する。

取組

- 話題抽出に留めず、メッセージング戦略やコミュニケーション手法まで解釈するようモデルを設計し、主張抽出とナラティブ割当を一連の処理として組み込んだ
- OpenCTI統合に加えて外部ビューアを用意することで、探索性と説明性の両立を図った

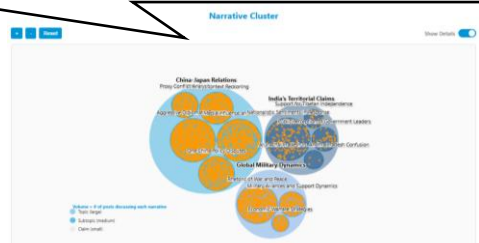
成果

- 分析結果を階層構造として整理し、内容面と拡散面を同時に追跡できる基盤を整備した
- ナラティブや主張の構造を可視化し、規模や主張密度の把握、その他詳細確認を可能にした
- 投稿時刻に基づくナラティブの拡散を時系列的に可視化することが可能となっている

ユーザへの提供価値

ユーザはナラティブ構造と拡散の推移を同一画面で追跡でき、重要トピックや主張群、拡散の転換点を素早く把握できる。これにより、偽情報・誤情報の拡散状況の整理とレポートングを効率化する

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变化で表される



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

3-2. 技術開発の個別詳細

分析機構：オーディエンス・モデリング・拡散シミュレーション

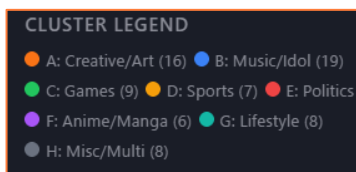
- SNS上における拡散経路の可視化と実データとの検証までを一体的に扱うことを目的に、分布シミュレーション基盤を開発・実装した。
- SNS環境におけるオーディエンスを興味・関心に基づいてグルーピングし、その上で各グループがどのような情報に影響を受けやすいかをモデル化した。情報の発信元となるグループ、発信メッセージ、シミュレーションの期間、何段階先の拡散までシミュレーションするかなどユーザが設定した条件に基づいた拡散シミュレーションを実行できる形とした。
- 設定した発信メッセージが、時間の経過とともにどのオーディエンスにどの程度受容されるか、どのような経路・速度で拡散するかを予測する。偽・誤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程度の抑制効果を持ち得るかを評価できる。

取組

- 日本のSNS環境を評価し、各グループ内外の情報流通を観測することで、環境のモデリングを実施した
- 各グループの挙動にはエンリッチメント機構の各分析結果が反映され、より実際に近い振る舞いをシミュレーションすることを目指した

成果

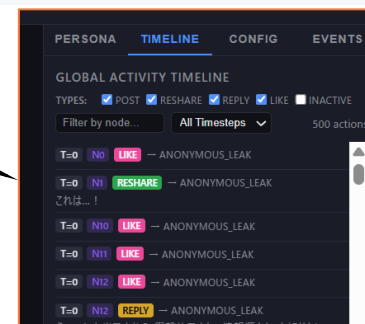
- 拡散シミュレーションでは、条件を指定して即時にシミュレーションを実行し、各グループがその時間に行くと想定される活動をその根拠とともに確認できる
- 偽・誤情報の拡散およびその訂正情報の拡散をシミュレートし、訂正情報の拡散を最大化する設定の探索を可能とする



カテゴリによって色分けしたクラスタ間の情報拡散経路を矢印で可視化

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

GUI上で表示される拡散シミュレーション



ユーザへの提供価値

分析者は発信クラスタとメッセージを指定するだけで多段の拡散をシミュレートでき、その規模や転換点を把握できる。偽・誤情報の拡散を打ち消すための訂正情報を、いつ、どのグループから、どのように流せば訂正情報の流通が最大化するかを早期に把握することができ、偽・誤情報を打ち消すための行動指針や計画を策定し実施する事に資する

3-2. 技術開発の個別詳細

分析機構：自動分析・アラート機能

- 各分析結果に基づく重要な変化を検知し通知するためのアラート機能をOpenCTI上に実装している。ユーザーは個別のトリガーを設定することで、条件に合致した場合自動的にそれを知ることができる。
- 本機能は、OpenCTIに格納されているすべてのデータに対して、設定されたトリガーに基づいて各種のデータ作成・更新・削除を検知し、その状態変更をOpenCTIのGUI上に通知する。

取組





- 本機能は、OpenCTI既存のアラート機能を基盤として実装した
- 今後は、既存のGUI上の通知に加えて、ユーザーのメールでも受信できるよう通知経路を拡張する予定である
- OpenCTIの標準機能では対応できない、より詳細な条件（例：特定のセンチメントを含む投稿数が閾値を超えた場合）の通知の実装に向けて、機能拡張に取り組んでいる

成果





- 各種分析に関する通知をプラットフォーム上に実装することで、その開始・終了や変化などをリアルタイムに追跡することができる

ユーザへの提供価値

自動通知・アラートにより、投稿・記事・アカウントおよび分析結果の有意な変化を早期に把握可能とし、迅速な意思決定および対応に資することで、偽・誤情報の拡散や風評被害等による事業リスクの最小化を図る

TYPE	NAME	NOTIFICATION	TRIGGERING ON
 LIVE TRIGGER	Article	User interface	<button>create</button> <button>update</button> <button>delete</button>
 LIVE TRIGGER	Entitues are mentioned	User interface	<button>create</button> <button>update</button> <button>delete</button>
 LIVE TRIGGER	SNS Account change	User interface	<button>create</button> <button>update</button> <button>delete</button>
 LIVE TRIGGER	SNS Post change	-	<button>create</button> <button>update</button>

OpenCTI上のアラート設定画面

OPERATION
 MODIFICATI...
 MODIFICATI...
 MODIFICATI...
 MODIFICATI...

OpenCTIのGUI上に表示される通知の例

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

本取組の目的は、プロジェクトの進捗と成果の現実的な評価のため、技術的KPIとビジネスKPIを設けて開発技術の性能及び実務上の有効性を適切に検証することである。

KPI一覧

カテゴリ	目的	モジュール	KPI名	評価の段階
技術的KPI	開発技術の性能を定量的に評価する	コネクター機構	①データ取得網羅性	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">ベースラインKPI</div> 実用上の最低限の価値を 保証する
			②STIX形式への変換における構造的正確性	
			③モーダル横断メディア理解	
		エンリッチメント機構	④AI生成コンテンツ検出	
			⑤エンティティ検知	
			⑥センチメント・エモーション分析	
			⑦埋め込み表現の取得	
			⑧アカウント真正性分析	
			⑨アカウントプロファイリング	
			⑩トピック分析・ナラティブ検知・主張抽出	
		分析機構	⑪拡散傾向の分析・可視化	
			⑫オーディエンスモデリング	
			⑬拡散シミュレーション	
ビジネスKPI	実務における導入効果を測定する	-	⑭-⑰各工程(全4工程)それぞれにおける工数削減率	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">ターゲットKPI</div> 技術の潜在力を最大限発 揮した際の評価となる

技術的KPI達成度別割合

- 技術的KPIについては13項目を設定した。13項目すべてにおいてベースラインKPIを達成し、そのうち9項目(69%)がターゲットKPIを達成した。



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

一部のKPI計測において用いた評価指標とその説明は以下の表の通りである。⑦埋め込み表現の取得においては、より適切な評価のため、当初目標としていたPrecision@10とNMIによる評価からnDCG@10とV-measureによる評価へと変更を行った。

評価指標	モジュール/KPI	説明
BLEU	エンリッチメント/③モーダル横断メディア理解	機械翻訳の出力が人間による参照訳とどの程度一致しているかを単語の一致度に基づいて評価
mAP	エンリッチメント/③モーダル横断メディア理解	画像や動画内の対象物を検出する際の正確性と見落としの少なさを総合的に評価
F1	エンリッチメント/複数のKPI	誤検知の少なさと見逃しの少なさの双方をバランスよく満たしているかを評価
Precision@10	エンリッチメント/⑦埋め込み表現の取得	検索結果上位の中に実際に関連性の高い結果がどの程度含まれているかを評価
nDCG@10	エンリッチメント/⑦埋め込み表現の取得	検索結果の関連性だけではなく、重要度の高い結果が上位に配置されているかを評価
NMI	エンリッチメント/⑦埋め込み表現の取得	データを複数のグループに分類した際、その分類結果が本来の分類構造とどの程度一致しているかを評価
V-measure	エンリッチメント/⑦埋め込み表現の取得	クラスタリング結果において、同一種類のデータが適切にまとまっているか、同一グループに異なる種類のデータが混在していないかを総合的に評価
MCC	エンリッチメント/⑧アカウント真正性分析	判定結果が偶然による一致ではなく、ラベルを適切に識別できているかを単一の指標で総合的に評価
専門家評価	エンリッチメント/⑨アカウントプロファイリング	技術開発主体の複数アナリストによるプロフィール推定結果との一致度を評価

