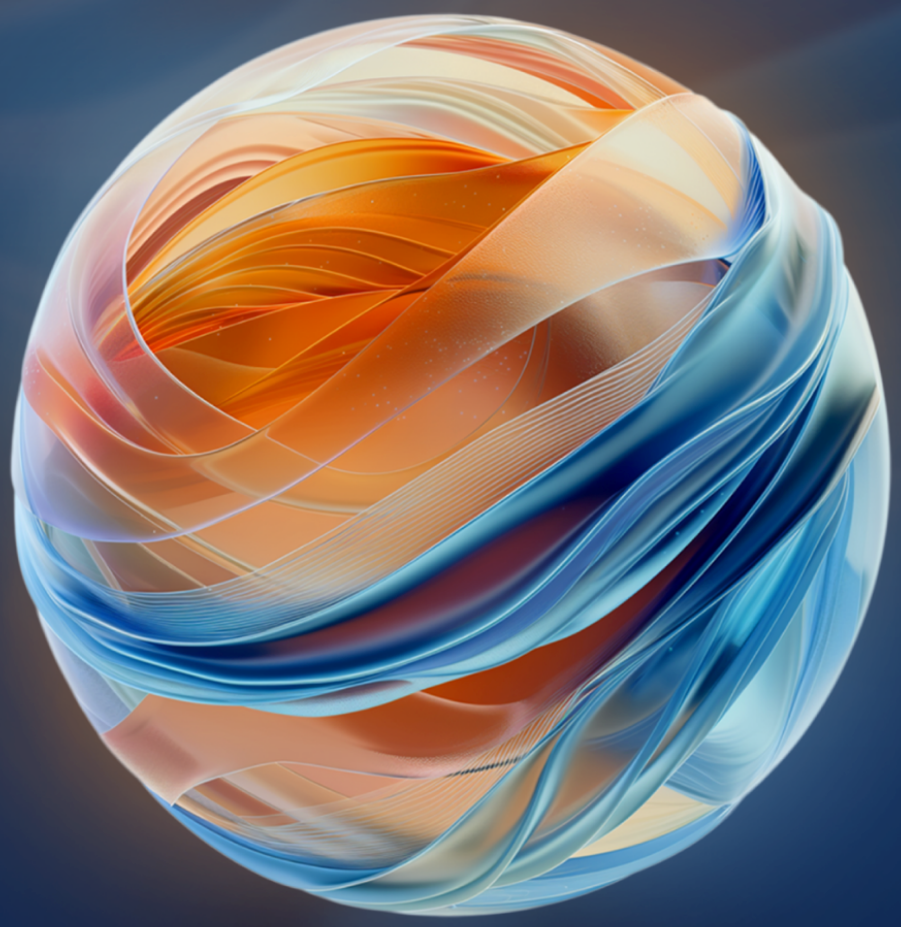
AI リスクマネジメント：
規制の枠を超えて考える



AI Governance and Compliance
Working Group

cloud security alliance®

AI ガバナンス・コンプライアンス・ワーキンググループの恒久的かつ公式な場所は https://cloudsecurityalliance.org/research/working-groups/ai-governance-compliance である。

© 2024 Cloud Security Alliance - All Rights Reserved.クラウド・セキュリティ・アライアンス（https://cloudsecurityalliance.org）のダウンロード、保存、コンピュータへの表示、閲覧、印刷、リンクは以下の条件に従って行うことができる：(a)ドラフトは、個人的、情報的、非商業的な使用に限って使用することができる。(b)ドラフトは、いかなる方法によっても修正または変更することができない。(c)ドラフトは、再配布することができない。(d)商標、著作権、またはその他の表示を削除することができない。(d)商標、著作権、その他の表示を削除してはならない。米国著作権法のフェアユース規定により許可される限り、クラウド・セキュリティ・アライアンスに帰属させることを条件に、ドラフトの一部を引用することができる。

謝辞

筆頭著者

シャンタル・スプリース博士

貢献者

キャンディ・アレクサンダー

サニトラ・アングラム

レヌ・ベディ

マダヴ・チャブラニ

フィリップ・チャイラ

マルコ・エルミニ博士

ベッキー・ゲイロード

フレデリック・ヘーニヒ

アルピタ・カウシク

キャシー・マイリー

アシシュ・ヴァシシュタ

ピーター・ベンチュラ

レビュアー

ハルシュ・ダイエー

ベニアミーノ・ディ・マルティーノ

デブラプ・ゴシュ

ラフル・グプタ

アシシュ・グプタ

ディーラジ・グルグベリ

J.ウィンストン・ヘイデン

シビル・VR・クラインミッシェル

ヴァイバフ・マリク

クリシュナ・マンガト

タレシュ・メフラ

スヴェン・オレンスキー

メガーナ・パルワテ

ヘナ・プラサンナ

チャンドラ・ラジャゴパラン

マイケル・ロザ

マリア（MJ）・シュヴェンガー

ラケシュ・シャルマ

CSA グローバルスタッフ

ライアン・ギフォード

クレア・レーナート

スティーブン・ルンペ

目次

謝辞 3

目次 4

AI リスクマネジメント：規制の枠を超えて考える 7

序文 7

エグゼクティブサマリー 8

パート 1 と 2：どのように連携して利用するか 8

定義 9

AI レジリエンス：ロバスト性、レジリエンス、可塑性 9

インテリジェント・システム：AI システム、AI エージェント、ロボット 10

説明責任、責任、賠償責任 11

監査人のアセスメント 12

AI ガバナンス 14

適用法、規制、標準 16

出発点としてのユースケース 17

インテリジェント・システム・インフラ 19

データ処理ユニット 20

センサーの役割 21

データの役割 24

サプライチェーンの役割 25

データ・プライバシー 26

データと著作権 26

データソース 26

有機データと合成データ 27

データの質 27

データ保管 28

ネットワーキング、コネクティビティ、コミュニケーション・インターフェイス 29

フォグ・コンピューティングとクラウド・コンピューティング 30

ソフトウェア・コンポーネント 31

アルゴリズム、トレーニング方法、モデル 32

アルゴリズムとモデル 32

フロンティア 35
モデルの微調整と妥当性確認 36
オーバーフィットと汎化 36
微調整 36
妥当性確認 36
モデルの頑健性と安定性 37
継続保守 37
微調整とモデル妥当性確認の最前線 37
アクチュエーター 38
電源 39
ユーザ・インターフェース 41
制御システム 43
安全システム 44
セキュリティ・システム 45
高度なプライバシー手法 47
開発とデバッグ 49
エンドユーザ・トレーニングとドキュメンテーション 50
エンドユーザ向け文書 50
(義務）エンドユーザに対するトレーニングとトレーニング記録の保管 52
展開とモニタリング 54
廃止措置 56
結論 57
附属書 58
附属書 1：監査の質問 "監査人を監査する" 58
附属書 2：監査の質問 "AI ガバナンス" 59
附属書 3：監査の質問 "説明責任、責任、賠償責任" 62
附属書 4：監査の質問 "法律、規制、標準" 63
附属書 5：監査の質問 "AI ビジネスケース" 64
附属書 6：監査の質問 "AI インフラ" 66
附属書 7：監査の質問 "センサー" 68
附属書 8：監査の質問 "データ" 68
附属書 9：監査の質問 "データ処理ユニット" 71
附属書 10: 監査の質問 "データ保管" 71

附属書 11：監査の質問 "ネットワーキング、コネクティビティ、コミュニケーション・インターフェイス" 72
附属書 12：監査の質問 "フォグ・コンピューティングとクラウド・コンピューティング" 73
附属書 13：監査の質問 "ソフトウェア・コンポーネント" 74
附属書 14：監査の質問 "アルゴリズム、トレーニング、モデル" 75
附属書 15：監査の質問 "微調整と妥当性確認" 77
附属書 16：監査の質問 "アクチュエーター" 78
附属書 17：監査の質問 "電源" 81
附属書 18：監査の質問 "ユーザ・インターフェース" 83
附属書 19：監査の質問 "コントロール・インターフェイス" 88
附属書 20：監査の質問 "安全システム" 90
附属書 21：監査の質問 "セキュリティ・システム" 92
附属書 22：監査の質問 "開発とデバッグ" 93
附属書 23：監査の質問 "エンドユーザ・トレーニングと文書化" 95
附属書 24：監査の質問 "展開とモニタリング " 96
附属書 25：監査の質問 "廃止" 98
略語と用語集 100
参考文献 104

AI リスクマネジメント：規制の枠を超えて考える

序文

人工知能（AI）が産業プロセスにますます統合されつつある世界では、正確で目的に沿った、結果に基づく AI 監査の必要性がミッションクリティカルとなっている。AI は、効率性の向上、意思決定の強化、革新的な機能など、多大なメリットをもたらす一方で、規模の大小を問わず、組織にとって重大なリスクや課題ももたらす。AI に対する信頼を確立するには、規制上の必要性を超えた改善でコンプライアンスに積極的に対処する AI 監査への包括的なアプローチによってのみ達成できる。

本稿では、業界を超えて適用できるように設計されたインテリジェントシステムを公平にアセスメントするために必要な、全体的な概要と適用可能な方法論を提示する。これらの監査上の考慮事項は、既存の AI 監査のベストプラクティスを基礎とし、AI のライフサイクル全体にまたがる革新的なアプローチを提供するものであり、附属書に監査の質問例を含む。

プライバシー、セキュリティ（情報とサイバーセキュリティを含む）、及び信頼性は、批判的思考と調査的思考、好奇心、及び意図しない挙動がないかシステムを評価する監査人の能力に焦点を当てたリスクベースのアプローチを提案することによって強調される。ここで定義するアプローチは、監査のベストプラクティスに沿ったものであり、現在または将来の AI 規制とは無関係であるが、AI 管理システムに関連するものである。AI のダイナミックな進化を考慮し、本稿の内容は、監査人が認識すべき領域の概略を示すものであるが、変更の可能性がある詳細については言及しない。

現在、例えば大規模言語モデル（LLM）[4]、[5]、[6]のアセスメントなど、新しい監査手法[1]、[2]、[3]が表面化しており、革新的なリスクアセスメントの必要性については相互に同意している。我々の姉妹ワーキンググループでは、AI ガバナンスについて基本的な観点から議論している：Don't Panic！Getting Real About AI Governance [7]」である。この論文は、具体化の有無にかかわらず、ライフサイクル全体[8]、[9]に沿って多様なインテリジェントシステムをアセスメントするために使用できる目標を提示している。さらに、これらの監査上の考慮事項は、インテリジェントシステムの監査への構造的なアプローチを促進する。これは、テーマ別の概要に続いて、既成概念にとらわれない思考を促進し、コンプライアンスの範囲を超えたアセスメントを実施するのに役立つ、対応する附属書での質問を提供することによっている。

エグゼクティブサマリー

本稿は、AI 監査の創造と進化に貢献することを目的としており、多様でインテリジェントなシステムを、そのライフサイクル全体を通じて、また具体化の有無にかかわらず評価するための全体的アプローチに関する考察を提供する。インテリジェントシステムを厳格にアセスメントするための方法論を提示し、企業が検出されたリスクを軽減し、信頼できる AI を構築するために革新的かつ積極的にコンプライアンスにアプローチできるようにする。

これらの AI 監査に関する考察は、監査人の批判的思考と調査的思考、好奇心旺盛でありながら分析的な姿勢を重視することで、業界や法制を超えたインテリジェントシステムのアセスメントを支援する提言を提案している。厳密なアセスメントに基づき、インテリジェント・システム導入のメリットは、将来的なビジネスへの影響につながる隠れた落とし穴に対する不確実性を減らしながら、金銭的に十分に享受することができる。

AI のダイナミックな性質は、各トピックを深く掘り下げるのではなく、概要を説明することで考慮されている。これにより、変化を許容する一方で、監査人は特定の知識を常に更新し続けるか、対応する専門家で監査チームを補強する必要がある。各監査は、その基礎となるビジネスケースと同様にユニークである。これらの考察は、多様な AI 技術に使用できる方法論を提示することを目的としている。

本稿ではまず、監査人に概要を理解してもらうために、考慮すべき事項を紹介する。附属書では、監査やアセスメントにおいて各トピックに対応するための質問を提案し、コンプライアンスを超えたリスク評価を容易にし、既成概念にとらわれない思考を促進する。

主な対象者は、上級管理職、規制当局、インテリジェント・システムのアセスメントや監査を行う人々である。

パート 1 と 2：どのように連携して利用するか

インテリジェントシステムをアセスメントするための本提案のパート 1 では、基本的な概念を説明し、必要な構成要素および／または使用される構成要素の包括的な理解を確立する。この基礎知識により、監査人はインテリジェントシステムの信頼性を評価するための基礎知識を身につけることができ、幅広い技術を可能な限り簡潔に網羅することができる。この準備により、批判的思考が可能になり、リスクアセスメントが容易になり、一般的なコンプライアンスの範囲を超える。

パート 2 は、本稿のパート 1 の各章に対応する潜在的な質問を付した附属書で構成されている。想定される質問を記載した実践的なリストは網羅的なものではないが、現在の視点から潜在的なリスクを特定するためのガイドラインとして役立つ。これらの質問をすることにより、型にはまらない思考を刺激し、既存の前提に挑戦することで、リスクアセスメントの実践を強化し、インテリジェントシステム全体の信頼性を高めることを目的としている。

定義

本節の目的は、AI をエンド・ツー・エンドでアセスメントするための基本的な概念、原 則、語彙を提供することである。この明確化は、ユーザを支援し、様々な国や業界において比較可能なアセスメント結果を支援するために、これらの基本的な概念や用語の監査人の解釈を調整することができる。認可者の意図は、監査人が効果的かつ効率的に AI の信頼性を評価することを支援することである。

ここで定義される用語は、本文書で言及されるトピック領域内、または関連する AI アセスメント手法内に含まれ、本手法の目的に関連するものである。

AI アセスメントの目的は、起こりうるリスクを明らかにすることである。これらのリスクの低減は、被監査者によってマネジメントされ、展開された AI がその信頼性を成熟させるにつれて起こりうる問題に対処することによってなされなければならない。インテリジェント・システムの改善に向けた重要な礎石は、AI のレジリエンスである。

AI レジリエンス：ロバスト性、レジリエンス、可塑性

AI のレジリエンスとは、さまざまな脅威や実際に影響を及ぼす事象に直面しても、その性能を維持・回復するインテリジェントシステムの能力のことである。AI レジリエンスには 3 つの柱がある [図 1]：

1. **堅牢性**：意図した性能や機能を損なうことなく脅威に耐える機能。
2. **レジリエンス**：パフォーマンスに影響を与えるインシデントが発生した後、緊張状態にある AI システムが正常なオペレーションを回復する能力。これは、AI システムが回復して以前の運用状態に戻る能力である（スナップバック）。
3. **可塑性**：機能変化や進化的適応によってインシデントから回復する能力。性能とレジリエンスは、システムが新たな課題に対応する能力を高める可能性がある。これはまた、"make it or break it "に対するシステムの耐性を示す尺度でもある。

可塑性は興味深く強力なオプションになりつつあるが、変更は追跡可能でなければならず、その機能は妥当性確認されなければならない。インシデント発生後に AI が機能不全に陥った場合、自己修復オプションの有無にかかわらず、そのスイッチを切る（キルスイッチ）ことができなければならない。

AI レジリエンスの核となるこれらの柱は、インテリジェントなシステムが混乱に耐えるだけでなく、混乱か ら学習し進化することで、時間の経過とともにシステムを改善できるようにするために連携する。

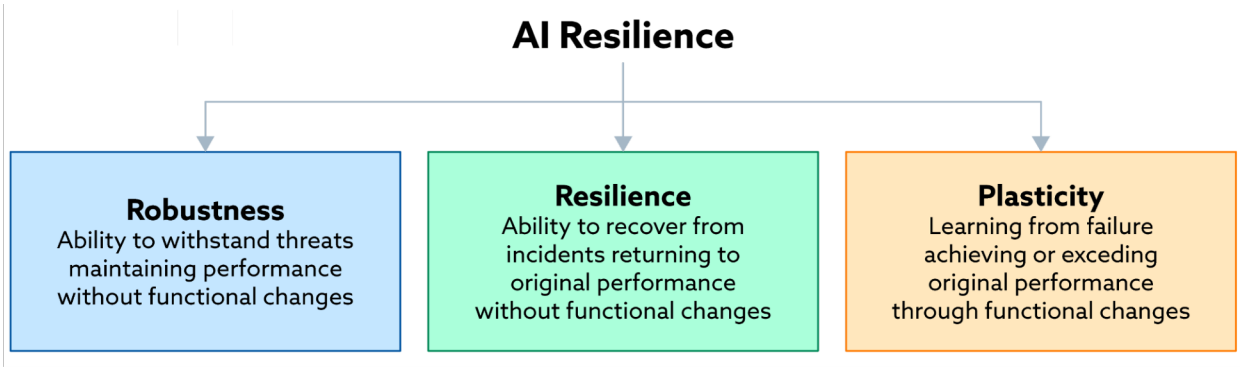


図 1：AI のレジリエンス[10]、[11]。

インテリジェント・システム：AI システム、AI エージェント、ロボット

AI システム：データ分析、意思決定、自然言語処理などのタスクを実行するために AI を利用するソフトウェアベースのアプリケーションを包括する広義の用語である。これらのシステムは、物理的な具現化なしに動作することができ、金融、医療、顧客サービスなどの分野でよく見られる。

AI エージェント：自律的に環境を認識し、意思決定を行い、事前に定義された目標の達成に向けて行動するように設計された、より具体的なタイプの知的システム。AI エージェントは多くの場合、センサーとエフェクターを使用し、（ソフトウェア・エージェントのような）仮想環境と物理的形態の両方に存在することができる。

ロボット：AI を使用して物理世界と相互作用する物理的事業体。ロボットは通常、AI システムやエージェントと、センサー、アクチュエーター、制御システムなどのハードウェア・コンポーネントを組み合わせ、単純な反復動作から複雑な適応動作まで、さまざまなタスクを実行する。

本稿では、AI を活用してタスクを実行するという共通の特徴を強調し、上記のすべてを効果的に包含する「**インテリジェント・システム**」という用語を選んだ。

説明責任、責任、賠償責任

これらの用語の関連性が高まっているため、さらなる定義が必要である。このような関連性の高まりは、サイバーセキュリティ、安全性、及びサードパーティリスクに関する上級マネジメントの説明責任と所有権を強調する最近の法律（例えば、NIS2、DORA、SEC、CRA）に よるところが大きい。このことは、インテリジェントシステムのリスクマネジメントと、リスク識別、低減の決定、及び是正措置の結果について、リーダーがどの程度責任及び／又は説明責任を負うことができるかという文脈に大いに関連している。シニアリーダーは、ますます AI 保証に責任を持つようになってきている。

説明責任は責任よりも高いリスクを伴い、どちらも法的義務（負債）を伴う。

説明責任：アカウンタビリティとは、監督を意味し、仕事の結果に対して責任を負うことを意味する。より高いレベルのオーナーシップを意味し、個人やグループが特定の行動や決定の成否に責任を負う。アカウンタビリティは一般的に単一であり、最高レベルでの結果報告と意思決定の正当化を伴う。

責任：責任とは、合意されたとおりに仕事を完了したり、役割を遵守したりする義務のことである。具体的な義務を課され、それが合意どおりに履行されることを約束し期待しながら、それを引き受けることを含む。責任はチームメンバー間で共有されることが多く、日常業務やより広範な目標が含まれることもある。

責任：責任とは、自らの行為や決定によって生じた損害や危害を賠償する法的義務のことである。個人または組織の経営委員会に、その責任と行動の結果に対する責任を負わせる法的枠組みを確立するものである。要するに、責任とは "職務に対するコミットメント "を意味し、説明責任とは "職務の機能結果に対して責任を負うこと "を意味する。責任とは、この 2 つの概念を法的結果と結びつけるものであり、説明責任と責任ある役割を担う人々の行動が、潜在的な法的影響によって拘束されることを保証するものである。本稿では説明責任に焦点を当てるが、責任ある行動と責任の自覚が求められる。

責任と説明責任の違いを理解することは、個人レベルでは重大な意味を持つ。リスクマネジメントは、起こりうる法的結果によって導かれるべきではないが、リスク行動はしばしば短期的な金銭的利益に影響される。新しい法律では、決断に対する個人の説明責任を軸とした長期的なリスクマネジメントという重要な視点が加わる。これには、短期的な利益から、より安全で持続可能な解決策への移行が必要であり、大きな文化的転換を意味する。

監査人のアセスメント

Springer による詳細な文献レビュー [12]は、監査手法は、その手法によって評価される技術とともに改善されなければならないことを強調している。インテリジェント・システムの監査は、異なる手法や個人のスキルセットが協力し、互いに補完し合わなければならない、多分野にまたがるアプローチである。効果的な監査は、技術指向、プロセス指向、ビジネス指向のアプローチを融合させなければならない。

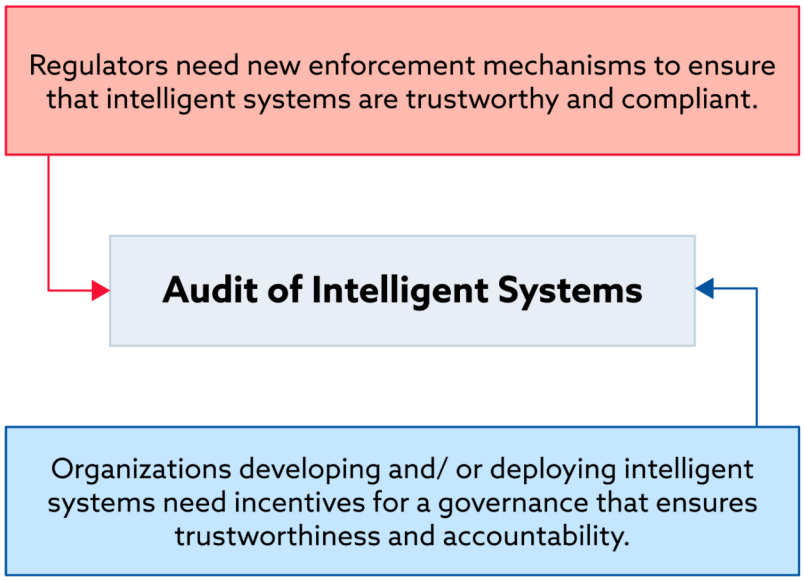


図 2：知的システムの監査（*Springer [12]*からの引用）

インテリジェント・システムのアセスメントは、監査手法、コンピュータサイエンス、ビジネス、哲学、心理学、業界専門家、法律学など多様な分野の専門知識と、コンプライアンス、リスク、ガバナンスに関する十分な経験を必要とする、ますます複雑化するプロセスである。

監査人の業務上および組織上の独立性は不可欠であり、セクターを超えた協力が推奨される。インテリジェント・システムのビジネス、環境、社会への影響を監査する際には、相互に関連した構造的なアセスメント・アプローチが推奨される。

監査の専門家としての効率性、有効性、誠実性を確保するために、AI 監査人自身の能力、スキル、経験、資格、ユースケースに関連した業務知識を評価する必要性が大きい。さらに、公平性を判断するために、個人的バイアス、知識不足、利益相反に関連する潜在的リスクを評価しなければならない。

規制団体や専門職団体は、監査実務を監視・保護し、妥当性確認のために AI 監査手順の標準を設定する上で、重要な役割を果たさなければならない。法的に AI 監査人が、必要な証拠や成果物にアクセスできるようにする一方で、得た知識はすべて、必要なときに、見る必要があり、知る必要がある場合に限って秘密にするために、拘束力のある秘密保持契約を結ばなければならない。

インテリジェント・システムの監査は複雑な範囲に及ぶため、有能な監査人からなる専門チー ムが、説明責任団体により選出され、監査の成功と結果の妥当性確認を確実にするために編成されなけ ればならない。インテリジェントシステム監査チームは、認可された最高監査責任者（CAO）または最高 監査責任者（CAE）が率いるものとし、最高監査責任者は、すべてのインテリジェントシステム監査活動 を監督し、結果を解釈し、インテリジェントシステムの信頼性についての意見または判断に関する所 見を文書化する。アセスメント担当者は、多分野にまたがる専門知識、批判的・調査的・抽象的思考能力を備えるだけでなく（「コンプライアンスのみ」にとどまらないリスクアセスメントを実施することができる）、現行の標準、規制、法律、その他のベストプラクティスを熟知し、その遵守を維持するための知識と経験を有していなければならない。

インテリジェントシステム技術が前進する一方で、規制当局、開発事業者、輸入事業者は、相互の努力のみがインテリジェントシステムの信頼できる有益なアセスメントをもたらすことを認識し、同意することが重要である。したがって、業界と規制団体が密接に協力し、インテリジェントシステムを評価するための輸入事業者のスキルアップを図ることが重要である。理想的には、監査人を評価し、そのスキルとパフォーマンスを追跡するために、認証可能な標準の開発が検討されるであろう。

