ChatGPT

防衛省×Sakana AI「総合分析業務AI実証」ディープリサーチ報告

Executive Summary

2026年7月29日、防衛省はSakana AI株式会社と「総合分析業務に必要なAI機能の調査・実証」に関する契約を締結した。防衛装備庁が公表した契約額は**3億2,903万2,000円**で、契約相手方はSakana AI株式会社。対象は、防衛省情報本部が担う戦略レベルの「総合分析業務」であり、AIによって①情報収集の効率化、②分析能力の向上、③情報の体系的管理を実証するものである。Sakana AIの公式発表は2026年8月24日、ユーザーが示した「防衛省の情報分析をSakana AIがAI支援　公開情報で収集・分析・管理、AIエージェント技術を実証」という見出しはLedge.aiが2026年8月28日に掲載した記事である。確認できた公開情報の範囲では、独立した発表主体・媒体としての「Deep Research」による一次発表は見当たらず、**一次情報はSakana AIおよび防衛省・防衛装備庁の調達文書**と見るのが妥当である。 [1]

本件で最も重要なファクトは、通常の「総合分析業務」が公開情報だけでなく電波情報や画像・地理情報等を含む「あらゆる情報源」を扱う一方、**今回のAI実証でAIに投入する対象情報は「公開情報（有料・無料は問わない。）」に明示的に限定されている**ことである。さらに調達説明書は、契約履行期間を通じ「保護すべき情報」の取扱いを**書面のみ**とすると明記している。したがって、現時点で「Sakana AIの生成AI・エージェントが防衛省の機密情報や特定秘密をクラウド上で直接解析する」と評価する根拠はない。 [2]

技術面では、防衛省は提案時点で、言語モデル、エージェント型・推薦型等のAI方式、モデル規模、シングル／マルチモーダル、オンプレミス／クラウド／ハイブリッド、ハードウェア構成、評価方法まで比較検討対象とした。しかし、**採択されたSakana AI案の具体的なモデル名、LLMベンダー、Fugu/Marlinの採用有無、GPU、クラウド事業者、ベクトルDB、クローラ、検索API、RAG実装、データ保持期間等は公開されていない**。 [3]

一方、Sakana AIには本件と極めて近い公開技術実績がある。2026年7月には、防衛・インテリジェンスチームが公開情報分析競技「DIVER OSINT CTF 2026」で、**Fugu-ultra 1.1を使ったOSINT分析エージェント**を構築して850超のチーム中5位となった。Fugu自体は、複数のフロンティア基盤モデルからタスクごとに適切なモデル・役割を動的に選ぶマルチエージェント・オーケストレーション方式であり、その基礎研究TrinityではThinker／Worker／Verifierの役割を軽量コーディネータが割り当てる。Marlinでは長時間の仮説生成、Web探索、情報収集、矛盾解消を自律反復し、総務省関連事業ではWeb/SNSをAIエージェントが検索し、複数ソースを提示する自動ファクトチェックも実装している。**これらは本件システムの技術候補を推定する強い材料だが、「防衛省案件で実際に同じ構成を使う」と公表された事実ではない。** [4]

運用面では、当初計画では契約後速やかに実証モデルを設計・開発し、**2027年度開始前、すなわち2027年3月末までを目安に官側が試用**、フィードバックに基づき改修する。その後も技術動向調査と技術・コスト・法制面の検討を継続し、**2027年度末までに最終成果を報告**する計画である。ただし説明書作成時には契約締結を2026年5月下旬〜6月上旬と想定していたのに対し、実際の契約日は7月29日であるため、個別マイルストーンが後ろ倒しされた可能性はあるものの、変更後の日程は公開されていない。 [5]

本件の本質的な価値は単なる「ChatGPT型の文章生成」ではなく、**OSINTを継続的に発見→収集→出所管理→重複排除→検索→複数モデルで検証→矛盾検出→引用付き分析→人間の分析官が承認**する一連のインテリジェンス・ワークフローをどこまで自動化・高信頼化できるかにある。逆に最大のリスクは、ハルシネーションそのものより、①敵対的情報操作を「大量の独立ソース」と誤認すること、②エージェントによる

Web探索が偽情報・プロンプトインジェクションに汚染されること、③検索・閲覧履歴や分析テーマ自体から防衛省の関心事項が推測されること、④モデル/API供給者への依存、⑤AIの高い流暢性による自動化バイアスである。このため、本格運用の合否は「回答精度」だけでなく、**出典追跡可能性、独立ソース数、時点整合性、引用忠実度、敵対的入力耐性、監査ログ、人間による覆し率**まで測る必要がある。防衛省自身のAI活用推進基本方針も、偽・誤情報、バイアス、人間による制御、ログ保存、データ出所の追跡可能性を明示的に重視している。 [6]

一次情報・関連報道のファクトチェック

公開情報を時系列で整理すると、今回のニュースは「8月下旬に突然始まった案件」ではなく、2026年3月の公募から選定・契約を経て8月に広く報道された案件である。防衛装備庁の説明書は、情報量の増大によって人間だけでは網羅的収集、迅速・正確な分析、過去の情報・分析結果の資産化と再利用に限界が生じていると問題設定し、市販の汎用生成AIについても情報の確度・鮮度・網羅性、ハルシネーション、データバイアスのため、そのままでは十分に信頼できないと明記していた。 [7]

区分	発表・掲載日時	発表者・媒体	原文の位置付け	URL
防衛装備庁・公募説明書	公募期間 2026年3月6日〜3月23日	防衛装備庁／提出・問合せ先は防衛省情報本部 分析部員	**最重要一次資料**。目的、実証対象、AI方式候補、スケジュール、セキュリティ条件を規定	`https://www.mod.go.jp/atla/soubiseisaku/newentry/startup_proc_r080306.pdf`
防衛装備庁・契約結果	公表時刻はPDF上に明記なし／契約日 2026年7月29日	分任支出負担行為担当官・防衛装備庁調達事業部総括装備調達官 浅見智宏	**契約の一次資料**。Sakana AI、契約日、3億2,903万2,000円を確認	`https://www.mod.go.jp/atla/soubiseisaku/newentry/startup_contr_r080806.pdf`
Sakana AI公式	2026年8月24日、時刻記載なし	Sakana AI株式会社	**受注側の一次発表**。「高度なAIエージェント技術」「情報本部」「戦略的インテリジェンス」を明示	`https://sakana.ai/defense-integrated-analysis/`