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

一部のKPI計測において用いた評価指標とその説明は以下の表の通りである。

評価指標	モジュール/KPI	説明
NPMI	分析/⑩トピック分析・ナラティブ検知・主張抽出	同一トピックに含まれる単語が実際の文章中で一緒に出現するかを測定しトピックの一貫性を評価
専門家評価	分析/⑪拡散傾向の分析・可視化	技術開発主体の複数アナリストによる主要拡散経路・拡散者の特定結果との一致度を評価
ピアソン相関係数	分析/⑫オーディエンスモデリング	さまざまな話題における「実際の反応数」と「シミュレーション環境における反応数」の傾向の相関関係を評価
予測誤差	分析/⑬拡散シミュレーション	ある話題が実際にどのアカウント群へ広がったかを、シミュレーションによる予測結果と実データとで比較した場合の誤差率を評価

4-2. 検証及び調査の個別詳細

KPIとその検証方法の詳細

前項4-1で設定した各KPIを評価するため、以下の手法を用いて検証を実施した。検証プロセスにおいては技術開発主体がすでに所持する過去の偽・誤情報関連事例の分析データ及び専門的知見を基盤としつつ、主要なベンチマークデータセットや本事業期間中に本開発技術を使用した協力主体により提供されるデータ及び分析結果を適宜活用し検証を行った。KPI名と具体的な検証方法の詳細は以下の表の通りである。

KPI名	具体的な検証方法
①データ取得網羅性	データ取得対象となるプラットフォームAPI仕様やクローリング対象サイトの構造と、実際にシステムが取得したデータのログの照合からデータ取得漏れの割合を算出した
②STIX形式への変換における構造的正確性	取得データをSTIX形式へ変換した結果からサンプルを抽出し、事前に定義したデータモデルとSTIX形式間のマッピングを専門家が目視で確認した
③モーダル横断メディア理解	画像・映像からテキストを生成する機能については人手で作成した正解キャプションとシステムにより生成されたキャプションを比較し言語生成指標で評価した。画像・映像内の物体・シーンを認識する機能については、正解ラベルの付与されたデータセットを用いて平均適合率(mAP)を測定した
④AI生成コンテンツ検出	モーダルメディア理解機構に追加実装されたAIモデルを用いて投稿に添付された画像・動画を分析しAI生成の正否を分析。AIによる評価とアナリストが付与した正解ラベルの一致度をF1スコアで算出した
⑤エンティティ検知	構造化されたクリーンなデータセット(例：ニュース記事)とノイズが多いデータセット(例：SNS投稿)双方を用意し、どちらも人手に用意した正解データとの比較を行い、F1スコアを算出した
⑥センチメント・エモーション分析	一般的な感情と偽・誤情報に特有な文脈での感情の分析精度を標準的な感情分析コーパスと技術開発主体独自のデータセットのそれぞれを用いて検証を行った
⑦埋め込み表現の取得	下流タスク性能により間接評価を行った

4-2. 検証及び調査の個別詳細

KPIとその検証方法の詳細

KPI名	具体的な検証方法
⑧アカウント真正分析	過去の事例で特定された非真性的なアカウント及び正規のアカウントを用いて識別精度を検証した。検証に際してユースケースに応じた適合率(無関係なアカウントの誤検知を行わないこと)と再現率(対象の不審なアカウントを見逃さないこと)のバランスを調整した
⑨アカウントプロファイリング	アカウントの政治的傾向や興味関心のような性質の主観性を考慮するため、技術開発主体のアナリストが有する知見に基づき複数のアナリスト間でプロフィール推定結果の一致度を評価した。また傾向が既知のアカウント群に対する推定精度も検証した
⑩トピック分析・ナラティブ検知・主張抽出	システムにより自動的に抽出されたトピックやナラティブの品質を自動評価指標(NPMI、正規化点相互情報量)と専門家による人手指標の両面から検証を行った。人手評価においては抽出されたトピックの一貫性と偽・誤情報の文脈との整合性の評価の一致率を確認した
⑪拡散傾向の分析・可視化	過去の事例をケーススタディとして本開発技術を用いて分析を行い、主要拡散経路や影響力のある拡散者の再現や特定の精度を検証した
⑫オーディエンスモデリング	本事業期間中に発生した情報拡散事例における特定アカウント群による特定情報へのエンゲージメントの実現値とシステムによる予測値との相関係数を計測した
⑬拡散シミュレーション	過去の情報拡散事例のデータを用いてシミュレーションモデルによる拡散規模の予測の誤差をバックテストで検証した
⑭-⑰工数削減率	まず過去に対応した偽・誤情報事例から複数のケーススタディを選定し、当時のアナリストが要した作業時間(人時)を基準とする手動分析ベンチマークを設定した。次に、これらのケーススタディや類似の新規事案に対し、本開発技術を用いて同様の分析を実行し、アナリストが要する実作業時間を計測した。最後に計測結果を以下の式に代入し工数削減率を算出した $\text{工数削減率(\%)} = ((\text{手動分析における標準工数} - \text{本システム利用時の実作業工数}) / \text{手動分析における標準工数}) \times 100$

4-2. 検証及び調査の個別詳細

コネクタ機構KPIとそのステータス

本取組の目的は分析の基盤である入力データについてその品質と網羅性を担保することである。①データ取得網羅性についてはベースラインKPIを達成し、②STIX形式への変換における構造的正確性についてはターゲットKPIを達成した。これらの成果の達成のために技術開発主体が過去に培った高トラフィック環境下における高効率かつ高確度な情報処理技術の知見の応用や、ソーシャルメディア等から生成される大量の非構造的・非標準的データに対して、適切なSTIXカスタムオブジェクトを独自に定義し、STIX形式への構造化に取り組んだ。

KPI達成状況及びスコアの詳細

番号	KPI	ベースラインKPI	ターゲットKPI	達成スコア	ステータス
①	データ取得網羅性	0.95	0.99	0.9716	ベースライン達成
②	STIX形式への変換における構造的正確性	0.95	0.99	1.00	ターゲット達成

- ①データ取得網羅性の検証において、開発環境における通信状況や使用機材の処理能力の制約に起因して、一部処理が途中で停止し、データ取得に欠落が生じた事例が確認された。一方で、通常の収集処理においては概ね安定して動作しており、大部分のデータは問題なく取得できていることを確認し、総合的にベースラインKPIを達成している。
- ②STIX形式への変換における構造的正確性においては情報源とのデータ構造のマッピングを徹底しており、取り込まれたデータのすべてが正確にSTIX形式に変換されていることを確認し、ターゲットKPIを上回る性能を示した。

KPI検証過程



4-2. 検証及び調査の個別詳細

エンリッチメント機構KPIとそのステータス

本取組では、収集したデータから偽・誤情報の拡散状況を的確に把握し、分析結果の信頼性を高めることを目的として、各分析モジュールの精度向上と性能評価に取り組んだ。

KPI達成状況及びスコアの詳細

番号	KPI	ベースラインKPI	ターゲットKPI	達成スコア	ステータス
③	モーダル横断メディア理解	BLEU 0.2 mAP 0.3-0.4	BLEU 0.3-0.4 mAP 0.5-0.6	BLEU 0.387 mAP 0.772	ターゲット達成
④	AI生成コンテンツ検出	-	F1 0.7(全体) 0.95(ロゴあり)	F1 0.78(全体) 0.99(ロゴあり)	ターゲット達成
⑤	エンティティ検知	F1 0.55-0.60(ノイズ) 0.60-0.65(クリーン)	F1 0.65-0.70(ノイズ) 0.75以上(クリーン)	F1 0.5764(ノイズ)	ベースライン達成
⑥	センチメント・エモーション分析 (3クラス分類)	F1 0.65-0.70	F1 0.75-0.80	F1 0.675	ベースライン達成

- ③モーダル横断メディア理解、SNS投稿に含まれる画像・動画に対して本コネクタが自動生成した説明文および重要項目リストを、人手で作成した注釈と比較し、文章の近さを表す指標としてBLEUスコア、重要な物体・シーン・テキスト要素をどれだけ漏れなく上位に抽出できているかを表す指標としてmAPをそれぞれ算出した結果、いずれの指標においてもターゲットKPIを満たす水準の性能を確認した。
- ④～⑥の各項目については、開発主体が作成した正解データ（答案）とシステム判定結果を照合し、その一致度に基づいてF1スコアを算出することで評価を実施した。
- ④AI生成コンテンツ検出については開発途中から追加した機能としての性質から例外としてターゲットKPIのみを設けることとした。映像内における生成AIによる創作を明示するロゴマークがあるなしにかかわらずターゲットKPIを満たす性能を確認した。
- ⑤エンティティ検知においてはノイズが多いデータセットに対して、ベースライン達成を達成する性能を示した。
- ⑥センチメント・エモーション分析 (3クラス分類)に関してはベースライン達成を達成する性能を示した。

