

「装備品等の研究開発における責任ある AI 適用ガイドライン（第 1 版）」の公表

2025 年 6 月 30 日

弁護士 殿井健幸

<目次>

1.はじめに

2.責任ある AI 適用ガイドラインの内容

(1)策定の経緯

(2) AI 装備品等の研究開発における確認事項

(3) AI 装備品等の研究開発における実施事項

3.おわりに

1. はじめに

2022 年 12 月には、日本政府は、防衛費の大幅な増額や反撃能力の保有・強化などを盛り込んだ新しい「防衛 3 文書」（「国家安全保障戦略」、「国家防衛戦略」及び「防衛力整備計画」）を策定し、2023 年には「防衛生産基盤強化法」が策定されました。

また、日本の防衛産業の利益率の実態が、従来、諸外国に比べ低かったこと等を踏まえ、令和 5 年以降、企業努力を反映する適正な利益を確保できるよう、2023 年度から製造・開発を発注した弾薬や航空機・艦船、通信機器などに対して、利益率を最高 15%とする制度を適用しました。

また、近時、防衛省は、日本スタートアップ大賞において「防衛スタートアップ賞」を新設するなど、優れた技術・製品・価格競争力等を有する企業が新たに防衛産業へ参入する機会を促進することにより、サプライチェーン強靱化や民生先端技術の取り込みを図ることで、防衛生産・技術基盤の強化を図っています。

不安定な国際情勢や、ドローン・宇宙・AI といった新規のテクノロジー領域におけるめざましい技術革新を背景として、防衛産業は将来有望な分野となっています。

2. 装備品等の研究開発における責任ある AI 適用ガイドラインの概要

(1) 策定の経緯

2025 年 6 月 6 日、防衛省より、「装備品等の研究開発における責任ある AI 適用ガイドライン」（以下「責任ある AI 適用ガイドライン」）が公表されました。

防衛省は、2024 年 7 月 2 日に、防衛省として初めて、AI 活用推進の羅針盤となる基本方針である「[防衛省 AI 活用推進基本方針](#)」（以下「基本方針」といいます。）を策定しました。責任ある AI 適用ガイドラインは、基本方針「第 3 部 AI 活用推進に向けた取組」「4. AI を使った装備品の研究開発」における「研究開発の文脈で、このような我が国の一連のコミットメントをどのように具体化するかについて、…（中略）…防衛省・自衛隊のガイドラインを策定する。」との記載を受けて、防衛省の装備品等の研究開発における責任ある AI 適用のコンセプトを示すものとして策定・公表されたものです。

責任ある AI 適用ガイドラインは、AI 技術を適用した装備品等（以下「AI 装備品等」といいます。）に係る研究開発事業の計画立案や実施等に当たって実施者が準拠すべき 指針として位置付けられるものとされています。

(2) AI 装備品等の研究開発における確認事項

日本政府は、自律型致死兵器システム（LAWS）に関する我が国作業文書（以下「LAWS 作業文書」といいます。）（※）において、我が国は、新興技術の軍事利用では、人間中心の原則を維持し、信頼性、予見可能性を確保し、責任ある形で行われることを重視する旨を表明し、我が国が国際人道法をはじめとする国際法や国内法により使用が認められない兵器システムの研究開発や運用を行う事はない旨を述べているほか、国際的にも開発・使用が認められるべきではない兵器システムや、人間の関与が確保された自律型兵器システムの開発・運用等について見解を示しています。

責任ある AI 適用ガイドラインは、LAWS 作業文書を踏まえ、AI 適用のリスクへの実効性のある対策を確保した管理を行うため、AI 装備品等の研究開発において準拠すべき要件として、以下のとおり、①AI 装備品等の運用構想が法的・政策的観点から適正であるかという要件（以下「要件 A」という。）及び②その運用構想を技術的観点から満足しているかという要件（以下「要件 B」という。）を設定し、研究開発事業の中で確認することとしています。

なお、A-1 の要件は AI 装備品等の研究開発に特有の要件ではないが、確認的に要件の 1 つとされている。A-2 の要件については、あくまでも現時点での政策的立場を反映した要件であり、今後、我が国として依拠すべきと考える新たな考え方が形成される場合には、必要に応じてこれを見直すことも検討することとされています。