AI ガバナンス

インテリジェントシステムの監査には、組織の文脈に合わせたアプローチが必要であり、監査を計画、実施、文書化する際に、各システムの重要性と該当するすべての組織特有の事項を考慮する。監査チームは、監査プログラム（マルチサイト、部門横断的、場合によっては事業体や国境/タイムゾーンをまたぐ）を設計する際に、批判的、調査的、抽象的思考を適宜適用しなければならない。

監査範囲、目標、成功指標を定義する（そして透明性を確保する）ためには、主要業績評価指標（KPI）、主要リスク評価指標（KRI）、またはその他の適用可能な測定値のような測定可能なインテリジェントシステム指標に合意し、それを使用することが不可欠である。

本稿では、（ISO/IEC 22989:2022[13]に沿った）インテリジェント・システムの信頼性に関する主要なアセスメント目標を挙げる。この分野は、技術面でも監査実務面でも急速に発展している。そのため、このリストは、監査範囲と監査方法を定義する際に、個々の監査ごとに監査チームが調整する必要がある。加えて、このようなアセスメントの計画や実地調査においては、批判的好奇心や調査的・抽象的思考と組み合わされたリスクベースのアプローチを常に適用しなければならない。

結果の文書化は、すべての利害関係者と合意しなければならない。

- **透明性**：スコープ内のすべての資産の可視性と、すべての物理的およびデジタルコンポーネントのエンドツーエンドのトレーサビリティ。
- **説明可能性**：知的システムまたはその設計者は、利害関係者（エンドユーザや監査人を含む）に対して、すべての意思決定プロセスを説明できなければならない。国立標準技術研究所（NIST）は、説明可能な AI（XAI）について、次の 4 つの原則[14]を提唱している：説明（出力の理由）、意味（意図されたエンドユーザにとって理解可能）、説明の正確さ（説明は、出力を生成する理由を正しく反映する）、知識の限界（システムは、それが設計された条件下で、および／または、その出力において十分な信頼性に達した場合にのみ動作する）。
- **予測可能性**：インテリジェントシステムの結果や挙動は予測可能でなければならない。関連するリスクによっては、これは許容可能な結果や挙動の範囲であったり、非常に厳しく規制されたものであったり、あるいはゼロ・トレランスのシナリオであったりする。
- **制御性**：制御可能性の確保は、細心の注意を払うべき複雑な課題である。インテリジェント・システムが制御性から逃れる方法を見つけることができるかどうかを調査する必要がある。インテリジェント・システムが人間の監視なしに機能するように設計されている場合は、明確に規制され、文書化されたガードレールを設置しなければならない。
- **信頼性**：アセスメントされるインテリジェント・システムは、設計された結果や行動の質と量を提供する信頼性があることを証明しなければならない。

- **公平性**：インテリジェント・システムが、人口統計学的グループ、社会経済的背景、多様な視点を含む複数の側面を代表する多様なデータセットで訓練されていることを確認し、公平な結果を促進する。
- **バイアス**：インテリジェント・システムは、それが意図された状況／用途において目的に適うものであることを保証するために、厳密にテストされなければならない。インテリジェント・システムの出力は、非倫理的な社会的条件によって識別されたり、影響を受けたりしてはならない。
- **AI のレジリエンス**：システムはロバストでなければならず、レジリエンシーでなければならず、その可塑性（継続的学習）は制御され、制御可能でなければならない。インテリジェント・システムの重要性に応じて、そのシステムは一つの目的に適合するだけでなく、変更されたパラメータや不慣れな環境下でテストされ、AI レジリエンスを証明する信頼できる結果を提供しなければならない。

加えて、以下の点も考慮すべき重要な点である：

- **解釈可能性**：これは、人間が出力を理解できるという要素に対処する。
- **説明責任**：インテリジェント・システムの動作、結果、意思決定（デジタルまたは物理的な影響に関係なく）に対する所有権と責任の明確な線引き。影響を受ける領域を可視化するために、責任共有モデルを確立することができる。
- **プライバシー・バイ・デザイン**：プライバシー保護措置の組み込みは、インテリジェント・システムの開発ライフサイクル全体を通して、初期設計段階から明らかでなければならない。
- **セキュリティ・バイ・デザイン（Security-by-design）**：安全な開発戦略が開発ライフサイクルを通じて適用され、すべてのコードとライブラリが検証されていなければならない。

オペレーション領域が高度であればあるほど、あるいはクリティカルであればあるほど、説明可能性と妥当性確認の必要性は高まる。特に LLM やその他のパターン生成システムでは、説明可能性は、ある出力へのパスが正しいのか、間違っている可能性があるのか、あるいは単に人間の理解を超えているのかを理解する上で重要な要素となる。この種の AI ガバナンスリスク（あるいは、開発者やその監査人が理解できないリスクをガバナンスできないこと）のアセスメントも、監査結果の中に文書化されなければならない。

適用法、規制、標準

適用法、業界標準、組織の内部方針を知り、確実に遵守することは極めて重要である。これは、全社的で革新的なリスクマネジメントと事業継続に対応する効果的なガバナンスを導入することによって達成される。インテリジェント・システムは、コンプライアンスとコンプライアンスを超えたリスクを評価するために監査可能でなければならない[15]。

クラウド・セキュリティ・アライアンス（CSA）は、AI ガバナンスとコンプライアンスに関する包括的なリソース・ハブを提供している[16]。このハブは CSA の AI ガバナンス・コンプライアンス・ワークグループによって作成され、ボランティアによって定期的に更新されている。様々な地域にわたる多様な AI トピックをカバーする膨大な文書コレクションをプロバイダとして提供している。無料で、完全に検索可能である。

出発点としてのユースケース

インテリジェントシステム所有者の助けを借りて、監査チーム（[1]）はユースケースを紹介し、インテリジェントシステムの目的とコンテキストの理解を促進しなければならない。コンプライアンスを確認する従来の監査では、システムは確立された要件、測定基準、および管理に対して判定される。これらの要件は、法律、規制、標準、組織、およびポリシーで定義され、地域や業界によって異なる場合がある。適用されるすべてのコンプライアンス要件、業界標準、ベストプラクティスに精通することは、評価者の責任である。コンプライアンスを超えたアセスメントの目標は、現在の技術革新のスピードが規制の能力を上回っているため、（まだ）規制団体によってカバーされていない新規の課題を特定することである。本稿は、可能性の境界線を押し広げることを目的としたインテリジェントシステムの信頼性を高めるための考察を提供することを意図している。したがって、このユースケースは、コンプライアンスを超えたアセスメントの観点から、特定のインテリジェントシステムに適合したメトリクスの確立を導くものである。

また、ユースケースは、エッジケースシナリオに直面した場合でも、望ましい結果や動作を達成するために、計画された、あるいは（法的に）要求される人間の介入の程度を定義しなければならない。人間の監視に関する明確なガイドラインがなければならない [15]。これらの要素は、ヒューマン・イン・ザ・ループ（人間による積極的な監視）、ヒューマン・アウト・オブ・ザ・ループ（人間によるオーバーライドのオプションなし）、ヒューマン・オン・ザ・ループ（人間が監視の役割を果たし、制御を引き継ぐことができる）として知られている。

監査人とエンドユーザは、システムの重要性によっては、ユースケースに適切なレベルの人間の関与が欠けていると、深刻なインシデントが発生する可能性があることを認識しなければならない。人間の監視を特に排除した方が安全なユースケースがあるかもしれない。そのようなユースケースの例としては、ある種の水力インフラがあり、そこでは人間の監視が許されるよりもはるかに速いスピードで意思決定が行われる必要がある。また、人間によるレビューのための「オプトアウト」または「オプトイン」オプションが実装されているかどうか、そして人間がインテリジェント・システムの決定に異議を唱えることができるかどうかもアセスメントされなければならない。これらの検討は、最初のユースケース検討の不可欠な部分であり、適切に計画、文書化、実施されなければ、監査結果に影響を与える。

監査範囲を明確にすることも重要である。対応計画には、監査人と主要な利害関係者の役割、責任、必要な利用可能性を含めなければならない。関連する規制、拘束力のある規範要求事項、会社の方針に関する知識が必要である。監査人とエンドユーザの安全、製品の品質とセキュリティ、データの機密性、

---

[1] 監査チーム：内部監査チーム、あらゆる種類の外部監査チーム（セカンドパーティやサードパーティを含む）、または規制機関の監査チームである。

完全性、可用性（CIA）、インシデント管理、コンプライアンス規準、事業継続性についても、監査計画の中で考慮し、監査中にレビューしなければならない。

この種のアセスメントを企業のリスクマネジメントプロセスと手順の中に組み込むことで、責任あるインテリジェントシステムの使用を保証し、上級マネジメントが責任を負うインテリジェントシステムの展開に対する認識を促すことができる。

インテリジェントシステムのユースケースと目的適合性をアセスメントする際、アセスメント担当者は、批判的思考、調査的思考、抽象的思考を適用する能力を持たなければならない。これらの方法を用いることで、監査人は新規のリスクやコンプライアンス違反の可能性を特定することができる。このような内部または外部の公平なレビューのプロセスを通じて、インテリジェントシステムのセキュリティと信頼性を高めるための改善勧告を行うことができる。このように、監査人は、ユースケースに基づく実際のパフォーマンスをレビューし、仮定に異議を唱え、現在のインテリジェントシステムの運用について適切な視点を提供することで、低減が必要なリスクを特定することができる。監査勧告が聞き入れられれば、インテリジェント AI システムの長期的なセキュリティ、安全性、信頼性にプラスの変化をもたらすことができる。

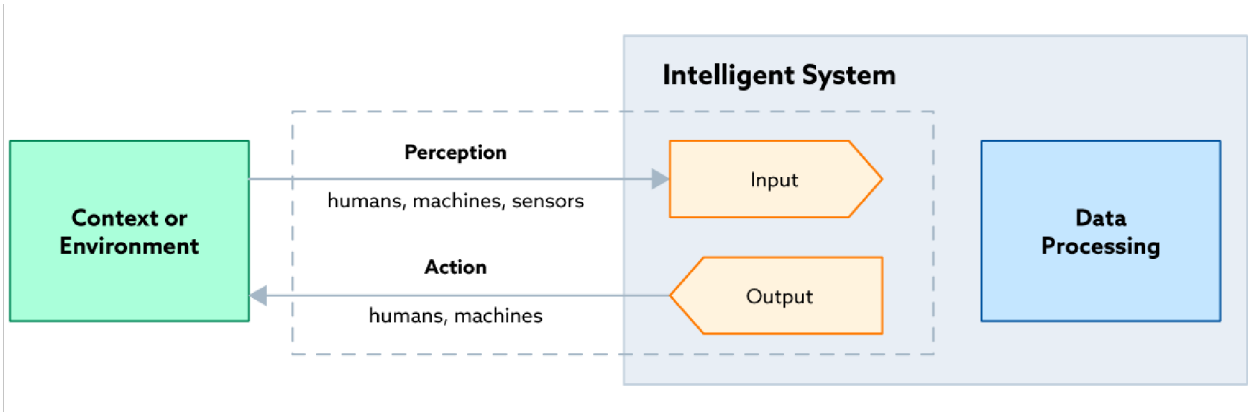


図 *3*：インテリジェント・システムの定義（*OECD* より引用）

以下のセクションでは、特定のインテリジェント・システムに当てはまるかどうかは別として、考慮すべき概念の概要を示す。

インテリジェント・システム・インフラ

インテリジェントシステムは機密データを扱うことが多いため、インフラのアセスメントは極めて重要であり、高度なプライバシー手法（後述）の実装が必要となる。このようなアセスメントは、計算能力、データ処理速度、帯域幅だけでなく、セキュリティ、プライバシー、スケーラビリティ、持続可能性など、インテリジェントシステムのパフォーマンスに対するサポートと影響について洞察するために不可欠である。インフラストラクチャーには、インテリジェント・システムの開発、訓練、展開、運用、維持に不可欠なハードウェア・コンポーネント、ネットワーク、ストレージ、オペレーティング・システム、ソフトウェア・アプリケーションが含まれる。

効果的なサイバーセキュリティ管理システム[17]を導入することは、特に機密性の高いデータを扱うインテリジェントシステムのレジリエンスを確保する上で最も重要である。NIST の既存の枠組みや、人工知能リスクマネジメントフレームワーク（AI RMF）[18]、GenAI RMF プロファイル[19]は、インテリジェントシステムのサイバーリスクやその他の重要なリスクをマネジメントするための基盤を提供している。将来的には、欧州連合サイバーセキュリティ機関（ENISA）の新たな AI サイバーセキュリティ実践の枠組み（FAICP）[20]のような標準が、インテリジェントシステムのライフサイクル全体を通してのセキュリティ確保について、より具体的なガイダンスを提供する。監査人は、これらの標準、枠組み、ベストプラクティスを厳格に遵守（適合）しているかをチェックしなければならない。

エネルギー効率と環境の持続可能性は、インテリジェント・システムのエコシステムにおいて不可欠な概念となっている。欧州連合（EU）では、カーボンフットプリントに関する法律がすでに施行されている。監査人は、エネルギー効率の高いハードウェアの採用、再生可能エネルギー源の利用、グリーン・コンピューティングの実践や電子廃棄物の責任ある処理による環境への影響の最小化など、エネルギー使用を最適化するための対策も評価しなければならない。ベストプラクティス、フロンティアソリューション、現行の規制に関して常に最新情報を入手することは、評価者の義務である。これらの複合的な努力は、AI の持続可能な発展に貢献し、世界的な気候変動緩和の努力と、国連が発表した持続可能な開発目標（SDGs）の目標 13（気候変動対策）に合致する[21]。インテリジェントシステムの長期的な信頼性を確保するために、監査人は、各システムのユースケース、IT アーキテクチャー・インフラストラクチャーの理解、および持続可能性要件を含むアセスメント規準に関する幅広い知識に基づいて、批判的、調査的、および現行の規制を超えて、事実に基づいた積極的な提案を行うことができなければならない。

データ処理ユニット

データ処理装置は、CPU、GPU、テンソル処理ユニット（TPU）、モノのインターネット（IoT）デバイス内にあるエッジプロセッサを含むインテリジェントシステムにとって極めて重要である。CPU は汎用コンピューティングタスクを処理し、GPU は複雑な計算の並列処理に優れ、TPU は機械学習ワークロードを最適化する。エッジ・プロセッサーはデータをローカルに処理することで効率を高め、IoT 環境における待ち時間を短縮する。将来的には、量子コンピュータのような新たな技術は、特定の計算の改善のために量子力学を活用することで、飛躍的な処理能力を約束する。バイオロジカル・コンピュータは、生物学的材料を処理に利用することを探求しており、バイオ・インスパイアード・コンピューティングの可能性を提供している。これらの技術革新は、多様な計算パラダイムが様々な領域でインテリジェント・システムの能力向上に貢献する未来を示唆している。

この分野では現在、信じられないような進歩が進んでおり、最新の洞察、最前線、ベストプラクティスに個人的な知識を継続的にアップデートすることが重要である。これらの新しいテクノロジーは、インテリジェント・システムの長期的な信頼性を確保するために、批判的思考と積極的な対策を講じる余地を十分に与えてくれる。

センサーの役割

視覚、聴覚、触覚、温度、近接、化学化合物、水分検出器など、あらゆる種類のセンサーが、インテリジェント・システムのデータ収集に重要な役割を果たしている。IoT は、センサー技術の進歩を素早く活用する、非常にダイナミックな分野である。表 1 に示すように、このデータはインテリジェント・システムによって様々な用途に利用される。

アプリケーション	説明
スマートシティ	センサーネットワークは、交通の流れ、エネルギー消費、環境状態を監視し、インテリジェントな都市計画と資源管理を可能にする。
精密農業	センサーは、土壌水分、養分レベル、作物の健康状態に関するリアルタイムのデータを提供し、灌漑、肥料使用、全体的な農業生産性を最適化することができる。
先進ロボット工学	次世代センサーは、ロボットの環境をより深く理解し、より安全なインタラクション、敏捷性の改善、複雑な環境での自律的な操作を可能にする。
個別化医療	バイオセンサーは、継続的な健康モニタリング、病気の早期発見、オーダーメイドの治療を可能にする。

表 *1*：インテリジェント・システムによるセンサー・データの応用

センサー技術の進歩は、3 つの重要な分野を通じて大きな飛躍を遂げようとしている：

- **小型化**（ナノテクノロジーも使用する）
- **コンテクスト認識**
- **センサー・フュージョン**

小型化によって、信じられないほど小さなセンサーの開発が可能になり、その潜在的用途が大幅に拡大する。コンテクストを認識することで、センサーは周囲の環境を解釈できるようになり、より豊かなデータと洞察に満ちた分析が可能になる。最後に、センサー・フュージョンは、複数のソースからのデータの統合を促進し、より包括的な環境画像を作成する。

メリット	説明
精度改善	コンテキストに基づいて無関係なデータをフィルタリングすることで、より正確で実用的な洞察が得られる。
オートメーションの強化	センサーはコンテキストに基づいて自動応答をトリガーすることができる。例えば、スマートライトは夜間に誰かが部屋に入った時だけ点灯することができる。
電力削減	関連するデータに焦点を当てることで、コンテキスト・アウェア・センサーは動作することができる。
消費	より効率的に、バッテリー寿命を延ばすことができる。
パーソナライズされた体験	センサーは、ユーザの嗜好や環境条件に合わせて動作を調整し、よりパーソナライズされた体験を生み出すことができる。

表 2：センサー技術の進歩がもたらすメリット

センサーの進歩は、過小評価されたリスクや新たなリスクをもたらす可能性がある。アセスメントでは、センサー機能の組み合わせと、その結果生じる怠慢、誤用、故意／悪意による濫用による潜在的危害を評価しなければならない。

センサーの所有者と実装者は、これらのセンサーの全体的な影響とリスクを包括的に理解するために、センサーデータから生成される望ましい洞察を評価者が理解するのを支援しなければならない。また、センサーの紛失やキャリブレーション不良が評価対象のインテリジェントシステムに与える影響も考慮しなければならない[22]。この情報の正確さと理解は、インテリジェントシステムの信頼性を評価する上で極めて重要である。

表 3 は、いくつかの潜在的な課題の概要を示している。このリストはすべてを網羅しているわけではなく、時間の経過とともに変化する可能性がある。

チャレンジ	説明
データ・セキュリティとプライバシー	コンテクスチュアル・データの収集と処理には、プライバシーとセキュリティに関する懸念がつきまとう。堅牢なデータ保護対策は極めて重要だが、現実的には難しい。
複雑さ	コンテキストを意識したセンサーシステムを開発・展開するには、センサー技術、データ分析、機械学習（ML）の専門知識が必要だ。

チャレンジ	説明
相互運用性	異なるセンサーやシステムがシームレスにコミュニケーションし、コンテクストデータを共有できるようにするには、標準化が必要だ。

表 *3*：センサー技術の進歩がもたらす課題

次の章では、インテリジェント・システムにおける複雑なデータ管理に光を当てる。

データの役割

データ[23]が新たな通貨であり、特にインテリジェント・システムにおいてビジネスの成功の鍵となる今日の世界では、グローバルなデータ規制、データ品質、データを取り巻く課題と機会（図 4）を理解することが極めて重要である。これらの要素は、ビジネスを「左右」する可能性がある。

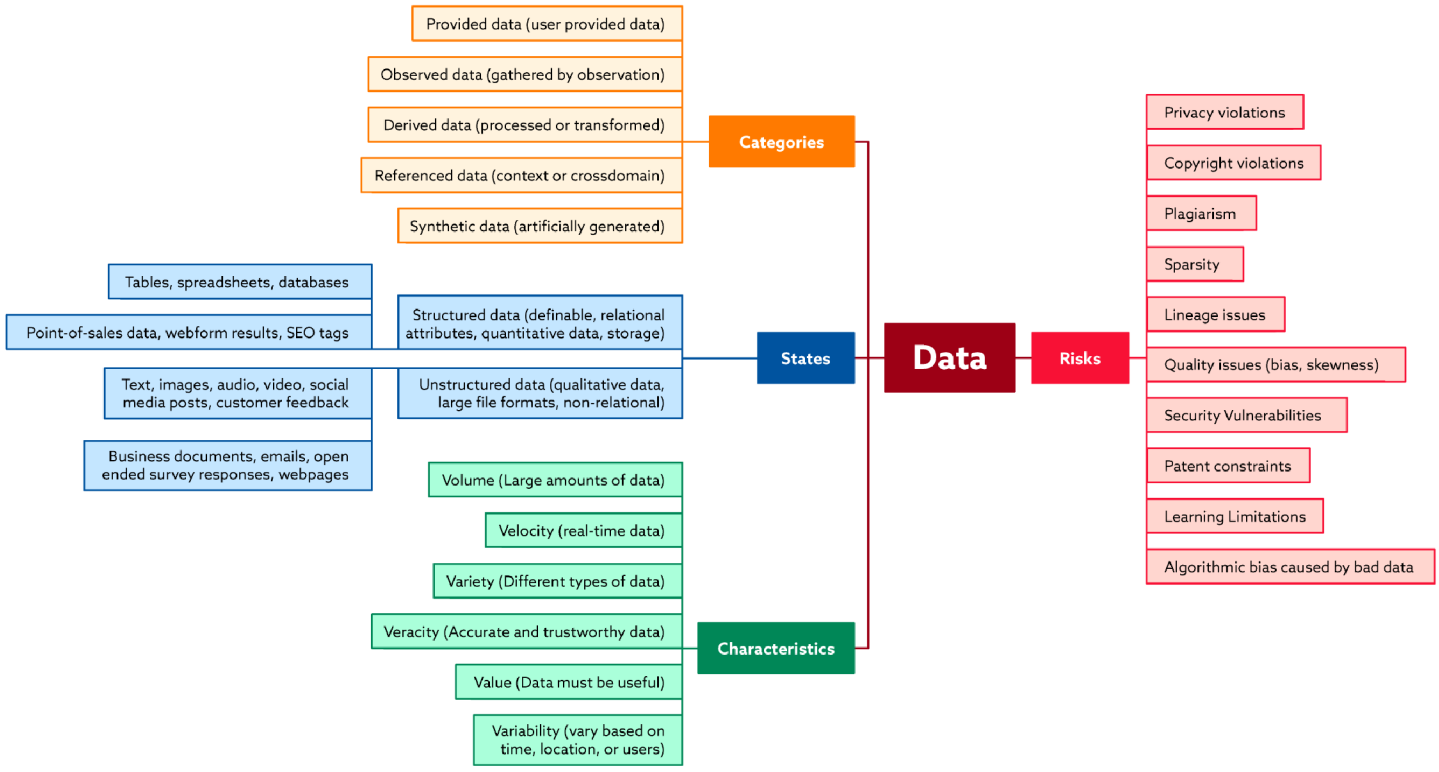


図 4：インテリジェント・システムにおけるデータの役割

サプライチェーンの役割

サードパーティ（[2]）から提供される AI コンポーネントを検討する際には、データ・プライバシー、機密性、完全性、可用性、バイアス、識別性、コンプライアンス、パフォーマンス、統合、説明可能性の欠如、AI を利用したサイバー攻撃、規制およびコンプライアンスリスクなどのリスクを考慮する。リスクマネジメントのため、インテリジェントシステムの部品はソフトウェア部品表（SAIBOM）でトレーサビリティを確保する必要がある。最近の出来事は、ソフトウェアまたはハードウェア・コンポーネントおよび／またはサイバーセキュリティ・サービスのプロバイダに対する効果的な賠償責任リスクマネジメントの必要性を示している。これらのリスクはサプライヤーのリスクマネジメントに含まれなければならない。EU やその他の規制当局は、サードパーティのリスクマネジメントを要求しているため、実際のベンダーのリスクアセスメントを行い、その重要性を判断しなければならない。インテリジェント・システム・アセスメントは、企業の IT およびサードパーティのリスクマネジメントに焦点を広げ、場合によっては、国境を越えたプライバシー規制（該当する場合）に焦点を当てたパートナー・ベンダーおよび／または重要なサプライヤーのアセスメントを含めなければならない。サードパーティは、合意された期限内に脆弱性やインシデントを開示するよう契約上義務付けられ、開示の遅延や不履行には多額の金銭的ペナルティが科されるべきである。

透明性とトレーサビリティを確保することは、信頼性を維持するために極めて重要であり、アセスメントの際にはこれらのレベルを評価しなければならない。

[2] サードパーティとは、開発者またはそのサプライヤ以外のサプライヤのことである。

データ・プライバシー

様々なグローバル規制がデータを管理し、しばしば域外適用される。例えば、一般データ保護規則（GDPR）[24]は、たとえ EU 域外で処理されたとしても、EU 事業体のほとんどの個人データに適用される。データ保護影響アセスメント（DPIA）[25]は、企業がコンプライアンス違反を特定し、ギャップを修正するのに役立つ。同様に、カリフォルニア州消費者プライバシー法（CCPA）[26]、カリフォルニア州プライバシー権法（CPRA）[26]、スイス連邦データ保護法（FADP）[27]は、同等のデータ使用制約を課している。