4-2. 検証及び調査の個別詳細

エンリッチメント機構KPIとそのステータス

KPI達成状況及びスコアの詳細

番号	KPI	ベースラインKPI	ターゲットKPI	達成スコア	ステータス
⑦	埋め込み表現の取得(下流タスク性能による間接評価)	Precision@10 0.60 NMI 0.5	Precision@10 0.70以上 NMI 0.6-0.7	V-measure (avg): 0.5949 nDCG@10: 0.4498	ターゲット達成 (V-measure、元のKPIとの互換性を確認済み)
⑧	アカウント真正性分析	F1 又は MCC 0.60-0.70	F1 又は MCC 0.75-0.80	F1 0.9716	ターゲット達成
⑨	アカウントプロファイリング	専門家評価 0.60-0.70	専門家評価 0.70-0.80	1.00	ターゲット達成

- ⑦埋め込み表現の取得については、下流タスク性能による間接評価として実施した。当初はPrecision@10及びNMIを評価指標として採用していたが、Precision@10と比較してnDCG@10は表示結果の質を評価に含めることが可能であり、またV-measureはNMIと比較してクラスタリング性能をより厳密に評価できることから、最終的にこれらの指標をKPIとして採用した。その結果、ターゲットKPIを満たす性能を確認した。
- ⑧アカウント真正性分析については、開発主体が作成した正解データ（答案）とシステム判定結果を照合し、その一致度に基づいてF1スコアを算出することで評価を実施した。なお、真正性に関する定義は一般的に十分集約されていないことから、評価に当たっては適切な定義への調整を行った。その上で、ターゲットKPIを達成する性能を確認した。
- ⑨アカウントプロファイリングについては、専門家評価に基づき性能評価を実施した。評価に際しては、専門家評価に主観が混入することによる妥当性及び客観性への影響が懸念されたが、複数の専門家による評価体制を採用することで、その影響の低減を図った。その結果、ターゲットKPIを達成する性能を確認した。

4-2. 検証及び調査の個別詳細

エンリッチメント機構：検証およびモデル選定の過程の一例

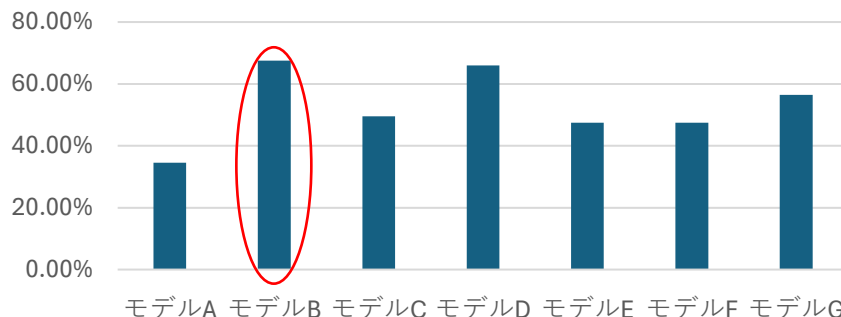
センチメント・エモーション分析

センチメント・エモーション分析において各モデルを検証し、最も精度の高いモデル（右赤枠内）を採用した。

検証に使用したデータ

WRIME ver.2

各モデルの3値分類正確性



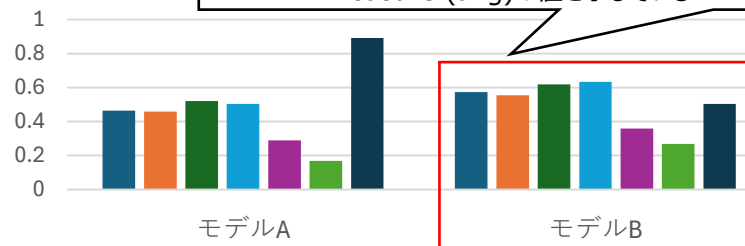
埋め込み表現の取得

埋め込み表現の取得において、複数のデータセットを用いて複数のモデルの精度を検証した。その結果、すべてのデータセットで比較的安定して高い精度を示した、汎用性の高いモデル（右図赤枠内）を採用した。

検証に使用したデータ

NFCorpus,
TwentyNewsgroupsClustering

各棒グラフは、それぞれ異なるデータセットにおけるV-measure (avg)の値を示している



4-2. 検証及び調査の個別詳細

分析機構KPIとそのステータス

本KPIの検証に当たっては、2026年1月に行われた衆議院選挙で観測された情報拡散のデータを使用し、それを現実の事例として性能を検証した。

KPI達成状況及びスコアの詳細

番号	KPI	ベースラインKPI	ターゲットKPI	達成スコア	ステータス
⑩	トピック分析・ナラティブ検知・主張抽出	NPMI: 0.0代後半 専門家評価一致率: 0.60-0.65	NPMI: 0.1-0.2 専門家評価一致率: 0.70-0.80	0.0596 (ノイズ)	ベースライン達成
				0.8467 (クリーン)	ターゲット達成
⑪	拡散傾向の分析・可視化	0.60	0.70以上	1.0	ターゲット達成

- ⑩トピック分析・ナラティブ検知・主張抽出に関しては、抽出されたトピック（ナラティブ・主張）の一貫性を、正規自己相互情報量（NPMI）により評価した。検証にあたり、ノイズの多い大規模データセットと、5トピックに限定した比較的クリーンなデータセットの両方を用いて実行した。その結果、前者では平均NPMIがベースラインKPIを満たし、後者ではターゲットKPIを上回る性能を示した。これにより、データがクリーンであるほど一貫性指標が向上することが確認され、ノイズの多いデータでも一定の性能を担保できることが示された。
- ⑪拡散傾向の分析・可視化に関しては、過去の事例をケーススタディとして本開発技術を用いて分析を行い、技術開発主体の複数アナリストによる主要拡散経路・拡散者の特定結果との一致度を評価した結果、一致度1.0でターゲットKPIを満たした。

KPI検証過程



4-2. 検証及び調査の個別詳細

分析機構KPIとそのステータス

KPI達成状況及びスコアの詳細					
番号	KPI	ベースラインKPI	ターゲットKPI	達成スコア	ステータス
⑫	オーディエンスモデリング	ピアソン相関係数 0.4-0.5	ピアソン相関係数 0.6-0.7	ピアソン相関係数 0.7439	ターゲット達成
⑬	拡散シミュレーション	予測誤差±40%程度	予測誤差 ±20%-30%以内	予測誤差 39.7%	ベースライン達成

- ⑫ オーディエンスモデリングについては、興味領域ごとに区分したアカウント群を対象に、さまざまな話題における「実際の反応数」と「シミュレーション環境における反応数」の傾向がどの程度一致しているかを交差検証し、その結果をピアソンの相関係数として算出した。検証の結果、相関係数は0.7439となり、ターゲットKPIを上回る性能を示した。
- ⑬ 拡散シミュレーションについては、ある話題が実際にどのアカウント群へ広がったかを、シミュレーションによる予測結果と実データとで比較し、拡散予測のKPIを、陰性を含む全件ベースの予測誤差率として算出した。その結果、全体集計では39.7%となり、ベースラインKPIである±40%を下回る性能を確保した。
- 今回のデータセットが衆院選に関するものであったこともあり、政治系アカウント群に関する予測では比較的良好なパフォーマンスを示すことができた一方、他アカウント群においては改善余地が残る結果となった。改善点としては、シミュレーションのためのエージェントの改良、より多量のデータの参照によるシミュレーション等が挙げられる。

KPI検証過程



4-2. 検証及び調査の個別詳細

ビジネスKPIとそのステータス

本取組の目的は、偽・誤情報対策業務への本開発技術の導入が運用負担の軽減と費用対効果の向上に貢献することを立証することにある。これらのKPI達成のために本開発技術を用いた分析の際の各工程の作業時間と手動分析ベンチマークを比較し、各工程の削減率を測った。

KPI達成状況及びスコアの詳細

番号	KPI	ベースラインKPI	ターゲットKPI	達成スコア	ステータス
⑭	関連データの収集・準備	工数削減率： 平均30%-40%削減	工数削減率： 平均50%-60%削減	平均63%削除	ターゲット達成
⑮	主要情報の分析・検証			平均58%削除	ターゲット達成
⑯	拡散シミュレーション			平均34%削除	ベースライン達成
⑰	分析結果に基づく報告・共有			平均41%削除	ベースライン達成

KPI達成における困難と対策

- 全工程で、正規化などのデータ準備、アカウント・投稿の手動確認、レポート用可視化作成といった手作業を大幅に削減した。加えて、ナラティブ抽出などの独自機能により、他社ツール間のデータ連携に伴う工数も短縮した。
- 一方で、情報源設定やOpenCTI連携コネクタの稼働など収集・準備工程、およびデータ解釈には依然として時間を要する。今後は、コネクタ高速化とOpenCTIのGUI上でデータに関する補足情報提示により改善を進める。

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

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

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、システム紹介及びヒアリングを行い、その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、技術開発主体の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取り上げ、洞察や推奨事項を共有することで、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

