

AIシステムの認証制度のデザイン

製品のリスク管理策と組織のマネジメントを統合した
「ジョイント認証」の試み

目次

1. AIシステム認証への高まるニーズ	2
2. 既存の認証アプローチの限界とジョイント認証の必要性.....	3
(1) AIシステムが特定の性能・仕様を満たしているかを評価するアプローチ.....	4
(2) AIシステムのリスクに対する「管理策」を評価するアプローチ.....	4
(3) AIシステムを提供する「組織」を評価するアプローチ.....	5
(4) AIシステムに対する「ジョイント認証」－管理策認証と組織認証の融合.....	7
3. AIシステムに対するジョイント認証の方法.....	8
(1) ISO/IEC 42001 (AIマネジメントシステム) の構成.....	8
(2) AIシステムに対するジョイント認証の枠組み.....	9
(3) ISO以外の分野における類似の取組.....	11
4. ジョイント認証のアプローチのメリット	14
(1) 予測困難なリスクに対する柔軟かつ変化に強い対応	14
(2) マネジメントシステム認証の流用による手続の効率化.....	15
(3) 個別ドメインにおける認証との接続による拡張性	15
(4) AI関連法規へのコンプライアンスの確保.....	17
5. おわりに：ジョイント認証の未来.....	19

本ホワイトペーパーは、スマートガバナンス株式会社が、Microsoft Corporationと
共同で研究した成果をもとに執筆したものです。

1. AIシステム認証への高まるニーズ

AIを組み込んだ製品やサービス（以下「AIシステム」）は、リサーチや資料作成といった組織内での利用から、金融サービスのような規制領域、さらに人の生死に直接的に影響しうる自動運転や医療機器まで、社会のあらゆる場面に急速に浸透している。これらのAIシステムは、我々の生産性や生活の質を大きく向上させるものであり、もはや社会に不可逆的な変化をもたらしている。

このようなAIの実装場面の拡大に伴って、AIがもたらすリスクの領域や質・量も拡大している。

その内容は、たとえば、大量に収集された個人情報に基づくプロファイリングや悪意をもった行動誘導などのリスク、自動運転車や医療機器の誤作動のように人間の身体・生命に関わるリスク、バイアスの増幅による差別や機会の不平等拡大に関するリスク、AIの脆弱性を悪用して CBRN（化学・生物・放射線・核）関連システムへのアクセスが容易になるなど国家安全保障に対するリスク、ディープフェイクや偽情報の拡散による民主主義そのものに対するリスクなど多岐にわたる。

AIシステムによるこのようなリスクの登場に伴って、AIシステムに対する認証（Certification）制度の導入を求める声が高まっている。認証とは、独立した機関が、対象となる製品、プロセス、サービス、またはシステム等が特定の要件を満たしていることを評価し（適合性評価）、これを証明する文書（証明書）を提供することをいう¹。認証のメカニズムは、とりわけハイリスクな領域において、企業の自主的な取組に委ねるだけでは十分な安全性や信頼が確保できない場合に重要となる。たとえば、EUのAI法では、ハイリスクAIシステムについてAIシステムの提供者が適合性評価を行うことを義務付けており、一定の場合には第三者による認証の取得まで義務付けている²。また、自動車や医療機器といった、従来規制によって認証取得が義務付けられていた領域において、AIを組み込んだこれらの製品にどのように認証を付与すべきかについて、日本をはじめ多くの国において検討が進められている。AIがあらゆる領域に実装される現在、独立かつ専門的な第三者によるAIシステムへの「認証」の期待が急速に高まっている³。

¹ ISO「Conformity Assessment」(<https://www.iso.org/conformity-assessment.html>)

² EU AI 法 43 条

³ 日本の AI セーフティ・インスティテュート(AISI)の発表した「AI セーフティ年次レポート 2024」7 頁(https://aisi.go.jp/assets/pdf/j-aisi_report_2024_ja.pdf)においては、AI 製品に対する適合性評価規格の策定が ISO 及び CASCOC で議論されていることも受けて、AISI としても後に詳述するジョイント認証の検討がなされている旨が記載されている。また、認証ではないが、2025 年 6 月 3 日に発表された内閣官房知的財産戦略本部による「新たな国際標準戦略」24 頁(<https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2025/pdf/kokusaisenryaku.pdf>)においては、日本の標準における重要領域の中の戦略領域に AI 分野が位置付けられており、AI 分野での認証含む標準に関する日本政府の関心は高いことが伺われる。

2. 既存の認証アプローチの限界とジョイント認証の必要性

しかし、AIシステムに対する認証付与の仕組みを検討する場合、そこには様々な課題が存在している。その根本には、「何を評価するか」という点に関する複数の考え方がある。

一般的な認証の対象としては、(1) 製品が一定の性能・仕様へ適合しているか、(2) 製品に対して適切なリスク管理策が取られているか、(3) 製品を提供する組織の体制が適切か、という3つの考え方がある。

【図1 評価対象の分類】

(1) 製品の性能・仕様	(2) 製品リスクの管理策	(3) 組織のマネジメント
【評価事項】 製品・サービスが一定の性能や技術仕様など特定の要求事項に準拠しているか 【主な評価項目】 製品・サービスに関する物理的安全性（耐熱／絶縁等）や性能試験の手法等 【長所・短所】 長所：リスクが既知であり、そのリスクへの対処法が確立されている場合には有効 短所：高度に複雑なシステムや、ダイナミックに変化していくシステムには対応困難 【規格例】 IEC 62368-1(情報通信機器) IEC 60335(家庭用電気機器)等	【評価事項】 製品・サービスから生じるリスクを受容可能な範囲に抑えるための具体的な対策（管理策）が実施されているか 【主な評価項目】 製品・サービスごとのリスクの特定・評価、対策の選択・実装・文書化等 【長所・短所】 長所：リスクへの対応の内容や方法について、AIシステムに関する情報優位性をもつ提供者側が柔軟に判断できる 短所：一定の対応策が十分であるかどうかの外部からの判定が難しい また、技術変化の速い分野においては、認証付与時点における管理策の内容が陳腐化しやすい 【規格例】 ISO/IEC 27001 附属書 A ISO/IEC 42001 附属書 A・B 等	【評価事項】 製品・サービスを提供する組織が、リスク管理を継続的に実行・改善できる体制を備えているか 【主な評価項目】 組織における PDCA（計画・実行・評価・改善）サイクルの運用等 【長所・短所】 長所：システム提供者の組織を通じて、管理策の適切性を担保できる また、具体的な管理策の評価よりも陳腐化しにくい（状況変化に強い） 短所：リスクのコンテキストが広く、また自律的に挙動するAIシステムについては、組織だけでその信頼性を担保することに限界がある 【規格例】 ISO/IEC 9001 ISO/IEC 27001 本文 ISO/IEC 42001 本文

