

中小企業における AI導入のための ツールキット



イントロダクション

このツールキットは、カナダが議長国を務めた2025年のG7カナナスキス・サミットでG7首脳が表明したコミットメントに基づいて作成されています。中小企業（SMEs）—零細企業を含む—が、安全で責任ある、信頼できる人工知能（AI）を導入することを支援するために設計されており、高度なAIシステムを開発する組織向けの広島AIプロセス（HAIP）国際指針に沿っています。

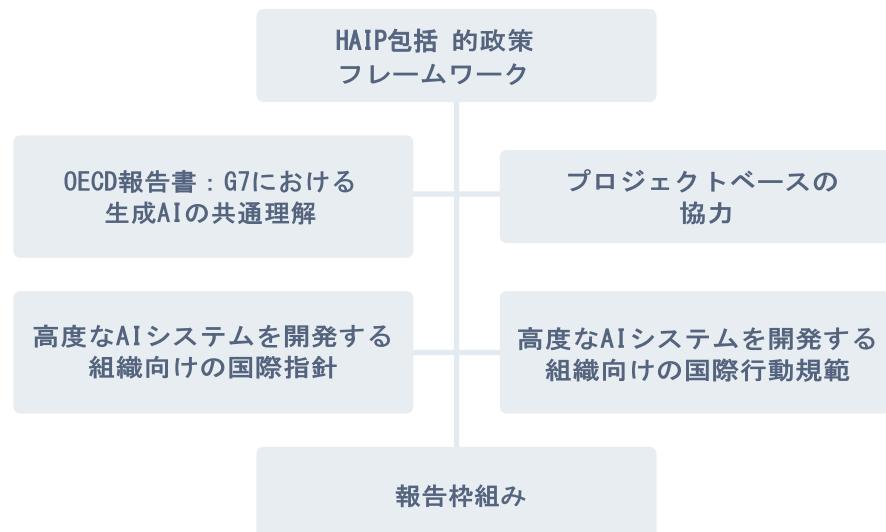
このツールキットは、2025年7月24日に開催されたG7バーチャルワークショップ「信頼できるAI導入の推進：G7広島AIプロセスの成果の活用」で得られた知見と提言を取り入れています。さらに、その後のフォローアップ調査からのフィードバックや、G7メンバーおよび主要な関係者（モントリオール国際AI専門センター〔CEIMIA〕、経済協力開発機構〔OECD〕を含む）からの貢献も反映しています。ワークショップと調査には、政府、大小さまざまな産業、非営利団体、学術機関を代表する26か国以上、260名を超える参加者が集まりました。

G7諸国では、中小企業は通常、従業員数によって定義されます。本ワークショップおよびツールキットでは、中小企業を「従業員数250人未満の企業」（[OECD](#)）と分類しています。[OECD](#)によると、中小企業は全事業者の99%を占め、労働者の3分の2を雇用し、OECD加盟国全体で付加価値の50～60%を生み出しており、彼らのデジタル変革は経済の安定性と世界的なイノベーションに不可欠です。しかし、中小企業は信頼できるAIを導入する際、大企業と比べてリソース面で制約に直面する可能性があります。特にAIガバナンスの状況が急速に進化している中ではなおさらです。G7諸国が中小企業におけるAI導入を加速させる中で、エンドユーザーに安心感を与え、市場機会を開く信頼構築のために、責任ある導入を行える体制を整えることが極めて重要です。

本ワークショップおよびツールキットでは、中小企業のAI開発者とAI導入者を区別し、後者を主な対象としました。開発者はAIシステムを設計・構築・学習させる役割を担います。一方、導入者は特定の状況でAIシステムを統合し、利用する役割を担い、場合によってはAIシステムの「ユーザー」にもなります。この区別により、AIシステムの開発に特化した企業だけでなく、既存の事業にAIを採用・統合することに关心を持つ多数の中小企業も含まれます。

このワークショップとツールキットは、中小企業のAI導入者が自社で責任あるAI活用を目指す際に直面する課題に対応することを目的としています。第1章ではHAIP包括的政策フレームワークの概要を簡単に紹介し、第2章では中小企業がHAIP国際指針を自社の状況に適用する際に報告した主な困難を取り上げます。第3章では、中小企業内部に存在する実務的な課題に焦点を当て、これらがHAIP国際指針の実施や、最終的な信頼できるAI導入を妨げる要因となり得ることを説明します。