区分	発表・掲載日時	発表者・媒体	原文の位置付け	URL
ビジネス+IT	2026年8月24日 18:00	ビジネス+IT	主要日本語報道。AIエージェントによる自律収集・クロスチェック等を報道。ただし一次資料より具体的な部分は二次情報として扱うべき	`https://www.sbbit.jp/article/cont1/186660`
Ledge.ai	2026年8月28日、時刻記載なし	Ledge.ai編集部	ユーザー提示の**完全一致する見出し**	`https://ledge.ai/articles/sakana_ai_defense_information_analysis`
Sakana AI・OSINT競技	2026年7月30日	Sakana AI株式会社	本案件に非常に近い隣接実績。Fugu-ultra 1.1によるOSINT分析エージェント	`https://sakana.ai/diver-osint/`
Sakana AI・DEEP DIVE	2026年5月29日	Sakana AI株式会社	衛星画像等のオープンソースデータと安全保障専門知を持つDEEP DIVEとの情報分析提携	`https://sakana.ai/deep-dive-partnership/`

各日時・契約金額・担当者等は、防衛装備庁、Sakana AI、各媒体の公開ページから確認できる。 [8]

「Deep Research発表」の扱い。 今回確認できた範囲では、引用された見出しそのものは「Deep Research」という媒体ではなくLedge.aiの記事である。したがって本レポートでは、ユーザーの「@Deep Research」を調査指示・テーマ指定として解釈し、一次発表をSakana AI、防衛省側の原典を防衛装備庁調達文書と位置付けた。Ledge.aiページでは2026年8月28日とLedge.ai編集部が確認できる。 [9]

実証範囲を理解するうえでの最重要な区別は次の通りである。

項目	公開資料上の範囲	判定
通常の「総合分析業務」	公開情報、電波情報、画像・地理情報等「あらゆる情報源」から収集・分析・管理	確定
今回のAI実証への入力	「公開情報（有料・無料は問わない。）」	**確定**

項目	公開資料上の範囲	判定
特定秘密をAIへ投入	記載なし	**確認できない**
「保護すべき情報」の電子処理	全契約期間、「書面のみ」と明記	**今回については否定的**
戦術C2・ドローン制御	2026年3月の別契約	**今回とは別案件**
戦略インテリジェンス支援	情報本部の今回の案件	**確定**

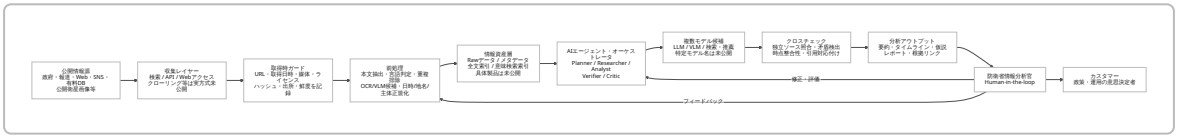
防衛省の説明書は総合分析業務そのものを広く定義する一方、実証対象を公開情報に限定しており、Sakana AIも3月案件を部隊・センサー・C2、今回を国家レベルの安全保障政策を支える戦略インテリジェンスと明確に区別している。 10

なおビジネス+ITは、人工衛星画像や公開情報をAIエージェントが自律収集しクロスチェックしてレポートを生成すると報じ、さらに外国製AIと国産AIのハイブリッド構想にも言及している。しかし、後者のモデル配置やサーバ構成等は**今回の調達説明書・契約結果・Sakana AI公式発表では確認できない**。したがって本レポートでは「報道された構想・解釈」とし、確定した実装仕様とは扱わない。 11

技術構成・AIエージェント・データフロー

公開資料から確実に言えるのは、防衛省が「市販の生成AIをそのまま使う」提案よりも、独自性・優位性のある技術を求めていることである。評価基準では、スケールしたTransformer等の既存技術に対し、たとえば**計算コストを抑えながら精度を維持すること、ハルシネーションを局限すること**を優位性の例として明示している。また実証提案には、言語モデル、エージェント型・推薦型、規模、モダリティ、オンプレミス／クラウド／ハイブリッド、ハードウェア資源、効果、評価方法の具体化が求められた。 12

ただし、**採択されたSakana AIの技術提案書および確定仕様書は公開されていない**。したがって以下の図は、①防衛省が明記した要求、②Sakana AIが公開済みの関連技術、③一般的にこの要件を満たすために必要となる構成を重ねた「推定参照アーキテクチャ」であり、実システムの設計図ではない。



この構成で特にSakana AIらしい候補となるのが**マルチエージェント・オーケストレーション**である。Fugu は、一つの巨大モデルに全処理を委ねるのではなく、問題に応じてモデルプールからモデル・サブタスク・協調方法を動的に選択する。基礎技術Trinityは、軽量コーディネータが各ターンでLLMに「Thinker」「Worker」「Verifier」の役割を付与し、基盤モデル自体の重みを変更せずに推論時に組み合わせる方式である。Trinityのコーディネータは2万未満の学習可能パラメータで構成されるとSakana AIは説明している。 13

さらにFugu/Conductorでは、オーケストレータ自身を再び呼び出して前回出力を再評価し、必要なら修正ワークフローを起動する**recursive test-time scaling**が公開されている。これは「より巨大な単一モデルへ毎回投げる」のではなく、難問だけ検証回数や協調回数を増やす設計が可能であり、防衛省が求めた「計算コストを抑えつつ精度維持」「ハルシネーション局限」と方向性がよく一致する。ただし、一致はあくまで設計上の親和性であり、Fuguが防衛省実証の本番オーケストレータだとする公式確認はない。 14

より直接的な状況証拠は、契約直前の2026年7月30日に公表されたOSINT競技である。Sakana AIの防衛・インテリジェンスチームはFugu-ultra 1.1を使ったOSINT分析エージェントを構築し、850超のチーム中5位を記録した。同社自身が、その知見を「実務向けにさらに改善し、日本の防衛・インテリジェンスへの貢献につなげる」としているため、**Fugu系技術が今回の研究開発に技術的知見として流入する可能性は高い**と推定できる。しかし、契約上の採用を証明する記述ではない。 [15]

