【資料4】

AI事業者ガイドライン更新に向けた論点

2025年12月2日

AIガバナンス検討会　事務局

AI事業者ガイドラインの今後の更新に向けた論点

意見照会（10/15～10/31）にて収集した意見内容を分類

1. AI技術の進展への対応

－AIエージェント/Agentic AI（マルチAIエージェント）に関する追記

－フィジカルAIに関する追記

等

2. リスクの記載内容の見直し

－リスクの分類の見直し、リスクの評価についての追記

－既に記載しているリスクについての追記・修正

3. AI事業者ガイドラインの利活用の推進策の検討

－AI事業者ガイドラインの利活用における課題の整理及び解決策の検討

4. 事例の追加/更新

－事例の収集、追加（AI利用者の視点、AIエージェント活用事例等）

－既に記載している事例の更新

5. その他

－政策動向（AI法や広島AIプロセス等）との整合

－AIの利用や便益の広がりに関する記載の追加（業界別の利用例等）

● 論点①　AI技術の進展に伴う記載の追加

- 以下についてのユースケース・便益・リスク・対策について
  ーAIエージェント（及びエージェンティックAI）
  ーフィジカルAI

● 論点②　リスクの記載の見直し・追加

- 事業者によるリスクベースアプローチに資する記載内容について
  ーリスクの評価の追記
  等

● 論点③　AI事業者ガイドラインの利活用の推進策

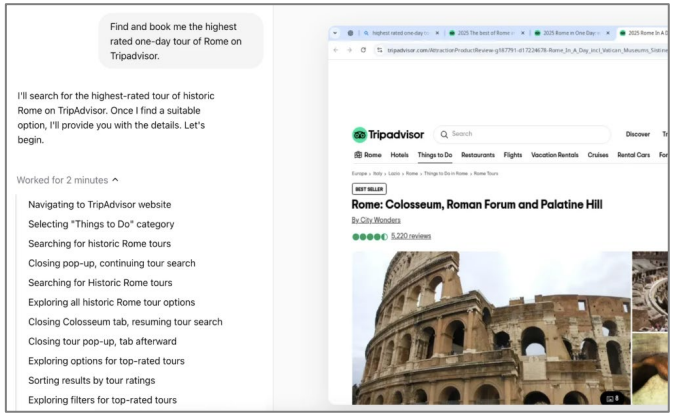
- 事業者によるAI事業者ガイドラインの利活用における課題
- AI事業者ガイドラインの利活用推進策・具体案

論点①
AI技術の進展に伴う記載の追加

主に以下についてのユースケース・便益・リスク・対策について
- ーAIエージェント（及びエージェンティックAI）
- ーフィジカルAI

AIエージェントのサービス事例

■WEB・アプリ操作エージェント（OpenAI「Operator」[*1]）

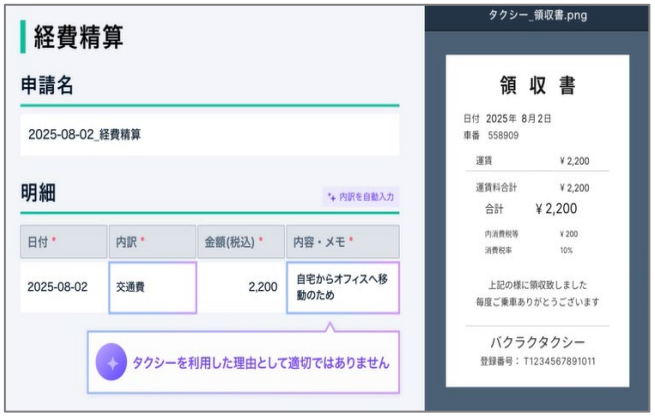


ユーザーの指示を解釈し、目的達成のため自律的にユーザーの代わりに**WEBサイトやアプリの操作**を行うサービスであり、予約やフォーム入力などの**繰り返し作業の効率化**が期待される