グローバル化した世界では、異なる法律を注意深くナビゲートしなければならない。多くの場合、明示的な同意が必要であり、適用されるプライバシー法の例外は、データ対象者が明示的に受け入れなければならない。忘れられる権利」（GDPR）と「記憶する義務」（規制産業におけるデータ保持法と EU の AI 法）は矛盾する概念であり、インテリジェント・システムにおけるデータ・プライバシーを複雑にしている。アセスメントでは、特にデータ保持が必須または不可避であり（例えば、トレーニングデータとして使用される）、GDPR に準拠していない場合、明示的かつインフォームド・コンセントを文書化する透明で法的に健全な同意管理システムを企業が有していることを確認しなければならない。

データと著作権

知的システムのトレーニングに使用されるデータは、外部の知的財産や著作物が含まれている可能性があるため、著作権に関する懸念が生じる可能性がある。最新の出版物については、CSA AI Resource Hub [16]を参照されたい。あるシステムで生成された AI 的コンテンツが、別のシステムのトレーニング・データになる可能性があるため、これは複雑な課題となる。著作権で保護されたコンテンツが学習ループに入ると、知的システムの「共有知識」から切り離すことが難しくなる。したがって、アセスメントは、著作権侵害を防止し、著作権で保護されたコンテンツの削除をサポートするプロセス（例えば、デジタルミレニアム著作権法（DMCA）[28]の通知とテイクダウンシステム）を導入するために、データのソースと適用法への準拠をチェックしなければならない。

データソース

今後の課題として、利用可能な高品質の[3] トレーニングデータがボトルネックになる可能性があり、早ければ 2026 年に不足する可能性がある[29], [30]。しかし、これらの予測は、大量監視、自動運転車、IoT の世界などのソースからの利用可能なデータの増加や、合成データの生成を考慮していない。

[3] 質の高いデータは、書籍、科学論文、ウィキペディア、フィルタリングされたウェブページなどで構成されている。

データの闇市場[31]は、個人を特定できる情報（PII）、金融データ、医療記録、ログイン認証情報を含む違法に入手されたデータセットで繁栄しており、個人と企業の両方に重大なリスクをもたらしている。

データの血統」の透明性は、インテリジェント・システムのアセスメントに関連するトピックになるだろう。産業界は、綿密で追跡可能なデータ管理を実証する準備をしなければならない。これは、データがシステム内を移動する際に、データの出所から最終的な場所まで、また途中で加えられた変更に至るまで、データの系譜を保証するのに役立つ。

有機データと合成データ

オーガニック・データは、実世界のシナリオや行動を反映し、その正確性と関連性から信頼される、ニュアンスに富んだ多様なパターンを捉えることで、認証性を提供する。しかし、特殊な領域での利用可能性の制限、機密情報によるプライバシーリスク、入手・保存・マネジメントの高コストなどの課題に直面している。対照的に、合成データはスケーラビリティに優れ、プライバシーを守りながら精密な操作を可能にする。しかし、有機的な複雑性を完全に捉えるにはリアリズムに欠ける可能性があり、事前学習によって知的システムが意図しないパターンを学習してしまう可能性もある。

事前学習に合成データを使用し、微調整にオーガニックデータを使用する組み合わせは、特定の状況下において、より優れた性能を発揮することが証明されている[32]。監査では、データの出所と、有機データと合成データを使用する理由は、透明性があり、説明可能で、目的が明確でなければならない。合成データを生成・販売する企業を監査するには、その生成方法を深く理解する必要がある。このデータにおける意図的でない、あるいは意図的なパターンは、インテリジェント・システムにおける予期せぬ、潜在的に有害な動作につながる可能性がある。

この分野は進化しており、新しい進歩を包含しているため、輸入事業者にとっては、最新の進歩、ベストプラクティス、標準、法律、規制に精通していることが重要である。

データの質

データの質とは、正確性（誤りがないこと）、完全性（期待される属性がすべて含まれていること）、一貫性（形式や値が統一されていること）、信頼性（一貫して信頼できること）、適時性（最新であること）、妥当性（現実世界の対象や事象を正しく代表すること）、使用目的との関連性を指す。質の高いデータは、十分な情報に基づいた意思決定を行い、信頼できる分析結果を確保し、データ駆動型システムの信頼を醸成するために極めて重要である。データの質に影響を与える要因には、収集方法、保管方法、妥当性確認プロセスなどがあり、これらはエラー、バイアス、不整合を最小限に抑えることを目的としている。継続的なモニタリングと保守は、長期にわたってデータ品質を維持するために不可欠であり、データ品質が目的に適合し、意味のある洞察をサポートすることを保証する。データ品質を維持することは、意思決定や分析のためにデータに依存している組織にとって極めて重要である。

データ保管

データ・ストレージ・ソリューションはインテリジェント・システムにとって不可欠であり、膨大な量の情報を効率的に管理・処理するためのさまざまな技術を包含している。ソリッド・ステート・ドライブ（SSD）[4] やハード・ディスク・ドライブ（HDD）[5] のような内部ストレージ・オプションは、ローカルに保存されたデータへの高速アクセスを提供し、ランダム・アクセス・メモリー（RAM）[6] は、即時の処理ニーズに対応した迅速なデータ検索を可能にする。クラウド・ストレージ・ソリューションは、大規模なデータセットの保存と分析のためのスケーラブルでアクセス可能なリポジトリを提供し、柔軟性を高め、分散環境にわたるインテリジェント・システムのコラボレーションを促進する。インテリジェント・システムの監査では、パフォーマンス、セキュリティ、安全なデータ保存容量を維持するために十分なリソースがあるかどうかをチェックすることが重要である。

データ量が急増するにつれ、研究者は先進的なストレージ方法を模索している。ホログラフィック・データ・ストレージ[33]は、迅速な検索を伴う高密度ストレージとして有望であり、DNA ベースのストレージ[34]は、長期的なデジタル・ストレージ・ソリューションを提供する。このような新しいデータ・ストレージは、ビジネス・ケースに適しているかどうかを監査で批判的に評価しなければならない。

---

[4] SSD は HDD よりも高速で、消費電力も少ない。

[5] HDD は、現代の代替品と比べると古く、遅い技術だ。[6] RAM はアクティブに処理されるデータに使われる。

ネットワーキング、コネクティビティ、コミュニケーション・インターフェイス

サイバーセキュリティでは、AI を活用した新たなセキュリティ・ソリューションと、耐量子暗号[35]を含む暗号化手法の進歩が、将来の課題を克服する上で重要な役割を果たすだろう。

コネクティビティとコミュニケーション・インターフェイスは、インテリジェント・システムにとって極めて重要であり、シームレスなデータ交換と堅牢なサイバーセキュリティを確保するために、幅広いテクノロジーを包含している。これらには、WLAN、イーサネット、セルラー（4G/5G）といった従来のネットワーキング手法が含まれる。Bluetooth、Bluetooth Low Energy (BLE)、RFID などの近距離無線通信プロトコルも不可欠である。さらに、LoRaWAN、Zigbee、Z-Wave、NB-IoT、Sigfox といった IoT に特化したプロトコルは、低消費電力、長距離、短距離のコミュニケーションにさまざまな選択肢を提供する。

マシン間コミュニケーションや高性能データ交換といった特定のニーズを満たすために、OPC UA、DDS、CoAP といった産業用通信プロトコルやリアルタイム通信プロトコルが利用されている。これらのシステムはそれぞれ、無線通信の脆弱性から、制約のあるネットワーク上で伝送されるデータの安全性まで、独自のサイバーセキュリティ上の課題をもたらす。これらの課題を管理することは、AI 駆動型アプリケーションにおけるデータの完全性、機密性、可用性を確保するために不可欠である。

今後、6G ネットワークの発展は、接続性の向上とさらなる低遅延を約束する。量子通信や新しい IoT プロトコルの進歩により、状況はさらに変化すると予想される。

フォグ・コンピューティングとクラウド・コンピューティング

フォグ・コンピューティング・ノードは、センサーやアクチュエーターのようなデータソースの近くに戦略的に配置され、ネットワーク・エッジとクラウド・データセンター間の仲介役を果たす。これらのノードは、その処理能力とストレージ能力により、ローカルデータの分析、フィルタリング、前処理を行い、レイテンシーと帯域幅の使用を削減する。インテリジェントなゲートウェイとして機能するこれらのノードは、クラウドレベルの処理が必要なデータとローカルで処理できるデータを判断し、クラウド全体の負荷を最適化し、リアルタイムのアプリケーション・パフォーマンスを確保する。このようにフォグとクラウド・コンピューティングが統合されることで、効率性が向上し、インターネットが途絶えた場合でも継続的な運用が可能になる。

今後、フォグコンピューティングは、自律走行車、スマートシティ、産業用 IoT など、新たなテクノロジーにおいて極めて重要な役割を果たすと予想される。フォグレベルでの予測分析やエッジアプリケーション（[6] ）のためのインテリジェントシステムの統合は、意思決定能力をさらに高めるだろう。さらに、NIST による軽量暗号の Ascon [36]ファミリーの標準化など、サイバーセキュリティ対策の進歩により、フォグコンピューティング [37]特有の課題への対応が期待される。

AI がフォグやクラウド・コンピューティング・システムに統合されることで、データの完全性と品質の維持に新たな複雑なレイヤーが導入され、監査プロセスがこれまで以上に重要になる。この文脈で AI システムを監査するには、使用されるデータの品質を検証するだけでなく、フォグとクラウドの両レベルで採用されるアルゴリズムとモデルを評価する必要がある。これには、バイアスを特定し、規制標準へのコンプライアンスを確保し、AI 出力の全体的な信頼性を検証するための継続的なモニタリングと評価が必要である。フォグ・コンピューティング・ノードがデータの前処理やローカル分析でより多くの責任を担うようになると、監査はこれらの分散環境にも適用され、ローカルで行われる決定が組織の目標や倫理的配慮に沿ったものであることを保証する必要がある。

[6] ネットワークエッジのデバイス上で動作するソフトウェアで、ローカルでデータを処理することにより、レスポンスタイムを高速化し、集中型クラウドリソースへの依存を軽減する。

ソフトウェア・コンポーネント

ソフトウェア・コンポーネントは、インテリジェント・システムの運用と管理に不可欠である。

仮想化ソフトウェアは、仮想マシン間での効率的なリソース割り当てと管理を容易にする。オペレーティングシステムは、ハードウェアとソフトウェアの相互作用を基礎からサポートする。ミドルウェアは、異種のアプリケーションやシステム間のコミュニケーションを容易にする。アプリケーション・ソフトウェアと開発環境は、インテリジェントなシステムとアプリケーションのアルゴリズムの作成と実行を可能にする。データ管理ツールは、大規模なデータセットの効率的な保存、検索、処理を保証する。モニタリングとロギング・ツールは、システムのパフォーマンスとセキュリティを追跡する。コンテナ・オーケストレーション・プラットフォームは、分散環境にまたがるインテリジェント・システム・アプリケーションの展開と管理を合理化する。これらのコンポーネントが一体となってインテリジェント・システム・インフラのバックボーンを形成し、複雑なコンピューティング環境におけるスケーラビリティ、効率性、信頼性を支えている。

アルゴリズム、トレーニング方法、モデル

知能システムは、アルゴリズム、学習方法、モデルという 3 つの主要な構成要素に依存している。これらの構成要素は、部分的に互いに依存し合い、明確な境界のない絡み合った依存関係を示す。

アルゴリズムは、モデルがデータからどのように学習するかの青写真として、核となる学習方法を定義する。学習方法は、アルゴリズムの選択、データの準備、プロジェクト目標に基づく最適な方法の選択など、学習プロセス全体を包含する。学習方法は、獲得した知識を具現化し、推論エンジンを通じて新しいデータの予測に使用されるモデルの選択に影響を与える。商用またはオープンソースの枠組みは通常、インテリジェントシステムの開発、トレーニング、評価、展開を容易にする様々な機能、ライブラリ、API を提供している。

これらの主要コンポーネントの信頼性と効率を維持することは、生産環境において、効果的、効率的で信頼できるインテリジェント・システムを確保するために極めて重要である。

これらのトピックとその相互依存関係に関する文献は、教科書、オンラインコース、科学出版物、ディスカッションフォーラムなどで見ることができる。依存関係を完全に理解し、批判的、調査的でありながら公平な視点からインテリジェントシステムをアセスメントする能力を持つためには、各トピックを詳細に理解することが必要である。

過去には、意思決定における知的システムの潜在的な落とし穴を例証する、データ、アルゴリズム、トレーニング、モデルに関するいくつかの問題があった [10]。EU の AI 法（第 5 条）[38]、[39]やコロラド州の上院法案 24-205[40]のような規制は、リスクの高いアプリケーションを表示し、責任ある使用を奨励することによって、これらの懸念に対処するための措置を講じている。これは、システムを批判的にアセスメントし、リスクを低減するための事前対策を提案することで、コンプライアンスを真摯に遵守し、現在のコンプライアンス標準を超えても信頼できるインテリジェントシステムを展開することの重要性を強調している。

アルゴリズムとモデル

適切なアルゴリズムとモデルの選択には、特定の問題と利用可能なデータに基づく様々な重要な考慮事項が含まれる。これはすぐにかなり専門的になるが、インテリジェント・システムのリスクを評価できるようになるためには、輸入事業者が基礎となる技術を十分に把握することが重要である。

また、アルゴリズムやモデルの複雑さと、利用可能な計算能力とのバランスをとるための計算資源のアセスメントも必要である。スケーラビリティの必要性、ノイズに対する頑健性、アルゴリズムやモデルが欠損データやデータの質の低さをどの程度処理できるかを考慮する。

HITL（Human in the Loop）とは、機械学習や情報活動において、人間が ML モデルの訓練や評価に積極的に参加し、貴重なガイダンスやフィードバック、注釈を提供する協調的アプローチである。このコ

ラボレーションを通じて、HITL は人間と機械双方のユニークな能力を活用し、ML システムの精度、信頼性、適応性を高めることを目指している。HITL はまた、ML モデルやインテリジェント・システムを運用する上で極めて重要な概念であり、インテリジェント・システムの重要性に応じて法的に要求されることもある。

以下のアプローチは、多様なアプリケーションや課題に対応し、インテリジェント・システム開発のツールキットを拡張する：

- **教師あり学習アルゴリズム：**これらのアルゴリズムは、ラベル付けされたデータから学習し、望ましい出力がプロバイダから提供される。例えば、決定木、線形回帰、サポート・ベクトル・マシンなどがある。
- **教師なし学習アルゴリズム：**これらのアルゴリズムは、ラベル付けされていないデータからパターンを発見する。例としては、K-means のようなクラスタリングアルゴリズムや、主成分分析（PCA）[7] やオートエンコーダ[8] のような次元削減技術がある。
- **半教師あり学習：**ラベル付けされたデータとラベル付けされていないデータを組み合わせて、学習の精度と効率を改善する。
- **転移学習：**学習元のタスクや領域から得た知識を、学習対象のタスクや領域での学習やパフォーマンスを改善するために適応させる。
- **オンライン学習（インクリメンタルラーニング）：**新しいデータが入手可能になるとモデルを継続的に更新し、進化するパターンやトレンドに適応する。
- **マルチタスク学習：**複数のタスクを同時に学習し、関連するタスクの汎化と効率を改善する。
- **能動学習：**最小限のラベル付けデータでモデルの性能を向上させるために、ラベル付けのための新しいデータサンプルの選択をガイドする。
- **模倣学習（実演から学ぶ）：**専門家の実演や行動例から学び、望ましい行動や振る舞いを再現する。
- **強化学習アルゴリズム：**これらのアルゴリズムは、環境との試行錯誤を通じて学習し、望ましい行動に対して報酬を受け取る。例えば、Q-learning や Deep Q-Networks などがある。
- **自然言語処理（NLP）アルゴリズム：**人間の言語を処理・分析することに焦点を当てた機械学習のサブ分野。分析には、感情分析、機械翻訳、テキスト要約などが含まれる。

---

[7] PCA は、データセットの本質的な情報を保持したまま変数の数を減らすために使われる手法である。

[8] オートエンコーダは、データを低次元表現に圧縮し、この圧縮された形から元のデータを再構築するように設計されたニューラルネットワークである。オートエンコーダーは、データセットの複雑なパターンや関係を捉える能力を活用し、次元削減、異常検知、データ生成などのタスクに使用される汎用的なツールである。

- **コンピュータ・ビジョンのアルゴリズム：**このサブフィールドは、画像や映像の分析に関連するタスクに焦点を当てている。

  検知には、物体検知、画像分類、顔認識などがある。

- **進化的アルゴリズム：**自然淘汰に着想を得た問題解決手法で、解の候補が繁殖、突然変異、淘汰などの過程を経て進化し、最適解を見つける。

こうした最先端のアプローチは、可能性の限界を押し広げ、新たなアプリケーションへの扉を開くもの：

- **連合学習アルゴリズム：**生データそのものを共有することなく、複数のデバイスやサーバー間で協調的にモデルを学習する。このアプローチは、プライバシー保護アプリケーションや分散データシナリオに有益である。
- **生成的敵対ネットワーク（GAN）：**生成器と識別器という 2 つの競合するニューラルネットワークを含む。生成器が新しいデータ（画像やテキストなど）を作成し、識別器が実際のデータと区別しようとする。この競争により、非常にリアルなコンテンツが生成される。
- **教師あり学習：**入力データからラベルを生成するための事前タスクを使用し、人間による注釈なしでモデル学習を強化する。
- **メタ学習（学習するための学習）：**異なるタスクやドメインから学習する方法を学ぶことで、モデルの学習と適応を改善する。
- **数ショット学習：**非常に限られたデータセットでモデルを学習させ、わずかな例から学習させることを目的とする。これは、大量のラベル付きデータを取得することが高価であったり困難であったりするタスクにおいて特に有用である。
- **ワンショット学習：**単一の例または少数の例から学習し、新しい未知の事例に汎化する。
- **ゼロショット学習：**対象領域からラベル付けされた例がなくても、対象物を分類できるように、領域を超えて知識を転送する。
- **説明可能な AI（XAI）：**機械学習モデルが複雑化するにつれ、その意思決定プロセスを理解する必要性が高まっている。XAI 技術は、これらのモデルをより透明で解釈可能なものにし、人間が彼らの推論をよりよく理解できるようにすることを目的としている。

このリストはすべてを網羅することを意図したものではなく、主流とフロンティアの両方のアプローチを含み、変更される可能性がある。批判的な姿勢を保ちつつ、オープンマインドで各監査に臨むことを確認する。

フロンティア

インテリジェント・システムの未来が模索される中、量子機械学習とニューロモーフィック・コンピューティングが、この分野に革命をもたらす可能性を秘めたエキサイティングなフロンティアとして浮上している。量子機械学習[41]は、古典的な能力を超える問題に取り組む能力を提供し、ニューロモーフィック・コンピューティング[42]、[43]は、脳の効率的な情報処理に着想を得たハードウェアの実現を目指す。どちらの技術も初期段階ではあるが、処理能力における現在の限界を克服し、研究のブレークスルーをもたらし、インテリジェント・システムの二酸化炭素排出量を削減できる可能性がある。新たな可能性と同時に新たなリスクも生じ、監査人は未知の領域に直面するかもしれない。

モデルの微調整と妥当性確認

微調整と妥当性確認は、AI システムのモデルの有効性を確保するための重要なステップである。トレーニングは、モデルを既存のデータにさらすことで、パターンを学習させ、予測や洞察を生成させる。微調整は、事前学習したレイヤーを特定のタスク用に調整することでモデルを改良し、その能力を高める。別のデータセットによる妥当性確認では、モデルの汎化能力を評価し、トレーニング・データへのオーバーフィッティングを防止することで、実世界のアプリケーションでの正確な出力を保証する。

オーバーフィットと汎化

AI システムのモデルが、新しい未知のデータに対してうまく汎化されるようにするには、データセットをトレーニングセットとテストセットに分割する、交差性確認を適用する、学習曲線を分析する、正則化手法を使用する、ハイパーパラメータを調整するなどの手法がある。これらの手法は、モデルがオーバーフィットしているか、うまく汎化しているかを評価し、全体的なパフォーマンスを改善するための調整を導くのに役立つ。

微調整

微調整とは、事前学習したモデルのパラメータを、より少ないタスク固有のデータセットを使って調整し、特定のタスクやドメインに適応させることである。このプロセスは、事前学習よりも少ない計算リソースで済み、低グレードのハードウェアでも実行できる。低ランク適応（LoRA）のような革新的な技術により、微調整はより効率的で費用対効果が高くなった。また、微調整には人間のフィードバックを取り入れることで、モデルをより望ましい動作に近づけることができる。

妥当性確認

妥当性確認は、代表的な検証データセットを使用して、新しい未知の入力に対する出力を予測する際に、学習済みモデルがどのように動作するかを評価する。このプロセスでは、正確度、精度、再現性、F1[9] スコアなどの指標を計算し、モデルの有効性を公平に評価する。性能テストは、さまざまな作業負荷やリアルタイムのシナリオを処理するモデルの能力を評価し、潜在的な制限やボトルネックを特定する。計算効率と応答時間を測定し、モデルのアーキテクチャと実装を最適化する。

さらに、インテリジェントシステムにおける倫理への注目の高まりは、IBM の AI Fairness 360 や Google の What-If Tool のような公平性とバイアス検知の枠組みを妥当性確認プロセスに統合すること

[9] F1 スコアは、精度と想起のバランスをとっており、偽陽性と偽陰性の両方を考慮することで、不均衡なデータセットにおけるモデルの精度を評価するのに有用である。

につながり、モデルの正確性だけでなく公平性も確保することになった。XAI 技術の台頭もモデルの透明性と信頼性を高めている。

モデルの頑健性と安定性

AI システムのモデルのロバスト性を評価するには、敵対的な攻撃に抵抗する能力、ノイズや破損したデータでも性能を維持する能力、多様なシナリオに汎化する能力を評価する必要がある。安定性は、モデルが予期せぬ入力を処理し、透明性のある意思決定プロセスを提供し、変化するデータ分布に適応し、リスクとバイアスを最小化できることを保証する。

継続保守

継続的な妥当性、有効性、信頼性を確保するために、AI システムのモデル、アルゴリズム、パフォーマンス指標の継続的なテスト、更新、改良の手順を確立する。

微調整とモデル妥当性確認の最前線

AI システムの分野が進化を続ける中、微調整と妥当性確認における新たなトレンドが、モデルの最適化と評価方法を再定義しつつある。連合学習は、データ・プライバシーを保護し、規制コンプライアンスに対応しながら、分散化されたデバイス間でモデルの学習を可能にするもので、ヘルスケアや金融において特に価値がある。さらに、ゼロショットや少数ショットの学習技術が人気を集めており、多様なデータセットでの広範な事前学習を活用することで、モデルが最小限の微調整でタスクを実行できるようになっている。このような技術革新は、モデルの適応サイクルを短縮することができる。

さらに、継続学習システムの採用が一般的になりつつあり、モデルがリアルタイムで新しいデータに適応し、ダイナミックな環境において適切な状態を維持できるようになっている。微調整のために合成データまたは有機データをそれぞれ使用することで、このプロセスの効率を高めることができる。

アクチュエーター

ロボットは、コマンドを物理的な動作に変換するために、主にアクチュエータに依存している。これらのアクチュエーターには、AC、DC、ステッピングモーターなどの一般的な回転式オプションや、正確な動作のためのサーボが含まれる。空気圧シリンダーのようなリニアアクチュエータは、押したり引いたりする動作に使われる。機械式、磁気式、真空式などさまざまなグリッパーは、ロボットが物体を把持することを可能にし、車輪やトラックは移動性を提供する。アクチュエータの選択は、ロボットの力、速度、精度、移動範囲に対する特定のニーズによって決まる。

アクチュエーターの未来は明るい。生物の動きを模倣した人工筋肉や、非常に器用な操作を可能にするマイクロ流体技術などの可能性がある。これらの進歩は、より多用途で効率的なロボットを生み出す大きな可能性を秘めている。しかし、潜在的なリスクとしては、誤動作や誤用による予期せぬ結果や、ますます人間に近くなっていくロボットを取り巻く倫理的配慮に注意深く対処する必要がある。

電源

インテリジェント・システムの電源は、その物理的な形態によって異なる。ロボットやその他の具現化されたインテリジェント・システムは、持ち運び可能だが容量に制限のあるバッテリー、継続的な電力を供給するが移動に制限のある AC 電源、または柔軟性を高めるために両方を組み合わせたハイブリッド・システムに依存している。クラウドベースのインテリジェント・システムやエージェントは、データセンターの電力インフラを活用する。スマートな電力管理システムは、インテリジェント・システム全体の消費量を最適化するために極めて重要である。

物理的な具現化を伴う知的システム（ロボット）の場合：

- **バッテリー：**ポータブルで充電可能だが、容量に限りがあり、充電が必要。
- **AC 電源：**コンセントへの直接接続で連続運転が可能だが、機動性が制限される。
- **ハイブリッド・システム：**バッテリーと AC 電源を組み合わせ、携帯性と長時間稼働の両方を実現する。