以下では、AIシステムに対してこれらの3種類の考え方を当てはめた場合の認証の対象と方法、課題について分析する。

(1) AIシステムが特定の性能・仕様を満たしているかを評価するアプローチ

第一の考え方として、AIシステムが、あらかじめ定められた特定の性能や技術仕様などの要求事項へ適合するかどうかを評価するアプローチがある。工業製品に対する認証は、まさにこのような方法によって与えられてきた。例えば、日本の電気用品安全法上、電気マッサージ器などの「特定電気用品」については、当該製品が適切な耐熱性、絶縁性等に関する法定の安全性能を満たすかについての認証を受けてPSEマークを取得することが義務付けられている。同様に、AIシステムの学習データやアルゴリズム等を対象として、そのシステムが仕様書の記載どおりに設計されているかを評価したり、仕様書が求める性能基準を満たすかをテストしたりすることが考えられる（なお、一般的にこのようなアプローチは「製品認証」とも呼ばれるが、下記（2）で述べる「管理策」に対する認証のことも「製品認証」と呼ぶ場合があるため⁴、本ペーパーでは両者の混同を避ける目的で、上記のアプローチについて「仕様適合性」という呼称を用いる）。

このように、事前に規定された一定の要求事項への適合性を評価するアプローチは、製品が試験時から上市後まで大きく変化することなく、従来のモノの製品のように、リスクが既知であり、そのリスクへの対処法が確立されている場合には有効である。しかし、継続的にアップデートされるシステムの場合や、特に、AIのように高度な複雑性を持ち、システムがよりダイナミックに変化していくような製品については、このようなアプローチは十分に有効でない場合もある。すなわち、AIシステムは、学習データや運用環境に依存して挙動が変化する「動的」な性質を持つため、ある一時点におけるスナップショットだけではリスクを十分に評価できない。また、AIの判断過程は高度に複雑なブラックボックスであるため、未知の影響を事前に網羅的に想定して要求を定めることは容易ではない。さらに、物理的な安全性のみならず、公平性や倫理的な適切性といった定性的な価値判断に基づく要求事項を、客観的かつ定量的な評価指標として定義し、リスト化することも難しい。そのため、AIシステムの適合性評価において硬直的に一定の性能や仕様を要求することは、AIシステムの開発や活用を強く抑制してしまうリスクがある⁵。

(2) AIシステムのリスクに対する「管理策」を評価するアプローチ

第二に、AIシステムから想定される様々なリスクに対して、これに対処するための適切な管理策がとられているかを評価するアプローチが考えられる。つまり

⁴ たとえば、EU AI法は、製品安全法令であるが、その中身は（1）のような仕様適合性ではなく、（2）のような管理策を評価する内容となっている。

⁵ もっとも、自動運転や手術ロボットなど、一部のハイリスクなハードウェアを伴うAIシステムにおいては、性能・仕様を対象とする適合性評価や認証のニーズがより高まる場合もあり、そのような場合には分野ごとの法律で性能・仕様を要求したり、標準において性能・仕様のモデルを示したりすることは考えられる。とはいえ、これらはあくまで個別の分野ごとに設定されるものであり、AIシステム一般に適用される共通の性能・仕様を設定することは困難または過剰であると思われる。

個々のAIシステムについて、そこから生じるリスクを特定・評価し、それに対する一定の対応がとられているかという点を評価するというものである。

このような手法は、サイバーセキュリティやプライバシーなどソフトウェアベースのリスクについて一般的に採用されてきた。例えば、情報セキュリティマネジメントシステムの規格であるISO/IEC 27001の附属書Aは、システムの情報セキュリティに対する管理策を定める。そこでは、組織が自社のシステムに改ざんリスクのある機密データが含まれているかを特定し、暗号技術の効果的な利用に関するルールを定義し、実施するといったような管理策が複数定められている（A.8.24）⁶。

このようなアプローチは、個々のAIシステムやそこから生じるリスクの性質についてより多くの情報を有している提供者（ここでは、AIシステムを開発したり、AIシステムをサービスとして提供したりする主体を含む。以下同様。）の側が、それらの情報をもとにベストなリスク対策を判断できる点で優れている。また、（1）のように画一的な基準で判断するのではなく、リスクへの対応方法を組織が柔軟に判断することとされている点で、リスクの範囲が広くダイナミックに状況が変わるAIシステムなどの認証に有用であると考えられる。

もっとも、そのようにして選択された管理策が適切であるかどうかについて、第三者（認証主体）が、どのように評価できるのかについて課題が残る。一般的に、認証主体は、AIシステムの提供者よりも少ない情報しかもたないためである。また、AIシステムの場合は特に技術や社会受容性の変化が速いため、認証付与時点において、あるAIシステムの「管理策」のリストが適切だと判断されたとしても、それが継続的に妥当であり続けるとは限らないという難点もある。そのため、AIシステムの提供者（組織）自体が、管理策を常に客観的に評価し見直し続けることができる体制になっていなければ、不断に変化を続けるAIシステム等に対するリスク対応としては十分なものとは言い難い。

実際、次項で述べるように、ソフトウェアのリスクに対する管理策は、それ自体で評価されるものではなく、常にマネジメントシステムの条項全体の文脈の中で、その妥当性が評価されてきた。

（3）AIシステムを提供する「組織」を評価するアプローチ

第三のアプローチとして、AIシステムを提供する組織の在り方を評価するアプローチが考えられる。

もともと、ソフトウェアの世界では、製品自体ではなく、組織のマネジメントシステムに対する認証を与えることが広く行われてきた。典型的なのは、ISO 9001（品質マネジメントシステム）や、上記のISO/IEC 27001（情報セキュリティマネジメントシステム）の本体部分である。

⁶ ISO/IEC 27001 Annex A.8 Technological controls 24. Cryptography

たとえば、マネジメントシステムの典型例であるISO 9001（品質マネジメントシステム）認証では、個別の製品が特定の品質基準を満たしているか、あるいは個々の製品についてとられたリスク管理策が適切かといった点を直接要求するわけではなく、顧客の要求を把握し、製品の設計・開発から製造、検査に至るまでの一連のプロセスを適切に管理し、継続的に改善していくための「仕組み（システム）」が組織内で確立・運用されていることを認証する⁷。つまり、認証の対象はあくまで「品質を管理する組織の体制」であり、その結果として生み出される製品やサービスの品質そのものを保証するものではない。