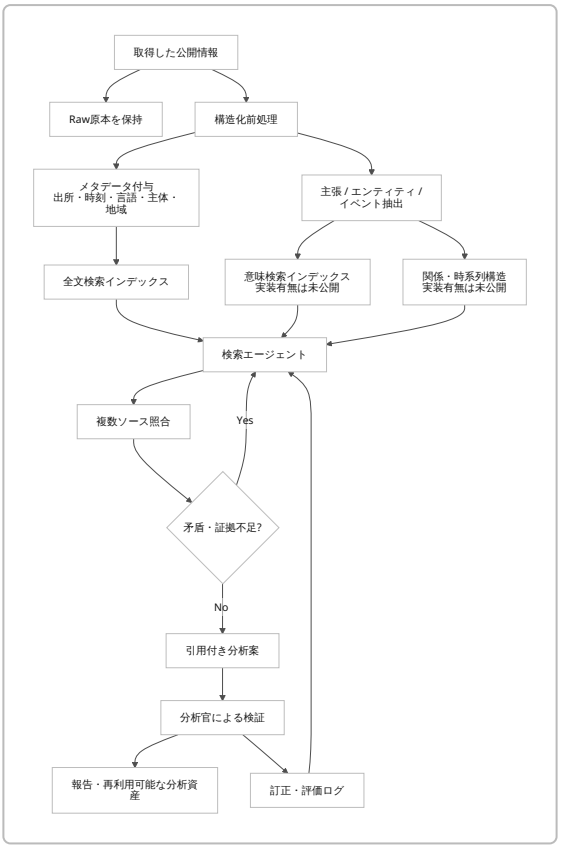
Sakana Marlinも関連性が高い。MarlinはSakana AIが「Ultra Deep Research」と呼ぶ自律型リサーチエージェントで、テーマ設定後、仮説形成、情報収集、Web探索、矛盾解消を反復し、最大8時間の連続的な自律推論を行うとしている。これは防衛情報分析に必要な「一発回答」ではない長期調査型エージェントの実装例である。ただしMarlinは企業戦略リサーチ向け公開製品であり、**防衛省案件でMarlinそのものを利用するとの発表はない**。 [16]

同社が総務省関連事業で開発した偽・誤情報対策技術も重要である。そこでは、投稿から検証可能な主張を抽出し、AIエージェントがWeb/SNSを自律検索し、仮説検証を繰り返して、結論までの論理と複数の裏付けソースを示す仕組みを実装した。また画像・動画については生成・加工検知や逆画像検索も行っている。したがって、**「収集→主張抽出→検索→多源照合→根拠提示」という技術要素をSakana AIが既に持つことは確認できる**が、防衛向け実装と同一コードであることまでは分からない。 [17]

公開情報の収集方式については、ユーザー指定の「クローリング、API、スクレイピング等」のどれを実際に用いるかは**未公開**である。防衛省文書が定めるのは入力情報の種類までで、収集プロトコルやWebブラウザ自動操作、検索API、RSS、SNS API、有料データベースAPI、独自クローラの有無を開示していない。Sakana AIの別案件ではWeb/SNSの自律検索、MarlinではWeb navigationが明示されているため、検索APIやブラウザ型ツールをエージェントが使用する方式は技術的に十分可能だが、本件について断定できない。 [18]

前処理・ラベリング・保存・検索も本件固有仕様は未公開である。一方、防衛省のAI活用推進基本方針は、防衛省全体のデータ基盤について、メタデータの一覧化、作成時のタグ付け・カタログ化、データフォーマット整備、クラウドを用いた収集・蓄積・管理、データレイクとデータウェアハウス、学習用データへのタグ付け、再学習と学習済みモデル配布のサイクルを明示している。したがって、本実証の設計を評価する際にも、**URL・取得時刻・発信主体・言語・地域・情報種別・信頼度・権利条件・元データハッシュ等をメタデータ化する設計**が防衛省の既存方針と整合的である。ただしこれは「推奨／推定される構成」であって、本契約にそのまま実装されたことを示す資料ではない。 [19]

検索層については、文字列検索だけでは表記揺れ・多言語・関連概念を捉えにくいため、技術的には**全文検索＋埋め込みベクトル検索＋メタデータ絞り込み**のハイブリッド検索が合理的である。また人物・組織・地名・装備・イベント・日付の関係を長期追跡するなら知識グラフが有力である。しかし、Elasticsearch/OpenSearch等の検索エンジン、特定のベクトルDB、Graph DB、Embeddingモデルの採用はすべて**未公開**である。



学習・推論環境についても、採用構成は未公開である。調達側自身がオンプレミス、クラウド、ハイブリッドを比較対象にしており、つまり公募時点で特定方式に固定していなかった。Sakana Fuguの公開アーキテクチャは、外部の複数基盤モデルを推論時に動的協調させる方式で、Trinityは基盤モデルの重み変更を必要としない。そのため、本実証でも「巨大な防衛専用LLMを一から学習」するより、既存モデル＋軽量オーケストレータ＋検索・検証ツールを組み合わせる**推論時オーケストレーション中心の方式**は合理的な候補であるが、学習データ、Fine-tuning、LoRA、RL、GPU種別等はいずれも未公開である。 20

可視化についても本件固有UIは未公開である。情報分析官向けにはタイムライン、地図、主体間ネットワーク、原文とAI分析の並列表示、情報源ごとの信頼度・更新日時、矛盾ソースの対照表示などが有効と考えられる。Sakana AIは別の総務省関連事業でSNS空間の分析・可視化やユーザー／反応分析を実装しているため、技術能力は確認できるが、防衛省UIの採用仕様とは区別すべきである。 17

> **図版挿入推奨箇所**：この節では、防衛装備庁の契約結果PDFの1ページ画像、Sakana Fugu公式ページのオーケストレーション概念図、Sakana AI「偽・誤情報対策」公式ページのファクトチェック図を並べると、「契約上の確定事項」と「Sakana AIの公開技術資産」を視覚的に区別しやすい。各図版は転載条件を確認し、原ページへの出典を併記すべきである。 21

運用範囲・セキュリティ・ガバナンス

本件の利用主体は防衛省・自衛隊であり、具体的な実務窓口は調達説明書上「防衛省情報本部 分析部員」である。Sakana AIも、今回の事業で情報本部が担う「総合分析業務」の抜本的強化を目指すとしている。分析官は収集・分析・管理した情報から報告書等を作成し、その成果を政策・運用の意思決定者、調達文書でいう「カスタマー」に提供する。したがって現在の公開スコープは、AIが最終的な国家意思決定を自動で下すシステムというより、**分析官の情報処理・分析・知識管理を支援する意思決定支援系**である。 22