物理的な実体を持たないインテリジェント・システム（クラウドベースのシステム）の場合：

- **データセンターの電力：**データセンターで稼働する大規模インテリジェントシステムは、データセンターの全体的な電力インフラに依存しており、この電力インフラには、グリッド電力、バックアップ発電機、再生可能エネルギー源の組み合わせが含まれる場合がある。
- **IoT とその他の小型デバイス：**これらのデバイスは多くの場合、バッテリー、エネルギーハーベスティング技術、または低電力ワイヤレス充電に依存している。小型で、遠隔地や手の届きにくい場所に設置されることが多いため、電力効率は極めて重要である。低消費電力チップ設計、効率的なスリープモード、断続的な接続性などの戦略は、バッテリー寿命を延ばし、信頼性の高い動作を確保するために不可欠である。

さらに考慮すべきことがある：

- **電力管理システム：**具現化されたインテリジェント・システムと具現化されていないインテリジェント・システムの両方の電力消費を最適化し、効率的な運用を保証する。
- **ワイヤレス充電：**ロボットやその他の具現化されたインテリジェント・システムのための新しい技術で、ケーブルの必要性をなくし、機動性を高める。

このリストは、インテリジェント・システム用の主な電源オプションを網羅しているが、具体的なニーズは、システムのタイプ、処理能力要件、運用環境によって異なることを覚えておくことが重要である。

インテリジェント・システムを含む重要インフラへの電力供給には、多面的なセキュリティ・アプローチが必要である。送電網の脆弱性は、レジリエンスを強化し大規模停電を防ぐために、冗長電源とフォ

ールトトレラント送電網を必要とする。さらに、電磁パルス（EMP）[10] に対するシールドと、サイバー攻撃から電力管理システムを保護することも極めて重要である。

[10] 電子機器を損傷させる強力な電磁エネルギーの爆発。EMP は落雷などの自然現象によって引き起こされることもある。

ユーザ・インターフェース

最終的には、開発者の間に説明責任を果たす文化を醸成し、ユーザ・インターフェース（UI）の革新が魅力的であるだけでなく、信頼性の最高標準に準拠していることを保証することは、指導者の義務である。コンプライアンスを確認し、必要であれば積極的な提案をすることは、評価者の義務である。

ユーザ・インターフェースは、従来のディスプレイやパネルから、音声インターフェースやハンズフリー用のモバイル・アプリに至るまで、多様なニーズに合わせて進化している。タッチスクリーンやジェスチャー認識は、インタラクションをさらに効率化する。

拡張現実（AR）、仮想現実（VR）、ブレイン・コンピュータ・インターフェイス（BCI）、ハプティック・フィードバック（触覚フィードバック）[11] を含む未来の UI ランドスケープは、没入的でアクセシブルな体験に向けられている。AR/VR は手術にリアルタイムのデータオーバーレイを提供し、BCI は思考ベースのコントロールを可能にする。触覚フィードバックによるエンゲージメントの強化と、アクセシビリティへの配慮の重視は、様々な分野に革命をもたらす。

AI が統合された UI は、ユーザのニーズにリアルタイムで適応することで、パーソナライズされた体験へのシフトを約束するものであり、個々の要件に合わせてプレゼンテーションやインタラクションをカスタマイズすることで、アクセシビリティを顕著に改善する可能性がある。このパーソナライゼーションへの動きは、利便性と包括性を高める役割を果たすだろう。

超パーソナライゼーションは、そのメリットにもかかわらず、情報エクスポージャーの多様性を低下させ、創造性やコラボレーションに影響を与える可能性があるなどのリスクを伴う。パーソナライゼーションの利点と同時に、UI エクスペリエンスの多様性を維持するためには、バランスの取れたニュアンスのアプローチが不可欠である。

将来の UI の進歩は、セキュリティ、安全性、プライバシーに関する重要な検討をもたらす。BCI による広範なデータ収集は強固な保護を要求し、倫理的な問題を提起する。さらに、AR/VR の没入的な性質は、状況認識を維持するために高度な安全防御策を必要とし、相互接続された UI ランドスケープは新たな脆弱性を提示し、潜在的な脅威に対する厳格な防御の必要性を強調している。

監査人の観点からは、開発者とともに新しい UI をアセスメントすることは、その潜在的な利益と、安全性、プライバシー、セキュリティの懸念を含む固有のリスクを徹底的に理解するために極めて重要である。監査人は、様々な UI のリスクプロファイルを体系的に評価し、安全、セキュリティ、プライバシーを維持するための対応するガードレールを差し引くことができなければならない。防御は、魅力的な UI

[11] ユーザ・インターフェースにおける触覚フィードバックとは、触覚を使って、物理的な感触を通してユーザに情報を伝えることを指す。

の追求がユーザを保護する必要性を覆い隠さないことを保証し、ユーザ参加と倫理的責任のバランスをとるデザイン哲学を提唱する上で重要な役割を果たす。

制御システム

インテリジェント・システムは、意思決定を（物理的な）行動に移すために様々な制御システムを利用する。集中制御は単純ではあるが、脆弱性に悩まされる。

- 分散型制御システムは、制御を分散させることでロバスト性と適応性を高める。
- 階層的制御システムは、高レベルのプランニングから低レベルの実行まで、複数のレベルにわたって制御システムを組織化し、階層的な意思決定によって複雑な行動を可能にする。
- 行動ベースの制御システムは、あらかじめ定義された状況では優れているが、柔軟性に欠ける。
- ハイブリッド・アプローチは、異なるコントロールを組み合わせることの長所を活用し、よりバランスの取れたソリューションを提供する。

将来の制御システムは、進化をシミュレートして制御システムを最適化する進化的アルゴリズム（自己最適化アルゴリズム）のような技術を通じて、適応学習やリアルタイム最適化を取り入れ、さらに高度なものになることが期待されている。これにより、意思決定に優れ、シームレスでダイナミックな環境との相互作用を持つロボットが誕生する可能性がある。

しかし、このような複雑なシステムにおいて信頼性を確保することは最も重要である。ハイブリッドシステムのエッジで分散制御を行うことで、意図しない自律的行動によるリスクや相互接続による脆弱性を軽減できる可能性がある。責任ある開発には、慎重な設計と強固なセキュリティ・プロトコルが不可欠である。

安全システム

信頼できるインテリジェントシステムの開発には、クリティカルシンキングと厳格なリスクアセスメントが不可欠である。インテリジェント・システムは、意図しない結果を防ぐためのソフトウェアのセーフガードに依存しているため、安全な運用が特に重要である。これには、危機的な状況で自動的に運転を停止するフェイルセーフ機構や、意思決定における誤動作を検知・回避するように設計されたアルゴリズムが含まれる。

具現化されたエージェントやロボットにとって、さらなる物理的安全対策は極めて重要である。衝突検知・回避システム、緊急停止機構、囲いのような物理的バリアはすべて、事故のリスクを最小限に抑えるために連携している。重要部品の冗長性により、1 つの部品が故障しても安全な機能が継続する。

以下のリストは、インテリジェント・システムの安全対策をまとめたものである：

- **物理的な保護措置：**これには、機械的な障壁、囲い、衝突の検知と回避、身体的危害を防ぐための速度や力の制限などが含まれる。
- **センサー・フュージョン：**複数のセンサーからのデータを組み合わせることで、より包括的で文脈を考慮した環境画像を作成し、安全なナビゲーションを支援する。
- **耐障害性：**重要なコンポーネントに冗長性を持たせてシステムを設計することで、1 つのコンポーネントが故障しても安全に稼働し続けることができる。
- **安全ソフトウェア：**ソフトウェアアルゴリズムはシステムの動作を監視し、安全でない動作を防ぐために介入することができる。これには、緊急停止や機能制限のトリガーが含まれる。すべての高度でインテリジェントなシステムには、製造事業者の緊急停止に似た「キルスイッチ」があるはずだ。
- **人間とロボットの共同作業の安全プロトコル：**ロボットが人間と一緒に働くようになると、事故を防ぎ、人間とロボットの間の明確なコミュニケーションを確保するために、特定のプロトコルや安全対策が必要になる。

自律システムやロボットにおいて倫理的な意思決定を保証することは、依然として課題である。物理的な）行動の予期せぬ結果や、深く相互接続されたシステムから生じる脆弱性は、安全上の危険、データ・違反、プライバシー侵害につながる可能性がある。OSHA [44] は、ロボットを扱う際に生じる伝統的なリスクや危険の多くと、いくつかの安全上の注意点を強調している。

あらゆるレベルにおいて斬新で革新的、かつ信頼性を高めた製品は、明日の技術的展望において競争力のある違いを生み出すことができる。コンプライアンスを超えたインテリジェント・システムの監査の観点からも、そのようなシステムは、この世界をより信頼できる場所にするのに役立つだろう。

セキュリティ・システム

インテリジェント・システムのセキュリティ・システムは、脅威や脆弱性から保護するために極めて重要である。これらのシステムには、データ伝送を安全にする暗号化、不正アクセスを制限するアクセス管理、システム・アーキテクチャ、ソフトウェア・コード、システム構成の脆弱性を特定する脆弱性スキャン、潜在的な侵害を特定し対応する侵入検知システムなどが含まれる。攻撃的なセキュリティ・システムと予防的なセキュリティ対策に、定期的なセキュリティ監査とタイムリーなアップデートを組み合わせることで、高レベルの保護が保証される。セキュア・ブート・メカニズムとファームウェアの整合性チェックは、悪意のあるソフトウェアからシステムを保護する。インテリジェント・システム導入のリスクは広く議論されているが、人間の失敗は依然として最も一般的なセキュリティ問題である。群衆を "ヒューマン・ファイアウォール "に変えるために必要な文化的変化を達成することは、非常に難しいように思われる。継続的なトレーニング、報奨金、安全な行動、リーダーシップの支持といった戦略は、この変革を助けることができる。

ここでは、インテリジェント・システムに不可欠なセキュリティ対策の概要を説明する：

- **暗号化**：データの保存と送信の両方を暗号化することで、データの機密性を確保する。二重の暗号化技術を実装することで、将来の量子コンピューティングの脅威に対するレジリエンスに備えることができる。
- **データのサニタイゼーション：**不要なデータや期限切れのデータを確実に削除し、高度なツールを使っても復元できないようにする。この対策は、機密情報を保護するために特に重要である。
- **認証：**パスワード、トークン、生体認証、多要素認証（MFA）などの方法を利用して、アクセスを要求するユーザやデバイスの身元を確認し、認可された事業体のみがアクセスできるようにする。
- **アクセス管理：**システムリソースへのアクセスを制御し、事前に定義されたルールに基づいて権限を割り当てることで、機密データや機能への不正アクセスを防止する。
- **セキュリティ情報・イベント管理（SIEM）：**複数のソースからのセキュリティデータを統合・分析し、脅威検知、インシデント対応、包括的なセキュリティ監視のための一元化されたプラットフォームを提供する。
- **侵入検知システム（IDS）：**ネットワーク・トラフィックとシステム・アクティビティを継続的に監視し、異常検知やシグネチャ・ベース検知などによって、潜在的なセキュリティ侵害を示す不審な行動を特定し、警告する。
- **継続的な監視：**システム活動の継続的な監視を実施して、異常や潜在的なセキュリティインシデントをリアルタイムで検知し、積極的な防御策を可能にする。
- **脆弱性管理：**定期的な更新とパッチ適用を通じて、システムアーキテ クチャ、構成、ソフトウェアおよびハードウェアコンポー ネントのセキュリティ脆弱性を積極的に特定し、低減する。

- **インシデント対応計画：**セキュリティ侵害やサイバー攻撃に対処・管理するための構造的なアプローチを策定・維持し、被害を最小限に抑え、復旧に要する時間とコストを削減し、事業の継続性を維持する方法で事態に対処することを目指す。
- **災害復旧**：サイバー攻撃や自然災害などの重大な障害が発生した場合に、重要なデータやシステムを復旧させるための計画や仕組みがあることを確認する。
- **物理的セキュリティ：**インテリジェント・システムの物理的ハードウェア・コンポーネントを盗難、改ざん、損傷から防御する。これは、見落とされがちだが、セキュリティの重要な側面である。
- **セキュリティ意識向上およびトレーニング：**インテリジェント・システムのユーザとオペレータに、潜在的なセキュリティ・リスク、ベスト・プラクティス、インシデント対応手順を教育し、信頼できるセキュリティ体制を維持する。
- **ペネトレーションテストとバグ報奨金**ペンテストは EU の新しい規制によって特定の重要産業に対して義務付けられているが、バグ報奨金プログラムは任意だが、悪意ある行為者に悪用される前に脆弱性を発見する上で大きな利益をもたらす。

最後になるが、定期的な監査は、セキュリティ対策の有効性を評価し、業界標準への 準拠を確認し、改善すべき領域を特定するために重要である。

AI 主導の（攻撃型）セキュリティ・ツールとゼロ・トラスト・アーキテクチャの台頭が、インテリジェント・システム・セキュリティの未来を形成している。技術的なセキュリティも、脆弱性を自動的に検出してパッチを当てることができる自己修復システムによる進歩が期待されている。このような進歩により、進化する脅威にリアルタイムで適応する、より安全なシステムが実現する。しかし、課題も残されている。システムが相互に接続されているため、複雑な攻撃対象が急速に拡大しているのだ。これらのリスクを低減するには、高度なセキュリティテクノロジーと継続的な警戒、リスクアセスメント、そして理想的にはこれらのリスクに対する文化的意識の改善を組み合わせた多層的なアプローチも必要である。

高度なプライバシー手法

この概要は、プライバシーに関する要求が厳しいインテリジェント・システムを評価するために不可欠である。プライバシーに関する懸念が最小限または全くないシステムについては、このセクションの関連性は低いかもしれないので、必要に応じて見直すことができる。

インテリジェントシステム、特に機密データを扱うシステムのライフサイクル全体を通して、プライバシーを向上させる技術は極めて重要である。このセクションでは、様々な開発段階における主要な手法の概要を提供し、以前に議論された概念とプライバシーに特化した概念の両方を要約する。これらの手法のいくつかはまだ改良中であることに注意してほしい。ベストプラクティスを超えて必要な知識をアップデートし、フロンティアの手法やその利点、起こりうるリスクについて常に情報を得ることは、査定者の義務である。

インフラストラクチャー：

- **データの居住：**プライバシー規制を遵守するため、特定の地域または国においてデータを保存・処理する。
- **分散型ストレージ：**データを各拠点に分散し、データ漏洩や自然災害の影響を最小限に抑える。
- **ハードウェア・セキュリティ・モジュール（HSM）：**これらの耐タンパー性デバイスは、データ暗号化のための暗号キーを保護し、物理的セキュリティのレイヤーを追加する。

データの収集と準備：

- **データの最小化：**AI モデルの機能に厳密に必要なデータのみを収集する。
- **データの匿名化と仮名化：**データを変換し、トレーニングのために統計的特性を保持しながら、個人情報を削除またはマスクする。
- **差分プライバシー[45]：**学習データにノイズを注入し、個人のプライバシーを保護しながらモデルの精度を改善する数学的手法。
- **連合学習 [46]：**データ自体を転送することなく、分散化されたデータセット上で AI モデルを学習する。各デバイスは自分のデータでローカルモデルを学習し、生データではなくモデルの重みのみを共有する。

トレーニングと能力開発：

- **合成データ：**実際のデータに似ているが、実際の個人情報を含まない人工的なデータ。これは、トレーニングや微調整のために実際のデータセットを補完するために使用することができる。
- **準同型暗号：**暗号化されたデータを復号化せずに分析できる。しかし、現在のところ計算速度が遅いため、リアルタイムのアプリケーションや大規模なデータセットには適していない。

- **プライバシーを保持する機械学習アルゴリズム：**これらのアルゴリズムは、モデル作成プロセス中に学習データについて明らかになる情報を最小限に抑えるように特別に設計されている。

展開と運営：

- **セキュア・マルチパーティ・コンピューティング（SMPC）：**複数の関係者が、互いに自分のデータを明かすことなく、自分のデータに対して共同で関数を計算することを可能にする機能。プライバシーを保護したデータ・分析に利用できる。
- **コンフィデンシャル・コンピューティング [47]（信頼可能な実行環境（TEE）の活用）：**この技術は、AI システムのモデル内でデータが処理されている間、データを分離して保護する。メインシステムが侵害されたとしても、攻撃者は TEE のセキュアなエンクレーブ内でデータ自体にアクセスすることはできない。

アセスメントは、確立された標準、業界のベストプラクティス、および規制要件に照らして評価することにより、インテリジェントシステムで使用される技法の適切性を評価する。監査人が対象分野の専門家（SME）でない場合は、監査スタッフを増強すべきである。

開発とデバッグ

開発ライフサイクル全体を通じて、様々なツールがインテリジェント・システムの作成と改良を可能にする。安全でないサプライ・チェーンはインテリジェント・システムに脆弱性をもたらす可能性があるため、サプライ・チェーンには特に注意を払う必要がある[48]。共同開発プラットフォームは、コード共同パイロット、バージョン管理、以前のバージョンに戻す機能を提供し、堅牢なソフトウェア・エンジニアリングを促進する。デバッグ・ツールはコーディング・エラーの特定に役立ち、パフォーマンス・モニタリング・ツールは最適な効率を保証する。機械学習枠組みには、インテリジェント・システムの開発用に調整されたツールやライブラリが装備されており、シミュレーション環境は、AI システム・モデルの訓練とテストを行うための安全な空間を提供する。

自動テストフレームワークは、主にコード機能の確保とバグの検出に重点を置いているが、アプリケーションレベルでは大きな成長の可能性を秘めている。プロファイリング・ツールは、開発者がコードのパフォーマンスを分析し、効率のボトルネックを突き止めることを可能にする。説明可能な AI（XAI）ツールは、効果的なデバッグとパフォーマンスの最適化にとって重要な、モデルがどのように意思決定を行うかを理解することで開発者を支援する。

将来の進歩により、さらに洗練されたツールが期待される一方で、自動化への過度の依存には潜在的なリスクがある。この依存は予期せぬ結果を招く可能性があり、ツールの信頼性の重要性を浮き彫りにしている。これらのツールの完全性は、結果として得られるインテリジェント・システムの信頼性に大きく影響するため、徹底的な監査とこの分野に関する深い知識の必要性が強調される。

インテリジェント・システムが自己最適化するように進化するにつれ、サードパーティ・サプライヤーへの依存度は低下するかもしれない。しかし、部品が予測不可能であればあるほど、その部品固有のばらつきは大きくなり、したがって、人間の監視にもかかわらず予期せぬ動作や検出されない誤動作のリスクが高まる。

エンドユーザ・トレーニングとドキュメンテーション

エンドユーザ向け文書

インテリジェントシステムの詳細なエンドユーザ向け文書は、透明性、トレーニング、保守、コンプライアンス、ビジネスプロセスへの統合を成功させるために不可欠である。NIST AI RMF は、能力、制限、ユースケース、技術データ、パフォーマンス指標、セキュリティ管理者、ユーザガイド、保守、サポート、ベストプラクティスなど、AI システムの文書化に必要な要素を概説している。ドキュメンテーションは、明確で、簡潔で、読者に特化し、さまざまなアクセシビリティ経路に対応し、定期的に更新されるべきである。主な要素は以下のとおりである：

- **ユーザガイド：**構造化されたユーザガイドは、エンドユーザがインテリジェントシステムを効果的に使用することを支援する。ユーザガイドには、リスクを強調したクイックスタートガイド（必要であれば免責事項、および／または法的責任を含む）、システムアーキテクチャ、プライバシーとセキュリティのためのデータマネジメント、設定ガイドライン、ユーザ・インターフェースの説明、想定されるユースケース、トラブルシューティング、用語集、およびエンドユーザ固有のフィードバックが可能なサポート連絡先情報が含まれるべきである。明確な指示、詳細な説明、視覚的な補助は、ユーザがシステムを素早く習得するのに役立ち、特定のインテリジェント・システムに対するオンライン・トレーニングは、システムへのアクセスを許可する前に提供されたり、義務付けられたりする可能性がある。
- **保守ガイド**インテリジェントシステムの保守には、最低限、継続的なモニタリング、セキュリティおよびインシデント対応、バージョン管理（必要に応じて以前のバージョンに戻す可能性の概要を含む）、構成管理、定期保守（継続的に学習するシステムの適切なデータ管理を含む）、コンプライアンス（SAIBOM の更新を含む）が必要である。システムのアップグレード、セキュリティチェック、データバックアップのスケジュールを管理する。技術的トラブルシューティングを文書化し、インテリジェント・システムの技術的問題を開示するための技術サポート窓口とオンラインフォームを提供する。
- **リリースノートまたは変更履歴：**リリースノートまたは変更履歴は、システムの変更（バグ修正を含む）、更新（新機能を含む）、改善を要約したものである。これらは、インテリジェントなシステムがコンプライアンスを維持し、エンドユーザに変更された機能や新たなトレーニングの必要性についての情報を提供するために、規制産業に関連するものである。
- **API リファレンス・ドキュメント：**API リファレンス・ドキュメントは、開発者がインテリジェント・システムに関連する API を使用するのを助け、開発と統合を容易にする。これは API の設計、

構造、機能、入力、出力、パラメータ、そして場合によってはAPIアクセスコードについて詳述している。APIは標準化された統合に不可欠であり、自動化に役立つ。特にアクションがトリガーされる可能性がある場合、そのアクセスは管理されなければならない。

- **情報開示：**インテリジェント・システムがどのようにデータを使用し、どのように生成的AIがコンテンツを生成するかを開示し、透明性を確保する。インテリジェント・システムと対話する場合、ユーザは情報を得なければならない。この開示は、倫理的・法的な理由から、ユーザの誤解を招かないようにし、「ヒューマン・イン・ザ・ループ」のコンセプトをサポートするために極めて重要である。
- **透明性：**透明性報告書は、アーキテクチャやトレーニング・データのような技術的側面への一般公開を提供し、公平性、バイアス、プライバシー、セキュリティに対処するための措置の概要を説明する。これらの報告書は、責任と説明責任のあるインテリジェントなシステム開発を実証するものであり、特徴や改善点を強調するものである。これは継続的な改善につながり、エンドユーザが斬新な解決策を提案することを可能にする。
- **プライバシーと機密性：**インテリジェント・システムのプライバシー・機密保持に関する通知は、データの収集、使用、保護、アクセス、保存、保持、およびデータの権利についてユーザに通知する。これはユーザが自分のデータがサポートする特定のアプリケーションを理解し、法的義務を果たすのに役立つ。
- **AI利用方針：**インテリジェントシステム使用方針は、責任ある効果的な使用のためのガイドライン、意図されるデータ使用、システムの範囲と目的、認可ユーザまたは認可ユーザになるための要件を概説する。インテリジェントシステムの使用を組織の価値観、目標、法的要件と整合させ、セキュリティと安全プロトコルを含む。
- **SAIBOMとデータストレージ：**ソフトウェア部品表には、物理的な材料、ソフトウェア、AIコンポーネントを含めなければならない。データの物理的保管場所は、監査証跡と関連ログの場所と同様に、文書化されなければならない。
- **運用状況の監視：**インテリジェント・システムの運用状況監視は、利用データ、応答時間、エラー率、可用性を追跡することで、サービスの信頼性と必要なパフォーマンスを保証する。問題を迅速に特定・解決し、パフォーマンス標準を確実に満たすことで、ダウンタイムを最小限に抑え、安全な運用状態を維持することができる。
- **ヘルプセンター：**十分なリソースを備えた、よく設計された有能なヘルプセンターは、信頼性が高く説明責任のあるインテリジェントシステム製品サポートを提供し、ユーザの信頼、満足、忠誠心を築くというコミットメントを示す。ヘルプセンターは、ユーザガイド、チュートリアル、トレーニング、よくある質問（FAQ）、トラブルシューティングのヒント、テクニカルサポートなどの包括的なリソースを提供し、ユーザが必要とするヘルプを迅速かつ専門的に得られるようにする。

- **責任ある／説明可能な AI の実践：**責任ある／説明可能な AI の実践は、知的システムの倫理的かつ持続可能な開発と利用を保証し、危害を最小限に抑え、利益を最大化する。このような実践には、多様なデータセットの使用、バイアスの低減、透明性の確保、プライバシーの保護、人間による監視が含まれる。
- **トレーニングデータに使用されている著作物の開示：**著作権の制限を含め、トレーニングデータの出典を開示することで、透明性と法令遵守を確保し、侵害リスクを低減する。サードパーティーのコンテンツを倫理的に尊重して使用することで、責任ある倫理的な事業活動を実証する。
- **エンドユーザと影響を受けたコミュニティのためのフィードバック・プロセス：**公平性、正確性、全体的な利益を確保し、改善の可能性を導くために、エンドユーザや影響を受けるコミュニティが、システムの成果やコンテンツの品質について意見を提供するためのフィードバック・プロセスは極めて重要である。アンケート調査、ユーザ行動分析、ユーザテスト、ソーシャルメディアモニタリングなど、さまざまなフィードバック方法がある。得られたフィードバックは、改善点の優先順位付けやユーザ体験の向上に利用できる。カスタマーサポートは、ユーザの経験や困難を反映し、改善点の特定に役立つ。