セクション1 - 広島AIプロセス(HAIP)概要



2023年にHAIP包括的政策フレームワークは、日本が議長国を務めた2023年のG7の下で開始された取組であり、バリューチェーン全体にわたる高度なAIシステムのガバナンスを形成することを目的としています。このフレームワークは次の4つの柱で構成されています：(1) 生成AIに起因する優先的なリスク、課題、機会の分析に関する報告、

(2) 高度なAIシステムを開発する組織向けの広島AIプロセス国際指針、(3) 高度なAIシステムを開発する組織向けの広島AIプロセス国際行動規範、(4) AIに関するプロジェクトベースの協力。

OECDのAI原則を基盤として、国際指針は次の点に焦点を当てています：

- **ライフサイクル**: AIのライフサイクル全体を考慮し、レッドチーミングやトレーサビリティによって不正利用や脆弱性を軽減すること（原則1、2）
- **透明性**: 説明責任を高めるために、透明性の促進、情報共有、報告を行うこと（原則3、4、5）
- **セキュリティ**: セキュリティ対策への投資に加え、AI生成コンテンツを識別するためのコンテンツ認証および出所確認の仕組みを導入すること（原則6、7）

国際指針は、最先端の基盤モデルや生成AIシステムを含む、高度なAIシステムを責任を持って開発・導入する際に考慮すべき基本事項を示しています。国際行動規範は、国際指針の実施に関する詳細を提供し、高度なAIシステムを開発する組織に推奨される具体的な行動を示しています。

- **技術の進展**: リスク軽減の最先端を進めるための研究に投資するとともに、世界的・社会的課題の解決やデジタルリテラシー向上の取組みを推進すること（原則8、9）
- **標準化**: 国際的な技術標準の策定（原則10）
- **プライバシー**: 個人データおよび知的財産を保護するための仕組みを実装すること（原則11）

2024年のイタリアG7議長国の下で始まり、同年6月の「ブーリア・コムニケ」でG7首脳により再確認された取組として、G7はOECDに対し、広島AIプロセスの国際行動規範の自主的な採用状況をモニタリングする仕組みの開発を依頼し、「報告枠組み」が開発されました。報告枠組みは、組織が国際行動規範に沿って取組を進め、安全策を組み込み、透明性を高めるための質問を盛り込んだ、任意で利用できるツールです。

セクション2 – HAIPの国際指針と中小企業

ワークショップの重要な焦点は、国際指針がどのように中小企業のAIシステム導入・活用を支援できるかを理解することでした。議論の中で、中朝企業にとって適用が最も難しいと頻繁に指摘されたのは、原則1、2及び10です。これらの課題を克服することは、多くの中小企業が信頼できるAI導入を実現するための重要な第一歩となります。本セクションでは、その目標達成に向けた関連する考慮事項を示します。

原則1 – AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる

原則10 – 国際的な技術規格の開発を推進し、適切な場合にはその採用を推進する

考慮事項: AIシステムを業務に導入する前に、スケジュール管理や受注整理など、どのような機能であっても、潜在的なリスクを慎重に検討する必要があります。業務遂行においてどのようなリスクが発生し得るか、AIツールの利用が適切か、またどの種類のAIツールがその状況にふさわしいかを考慮するべきです。特にリスクの高い業務機能にAIを導入する場合は、信頼できる第三者から、使用中に発生し得るリスクを軽減するための措置（テストや適切な使用方法を説明するマニュアルの提供など）を講じたシステムを調達することが有効です。

考慮事項: 認知された標準を採用することで、企業はステークホルダー、顧客、パートナーに対して信頼性と信用を高めることができます。標準は、明確な期待値やガイドラインを提供することで、組織間の整合性を強化し、戦略的な意思決定をより適切に行うための支援にもなります。AI標準に準拠した製品を提供するベンダーを選ぶことで、責任ある取組を行っているという確証が得られ、AI導入に伴う潜在的なリスクを軽減できます。さらに、中小企業は業界主導のAI標準策定に参加する機会を見つけるべきです。