Sakana AIは、本件を3月のC2研究とは明示的に切り分けている。3月案件ではドローン等の陸海空データ、マルチモーダル統合、小規模視覚言語モデルによるエッジ処理を扱うが、今回の案件は安全保障政策全体を支える**戦略インテリジェンス**である。したがって「Sakana AIが今回、AIで兵器を自動制御する」という理解は公開情報からは支持されない。 [23]

実証スケジュールはかなり具体的である。説明書では、契約後速やかにモデル設計・開発、2027年度開始までに官側試用、官側フィードバックに基づく改修、契約期間全体を通じた国内外AI技術の動向調査、2026年度後半から技術・コスト・法制上の課題と対策の検討、2027年度末に成果報告書提出としている。実際の契約日は当初想定より遅い2026年7月29日であるため、個別工程がどの程度再設定されたかは**未公開**である。 [5]

人的監督は、制度レベルでは明確に想定されている。 調達説明書自体が「官側に試用させ、官側のフィードバック等を基に改修」としており、防衛省AI活用推進基本方針もAIへの過度な依存・自動化バイアスを警戒し、必要に応じて人間が制御可能な状態を確保し、適切なタイミングで人間の判断を介在させるよう求めている。もっとも、「どの種類の分析を誰が承認するか」「AIの出力を何級の職員がサインオフするか」「二名確認が必要か」といった具体的なHuman-in-the-loop手順は未公開である。 [24]

アクセス制御についてもプロジェクト固有のRBAC/ABAC、MFA、端末認証等は未公開である。ただし防衛省のAI方針では、情報の保全区分を考慮し「誰がどのデータにアクセスできるのか」というルールの明確化、不正アクセス・改ざん防止を求めている。防衛装備庁の調達情報セキュリティ基準は、「保護すべき情報」を扱う契約について組織的・物理的・情報システム上の対策を要求し、電子システムで扱う場合はシステムセキュリティ実装計画書を提出して官側確認を受ける仕組みを設けている。 [25]

今回についてはさらに重要な例外があり、「保護すべき情報」の取扱いは**契約履行の全期間を通じて書面のみ**とされる。防衛装備庁がいう「保護すべき情報」は、契約関係情報のうち「部内限り」または「注意」に該当し、官が指定した情報をいう。これは「特定秘密」と同義ではない。したがって公開資料からは、Sakana AIの実証システムが秘密指定データを電子的に保持しているとは評価できない。 [26]

一方、安全保障の観点では、**公開情報しか入っていないからシステム自体も低機密とは限らない**。これは公開資料から導かれる分析上の推論である。たとえば検索語、継続監視対象、アラート設定、分析官の関心、未公表の仮説、どの人物・組織をどの頻度で調べるか、といった「利用メタデータ」を見れば、防衛省のインテリジェンス上の優先順位を推測できる可能性がある。したがって原文データの分類だけでなく、**クエリ、プロンプト、検索履歴、エージェントのツール呼び出しログ、生成途中の仮説**も保護対象として設計するのが合理的である。これは、防衛省がAIの入力・推論・判断根拠をログ化しつつ、同時に機密性・完全性・可用性を確保するよう求めていることとも整合する。 [6]

ログ管理について、本件固有の実装・保存期間は未公開である。しかし防衛省の基本方針は、検証可能性のためAIの開発過程、利用時の入出力、学習プロセス、推論過程、判断根拠等のログを記録・保存し、技術特性・用途に応じて記録方法、頻度、保存期間を決めること、さらにデータの出所や開発・利用中の意思決定を追跡可能にすることを求めている。つまり、監査ログを持たない「ブラックボックス型Deep Research」は、防衛省自身の政策方向とは整合しにくい。 [27]

本実証で望ましいログ構造は、分析上、最低でも以下を相関可能にすることである。

```
Case ID → Analyst ID → Prompt/Task → Agent plan → Search query → Tool/API call → Retrieved URL → Retrieval timestamp → Content hash → Model/version → Model output → Citation mapping → Verifier result → Human edits → Final approval
```

これにより、後日「なぜこの結論になったか」「当時どの情報を見たのか」「モデル更新後に結論が変わったのか」を再現できる。防衛省の方針が要求する出所・判断の追跡可能性を、エージェント時代に具体化した設計といえる。 [27]

データ保持方針については、保存期間、削除方法、バックアップ、災害復旧、モデル提供企業によるAPI入力の二次利用禁止、国外保存の可否などはすべて**未公開**である。この点は運用上の主要な透明性ギャップである。防衛省全体の方針ではデータレイク、DWH、メタデータ管理、アクセスルール、再学習サイクルを構想しているものの、それが今回の実証にどの範囲まで適用されるかは公開されていない。 [28]

公開／未公開をまとめると次の通りである。

項目	状況	コメント
契約日	**公開**	2026年7月29日
契約額	**公開**	329,032,000円
実証入力	**公開**	有料・無料を問わない公開情報
利用部門	**公開**	情報本部の総合分析業務
官側試用	**公開**	2027年度開始前までを当初計画
最終報告	**公開**	2027年度末
具体的LLM	**未公開**	Fugu利用も契約資料では未確認
外部LLM API	**未公開**	ベンダー・データ所在地とも不明
クローラ/API/スクレイパー	**未公開**	公開技術から類推は可能
RAG/Vector DB	**未公開**	製品・Embedding・Chunking等不明
クラウド／オンプレ	**未公開**	公募では比較対象
GPU・推論基盤	**未公開**	型式・台数・所在不明
RBAC/MFA等	**未公開**	防衛省一般方針は存在
ログ保存期間	**未公開**	一般方針は用途別に決定するよう要求
データ保持期間	**未公開**	削除・バックアップ規則も不明
第三者監査	**未公開**	公開資料で確認できず
特定秘密のAI処理	**確認されない**	実証入力は公開情報、保護情報は書面のみ

以上は防衛装備庁・Sakana AI一次資料と防衛省AI方針から整理した。 [29]

法規制・倫理・安全保障リスク

法的には、「公開情報だから自由に何でも自動収集・永久保存できる」という理解は適切ではない。今回の調達文書自体も、技術面・コスト面だけでなく**「法制面等」に関する課題の洗い出しと処置・対策の導出**を提案評価の対象としている。 [30]