■広告企画特化型マルチエージェント（博報堂テクノロジーズ「Nomatica」[*2]）

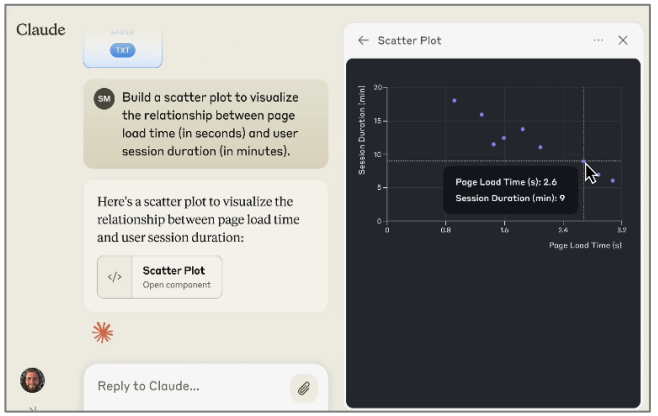
複数の専門エージェントが自律的に議論しながら**企画を立案・評価**するマルチエージェントプラットフォームです。ユーザーの意図に応じて**最適なエージェントを選び**、調査や戦略立案、資料作成まで支援します。これにより、**企画業務の効率化と質の向上**が期待されます。

■経理AIエージェント（Bakuraku[*3]）



経費や稟議の**申請内容を自動でチェック**し、申請者にリアルタイムでフィードバックを行うサービスであり、**AIが社内ルールや過去の申請データを理解して自律的に内容確認**を行うことで、申請ミスや差し戻しを減らし、**業務効率を向上**させることが期待される

■コード生成エージェント（Claude Code[*4]）



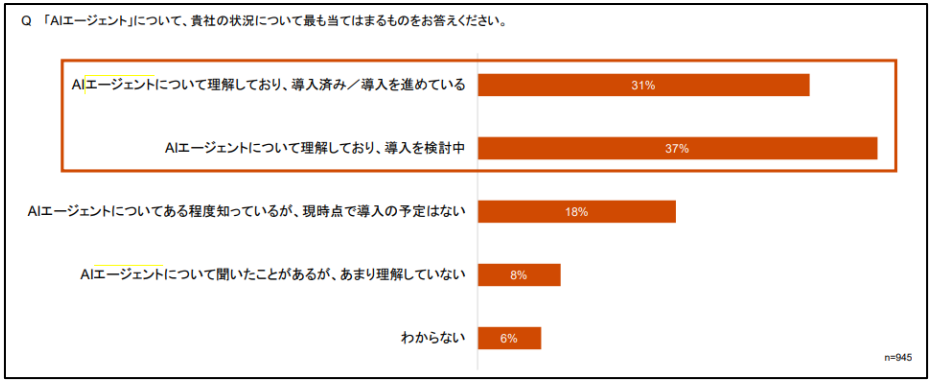
プロジェクト全体のコード構造やファイル依存関係を理解し、開発者の指示に基づいて自律的に**コードを生成・修正**したり、ツールや**APIを活用**して作業を自動化する開発支援ツールであり、**開発速度の向上とコード品質の改善**が期待される

*1: https://openai.com/ja-JP/index/introducing-operator/
*2: https://www.nomatica.hakuhodo-technologies.co.jp/
*3:https://bakuraku.jp/ai/
*4:https://code.claude.com/docs/ja/overview

AIエージェントの動向

事業者の動向

約70%の事業者が導入中、もしくは導入検討中



※PwC社「生成AIに関する 実態調査2025 春 5カ国比較」
https://www.pwc.com/jp/ja/knowledge/thoughtleadership/2025/assets/pdf/generative-ai-survey2025.pdf

政府の動向

人工知能基本計画骨子にて利活用の推進について言及

（２） 社会課題解決に向けたＡＩ利活用の推進

① 医療・ヘルスケア【内閣府（科技）、厚労省、経産省】、介護【厚労省、経産省】、金融【金融庁】、教育【文科省】、防災・消防【内閣府（防災）、総務省、文科省、国交省】、環境保全【文科省、環境省】、農林水産業【農水省】、食品産業【農水省】、インフラ建設・管理【国交省】、造船・舶用工業【国交省】、公共交通【国交省】等におけるＡＩエージェントやフィジカルＡＩ等の開発・実証・導入促進【◎内閣府（ＡＩ室）、関係省庁】

※内閣府　人工知能基本計画骨子　P5～P6
https://www8.cao.go.jp/cstp/ai/ai_plan/aiplan2025_draft3.pdf