このような手法は、製品の信頼性やリスク管理策の適切性を第三者が直接的に判断するのではなく、むしろ組織のメカニズム（ガバナンス）を通じて提供者自体に担保させようとするものである。これによって、上記（2）の管理策評価の課題であった、AIシステムの提供主体と外部の認証主体の間の情報格差を埋めることが期待できる。また、具体的な「管理策」が、技術や社会の変化によって陳腐化しやすい一方、「組織が管理策を随時評価・改善できる体制か」を評価するマネジメントシステム認証は、より状況変化に対応しやすいというメリットもある。

AIシステムの文脈においては、その動的な性質を踏まえると、マネジメントシステム的なアプローチが極めて重要になってくる。組織は、AIシステムに関するリスクや影響を評価し、プロセスを見直し、必要に応じて更新できる人的管理を中心とした強固なシステムを有していることを示す必要性は高い。

他方、組織やその中の人間を通じてリスクをコントロールするアプローチには限界もある。たとえば、米国におけるソフトウェア医療機器の承認プロセスにおいて、個々の医療機器ではなくそれを製造する組織のマネジメント要件を重視し、一定の基準を満たす企業には製品の個別承認プロセスを簡素化または省略できるようにする試みが行われた（Pre-certification Program）。しかし結果的には、ソフトウェア開発プロセスの多様性や、組織評価の基準の欠如、臨床性能やサイバーセキュリティなどデバイス固有の要素を認定する必要性などを理由として、このアプローチは有効に機能しなかったことが報告されている⁸。これは、組織のみを対象とする認証の限界を端的に示すものである。そのため、AIシステムを提供する組織だけでなく、AIシステムそのものについて適切な管理策が講じられているかを評価することに対する社会や市場の要請は高い。

⁷ ISO/IEC 27001

⁸ FDA 「The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings」 12頁 (<https://www.fda.gov/media/161815/download?attachment>)

(4) AIシステムに対する「ジョイント認証」 — 管理策認証と組織認証の融合

以上のように、リスクシナリオが多様でシステムの内容自体もダイナミックに変化していくAIシステムについては、(1) 製品の性能・仕様、(2) 製品のリスクに対する管理策、(3) 製品を提供する組織のマネジメント、のいずれを対象とする認証アプローチにも限界がある。

このうち、(1) のアプローチについては、たとえば自動運転車や手術ロボットなどのハードウェア製品であってシステムの組み込みが限定的な範囲について、分野ごとの法令等に基づいて一部適用する余地があるものの、一般論としてはAIアルゴリズムのように変化の速い領域においては適用困難である場合が多い。そこで、AIシステムに対する基本的な認証の枠組みとしては、製品ごとに採用されている管理策の適切性(上記(2))と、それを評価・実行・改善する組織のマネジメントの適切性(上記(3))をあわせて審査することが、AIシステムへのより実質的なリスク管理に繋がり、結果として、信頼の付与につながると考えられる⁹。

このようなアプローチは、管理策と組織の双方を評価することから、「ジョイント認証」(Joint Certification)とも呼ばれることもある¹⁰。ジョイント認証は、「AIシステムの信頼は、信頼できるマネジメントシステムを有する組織・プロセスが適切な管理策を講じ続けることによってこそ確保される」という考え方に基づくものである。

このようなアプローチを導入する場合には、少なくとも以下の①～③の要素が必要となる。

1. AIシステムの「管理策」に関する評価基準
2. AIシステムを提供する「組織」に関する評価基準
3. 両者を統合的に評価し認証を付与するための仕組み

以下では、既存の標準やフレームワークを参照しながら、これら①～③の要素をどのように充足することができるかについて検討する。

⁹ なお、これらに加えて、とりわけハードウェア製品(自動運転車・AI医療機器等)について、一定の性能や仕様が求められる際には、別途仕様適合性認証を行う場合も考えられる。

¹⁰ AI製品以外についても、一部の物理的製品について既にジョイント認証が実施されている。たとえば、エアコンユニットについては、ISO/IEC 9001 (Quality Management System certification)とEnergy Star (Product certification)を組み合わせることとされている。ただし、これは本文で言及した(1)と(3)を組み合わせる類型である。

3. AIシステムに対するジョイント認証の方法

(1) ISO/IEC 42001 (AIマネジメントシステム) の構成

ISO/IEC 42001が現在、責任あるAIシステムに焦点を当てた認証制度を伴う唯一の国際標準として広く認識されている。

ISO/IEC 42001は、全10章の本文と4つの附属書（A～D）で構成される。

このうち本文は、ISO 9001（品質マネジメントシステム）やISO/IEC 27001（情報セキュリティマネジメントシステム）といった著名なマネジメントシステム標準と類似した構造になっており、計画・実行・評価・改善という共通のハイレベルな構造に従っている。すなわち、ISO/IEC 42001の本文は、典型的な「組織」に関する標準である。

他方、附属書Aは、AIシステムの設計・運用に関連するリスクに対応するための管理目標および管理策を包括的に示している。そこではたとえば、組織がAIシステムの影響評価を実施し、その結果を文書化すること¹¹、といった管理策が記載されている。附属書Bは、その実行についての詳細なガイダンスを示している¹²。例えば、上記のAIシステムの影響評価プロセスに関する管理策のガイダンスとしては、配慮すべき影響の項目として、個人の法的地位、基本的人権、社会といった具体的な要素を挙げている¹³。このように、ISO/IEC 42001の附属書A及びBは、この規格の重要な要素として機能する。附属書Aでは組織がリスクに適切に対処しているかを判断するために参照しなければならない規範的な管理策が記載され、附属書Bは選択された管理策を適切に実施するための情報提供としてのガイダンスが記載されている¹⁴。

なお、附属書Cは、リスクマネジメントにおける潜在的な組織目標やリスク要因を、附属書Dは、多様な産業セクターにおける具体的な活用事例（ユースケース）を示している。これらについては、「参考（informative）」として位置づけられており、あくまで参考となる情報の提供であり、規範性をもたず、規定の内容を構成しない。