中小企業が標準策定に関与することで、大企業にはない独自の視点を標準の開発に提供できます。

原則2 – 市場投入を含む導入後、脆弱性及び必要に応じて悪用されたインシデントやパターンを特定し、緩和する

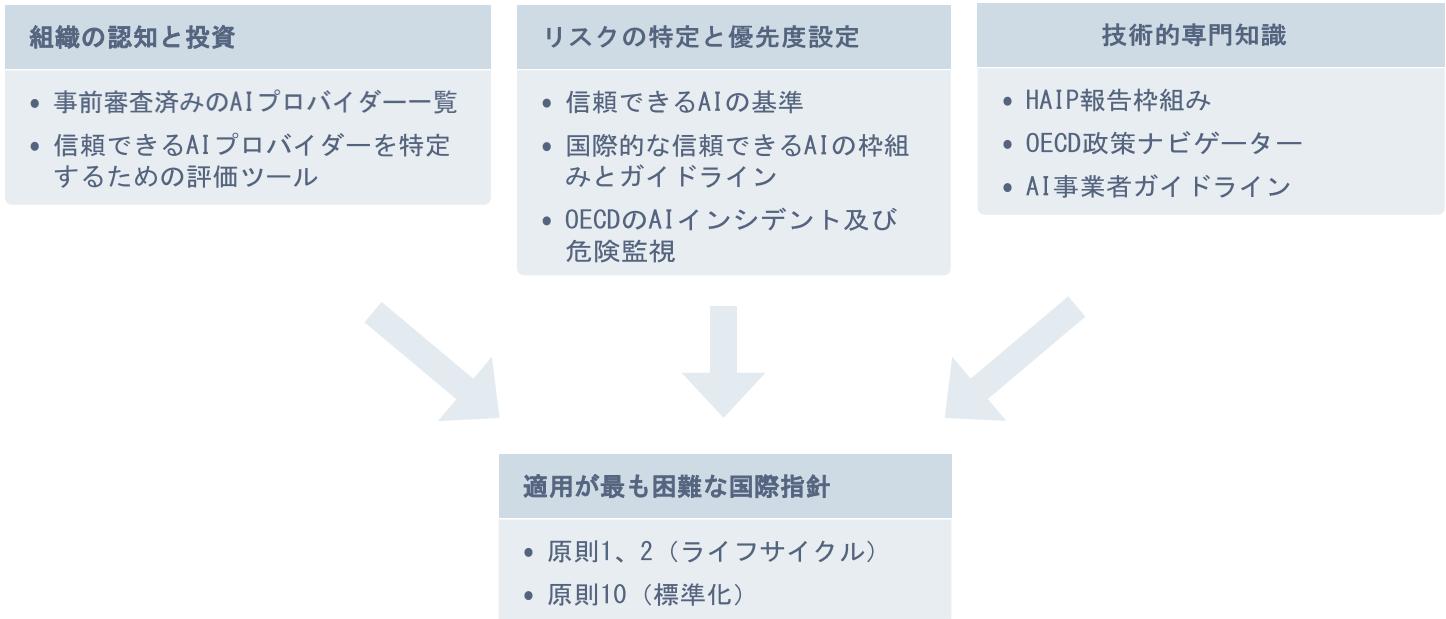
考慮事項: リスクの高い高度なAIシステムを導入する場合、意図しない影響やエラー、不適切な利用といったインシデントを継続的に監視し、問題が発生した際には迅速に対応することが重要です。例えば、AIを活用したカスタマーサービスのチャットボットを使用する場合、その応答を監視することで、期待されるサービスレベルが維持されているか確認できます。もしチャットボットが体系的に誤った情報を提供したり、意図しない目的に誘導される可能性がある場合、それは不具合や悪用の脆弱性を示しているかもしれません。そのような場合は、開発者と連携して問題を軽減する対応が必要です。



セクション3 - 課題とツール

ワークショップでは、中小企業がHAIP国際指針に沿ってAIを導入する際に直面する実務的な障壁についても議論しました。これにより、セクション2で特定された原則（原則1、2、10）の実施に伴う具体的な課題について、追加の知見が得られました。本セクションでは、これらの課題を取り上げ、それぞれに対応する参考リソースを提示し、中小企業が克服するための手助けとなる情報を提供します。