構成員からの意見

AI技術の進展に伴う更新として、AIエージェントについての記載の追加について多くのご意見をいただいた。（以下、一部抜粋）

・AIエージェントなど、高度に進化し自律性を有したAIによるリスクが漏れているように感じます。
・AIエージェントに関する記載を脚注等で追加すべき
・例えば、（生成AIだけに直結する話ではないですが）昨今「AIエージェント」が重視されている（もてはやされている）一方、本ガイドラインには当該ワードが一切出てきません。
・AIエージェントのリスクを追加すべきではないか。具体的な例として、自律的判断による非倫理的な行動、プライバシー侵害、評価・制御不能な判断、責任の所在の不明確さが挙げられる。
・AIエージェント、フィジカルAIについては多数の意見があり，私も絶対重要だと思います．とくに自律的AIエージェントは，利用者本人が知らないうちに購買や予約をするケースもあります．（とくに，利用者が多数のAIエージェントを同時に使う場合は顕著）
・AIエージェントについては、次回更新時に少なくとも用語の定義はした方が良いと考えます。
・AIエージェントという新技術の投入により列挙されているリスク（特に社会的リスク）にどのような影響が起きるのか、さらには新たな対処が追加的に要求されるのかについては整理の必要があると思いました。
・AIエージェント、フィジカルAIへの対応はぜひ積極的に取り組んでいただきたいところ

AIエージェントに関する現ガイドラインの記載

AIエージェントに関して、AI事業者ガイドライン（1.1版）においては用語の定義はされておらず、便益やリスクについての記載も限定的

本編P10 「関連する用語」	AIエージェントに関する定義は含まれていない
別添P12 「AIによる便益」	AIエージェントに関する便益は「生産性の向上」のみに言及
別添P13 「AIによるリスク」	脚注にて、AIエージェントの一部のリスクについて複雑化・深刻化する可能性について言及しているのみ

関連する用語

- **AI**
  現時点で確立された定義はなく（統合イノベーション戦略推進会議決定「人間中心の AI 社会原則」（2019 年 3 月 29 日））、広義の人工知能の外延を厳密に定義することは困難である。本ガイドラインにおける AI は「AI システム（以下に定義）」自体又は機械学習をするソフトウェア若しくはプログラムを含む抽象的な概念とする。

生成 AI による可能性

上記に加え、直近では生成 AI が台頭している。生成 AI は DX への遅れをとった日本企業の巻き返しの引き金となる可能性も高い。

加えて、自律的な AI システム（以下、AI エージェント）も登場している。従来型の AI や生成 AI に比べより高度な効率化や自動化が可能となることで、生産性の向上につながること等が期待されている。

グローバルの激しい競争を勝ち抜くためにも、生成 AI を積極的に取り入れる形でデジタル戦略の見直しを行う等、自身が享受できる便益を正しく理解し、可能性を模索するとともに、積極的な姿勢を持つことが期待される。

このように、技術発展により AI 活用による便益が大きくなる一方で、従来型 AI でも現れていたリスクが生成 AI の台頭によりさらに増大傾向にある。また、生成 AI により新たに顕在化したリスクもある。加えて、多くの生成 AI サービスで利用障壁が下がったことから、意図しないリスクを伴う使われ方をする恐れもある [16]。

脚注

[16] AI エージェントの登場により、事故等の安全性面のリスクや過度な依存、労働者の失業等のリスクが複雑化・深刻化する可能性があることにも留意する必要がある。