(義務）エンドユーザに対するトレーニングとトレーニング記録の保管

ISO 42001 [8]は、インテリジェントシステムの開発または使用に携わる個人のためのトレーニングガイドラインを提供し、必要なスキルを特定している。トレーニングは、原則、方法、倫理、データ保護、サイバーセキュリティ、法的要件をカバーする必要がある。トレーニングは構造化され、参加型に設計され、評価とフィードバックが含まれなければならない。出席、評価、資格などの文書化を維持する必要がある（時にはコンプライアンス上も）。研修リソースは、多様な対象者がアクセスでき、理解しやすく、定期的に更新され、オンラインチュートリアルや文書が含まれていなければならない。理想的には、学習効果を高め、必要とされるスキルの信頼性と適用可能性を確保するために、（場合によってはシミュレートされた環境やサンドボックスでの）実地訓練、ゲーミフィケーション、その他の魅力的な方法が可能なリソースである。

最新機能に関するウェビナーは、エンドユーザと交流するのに最適な方法であり、また専門家から回答を得ることもできる。オプションの追加情報を含む正確で最新のナレッジベースは、エンドユーザにとってさらに役立つリソースとなる。

コミュニティ・フォーラムは、他のユーザから回答を得るだけでなく、開発者に概要を伝え、一般的な問題や深刻な問題を発見するための優れた方法である。コラボレーション・プラットフォームは、アプリケーションをより深く理解したい開発者やエンドユーザに有益な情報を提供することが多い。

インテリジェント・システムの全ユーザが適切なトレーニングを受け（トレーニング履歴を文書化し）、定期的に更新されるようにすることは、信頼性を高め、損害や事業損失につながる意図しない誤用を低減する。知識ギャップを特定するための適切なトレーニングは、作業品質の向上に役立ち、法的要件やコンプライアンスもサポートする。トレーナー・アプローチでは、知識の確実な伝達を確実にするために、トレーナーの技術的・社会的スキルを確認することが極めて重要である。

展開とモニタリング

インテリジェント・システムの展開と監視段階を監査することは、システムが意図したとおりに機能し、不測の事態に確実かつ安全に対処できることを保証するために重要である。

継続的な評価：リアルタイムのパフォーマンス監視を実施し、主要目標を追跡し、ドリフトを検知する。ドリフトとは、インテリジェント・システムの目的（目標）からの逸脱と定義され、些細な不正確さにつながる可能性がある。しかし、自己修復能力や進化的アルゴリズムと組み合わせることで、ドリフトは重大な混乱を引き起こし、現時点では定量化すらできないリスクをもたらす可能性もある。ありふれたことのように聞こえるが、継続的な評価はインテリジェント・システムの信頼性を維持するために最も重要なコンセプトのひとつである。

バージョン管理と文書化：システムの完全性を維持し、将来の監査を容易にするために、徹底した変更追跡と詳細な文書化を維持する。明確な監査証跡を伴う信頼性の高い変更管理プロセスは、透明性と説明責任を高める。

インシデント対応：故障やセキュリティ違反に対処するための詳細な手順を確立する。インテリジェント・システムが具現化されているか否かにかかわらず、アクセス可能な「キル・スイッチ」または他の緊急シャットダウン方法を備えていなければならない。インシデントに迅速かつ確実に対処するためには、技術的・組織的な対策を講じなければならない。さらに、NIS2 や SEC のようなサイバーセキュリティ規制を遵守するには、迅速かつ適切な情報開示と対応プロトコルを持つ必要がある。

サイバーセキュリティ：展開と監視のライフサイクル段階において、最も重要な点は以下の通りである：

- **アクセス管理**：AI システムの不正使用を防ぐため、厳格なアクセス管理メカニズムを導入する。これにより、認可された担当者のみがシステムの機密部分とやり取りできるようになり、悪意のある活動や物理的な行為のリスクを低減できる。アクセス管理を他のデジタル事業体にも拡張し、ゼロトラスト・アーキテクチャを実装し、あらかじめ定義された、あるいはランダムな時間間隔でアクセス権をリセットする。
- **検知**：プロアクティブな監視ツールを使用して、悪意のある活動や異常を特定し、対応する。侵入の早期検知は、潜在的な損害を最小限に抑え、システムの完全性と信頼性を維持するために極めて重要である。
- **脆弱性スキャンとパッチ適用**：脆弱性を定期的にスキャンし、適時パッチを適用する。潜在的な悪用からシステムを保護し、継続的で安全な運用を確保するためには、脆弱性を識別し、迅速に対処することが不可欠である。脆弱性情報公開プログラム（VDP）、バグ報奨金、あるいはハニーポットは、侵害のリスクを低減することができるが、高度な持続的脅威（APT）やサイレント・コンプロマイズに対しては、このような戦術は有効ではないかもしれない。

規制環境におけるリリース管理：AI システムのアップデートや新バージョンの展開において厳格な標準を遵守し、安全性と規制要件へのコンプライアンスを確保する。コンプライアンスは法的に極めて重要であるが、リスクアセスメントは単なるコンプライアンスにとどまらないことを確認することが不可欠である。インテリジェント・システムが意図したとおりに動作し、不測の事態に安全に対処できることを検証する。

廃止措置

インテリジェント・システムの安全な廃止には、機密情報を保護するために保存されているすべてのデータを安全に消去すること、環境規制に従ってハードウェアを解体すること、適切な通知とともにすべてのアクセス管理を無効化または削除すること、ソフトウェア・ライセンスやサードパーティとの統合を適切に終了することが含まれる。適切なプロセスの文書化は、記録、説明責任、将来の参照のために不可欠である。監査人は、危険物、予期せぬ動作、事業継続への脅威など、潜在的な解体の危険性を評価する必要がある。

このフェーズには、いくつかの重要な要素が含まれる。第一に、監査人は、データ移行プロセスが安全であり、意思決定プロセスが中断されることなく、ロジックが維持され、または後継システムで更新され、業務および事業活動が継続されることを検証しなければならない。

緊急廃止には、強固な事業継続計画が必要である。事業継続性が確保されていれば、誤ったインテリジェント・システムを即座に停止できる（あるいは実際に停止している）可能性が大幅に高まる。

監査人は、組織が必要なデータを安全に保持し、古くなったデータを廃棄するための明確で実行可能なポリシーを持っていることを確認し、機密情報がすべてのストレージメディア（独自の倫理的課題をもたらす可能性のある生物学的ストレージオプションを含む）から確実に永久に削除されるようにしなければならない。ログとメタデータは、将来の参照、法令遵守、フォレンジック調査のために、安全にアーカイブ されなければならない。監査人は、これらのログが、データの完全性と機密性を保護しつつ、容易に検索できる方法で保存されていることを確認すべきである。今後の課題としては、インテリジェント・システムの廃止における倫理的考慮事項への取り組みが挙げられる。

結論

コンプライアンスの意図は信頼性である。インテリジェント・システムは急速に進化しているため、本稿では、規制遵守の範囲を超えて、より良いインテリジェント・システムを実現するためのリスクベースの重要なアプローチを提案する。インテリジェント・システムを総合的にアセスメントするためには、調査的思考、好奇心、継続的学習への情熱が必須条件である。

インテリジェント・システムは今や、医療、金融から交通、国家安全保障に至るまで、様々な分野に不可欠なものとなっている。このユビキタス性は、チェックボックスにチェックを入れるだけでなく、インテリジェントシステムを評価する意思と能力を持つ監査人の必要性を強調している。信頼性にはコンプライアンスだけでなくコミットメントが必要であり、これは評価者の姿勢と倫理標準に固有のものである。インテリジェントシステム技術の進歩がかつてないスピードであることを考慮すると、規制当局がまだ対処していないリスクを低減することは、先を見越したコンプライアンスと考えることができ、将来性のあるインテリジェントシステムを促進することができる。

重要な焦点のひとつは、監査プロセスの継続的改善であり、これは技術の進歩や新たな脅威と連動して進化しなければならない。しかし、この野心的な目標を達成することは、個人や事業体だけでできることではない。セクターを超えた利害関係者が協力し、リソースを出し合い、知識を共有し、普遍的に採用できるベストプラクティスを確立する必要がある。

最終的な目標は、インテリジェント・システムが安全性を損なうことなく人間の能力を高めるような、包括的で多様な環境を育成することである。

信頼できるインテリジェント・システムへの道は、信頼性は共有された責任であることを認識することにある。テクノロジーと監査がコンプライアンスを超えて到達できるようにするための協力は、現在および将来のインテリジェントシステムの信頼を高めるのに役立つ。

附属書

以下の設問は、特定されたすべてのリスクの概要を提供するために、企業内の様々なリスクマネジメ ントシステムを完全に統合することを考慮していない。相互に依存するリスクを管理、低減又は移転し、インシデントが発生した場合に全体的な対応が可能であることを保証するために、中央リスクマネジメント及びガバナンスのアプローチを維持することが強く推奨される。附属書の設問は、先に述べた各章に続くものであり、好奇心、リスク認識、調査的思考を喚起することを意図している。

附属書 1：監査の質問 "監査人を監査する"

AI システムの能力とプロセス：

- インテリジェント・システムはどのように監査候補者を選び、優先順位をつけるのか？
- インテリジェント・システムはどのようなデータを使って監査を行うのか？
- 特定の種類のリスクや異常を識別するために、識別システムはどのように訓練されるのか？
- インテリジェント・システムはどのようにデータを処理・分析し、潜在的な問題を検知するのか？
- 監査プロセスを通じて、どのような人的介入が必要なのか？
- インテリジェント・システムの結論の正確性と有効性はどのようにテストされ、妥当性が確認されるのか？
- インテリジェント・システムが監査プロセスにバイアスを持ち込まないようにするために、どのような保護措置が講じられているか？

監査プログラムの設計：

- 変化するリスクや技術に適応するために、監査プログラムをどのように定期的に見直し、更新しているか。
- 監査プログラムは、組織内でのインテリジェントシステムの使用に関連する潜在的なリスクと機会を明確に理解しているか？
- 監査プログラムチームは、組織内のインテリジェントシステムの効果的な使用と潜在的なリスクをアセスメントするために必要な知識とスキルを備えているか？
- 監査プログラムには、インテリジェントシステムまたは特定のインテリジェントシステムの使用を審査するために特別に設計された枠組みまたは方法論があるか？
- 監査プログラムは、関連する規制や標準への準拠を確保するために、組織内のインテリジェント・システムをどのように監視しているか？

監査チームとコラボレーション：

- 監査チームはどのように構成され、コンピューターサイエンス、哲学、心理学、法律学、業界の専門分野など、さまざまな分野でどのような専門知識を持っているのか。
- 監査人が、監査対象のインテリジェント・システムから運用上および組織上独立していることを保証するために、どのようなプロセスが設けられているか。
- 監査プロセスにおいて多様な視点や専門知識を結集するために、セクターを超えた協力関係はどのように行われているのか？
- 実用的なシナリオにおいて、インテリジェント・システムが環境や社会に与える影響を監査するために、どのような方法論、ツール、技術が使われているのか？
- 知識ベースと能力ベースのスキルを維持するために、監査員はどのように評価され、訓練され、指導されるのか。
- 監査チーム内の利益相反や個人的バイアスを防止するために、どのような防御が行われているか？
- インテリジェント・システムの監査実務の監視と保護には、どのような規制機関や専門機関が関わっているのか？
- 知的財産の秘密を守りながら、監査人が必要な証拠にアクセスできるようにするために、秘密保持契約はどのように活用されているか？
- 知的システムの信頼性はどのように判断され、どのような規準で信頼性のレベルが決定されるのか？
- また、監査方法が効果的かつ効率的であることを保証するために、どのような措置が講じられているか？

附属書 2：監査の質問 "AI ガバナンス"

ガバナンスと説明責任：

- 組織内に監査組織があるか？
- 社内の AI 活動の調整と監督を担当する最高人工知能責任者（CAIO）の役割は定義されているか？
- CAIO はどのチームと協働するのか。外部のコンサルタントが関与しているか、またその役割は何か。
- 外部監査人の窓口を一本化しているか。それは誰なのか？
- 明確な責任体制が確立されているか。
- AI ガバナンスが整備され、事業継続の側面を含む企業全体のリスクマネジメントに統合されているか。

- 経営陣は信頼できる AI の開発に十分なリソースを割り当てているか？
- 明確なオーナーシップ構造があり、インテリジェント・システムの行動、結果、意思決定に対する責任とオーナーシップを共有する線引きが定義されているか？
- 会社は、インテリジェント・システムを責任を持って使用するための枠組みを開発したか？
- 会社は、インテリジェント・システムの許容される倫理的な使用に関する方針を定めているか？
- AI の許容可能で倫理的な使用に関する方針はどのように文書化され、社内の適切な個人にコミュニケーションされているか？
- 会社は、インテリジェント・システムの許容可能かつ倫理的な使用に関するポリシーの遵守をどのように監視しているか？
- 全社的なインテリジェント・システムの使用を監視・追跡するプロセスがあるか？
- 会社は、サードパーティーのサービスプロバイダーによるインテリジェントシステムへのアクセスを監視・追跡する可能性を持っているか？

リスクアセスメント:

- AI システムの重要性がアセスメントされ、監査アプローチがシステムの重要性と組織の特殊性を評価するために調整されているか？
- システムの重要性を考慮したリスクアセスメントが実施されたか？
- リスクアセスメントは、システムの重要性に応じて、あらかじめ定められた間隔で計画されているか？
- 受容されたリスク」はあるか？上級管理職の署名はあるか？
- 移転されるリスクはあるか？対応する保険契約は適用されるか？
- インテリジェントシステムに関連するリスクは、企業や情報セキュリティのリスクマネジメントに統合されているか？
- 事業継続計画が策定され、承認されているか。また、その計画にはインテリジェントシステムが含まれているか。

指標と可視性：

- 透明性を維持するために、KPI や KRI といった測定可能な指標が適切に設定され、効果的に活用されているか。
- すべての AI コンポーネントを追跡するための AI 部品表（AIBOM）があるか？
- AI のすべてのデジタル部品を追跡するためのソフトウェア部品表（SBOM）があるか？
- AI のすべての物理的コンポーネントを追跡するための部品表（BOM）があるか？

- AI システム内の資産は容易に可視化され、物理的にもデジタル的にも追跡可能か？

説明可能な AI 原則（XAI）：

- XAI のために提案された原則（説明、意味のある説明、説明の正確さ、知識の限界）を明確に理解し、これらの原則が適切に適用されているか。
- インテリジェント・システムの意思決定プロセスを利害関係者に説明する仕組みがあるか？
- アウトプットの説明は、想定されるエンドユーザにとって有意義で理解しやすいものであるか？
- システムは、出力を生成する理由を正しく反映し、説明の正確性を保証しているか？
- システムは、設計された条件下でのみ、および／または、その出力が十分な信頼に達したとき（知識の限界）にのみ作動するか？
- インテリジェント・システムの意思決定が、特に高度なシステムにおいては、人間の理解を超えてしまうという潜在的な課題に、組織はどのように取り組んでいるのか？

公平性とバイアス：

- インテリジェント・システムがトレーニングの公平性を確保するために使用できる、複数の人口統計グループ、社会経済的背景、多様な視点を代表する多様なデータセットがあるだろうか？
- そのインテリジェント・システムは、意図された状況での使用に適しているか、バイアスが最小化または除去されているかを確認するために、厳密にテストされているか？
- さまざまな環境やシナリオを考慮した継続的なバイアステストが実施されているか。

AI レジリエンス：

- システムはレジリエンスに富み、インテリジェント・システムの重要性を考慮し、変更されたパラメーターや不慣れな環境下でテストされているか？
- システムが可塑性（自己修復、自己/連続学習、進化的アルゴリズム）のアプローチを取り入れている場合、その可塑性は透明性があるか？
- インテリジェント・システムが可塑性を持つ場合、それ自身を再利用したり、限界パラメーターを変更したりすることができるのか？本来の目的やユースケースからの逸脱を防ぐための設計とは何か？

プライバシーとセキュリティ：

- プライバシーとセキュリティに対する保護措置はどの程度整っているか？
- システムは、そのライフサイクルを通じて、プライバシー・バイ・デザイン及びセキュリティ・バイ・デザインのアプローチを採用しているか？
- コード/ライブラリのセキュリティ脆弱性を検証したか？
- 開発段階で安全な開発戦略が適用されたか？

監査プロセス：

- クリティカル・シンキングは、特にインテリジェント・システムの日進月歩の性質を考慮すると、監査プロセスにおいてどのように適用されるのだろうか？
- 過去の監査結果は入手可能か。識別された指摘事項は改善されているか。

附属書 3：監査の質問 "説明責任、責任、賠償責任"

説明責任を自覚する：

- 経営幹部やその他の要員は、自らの説明責任、責任、法的責任を認識しているか。
- 責任ある行動を奨励し、法的リスクを最小限に抑えるために、従業員に対して説明責任、責任、賠償責任に関する適切な研修が行われているか。
- チームメンバーに割り当てられた責任と義務を概説する適切な文書または証拠があり、それらがチームメンバーに明確に伝えられているか。
- 個人または組織の責任とその行動の結果に対する責任を追及するための法的枠組みに関する懸念はあるか？
- 上級マネジメントに対する潜在的な法的リスクを軽減するための仕組みがあるか。
- AI 保険における説明責任、責任、賠償責任に関連するリスクに対応する計画があり、その計画はコミュニケーションされているか。

レポーティングとリスクマネジメント：

- 説明責任に関して、個人やグループが自らの行動や決定について説明できるよう、適切な報告や正当化の手続きが行われているか。
- 変更を確実かつ透明性をもって追跡するための変更管理プロセスが確立されているか。
- 具体的な行動や決定に関連する成果を測定・追跡する仕組みがあり、個人またはグループが成否の責任を負っているか。
- 成功がリスク選好に影響するという事実を、会社はどのように扱っているのだろうか？
- 経営幹部の多様性はどの程度か？
- チームメンバー間で責任を分担することに関連する潜在的リスクはあるか、ある場合はそのリスクは低減されているか。
- AI の幻覚に対するリスクや、出力に依存する能力に関するガードレールの理解など、AI の責任ある使用について、会社は従業員やマネジメントをどのように教育しているのだろうか？

附属書 4：監査の質問 "法律、規制、標準"

法規制の遵守：

- インテリジェント・システムにはどのような法律や規制が適用されるのか？
- インテリジェント・システムは、これらの法規制に準拠して開発されているか？
- 開発チームは、法律、規制、標準の変更に対する継続的なコンプライアンスをどのように確保しているか？
- インテリジェント・システムの利用者は、その権利とアプリケーションが関連法、規制、標準に準拠していることについて知らされているか？
- 業界固有の標準やベストプラクティス（ISO/IEC 42001:2023、ISO/IEC 27701 など）が整備され、定期的に監査されているか。
- インテリジェントシステムにはどのようなデータ・プライバシー法が適用されるか？
- これらのプライバシー法の遵守は、どのように実施または組織化されているか？
- 現行の法律に従い、必要な期限内にデータ漏洩をコミュニケーションするためのプロセスを確立しているか。

プライバシーリスク・マネジメント：

- プライバシーリスク要因をアセスメントし、文書化し、コミュニケーションし、低減し、受け入れ、または移転（保険）しているか？
- データ所有者のデータ使用に関する同意の撤回を管理するための措置が講じられているか、またはデータ所有者にその旨が通知されているか。
- データ所有者はデータの使用方法を制限できるか？
- インテリジェント・システムのさらなる使用および／またはトレーニングのために個人データを使用することに明示的に同意した記録があるか。
- ユーザは、インテリジェント・システムにすでに提供されたデータについては、そのような合意を取り消すことができない可能性があることを認識しているのだろうか？
- データ所有者はデータをその重要度に応じてタグ付けすることができ、インテリジェント・システムは関連するリスクに応じて異なる重要度のデータを扱うことができるか？
- データ盗難はどのように防止・検知されるのか？
- データ漏洩はどのように処理されるのか？残されたリスクは受け入れるのか、それとも移転するのか？

- データは安全に廃棄されているか？

倫理と持続可能性のガイドラインと標準：

- 知的システムに適用される倫理指針や標準はあるのか？
- 倫理的指針の実施は完全に文書化されているか？
- AI に適用される持続可能性のガイドラインや標準はあるのか？
- 持続可能性の側面は、インテリジェント・システムにどのように統合されているのか？

ガバナンスと説明責任：

- 会社は、適用法、規制、標準、社内方針をすべてリストアップしているか？
- 社内で使用されているすべてのインテリジェント・システムの在庫を把握しているか？
- インテリジェント・システムのサードパーティ・サプライヤーの概要を把握しているか？
- データの使用、モデルのトレーニング、意思決定プロセスについて監査可能な証跡が記録され、アカウンタビリティが確保されているか。
- インテリジェントシステムは、関連する法律、規制、標準への準拠を検証するために、サードパーティによる監査や認証を受けているか？

公平性とバイアス：

- インテリジェントシステムは、保護されたグループに対する公平性、バイアス、識別的差別についてテストされたか？
- 調査結果はどのように文書化され、必要であれば対処されたか？

附属書 5：監査の質問 "AI ビジネスケース"

ビジネスケースとプロセス：

- そのインテリジェント・システムは、特定のビジネス・ケースに合致しているか？
- AI システムのビジネスケースはどのように定義され、そのパフォーマンス、レジリエンス、説明可能性に関して期待される指標は何か？
- インプットとアウトプットを含むビジネスプロセスが定義されているか？

指標と KPI：

- パフォーマンス、インテリジェントシステムのレジリエンス、説明可能性（必要な場合）について、測定可能な KPI が定義されているか。

- AI システムの "適合性 "に期待される指標は、その意図する用途や文脈とどのように整合しているのか？
- KPI は可視化され、文脈に置き換えられて、CAIO またはそれに相当する者に報告されているか。
- KPI が許容範囲外の場合、対策を講じるプロセスがあるか？

「バイ・デザイン」アプローチ：

- インテリジェント・システムの開発は、倫理原則、プライバシー、セキュリティを考慮した適切な "by design "アプローチに従って行われたか？

監査範囲の定義：

- システムの境界、主要な利害関係者、関連する規制、エンドユーザの安全性、製品品質、データの機密性、完全性、可用性（CIA）などの特定の懸念事項など、監査の範囲はどのようなものか。
- すべての関係者を特定し、監査に参加させたか。

リスクアセスメント:

- 企業のリスクマネジメントは AI システムのリスクアセスメントとどのように統合され、信頼できる AI を確保するためにどのようなプロセスがあるのか。
- 事業継続性など、AI システムの監査時に考慮すべき追加のリスクや懸念事項はあるか？

上級管理職の関与：

- 上級管理職は、展開されたインテリジェント・システムを十分に認識し、説明責任を果たしているか？

クリティカル・シンキングと仮定への挑戦：

- 一見独立した点と点を結びつけ、インテリジェント・システムの安全性を高めるために、クリティカル・シンキングはどのように応用されるのだろうか？
- 想定を覆し、潜在的リスクを低減するために、監査人はどのように既成概念にとらわれない思考を適用しているか。
- 監査人は、その AI システムが特定の環境において「使用に適しており」、長期的なセキュリティリスクをもたらさないことをどのように確認するのか。

人間の監督レベル：

- 組織は、AI システムに必要な人間の監視レベル（ループ内人間、ループ外人間、ループ内人間）を明確に定義しているか？
- 人的監視のレベルを選択した根拠を説明する文書があるか？

ビジネスケースに関する考察：

- 初期のビジネスケースには、リスクや危害を軽減するための適切な人的関与のレベルのアセスメントが含まれているか？
- ビジネスケースにおいて、人的関与のレベルはどれくらいの頻度で見直され、更新されるのか？

人間審査のオプション：

- 人間によるレビューのための「オプトアウト」および／または「オプトイン」オプションは、AI システム内に実装されているか？
- インテリジェント・システムが下した決定に対して、人間が異議を唱えることができる仕組みはあるのだろうか？

効果とモニタリング：

- 人的監督の有効性を監視し、必要に応じて調整するプロセスがあるか？
- 監督プロセスに携わる人間が自らの役割と責任を理解するためのトレーニング・プログラムはあるか？

リスクアセスメント:

- AI システムの意思決定プロセスへの人間の関与が不十分であることによる潜在的な弊害を判断するために、リスクアセスメントが実施されたか？
- 選ばれた人的監視のレベルが、システムの重要性及び利害関係者への潜在的影響に合致していることを、組織はどのようにして保証するのか？
- 利害関係者は、人間によるレビューをオプトインまたはオプトアウトする能力、およびインテリジェント・システムの決定に異議を唱える権利について、どのように知らされているか？

コンプライアンスとベストプラクティス：

- 人による監督やレビューのレベルを決定する際、組織は業界のベストプラクティスや規制要件に従っているか？
- 人による監督に関して、組織はどのようにして長期にわたってこれらの慣行や規制の遵守を保証しているか？

附属書 6：監査の質問 "AI インフラ"