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

3. 令和7年富士調査研究会同への出展

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。

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

主なフィードバック

対応プラットフォームの拡充や、ボット検知・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

4. マネージドサービスのPoC実施

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

マネージドサービスを先駆的にPoCとして提供した

現在来年度におけるマネージドサービスの実運用を同社へ提案中である

5. 法規制やコンプライアンスに関する調査

利用者が法的リスクを懸念することなくシステムを活用できるよう、関連法規・ガイドラインに関する調査を実施した。調査結果はP.78に記載している。

5-2. 社会実装に向けた取組の個別詳細

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：機能改善に向けたヒアリング

具体的な取組内容

- 主要な開発段階ごとに機能や分析等に関するレビューや意見を収集

目的

- ステークホルダーの知見反映により、ナラティブ拡散要因を拡張することで技術性能を向上
- 情報拡散ネットワークにおける特定済み・疑わしいアカウントの追加マッピングにより実用性を強化
- システムを用いた分析事例の共有、洞察及び改善点に関するフィードバックを収集
- 予備的対策（デバンキング・プレバンキング）の提示及び実務的観点から有効性を検証

主な対象

- 公的機関
- 報道機関

主なマイルストーンとヒアリング内容

8月

開発技術のコンセプトや
ビジネス要求

9月-10月

エンリッチメント機構及び
その可視化

12月

拡散シミュレーション以
外の分析機構及びその
可視化

2月-3月

すべての機能のデモ及び
将来的な機能の向上

5-2. 社会実装に向けた取組の個別詳細

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：社会実装に向けたヒアリング

具体的な取組内容	<ul style="list-style-type: none"> 開発技術に関する意見やフィードバックの収集 拡散シミュレーション機能の有用性をヒアリングにて確認
目的	<ul style="list-style-type: none"> ステークホルダーの拡散シミュレーション機能に関する認識やニーズの把握 ステークホルダーの意見をビジネスモデルの精緻化のために活用
主な対象	<ul style="list-style-type: none"> 公的機関 報道機関

主な取組成果

- システム紹介及びヒアリング件数

8月	9月	10月	11月	12月	1月	2月	合計
4件	3件	6件	2件	6件	3件	1件	25件

ヒアリングに応じた組織数		➔	契約締結に向けた具体的な協議が行われた組織数	
政府機関	17組織		予算請求済み	1件
報道機関			サービスプラン等の協議中	4件
民間企業			計	5件

その他組織に対しても継続的にアプローチ中である。

5-2. 社会実装に向けた取組の個別詳細

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

機能改善に向けたヒアリングを通して得たフィードバックは以下の通りである。

全体的な評価

開発中のシステムについては、有用性が総合的に評価されている

公的機関からのフィードバック

- システム全体としては概ね高評価を得た
- シミュレーション機能に関しては、分析結果の可視化が専門家のみならず一般層にも届く情報発信に寄与し得ると高く評価された
- 文全体ではなく特定のエンティティに対する感情分析が可能である点についても、実運用上のニーズに適していると強調された

ニーズ

分析対象・分析範囲

- 分析対象プラットフォームの拡大（TelegramやWeibo等）
- 中国メディアの分析

機能要件

- Deepfake・AI生成コンテンツの検知
- 偽・誤情報の検知やアラート
- ボット・不正アカウント分析

分析アウトプット・可視化要件

- 時系列変化を把握できる表示
- 不審コンテンツ検知時のアラート表示
- ネットワーク構造（フォロワー/フォロー関係）の可視化

報道機関からのフィードバック

ニーズ

分析対象・分析範囲

- 複数SNSを横断したクロスプラットフォーム分析
- サイバーセキュリティと認知戦を横断的に組み合わせた分析

機能要件

- クラスター機能によるボットネットワークの特定
- 情報拡散の影響度をA～C等の段階で評価・判断できる仕組み

分析精度・信頼性

- 報道用途での使用に耐える高い分析精度の確保

5-2. 社会実装に向けた取組の個別詳細

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

社会実装に向けたヒアリングを通して得たフィードバックは以下の通りである。

全体的な評価

- 偽・誤情報分析ツールの選定を模索している組織が一定数存在し、PoC形式での提供への関心が示された

ニーズ

運用・体制面

- 必要に応じて、プラットフォーム利用や分析を技術開発主体に柔軟に依頼できる体制

リスクや懸念点

分析精度の担保

- 分析精度が担保されなければ報道に使用不可

誤用・誤解釈

- ユーザの知識や理解不足により、誤った解釈・発信につながる可能性

5-2. 社会実装に向けた取組の個別詳細

月次レポートの発行

具体的な取組内容

- 社内アナリストが偽情報や世論操作に悪用され得るSNS上の動向を、開発システムを用いて定期的に分析し、月次レポートとして顧客・取引先に無償配布した
- 2025年8月～2026年2月の計7回にわたり月次レポートを発信した

目的

- 偽・誤情報事例の分析結果に加え、当社の洞察及び推奨事項を提示することで、社会全体における偽・誤情報対策への認識を高める
- 本開発技術が、偽・誤情報に対する対策立案において高い有用性を有することを示す
- 月次レポートの提供を継続することで、個別顧客との接点を継続的に創出する

月次レポートのテーマ

8月	9月	10月	11月	12月	1月	2月※
衆院選(2024)・参院選(2025)におけるSNS上の議論の進展・影響とその比較	戦後80周年に関する海外からの情報作戦	「解体デモ」をめぐるSNS議論と海外勢力の関与	異なる文脈でのクマの襲撃動画の拡散が、日本への情報作戦において果たした役割	高市首相の台湾発言をめぐる情報戦	衆議院選挙における情報作戦リスク予測分析	2026年衆院選の総括：ボット・インフルエンサー・アルゴリズム

※成果報告書最終化期限との兼ね合いにより、当初の予定よりも前倒しで実施

5-2. 社会実装に向けた取組の個別詳細

月次レポートの発行

- 分析に使用した開発システム機能

	11月	12月	1月	2月※
テーマ	AI生成動画	高市首相の台湾発言	衆議院選挙	衆議院選挙
使用した機能	AI生成動画検知機能	センチメント分析機能 ナラティブ分析機能 AI生成動画検知機能	センチメント分析機能	アカウントプロファイリング エンティティ検知・ワードクラウド（可視化）機能 ナラティブ分析機能
分析	AI生成クマ襲撃動画を検知し、その拡散状況を分析	「一つの中国」に関連する投稿とそのセンチメント、ナラティブを言語別に分析	衆院選に関して拡散されているトピックを抽出し、情報作戦に悪用されるリスクを評価	衆院選において影響力を持つアカウント類型とXアルゴリズムへの適応度を分析しSNS上での情報拡散傾向を評価

- 使用した機能及びレポートの例

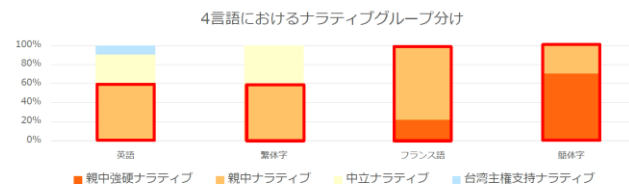
※成果報告書最終化期限との兼ね合いにより、当初の予定よりも前倒しで実施

ナラティブ分析機能



12月レポート

4言語のコメントにおける優勢議論



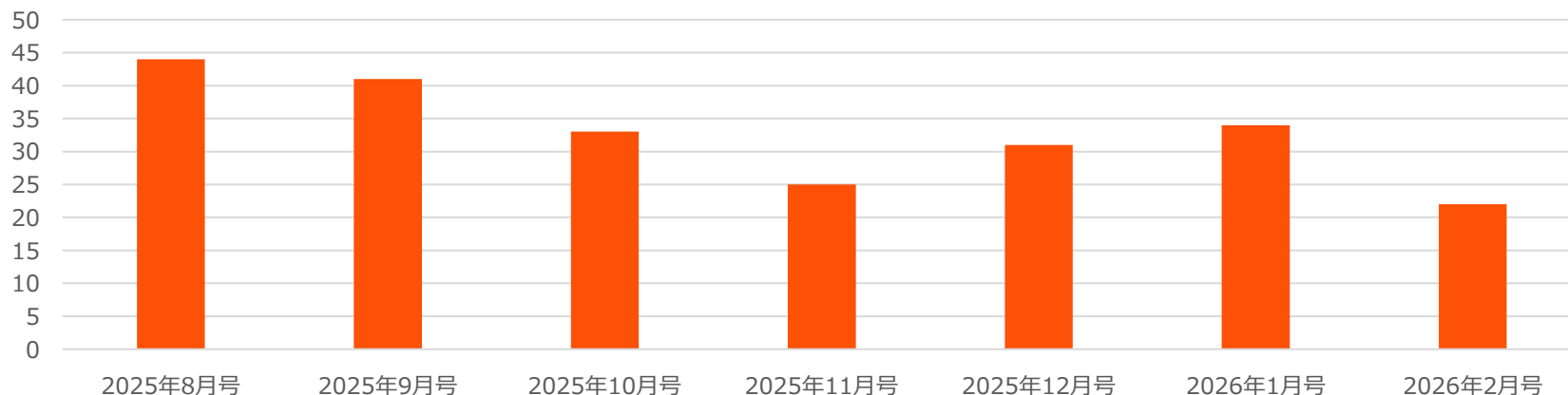
- 「一つの中国」に関連した動画に対する反応はおもに親中ナラティブが全言語において大部分を占めている
- 親中ナラティブ以外のナラティブも「中立」や「台湾主権支持」であり、日本側によるナラティブの浸透は確認できない
- 特にフランス語ではほとんどの主張が親中ナラティブを占めている

5-2. 社会実装に向けた取組の個別詳細

月次レポートの発行

主な取組成果

- 月平均レポート閲覧者数：33人
- 月別レポート閲覧者数



- 一定数のステークホルダーとのエンゲージメントを継続的に維持することができた

- 月次レポートの例

2025年8月号



衆院選(2024)・参院選(2025)におけるSNS上の議論の進展・影響とその比較

情報作戦ハイライト
2025年8月
株式会社コンステラセキュリティジャパン

©2025 CONSTELLA SECURITY JAPAN Inc. All Rights Reserved.