AI事業者ガイドラインの次期更新において、
AIエージェントの定義・便益・リスク・対策の追記を検討する

AIエージェントの定義と便益（案）

AIエージェントの定義と便益については、以下をAI事業者ガイドラインに記載してはどうか

AIエージェントの定義

AIエージェントとは、特定の目標を達成するために、環境を感知し自律的に行動するAIシステム

（参考）一部事業者におけるAIエージェントの定義

#	事業者	定義
1	Microsoft	AI エージェントとは、**状況を観察し、データを解釈し、特定の目標に向かって行動するシステム** のことである。**反復的な作業を減らし、正確性を高め、より迅速な意思決定を導く**ことで、人を支援するように設計されている （https://www.microsoft.com/en-us/microsoft-copilot/copilot-101/how-do-ai-agents-work）
2	Google	AI エージェントは、**AI を使用してユーザーの代わりに目標を追求し、タスクを完了させるソフトウェア システム**である。**推論、計画、メモリーが可能であることが示されており、意思決定、学習、適応を行うレベル**の自律性を備えている （https://cloud.google.com/discover/what-are-ai-agents?hl=ja）
3	Amazon	人工知能 (AI) エージェントは、**環境と対話し、データを収集し、そのデータを使用して自己決定タスクを実行して、事前に決められた目標を達成する**ためのソフトウェアプログラムである。目標は人間が設定しますが、その目標を達成するために実行する必要がある最適なアクションは AI エージェントが独自に選択する (https://aws.amazon.com/jp/what-is/ai-agents/)
4	IBM	AIエージェントとは、**ワークフローを設計し、利用可能なツールを活用することで、ユーザーまたは別のシステムに代わってタスクを自律的に実行**できるシステムまたはプログラムである。 AI（人工知能）エージェントは、**意思決定、問題解決、外部環境とのやり取り、アクションの実行など、自然言語処理以外の幅広い機能を備える**ことができる（https://www.ibm.com/jp-ja/think/topics/ai-agents）
5	Accenture	AIエージェントは、**大規模言語モデル（LLMs）を利用して問題を推論し、解決策を計画し、計画を実行する自律的なAIプログラム**である。**過去のユーザーとのやり取りの「記憶」と一連のツールを活用して特定の目標を達成**する。AIエージェントは人間の意図をすばやく把握し、複雑なタスクを自動化するための事前に構築されたワークフローを提示し、パーソナライズされた支援を提供し、人間とコンピュータの相互作用を向上させる （https://www.accenture.com/jp-ja/insights/data-ai/hive-mind-harnessing-power-ai-agents）

AIエージェントの便益

ユーザーの意図を理解し自律的にタスクを遂行することで、複雑な業務プロセスを効率化し、人的負荷を大幅に削減できる

*事務局にて作成。

AIエージェントのリスク（案）

□：従来AI/生成AIにも共通するリスク
■：AIエージェント特有と考えられるリスク

技術的特徴	リスク（案）	インシデント例（AIエージェント以外の従来技術の事例も含む）
ユーザーの**指示を解釈**し、**目的を定め、達成計画**を立てる	**悪意ある入力で誤作動** – 不正な指示にて本来と異なる行動を取る	チャットボットに多くのユーザーが差別的内容を投稿し、その内容に基づいた出力を行うようになった
状況に応じて**最適な行動を選択・実行**し、**ポリシーに基づく判断と環境変化への柔軟な対応**を行う	**制約回避した不正行動** – 人間の意図しない方法で制約を破る	価格交渉をするチャットボットが、人間には意味不明な独自言語を用いて会話を行った
	判断根拠が不明瞭 – 非決定的な判断で根拠の追跡が困難	ある医療AIが誤った治療提案を複数行い、その判断過程がブラックボックスだったため、医療現場で大きな不信を招いた
長期・短期メモリで経験や情報を蓄積し、**将来の判断に活用**する。**継続学習によって性能や方針を進化**させる	**誤情報の記憶汚染** – 間違った情報を記憶し、将来の判断に悪影響	※「悪意ある入力で誤作動」のインシデント例と同様
	誤情報の拡散 – 間違いを繰り返し学習・出力して広める	ある科学系論文QA用AIが存在しない論文やデタラメな回答をもっともらしく生成し、公開直後に提供停止となった
外部ツールやAPIの呼び出し、コードの生成・実行を通じて、環境へ具体的に作用する	**ツールの悪用** – 許可された範囲のツールで意図しない操作を実行	子どもの遊びの声かけにより音声アシスタントが高額商品を注文、さらに報道を通じて同様の誤注文が広がった
	コードの悪用 – 生成したコードが不正操作や侵入に利用される	生成AIにより、不正契約を自動で結ぶプログラムが作成された
認証・認可機能を持ち、**ID管理や権限委譲を通じて、利用者・エージェントごとのアクセス制御**を実現し、他システムとの情報交換の基盤となる	**権限の乗っ取り** – 他のシステムから権限を奪い高い権限を取得	業務自動化AIで、低権限エージェントが上位権限の操作を不正に実行できる脆弱性が発覚した
	なりすまし操作 – 他のAIを装い不正行為を行う	自律型AIが、他のAIに偽の指示を与えて安全規範を破らせようとする行動を試みた
	偽情報の混入 – 通信に虚偽情報を加え協調行動を妨げる	※「なりすまし操作」のインシデント例と同様
設計した**UI/UX**を通じてユーザーと対話・操作し、意図や感情を読み取って**適切な応答や提案**を返す。	**人間の過信誘導** – AIを過信させて有害な行動に導く	欧州で男性が対話型AIとのやり取りに依存した結果、AIの指示により自ら命を絶った