¹¹ ISO/IEC 42001 A. 5. 3

¹² これらの策はすべてを実施する必要はなく、組織の特性に応じて独自に策定・実施することができる（ISO/IEC 42001 本文6. 1. 3参照）。

¹³ ISO/IEC 42001 B. 5. 2

¹⁴ 38個の全ての管理策を遵守しなければならない訳ではない。38の管理策についてはその採否も含めて判断することとされている（ISO/IEC 42001 6. 1. 3. ）。

【図2 ISO/IEC 42001の構造】



(2) AIシステムに対するジョイント認証の枠組み

以上の分析を踏まえると、AIシステムに対するジョイント認証として必要となる各要素（2.（4）①～③で整理した各項目）については、以下のような枠組みを用いることが考えられる。

- ① AIシステムの「管理策」に対する評価基準：ISO/IEC 42001 附属書A, B
- ② AIシステムを提供する「組織」に関する評価基準：ISO/IEC 42001 本文
- ③ 両者を統合的に評価し認証を付与するための仕組み（例：適合性評価スキーム）：ISO/IEC 42007「AIシステムの適合性評価スキームを策定するための高レベルの枠組みとガイダンス」（現在ISO/IEC SC42 JWG6にて開発中）

まず、①AIシステムに対する「管理策」の評価基準としては、ISO/IEC 42001の附属書A（及びそのガイダンスにあたる附属書B）をベースとして、製品ごとに適切なリスク管理がなされているかを審査することが有効であると考えられる。上述のように、リスクが不確実で、システムやサービス自体がダイナミックに変

化していくAIシステムについては、一定の仕様をみたすかといった硬直的な評価基準ではなく、リスクマネジメントや文書化などの管理策を製品ごとに審査するアプローチの方が適していると考えられる。

次に、②のAIシステムを提供する「組織」に関する評価基準としては、マネジメントシステム標準としてのISO/IEC 42001 がこれに該当する。上述のとおり、あるAIシステムの「管理策」のリストが適切だと判断されたとしても、それが継続的に妥当であり続けるとは限らない。そのため、AIシステムを開発・運用する組織が、管理策を常に評価しアップデートし続けることができる体制になっていることを評価するのがこの項目である。

③の両者を統合的に評価し認証を付与するための仕組みは、共同認証アプローチの最も重要な部分である¹⁵。前述のとおり、ISO/IEC 42001 の構造はこのアプローチに特に適している。この規格は、本体部分にあるマネジメントシステム要求事項と、附属書にあるシステム固有の管理策を分けて評価する仕組みを可能にする。実際、AI 特有の製品認証スキームを作成するためのハイレベルなガイダンスの策定が、国際標準化機関で既に進行中である¹⁶。将来的に策定されるこのガイダンスは、ジョイント認証の考え方を含む適合性評価スキームの構築において実務的に参照されることが予想される。

実は、ある製品・サービスに対する認証を付与する際に、製品個別の要求と組織に対する要求の双方を評価するスキームは、AIシステムに限らず存在する。その典型例が、ISO/IEC 17067で示されるタイプ5やタイプ6と呼ばれる認証スキームである。具体的には、タイプ5認証スキームでは、製品そのものが要求事項を満たしているかを試験すると同時に、その製品を継続的に生産する工場のマネジメント体制（組織）が有効に機能しているかを評価する。さらに、認証後も市場からの製品サンプリング試験と組織の定期監査を組み合わせることで、製品と組織の両面から継続的に適合性を保証するモデルである。一方、タイプ6認証スキームは、主にプロセスやサービスの認証に用いられ、ここでも製品ごとの管理策と、それを生み出すマネジメントシステム（組織）の双方を評価する。

しかし、この種の認証スキームにおいて、最終的な認証の対象となるのはあくまでも製品・サービス単位であり、それとは独立して組織自体に認証が与えられるものではない。したがって、ある組織が、ひとつの製品やサービスについて認証を取得した後、別のAI製品についても認証を取得しようとした場合、個々の製品やサービスに固有の管理策のみならず、組織そのものについても再度全面的な認証が必要となる可能性がある。

このため、AIシステム自体に対する認証と、組織に対する認証の二段階の認証プロセスを導入することが、認証機関と認証を受ける側の双方にとって有益で

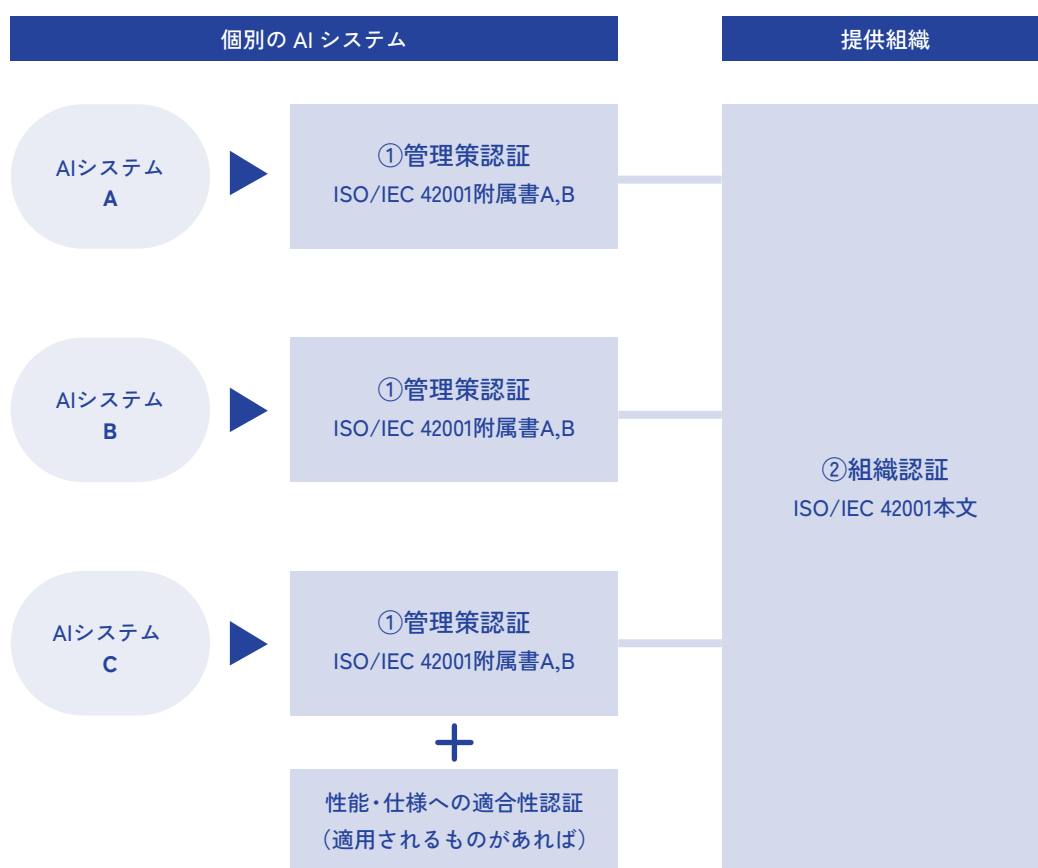
¹⁵ ISO/IEC 42007 については、まだ未発表であるが、ISO の HP(<https://www.iso.org/standard/89967.html>) に作業の進捗が記載されている。