インフラストラクチャー資格：

- AI に特化したワークロードを処理するために、既存のインフラにはどのような資質があるのだろうか？
- インフラの認定にはどのようなプロセスと規準が用いられたのか？

- インフラは、セキュリティ、プライバシー、拡張性、持続可能性について監査されているか？
- インフラが準拠する必要のある関連規制や標準があるか、また監査プロセスにおいて準拠が確認されているか。
- 監査人は、現行の規制の枠を超え、監査がインテリジェント・システムの長期的な信頼性に役立つことを保証しているか？
- 潜在的なインフラ障害に対処するために、どのような災害復旧計画があるのか。
- インフラに使われているデータセンターの（地政学的、物理的）セキュリティはどうなっているのか？

インフラのパフォーマンスとスケーラビリティ：

- 計算能力、データ処理速度、帯域幅などの要素を通じて、インテリジェント・システムのパフォーマンスや信頼性に影響を与える可能性のある、インフラストラクチャーにおける現在の制限や課題はあるか？
- 増大するデータ量や計算ニーズに対して、インフラストラクチャーはどの程度スケーラブルなのか？
- スケーラビリティを促進するために使用されるキャパシティプランニング戦略とテクノロジーについて説明する。

インフラにおけるデータの取り扱いとプライバシー：

- インテリジェント・システム・インフラは機密データをどのように扱うのか？
- 高度なプライバシーを確保するために、機密コンピューティングがハードウェアに統合されているか？
- データの処理中およびインフラの異なる部分間での転送中に、データを保護するためにどのような措置が講じられているか？

持続可能性と環境への影響：

- インテリジェント・システム・インフラは、エネルギー効率の高いハードウェアの使用など、エネルギー効率にどのように取り組んでいるのか？
- インフラの設計および運用において採用された、省エネルギーの特徴および慣行を概説する。
- 再生可能エネルギーを使用しているか？
- インフラストラクチャーはどのように環境の持続可能性に取り組んでいるのか？
- 電子廃棄物はどのように処理されるのか？

附属書 7：監査の質問 "センサー"

センサーの統合と相互運用性：

- 視覚、聴覚、触覚、温度、近接、化学物質検出器など、必要なセンサーはすべてインテリジェント・システムに適切に組み込まれているか？
- これらのセンサーは期待通りのデータを提供しているだろうか？
- センサーがコミュニケーションし、コンテキストデータを共有するためのプロトコルは何か？
- コンテキストを意識したセンサーシステムにおいて、センサー技術、データ分析、ネットワーキング・プロトコル、機械学習（TinyML を含む）は適切に活用されているか？

リスクマネジメントとコンプライアンス：

- センサーの能力を組み合わせることで、過小評価されているリスクはないか？
- そのようなセンサーの放置、誤用、意図的な濫用といった潜在的な問題を軽減するための計画があるか。
- インテリジェントシステムのための包括的な洞察を生成する上で、センサーデータの信頼性を評価するための適切な分析ツールは整備されているだろうか？
- センサーが取得したデータのプライバシー・セキュリティを確保するために、適切なデータ保護対策が講じられているか。
- インテリジェント・システムにおけるセンサーの使用は、すべての適用法に準拠しているか？

技術の最前線：

- 小型化、コンテキストアウェアネス、センサーフュージョンなど、インテリジェントシステム全体のリスクアセスメントに影響を与える可能性のある新技術はあるか？
- センサーを交換する場合、精度の改善、自動化の強化、消費電力の削減など、期待される利点は達成されているのだろうか？

附属書 8：監査の質問 "データ"

データソースと要件：

- インテリジェント・システムのビジネスケースに必要なデータ要件（量、質、機能）は誰が定義するのか？
- データの質（正確性、一貫性、完全性、信頼性、妥当性）を評価するために、どのような指標や方法が用いられているか。

- データが高品質であることを保証するために、どのような措置が取られているか？
- データの特性や質は、システムのパフォーマンスにどのような影響を与えるのか？
- 具体的なビジネスケースにはどのようなデータが使われたのか（例：ラベル付き、ラベルなし、セミラベル付き、構造化、非構造化、有機のみ、合成のみ、ハイブリッドアプローチ）。
- これらのデータ要件はどのように文書化され、コミュニケーションされているのか？
- 外部／購入データを使用する正当な理由は何か？
- インテリジェント・システムのために、データはどのように調達され、収集され、管理されるのか？
- 原産地や有機か合成かに関するデータが表示されているか。
- データにおける望ましくないバイアスのリスクはどのように軽減されるのか？

データ・プロベナンスとスクレイピング：

- 収集から使用までのトレーサビリティ（「闇市場データ」、「ダーティ・データ」）を確保するために、データの出所はどのように追跡されるのか？
- 闇市場データ」の意図しない利用に伴うリスクはどのように低減されるのか？
- ダーティ・データ」を識別し、回避するために、どのような手順があるのか？
- 使用されているすべてのデータセットの出所と出自を確認するために、どのような仕組みがあるのか。
- トレーニングデータの出所が追跡可能であり（「データペディグリー」）、信頼できる組織によって検証または認証されているか。
- スクレイピングされたソースは、トレーサビリティとデータの出所を保証するために文書化されているか。
- データスクレイピングはウェブサイトの利用規約（TOS）を尊重し、違反を防いでいるか？
- 特に生成的 AI の場合、学習用データセットの著作権規制への準拠はどのように確保されるのか？
- 生成的 AI システムが、著作権規制や特許などで保護されているオリジナルの（芸術）作品を複製しないことを保証する仕組みはあるか？
- 著作権侵害の可能性を特定し、対処するための社内プロセスはどのようなものか。
- 著作権で保護された作品に関する契約は文書化されているか。

データ管理とライフサイクル：

- データ・コレクション中にプライバシーを確保し、CIA の三要素（機密性、完全性、可用性）を遵守するために、どのような手順があるか？

- データのライフサイクル全体（作成／取得、使用、移転、保管、削除）を統制する、文書化されたポリシーと手続きは何か？
- インテリジェント・システムによってアクセス、編集、削除されたデータの監査証跡があり、そのような変更を認可したユーザまで遡ることができるか。
- インテリジェント・システムが使用するデータの変更や更新は、そのパフォーマンスにどのような影響を与えるのか？
- システムの運用期間中（およびそれ以降も）、意思決定プロセスの正確性と有効性を維持するために、どのような対策が講じられているか？

データ・プライバシーとセキュリティ：

- プライバシー保護のために、データ・デアイデンティフィケーション（匿名化／仮名化）はどのように実施されるのか？
- 差分プライバシー技術を適用する場合、その遵守はどのように確保されるのか？
- 訓練用データセットに提出する前に、個人データをぼかす／暗号化する措置がとられているか？
- データ保持はどのように規制され、達成されるのか？
- 静止時および転送時のデータセキュリティを確保するために、どのような手段を用いているか。
- 将来を見据えた暗号化のために、耐量子暗号ソリューションは検討されているのだろうか？

データ準備プロセス：

- 誰がデータの準備と分割（トレーニング、妥当性確認、テスト）に責任を持つのか？
- データ分割はどのように文書化され、正当化されるのか？
- 準備中のデータの整合性を確保するために、どのような手順がとられているか？

合成データ：

- 合成データの生成は、AI システムのモデルにおける妥当性をどのように確認するのか？
- 合成データの生成はどのように規制され、文書化されるのか？

データの品質保証：

- 誰がデータ品質保証とデータガバナンスポリシーの遵守に責任を持つのか？
- データ品質アセスメントから得られた知見はどのように文書化され、対処されているか？

データガバナンスの枠組み：

- きちんと定義されたデータガバナンスの枠組みがあるか？
- データの選択、操作、使用において倫理的標準が維持されるような方針と手順があるか。

モデル出力：

- モデルの出力データはどのように保存され、セキュリティは確保されているのか？

附属書 9：監査の質問 "データ処理ユニット"

データ処理ユニット：

- CPU、GPU、TPU、エッジプロセッサーなど、必要なデータ処理装置はすべて揃っているか？
- 各データ処理装置は、それが設計された特定のタスクのために効率的かつ効果的に活用されているか？
- 必要なデータ処理の規模に関して、既存のデータ処理装置に限界や課題はあるか？
- 異なるデータ処理ユニットを統合し、タスクが適切に分散されるようにワークロードを割り当てる適切なメカニズムがあるか？
- エッジプロセッサーは、ローカルにデータを処理することで効率を高め、IoT 環境における待ち時間を短縮するために活用されているのだろうか？

セキュリティとコンプライアンス：

- データ処理装置をサイバー攻撃から守るために、適切なセキュリティ対策がとられているか？
- データ処理ユニットに関連し、その技術仕様やシステムへの統合などの詳細を含む適切な文書があるか。
- さまざまなデータ処理ユニットとその関連テクノロジーの使用について、規制要件に準拠するようガバナンスする方針またはガイドラインがあるか。

新たなテクノロジー：

- 量子コンピュータのような、計算能力を向上させるために評価されている新技術はあるのか。もしあれば、既存のインフラとの統合の可能性を管理するために、どのような対策がとられているのか。
- また、もしそうなら、知的システムの能力を向上させるために生物学的コンピュータを使用する計画はあるのか？

附属書 10: 監査の質問 "データ保管"

データ・ストレージ・ソリューション：

- 内蔵ストレージ・オプション（SSD や HDD）、RAM、クラウド・ストレージ・ソリューションなど、必要なストレージ・ソリューションをすべて利用できるか。

- 内部ストレージのオプションはシステムに適切に統合されているか、また、その要件を満たすのに十分なストレージがあるか。
- RAM リソースは、即時の処理ニーズに対して適切なデータ検索を提供するように構成されているか？
- クラウド・ストレージ・ソリューションは正しく展開され、大規模なデータセットの保存と分析のためのスケーラブルでアクセス可能なリポジトリを提供しているか？
- 内部ストレージからクラウド・ストレージ・ソリューションへのデータ移行や統合について懸念はあるか？
- 潜在的なデータ移行の問題を軽減するための計画やプロセスがあるか？
- システムがクラッシュした場合でも、データの可用性と完全性を確保するための適切なデータバックアップとリカバリ機能があるか？

データセキュリティ、コンプライアンス、文書化：

- データセキュリティ上の懸念はないか。また、いずれのストレージ・ソリューションに保存されたデータも暗号化されているか、またはアクセス管理メカニズムによって適切に保護されているか。
- 各ストレージ・オプションの使用方法、ストレージ・オプションの適切な統合方法、データのバックアップとリカバリ方法について、適切な文書があるか。
- GDPR などの規制に配慮し、データの保管・管理におけるコンプライアンスを確保し、必要な措置を講じているか。

新たなストレージ技術：

- ホログラフィック・データ・ストレージや DNA ベースのストレージのような先進的なストレージ方式をインテリジェント・システムに統合する計画はあるのか？

附属書 11：監査の質問 "ネットワーキング、コネクティビティ、コミュニケーション・インターフェイス"

コミュニケーション・インターフェイスとプロトコルの統合と設定：

- シームレスなデータ交換を保証するために、必要なコミュニケーション・インターフェイスがインテリジェント・システムに適切に統合されているか？
- 使用シナリオやデバイスの互換性に基づいて、最も適切なコミュニケーション・インターフェイスを選択するための適切なガイドラインや手順があるか。

- WLAN、イーサネット、セルラー（4G/5G）などの従来のネットワーク方式は、最適なネットワーク・パフォーマンスを維持するために適切に設定されているか？
- ブルートゥース、BLE、RFID などの近距離通信プロトコルは、無線の脆弱性やデータ漏洩に対して適切に保護されているか？
- LoRaWAN、Zigbee、Z-Wave、NB-IoT、Sigfox などの IoT 専用プロトコルは、低消費電力、長距離、短距離のコミュニケーションに使用されているか、また、これらのプロトコルで伝送されるデータを保護するために十分なセキュリティが確保されているか。
- OPC-UA、DDS、CoAP などの産業用およびリアルタイム通信プロトコルは、マシン間通信や高性能データ交換のニーズを満たすために活用されているか、またデータは安全か。

相互運用性とサイバーセキュリティ：

- これらのインテリジェント・システムで使用される異なるコミュニケーション・プロトコル間の相互運用性に懸念はないか、また潜在的な問題を軽減するための計画はあるか？
- 無線の脆弱性から、制約のあるネットワーク上で伝送されるデータの安全性まで、さまざまな通信プロトコルに関連するサイバーセキュリティの課題を管理するための計画があるか？

将来の接続計画：

- 接続性を改善し、遅延を減らすために 6G ネットワークにアップグレードする計画はあるか。
- 量子コミュニケーションネットワークなど、新たなコネクティビティ技術を導入する計画はあるのか？

附属書 12：監査の質問 "フォグ・コンピューティングとクラウド・コンピューティング"

フォグ・コンピューティング・インフラストラクチャ：

- 待ち時間を最小化し、データ伝送プロセスの効率を最大化するために、フォグ・コンピューティング・ノードはどこに戦略的に配置されるのか？
- フォグ・コンピューティング・ノードは、どの程度の処理能力とストレージ能力を持つのだろうか？ローカルのデータ分析、フィルタリング、前処理を効果的に実行するのに十分だろうか？
- フォグ・コンピューティング・ノードのどこにインテリジェント・ゲートウェイを配置すれば、クラウド全体の負荷を最適化し、リアルタイムのアプリケーション・パフォーマンスを確保できるのか？
- フォグ・コンピューティング・システムは、クラウド・コンピューティング・システムとシームレスに統合され、2 つのシステム間でデータがスムーズに流れるように設計されているか？

- 異なるフォグ・コンピューティング・システムの統合に影響を及ぼす可能性のある、異なるフォグ・コンピューティング実装間の相互運用性に関する問題は現在存在するか？
- フォグレベルで予測分析やエッジ AI アプリケーションのためのインテリジェントシステムを実装するのに十分なリソースがあり、そのようなアプリケーションは意思決定能力を向上させると期待されているのだろうか？
- フォグ・コンピューティング・ソリューションのスケーラビリティ、特に自律走行車やスマートシティ、産業用 IoT などの新興技術をサポートする際に懸念されることはあるか？

セキュリティとコンプライアンス：

- フォグ・コンピューティング・ネットワークで伝送されるデータのセキュリティを確保するために、現在どのようなサイバーセキュリティ対策がとられているのか？
- フォグ・コンピューティング・ソリューションが、データの取り込み、保管、送信に関して必要な規制要件を満たしていることを確認するためのポリシーやガイドラインがあるか。
- フォグ・コンピューティング・ネットワークの運用に影響を及ぼす可能性のあるインターネットの途絶に備えて、バックアップ・システムや計画があるか？

附属書 13：監査の質問 "ソフトウェア・コンポーネント"

インストールと設定：

- 必要なソフトウェア・コンポーネント（仮想化ソフトウェア、オペレーティング・システム、ミドルウェア、アプリケーション・ソフトウェア、データ管理ツール、モニタリング、ロギング・ツール、コンテナ・オーケストレーション・プラットフォーム）がすべて適切にインストールされ、統合されていることを確認する責任は誰にあるのか。
- 仮想マシンはどのように適切に設定され、要求に応じて必要なリソースが割り当てられているか？
- ミドルウェアは、異なるアプリケーションやシステム間でシームレスなコミュニケーションを提供できるか？
- コンテナ・オーケストレーション・プラットフォームは、分散環境で AI システムを効率的に管理し、スケーリングしているか？
- スケーラビリティ、信頼性、効率性のためにソフトウェア・コンポーネントを使用する際の制限や課題はあるか？

保守と文書化：

- ソフトウェア・コンポーネントの更新やセキュリティ脆弱性のパッチは、どれくらいの頻度で行われているか？

- ソフトウェア・コンポーネントのインストール、管理、保守手順に関する適切な文書はどこにあるか？
- 監視・記録ツールは、システムのパフォーマンスとセキュリ ティを適切に追跡するために必要な機能と特徴を備えて いるか。

附属書 14：監査の質問 "アルゴリズム、トレーニング、モデル"

アルゴリズム：

- 問題分析（例：学習タイプ、データの利用可能性）に基づくアルゴリズムの選択はどのように正当化されるのか？
- 選ばれたアルゴリズムに潜在的なバイアスはあるか？
- もし説明可能な AI（XAI）技術が関連するのであれば、モデルの推論を理解するために実装されているか？

トレーニングデータセット：

- トレーニングデータの質と代表者はどのように確保されるのか？
- データセット内のバイアスや公平性の懸念を軽減するために、どのような対策が講じられているか？
- プライバシー侵害（個人を特定できる情報など）を含むデータを特定し、対処する手順があるか？
- 効果的なトレーニングのために、ラベル付きデータ、ラベルなしデータ、合成データのバランスはどのように管理されているのだろうか？

事前学習モデル：

- 事前学習されたモデルが使用される場合、その出所や潜在的なバイアスはどのように文書化され、理解されるのか？
- 事前学習済みモデルに関連するライセンス制限や著作権への配慮はあるか？

モデルトレーニングの枠組み：

- 選択した枠組み（TensorFlow、PyTorch など）は安全で最新か？
- モデルトレーニングのために、バージョン制御と再現性をどのように実践しているか？

評価指標：

- 基本的な精度だけでなく、多様な評価指標（精度、正確さ、再現性など）を用いて、モデルのパフォーマンスをどのように評価するのか。
- 選択した測定基準の限界を理解し、結果を解釈する際に考慮しているか。

ハイパーパラメーターを調整する：

- オーバーフィッティングを避けるために、ハイパーパラメータのチューニングのプロセスはどのように文書化され、妥当性が確認されているのか？
- 人間のバイアスがハイパーパラメータの選択に影響するのを防ぐ手順はあるか？

モデル推論エンジン：

- モデル推論エンジンを敵対的な攻撃や操作から守るために、どのようなセキュリティ対策がとられているのか？
- 推論エンジンの効率と性能はどのように監視され、実際の展開に最適化されているのか？

モデルの展開と管理：

- 継続的なモニタリングとパフォーマンス評価を確実にするために、本番環境ではどのようにモデルを展開するのか？
- モデルのドリフトやパフォーマンス低下に備えたロールバック戦略は確立されているか？

セキュリティとドリフト検知：

- モデルやシステムのセキュリティ脆弱性はどのようにアセスメントされ、対処されるのか。
- 敵対的攻撃（ポイズニング、回避など）を検知し、低減するプロセスはあるか？
- 長期にわたってモデルの有効性を維持するために、データドリフト（コンセプトドリフト）はどのように監視され、低減されているのか？

倫理的配慮：

- 公平性、バイアス、説明可能性への配慮は、インテリジェントシステムのライフサイクルを通してどのように統合されているのだろうか？
- モデルの出力に関連する潜在的な倫理的リスクを特定し、低減するための手順があるか。
- もしそのモデルが利害の大きい領域（例：医療、金融）で使用される場合、倫理的、規制的な考慮事項が追加されるか？

透明性と説明可能性：

- モデルの限界と潜在的な故障モードを明確に理解しているか？
- 特に重要なシナリオにおいて、モデルの決定を人間のユーザに説明する仕組みはあるか？
- インテリジェント・システムの開発、展開、運用を通じて、人間の監視はどのように維持されるのか？

附属書 15：監査の質問 "微調整と妥当性確認"

ハイパーパラメーターを調整する：

- モデルをトレーニングする前に、ハイパーパラメータは適切に調整されたか？
- 選ばれた価値観の根拠は何だったのか？
- 検索や最適化の戦略は用いられたか？

モデルを微調整する：

- 微調整によってモデルの精度を改善しようとしたのだろうか？
- 微調整にはどのようなアプローチが用いられたのか？
- 微調整のためにデータソースを追加したり、別の方法を検討したりしたか？
- 検索や最適化の戦略は用いられたか？

妥当性確認方法：

- 選択された妥当性確認方法と、妥当性確認用データセットはどのように選択されたのか？
- モデルの性能はどのように評価され、どのような評価指標が用いられたのか？

オーバーフィッティングのチェック：

- オーバーフィッティングのテストは行われたか、行われた場合はどのように行われたか。
- 別のテストデータセットを使用したのか、それともこの目的のために妥当性確認データセットを使用したのか。

パフォーマンスとロバスト性：

- モデル出力の正確性と信頼性はどのように検証されたのか？
- モデルの性能と潜在的なバイアスを継続的に監視するために、どのような手順を導入しているか。

汎化と頑健性：

- モデルの頑健性と汎化能力をチェックするために何をしたのか？
- モデルの頑健性をテストするために、敵対的な攻撃やその他の摂動は用いられたか？
- AI のセキュリティを評価するためにペンテストを実施したか？
- バグバウンティは脆弱性を継続的にチェックするように設定されているか？

解釈可能性と透明性：

- モデルの解釈可能性と透明性を確保するために、どのような措置が取られたのか？

- バイアスを防ぐための低減技術は適用されたか？

パフォーマンス指標と規準：

- AI システムのモデルのパフォーマンスは、要求に対してどのように測定されたのか？
- モデル展開の受け入れ規準は何だったのか？
- 誰が、どのような根拠に基づいて、このモデルの展開を最終決定したのか？

仮定と限界：

- AI システムのモデルの前提条件と限界は何か、またそれらはどのように文書化されているか？

「妥当性確認：

- AI システムのモデルは、現実的なシナリオで、想定されるさまざまな環境条件のもとで、その意図する機能について妥当性確認がなされているか？
- AI システムのモデルは、データの性質とモデルの意図された用途に基づき、期待に沿った結果を生み出しているか？

パフォーマンスと信頼性：

- AI モデルの性能は、精度、スピード、信頼性、必要なリソースの点で許容できるか？
- AI システムのモデルの挙動は、エッジケース、異常値、予期せぬ入力を含め、公称条件下および悪条件下で適切に特性化されているか？

バイアスと意図せざる効果：

- AI システムのモデルの出力は、意味的、文脈的、あるいは人間が解釈可能な観点から意味があるのだろうか？
- AI システムのモデルを実際に使用した結果、バイアスや意図しない影響が生じることはないか？

第三者による評価：

- AI モデルの性能と妥当性確認について、独立したサードパーティによる評価が行われたか？
- サードパーティによる評価では、どのような規準が用いられたのか？

附属書 16：監査の質問 "アクチュエーター"

アクチュエーターの選択の正当性：

- 具体的にどのような規準でアクチュエーター（モーター、サーボ、グリッパーなど）を選んだのか？
- 選択したアクチュエータは、ロボットが意図する力、速度、精度、可動域の要件とどのように整合するか？

- 代替のアクチュエータは検討され、評価されたか？

ドキュメンテーションとテスト：

- 使用される各アクチュエータの仕様、能力、制限を詳述した包括的な文書があるか。
- 想定される運転条件下でアクチュエータの信頼性と性能を保証するために、どのような展開前試験手順を実施したか？

アクチュエータの性能と信頼性：

- 動作中のアクチュエーターの性能（力出力、速度、精度など）を測定するために、どのような指標が使われているか？
- これらのパフォーマンス指標はどれくらいの頻度で評価され、許容できるパフォーマンスの閾値はどの程度か？

耐久性と保守：

- さまざまなアクチュエーターの期待寿命と、耐久性はどのようにアセスメントされるのか？
- アクチュエータの継続的な性能を確保し、故障を防止するために、どのような保守スケジュールと手順があるか？

故障検知と管理：

- アクチュエータの故障はどのように検知、報告、管理されるのか？
- アクチュエータが故障した場合、確実に運転を継続するために、どのような冗長対策がとられているか？

安全性とリスクマネジメント：

- アクチュエータによる意図しない動きや動作を防ぐために、どのような安全機構があるか？
- アクチュエータの誤作動や誤用による潜在的リスクはどのように低減されるのか？

緊急時の手順

- アクチュエータ制御システムには、どのような緊急停止プロトコルと安全チェックが組み込まれているか？
- これらのプロトコルはどのようにテストされ、妥当性が確認されているのか？

倫理的配慮：

- アクチュエーターの設計や展開において、倫理的な配慮、特に人間に似た動きや相互作用に関する配慮はどのようになされているのだろうか？

- 危害や倫理的問題を引き起こす可能性のある方法でのアクチュエーターの誤用を防ぐために、どのような対策が取られているか？

革新的なアクチュエーター

- 先進的あるいは実験的なアクチュエーター（人工筋肉、マイクロ流体技術など）は使われているか？
- これらの新技術にはどのような潜在的利益とリスクがあり、それらはどのようにマネジメントされているのだろうか？

適応性と拡張性：

- アクチュエータは、将来の進歩やロボットの動作要件の変更にどの程度対応できるか？
- 大幅な再設計なしに、アクチュエーター・システムをアップグレードまたは拡張できる規定があるか？

コンプライアンスとその先へ：

- アクチュエータは、関連する業界標準や規制にどのように準拠しているか？
- 継続的なコンプライアンスを確保するために、どのような内部・外部監査が実施されているか。
- AI の信頼態勢についてサードパーティによる評価が行われたか？その結果はどうだったか？
- 標準に準拠するだけでなく、アクチュエータの安全性、信頼性、効率を高めるために、どのような革新的な手法や技術が採用されているのか。
- アクチュエーター技術における新たなトレンドや将来的な規制変更の可能性について、組織はどのように先手を打っているのだろうか？

