様式３

警察活動の高度化・効率化のためのＡＩの試験的導入及び実導入に向けた検討

人工知能等を用いたサイバー空間における違法・有害情報の探索・分析技術の実用化に向けた検討

官民研究開発投資拡大プログラム（PRISM）

「ＡＩ領域」

令和４年度成果

令和５年３月

警察庁

資料１ 「人工知能等を用いたサイバー空間における違法・有害情報の探索・分析技術の実用化に向けた検討」の概要

アドオン額：60,000千円（警察庁）

元施策・有／PRISM事業・新規

課題と目標

■（課題）膨大かつ多種多様に存在するサイバー空間上の違法・有害情報による被害が深刻化
これら情報を人手で全て対処することは困難

■（目標）汎用的なサイバーパトロールの仕組みを構築・実装し、サイバー空間における違法・有害情報対策を高度化・効率化することで、効果的な取締り・抑止を実現する。

「人工知能等を用いたサイバー空間における違法・有害情報の探索・分析技術の実用化に向けた検討」の概要

■元施策：
将来構想である汎用的なサイバーパトロールシステムの構築に向け、キーワード等により対象を絞ったwebクローリングと自然言語処理等によるコンテンツの違法性判断を組合わせたサイバーパトロール手法の開発・検証を行う。（R4年度：35,006千円）

■PRISMで実施する理由：
サイバー空間における違法・有害情報の探索・分析に係る業務の高度化・効率化により、警察における当該情報への対策に要する財政支出の効率化が見込まれるほか、本事業の成果は、民間企業・権利者団体による不正商品対策等にも活用できる可能性があり、官民双方の取組によりサイバー空間の安全・安心の確保が進むことが期待され、民間企業等（ネットオークション企業等）の投資誘発効果も見込まれるため、PRISMで実施する。

■テーマの全体像 ：

【元施策】
ＡＩの自然言語処理技術を活用したWebコンテンツの違法性判断に関する実証実験（委託事業）

＋

【PRISM事業】
画像分析技術等による違法・有害情報の探索・分析に関する調査

▶

＜警察業務への実装＞
○サイバー空間における違法・有害情報の探索・分析の高度化・効率化
○限られた警察力の最適配分を実現

▶

○サイバー空間の安全・安心の確保
○公正な経済活動

出口戦略

○ AI等を活用した効率的なサイバー空間上の違法・有害情報の探索・分析を実現し、警察業務に実装する。

民間研究開発投資誘発効果等

○民間投資誘発効果：
PRISM事業を通じて、不正商品対策に活用可能な技術の現状が整理されるとともに、警察や民間企業・権利者団体における不正商品対策に係るニーズがITベンダーや研究機関にも広く知られることで、民間投資誘発効果が生じる可能性

※ 海賊版対策費（一社あたり平均）：約490万円、対策実施法人数：約３万９千法人

○民間からの貢献額：
15,000千円程度以上の貢献を見込んでいる。

資料２「人工知能等を用いたサイバー空間における違法・有害情報の探索・分析技術の実用化に向けた検討」の概要

アドオン（警察庁）：60,000千円
元施策名：人工知能等を用いたサイバー空間における違法・有害情報の探索・分析に係る実証実験
（R４：35,006千円）

【実証実験】

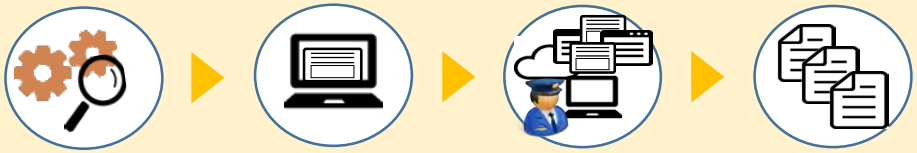
〈目的〉
汎用的なサイバーパトロールシステムの構築に向け、その趣旨に沿った仕組みを小規模に構築・検証

〈内容〉
不正商品関連（海賊版サイト、偽ブランド品売買等）、ヤミ金関連、チケット不正転売関連等の情報を効率的に探索・分析する仕組みを検討・構築し、その効果、課題を業者委託による実証実験により確認する。