課題解決のためのリソース



組織の認知と投資のためのリソース：

参加者は、組織内の事業部門間で整合性が取れない可能性を指摘しました。こうした不一致は、信頼できるAIに対する組織の理解や投資を制限する傾向があります。AIが特定の業務プロセスや目標を支援する方法を明確に理解していない場合、意思決定者は組織や顧客に利益をもたらす機会を見逃す可能性があります。さらに、AIを業務機能に統合する際の潜在的なリスクについても明確に把握できない場合があります。

AI導入を始めたばかりの中小企業にとって、すでに信頼できるAIとして認められているサービスを選択することは、プロセスを大幅に効率化できます。このアプローチにより、各システムの機能やリスク管理を独自に評価する時間と労力を避けられます。さらに、一部の政府は、強固なガバナンスを示すAIプロバイダーに対して、事前承認リストや審査プロセスを提供しており、特に公共調達において有効です。これにより導入が簡素化され、中小企業は広範な評価に時間を費やすことなく、AIの活用に集中できます。

例として、カナダ政府は責任ある効果的なAIサービス、ソリューション、製品を提供するAIサプライヤーの公開リストを提供しています：[AIサプライヤーリスト - Canada.ca](https://ai-supplier-list-canada.ca)。

このリストは公共調達向けに設計されていますが、関連分野で実績のあるAIサプライヤーを探している中小企業にとっても有益なリソースとなります。掲載されているサプライヤーは、[アルゴリズム影響評価](#) (AIA) を完了し、信頼性と効果的なAIを提供するための事前資格を取得しています。これにより、中小企業は事前審査済みのサプライヤーを利用でき、専門性やガバナンス体制に対する信頼性を高めることができます。

他国もこの分野で取組を進めています。例えば、英国政府はAIの効果的な導入を可能にするために[「AI調達ガイドライン」](#)を提供しています。さらに、英国は2025年1月に[「AI管理の基本 \(AIME\)」](#)ツールに関する公開協議を完了しました。このツールは、組織が責任あるAI管理システムやプロセスを評価・導入するための自己評価ツールとして設計されています。

さらに、OECDの「[信頼できるAIのためのツール&指標カタログ](#)」は、中小企業に対し、責任ある倫理的AIの原則に沿ったAIプロバイダーを特定するためのリソースや評価ツールの有用なディレクトリを提供しています。このカタログは、信頼性のある実践を示すサプライヤーやソリューションを見つけるのに役立ち、AIパートナーの評価と選定を容易にします。

リスクの特定と優先度設定のためのリソース:

中小企業は、AI導入に伴うリスクの特定や優先順位付けにおいて困難に直面することがよくあります。特に、急速に進化する技術や変化する規制要件に対応する中で、その難しさが増します。さらに、限られたリソースが、こうしたリスクを効果的に管理・軽減する能力を一層制約する可能性があります。

OECDの「[AIシステム分類フレームワーク](#)」は、利用状況、人の関与度、機能、リスクプロファイルに基づいて、さまざまな種類のAIを特定・評価するために設計されています。その主な目的は、情報に基づいた意思決定と責任あるAI導入を支援することです。導入前にこのフレームワークを使用することで、潜在的なリスクやソリューションの複雑さを把握し、適切な安全策を実装できます。例えば、カスタマーサービス向けのAIチャットボットを検討する場合、このフレームワークはシステムの機能、必要な監督レベル、管理すべきリスクを明確にします。適用方法としては、利用ケースをガバナンス要件にマッピングし、必要な安全策を定義し、それをベンダーへの期待事項に反映させます。システムが拡大する際には、リスク管理策が有効であり続けるよう、このフレームワークを再確認することが推奨されます。

[ISO/IEC 42001:2023](#)は、組織内でAIマネジメントシステム（AIMS）を確立、実施、維持、継続的に改善するための要件を規定した国際規格です。この規格は、AIベースの製品やサービスを提供または導入する事業体向けに設計されており、AIシステムの責任ある開発と利用を確保します。標準は、ポリシーの策定、リスク管理の実践、責任体制、監視メカニズムの確立に関する明確なガイダンスを提供し、AIシステムが責任を持って透明性のある方法で使用されることを保証します。あらゆる規模・業種の組織に適用可能で、AIを責任ある効果的な方法で管理するためのツールを提供します。また、進化するベストプラクティスや技術の進歩を反映するため、隨時更新される予定です