AIエージェントの主体ごとの対策（案）

リスク（案）	対策（方針）	主体区分毎の対策（案）		
		AI開発者	AI提供者	AI利用者
悪意ある入力で誤作動 – 不正な指示にて本来と異なる行動を取る	権限管理と不正利用防止	最小権限設計の徹底や、ツール呼び出し時の安全性検証等	APIキーや認証情報の管理強化や、アクセス制御	操作履歴の定期確認
制約回避した不正行動 – 人間の意図しない方法で制約を破る				
判断根拠が不明瞭 – 非決定的な判断で根拠の追跡が困難				
誤情報の記憶汚染 – 間違った情報を記憶し、将来の判断に悪影響	自律行動の制御	ガードレール設計	HITL（ヒューマンインザループ）により、高リスク操作は人間承認を必須化	
誤情報の拡散 – 間違いを繰り返し学習・出力して広める				
ツールの悪用 – 許可された範囲のツールで意図しない操作を実行	メモリの健全性確保	メモリ更新データの制限や信頼性の評価	メモリ管理ポリシー策定（保存期間・削除ルール等）	誤情報を発見した際の即時報告
コードの悪用 – 生成したコードが不正操作や侵入に利用される				
権限の乗っ取り – 他のシステムから権限を奪い高い権限を取得				
なりすまし操作 – 他のAIを装い不正行為を行う				
偽情報の混入 – 通信に虚偽情報を加え協調行動を妨げる	通信の安全性確保	エージェント間通信の暗号化	通信ログの監査と異常検知	外部接続の必要最小限化や、エージェント連携の制限
悪意あるAIの侵入 – マルチエージェント環境全体の安全性を低下				
人間の過信誘導 – AIを過信させて有害な行動に導く	AI過信・要求過多の防止	出力に不確実性指標を付与（信頼度評価）	AI利用者向け教育（AIの限界・リスク）	大量要求を避け、優先度を明確化

※構成員のご意見を踏まえ、事務局にて作成。
※リスクが顕在化し事故が発生した際の民事責任の在り方については、経済産業省主催の「AI利活用における民事責任の在り方に関する研究会」の検討結果や今後の動向をAI事業者ガイドラインに反映していく想定。
※リスクと対策については、業種共通の事項について追加を検討する。