サイバー空間上の膨大な情報 ▶ キーワード等により対象を絞ったWebクローリング ▶ 関連情報の抽出 ▶ ＡＩによる違法性の判定・分類 ▶ 人による確認後、削除要請、警告等内容に応じた対応

【PRISM】

民間企業等で実践されている、画像分析技術等を活用したサイバー空間上の違法・有害情報の探索・分析に係る手法について調査を行い、技術的課題を把握した上で、警察業務への実装に向けた検討を行う。



技術調査 ▶ 技術開発の必要性等の整理 ▶ 実用化に向けた課題の整理 ▶ 導入に向けた計画立案、仕様・運用に係る検討

【開発のイメージ】

令和４年度	令和５年度以降
元施策 自然言語処理技術等を活用したWebコンテンツの違法性判断に関する実証実験	**モデル開発・実装** 技術調査・実証実験の成果をもとに、学習データセットの整備、モデル開発を行い、概念実証を経て、実装を目指す。
PRISM 画像分析技術等による違法・有害情報の探索・分析に関する調査・検討	自然言語処理 ＋ 画像認識

① 技術調査
- 民間企業等（プラットフォーム事業者等）で実践されている探索・分析手法、既存の技術・サービス、人材育成等に関する調査を実施

② 現状把握及び課題の整理
- サイバーパトロールに係る警察業務の現状整理
- 民間企業のサイバー空間上の違法・有害情報対策に関するニーズの把握

③ 開発研究・実装に向けた検討
- 警察が行うサイバーパトロールへの技術の導入に向けた課題の整理
- 警察における運用に向けた計画立案、仕様・運用に係る検討

資料３「人工知能等を用いたサイバー空間における違法・有害情報の探索・分析技術の実用化に向けた検討」の目標達成状況

〇施策全体の目標
　汎用的なサイバーパトロールの仕組みを構築・実装し、サイバー空間における違法・有害情報対策を高度化・効率化することで、効果的な取締り・抑止を実現する。

事業名等	当年度目標	目標の達成状況
①技術調査	• サイバー空間上の不正商品等違法・有害情報の探索・分析に活用可能な既存の技術・サービスについて文献調査・動向調査を実施 • 自社で提供するプラットフォームサービスの運営に当たりサイバー空間上の不正商品等の違法・有害情報対策を実施している民間企業等にヒアリングを行い、その手法や人材育成等に関して調査を実施	【基礎調査】 • ３つの技術分野（画像処理・自然言語処理・音声認識）と23のサブカテゴリを対象に網羅的な調査を実施し、違法・有害情報ごとの代表的な手法や技術進展状況について整理を実施 • 「技術分野カテゴリ」単品もしくは組み合わせによるAI支援方法があるか検討を実施 【個別調査】 • 警察における業務担当者に対するヒアリングを実施 • サイバー空間上の不正商品等の違法・有害情報対策を実施している民間企業等を対象に、関連企業・組織の網羅的な抽出、机上調査を行った上で選定し、対策手法等について４社に対するヒアリングを実施
②現状把握及び課題の整理	• 警察で実施しているサイバーパトロールについて、業務フロー等の現状を整理 • サイバー空間上の違法・有害情報対策を実施する民間企業等に対してヒアリングを行い、AI等先端技術の活用による対策の高度化に係るニーズを把握・整理 • 技術調査の結果も踏まえ、警察業務への実装に向けた課題を整理	【サイバーパトロールの現状整理】 • AsIs/ToBe分析により、現状の業務課題と技術課題に分けて整理 • ヒアリング前後に調査目的の説明と検討結果の報告を行い、有意義となる調査を実施 【民間企業等のニーズ把握】 • サイバー空間上の違法・有害情報対策を実施している民間企業等を対象に、関連企業・組織の網羅的な抽出、机上調査を行った上で選定し、対策の高度化に係るニーズについて２社に対するヒアリングを実施
③開発研究・実装に向けた検討	• サイバー空間上の違法・有害情報対策の高度化・効率化に向け、システム構想、運用方策等を検討 • 汎用的なサイバーパトロールシステムの構築やその継続的な活用・運用に係るロードマップを策定	• ①技術調査結果と②現状把握及び課題の整理を基に、プロセス・アクティビティ・タスクを軸としたフレームワークを活用してサイバーパトロールシステムの要件を整理 • 本施策の出口イメージとして今後のサイバーパトロールのシステムの構築等に向けたロードマップ案を作成