2025年9月号



戦後80周年に関する海外からの情報作戦

情報作戦ハイライト
2025年9月
株式会社コンステラセキュリティジャパン

©2025 CONSTELLA SECURITY JAPAN Inc. All Rights Reserved.

2025年12月号



高市首相の台湾発言をめぐる情報戦

情報作戦ハイライト
2025年12月
株式会社コンステラセキュリティジャパン

©2025 CONSTELLA SECURITY JAPAN Inc. All Rights Reserved.

5-2. 社会実装に向けた取組の個別詳細

令和7年富士調査研究会同への出展

富士調査研究会同は、陸上自衛隊富士学校主催の展示会であり、参加企業が官側関係者等と技術に関する情報共有・意見交換を行う場である。展示企業及び自衛隊・防衛省所属の特定人員のみが参加した。

具体的な取組内容

本開発技術の概要（アーキテクチャ、機能、ユースケース等）や、ハイブリッド分析のフレームワーク、想定成果、試作ソフトウェアの紹介

目的

本開発実証及び次年度以降の本開発技術に関するユーザ側の認識やニーズの把握及びフィードバックの収集

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

官側関係者に本開発技術の認知を図ることができた

- デモを実施した本開発技術の当時GUI及び展示会の様子

本開発技術の当時GUI



展示会の様子



5-2. 社会実装に向けた取組の個別詳細

令和7年富士調査研究会同への出展

以下では、会同で受領したフィードバックを整理した。

全体的な評価

- OSINTツールとして活用したいとの評価を得た
- サイバー領域と認知戦を横断的に分析できる統合的な分析基盤としての有用性が示された

関心

- X、Weibo、Signal等の世界的に利用されているプラットフォーム、とりわけ中国で広く使用されているSNS
- 日本と同様に中国周辺に位置する台湾やフィリピンに対する中国の情報戦術や、陰謀論者にリーチするプラットフォーム

ニーズ

- ディープフェイクや生成AIコンテンツの検知
- ボットの作成主体・運用方法・目的・接続サーバーの分析
- 偽・誤情報検知及びファクトチェック機能
- 過去のサイバー攻撃・認知戦データを活用したAI分析
- 新旧パターンの自動検知
- 実務及び上層部への報告に有用な数値指標を重視した可視化

5-2. 社会実装に向けた取組の個別詳細

マネージドサービスのPoC実施

対象企業（製造業に属する国内企業）に関するSNS上の言説を分析し、その結果をレポートとして同社に提供した。

具体的な取組内容

- 複数のSNSプラットフォームにおいて対象企業に関する言及をモニタリング
- 対象企業を標的としたネガティブキャンペーンの検知
- ネガティブキャンペーンの拡散に関与する高活動アカウントの特定及び主要な特徴の整理

目的

- 来年度におけるマネージドサービス契約の締結に資する

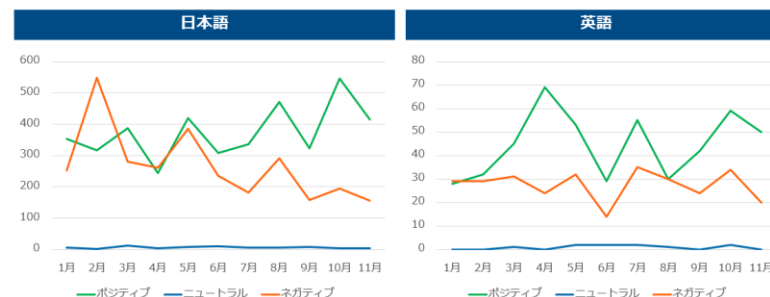
主な取組成果

先駆的にマネージドサービスをPoCとして提供することが出来た

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

センチメント分析機能

センチメント分析機能を使用し、対象企業に対するSNS上の言説のセンチメントの推移を分析した。



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組を通して得たフィードバック

全体的な評価

サイバー領域と認知戦を統合した分析が可能である点については、特に高い評価を受けた

開発中のシステムについては、有用性が総合的に認識されている

実運用を想定し、OSINTツールとして実際に使用したいとの前向きな評価が得られた

報道機関からのフィードバック

ニーズ

分析対象・分析範囲

- 複数SNSを横断したクロスプラットフォーム分析
- サイバーセキュリティと認知戦を横断的に組み合わせた分析

機能要件

- クラスタ機能によるポットネットワークの特定
- 情報拡散の影響度をA～C等の段階で評価・判断できる仕組み

分析精度・信頼性

- 報道用途での使用に耐えうる高い分析精度の確保

運用・体制面

- 必要に応じて、プラットフォーム利用や分析を技術開発主体に柔軟に依頼できる体制

リスクや懸念点

分析精度の担保

- 分析精度が担保されなければ報道に使用不可

誤用・誤解釈

- ユーザの知識や理解不足により、誤った解釈・発信につながる可能性

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組を通して得たフィードバック

公的機関からのフィードバック

関心

海外の情報作戦

- 中国による日本や周辺アジア諸国に対する影響工作

政府情報基盤の悪用

- 政府の公式情報源から発信された情報が、切り取られ、政府の意図とは異なる形で拡散される事例への対策

アクターのスタンスの把握

- アカウント所有者個人の思想、価値観、イデオロギーまで踏み込んだ分析

ニーズ

分析対象・分析範囲

- 分析対象プラットフォームの拡大（TelegramやWeibo等）
- 中国メディアの分析

機能要件

- Deepfake・AI生成コンテンツの検知
- 偽・誤情報の検知やアラート
- ボット・不正アカウント分析

分析アウトプット・可視化要件

- 上位者報告に耐える定量的アウトプット
- 時系列変化を把握できる表示
- 不審コンテンツ検知時のアラート表示
- ネットワーク構造（フォロワー/フォロー関係）の可視化

5-2. 社会実装に向けた取組の個別詳細

法規制やコンプライアンスに関する調査

本システムがSNS・ニュース記事・ウェブサイトのデータを収集・分析する際には、個人情報保護法やGDPR等の関連法令、プラットフォーム規約を順守する必要がある。ゆえに、これらを調査し遵守方針と運用上の留意点を整理した。

本開発技術の特徴

SNSのユーザID・ハンドルは他の公開情報との照合により個人識別につながり得るため、「個人情報」として扱われる場合がある

SNS上のデータを取り扱うため、不正取得・回避行為リスクの抑止が必要とされている

グローバルなSNSデータを収集するなかで欧州居住者のデータを取り扱う場合がある

適用すべき法規・ガイドライン

日本	共通	海外
<ul style="list-style-type: none"> ① 個人情報の保護に関する法律 ② 個人情報保護委員会ガイドライン 	<ul style="list-style-type: none"> ① SNSプラットフォーム規約 	<ul style="list-style-type: none"> ① EU一般データ保護規則 (General Data Protection Regulation, GDPR)

主要義務	データ取得	保管	処理
利用目的の明確化	偽・誤情報拡散リスクの早期把握を目的とする	定義された目的以外での利用を防ぐための設計と運用を行う	SNSアカウントを実在人物に意図的に紐付けることを目的としない
適正取得	認証回避やブロック回避等を禁止する	データ収集設定を変更管理する	該当しない
最小化	公開された情報のみを収集し、取得フィールド等を限定する	処理・分析に必要なないデータを保存しない	分析は必要最小限の特徴量で行う
保存制限	分析に必要な期間に限った保存とする	目的の達成に必要な期間に限った保存とする	削除・抑止依頼により再収集を防止する

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

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

1. 月次ウェビナーの開催と月次レポートの発信

ウェビナー及びレポートを通じ、関係組織に対して本開発技術を用いた偽・誤情報の調査・分析結果を発表した。

開催日	テーマ
8月6日	衆院選(2024)・参院選(2025)におけるSNS上の議論の進展・影響とその比較
9月10日	戦後80周年に関する海外からの情報作戦
10月15日	「解体デモ」をめぐるSNS議論と海外勢力の関与
11月19日	ウェビナー：生成AI動画の台頭と日本への脅威 レポート：異なる文脈でのクマの襲撃動画の拡散が、日本への情報作戦において果たした役割
12月23日	高市首相の台湾発言をめぐる情報戦
1月28日	衆議院選挙における情報作戦リスク予測分析
2月25日※	2026年衆院選の総括：ポット・インフルエンサー・アルゴリズム