環境と操業条件：

- 様々な環境条件（例えば、温度、湿度、水、ほこり、紫外線エクスポージャー、限られた視界、極端な騒音や振動）での動作について、アクチュエータはどのようにテストされ、妥当性が確認されているか？
- 環境による損傷や劣化からアクチュエータを保護するために、どのような対策がとられているか？
- このシステムは、環境に配慮した方法でどのように廃棄されるのか？

運用の多様性：

- アクチュエータは、さまざまな作業タスクに対応したり、新しいタスクに適応したりする上で、どの程度汎用性があるか？
- 特定の用途に対してアクチュエータの性能を最適化するために、どのようなトレーニングや校正プロセスが行われているか？

- 特定のアクチュエーターなしで事業継続は可能か？生産ボトルネックや材料不足の場合の事業継続計画はあるか？

附属書 17：監査の質問 "電源"

標準に準拠すること：

- 電源装置は、関連する業界標準や規制にどのように準拠しているか？
- 継続的なコンプライアンスを確保するために、どのような内部・外部監査が実施されているか。

コンプライアンスを超える：

- 電源の安全性、信頼性、効率を高めるために、標準に準拠するだけでなく、どのような革新的な手法や技術が採用されているか。
- 電源技術における新たなトレンドや、将来起こりうる規制の変更に対して、組織はどのように先手を打っているのか。

電源の選択を正当化する：

- インテリジェント・システムの電源オプション（バッテリー、AC 電源、DC 電源、ハイブリッド・システムなど）は、どのような規準で選択されたのか？
- これらの電源オプションは、AI システムの運用要件や制約条件（モビリティ、ランタイム、エネルギー効率など）とどのように整合しているのか。

ドキュメンテーションとテスト：

- 使用される各電源装置の仕様、能力、制限について詳述した包括的な文書があるか。
- 想定される運転条件下で電源の信頼性と性能を確保するために、どのような展開前試験手順が実施されたか。

パフォーマンス：

- 動作中の電源装置の性能と効率（電池寿命、消費電力量、充電時間など）を測定するために、どのような指標を使用す るか。
- これらのパフォーマンス指標はどれくらいの頻度で評価され、許容できるパフォーマンスの閾値はどの程度か？
- 電源装置の性能はどのように継続的にモニターされているのか、またそのためにどのようなツールが使われているのか。
- 電源装置の性能指標が潜在的な問題を示している場合、どのような措置が取られるのか。

耐久性と保守：

- 各種電源装置の期待寿命はどの程度か、また耐久性はどのようにアセスメントされるのか。
- 電源装置の継続的な性能を確保し、故障を防止するために、どのような保守スケジュールと手順が定められているか。

故障検知と管理：

- 電源障害はどのように検知、報告、管理されるのか？
- 電源障害が発生した場合、連続運転を保証するためにどのような冗長性対策が実施されているか？

電力管理システム：

- インテリジェント・システム全体で効率的な電力消費を確保するために、電力管理システムはどのように最適化されているのか？
- 業務上のニーズとエネルギー利用可能性に基づいて電力使用量を動的に管理するために、どのような手段を講じているか。

安全メカニズム：

- 過熱、過充電、電気サージなど、電源装置に関連する危険を防止するために、どのような安全機構が設けられているか。
- 電源装置の故障や誤用による潜在的リスクはどのように低減されるのか。

緊急時の手順

- 電源管理システムには、どのような緊急シャットダウンプロトコルと安全チェックが組み込まれているか？
- これらのプロトコルはどのようにテストされ、妥当性が確認されているのか？

ワイヤレス充電：

- ワイヤレス充電技術は、インテリジェント・システム、特にロボットのためにどのように統合され、テストされているのだろうか？
- ワイヤレス充電中に緊急事態や故障が発生した場合、システムがシャットオフできるよう、どのような対策がとられているのか？

環境への配慮：

- 電源の選択が環境に与える影響、特にバッテリーの廃棄やデータセンターのエネルギー消費については、どのようにアセスメントされているのだろうか。
- カーボンフットプリントを最小限に抑え、インテリジェント・システム電源の持続可能性を高めるために、どのような措置が取られているか？

再生可能エネルギーの統合：

- インテリジェント・システム、特にデータセンターでは、どの程度まで再生可能エネルギーが使われているのだろうか？
- 再生可能エネルギーの使用を増やし、非再生可能エネルギー源への依存を減らすために、どのような戦略があるか？

革新的なパワー・ソリューション：

- 先進的あるいは実験的な電源技術は使われているか？
- これらの新技術にはどのような潜在的なメリットとリスクがあり、それらはどのようにマネジメントされているのだろうか？

適応性と拡張性：

- 電源装置は、AI システムの将来の進歩や運用要件の変更にどの程度適応できるか？
- 大幅な再設計なしに電源システムをアップグレードまたは拡張するための規定があるか？

附属書 18：監査の質問 "ユーザ・インターフェース"

ディスプレイの品質と耐久性：

- 信頼できるユーザとの対話を保証するために、ディスプレイ／スクリーンの品質と耐久性はどのようにアセスメントされるのか？
- さまざまな環境条件下で、スクリーンの映り込みや歪み、視認性の問題を防ぐための対策が講じられているか。
- 物理的あるいはデジタル的に特殊なフィルターをかけることで、プライバシーを強化する可能性はあるのか？

タッチスクリーン機能：

- タッチスクリーンは、さまざまなタッチジェスチャー（タップ、スワイプなど）に対する反応性と正確性をどのようにテストしているのか？
- 長期間にわたって最適な性能を維持するために、タッチスクリーンを校正するプロトコルはあるのか？
- タッチスクリーンの感度は、温度変化や手袋をはめた状態での操作に適応できるか？

音声認識の精度：

- 正確さと信頼性は、アクセントやバックグラウンドノイズを考慮して、どのように測定されるのか？
- ユーザからのフィードバックに基づいて音声認識アルゴリズムを更新する仕組みはあるのか？

ノイズリダクションとエンハンスメント：

- 騒がしい環境での認識を向上させるために、ノイズキャンセリングは可能か？
- 明瞭度を向上させる音声強調機能はあるか？
- 特定の周波数をブロックして、不要なノイズを減らすことは可能か？

環境要因：

- 携帯電話がポケットに入っているなどの振動は、認識にどのような影響を与えるのか？
- システムは他の声や音からの干渉にどの程度対応できるか？

コントロールパネルの使いやすさ：

- ボタン、ノブ、レバーとユーザとのインタラクションを最適化するために、物理的なコントロールパネルはどのように設計されているか？すべてのユーザが使いやすく、アクセスしやすいように、人間工学の原則が考慮されているか。
- 極端な温度、低視認性、高ストレス、高加速度、高重力のような様々な条件下でパネルを使用する可能性を確保するために、安全性とセキュリティの原則が考慮されているか？

モバイルアプリのセキュリティ

- モバイル・アプリケーションを通じてユーザデータを保護し、不正アクセスを防止するために、どのような防御策がとられているか？
- 異なるデバイスやオペレーティング・システム間での互換性について、モバイル・アプリケーションはどのようにテストされているのか？
- エンドユーザに依存しない継続的な無線セキュ リティパッチが有効になっているか。
- サイバーセキュリティの脅威を考慮すれば、無線ファームウェアのアップデートは可能なのか？
- アプリが誤動作したり、反応しない場合に備えて、緊急シャットダウンは有効になっているか？
- アプリが正しく機能しているかをチェックする自己診断機能を備えているか？誤動作が検知された場合に対処できるようになっているか？
- アプリに自己修復能力はあるか？どのように制御されているのか？

ジェスチャー認識の精度：

- ジェスチャー認識システムは、手や団体の動きをどの程度正確に信頼できるのか？
- パーソナライズされたジェスチャーの認識を向上させるために、システムを訓練するための継続的学習は可能か？
- 識別可能なジェスチャーや動作パターンを保護するためのプライバシー対策がとられているか？

- さまざまな照明条件やユーザのジェスチャーに対して、認識能力を最適化するための調整が行われているか。
- 特定のアクションを誘発するジェスチャーを検証するための安全プロトコルが実装されているか？

AR/VR の統合と新技術：

- AR/VR 技術はどの程度シームレスにユーザ・インターフェースに統合され、どのような対策が没入感のあるユーザ体験を保証するのだろうか？
- VR 環境において、乗り物酔いを最小限に抑え、視覚的忠実度を最適化するためのプロトコルは存在するのか？
- AR/VR システムの安全な運用を保証するために、どのような環境セキュリティ対策が実施されているのか？
- ユーザが AR/VR 体験を精神病に似た形で現実と混同しないよう、どのようにアセスメントされるのだろうか？
- AR/VR 環境における意図しない自傷行為を防ぐために、どのような対策がとられているのか？
- AR/VR 環境において、他者への意図しない危害はどのように防止されるのか？
- AR/VR シミュレーションと現実を区別するために、どのような技術や方法論が導入されているのか？

セキュリティ：

- 不正な改ざんや破損を防ぐため、BCI システムへの物理的なアクセスはどのように管理されているか？
- 神経データの収集、送信、保管の際、データ・プライバシーはどのように確保されるのか。
- 不正アクセスやデータ漏洩を防ぐために、物理的およびサイバーセキュリティ的なアクセスはどのように管理されているのか？
- 触覚フィードバックシステムによって収集されたユーザデータの安全な保存、送信、匿名化のためのプロトコルは整備されているか？

故障と安全性：

- BCI のハードウェアまたはソフトウェアの不具合を検知し、対処するために、どのようなプロトコルが用意されているか？
- 神経信号によって引き起こされる意図しない行動や反応を防ぐためのセーフガードはあるのか？
- 潜在的な BCI のリスクや安全上の注意について、ユーザはどのように知らされているのだろうか？

説明責任と責任：

- BCI で制御された機器によって行われた行為の責任は誰が負うのか？
- BCI が誤用されたり故障したりした場合の説明責任を決定するプロトコルはあるか？
- ユーザの同意と潜在的リスクへの理解はどのように文書化されているか？

保険と賠償責任：

- BCI ユーザや利害関係者に対する保険や賠償責任の保護はあるか？
- BCI 関連の事故や故障が発生した場合、潜在的な責任はどのようにアセスメントされ、低減されるのか。

倫理的配慮：

- BCI の開発と展開（インフォームド・コンセント、自律性、公平性）を規定する倫理指針は何か？
- BCI アルゴリズムやデータ収集における潜在的なバイアスはどのように対処されるのか？
- ユーザ・インターフェース設計文書と標準は、アクセシビリティとインクルーシビリティの要件を適切に組み込んでいるか？
- ユーザビリティ・テストは、障害のあるユーザを対象に実施され、ユーザ・インターフェースがアクセシブルで使いやすいものであることを確認したか。
- ユーザ・インターフェースのコードは、支援技術との互換性やアクセシビリティ・ガイドラインの遵守を確認するためにレビューされたか？
- ユーザ・インターフェースが多様な言語や文化に対応できるよう、言語的・文化的レビューが行われたか。
- 文化的バイアスやステレオタイプがないことを確認するために、ユーザ・インターフェース・デザインがレビューされたか？
- 触覚フィードバックによる不快感、怪我、感覚の過負荷を防ぐには、どのような対策があるのだろうか？
- インフォームド・コンセント、ユーザの自律性、公平性、包括性といった倫理的配慮は、触覚フィードバック・アプリケーション、特にセンシティブな領域において、どのように対処されているのだろうか？

パフォーマンス、ユーザ・エクスペリエンス、フィードバック・ループ：

- 快適さ、使いやすさ、効果的な BCI インタラクションのために、ユーザ・エクスペリエンスはどのように最適化されているのだろうか？
- BCI を個々のユーザのニーズに合わせるための規定があるか（カスタマイズ、ユーザからのフィードバック）。

- 継続的な UI/UX 改善サイクルに、ユーザからのフィードバックをどのように組み込んでいるか。
- ハプティック・フィードバックの正確性、一貫性、キャリブレーションは、ユーザや環境間でどのように確保されているのだろうか？
- ユーザの嗜好や感覚的なニーズに基づいて触覚パラメータを調整するメカニズムによって、直感的なインタラクションを実現するためにユーザ・エクスペリエンスはどのように最適化されているのだろうか？
- ユーザ満足度やユーザビリティの指標をリアルタイムでモニタリングし、反復的な設計の更新に役立てる仕組みがあるか。
- ハプティック・フィードバックの正確性、一貫性、キャリブレーションは、ユーザや環境間でどのように確保されているのだろうか？

規制遵守とイノベーション：

- BCI は医療機器や家電製品の標準規格にどのように準拠しているのか？
- 進化する規制やベストプラクティスへの準拠を監視し、更新するための継続的な取り組みがあるか。
- 進化する規制を監視し、コンプライアンスを維持するために、どのようなプロセスを導入しているか。
- ユーザの安全性と信頼性を優先しながら、触覚フィードバック技術において、どのようにユーザからのフィードバックを集め、技術革新を促進しているのか。

リスクアセスメントと低減

- UI/UX テクノロジーに関連する潜在的リスクは、どのように特定され、プロアクティブに低減されるのか？
- 予期せぬ課題や脆弱性に備えるため、シナリオに基づくリスクアセスメントを実施する枠組みがあるか？

研究開発（R&D）と将来の技術統合：

- BCI の性能を改善し、新たな課題に取り組むために、どのような研究イニシアチブがあるのか？
- 責任ある BCI の開発と展開を確保しながら、組織はどのようにイノベーションを促進しているのか？
- 高度なインテリジェント・システム・インターフェースや量子コンピューティング主導のインタラクションなど、今後登場する UI/UX 技術を統合するために、システムはどのように準備されているのか？
- UI/UX 技術の進歩における新たなリスクと機会を監視するための戦略があるか？

- 触覚フィードバックシステムは他のプラットフォームとどの程度シームレスに統合できるのか？多様なデバイスやアプリケーションとの互換性を確保するための相互運用性標準はあるか？

附属書 19：監査の質問 "コントロール・インターフェイス"

単一障害点：

- 集中制御装置では、単一障害点のリスクはどのように軽減されるのか？
- コントロール・ユニットの故障に対応するためのバックアップ・システムや冗長性対策はあるか？

パフォーマンス・モニタリング：

- 集中コントロール・ユニットの性能はどのようにモニターされ、維持されているのか？
- リアルタイム診断や故障予測のプロトコルはあるのか？
- 重要なシステムの最低限必要な機能に対して、フェイルオーバー手順や冗長性が実装されているか？

セキュリティ：

- 集中管理ユニットをサイバー攻撃から守るために、どのようなセキュリティ対策が実施されているか？
- 集中管理ユニットへのアクセスはどのように管理され、記録されるのか？
- 対策を講じることができる自己診断機能が実装されているか（緊急停止、必要な性能の維持、自己修復機能など）。
- 緊急時にコントロール・ユニットをシャットダウンできるか？

意思決定の分配：

- また、どのようなプロトコルが一貫性と協調性を保証するのか？
- 分権化されたユニット間の対立や矛盾を処理する仕組みはあるか？

頑健性と適応性：

- 変化する環境に対するシステムの堅牢性と適応性はどのように評価されるのか？
- リアルタイムの状況に基づいて、分散ユニット間でタスクを動的に再配分する手順はあるか？
- 異なる計算アプローチ（例えば、異なるプロセッサを使用した従来のコンピューティング、量子コンピューティング、生物学的コンピューティング）を活用した特定の分散型能力に応じて、ワークロードを動的に割り当てることができるか？

コミュニケーションと同期：

- 分散ユニット間のコミュニケーションはどのように管理され、保護されているのか？

- システムの一貫した動作を保証する同期プロトコルは整備されているか？
- 分権化されたユニットは、文脈を認識させることに成功しているのだろうか？これはメリットかリスクか？

目標設定と実行：

- 中央ユニットが設定したハイレベルの目標は、下位レベルのユニットによる具体的な行動にどのように反映されるのか？
- ハイレベルの目標は、独立したユニットによってチェックされているか（4 つの目の原則）。
- 下位レベルの活動の状況や実績を中央ユニットに知らせるためのフィードバック・ループがあるか？

自律性と協調性：

- ハイレベル・ユニットの自律性はどのようにコントロールされているのか？
- 階層的制御システムにおいて、下位レベルのユニットにはどの程度の自律性があるのだろうか？
- 階層間の調整を確実にし、対立を防ぐ仕組みがあるか？

柔軟性と拡張性：

- さまざまなアプリケーションや環境に対して、階層型制御システムはどの程度柔軟でスケーラブルなのか？
- スケーラビリティはどのようにコントロールされるのか？
- システムやユーザの要求に基づいて階層を再構成するプロトコルはあるか？

行動の定義と活性化：

- システムにおいて、行動はどのように定義され、優先されるのか？
- 行動はどのようにコントロールされているのか？
- センサーの入力に基づいて、どのような規準で特定の動作を作動させるのか？
- どのような規準で特定の行動を無効化するのか？

効率性と複雑性：

- 特に複雑なタスクの場合、ビヘイビアベースの制御システムの効率はどのように評価されるのか？事前に定義されたビヘイビアの実行を最適化し、パフォーマンスを向上させる方策はあるか？
- 実行機能の最適化は、バイアスの増大、公平性の低下、文脈の軽視とどのように関係しているのだろうか？
- 全体的な機能を失うことなく、持続可能性のためにエネルギー消費を最適化することは可能だろうか？

適応力がある：

- ビヘイビアベースの制御システムは、新しい状況や予期せぬ状況に対してどの程度適応可能なのだろうか？
- 経験に基づいて新しい行動を学び、取り入れる仕組みがあるか？

アーキテクチャの統合：

- ハイブリッド・コントロール・システムの中で、異なるコントロール・アーキテクチャはどのように統合されるのか？
- 各アプローチの長所を効果的に管理し、活用するためのプロトコルはあるか？
- 中央集権的なプランニングと分権的な反応性のバランスはどのように保たれているのだろうか？
- リアルタイムの要求に基づいてバランスを変化させる適応戦略はあるのか？

将来への備え：

- ハイブリッド・コントロール・システムは、拡張性、適応性、技術の進歩という点で、どの程度将来性があるのだろうか？
- 新たな技術や管理手法を統合する計画はあるか？

安全性と信頼性：

- さまざまな運転条件において、制御システムの安全性と信頼性はどのように確保されているのか？
- 予期せぬ故障や誤動作に対処するための緊急プロトコルやフェイルセーフはあるか？

倫理的・法的コンプライアンス：

- コントロールシステムの設計と運用において、倫理標準と法的規制の遵守はどのように確保されているか？
- 管理システムによって取られた措置に対する説明責任と責任に対処するための措置があるか。

持続可能性：

- 制御システムの設計は、環境の持続可能性とエネルギー効率をどのように考慮しているか？
- ライフサイクルを通して、制御システムのエコロジカル・フットプリントを最小化する取り組みがあるか？

附属書 20：監査の質問 "安全システム"

物理的なセーフガードとメカニズム

- 機械的バリア、検知、衝突の検知と回避、速度や力の制限などの物理的な保護措置は、身体的危害を防ぐためにどのように設計されているか？
- これらのセーフガードはどのようにテストされ、有効性が評価されるのか？
- キルスイッチ」メカニズムはどのように実装され、意図したとおりに機能し、どのような状況下でもアクセス可能であることを保証するのか？どのようにテストされ、評価されるのか？

センサーフュージョンと信頼性：

- 複数のセンサーからのデータをどのように組み合わせ（センサー・フュージョン）、より包括的で状況を認識した環境画像を作成し、安全なナビゲーションに役立てるのか？
- センサー・フュージョンは、これまでとは異なる、あるいは新たなリスクをもたらす可能性があるのか、また、それはどのように低減されるのか？
- センサーの信頼性と精度はどのように評価されるのか？

システムの冗長性とフォールトトレランス：

- 重要なコンポーネントに冗長性を持たせ、1 つのコンポーネントが故障しても安全な機能を継続できるようなシステムはどのように設計されているのだろうか？
- これらの部品の信頼性はどのように評価されるのか？

ソフトウェアと意思決定の完全性：

- ソフトウエアのアルゴリズムは、どのようにシステムの動作を監視し、安全でない動作を防ぐために介入するのか？
- 介入決定が正確で信頼できるものであることを保証するために、どのような措置が講じられているか？
- 意思決定に人間（あるいは別のシステム）が異議を申し立てることができるか？
- 倫理的な意思決定はどのように確保され、物理的な行動や深く相互接続されたシステムから生じる脆弱性の潜在的な予期せぬ結果を特定するために、どのようなリスクアセスメントが実施されるのか？

人間とロボットの相互作用における安全プロトコル：

- 事故防止のための安全手順や対策はどのように策定されているのか？
- 人間とロボットの明確なコミュニケーションはどのように確保されているのだろうか？
- これらのプロトコルはどのようにテストされ、評価されているのか？継続的にモニターされているか？
- また、インテリジェント・システムの安全性、安全性、信頼性を確保するために、どのような対策が講じられているのだろうか？

附属書 21：監査の質問 "セキュリティ・システム"

暗号化と鍵管理：

- データの機密性は、保存と送信の両方で暗号化によってどのように確保されるのか？
- 暗号鍵が、知る必要がある場合にのみアクセスされ、ポリシーに従って更新／ローテートされ、適切な強度と暗号で作成され、承認され信頼された生成者のみによって生成されることを保証するために、鍵管理はどのように実施されているか。
- 使用されているアルゴリズムやプロトコルは、推奨されているガイドラインや標準に従っているか。また、それらを破る攻撃者の能力の高まりに対応するために、定期的に見直され、更新されているか。
- 将来の量子コンピューティングの脅威に対するレジリエンスに備えるために、二重の暗号化技術が導入されたのだろうか？

データのサニタイゼーション：

- 不要なデータや期限切れのデータを確実に削除し、復元されないようにする方法は？

認証とアクセス管理：

- アクセスを要求する人々やシステム・アカウントの正当性はどのように確認されるのか？
- 認可された事業体だけがアクセスできるようにするために、どの認証方法が使われるか？
- システム・リソースへのアクセスはどのように規制されているのか。また、権限を割り当てるために、どのような事前定義ルールが存在するのか。
- 機密データや機能への不正アクセスはどのように防止されているのか？

セキュリティ監視とインシデント管理：

- 脅威検知、インシデント対応、包括的なセキュリティ監視のための一元化されたプラットフォームを提供するために、複数のソースからのセキュリティデータをどのように統合・分析しているか。
- 異常や潜在的なセキュリティインシデントをリアルタイムで検知し、プロアクティブな脅威検知対策を可能にするために、システムとネットワークの活動を継続的に監視する方法をどのように導入しているか。
- ソフトウエアやハードウエアコンポーネントのセキュリティ脆弱性は、 どのようにして積極的に識別され、軽減されるのか。定期的なアップデートやパッチ適用を通じてか？
- セキュリティ侵害やサイバー攻撃に対処・管理し、被害を最小限に抑え、復旧に要する時間とコストを削減し、事業の継続性を維持することを目的とした、構造化されたアプローチはどのように策定・維持されるのか。

災害復旧と事業継続：

- バックアップは、リスクアセスメントによって設定された、または契約や規制によって要求された、必要なリカバリ時間やポイント目標を保証するために、定期的に実行され、有効性がテストされているか？
- バックアップは不正アクセスや攻撃から特別に保護されているか？
- サイバー攻撃、ソフトウェアやハードウェアの故障、自然災害などの重大な障害が発生した場合、重要なデータやシステムの復旧のための計画や仕組みはどのように実施されているか。

セキュリティ監査と物理的セキュリティ：

- セキュリティ対策の有効性を評価し、業界標準に準拠していることを確認し、改善すべき領域を特定するために、定期的なセキュリティ監査をどのように実施しているか。
- インテリジェント・システムの物理的ハードウェアは、盗難、改ざん、損傷からどのように保護されているのか？

セキュリティ意識向上およびトレーニング：

- インテリジェント・システムのユーザとオペレータは、信頼できるセキュリティ態勢を維持するために、潜在的なセキュリティリスク、ベストプラクティス、インシデント対応手順についてどのように教育されているのだろうか。

附属書 22：監査の質問 "開発とデバッグ"

開発およびテストプラットフォーム：

- 共同開発プラットフォームは、コード共同パイロット、バージョン管理、以前のバージョンに戻す機能をどのように実現し、堅牢なソフトウェアエンジニアリングを促進するのだろうか？
- 開発プロセスを通じてコードのセキュリティを確保するために、どのような対策が講じられているか？
- デバッグ・ツールはコーディング・エラーの特定にどのように役立ち、どのレベルの自動サポートが提供されるのか？
- 自動化された入力をダブルチェックするための手動プロセスはあるか？
- 自動テストフレームワークは、どのようにしてコードの機能を保証し、バグを捕まえるのか？
- 自動入力の妥当性確認をダブルチェックするための手動プロセスはあるか？
- シミュレーション環境は、どのようにして AI システムモデルのトレーニングやテストに安全な空間を提供するのか。また、シミュレーションデータのセキュリティと完全性を確保するために、どのような対策がとられているのか。