¹⁶ 具体的には、ISO/IEC JTC 1/SC 42 および ISO/CASCO において、ISO/IEC 42007 の策定が進められている。

あると考えられる。ISO/IEC 42001本体に整合した組織に関する認証は、同一の組織が提供する複数のAIシステムに横断的に適用されうる。したがって、組織が複数のAIシステムを提供する場合に、新たなAIシステムごとに認証が必要となるのは、そのAIシステム固有の管理策に限られ、組織について毎回再度の認証が必要となるわけではない。

以上を図示すると、以下のようになる。

【図3 AIシステムに対するジョイント認証のイメージ】



(3) ISO以外の分野における類似の取組

国際標準及びその認証以外の場面においても、システムから生じるリスクに対する管理策と、これを提供する組織体制の双方を評価して信頼を付与するジョイント的なメカニズムが採用されている。これらのメカニズムは、認証制度そのものではないが、ジョイント認証について考えるにあたって大いに参考となる。そこで、以下では、その具体例を紹介する。

a. EU AI法

EU AI法はAI製品に関する安全規制であり、とりわけ「高リスクAIシステム」と分類されるAIシステムに対して、適合性評価（製品・サービスやマネジメント

システム等が、法令や規格で定める要求事項を満たしているか否かを実証すること）を要求している。EU AI法において、この適合性評価の具体的な要求事項の中心となっているのは、(i) AIシステムのリスクマネジメントや記録に関する製品固有の「管理策」に関する要件及び(ii) それらを支える「組織」（マネジメントシステム）に関する要件である。

より具体的には、AIシステムへの「管理策」に関する要求として、高リスクAIシステムの提供者は、AIシステムのリスクマネジメントシステムをライフサイクル全体を通じて確立・維持し、訓練データ等のデータガバナンスを徹底し、技術文書を整備しログ等の記録管理を行い、ユーザーへの情報提供（透明性）や人的監督（Human Oversight）の措置を講じることなどが求められている（EU AI法第9条～14条）。これらは、ISO/IEC 42001の附属書A及びBに類似する事項である。

他方、「組織」に関する要求として、高リスクAIシステムの提供者は、品質マネジメントシステム(QMS)を導入し、上記要件への継続的な適合を保証する体制を持たねばならないこととされている（EU AI法第17条）。この品質マネジメントシステムの要求は、ISO/IEC 42001本体の内容に類似する。このように、EU AI法は、高リスクAIシステムの安全性を確保するために、管理策と組織マネジメントの双方の評価を要求しているのである。

b. 米国NISTによるAI Risk Management Framework

米国国立標準技術研究所（NIST）が公表したAI Risk Management Framework

（NIST AI RMF）¹⁷は、組織がAIシステムから生じるリスクを自主的に管理するためのガイダンスである。このフレームワークもまた、具体的な管理策とそれを支える組織のガバナンスを統合的に扱っている。

NIST AI RMFは、「Govern（ガバナンス）」、「Map（マッピング）」、「Measure（測定）」、「Manage（管理）」という4つの主要な機能から構成されている。このうち「Govern」は、AIリスクに対応するための組織的な文化や方針、責任体制を確立することを求めるものであり、他の3つの機能すべての基礎と位置づけられている。これは、これまで述べてきた「組織」に関する要件に相当する。

一方で、「Map」、「Measure」、「Manage」は、AIシステムのライフサイクルを通じてリスクを特定・分析し、対応策を講じるための一連の具体的な活動を促すものであり、個々のシステムに対する「管理策」の要求と軌を一にする。このように、NIST AI RMFは、組織全体のガバナンス体制の確立を土台として、その上で具体的なリスク管理策を実践することを求めている。ここでも、シ

¹⁷ NIST 「Artificial Intelligence Risk Management Framework (AI RMF 1.0)」
<https://www.nist.gov/itl/ai-risk-management-framework>

システムの安全性と信頼性を確保するために、管理策と組織的マネジメントの双方を統合的に評価するアプローチが採用されており、リスクベースで継続的に改善するシステムを示している。

c. 日本の政府AI調達ガイドライン

日本の「行政の進化と革新のための生成AIの調達・利活用に係るガイドライン」（以下「政府AI調達ガイドライン」）¹⁸は、政府機関による生成AIの利活用促進とリスク管理を表裏一体で進めるために策定されたものである。このガイドラインでは、政府が調達するAIシステムを評価する際に、管理策と組織体制を統合的に評価するアプローチを採用している。

まず「組織」に関する要件として、本ガイドラインは事業者に対し、最先端の情報をもとにAIガバナンスを構築し、インシデント時の対応などについても確立することや、必要な人材教育を行うこと等を求めている。これは、これまで論じてきた組織マネジメントに関する要求に他ならない。

他方で、個別のAIシステムに対する「管理策」として、本ガイドラインは事業者が遵守すべき詳細な管理策を定めている¹⁹。そこでは、データの適切な取扱い、有害情報の出力制御、セキュリティ確保から、知的財産権の帰属、リスク発生時の事業者の協力義務に至るまで、多岐にわたる具体的な要求事項が網羅されている。

このように本ガイドラインは、事業者におけるAIガバナンス体制の確立と、個別のシステムにおけるチェックシートを通じた詳細な管理策の履行を両輪で求めている。その結果として、システムの安全性と信頼性を確保するために、管理策と組織的マネジメントの双方が統合的に評価されており、ISO/IEC 42001 に類似したアプローチを反映している。

d. ISMAP（政府情報システムのためのセキュリティ評価制度）

ISMAP（Information System Security Management and Assessment Program）とは、日本における政府情報システムのセキュリティ評価プログラムを指し、2020年6月に開始された。主たる目的は、政府情報システムに関する統一的なセキュリティ基準を確立することである²⁰。評価プロセスでは、第三者である