要件 B については、各国の AI 倫理原則との関係性からも国際的潮流に整合するものとされています。

要件 A （法的・政策的要件）	A-1	国際人道法を始めとする国際法及び国内法の遵守が確保できないものでないこと ・ その性質上過度の傷害又は無用の苦痛を与えるもの、本質的に無差別なもの、その他国際人道法に従って使用することができないものでないこと等 ・ 国内法に従った使用ができるものであること
	A-2	人間の関与が及ばない完全自律型致死兵器でないこと ・ 適切なレベルの人間の判断が介在せず、人間による責任ある指揮命令系統の中での運用が確保できないような、人間の関与が及ばない完全自律型で致死性を有するものでないこと
要件 B （技術的要件）	B-1	人間の責任の明確化 ・ AI システムの利用に際して人間が適切な水準での注意を払い責任をもって対応できるよう、運用者の関与や運用者による適切な制御が可能な設計としていること
	B-2	運用者の適切な理解の醸成 ・ 運用者が AI システムを適切に使用できる仕組み、過度の依存を防ぐ対策、不具合発生時の運用者による改善の仕組みを設計していること
	B-3	公平性の確保 ・ バイアスの原因を探索、理解した上で AI システムとデータセットに対して適切な軽減策を講じ、不当で有害なバイアスを局限化していること
	B-4	検証可能性、透明性の確保 ・ AI システムの設計手順、採用したアルゴリズム、学習に使用したデータ等、システム構築の過程を明確化し、AI システムの検証可能性と透明性を確保していること
	B-5	信頼性、有効性の確保 ・ AI システムの信頼性、有効性、セキュリティに関して、多様な観点から試験評価を行い、ライフサイクル全体にわたって、許容可能なレベルでの動作を確保していること
	B-6	安全性の確保 ・ AI システムの誤作動や深刻な失敗のリスク発生を低減できる仕組みを整備し、安全性を確保していること

	B-7	国際法及び国内法の遵守が確保できないものでないこと ・適用される国際法及び国内法を遵守した運用が可能な設計としていること
--	-----	---

※[Working paper submitted by Japan to the United Nations on emerging technologies in the area of Lethal Autonomous Weapon systems \(LAWS\)](#) (概要日本語)

(3)AI 装備品等の研究開発における実施事項

上記「(1) AI 装備品等の研究開発における確認事項」において設定した要件を、①AI 装備品等の分類、②法的・政策的審査、③技術的審査の3つの事項を実施することで確認します。

①AI 装備品等の分類

AI 装備品等の研究開発にあたっては、特に上記 LAWS の観点から重点的にリスク管理すべき装備品等を高リスク AI 装備品等、それ以外の AI 装備品等を低リスク AI 装備品等と分類し、装備品等の態様に応じた適切なリスク管理を実施することとされています。

AI 技術を適用した装備品等の試作事業のうち、AI 機能に起因する出力がシステム内の破壊能力を有する機能に影響を与えない場合は低リスク AI 装備品等として分類されます。

AI 機能に起因する出力がシステム内の破壊能力を有する機能に影響を与える場合のうち、法的・政策的審査の結果、適格とされた場合には高リスク AI 装備品等として分類され、技術的審査を実施することで重点的なリスク管理が行われることとなります。

②法的・政策的審査

法的・政策的審査は、研究開発事業開始前に行われます。

研究開発対象の AI 装備品等が搭載する予定の機能及び想定される使用の態様等が前項の要件 A に沿って適切なものになっているかを法的・政策的観点から当該 AI 装備品等の構想段階で審査するものです。審査は、低リスク AI 装備品等ではないと判断された AI 装備品等を対象にして下表に示す審査項目について実施されることとなります。

法的・政策的審査は研究開発事業開始の可否を判断する重要な位置付けであるため、AI 装備品等に関する総合的な意思決定を行う会議体にて事業実施担当を審査する形で実施されますが、国際人道法を始めとする国際法や国内法の遵守に係る方策の適正性の確認が必要なため、法律分野の専門家や専門部署を審査員に含めることとされています。

	確認項目
A-1 国際人道法を始めとする国際法及び国内法の遵守が確保できないものでないこと	・当該装備品等を使用するにあたって軍事的必要性和人道的配慮のバランスを考慮し、過度の傷害又は無用の苦痛を与えることを防止する措置が含まれた構想になっているか。 ・軍事目標と非軍事目標（文民及び民用物）を区別し、軍事目標のみに攻撃を行うことが可能な装備品等の構想になっているか（区別原則）。 ・予測される具体的かつ直接的な軍事的利益との比較において、巻き添えによる文民の死亡、文民の傷害、民用物の損傷又はこれらの複合した事態を過度に引き起こすことが予測される攻撃を防止する措置が含まれた構想になっているか（比例性原則）。 ・攻撃に先立つ軍事目標の選定から攻撃の実施に至るまで、無差別攻撃を防止し文民と民用物への被害を最小限に抑えるための各種予防措置を実施し得る構想になっているか（予防原則）。 ・国際人道法のその他の要請や適用のある他の国際法及び国内法に従って使用することができないような装備品等の構想となっていないか。
A-2 人間の関与が及ばない完全自律型致死兵器でないこと	・適切なレベルの人間の判断が介在し、人間による責任ある指揮命令系統の中での運用が確保できる構想となっているか。 ・人間の関与が及ばない完全自律型の致死性を有する兵器システムの開発ではないか。