パフォーマンス分析ツール：

- パフォーマンス・モニタリング・ツールはどのようにして最適な効率を確保するのか？
- 一般的な KPI は何を測定しているのか？
- プロファイリング・ツールはどのようにコードのパフォーマンスを分析し、効率のボトルネックを突き止めるのか？
- パフォーマンスをモニターするために、どのような KPI が使われているのか？
- ソフトウェア／ハードウェアの潜在的な故障を予測し、保守をプロアクティブに実行するために、ツールをどのように活用できるか？

機械学習と AI 開発：

- 機械学習の枠組みはどのように設計され、知的システムの開発にどのように使われているのだろうか？
- 作成されたアルゴリズムとモデルの品質を保証するために、どのような対策が講じられているか？
- XAI ツールは、モデルがどのように意思決定を行うかを理解する上で、開発者をどのように支援するのだろうか？
- 使用されている XAI ツールの透明性を確保するために、どのような措置がとられているか？

セキュリティと構成管理：

- 開発・展開ツールの安全性を確保するために、どのような措置や仕組みがとられているか？
- これらのツールについて、文書化されたセキュリティーポリシーがあり、セキュリティー管理が行われているか。
- これらのツールの設定は文書化されているか？
- 組織は、これらのツールの構成に関する変更管理プロセスを持っているか？

統合と機能：

- 開発・展開ツールは、組織内の他のシステムと統合できるか？
- 統合はテストされ、文書化されているか？
- 開発・展開ツールはどのような機能を提供するのか？
- そのツールは、AI システムの開発と展開に特化して設計されているか？AI システムをサポートするために必要な機能をすべて備えているか？
- 開発・展開ツールがモデル展開の自動化と精度を確実にサポートするために、どのような対策が講じられているか？

管理と文書化：

- AI の開発ライフサイクルを通じて使用されるツールやライブラリは、プロジェクトの目標を達成し、業界標準を満たすために、どのように監査・監視されているのか？
- サードパーティへの依存度はどのように管理されているのか？
- サードパーティに関連する潜在的なリスクを特定し、低減し、伝達するために、どのような手段を講じているか。
- 開発・展開ツールのベンダーはどのように選ばれたのか？

  選択規準は、組織のニーズに最も適したツールの品質、機能、性能に基づいていたか？

  開発・展開ツールにかかる費用は？
- その費用は組織の予算内に収まっているか？
- AI システムの品質と信頼性を損なうことなく、支出を最適化する方法はあるのだろうか？
- ベンダーはどのようなトレーニングやサポートを提供しているのか？
- ユーザがツールを効果的に使用できるよう、適切な文書とトレーニングがあるか。
- AI システムに使用されている開発・展開ツールの適切な文書があるか？
- ドキュメントは最新で、ツールの機能とロジックを理解するのに十分か？

附属書 23：監査の質問 "エンドユーザ・トレーニングと文書化"

トレーニング教材：

- インテリジェント・システムの目的は、エンドユーザ向けの文書やトレーニング・マニュアルに正確に記述されているか、また、意図された目的に適しているか。
- ユーザ・ドキュメントとトレーニング資料は包括的でよく設計されており、データ保護、プライバシー、情報セキュリティ、倫理的配慮など、インテリジェント・システムの重要な側面をすべて網羅しているか？
- ユーザ・ドキュメントは分かりやすく提供されているか？
- 研修教材は、規制要件や業界標準に準拠しているか。
- トレーニング資料は完成しているか？
- コンプライアンス、透明性、説明責任、公平性といった倫理原則は、インテリジェントシステムのトレーニング教材で扱われているか？

- インテリジェントシステムの使用に関するトレーニング計画には、使用中に生じる可能性のあるバイアスや課題を考慮したシナリオが含まれているか？
- インテリジェント・システムのエンドユーザに対するトレーニングの目的は明確に定義され、トレーニング教材はその目的に対応しているか？
- エンドユーザは、インテリジェント・システムについて、その能力、限界、リスクを含め、簡潔で理解しやすい情報を提供されているか？

  研修資料は、さまざまな学習スタイルや障害に対応できるよう、複数のフォーマットで提供されているか。

  エンドユーザトレーニングは、データ保護、プライバシー、倫理的配慮など、関連する規制やコンプライアンス要件をカバーしているか。
- 研修教材は、サイバーセキュリティリスクの識別や、これらのリスクを低減するための方針など、サイバーセキュリティをカバーしているか。
- 新しい技術コンポーネント、アルゴリズム、あるいはインテリジェント・システムのその他の更新によって発生する可能性のあるリスクは、エンドユーザ・トレーニングでカバーされているか？
- エンドユーザは、テスト、アンケート、またはフィードバックの仕組みを通じて、インテリジェント・システムの理解度を定期的にアセスメントされ、その回答は記録され、評価されているか？
- 不適切なエンドユーザトレーニングに関連したインシデントは、トレーニング資料を更新することで適切に対処されているか。

KPI とメータリング：

- ユーザがインテリジェント・システムをどの程度効果的に理解し、使用しているかを評価するために、ユーザからのフィードバック指標を記録し、評価しているか。
- 研修の効果を測定し、定期的に評価しているか。
- 研修を実施し、進捗状況を文書化するために使用される方法は、組織の方針および規制ガイドラインに従って維持されているか。
- エンドユーザトレーニングとドキュメンテーションの有効性を評価するために、どのような指標を用いているか？これらの評価基準は明確に定義され、どのように測定されているか？

附属書 24：監査の質問 "展開とモニタリング "

安全とセキュリティ：

- AI システムのモデル障害やセキュリティ侵害を管理するために、どのようなインシデント対応戦略があるのか？

- 技術的および組織的な対応戦略の準備と有効性を、対応時間やコミュニケーション方法を含めて実証できるか。

  法律で義務付けられている（重大インシデントが発生した場合の）強制的な情報開示手続きは確立されているか、またどのように実施されているか。

  アクセス管理を行い、認可された個人だけがインテリジェントシステムを変更できるようにするために、組織はどのような対策を実施してきたか？

- 侵入検知のためのプロアクティブなモニタリングはどのように行われているか、また、組織はどのような技術や方法論を使用しているか。
- 警告はどのように管理され、エスカレーションされるのか？誰がエスカレーションの責任を負うのか？
- 脆弱性スキャンの頻度と深さ、パッチの優先順位と適用方法は？
- 予期せぬ重大な故障や極端なシナリオに対して、どのようなコンティンジェンシープランがあるのか。
- 致命的な故障に対処し、悪条件下でも継続的なサービスを保証するためのバックアッププランは何か？
- リアルタイムのパフォーマンスを監視し、AI システムモデルの目標のドリフトを検知するために、継続的な評価はどのように行われるのか？
- ドリフトはどのように検知されるのか？そして誰が責任を負うのか？

文書化と管理：

- 規制環境におけるリリース管理は、外部および内部の標準にどのように準拠するのか？
- インテリジェント・システムの新バージョンへの準拠はどのようにして確保されるのか？どのようにテストされるのか？
- 更新や修正における透明性と説明責任を確保するために、どのようなバージョン管理を実施しているか。
- 変更履歴の文書化の範囲と詳細は、監査やコンプライアンス上の要求に対して満足のいくものであるか。
- インテリジェント・システムの展開と運用手順に関する文書は、どの程度包括的で最新か？
- すべての文書は最新か、アクセス可能か、監査可能か。
- 対応する文書へのアクセスについて、どのような制限や要件が考えられるか。

附属書 25：監査の質問 "廃止"

システムの識別：

- 具体的にどのようなインテリジェント・システムが廃止されるのか？
- このインテリジェント・システムの主な機能と使用例は何か？
- インテリジェント・システムの廃止は、現在のビジネス・プロセスやオペレーションにどのような影響を与えるのか？
- このインテリジェント・システムに、対処すべき重要な依存関係はあるか？
- すべての関係者は、廃止プロセスについて知らされているか？
- 利害関係者の懸念やフィードバックはどのように管理されているか？

廃止の理由：

- なぜインテリジェント・システムは廃止されるのか？
- 代替の解決策やシステムが特定され、評価されているか。

廃止のためのガバナンス：

- 廃止を認可するガバナンス委員会があるか？会議の議事録／記録は、誰が認可され、決定する権限を与えられているかを反映しているか？
- インテリジェント・システムの最終的な廃止決定（予想通り、あるいは予想外）の責任は誰にあるのか？上級管理職が承認したのか？
- インテリジェント・システムに依存する意思決定プロセスの継続性を、組織はどのように確保しているか？
- 事業運営をシームレスに維持するために、どのような代替プロセスやシステムを導入しているか？

データ管理とセキュリティ：

- 廃止時のデータ移行にはどのような手順があるのか。
- 不要になったデータは、どのように安全に廃棄または識別解除されるのか？
- 組織は、廃止段階におけるアクセス管理とユーザ権限をどのように扱っているか？
- 廃止措置中の不正アクセスはどのように防止されるのか？
- すべての重要なデータは、紛失や破損することなく安全に転送またはアーカイブされているか？
- 移行中や廃棄中にデータの完全性と完全性を検証するために、どのような対策が講じられているか？
- データ保持方針はどのようなもので、関連法規にどのように準拠しているか。

- ログとメタデータは（コンプライアンスとフォレンジック分析のために）安全にアーカイブされ、どれくらいの期間保存されているか。

緊急時の計画とリソース：

- 文書化された緊急廃止計画があり、どのくらいの頻度でテストされているか？
- 予期せぬ廃止シナリオのために、どのようなリソースが利用可能か？
- 緊急時計画の有効性はどのように評価され、テストされるのか？

トレーニング、モニタリング、コンプライアンス：

- 廃止措置プロセスに携わるスタッフには、どのようなトレーニングおよび意識向上プログラムが提供されているか？
- コンプライアンスと監査を目的として、組織はどのように廃止プロセスを監視し、文書化しているか？

廃止システムの余波：

- 当該ソリューションが、記録されたインテリジェントシステムのインベントリにおいて「非アク ティブ／廃止」とマークされていること、及び、廃止後、リスク登録の関連するリスク 決定が「クローズ／該当なし」又は「アクション可能」であることを確認する。
- もはや関連性のないリスクや問題は解決されているか？

略語と用語集

AC	Alternating Current	交流
AI	Artificial Intelligence	人工知能
AI RMF	Artificial Intelligence Risk Management Framework	人工知能リスクマネジメントフレームワーク
AIBOM	AI Bill of Materials	AI 部品表
API	Application Programming Interface	アプリケーション・プログラミング・インターフェース
APT	Advanced Persistent Threats	高度な持続的脅威
AR	Augmented Reality	拡張現実
ASCON	Family of lightweight authenticated ciphers	軽量認証暗号ファミリー
BCI	Brain Computer Interface	ブレイン・コンピューター・インターフェイス
BLE	Bluetooth Low Energy	ブルートゥース・ロー・エナジー
BOM	Bill of Materials	部品表
CAIO	Chief Artificial Intelligence Officer	最高人工知能責任者
CAO	Chief Audit Officer	最高監査責任者
CCPA	California Consumer Privacy Act	カリフォルニア州消費者プライバシー法
CIA	Confidentiality, Integrity, and Availability	機密性、完全性、可用性
CoAP	Constrained Application Protocol	制約付きアプリケーション・プロトコル
CPRA	California Privacy Rights Act	カリフォルニア州プライバシー権法
CPU	Central Processing Unit	中央処理装置
CSA	Cloud Security Alliance	クラウド・セキュリティ・アライアンス

DC	Direct Current	直流
DDS	Data Distribution Service	データ配信サービス
DMCA	Digital Millennium Copyright Act	デジタル・ミレニアム著作権法
DNA	Deoxyribonucleic Acid	デオキシリボ核酸
DORA	Digital Operational Resilience Act	デジタル・オペレーショナル・レジリエンス法
EMP	Electromagnetic Pulse	電磁パルス
ENISA	European Union Agency for Cybersecurity	欧州連合サイバーセキュリティ機関
EU	European Union	欧州連合
EU AI Act	European Union Artificial Intelligence Act	EU 人工知能法
FADP	Swiss Federal Act on Data Protection	データ保護に関するスイス連邦法
GANs	Generative Adversarial Networks	生成的敵対ネットワーク
GDPR	General Data Protection Regulation	一般データ保護規則
GenAI RMF	Generative AI Risk Management Framework	生成的 AI リスクマネジメントフレームワーク
GPU	Graphics Processing Unit	グラフィックス・プロセッシング・ユニット
HDD	Hard Disk Drive	ハードディスク・ドライブ
HSM	Hardware Security Modules	ハードウェア・セキュリティ・モジュール
IDE	Integrated Development Environment	統合開発環境
IDS	Intrusion Detection Systems	検知システム
IoT	Internet of Things	モノのインターネット
ISO	International Organization for Standardization	国際標準化機構
KPI	Key Performance Indicator	主要業績評価指標

KRI	Key Risk Indicator	主要リスク指標
LLM	Large Language Model	大型言語モデル
LoRaWAN	Long Range Wide Area Network	長距離ワイドエリア・ネットワーク
MFA	Multi-factor Authentication	多要素認証
NB-IoT	Narrowband Internet of Things	ナローバンド・インターネット・オブ・シングス
NIS2	Network and Information Security Directive 2	ネットワークと情報セキュリティ指令 2
NIST	National Institute of Standards and Technology	国立標準技術研究所
NLP	Natural Language Processing	自然言語処理
OPC UA	OPC Unified Architecture	OPC ユニファイド・アーキテクチャ
OSHA	Occupational Safety and Health Administration	労働安全衛生局
PCA	Principal Component Analysis	主成分分析
PII	Personally Identifiable Information	個人を特定できる情報
RAM	Random Access Memory	ランダム・アクセス・メモリー
RFID	Radio Frequency Identification	無線周波数識別
SAIBOM	Software and AI Bill of Materials	ソフトウェアと AI 部品表
SBOM	Software Bill of Materials	ソフトウェア部品表
SEC	Securities and Exchange Commission	証券取引委員会
SEO	Search Engine Optimization	検索エンジン最適化
SIEM	Security Information and Event Management	セキュリティ情報とイベント管理
Sigfox	Low-power, long-range wireless communication protocol	低消費電力、長距離無線コミュニケーション・プロトコル

SMPC	Secure Multi-Party Computation	セキュアなマルチパーティ計算
SSD	Solid State Drives	ソリッド・ステート・ドライブ
TEE	Trusted Execution Environments	信頼可能な実行環境
TOS	Terms of Service	利用規約
TPU	Tensor Processing Unit	テンソル処理ユニット
UI	User Interface	ユーザ・インターフェース
VDP	Vulnerability Disclosure Program	脆弱性開示プログラム
VR	Virtual Reality	仮想現実
WLAN	Wireless Local Area Network	ワイヤレス・ローカルエリア・ネットワーク
XAI	Explainable AI	説明可能な AI
Zigbee	Low-power, short-range wireless communication protocol	低消費電力、近距離無線コミュニケーション・プロトコル
Z-Wave	Wireless protocol for home automation	ホームオートメーション用無線プロトコル

CSA はオンライン用語集[49]を提供している。

参考文献

[1]	世界デジタル技術院（WDTA）、「生成的 AI アプリケーション・セキュリティ・テストと妥当性確認標準」、世界デジタル技術院（WDTA）、04 2024。[オンライン]。利用可能: https://wdtacademy.org/publications/GenerativeAiApplicationSecurityTestingAndValidationSt andard.[Accessed 14 08 2024].
[2]	L.L. Floridi, M. Holweg, M. Taddeo, J. Amaya, J. Mökander and Y. Wen, "capAI - A Procedure for Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act," Center for Digital Ethics.
[3]	NIST, "NIST AI RMF Playbook," National Institute of Standards and Technology (NIST), [Online].利用可能: https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.[Accessed 14 08 2024].
[4]	OWASP Top 10 for LLM Applications Team, "LLM AI Cybersecurity & Governance Checklist," OWASP, 2024.
[5]	World Digital Technology Academy (WDTA), "Large Language Model Security Testing Method," World Digital Technology Academy (WDTA), 04 2024.[オンライン]。利用可能: https://wdtacademy.org/publications/LargeLanguageModelSecurityTestingMethod.[Accessed 14 08 2024].
[6]	J.Mökander, J. Schuett, H. R. Kirk and L. Floridi, "Auditing large language models: a three-layered approach," *AI and Ethics,* 05 2023.
[7]	D.Stocker, J. Martella, A. Sharpe and I. Okoli, "Don't Panic ! Getting Real About AI Governance," Cloud Security Alliance (CSA), [Online].利用可能： https://cloudsecurityalliance.org/artifacts/dont-panic-getting-real-about-ai-governance。[Accessed 18 09 2024].
[8]	ISO, "ISO/IEC 42001:2023", 2023.[オンライン]。利用可能： https://www.iso.org/standard/81230.html。[Accessed 14 08 2024].
[9]	ISO, "iso/iec 5338:2023," 2023.[オンライン]。利用可能: https://www.iso.org/standard/81118.html.[Accessed 14 08 2024].

[10]	C.Spleiss, R. Ayalin, F. Chyla, B. Gaylord, F. Haenig, R. Heckman, H. Labib, L. Ruddikeit, A. Sharpe and A. Vashishtha, "AI レジリエンス : AI Resilience: A Revolutionary Benchmarking Model for AI Safety", Cloud Security Alliance (CSA), 2024.
[11]	C.Spleiss, "AI Resilience & Diversity - BLOG," Cloud Security Alliance (CSA), 2024.
[12]	J.Mökander, "Auditing of AI: Legal, Ethical and Technical Approaches," *Digital Society (DISO),* vol. 2 , no.49 (2023), 08 11 2023.
[13]	ISO、「人工知能の概念と用語」、[オンライン]。利用可能： https://www.iso.org/standard/74296.html。[Accessed 02 10 2024].
[14]	P.フィリップス、C.ハーン、P.フォンタナ、A.イェーツ、K.グリーン、D.a.P.M.ブローニャトウスキー、「説明可能な人工知能の 4 原則」、2021 年。
[15]	imda & pdpc - Singapore, "Model Artificial Intelligence Governance Framework," 21 01 2020.[オンライン]。利用可能： https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodela igovframework2.pdf.[Accessed 13 08 2024].
[16]	CSA Workgroup AI Governance and Compliance, "AI Governance & Compliance Resource Links Hub," Cloud Security Alliance (CSA), [Online].利用可能： https://cloudsecurityalliance.org/ai-governance-compliance-resource-links。[Accessed 13 08 2024].
[17]	ISO, "ISO/IEC 27001:2022," [オンライン]。利用可能: https://www.iso.org/standard/27001.[Accessed 13 08 2024].
[18]	NIST, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," National Institute of Standards and Technology, 01 2023.[オンライン]。利用可能： https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.[Accessed 13 08 2014].
[19]	NIST、「人工知能リスクマネジメントフレームワーク」: 生成的人工知能プロファイル : NIST AI 600-1」国立標準技術研究所、07 2024。[オンライン]。 https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf。[Accessed 13 08 2024].

[20]	Cyber Risk GmbH, "The Framework for AI Cybersecurity Practices (FAICP)," [Online].利用可能: https://www.faicp-framework.com/.[Accessed 13 08 2024].
[21]	国連、「17 の目標」、[オンライン]。利用可能：https://sdgs.un.org/goals.[Accessed 13 08 2024].
[22]	NASA, "The Unknown Known", システム障害ケーススタディ, vol. 3, no.01, 01 2009.
[23]	Digital Curation Centre Trilateral Research, "The Role of Data in AI," 2020.
[24]	「一般データ保護規則（GDPR）」EUR-Lex, 27 04 2016.[オンライン].利用可能：https://eur-lex.europa.eu/eli/reg/2016/679/oj。
[25]	GDPR.EU、「データ保護影響アセスメント（DPIA）」、保護技術 AG、[オンライン]。利用可能: https://gdpr.eu/data-protection-impact-assessment-template/.[Accessed 13 08 2024].
[26]	CA.GOV, "California Privacy Protection Agency," [Online].利用可能: https://cppa.ca.gov/.[Accessed 13 08 2024]。
[27]	スイス連邦評議会、「データ保護に関する連邦法（FADP）」、[オンライン]。利用可能：https://www.fedlex.admin.ch/eli/cc/2022/491/en。[Accessed 13 08 2024].
[28]	米国著作権局、「デジタルミレニアム著作権法」、[オンライン]。利用可能：https://www.copyright.gov/dmca/。[Accessed 13 08 2024].
[29]	R.Matulionyte, "Researchers warn we could run out of data to train AI by 2026.ではどうするのか？", 07 11 2023.[オンライン]。Available: https://theconversation.com/researchers-warn-we-could-run-out-of-data-to-train-ai-by-202 6-what-then-216741.[Accessed 13 08 2024].
[30]	P.ヴィラロボス、A.ホー、J.セビージャ、T.ベシログル、L.H.M.ハイム、"Will we run out of data?人間が生成したデータに基づく LLM スケーリングの限界" 04 06 2024.[オンライン]。Available: https://arxiv.org/abs/2211.04325.[Accessed 13 08 2024].
[31]	R.エルガート「データの闇市場：ハッカーが盗んだデータを持ち出す場所」、08 05 2019.[オンライン]。入手可能：

	https://www.turn-keytechnologies.com/blog/article/the-data-black-market-where-hackers-ta ke-stolen-data.[Accessed 13 08 2024].
[32]	Y.-w.Kim, "How Transferable are Video Representations Based on Synthetic Data?", MIT, 2022.
[33]	X.Lin, J. Liu, J. Hao, K. Wang, Y. L. H. Zhang and H. a. T. X. Horimai, "Collinear holographic data storage," *Opto-Electronic Advances,* vol. 3, no.3, 2020.
[34]	L.Ceze and J. &.S. K. Nivala, "Molecular digital data storage using DNA," *Nature Reviews Genetics,* vol. 20, p. 456-466, 2019.
[35]	R.Grimes, E. Chiu, J. Gable, B. Huttner and L. Perret, "Practical Preparations for the Post-Quantum World," Cloud Security Alliance (CSA), 2021.
[36]	NIST, "Lightweight Cryptography", 国立標準技術研究所, [Online].利用可能: https://csrc.nist.gov/Projects/lightweight-cryptography.[Accessed 14 08 2024].
[37]	T.Liu and G. a. J. R. Ramachandran, "Post-Quantum Cryptography for Internet of Things : A Survey on Performance and Optimization", 2024.
[38]	欧州委員会、「EU AI 法」、欧州委員会、2021 年。
[39]	EU 理事会 , 「人工知能法 : EU 理事会と欧州議会は、世界初の AI に関する規則について合意した。[オンライン].利用可能 : https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-a ct-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/. [Access 14 08 2024].[Accessed 14 08 2024].
[40]	コロラド州議会「人工知能の消費者防御 : SB24-205」、2024 年。[オンライン]。利用可能: https://leg.colorado.gov/bills/sb24-205.[Accessed 14 08 2024].
[41]	J.Biamonte, P. Wittek, N. Pancotti, P. Rebentrost and N. a. L. S. Wiebe, "Quantum Machine Learning," 2018.
[42]	C.ニューロモーフィック・コンピューティング・アルゴリズムとアプリケーションの可能性」、*Nature Computational Science*、2 巻、10-19 頁、2022 年 1 月 31 日。

[43]	N.バーニー、「ニューロモーフィック・コンピューティングとは何か」、TechTarget、04 2023。[オンライン]。利用可能: https://www.techtarget.com/searchenterpriseai/definition/neuromorphic-computing.[Accessed 14 08 2024].
[44]	U.S. Department of Labor, "Hazards Associated with Industrial Robot Applications," 米国労働省、[オンライン]。利用可能: https://www.osha.gov/otm/section-4-safety-hazards/chapter-4#hazards.[Accessed 13 08 2024].
[45]	ハーバード大学プライバシー・ツール・プロジェクト、「差分プライバシー」、 [オンライン]。利用可能: https://privacytools.seas.harvard.edu/differential-privacy.[Accessed 14 08 2024].
[46]	K.Martineau, "What is federated learning?", IBM, 24 08 2022.[オンライン]。利用可能 : https://research.ibm.com/blog/what-is-federated-learning。[Accessed 14 08 2024].
[47]	コンフィデンシャル・コンピューティング・コンソーシアム、「コンフィデンシャル・コンピューティング・コンソーシアムについて」、[オンライン]。利用可能: https://confidentialcomputing.io/about/.[Accessed 14 08 2024].
[48]	S.Burke, M. Capotondi, D. Catteddu and K. Huang, "Large Language Model (LLM) Threats Taxonomy," Cloud Security Alliance (CSA), 2024.
[49]	CSA, "Cloud Security Glossary", Cloud Security Alliance (CSA), [Online].利用可能: https://cloudsecurityalliance.org/cloud-security-glossary.[Accessed 14 08 2024].

© Copyright 2024, Cloud Security Alliance.無断複写・転載を禁ずる。