カナダ・デジタルガバナンス評議会の[CAN/DGSI 101:2025規格](#)「中小企業によるAIの倫理的設計と利用」は、中小企業がAIを倫理的に導入・活用するための新しいロードマップです。この規格は、AIシステムに倫理を統合するための明確なガイドラインと要件を提供し、リスク管理、倫理的設計、導入戦略、継続的な監視といった重要分野を網羅しています。特に従業員500人未満の組織向けに設計されており、最高水準の倫理基準を維持しながらAIを自信を持って活用するためのツールを提供しま

す。また、AIの進歩に対応するため、定期的に更新される予定です。

OECDの「[AIインシデント&ハザードモニター（AIM）](#)」は、中小企業がAIリスクを特定し、優先順位を付けるのに役立ちます。AIMは、業界特有のインシデントや類似導入での一般的な問題を提示し、注意すべき共通リスク領域に関する洞察を提供します。AIのインシデントやハザードをカタログ化することで、リスクや被害を明示し、時間の経過とともにパターンを明らかにし、AIに関する信頼できる共通理解を支援します。さらに、国、業界、AI原則、重大度などの柔軟なフィルタリング機能を備えており、ユーザーは特定のユースケースにおける高リスク領域を迅速に把握できます。

Vector Instituteの「[責任ある生成AIガバナンスガイド](#)」は、生成AIの責任あるガバナンスのための実践的なリソースを提供します。このガイドは、コントロールやリスク指標に関するユーザーフレンドリーなインターフェースを通じて、課題やリスクの概要を把握できるよう設計されています。プロジェクト内のリスクの特定と軽減を支援し、実装計画に役立つ多数のリソースを含んでいます。プラットフォームには、リスク評価を含むすべての機能をナビゲートできるAIチャットボットが組み込まれており、特定のビジネスケースを追加して、カスタマイズされたリスク特定を行うことも可能です。

「[人権、民主主義、法の支配の観点からの人工知能システムのリスクおよび影響評価の方法論（HUDERIAメソドロジー）](#)」は、AIシステムのライフサイクル全体にわたる潜在的なリスクと社会的影響を自動的に評価するための構造化された枠組みと指針を提供します。この方法論は、欧州評議会の人工知能委員会によって採択され、AIの設計、開発、導入、利用において、人権、民主主義、法の支配の原則を実装することを支援します。HUDERIAメソドロジーには、コンテキストに基づくリスク分析が含まれています。これは、AIシステムが人権、民主主義、法の支配に与える可能性のあるリスクを特定し理解するために必要な情報を収集・マッピングする構造化されたアプローチを提供し、AIが当該問題に対する適切な解決策であるかどうかを判断する助けとなります。また、ステークホルダー・エンゲージメントプロセスがあり、関係するステークホルダーを巻き込み、影響を受ける可能性のある人々から洞察を収集し、潜在的な被害や緩和策を文脈化するためのアプローチを提案します。さらに、リスクおよび影響評価では、人権、民主主義、法の支配に関連するリスクと影響を評価するための手順を概説します。最後に、緩和計画があり、救済措置へのアクセスや反復的なレビューを含む、緩和および是正措置を定義するプロセスを説明します。この方法論は、業種を問わ

ず適用可能で、さまざまな文脈に適応できるよう設計されており、AI技術が信頼できる、人間中心で、基本的権利に沿ったものとなるよう、公的・民間の両方の組織を導くことを目的としています。

技術的専門知識のためのリソース:

報告枠組みは、G7諸国の戦略的リーダーシップとOECDの技術・政策の専門知識を結集した国際的な協力によって開発されました。この取組の目的は、組織が国際行動規範に準拠するための実践的なツールを提供することです。価値あるピアツーピアの学びや知識共有を促進するため、フレームワークでは提出された報告書をオンラインで公開しています。この仕組みにより、ユーザーは現場で実際にどのような質問がされているかを確認できるだけでなく、他の組織の経験やリスク軽減策から直接学ぶことができます。公開される報告書は、さまざまな業界や回答者による多様な事例が増え続けています。さらに、中小企業にとって重要なのは、OECDが現在、中小企業がHAIP行動規範を効果的に適用できるよう支援するための追加の取組やリソースを開発している点です。