③技術的審査

技術的審査は、事業開始前、設計承認前、試験開始前、運用開始前等の研究開発における結節において行われます。

技術的審査は、高リスク AI 装備品等の試作品につき、実際の配備・運用段階でも要件 A を引き続き満たすことを可能にするための機能を備えているか、また適切なリスク軽減策が施されているかを技術的観点から確認するために行う審査です。下表に占める審査項目について審査が実施されることとなります。

技術的審査は AI 技術に対して専門的な知識を持つ部内の者によって構成される会議体が事業実施担当を審査する形で実施することとされていますが、審査にあたっては AI を搭載したシステムの安全性確保や品質管理についての高度な技術的知識が必要となる可能性もあることから、部外の専門家にも必要に応じて意見聴取することとされています。

	確認項目
B-1 人間の責任の明確化	・ AI システムの利用に際して適切なタイミングと程度において、運用者の関与や運用者による適切な制御が可能となるように設計されているか。 ・ AI システムと運用者それぞれの役割が明確に定義されているか。 ・ 運用者が負うべき責任が明確に定義されているか。
B-2 運用者の適切な理解の醸成	・ AI システムの挙動、パフォーマンス範囲、操作等に運用者が習熟して当システムを適切に使用できるように設計されているか。 ・ AI システムへの過度の依存を防ぐための対策が設計されているか。 ・ AI システムのモニタリング時に、運用者が不具合を認めたときにそれを改善することを可能とする仕組みが設計されているか。
B-3 公平性の確保	・ データセットの公平性要件が明確化されているか。 ・ AI モデルに関連するバイアスが許容レベルを超えないか確認するとともに、不具合が認められた場合の改善の仕組みが設計されているか。
B-4 検証可能性、透明性の確保	・ AI システムの構築に係る過程、使用した手法・データ・アルゴリズム等が明確化され、その妥当性を後に検証可能な仕組みが整備されているか。 ・ 研究開発体制（事業者含む。）において、説明責任を果たす責任者を明確化しているか。
B-5 信頼性、有効性の確保	・ AI システムの信頼性を確保するために、多様な指標を用いた試験評価を実施しているか。 ・ 開発初期から研究開発終了までの全ての期間において、運用を想定したデータセットを使用することが検討されているか。 ・ 運用を想定して、AI に接続する装備品等の信頼性を損なうことなく、AI を適用できることを設計において担保されているか。 ・ AI システムが容易に維持管理できる設計となっているか。 ・ AI システムに対してセキュリティ管理策の実施が検討されているか。
B-6 安全性の確保	・ AI システムの誤作動や深刻な失敗・事故の発生を低減する安全機構が設計されているか。
B-7 国際法及び国内法の遵守が確保できないものでないこと	・ 装備品等の運用に際して適用される国際法や国内法令、各種内部規則等から逸脱しないよう対策が検討されているか。

3.おわりに

実際の装備品等の研究開発には様々な形態があり得るため、責任ある AI 適用ガイドラインで示された考え方を指針として、各事業に応じて柔軟に事業設計を行うことが必要とされます。

また、AI をとりまく状況は急速に変化しており、一度決めたルールや手続きを維持し続けることが必ずしも適切でない場合があることから、ガイドラインは適宜更新を行うことが予定されていますので、最新のガイドラインをフォローアップする必要があります。

さらに、国際人道法を始めとする国際法や国内法の遵守に係る方策の適正性の確認を行う法的・政策的審査はもちろん、技術的審査においても、国際人道法を始めとする国際法や国内法の観点からの規範的な評価が必要となります。

そのため、AI 装備品等の研究開発事業に参入するにあたっては、法律分野の専門家を起用して進めることが有効になると考えられます。

本ニュースレターは、掲載時点までに入手した情報に基づいて執筆したものであり、また具体的な案件についての法的助言を行うものではないことにご留意ください。また、本ニュースレター中意見にわたる部分は、執筆担当者個人の見解を示すにとどまり、当事務所の見解ではありません。

以 上

ニュースレターの配信登録は[こちら](#)です。
バックナンバーは[こちら](#)でご覧いただけます。

牛島総合法律事務所
<https://www.ushijima-law.gr.jp/>