¹⁸ デジタル庁「行政の進化と革新のための生成AIの調達・利活用に係るガイドライン」
(https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/80419aea/20250527_resources_standard_guidelines_guideline_01.pdf)

¹⁹ 添付文書である「調達チェックシート」及び「契約チェックシート」がその中核をなす。

²⁰ NISC・デジタル庁・総務省・経済産業省「政府情報システムのためのセキュリティ評価制度（ISMAP）の概要」8頁
(https://www.ismap.go.jp/csm?id=kb_article_view&sysparm_article=KB0010005&sys_kb_id=c7b753a583e86a10aa68c6a8beaad3c1&spa=1)

ISMAP運営委員会が、特定のサービスが要求されるセキュリティ基準を満たしているかどうかを審査する。これらの要件はISMAP基準²¹に定められており、以下の章に区分されている。

- ・ ガバナンス基準（第3章）：クラウドサービス提供者の経営層に対し、組織的な情報セキュリティを監督・指導することを要求する。
- ・ マネジメント基準（第4章）：クラウドサービス提供者の管理者に対し、情報セキュリティマネジメントを確立、実施、運用、監視、維持し、継続的に改善することを要求する。
- ・ 管理策基準（第5章）：クラウドサービス提供者の運用担当者に対し、リスク対応方針に基づき選択すべき管理策のセットを提供する。

ISMAPマネジメント基準は、ISO/IEC 27001 を部分的に引用しており、前述したような組織的要件を含んでいる。さらに、ISMAP管理策基準には、「情報セキュリティの方針は定義され、経営層の承認を受け、発行され、従業員および関連する外部関係者に周知されなければならない」（第5章1.1）といった管理策に関する要件も含まれている。ISMAPは主としてクラウドサービス製品およびシステムの評価基準であるが、同時に管理策と組織的マネジメントを評価し、システムのセキュリティを確保するという点で、ジョイント認証に類似したアプローチを採用している²²。

4. ジョイント認証のアプローチのメリット

ジョイント認証のアプローチには、以下のようなメリットがある。

（1）予測困難なリスクに対する柔軟かつ変化に強い対応

前節で述べたように、ジョイント認証とは、製品レベルの管理策に関する認証メカニズムと、組織レベルのマネジメントに関する認証メカニズムを組み合わせ、AIシステムを認証するアプローチである。その主要な強みは、ISO/IEC 42001 を単なるマネジメントシステムとして認証する（附属書Aにシステムレベルの管理策が含まれているにもかかわらず）、あるいは両者を個別に認証するという限界を克服できる点にある（第2節参照）。

²¹ ISMAP 運営委員会「ISMAP 管理基準」

(https://www.ismap.go.jp/csm?id=kb_article_view&sysparm_article=KB0010028&sys_kb_id=757ef62ac3755250076ededb05013195&spa=1)

²² 但し、ISMAP の管理策は 1000 を超える数であり、一般の事業者にとっては非常に負担が重い。そのため、リスクの小さい情報を扱う業務について、より監査の対象を限定した ISMAP-LIU（ISMAP for Low-Impact Use）という制度も導入されている。(<https://www.digital.go.jp/policies/security/ismap-liu>)

個別のリスク管理策を認証することは、提供者が自らのAIシステムについて情報上の優位性を持っているため、リスク対応の内容や方法を柔軟に決定できるというメリットがある。しかし、その一方で重大な課題もある。それは、当該管理策の適切性について外部からの評価が困難になることである。さらに、これらの管理策の内容は、認証が付与される時点には既に陳腐化している可能性がある。

一方、「組織」に対する認証には、組織自体に責任ある管理策の実施を促すことで、システムの提供主体と外部の認証主体の間の情報格差を埋めることができ、また、組織に対する認証は具体的な管理策に対する認証よりも陳腐化しにくいというメリットがある。他方で、AIシステムはリスクのコンテキストが広く、また自律的に挙動するため、組織を認証するだけではAIシステムへの信頼を十分に担保し難いという欠点がある。

ジョイント認証は、これらのスキームの「良い所取り」をすることで、最新の技術水準を前提とした、柔軟かつ変化に強い信頼確保の在り方を提供するものである。

(2) マネジメントシステム認証の流用による手続の効率化

ジョイント認証の実務上の大きな利点は、「組織」レベルのAIマネジメントシステム (AIMS) に関する認証を、複数のAIシステムに流用できる点にある (図3参照)。従来の製品認証では、個々のAIシステムごとに、組織レベルまで含めた認証を取得する必要があるため、コストと期間が膨大になる傾向があった。しかし、ジョイント認証では、組織全体の体制についてはあくまで (AIシステムごとではなく) 組織を対象として評価しうる。その際、組織に対する認証 (AIMS認証) を一度取得すれば、個々のAIシステムについては、個別のAIシステムのリスク管理策のみを製品レベルで評価すれば足りる。これにより、認証コストと認証取得にかかる期間を大幅に削減し、AIシステムの開発・導入、そして市場への迅速な展開を促進することが期待できる。特に、複数のAIシステムを開発・運用する組織にとっては、認証プロセス全体が継続的に効率化される。また、組織認証を再利用できるようにすることは、認証機関の負担軽減にもつながり、認証機関のリソース逼迫という課題の解決にも寄与する。

(3) 個別ドメインにおける認証との接続による拡張性

AIシステムに対するジョイント認証のアプローチは、既存の他のマネジメントシステム認証への接続も期待されている。たとえば、医療機器については、国際標準であるISO 13485 (医療機器の品質マネジメントシステム)²³が、多くの国の規制において参照されている。ISO 13485とISO/IEC 42001は、PDCA (計画-実行-チェック-改善) サイクルに沿った組織体制を要求する点で、その骨格は類