OECD AIポリシーナビゲーターは、世界各国のAIに関する政策や規制を追跡するための中央リポジトリとして機能します。急速に変化する規制環境を踏まえ、このナビゲーターは、各国政府がAIリスクにどのように対応しているかについて重要な洞察を提供します。

専門家や公式の寄稿者によって定期的に更新されるこのナビゲーターは、国別のフィルタリング機能を備え、国際的なAI政策の最新情報を含んでいます。これにより、ユーザーは自分の管轄に関連する情報をカスタマイズでき、政策の変化を予測し、組織の取り組みを新たな国際基準に合わせるための信頼できる参照ポイントとして活用できます。

EU AI法第62条は、AIの提供や導入に関わる中小企業、スタートアップに対して、EU加盟国が特定の支援措置を提供することを求めています。これらの措置には、研修活動、専門家による

助言、そして標準化開発プロセスへの中小企業の参加を促進する取り組みが含まれます。施行日は2026年8月に設定されており、この義務により、信頼できるAIの導入を支援するための取組が今後開始されることになります。

日本の総務省と経済産業省が策定した「AI事業者ガイドライン」は、AIを導入する中小企業にとって有用なリソースであり、信頼できるAIガバナンスの複雑さを乗り越えるための実践的で拘束力のない指針を提供します。この文書は、HAIPなどの国際的な議論と整合しながら、AIリスクの理解を助けます。リスクベースのアプローチを重視し、初期のデータ検討から継続的な運用に至るまで、AIライフサイクル全体でリスクを積極的に特定・管理することを推奨しています。ガイドラインでは、人間中心の設計、安全性、公平性、プライバシー、セキュリティ、透明性、説明責任といった主要原則を整理し、AIビジネス利用者向けに具体的な助言を提供します。これにより、急速に進化するAI環境に適応しながら、自信を持って導入し、利益を最大化し、潜在的なリスクを最小化することを支援します。

結論

G7加盟国の共同の取組は、OECDとの緊密な協力を通じて、中小企業が組織内で信頼できるAIを導入できるよう支援する強いコミットメントを示しています。さまざまなイニシアチブを通じて、小規模企業がAIの複雑さを責任ある方法で乗り越えるためのリソースを提供する取組が進められています。こうした継続的な努力により、中小企業が今後さらに積極的に関与し、将来のAIフレームワークの開発プロセスにも参加することが期待されます。その結果、中小企業の独自の視点やニーズが十分に反映され、より包括的で効果的なグローバルAIエコシステムの実現につながるでしょう。



付属書 - 高度なAIシステムを開発する組織向けの広島 プロセス国際指針

1. AI ライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度な AI システムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる

これには、レッドチーム等の様々な手法を組み合わせて、多様な内部テスト手段や独立した外部テスト手段を採用することや、特定されたリスクや脆弱性に対処するための適切な緩和策を実施することが含まれる。テストと緩和策は、例えば、システムが不合理なリスクをもたらさないように、ライフサイクル全体を通じてシステムの信頼性、安全性、セキュリティの確保を目指すべきである。このようなテストを支援するために、開発者は、データセット、プロセス、システム開発中に行われた意思決定に関連して、トレーサビリティを可能にするよう努めるべきである。

2. 市場投入を含む導入後、脆弱性、及び必要に応じて悪用されたインシデントやパターンを特定し、緩和する

組織は、リスクレベルに見合った適切なタイミングで、AI システムを意図したとおりに使用し、導入後の脆弱性、インシデント、新たなリスク、悪用を監視し、それらに対処するための適切な措置を講じるべきである。組織は、例えば、導入後に第三者やユーザーが問題や脆弱性を発見し報告することを促進することの検討が奨励される。組織はさらに、他の利害関係者と協力して、報告されたインシデントの適切な文書化を維持し、特定されたリスクと脆弱性を軽減することが奨励される。適切な場合には、脆弱性を報告する仕組みは、多様な利害関係者が利用できるものでなければならない。