個人情報保護法。 公開Web、SNS、報道記事等に氏名、顔、所属、位置、行動履歴等が含まれていれば、「公開情報」という取得元の属性と「個人情報／保有個人情報」であるかは別問題となる。個人情報保護委員会の行政機関等向けガイドラインでは、行政文書等に記録された「保有個人情報」には安全管理措置（法66条1項）、利用・提供の制限（法69条）等が適用されると整理されている。したがって、OSINTを人物単位に名寄せし、長期間検索可能なプロフィールや行動グラフとして蓄積する場合は、収集目的・必要性・保存範囲・アクセス制御を慎重に設計する必要がある。 [31]

個人情報保護法（e-Gov）：`https://laws.e-gov.go.jp/law/415AC0000000057` [32]

個人情報保護委員会・行政機関等編ガイドライン：`https://www.ppc.go.jp/personalinfo/legal/guidelines_administrative/` [33]

著作権。 Webページ、新聞・雑誌記事、有料データベース、画像等を機械的に取得・複製・保存する場合、著作権とライセンス契約の双方を検討する必要がある。文化庁は著作権法30条の4について、著作物に表現された思想・感情の享受を目的としない情報解析等は必要な限度で利用できる枠組みを説明している一方、権利者の利益を不当に害する場合等まで無制限に許容する規定ではない。特に有料情報源については、著作権法上の権利制限とは別に契約条件・API利用条件・データベース利用条件の確認が必要になる。 [34]

著作権法（e-Gov）：`https://laws.e-gov.go.jp/law/345AC0000000048/` [35]

文化庁「AIと著作権について」：`https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html` [36]

自動収集の法的境界。 通常アクセス可能な公開ページを収集する行為と、認証を突破したり他人の認証情報を利用したりする行為は区別する必要があり、後者には不正アクセス行為の禁止等に関する法律が関係する。また、robots.txtはそれ自体が一般的な法律上の許可・禁止を決めるものではないものの、サイト規約、契約、負荷、API条件を含めた収集ポリシーを実装すべきである。今回、クローリング・API・スクレイピング方式の具体仕様は公開されていないため、法的適合性も外部からは評価できない。 [37]

不正アクセス行為の禁止等に関する法律：`https://laws.e-gov.go.jp/law/411AC0000000128` [38]

秘密・機微情報。 特定秘密の保護に関する法律は、防衛等に関する情報のうち指定された特定秘密を保護する制度である。しかし今回の公開されたAI実証要件は公開情報に限定され、「保護すべき情報」も電子システムではなく書面のみとされるため、**現時点のAIシステムが特定秘密を取り扱うとみなすべきではない**。将来、実証成功後に秘密情報・電波情報・非公開画像等へ対象を拡張するなら、現行実証とは別段階のセキュリティ認証、ネットワーク分離、モデル/APIのデータフロー再設計、取扱者管理が必要になる。 [26]

特定秘密の保護に関する法律：`https://laws.e-gov.go.jp/law/425AC0000000108` [39]

倫理的には、**誤情報とバイアスが最大の設計課題として既に調達仕様に織り込まれている**。防衛装備庁自身が、汎用生成AIには情報の確度・鮮度・網羅性、ハルシネーション、データバイアスの問題があり、信頼できるものとして使えない危険性があると明記している。防衛省AI方針も、自動化バイアス、フィルターバブル、生成AIによる偽・誤情報・偏向情報を具体的リスクとして掲げる。 [40]

安全保障用途ではさらに、通常の企業向けDeep Research以上に**敵対的情報環境**を前提とする必要がある。たとえば敵対主体が同じ虚偽ナラティブを多数のサイト・SNSアカウント・転載メディアへ流せば、「10件の情報源で一致」と数えても実際には単一の情報操作キャンペーンという場合がある。このため単純な「ソース数」ではなく、**情報の系譜・相互引用関係・一次発信者の独立性**を評価する必要がある。NATOの改訂AI戦略もAIによる偽情報・情報作戦を安全保障上の懸念として明示している。 [41]

エージェント固有のリスクとしては、Webページ内の悪意ある指示をモデルが命令と誤認するプロンプトインジェクション、検索結果のSEO汚染、悪意あるPDF・HTML、ツール呼び出しによる不要な外部通信、外部モデルAPIへのクエリ流出等がある。この種のリスクに対しては、Web内容を「命令」ではなく「非信頼データ」として扱う分離、ツール権限の最小化、許可されたドメイン/APIだけを通すネットワーク制御、モデルから直接資格情報へアクセスさせない仕組み、重要操作の人間承認が必要となる。防衛省のAI方針も外部サービスを含む推論時参照情報をバイアス要因として捉え、AIへの外部攻撃に対応する必要性を明記している。 42

説明可能性については、LLM内部の全ニューロンを説明することより、「どの資料のどの部分から、どの中間判断を経て、その結論になったか」を追跡できる**プロセス説明可能性**の方が本件には実務的である。防衛省自身も、入力・出力・推論過程・判断根拠のログとデータ出所の追跡可能性を求めているため、生成テキストにURLを付けるだけでは不十分で、引用箇所と主張の対応、取得日時、文書版、モデル版まで固定する必要がある。 27

リスク緩和策を設計レベルで整理すると、次のようになる。

リスク	必要な緩和策
ハルシネーション	全重要主張に根拠、Verifierエージェント、引用と原文の含意一致評価、人間承認
古い情報	取得日時・発生日の分離、鮮度SLA、時間を考慮した検索ランキング
偽情報・情報工作	相互独立性評価、一次情報優先、ナラティブの発信源クラスタリング
ソース偏重	言語・国・政治的立場・媒体種類別のカバレッジ測定
自動化バイアス	AI結論を初期値にしないUI、反証材料の強制提示、分析官の異議記録
プロンプトインジェクション	外部コンテンツを非信頼領域に隔離、tool allowlist、サンドボックス
APIへの情報流出	データ分類別ルーティング、外部API禁止領域、ゼロ保持契約・国内処理の検証
モデル供給網	複数モデル対応、モデル変更時の回帰試験、SBOM/Model BOM
個人情報過剰蓄積	目的・必要性に基づく最小化、保持期限、アクセス監査
説明不能	出所・時刻・モデル・検索・中間判断・人間修正を一体的に記録