資料4 「人工知能等を用いたサイバー空間における違法・有害情報の探索・分析技術の実用化に向けた検討」の成果

技術調査

調査対象技術分野

画像処理（10カテゴリ）
- セグメンテーション
- 画像分類
- 物体認識
- 画像生成
- 画像検索
- 顔認識
- オブジェクトトラッキング
- 姿勢推定
- 異常検知
- OCR

自然言語処理（9カテゴリ）
- 文書分類
- 情報抽出（固有表現抽出）
- 質問応答
- 機械翻訳
- 文書要約
- 文書検索
- 感情分析
- テキスト生成
- 言語モデル

音声認識（4カテゴリ）
- 音声認識
- 音声分類
- 音声合成
- 感情認識

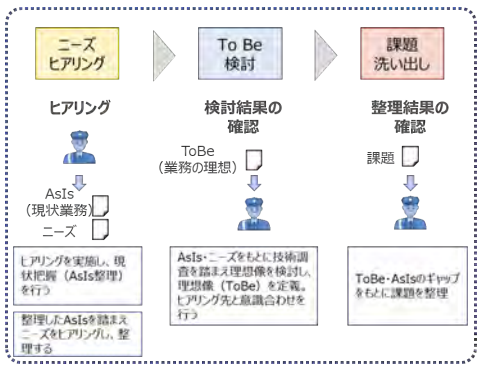
課題	支援方法：ルールベース・システムの作り込み	支援方法：機械学習
重要度のランク付け	ウェブサイトの種類や入っている文言の中でも特に優先度が高いとノウハウが得られているものからルールベースで並べ替えを行う。	■支援① ランキング 収集されたデータを過去の検索クエリ（文章及び画像）とその判断結果から学習したモデルによって優先度を予測する。 ■支援② 説明可能AI プラスαの機能拡充として、説明可能AIによって判断根拠を提示することで、後続の人手による知的作業（違法有害性判定）の補助として活用することが考えられる。
以前調査した対象（重複）の排除	システムで、過去の検索結果と照合することで新たに見つかった対象のみをシステムに追加するとともに、チェック済か否かの表示を可能にすることによって二度手間を防止する。	－
学習データの整備	システム上で人手により違法有害性を判断したときの、URLと文章・画像と判断結果のペアを残すことで、学習データとして蓄積することができる。 さらに過去の違法有害性判定結果から学習データを整備することで、学習データを効率的に拡充できると考えます。	－

支援方法	画像検索×文書検索
概要	入力された検索キーワード及び画像を基にマルチモーダル的に類似する事例を調査しランク付けを行う。部分一致やニュアンス，画像の一部特徴などを基に，より関連度の高い事例を選出する。
代表的技術	BERT 文脈を考慮した単語分散表現を出力する技術。単語分散表現を用いることで、表記が異なっても文脈で意味の異なる単語を区別できるようになる（例えば、食物の文脈で現れる「アップル」とコンピュータの文脈で現れる「アップル」が区別できるようになる）。 GCM ニューラルネットワークベースの類似画像検索としてクエリ画像とDB側の画像マッチングを行う。その中でも次元的な画像クラスタリングを前処理で行うことによって精度向上を果たしている。
代表的製品・ソリューション	OpenAI CLIP 画像とテキストのマルチモーダルで学習した事前学習モデル。ただし日本語モデルの対応が必要
技術進展状況	• 個別技術としては商用利用可能なOSSでの提供が存在 • CLIPについてマルチモーダル検索に最適化したモデルではないため利用にあたってはドメイン・言語に合わせたファインチューニングは必要となる想定 • 本対象ドメインにおいて利用するためには独自での学習データ収集と学習を行い、検証が必要
技術成熟度	TRL4（特定条件下での効能が検証されている）