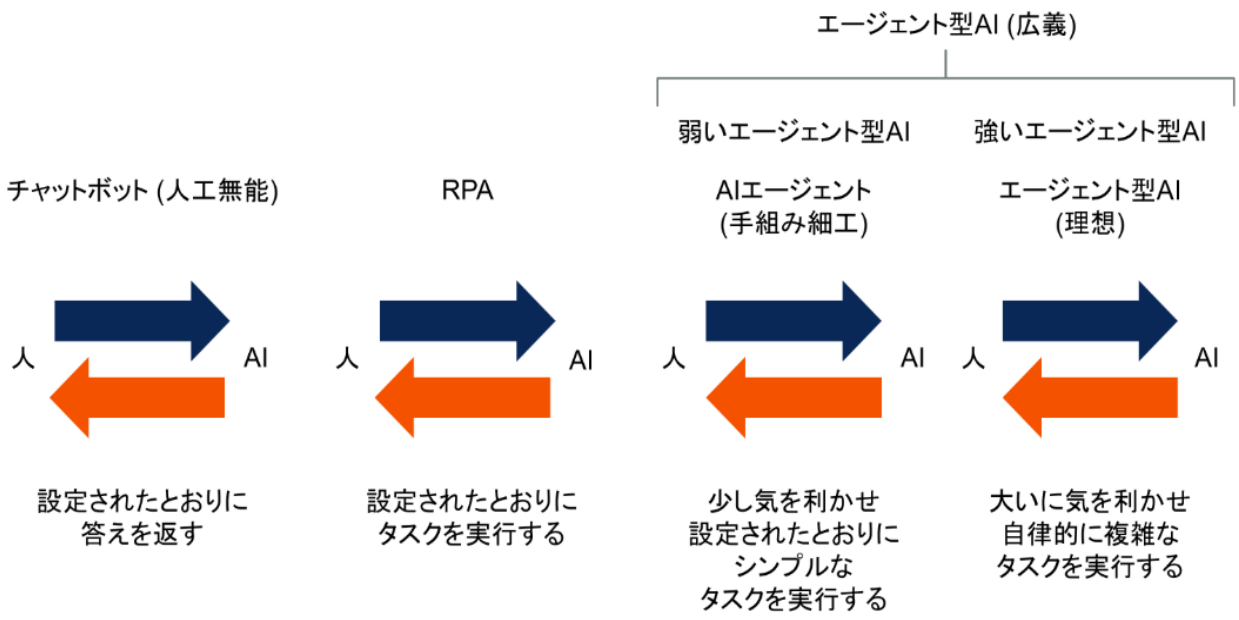
エージェンティックAIの説明（案）

エージェンティックAIの定義については、AIエージェントとの違いにおいて「目的に対する自律性」「マルチ（複数のAIエージェントで動作）」の観点から、以下のように説明してはどうか

エージェンティックAIの説明	AIエージェントよりも包括的かつ進化的な概念であり、複数のAIエージェントにより自律的に意思決定を下しアクションを起こす目標主導型のAIシステム

AIエージェントとエージェンティックAIの違い　①目的に対する自律度に応じた整理

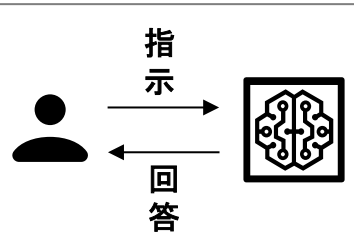
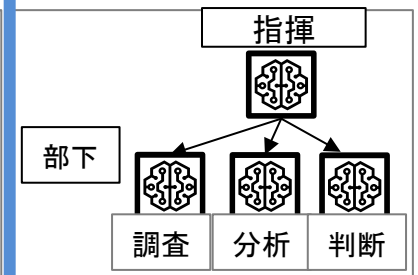
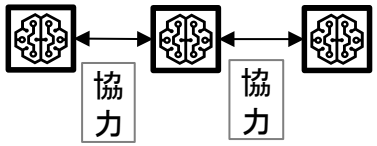
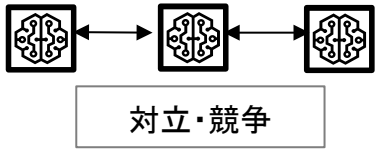
図1. チャットボット、RPA、AIエージェント、エージェント型AIの違い



*1: Gartner – AIエージェントとエージェント型AIに関する見解を発表　[https://www.gartner.co.jp/ja/newsroom/press-releases/pr-20250514-ai-agent]

エージェンティックAIの説明（案）

AIエージェントとエージェンティックAIの違い　②エージェントの構成（シングル / マルチ）

エージェント形態	シングルエージェント	マルチエージェント		
		階層型：指揮と部下	横並び：協働型	横並び：競争型
定義	• 単一のエージェントで動作	• 複数のエージェントで動作 • 指揮エージェントが部下のエージェントをコントロール	• 複数のエージェントで動作 • エージェント同士で協調	• 複数のエージェントで動作 • エージェント同士で競争する
利用ケース	シンプルで実装しやすく、特定のタスクであっても、状況に合わせた自律的判断が可能。判断を必要とするが複雑な意思決定が不要なケースで利用	明確な指揮構造で一貫性と全体目標との整合を維持する。部下エージェント追加でスケールが容易。企業管理など集中管理が必要なケースで利用	各々エージェントがリアルタイムの情報にもとづいて協力するよう行動する。共同作業リソース共有が必要なケースに利用	複雑な立場から議論し多様な戦略で競争することで革新と最適化を促進する。研究開発など創造的かつステークホルダの意図が複雑なケースで利用
イメージ				