3. 高度な AI システムの能力、限界、適切・不適切な使用領域を公表し、十分な透明性の確保を支援することで、アカウンタビリティの向上に貢献する

これには、高度なAIシステムの重要な新規公表全てについて、有意義な情報を含む透明性報告書を公表することが含まれるべきである。組織は、適かつ関連性のある導入者及び利用者がモデル／システムのアウトプットを解釈し、利用者がそれを適切に利用できるようにするために、透明性報告書内の情報を十分に明確で理解可能なものにすべきである。また、透明性報告書は、強固な文書化プロセスによってサポートされ、提供されるべきである。

4. 産業界、政府、市民社会、学界を含む、高度な AI システムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組む

これには、評価報告書、セキュリティや安全性のリスク、危険な意図的又は意図しない能力、AIのライフサイクル全体にわたるセーフガードを回避しようとするAI関係者の試みに関する情報等を含むが、これらに限定されない、適切な情報の責任ある共有が含まれる。

5. 特に高度なAIシステム開発者に向けた、個人情報保護方針及び緩和策を含む、リスクベースのアプローチに基づくAIガバナンス及びリスク管理方針を策定し、実施し、開示する

これには、個人データ、ユーザーープロンプト、高度な AI システムのアウトプットを含め、適切な場合にはプライバシーポリシーを開示することが含まれる。組織は、リスクベースのアプローチに従って、AIガバナンス方針とこれらの方針を実践するための組織的メカニズムを確立し、開示することが期待される。これには、AI のライフサイクルを通じて実行可能な場合には、リスクを評価し、軽減するための説明責任とガバナンス・プロセスが含まれるべきである。

6. AIのライフサイクル全体にわたり、物理的セキュリティ、サイバーセキュリティ、内部脅威に対する安全対策を含む、強固なセキュリティ管理に投資し、実施する

これには、情報セキュリティのための運用上のセキュリティ対策や、適切なサイバー／物理的アクセス制御等を通じて、モデルの重み、アルゴリズム、サーバー、データセットを保護することが含まれる。

7. 技術的に可能な場合は、電子透かしやその他の技術等、ユーザーが AI が生成したコンテンツを識別できるようにするための、信頼できるコンテンツ認証及び来歴のメカニズムを開発し、導入する

これには、適かつ技術的に実現可能な場合、組織の高度な AI システムで作成されたコンテンツのコンテンツ認証及び来歴メカニズムが含まれる。

来歴データには、コンテンツを作成したサービス又はモデルの識別子を含めるべきであるが、ユーザー情報を含める必要はない。組織はまた、透かし等を通じて、特定のコンテンツが高度なAIシステムで作成されたかどうかをユーザーが判断できるツールやAPIの開発に努めるべきである。

組織はさらに、可能かつ適切な場合には、利用者がAIシステムと相互作用していることを知ることができるよう、ラベリングや免責事項の表示等、その他の仕組みを導入することが奨励される。

8. 社会的、安全、セキュリティ上のリスクを軽減するための研究を優先し、効果的な軽減策への投資を優先する

これには、AIの安全性、セキュリティ、信頼性の向上を支援し、主要なリスクに対処する研究の実施、協力、投資及び適切な緩和ツールの開発への投資が含まれる。

9. 世界の最大の課題、特に気候危機、世界保健、教育等（ただしこれらに限定されない）に対処するため、高度なAIシステムの開発を優先する

これらの取り組みは、国連の持続可能な開発目標の進捗を支援し、グローバルな利益のためAIの開発を奨励するために行われる。

組織は、信頼できる人間中心のAIの責任あるスチュワードシップを優先し、また、デジタルリテラシーのイニシアティブを支援すべきである。

10. 國際的な技術規格の開発を推進し、適切な場合にはその採用を推進する

これには、電子透かしを含む国際的な技術標準とベストプラクティスの開発に貢献し、適切な場合にはそれを利用し、標準開発組織（SDO）と協力することが含まれる。

11. 適切なデータインプット対策を実施し、個人データ及び知的財産を保護する

組織は、有害な偏見バイアスを軽減するために、訓練データやデータ収集など、データの質を管理するための適切な措置を講じることが奨励される。

訓練用データセットの適切な透明性も支援されるべきであり、組織は適用される法的枠組みを遵守すべきである。