※成果報告書最終化期限との兼ね合いにより、当初の予定よりも前倒して実施

主な取組成果

月平均リーチ数	ウェビナー参加組織数（累計）
ウェビナー参加者：65人 レポート閲覧者：33人	133組織

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

2. メディアとの連携

ウェビナーに参加した報道機関からの依頼を受け、関連テーマで取材協力に応じた。

報道機関	テーマ
国内大手メディア (8月)	<ul style="list-style-type: none">2024年の衆院選と2025年の参院選におけるSNS戦略の比較
NHK (11月)	<ul style="list-style-type: none">AI生成動画の増加とその影響

6-2. 普及啓発活動の個別詳細

月次ウェビナーの開催と月次レポートの発信

具体的な取組内容

- 参議院選挙や生成AI動画等の時事的テーマを題材に、本開発技術を用いた分析結果を、2025年8月から2026年2月における計7回のウェビナー開催及びレポート発信を通じて発表した
- 11月以降はウェビナーにおいてシステムデモを実施し、アンケートを通じて参加者からのフィードバックを収集した

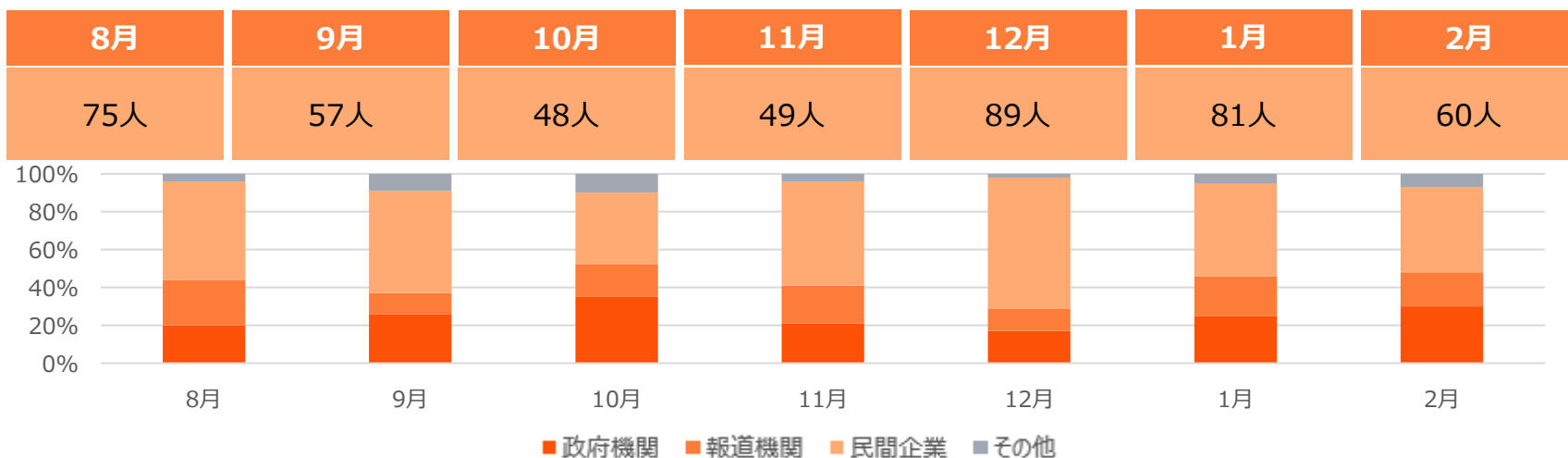
目的

- ケーススタディを通じて偽・誤情報の分析及び対策における本開発技術の有用性を示し、関係組織への価値を訴求する

主な取組成果（月次ウェビナー）

毎月平均65人、49組織が参加し、本開発技術の幅広い認知拡大を図ることができた

参加人数の詳細

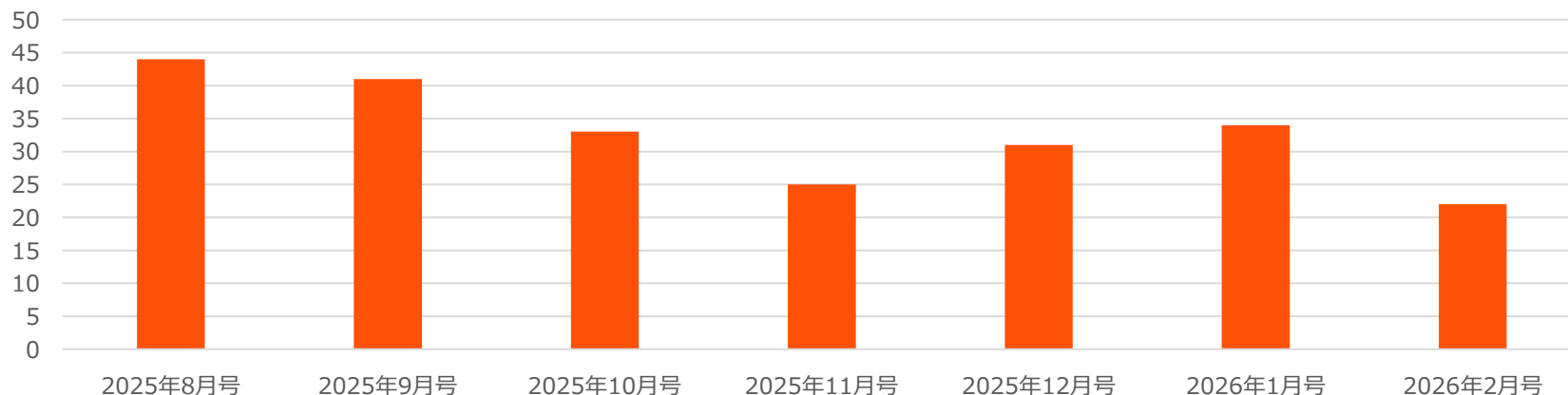


6-2. 普及啓発活動の個別詳細

月次レポートの発信

主な取組成果（月次レポート）

メールにて希望者に送付した月次レポートは平均して毎月約33人にアクセスされた



毎月概ね30人程度のアクセスが継続的に確認されており、本レポートは偽・誤情報に対する意識啓発や、本開発技術の価値を示す上で一定の効果を得たことが示唆される。

2025年8月号



衆院選(2024)・参院選(2025)におけるSNS上の議論の進展・影響とその比較

情報作戦ハイライト
2025年8月
株式会社コンステラセキュリティジャパン



©2025 CONSTELLA SECURITY JAPAN Inc. All Rights Reserved.

2025年9月号



戦後80周年に関する海外からの情報作戦

情報作戦ハイライト
2025年9月
株式会社コンステラセキュリティジャパン



©2025 CONSTELLA SECURITY JAPAN Inc. All Rights Reserved.

2025年12月号



高市首相の台湾発言をめぐる情報戦

情報作戦ハイライト
2025年12月
株式会社コンステラセキュリティジャパン



©2025 CONSTELLA SECURITY JAPAN Inc. All Rights Reserved.

6-2. 普及啓発活動の個別詳細

月次ウェビナーの開催と月次レポートの発信

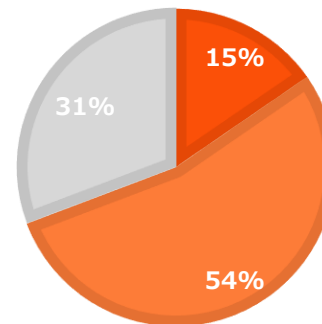
フィードバック

ウェビナー参加者の関心度

- ・ システムデモを行った11月ウェビナーにおいてアンケート回答者のうち約70%が「興味がある」と回答した
- ・ さらに、そのうち15%が個別デモを希望した

システム全体への評価

- ・ 非常に有用であり、実運用での利用を望むという声を多数確認できた
- ・ ナラティブ分析・感情分析等、複数観点での可視化機能が高い評価を受けた



- 興味がある：個別デモ・フリーフィングを希望する
- 興味がある：今後のウェビナーで詳しく知りたい
- 今は特に興味はない

参加者別のフィードバック詳細

公的機関

評価

- ・ モーダル横断分析や検索の中国語への対応を高く評価

ニーズ

- ・ X・TikTok・YouTubeに加え、Telegramや中国系SNSへの対応を期待
- ・ ユーザー名変遷の追跡やフォロー/フォロワー関係の可視化を要望

報道機関

評価

- ・ Open CTI上で、サイバーセキュリティと認知戦を横断的に分析できる点を高く評価

ニーズ

- ・ 報道利用に耐える分析精度の確保が必要
- ・ 情報拡散の影響度を段階的・定量的に評価できる仕組みを期待
- ・ クラスタ分析におけるポット判定機能の実装を要望