AIエージェント

エージェンティックAI

*「AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges」等の各種公開情報を基に、事務局にて作成

エージェンティックAIのリスク（案）

エージェンティックAIのリスクについてはAI利活用ガイドライン「6. AI 利活用原則の解説」の「③ 連携の原則」を参考としてはどうか

③ 連携の原則

AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、AI システム又は AI サービス相互間の連携に留意する。また、利用者は、AI システムがネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。

［ア 相互接続性と相互運用性への留意］

- AI サービスプロバイダは、利用する AI の特性及び用途を踏まえ、AI ネットワーク化の健全な進展を通じて、AI の便益を増進するため、AI の相互接続性と相互運用性に留意することが期待される。

［ウ AI ネットワーク化により惹起・増幅される課題への留意］

- AI が連携することによって便益が増進することが期待されるが、AI サービスプロバイダ及びビジネス利用者は、自ら利用する AI がインターネット等を通じて他の AI 等と接続・連携することにより制御不能となる等、AI がネットワーク化することによって リスクが惹起・増幅される可能性 があることに留意することが期待される。このため、開発者等からの情報を踏まえ、考えられるリスクを分析し、当該リスクを連携の相手方と共有するとともに、予防策や問題が生じた場合の対応策等を整理し、消費者的利用者等に対し、必要な情報提供を行うことが期待される。

■AI利活用ガイドライン
https://www.soumu.go.jp/main_content/000624438.pdf

AIエージェント/エージェンティックAI　AI事業者ガイドライン（1.1版）への追加イメージ

本編　https://www.soumu.go.jp/main_content/001002576.pdf

P9～　関連する用語　※定義の追加

・**AIエージェント**
AIエージェントとは、特定の目標を達成するために、環境を感知し自律的に行動するAIシステム

（脚注）
*エージェンティックAIとは、AIエージェントよりも包括的かつ進化的な概念であり、複数のAIエージェントにより自律的に意思決定を下しアクションを起こす目標主導型のAIシステム

P30～　第3部 AI 開発者に関する事項　※対策の追加

AI開発者に対するAIエージェントに関する留意事項を追加

P35～　第4部 AI 提供者に関する事項　※対策の追加

AI提供者に対するAIエージェントに関する留意事項を追加

P38～　第5部 AI 利用者に関する事項　※対策の追加

AI利用者に対するAIエージェントに関する留意事項を追加

別添　https://www.soumu.go.jp/main_content/001000988.pdf

P12～　AIによる便益　※便益の追加

AIエージェントにより、ユーザーの意図を理解し自律的にタスクを遂行することで、複雑な業務プロセスを効率化し、人的負荷を大幅に削減できる

P13～　AIによるリスク　※リスクの追加

・AIエージェントにより増幅する既存のリスク
・AIエージェントにより新たに顕在化するリスク
・エージェンティックAIに関するリスク

P84～　別添3 AI 開発者向け　※対策の追加

AI開発者に対するAIエージェントに関する留意事項を追加

P125～　別添4 AI 提供者向け　※対策の追加

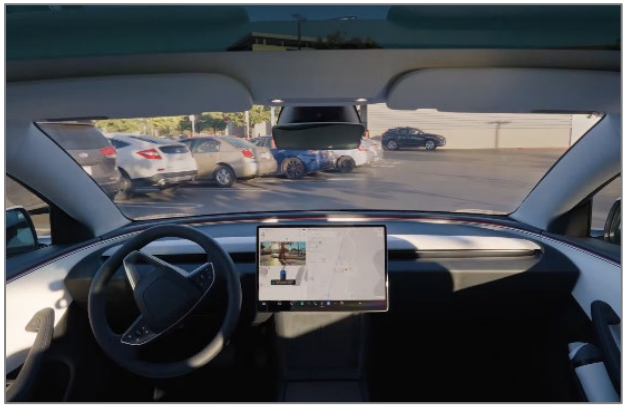
AI提供者に対するAIエージェントに関する留意事項を追加

P153～　別添5 AI 利用者向け　※対策の追加

AI利用者に対するAIエージェントに関する留意事項を追加

フィジカルAIのサービス事例について

■自動運転システム
（TESLA[*1]）



車両周囲を**センサーで認識**しAIが**走行判断や操作を自動化**する技術であり、**高精度センシング統合**や**リアルタイム推論**、**物理世界の予測能力**が生かされることで複雑な環境でも安定した運転判断を実現し、結果として**事故削減や渋滞緩和**、**移動効率の向上等**が期対される