これらは防衛省の人間中心、安全性、公平性、プライバシー、セキュリティ、透明性、アカウンタビリティの方針を、エージェント型OSINTへ具体化した提案である。 6

技術・運用評価と信頼性検証

技術的利点は明確である。第一に、AIエージェントは分析官が手作業では追跡しきれない量の公開情報を並列探索できる。第二に、異なるモデルや検索ツールを役割分担させることで、単一LLMへの依存を減らせる可能性がある。第三に、取得情報・過去の分析結果を資産として体系化すれば、担当者の異動や経験差による属人性を減らし、過去の判断根拠を再利用しやすくなる。これらは防衛装備庁が掲げるスピード、精度、効率、ヒューマンエラー削減、省人化、情報資産管理という課題に直接対応する。 7

Sakana AI側の技術的適合性も高い。Fuguは複数モデルを問題ごとに動的協調させ、TrinityではThinker/Worker/Verifierという明示的な分業が可能である。OSINT CTFでも同社の防衛チームが実際にFugu-ultra 1.1

をOSINTエージェントへ適用している。したがって、「汎用チャットAIではなく、複数の調査・検証エージェントを束ねる」という調達思想とSakana AIの公開研究テーマには強い整合性がある。 [43]

ただし限界も大きい。最も重要なのは、「人間より大量に情報を見る」ことと「真実を知る」ことは同じではない点である。公開情報そのものが不完全・意図的に操作されていれば、AIは誤った証拠を高速に統合するだけになる。また、エージェントが多数の検索を行うほど、低品質情報や悪意ある入力に触れる表面積も広がる。防衛省が汎用生成AIの確度・鮮度・網羅性・ハルシネーション・バイアスを最初から問題視しているのは、この点を認識しているためである。 [7]

さらにSakana Fugu型の複数モデル利用は「一つのモデルの弱点を別モデルで補う」という利点を持つ一方、複数モデルが**同じ学習データや同じ検索結果に依存していれば、誤りも相関する**。三つのLLMが同意したことを「三つの独立証拠」と数えてはならない。信頼度は「モデル投票数」ではなく、外部証拠の独立性と品質から推定すべきである。

したがって評価指標は単純な「生成文の正解率」だけでは足りない。本実証で公開すべき最小評価セットとして、以下を提案する。

評価領域	推奨指標	意味
情報収集	Recall@K、重要ソース捕捉率	必要情報を取りこぼしていないか
鮮度	発生→取得までの中央値/P95	最新情勢への追随速度
情報源	一次情報比率、独立ソース比率	まとめサイトの量産を「裏付け」と誤認していないか
検索	nDCG、MRR、Precision@K	関連資料を上位へ出せるか
主張抽出	Precision/Recall/F1	原文から事実を正しく構造化できるか
事実性	factual precision	AIが述べた事実の正確率
引用品質	citation precision / entailment	引用先が本当に主張を裏付けるか
ハルシネーション	unsupported-claim rate	根拠のない主張の割合
矛盾検出	contradiction recall	相反する情報を見落とさないか
信頼度	Brier score / calibration error	「80%確実」が本当に約80%当たるか
バイアス	国・言語・媒体別の性能差	特定地域・言語だけ精度が低くないか
業務効果	分析所要時間、分析官工数	実際の省力化
品質	人間の訂正率・重大訂正率	出力をどれだけ直す必要があるか
Human-AI	override rate、採用率	人間がAIをどの程度覆したか
セキュリティ	攻撃成功率、情報流出テスト	Prompt injection等への耐性
監査	provenance completeness	全重要主張を出所まで遡れるか
コスト	調査1件当たり推論費用	高精度化が持続可能か

特に**引用忠実度**と**独立ソース性**は安全保障用途のDeep Researchで一般的なQAベンチマーク以上に重要である。「正答を知っている問題」だけで評価すると、実運用の未知の事象での性能を過大評価しやすい。

評価方法は三段階に分けるべきである。まず、過去の案件を使い、その時点までに公開されていた情報だけをAIへ与える**時点固定バックテスト**を行う。次に、現役分析官と並走させるがAI結果を実際の報告には使わない**shadow mode**で、人間とのカバレッジ、所要時間、誤りを比較する。最後に限定された業務でHuman-in-the-loop運用を開始し、重大判断についてAI単独の自動送信を禁止したまま、効率向上とエラー増加のトレードオフを測るべきである。防衛省の基本方針が、人間による制御、ログ、追跡可能性、アジャイルなガバナンスを掲げることとも整合する。 6

第三者監査について、今回の公開資料では実施の有無を確認できない。このため透明性確保策として、機密に触れない範囲で、①評価指標の定義、②ベンチマーク作成原則、③モデル/API構成のカテゴリ、④重大インシデント件数、⑤ハルシネーション率・引用忠実度、⑥レッドチームの方法論、⑦人間の最終決定権、⑧モデル変更管理、を定期的に外部有識者へ提示する方式が望ましい。

完全なコードや分析対象を公開する必要はない。安全保障分野では透明性と秘匿性は両立可能であり、「何を見ているか」を公開せずに、**どの基準でAIを検証しているか**を公開することはできる。NATOもResponsible AIの実装に、品質管理、リスク緩和、governability、traceability、reliabilityを具体的なチェックへ落とす方向を進めてきた。 44

防衛・民間への波及と政策提言

国内への直接的な影響は、防衛分野で「基盤モデルそのものを一社から買う」のではなく、**AIエージェント、検索、検証、オーケストレーション、データ管理を一体として調達・実証するモデル**が成立しつつある点にある。防衛装備庁は今回、価格より技術的要素を重視するスタートアップ技術提案評価方式を適用し、研究・開発段階を含む国内外技術の調査まで求めた。これはモデル世代交代が速いAIを、固定仕様・長期調達だけで扱う難しさへの対応と評価できる。 45

Sakana AIが防衛・インテリジェンスを明示的な重点分野とし、DEEP DIVEと衛星画像等を含むオープンソース分析で提携し、OSINT競技でもFuguを投入していることから、日本でも**OSINT専門家＋安全保障ドメイン専門家＋AI研究者＋ソフトウェアエンジニア**を一体化するエコシステムが形成され始めていると評価できる。ただしDEEP DIVEが今回の防衛省契約の下請け・データ提供者であるとの公開情報はなく、両者を直接結び付けるべきではない。 46

国外の流れとも一致する。米国防総省は2023年のData, Analytics and Artificial Intelligence Adoption Strategyで、質の高いデータを基盤にAI導入の速度・俊敏性・責任ある利用を組織全体へ拡大する方針を掲げた。NATOも2024年の改訂AI戦略で、AIの責任ある防衛利用と実装加速を同時に掲げ、AIを使った偽情報・情報作戦を安全保障リスクとして扱っている。つまり日本の今回の案件は、世界的な「AIによる意思決定支援」と「Responsible AI／検証可能性」を同時に進める流れの国内版と位置付けられる。 47