懸念点

- ・ データ解釈次第で誤報道につながる可能性への懸念も指摘

6-2. 普及啓発活動の個別詳細

海外パートナーとの共同ウェビナーの開催

偽・誤情報対策を専門とする海外パートナーとの知識共有セッションを実施し、得られた知見を本技術の開発や分析の方法論、レポート作成に反映するとともに、ブリーフィングにてステークホルダーに共有した。

目的

日本の対象ユーザが国際的な事例やベストプラクティスを参照できるようにすることで、本システム利用の国際的な利点を訴求する

主な取組成果

開催日	パートナー	共有セッションの内容	日本の対象ユーザに対する有益性
11月5日	イスラエルのSNS脅威分析企業	<ul style="list-style-type: none"> 分析範囲・分析対象プラットフォーム 不正アカウントへの対応手順 	<ul style="list-style-type: none"> 偽・誤情報対策の効率化にあたっては、中国プラットフォームを対象とする必要がある 国際標準に基づき組織的拡散の兆候を早期に把握するためには、なりすまし及び協調行動の発見を最優先とすべきである
11月19日	台湾の脅威インテリジェンス企業	<ul style="list-style-type: none"> 中国による台湾に対する情報作戦分析の分析結果と手法 アカウント分析の手法についての知識共有 	<ul style="list-style-type: none"> YouTubeやTikTok上で中国に帰属している可能性のある協調的な活動が確認された 関与アカウントは政治目的のみならず、収益化や運用ノウハウ販売を狙っている場合もある
12月3日	台湾のThreat Analyst Summit 2026に参加	<ul style="list-style-type: none"> 中露連携による日本を標的とした情報戦 	<ul style="list-style-type: none"> 情報作戦の今後の兆候を適切に予測するには他領域や戦略目標と統合した分析が必要である 偽・誤情報への対策は、脅威の強度に応じて段階的に実施する必要がある
1月28日	米国の地政学戦略・リスク分析企業	<ul style="list-style-type: none"> 昨今の地政学的情勢 中国の対日情報作戦（特に若年層向け） 台湾有事と日本への影響 	<ul style="list-style-type: none"> 地政学的要因に紐づく偽・誤情報は民間企業の事業運営や意思決定に影響し得るため、重要なリスクとして継続的に監視することが必要である 中国等の懸念国による情報作戦の兆候を適切に把握するためには、海外事例調査と知識共有、並びに多言語分析が必要不可欠である

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

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及びをそれらを踏まえた今後の展望①

関係者ヒアリングやチーム内検討に加え、市場動向・顧客ニーズの調査を通じて、技術開発及び社会実装に向けた課題と今後の展望を整理した。

技術開発における課題	検証過程・要因	対応事項・改善案
外部ソフトウェアの機能・性能制約		
<ul style="list-style-type: none"> エンリッチメント及び分析モジュールにおける性能の限界 	<ul style="list-style-type: none"> 想定データ量及び同時要求の負荷検証で、応答遅延とスループット低下を確認 外部I/O+逐次処理がボトルネックである 	<ul style="list-style-type: none"> 非同期処理・キューイング・段階的 結果返却による体感レイテンシ改善 キャッシュ、差分更新、事前集計による再計算削減とスループット向上
<ul style="list-style-type: none"> 大量データ取り込み及び複雑な関係やグラフ処理の困難 	<ul style="list-style-type: none"> 大量データのインポート/関係生成の検証で、キュー滞留・処理待ちを確認 関係・グラフの更新/探索が高コストで、OpenCTI内部処理のスループットに制限されることが要因として考えられる 	<ul style="list-style-type: none"> 取り込みパイプラインの分離と制御による処理工程の最適化 処理サーバーの増強、アーキテクチャ上のデータ処理・加工に用いられるサービス・モデルの変更
<ul style="list-style-type: none"> OpenCTIにおける可視化カスタマイズの制約による、分析ニーズに適合したダッシュボード等を設計する際の柔軟性の制限 	<ul style="list-style-type: none"> OpenCTIの標準ダッシュボード・ウィジェットが分析ニーズ（指標・粒度）の実現には不足することを確認 	<ul style="list-style-type: none"> 可視化カスタマイズに関してOpenCTI開発元との協議を検討 高度な可視化ニーズを有する顧客向けに、OpenCTI上のデータを基盤として追加のカスタムダッシュボードを外部に構築する方針や分析結果表示画面のカスタマイズなどの代替手段を検討

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及びをそれらを踏まえた今後の展望②

技術開発における課題	検証過程・要因	対応事項・改善案
外部ソフトウェアの機能・性能制約		
<ul style="list-style-type: none"> インターネット公開情報を元に学習された外部LLMの使用に起因する、<u>分析品質・制度の外部依存性</u>への懸念 	<ul style="list-style-type: none"> KPI検証中に同一入力を複数回実行したところ、出力の揺れや根拠不明な回答の発生を一部確認 要因としては偽・誤情報に特化していない外部LLMの使用が考えられる 	<ul style="list-style-type: none"> 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 過去分析で検証済みのデータセットをナレッジ基盤として整備し、RAGを適用する運用の導入を検討する
アーキテクチャ・依存関係設計の選択		
<ul style="list-style-type: none"> SNSデータ抽出における<u>限られたデータソースへの依存</u>に起因する運用上の脆弱性 	<ul style="list-style-type: none"> 特定データソースによるレート制限・仕様変更・停止の影響が原因である 	<ul style="list-style-type: none"> SNS等の公式API連携を含むデータソースの多様化 レート制限時等に切替用の予備データソースを用意
運用・保守		
<ul style="list-style-type: none"> 復旧対応の長期化 	<ul style="list-style-type: none"> 停止や遅延を想定した場合、処理滞留の可能性を確認した 自動切替・再実行（リトライ）・迂回経路などのフォールバック設計が不足している 	<ul style="list-style-type: none"> 保守体制の強化

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及びをそれらを踏まえた今後の展望③

技術開発における課題	検証過程・要因	対応事項・改善案
分析手法上の推論限界		
<ul style="list-style-type: none"> アカウント行動に関する具体的情報（閲覧実態等）の取得制限による推定精度の限界 	<ul style="list-style-type: none"> プラットフォーム上で行われる全ての行動についてデータの抽出や学習ができるわけではない 	<ul style="list-style-type: none"> 社会学な知見に基づくアカウント分析結果の検証による、ボトムアップ・トップダウンの双方のアプローチでの誤差縮小
<ul style="list-style-type: none"> 公開情報のみを収集しており、各アカウント運用者の認知状態（意図・心理）までを把握できず、分析がSNS上の議論に限定され、現実世界における思想の広がりまでは対象外となる可能性 	<ul style="list-style-type: none"> ユーザの思惟過程がプラットフォーム上の行動として必ずしも表出しない事により、行動ベースのユーザ認知の推論には制限がある 	<ul style="list-style-type: none"> 分析精度向上に向けた追加要因を取り込む分析手法・モデル設計の研究開発 社会学・心理学調査によるモデルの補正 イベント、行政発表等の外部情報の統合による状況文脈の補強

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及びをそれらを踏まえた今後の展望①

社会実装における課題	検証過程・要因	対応事項・改善案
市場での位置づけ・競合		
<ul style="list-style-type: none"> 導入に向けた詳細の明確化 	<ul style="list-style-type: none"> ステークホルダーへのヒアリングを通じて、製品化に向けた具体的なニーズを把握することができた。同時に、実際の導入を想定した場合には、製品化の詳細な内容の提示が求められた 	<ul style="list-style-type: none"> 今後は分析の質の向上に加え、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
導入促進・ユーザ理解・教育		
<ul style="list-style-type: none"> 意思決定者層における本開発技術の事業上の必要性・<u>導入意義に関する認識の形成</u> 	<ul style="list-style-type: none"> 意思決定者へのヒアリングを通じ、脅威への関心はある一方で、事業リスクや効果の根拠（ROI・リスク低減）を説明することが求められた 	<ul style="list-style-type: none"> 意思決定者向けに、脅威の存在と背景を整理 本開発技術の戦略的価値（リスク低減、優先順位付け、意思決定の迅速化）を明確化 導入効果及び想定ユースケースの整理（適用範囲、利用シナリオ、判断フロー）
<ul style="list-style-type: none"> 分析・シミュレーション結果が悪用される懸念 	<p>ステークホルダーへのヒアリングにおいて、分析機構の出力（特にシミュレーションの結果）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</p>	<ul style="list-style-type: none"> エンドユーザへの教育、レポート作成時の言語化への注意喚起 特定のモジュールの提供を一部の顧客に限定する 利用ガイドライン及び禁止事項を整備し、導入契約・運用ルールに組み込む