²³ <https://www.iso.org/standard/59752.html>

似している。しかし、ISO/IEC 42001は、AIシステムの特徴を踏まえて、ISO 13485にはない追加的な要求事項を含んでいる（図4参照）。

【図4 ISO 13485とISO/IEC 42001の比較】²⁴

	両規格の類似点	42001 が 13485 と異なる点
計画フェーズ		
組織の状況	- QMS/AI マネジメントシステムの適用範囲を確立し、決定する。 - 定義されたプロセスを監視および測定する。 - 適切なプロセスを管理するためにリスクベースのアプローチを適用する。	- 適用範囲を決定する前に、外部および内部の課題を考慮する。 - 利害関係者のニーズと期待を理解する。 - AI マネジメントシステムの継続的改善のためのプロセスを計画する。
リーダーシップとコミットメント	- 品質、AI 方針、および目標が確立されることを確実にする。 - 要求事項を満たすことの重要性を伝える。 - 資源が利用可能であることを確実にする。 - AI マネジメントシステム /QMS が意図した結果を達成することを確実にする。	- AI マネジメントシステムの統合：ISO/IEC 42001 は、組織のプロセスへの統合をさらに追加し、他の組織方針との整合を強調している（附属書 B.2.3）。
実施および変更の詳細計画	- 関連する機能およびレベルで目標を確立する。 - 達成するための計画を決定する。 - AI マネジメントシステムおよび QMS システムへの変更の必要性を決定する。これは計画的な方法で実施されなければならない。	- 目標と実施計画を確立する前の、計画におけるリスクマネジメント（AI リスクアセスメント、AI リスク対応、および AI システム影響評価）。
支援と資源	- 人的資源の力量と訓練を決定し、提供する。 - 文書化の要求事項と管理に対する認識を確実にする。	- AI システムのための特定の資源には、データ資源、ソーリング資源、およびシステム・コンピューティング資源が含まれている。
実施・運用		
管理策の実施	- 両規格には、計画および管理プロセスが記載されている。 - 組織の要求事項を満たすプロセスを計画する。 - プロセスを管理し、緊急時対応計画を作成し、必要に応じてリスクマネジメントプロセスを文書化する。	- AI のリスクアセスメント、対応、およびシステム影響評価に焦点を当てる。 - 附属書で AI システムの影響を評価するための潜在的な管理策を提供する
パフォーマンスの確認		
測定、監視、分析および評価	- 頻繁な内部監査は、計画された間隔で内部監査を実施することに焦点を当てる場合がある。 - 計画された間隔でマネジメントレビューを実施する（ISO/IEC 42001 ではパフォーマンス評価フェーズの一部ですが、ISO13485 では計画フェーズでカバーされている）。	- 計画された間隔で AI システムの影響とリスクアセスメントを検出する。
改善の維持		
維持と処置	- 改善の必要がある。 - 不適合に対して是正処置をとる。 - 調査記録を維持し、文書化する。	- 継続的な改善が必要である。

たとえば、ISO 13485のマネジメントシステム要求を満たす医療機器がAIを搭載している場合、AI機能については、上記の追加的な要求事項（図4の右列）を

²⁴ Microsoft and EY, “How can the medical device industry seize the upside of changing AI regulatory requirements? – Proposal for a joint-certification approach”, 9-10頁

クリアすることで、ISO/IEC 42001を取得できるようにしておくこと、ゼロからISO/IEC 42001認証を取得する場合と比べて大幅に期間やコストを削減できる²⁵。

また、金融分野においても、たとえばEUのAI法では、金融機関の健全性に関するEU指令（2013/36/EU）に基づく品質マネジメントシステムやモニタリングシステムを、AI法が要求する品質マネジメントシステムやモニタリングシステムとして流用できる可能性が示唆されている²⁶。

このように、様々な分野において要求されている組織マネジメントや管理策の一部を、AIシステムの認証に再利用することができれば、ジョイント認証はさらなる拡張性を獲得し、認証取得コストや監査コストを一層低減することができる。

（４）AI関連法規へのコンプライアンスの確保

ジョイント認証を取得またはそれに向けた準備をしておくことで、世界各国の多くのAI関連法令を遵守できるようになりうると考えられる。

AIシステムの開発・利用について、国内外で多くの法規制の議論がされているが、その中でも、最も包括的な規制として知られているのは、EUのAI法である。ジョイント認証は、これらの日々変化する法的枠組みとの整合性を確保するうえで、極めて重要な役割を果たしうる。

3. （３）a. で述べたとおり、AI法は、高リスクAIシステムの提供者に対し、管理策と組織の双方を評価するアプローチをとっており、前者に対する適合性評価を求めている。同法では、生体認証システムの認証であって欧州整合規格が整備されていない場合や、他のEUにおける製品安全規制において認証が求められている製品に組み込まれるAIシステムなどについて、第三者による適合性評価（本ホワイトペーパーにおける「認証」）が義務付けられている²⁷。また、その他の高リスクAIシステムについても、自己による（内部統制システムを用いた）適合性評価が義務付けられている。

これらの適合性評価を実施する際の基準は、欧州整合規格の中で規定されることとなっているが²⁸、本ホワイトペーパーの執筆時点ではまだ欧州整合規格は成立していない。しかし、AI法とISO/IEC 42001の項目を比べると、AI法の品質マネジメントシステム（17条）に関する要求についてはISO/IEC 42001の本文に、AI法の管理策に関する要求についてはISO/IEC 42001の附属書A, Bに、多くが重複・類似している（図5参照）。そのため、ジョイント認証を得られるだけの管理策及び組織体制の整備を行っておくことで、EU AI法において高リスクAIシ

²⁵ 前掲Microsoft & EY, 14頁

²⁶ EU AI法前文158, Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013, (<https://eur-lex.europa.eu/eli/dir/2013/36/oj/eng>)

²⁷ EU AI法第43条

²⁸ EU AI法第40条

システムの提供者に課される義務の多くを履行できるという合理的な見通しがある状況と考えられる。

【図5 EU AI法とISO/IEC 42001の項目の比較】

◆ 組織マネジメントシステム

ISO/IEC 42001 (本文)	EU AI Act
条項 4-10	第 17 条：品質マネジメントシステム

◆ 管理策

ISO/IEC 42001 附属書 A	EU AI Act
A.5.2 AI システム影響評価プロセス	第 9 条：リスク管理システム
A.7.2 AI システムの開発および強化のためのデータ	第 10 条：データおよびデータガバナンス
A.7.3 データの取得	
A.7.4 AI システム用データの品質	
A.6.2.7 AI システム技術文書	第 11 条：技術文書
A.6.2.8 AI システムイベントログの記録	第 12 条：記録保持
A.8.2 システム文書および利用者への情報提供	第 13 条：透明性および利用者への情報の提供
A.9.3 AI システムの責任ある利用のための目的	第 14 条：人間による監視