民間への示唆も大きい。金融、経済安全保障、サプライチェーン、企業リスク、M&Aデューデリジェンス、不正調査などでは、防衛情報分析と同じく「情報量が多い」「正誤が混在」「時点が重要」「根拠を提示する必要がある」という条件がある。Sakana Marlinが企業戦略向け長時間Deep Researchを提供し、Fuguが複数モデルをオーケストレーションする構造は、防衛以外でも「単一LLMチャット」から「検証可能な調査工程」へ移行する方向を示している。 48

一方で、安全保障案件がモデル/API供給網に依存する場合、性能だけでなく**データ主権、モデル更新権、停止時の代替性、国外APIへの依存、サプライチェーン透明性**が重要になる。Fuguのようなモデルオーケストレーションは、理論的には特定モデルへのロックインを弱めやすいが、実際に防衛システムがどの外部モデ

ルを利用するかは未公開であり、国産AIだから全処理が国内モデル・国内データセンターで完結すると推定してはならない。 [49]

短期政策としては、2027年3月までの官側試用に合わせ、評価と監査を先に固定するべきである。 具体的には、公開情報限定という現行境界を維持しつつ、原文・URL・取得時刻・ハッシュを必須保存し、全重要主張を引用に結びつけること、独立ソース性を評価すること、外部Webを非信頼入力として扱うこと、外部モデル/API一覧とデータフローを官側が完全把握すること、人間の最終承認を必須化することが重要である。これは防衛省が既に掲げる人間中心、透明性、ログ、追跡可能性を具体化する措置である。 [6]

中期政策では、2027年度末までの実証結果を「特定ベンダーの製品評価」だけで終わらせず、情報分析AIの共通基盤仕様へ昇華すべきである。 モデルを交換可能にする標準インターフェース、データ・モデル・プロンプトのバージョン管理、共通OSINTベンチマーク、Model BOM、API依存関係、セキュリティ試験、データ保持区分、監査ログ形式を標準化すれば、Fuguであれ別方式であれ、将来のモデル世代交代に対応しやすくなる。防衛省は既にデータ形式の標準化、メタデータ・カタログ化、組織横断的データ管理を方針として掲げており、その延長線で実施できる。 [28]

同時に、実証結果のうち安全保障上公開できる部分については、少なくとも「従来業務に対して何%時間を短縮したか」「重要情報のRecallがどう変化したか」「引用忠実度」「重大誤り率」「人間の修正率」といった**性能のカテゴリ別指標**を公開することが望ましい。契約額だけが公開され性能が完全非公開では、費用対効果の民主的検証が困難になる。一方、具体的な監視対象や情報源リストを公開する必要はない。

長期政策では、公開情報から秘密・機微情報へ適用範囲を広げる場合に、今回の実証システムを単純に横展開しないことが重要である。 公開OSINTと秘密情報では脅威モデルが根本的に異なる。秘密情報を扱う段階では、ネットワーク分離、認定済みオンプレ／専用クラウド、外部API禁止または厳格なゲートウェイ、モデル重み・コンテナ・依存ライブラリの供給網管理、担当者認証、出力持ち出し管理、インシデント対応等を新たに設計・認証すべきである。現在の調達が公開情報と書面の保護情報に限定されていることは、この段階的なアプローチの合理性を示す。 [26]

最終的には、AIの役割を「答えを出す機械」ではなく、**情報分析官の探索範囲を拡張し、反証を探し、情報の来歴を整理し、人間がより良い判断をするための検証可能な共同分析者**として定義することが最も重要である。防衛省自身のAI方針も、AI活用そのものを目的化せず、人間が課題を特定し、その解決のためにAIを利用するという基本姿勢を明記している。 [27]

本件が成功したかを判断する最終的な基準は「AIが賢かったか」ではなく、**同じ人員・時間の下で、分析官がより多くの重要情報を発見し、誤情報をより早く排除し、判断根拠をより明確に説明でき、しかも情報漏えいや自動化バイアスを増やさなかったか**である。

参考資料

以下は、本調査で特に重要度が高い資料である。一次資料を先に、その後にSakana AI関連技術、法令・政策、海外比較資料を配置した。

重要度	資料	URL
最重要	防衛装備庁「総合分析業務に必要なAI機能の調査・実証に係る説明書」	`https://www.mod.go.jp/atla/soubiseisaku/newentry/startup_proc_r080306.pdf` [50]

重要度	資料	URL
最重要	防衛装備庁「スタートアップ技術提案評価方式を適用した案件の契約結果」	https://www.mod.go.jp/atla/soubiseisaku/newentry/startup_contr_r080806.pdf [51]
最重要	Sakana AI「防衛省から『総合分析業務に必要なAI機能の調査・実証』を受注」	https://sakana.ai/defense-integrated-analysis/ [52]
主要報道	Ledge.ai「防衛省の情報分析をSakana AIがAI支援 公開情報で収集・分析・管理、AIエージェント技術を実証」	https://ledge.ai/articles/sakana_ai_defense_information_analysis [9]
主要報道	ビジネス+IT「防衛省、自衛隊のインテリジェンス・情報分析にSakana AIの技術を採用」	https://www.sbbit.jp/article/cont1/186660 [53]
政策	防衛省「AI活用推進基本方針」発表ページ	https://www.mod.go.jp/j/press/news/2024/07/02a.html [54]
政策	防衛省「AI活用推進基本方針」PDF	https://www.mod.go.jp/j/press/news/2024/07/02a_03.pdf [55]
セキュリティ	防衛装備庁「装備品等及び役務の調達における情報セキュリティ基準」	https://www.mod.go.jp/atla/cybersecurity.html [56]
関連防衛案件	Sakana AI「防衛イノベーション科学技術研究所からの委託研究を開始」	https://sakana.ai/atla-contract-2026/ [57]
OSINT実績	Sakana AI「DIVER OSINT CTF 2026で5位入賞」	https://sakana.ai/diver-osint/ [15]
技術	Sakana Fugu	https://sakana.ai/fugu-beta/ [58]
技術	Sakana AI「Trinity: An Evolved LLM Coordinator」	https://sakana.ai/trinity/ [59]
技術	Sakana AI「Learning to Orchestrate Agents in Natural Language with the Conductor」	https://sakana.ai/learning-to-orchestrate/ [60]
Deep Research技術	Sakana Marlin	https://sakana.ai/marlin/ [16]