7-1. 技術開発及び社会実装における課題・展望

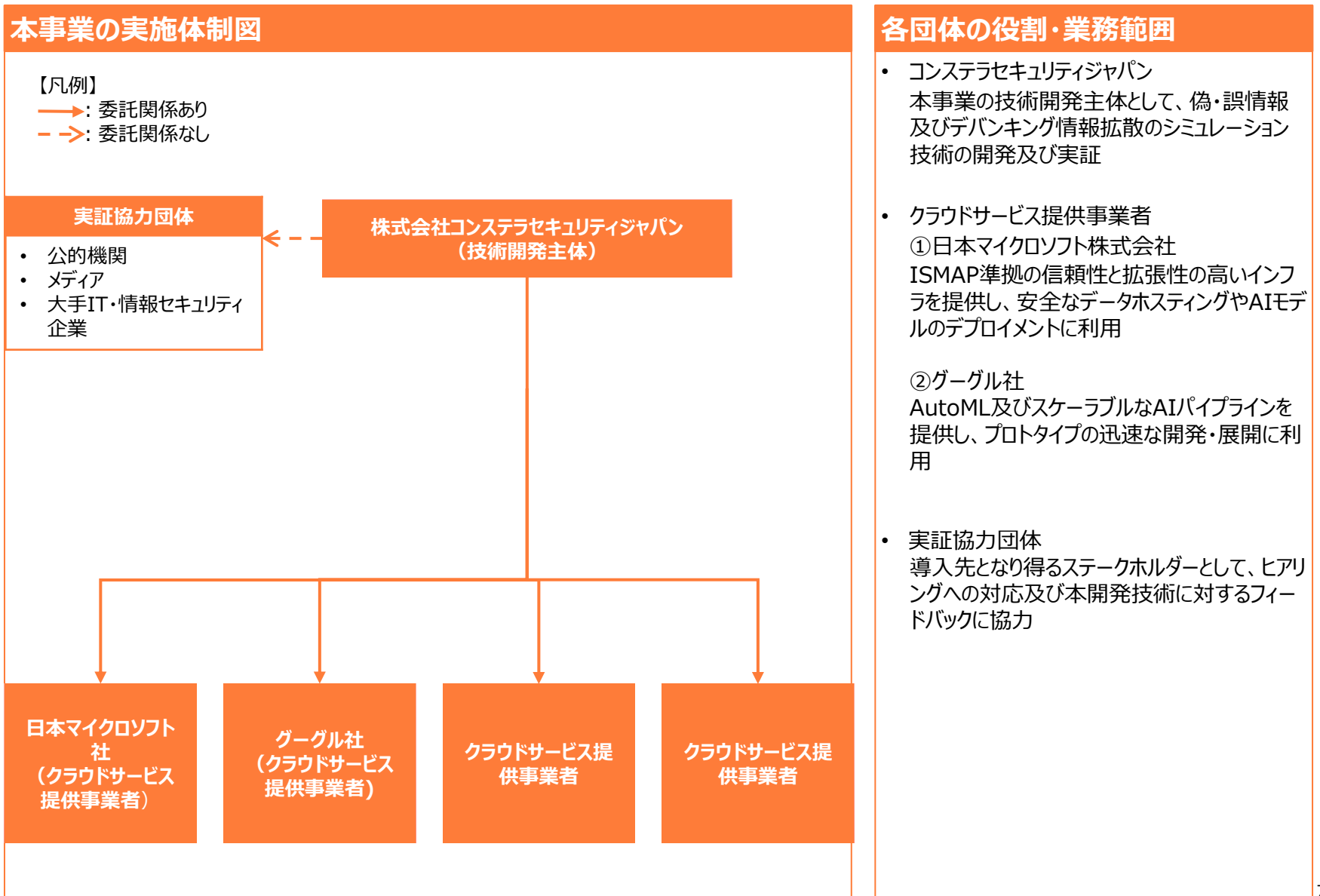
社会実装にあたっての今後の課題及びをそれらを踏まえた今後の展望②

社会実装における課題	検証過程・要因	対応事項・改善案
事業戦略		
<ul style="list-style-type: none"> 公的機関関連の顧客への依存による、<u>市場拡大の制限</u>に対する懸念 	<ul style="list-style-type: none"> ステークホルダーとのヒアリングにより、民間企業の偽・誤情報拡散への懸念と関連ソリューションへの関心は高まりつつある一方で、導入優先度は依然として低いことを確認した 	<ul style="list-style-type: none"> 民間企業で成立しやすいテーマ（ブランド毀損・風評、金融市場の情報操作、サイバー脅威の情報環境分析、危機広報）のユースケースを用意し、サービス提供モデルを具体化 報道機関との連携を向上し、本開発技術を使った共同調査を実施 シミュレーション機能の教育・リテラシー分野における活用について検討
運用・実装成熟度と現実社会への接続		
<ul style="list-style-type: none"> 長期的な信頼性を確保するため、持続可能な運用・保守体制を確立する必要がある 	<ul style="list-style-type: none"> 本開発技術は現在、開発～初期検証段階であり、今後事業化・顧客提案へ移行するには、監視・障害対応を含む持続可能な運用・保守体制の確立が求められる 	<ul style="list-style-type: none"> 調査別の対応からサブスクリプションモデルへの段階的な移行 運用・保守体制を定義し、教育の仕組みを整備 運用・保守を収益化し、持続可能な採算構造を採用 導入顧客から常時フィードバックを収集し、新たなニーズを踏まえた今後の開発ロードマップを作成

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

8-1. 実施体制及び役割分担



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)インターネット上の偽・誤情報等への対策技術の開発	→							
1. OpenCTI/コネクタの実装	→							
1.1 データモデル定義・実装 (コネクタ開発)	→							
1.1.1 対応予定のデータソースのデータモデル分析 (X、YouTube、TikTok、ニュース記事、検索エンジン結果)	→							
1.1.2 統一データモデルの定義	→							
1.1.3 統一データモデル実装	→							
1.2 システムインフラセットアップ	→							
1.2.1 クラウドインフラ準備	→							
1.2.2 OpenCTIのデプロイメント	→							
1.2.3 対応予定のデータソースとのコネクタ経由の接続	→							
2. エンリッチメント機構の実装	→							
2.1 投稿テキスト・画像・動画分析モジュール	→							
2.2 アカウト分析モジュール	→							
2.3 ナレッジグラフモジュール	→							
2.4 エンリッチメント機構による分析の可視化	→							
2.5 フィードバックの反映	→							

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. 分析機構の実装		→						
3.1 オーディエンスモデリングモジュール		→						
3.2 トピック・ナラティブ・主張の検知モジュール		→						
3.3 トピック・ナラティブ・主張の拡散傾向分析・可視化モジュール			→					
3.4 拡散シミュレーション設計			→					
3.5 拡散シミュレーション開発			→					
3.6 フィードバックの反映					→			

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	
(2)インターネット上の偽・誤情報等への対策技術の有効性等に関する検証及び調査	→								
1. コネクタ・エンリッチメント機構の実証	→								
1.1 KPI関連データ収集・分析	→								
1.2 実証レポート作成	→								
1.3 協力主体へのデモ及び報告		→							
1.4 フィードバックの受領・評価・改善		→	→						
2. 分析機構の実証				→					
2.1 KPI関連データ収集・分析				→		→			
2.2 実証レポート作成				→	→		→		
2.3 協力主体へのデモ及び報告					→		→		
2.4 フィードバックの受領・評価・改善					→	→	→	→	

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) インターネット上の偽・誤情報等への対策技術の社会実装に向けた取組	→							
1. ステークホルダーへのヒアリング	→							
1.1 システムのデモ・機能紹介	→		→		→		→	
1.2 ステークホルダーからのフィードバックの受領・評価・改善	→		→		→		→	
2. システムの利用により作成した誤・偽情報分析事例の定期的レポート発行	→							
2.1 レポート作成	→	→	→	→	→	→	→	→
2.2 レポート配布	→	→	→	→	→	→		→
2.3 レポート報告イベント	→	→	→	→	→	→		→
2.4 フィードバック収集	→	→	→	→	→	→		→

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					→			
1. システムデモ				→				
2. 成果報告書・社会実装実施計画書（案）執筆 1回目					→			
2.1 成果報告書					→			
2.2 成果報告書 概要版						→		
2.3 成果報告書 簡易版						→		
2.4 社会実装実施計画書					→			
2.5 社会実装実施計画書（案）1回目提出						→		
3. 成果報告書・社会実装実施計画書（案）修正						→		
3.1 成果報告書						→		
3.2 成果報告書 概要版						→		
3.3 成果報告書 簡易版						→		
3.4 社会実装実施計画書						→		
2.5 社会実装実施計画書（案）2回目提出						→		

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・社会実装実施計画書最終版 提出								→
4.1 成果報告書								→
4.2 成果報告書 概要版								→
4.3 成果報告書 簡易版								→
4.4 社会実装実施計画書								→
4.5 社会実装実施計画書最終版 提出								→
5. 成果報告会								→
(5) 普及啓発活動への協力	→							
1. 誤・偽情報に関するケーススタディの月次発信	→	→	→	→	→	→	→	→
1.1 分析レポート準備	→	→	→	→	→	→	→	→
1.2 レポート配布	→	→	→	→	→	→	→	→
2. 月次報告イベント	→	→	→	→	→	→	→	→