支援方法	説明可能AI（XAI）
概要	AIモデルを用いて予測を行った際に、どの要因（文章であれば部分・キーワード，画像であればエッジや形状）でその予測になったかを示す技術。
代表的技術	SHAP 局所説明を行うことができるPythonライブラリ。様々なデータ形式（テーブルデータ，画像，テキスト），モデルに対して用いることが可能。特徴量の貢献度をゲーム理論的な指標を用いて按分して説明する。
代表的製品・ソリューション	SHAP Pythonライブラリとして提供。
技術進展状況	• 技術としては，商用利用可能なOSSでの提供が存在 • ただし，マルチモーダルに最適化されたXAIはないため，使うモデルに制約が生じる可能性あり • 技術発展によって解決できる可能性あり
技術成熟度	TRL4（特定条件下での効能が検証されている）

- 警察における業務担当者に対するヒアリング結果等を踏まえ、調査対象技術による警察業務の支援方法を検討
- AIを活用した警察業務の支援方策として「ランキング機能」及び「判断根拠を提示する機能（説明可能AI）」による支援の可能性を示唆

現状把握及び課題の整理



	ヒアリング対象事業者候補の事前調査（公開情報調査）	ヒアリング／知見の取りまとめ
目的	ヒアリングの効率的・効果的な実施に向けた事前情報収集、ヒアリング対象事業者候補の選定	サイバーパトロールシステムの構築に向けた仕様、運用策等の具体化並びにサイバーパトロールシステムの実現に向けたロードマップ作成に資する情報の収集・整理
取組内容（概要）	• 公開情報調査 • ヒアリング対象事業者候補の優先順位付け＋候補の選定	• ヒアリング実施 • ヒアリング結果の取りまとめ

- 警察における業務担当者へのヒアリング結果を基に5W1HでAsIs/ToBeを整理し、業務/技術の観点で課題を抽出
- 民間企業等に対し、違法・有害情報対策に係るヒアリングを実施

開発研究・実装に向けた検討

- 要件定義フレームワークを活用し、サイバーパトロールシステムの要件を整理
- システムの構築等に向けたロードマップ案を作成

資料５「人工知能等を用いたサイバー空間における違法・有害情報の探索・分析技術の実用化に向けた検討」の民間からの貢献及び出口の実績

〇民間からの貢献額：令和４年度　1,551 万円（税込）
（内訳）技術調査（事例、動向調査等）等（1,551 万円（税込））

当年度当初見込み	当年度実績
・技術調査等：15百万円相当	・技術検証：1,551 万円（税込） （内容）検索エンジン利用、検証データ利用（Twitterデータ）、検証・評価作業

〇出口戦略
AI等を活用した効率的なサイバー空間上の違法・有害情報の探索・分析を実現し、警察業務に実装する

当年度当初見込み	当年度実績
（既存技術・サービスの警察活動への実装） • 民間企業等で実践されている、画像分析技術等を活用したサイバー空間における違法・有害情報の探索・分析に係る手法や既存の技術・サービス、人材育成等について調査し、警察が行うサイバーパトロールへの技術導入に向けた課題を整理する。 • サイバー空間上の違法・有害情報対策の高度化・効率化に向け、システム構想、運用方策等を検討し、汎用的なサイバーパトロールシステムの構築に係るロードマップを策定する。 （民間技術の継続的な導入） • 当該仕組みを継続的に活用する上で必要となる人材育成や民間企業、関連団体等との協力関係の在り方等について検討する。	（既存技術・サービスの警察活動への実装） • ①技術調査結果と②現状把握及び課題の整理を基に、プロセス・アクティビティ・タスクを軸としたフレームワークを活用して要件を整理 • 本施策の出口イメージとして今後のサイバーパトロールのシステムの構築等に向けたロードマップ案を作成 • 本施策で把握した画像処理技術（先端技術を活用した画像分類）や、違法有害情報検知技術（民間企業の技術やサービス、運用ノウハウ）及び元施策の実証実験成果を有効活用し、サイバーパトロールシステムを構想 （民間技術の継続的な導入） • 調査・検討結果を踏まえ、中長期視点における将来的な要件について検討