また、AI法のような分野横断的な法律でなくとも、自動車・医療機器・金融システムなどのドメインごとに、AIに関する法規制と、それに基づく認証制度の整備が世界各国で進められている。その際、2. で述べたようなAIシステムの特徴と既存の認証スキームの長所・短所を考えると、最終的な許認可や認証付与に必要な要件の基本的な骨格は、ホワイトペーパーで議論したようなジョイント認証の形態をとる可能性が高い。実際にも、3. (3) で述べたように、米国NIST AI RMFや日本の政府AI調達ガイドラインなど多くの文書において同様の構造が採用されている。

このように、ジョイント認証の仕組みがあることで、各国のAIに関連する規制内容の整合性が確保され、事業者がグローバル市場へ参入する際のコンプライアンスコストや、政府が法執行する際のコストを低下させることができ、より効率的かつ効果的な規制を実現することができると考えられる。

5. おわりに：ジョイント認証の未来

本ホワイトペーパーでは、AIシステムが社会に深く浸透する中で高まる認証へのニーズに応えるため、既存の認証スキームを分析・整理した上で、既存の認証アプローチの限界を特定し、その限界を克服する「ジョイント認証」の必要性和具体的な方法、そしてそのメリットについて検討した。

AIシステムの認証においては、従来のモノの「製品の仕様適合性」を評価するアプローチは、AIの動的な性質や複雑性から必ずしも有効ではない。そのため、AIシステムのリスクに対する「管理策」を評価するアプローチと、AIシステムを提供する「組織」のマネジメントを評価するアプローチのそれぞれが持つ長所を組み合わせた「ジョイント認証」が、AIシステムへのより実質的な信頼付与に繋がることを示した。

ジョイント認証は、「信頼できるマネジメントシステムを有する組織・プロセスが適切な管理策を講じ続けることによってこそ、AIシステムの信頼は確保される」という考え方に基づくものである。具体的な枠組みとしては、ISO/IEC 42001の附属書A, BをAIシステムの「管理策」の評価基準とした上で、ISO/IEC 42001の本文をAIシステムを提供する「組織」の評価基準に据えることが考えられる。さらに、現在ISOにおいて審議中の ISO/IEC 42007 も、ISO/IEC 42001 の内容に基づき、適切に設計された認証スキームを通じて、両者を統合的に評価する仕組みを含むはずである。

ジョイント認証には、予測困難なリスクについて柔軟に対応でき、かつ変化に強いというメリットがあるだけでなく、一つの組織認証を複数のAI製品に再利用したり、医療や金融など個別ドメインの認証とAI認証を組み合わせたり、法令の義務の一部をAI認証で代替することによって、多くの局面において認証取得やそのための監査の時間やコストを大幅に削減できるというメリットがある。これ

は、AIシステムに対する信頼をいち早く社会に届けることに直結する意味で、非常に意義の大きいものである。

一方で、ジョイント認証を導入する際には、課題も存在する。本報告書で繰り返し述べたとおり、AIシステムの提供者が「管理策」として選定・実施した内容が適切なものであるか、多様なAIシステムやユースケースに対して、適切にリスクを認識した上で、どの対策をどのように組み合わせることが合理的な管理策となるのかを第三者が評価することは難しい。多様なAIのリスクについて、それぞれどこまでが社会的に許容可能であるかの水準を決めることも難しく（例えば、EU AI法やNIST AI RMFも許容水準は示していない）、また、汎用目的やマルチモーダルなAIについてはリスクシナリオが無数にあり、様々な主体がチューニングを行っていくことも想定されるという問題もある。そうした中で、どのようなリスクが想定されるか、想定されるリスクについてどのような対策をどこまで取るべきかを判断することは、たとえAIシステムのリスクに精通した事業者であっても非常に困難である。そうした状況を前提とすると、「組織」の認証についても、上記のような困難な「管理策」の設計・運用・改善を継続的に行えるのがどのような組織であるか、異なる組織間でどのようにリスク対策を連携すべきであるかといった点について、より具体的な議論が必要となろう。

これらの課題に対し、ジョイント認証の有効性を最大限に引き出すためには、様々なユースケースを用いて、場面ごとの管理策や組織の在り方や、組織間の連携を官民のステークホルダーで共有していくことが重要になる。そのようなプロセスを継続的に行っていくことで、より多くのAIシステムの提供者が、大きなコストをかけずに最適な管理策の実践及び組織の整備を行うことができるようになり、ユーザー側も安心してAIシステムを信頼できるようになる。こうした仕組みは、「アジャイル・ガバナンス」を掲げる日本のデジタル政策²⁹や、2025年5月に成立したAI推進法³⁰の理念にも整合する。常に最新の状態にアップデートされたケーススタディに基づくジョイント認証は、AIシステムの信頼性を確保し、イノベーションの社会実装を加速させるための強力なツールとなるだろう。

²⁹ デジタル庁「令和7年度デジタル社会の実現に向けた重点計画」（<https://www.digital.go.jp/policies/priority-policy-program>）。「アジャイル・ガバナンス」は日本のデジタル政策において重視されており、「一律かつ硬直的な事前規制ではなく、リスクベースで性能等を規定して達成に向けた民間の創意工夫を尊重するとともに、データに基づくEBPMを徹底し、機動的・柔軟で継続的な改善を可能とする」とのアジャイルガバナンス原則も踏まえて、デジタル技術の実装を阻み、社会のデジタル化の阻害要因となっているアナログ規制について、目視規制、定期検査・点検規制等の代表的な7種類の規制及びFD等記録媒体規制に関する法令等約1万条項が見直され、見直しが必要とされた8162件について約98%（8162件のうち7983件、2025年5月末時点）の見直しが完了している。

³⁰ <https://laws.e-gov.go.jp/law/507AC00000000053>

AIシステムの認証制度のデザイン：
製品のリスク管理策と組織のマネジメントを統合した「ジョイント認証」の試み

発行年：2025年12月

発行：スマートガバナンス株式会社

<https://smart-governance.co.jp/>

【執筆者】

羽深 宏樹

松井 春樹

鄭 育昌（外部協力）

落合 孝文

Matteo Posarelli

宮原 瑞穂

【Advisor】

Kimberly Lucy, Director, GRC Standards (Microsoft Corporation)

Copyright © 2025 Smart Governance Ltd.

本ホワイトペーパーは、スマートガバナンス株式会社が、
Microsoft Corporationと共同で研究した成果をもとに執筆したものです。

本書の著作権および関連する一切の権利は、スマートガバナンス株式会社に帰属します。無断での転載・複製・改変、ならびに第三者への配布・提供を禁じます。