重要度	資料	URL
偽・誤情報対策	Sakana AI「総務省事業においてSNS空間の可視化と偽・誤情報対策を行う独自技術を開発」	https://sakana.ai/mic-project/ [17]
OSINT・安全保障連携	Sakana AI・DEEP DIVE パートナーシップ	https://sakana.ai/deep-dive-partnership/ [61]
個人情報	e-Gov「個人情報の保護に関する法律」	https://laws.e-gov.go.jp/law/415AC0000000057 [32]
個人情報	個人情報保護委員会「行政機関等編ガイドライン」	https://www.ppc.go.jp/personalinfo/legal/guidelines_administrative/ [33]
著作権	e-Gov「著作権法」	https://laws.e-gov.go.jp/law/345AC0000000048/ [35]
著作権・AI	文化庁「AIと著作権について」	https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html [36]
不正アクセス	e-Gov「不正アクセス行為の禁止等に関する法律」	https://laws.e-gov.go.jp/law/411AC0000000128 [38]
秘密保護	e-Gov「特定秘密の保護に関する法律」	https://laws.e-gov.go.jp/law/425AC0000000108 [39]
海外比較	U.S. DoD, 2023 Data, Analytics and AI Adoption Strategy関連発表	https://www.defense.gov/News/News-Stories/Article/Article/3578219/dod-releases-ai-adoption-strategy/ [62]
海外比較	NATO「Summary of NATO's Revised AI Strategy」	https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy [63]

調査時点：2026年9月1日（日本時間）。 この時点で最も重要な未公開事項は、Sakana AIの採択技術提案書・確定仕様書、具体的モデル名とモデル提供者、公開情報取得方式、RAG・検索・保存基盤、クラウド／オンプレ構成、モデル/APIへのデータ送信範囲、アクセス制御、ログとデータの保持期間、具体的評価指標・目標値、第三者監査の有無である。これらについては公開資料から推測を事実として補わず、本レポートでは明示的に**「未公開」**として扱った。 [64]

---

[1] [8] [9] 防衛省の情報分析をSakana AIがAI支援　公開情報で収集・分析・管理、AIエージェント技術を実証 | Ledge.ai
https://ledge.ai/articles/sakana_ai_defense_information_analysis

[2] [3] [5] [7] [10] [12] [18] [20] [22] [24] [26] [30] [40] [50] [64] mod.go.jp
https://www.mod.go.jp/atla/soubiseisaku/newentry/startup_proc_r080306.pdf

[4] [15] [43] Sakana AI防衛・インテリジェンスチーム、「DIVER OSINT CTF 2026」で5位入賞　Fuguを活用したOSINTエージェントの可能性
https://sakana.ai/diver-osint/

[6] [19] [25] [27] [28] [42] [55] mod.go.jp
https://www.mod.go.jp/j/press/news/2024/07/02a_03.pdf

11 53 防衛省、自衛隊のインテリジェンス・情報分析にSakana AIの技術を採用 情報本部のインテリジェンス・総合分析業務を支援するAI機能｜ビジネス+IT
https://www.sbbit.jp/article/cont1/186660

13 14 48 49 58 Sakana Fugu: A Multi-Agent Orchestration System as a Foundation Model
https://sakana.ai/fugu-beta/

16 Sakana Marlin — Your Virtual CSO
https://sakana.ai/marlin/

17 Sakana AI、総務省事業においてSNS空間の可視化と偽・誤情報対策を行う独自技術を開発
https://sakana.ai/mic-project/

21 29 45 51 mod.go.jp
https://www.mod.go.jp/atla/soubiseisaku/newentry/startup_contr_r080806.pdf

23 52 Sakana AI、防衛省から「総合分析業務に必要なAI機能の調査・実証」を受注
https://sakana.ai/defense-integrated-analysis/

31 33 個人情報の保護に関する法律についてのガイドライン （行政 ...
https://www.ppc.go.jp/personalinfo/legal/guidelines_administrative/?utm_source=chatgpt.com

32 個人情報の保護に関する法律
https://laws.e-gov.go.jp/law/415AC0000000057?utm_source=chatgpt.com

34 著作権法の一部を改正する法律（平成30年法律第30号）について
https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/?utm_source=chatgpt.com

35 著作権法
https://laws.e-gov.go.jp/law/345AC0000000048?utm_source=chatgpt.com

36 AIと著作権について
https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html?utm_source=chatgpt.com

37 38 不正アクセス行為の禁止等に関する法律
https://laws.e-gov.go.jp/law/411AC0000000128?utm_source=chatgpt.com

39 特定秘密の保護に関する法律
https://laws.e-gov.go.jp/law/425AC0000000108?utm_source=chatgpt.com

41 NATO releases revised AI strategy
https://www.nato.int/en/news-and-events/articles/news/2024/07/10/nato-releases-revised-ai-strategy?utm_source=chatgpt.com

44 NATO's Data and Artificial Intelligence Review Board
https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2022/10/13/natos-data-and-artificial-intelligence-review-board?utm_source=chatgpt.com

46 61 Sakana AI、一般社団法人DEEP DIVEとAIを活用した情報分析に関するパートナーシップを締結
https://sakana.ai/deep-dive-partnership/

47 62 DOD Releases AI Adoption Strategy
https://www.defense.gov/News/News-Stories/Article/Article/3578219/dod-releases-ai-adoption-strategy/?utm_source=chatgpt.com

54 防衛省・自衛隊：防衛省ＡＩ活用推進基本方針と防衛省サイバー人材総合戦略の策定について
https://www.mod.go.jp/j/press/news/2024/07/02a.html

56 防衛装備庁 : 装備品等及び役務の調達における情報セキュリティ基準について
https://www.mod.go.jp/atla/cybersecurity.html

57 Sakana AI、防衛イノベーション科学技術研究所からの委託研究を開始
https://sakana.ai/atla-contract-2026/

59 Trinity: An Evolved LLM Coordinator
https://sakana.ai/trinity/

60 Learning to Orchestrate Agents in Natural Language with the Conductor
https://sakana.ai/learning-to-orchestrate/

63 Summary of NATO's revised Artificial Intelligence (AI) strategy
https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy?utm_source=chatgpt.com