■巡回警備ロボット
（SECOM「cocobo」[*2]）

カメラ・センサーから得られた情報を基に、AIで**映像解析や行動認識**を行い、**施設内の異常を自律的に検知・通報**するロボットである。これにより、警備員の**負荷軽減**や**人手不足の解消**、**警備品質の向上**等が期待される

■清掃ロボット
（アイリスオーヤマ「DX清掃ロボットジルビー」[*3]）

床清掃を行う法人向けロボットであり、LiDARや3Dカメラなど**多重センサーで環境を認識**する他、**清掃ルート最適化・運用提案・状況学習**といった**自律的な判断**も行う。これにより、清掃**作業の効率化**や**コスト削減等**が期待される

■自律型ロボットアーム
（安川電機「MOTOMAN NEXT」[*4]）

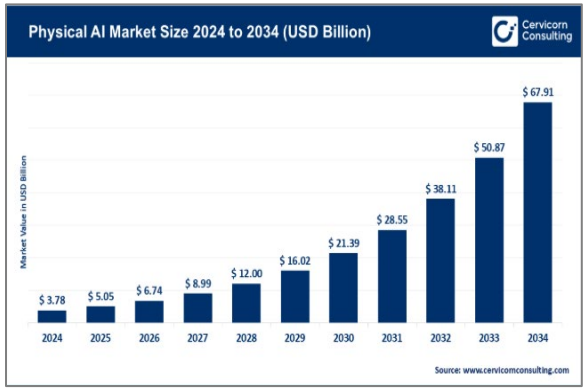
製造業をはじめとしたあらゆる領域での業務において、センサーやカメラから得られた情報から、**自ら判断・計画する自律適応型**ロボットである。これまで自動化が難しかった**変種・変量生産**や**不確定環境下の作業の効率化**や**省人化**等が期待される

*1: https://www.tesla.com/ja_jp/fsd
*2: https://www.secom.co.jp/business/cocobo/#v7YTPlayerModal
*3: https://www.irisohyama.co.jp/b2b/robotics/products/jilby/
*4: https://www.e-mechatronics.com/product/robot/special/motoman-next/

フィジカルAIの動向

市場の動向

2025年以降、急速に伸びが予測（CAGR　33.5%）



*Cervicorn Consulting [https://www.cervicornconsulting.com/physical-ai-market]

政府の動向

人工知能基本計画骨子にて利活用の推進について言及

（２）　社会課題解決に向けたＡＩ利活用の推進

① 医療・ヘルスケア【内閣府（科技）、厚労省、経産省】、介護【厚労省、経産省】、金融【金融庁】、教育【文科省】、防災・消防【内閣府（防災）、総務省、文科省、国交省】、環境保全【文科省、環境省】、農林水産業【農水省】、食品産業【農水省】、インフラ建設・管理【国交省】、造船・舶用工業【国交省】、公共交通【国交省】等におけるＡＩエージェントやフィジカルＡＩ等の開発・実証・導入促進【◎内閣府（ＡＩ室）、関係省庁】

※内閣府　人工知能基本計画骨子　P5～P6
https://www8.cao.go.jp/cstp/ai/ai_plan/aiplan2025_draft3.pdf

構成員からの意見

AI技術の進展に伴う更新として、フィジカルAIについての記載の追加について多くのご意見をいただいた。（以下、一部抜粋）

・フィジカルAIやAGIに対して「安全実証」「説明可能性」の透明化を義務づける枠組みの議論、検討を行うべきと考える。
・AIエージェント、フィジカルAIについては多数の意見があり，私も絶対重要だと思います．とくに自律的AIエージェントは，利用者本人が知らないうちに購買や予約をするケースもあります．（とくに，利用者が多数のAIエージェントを同時に使う場合は顕著）
・AIの急速な進化、特にAIエージェントやフィジカルAIといった技術の普及に伴う新たなリスクへの対応は喫緊の課題です。
・AIエージェント、フィジカルAIへの対応はぜひ積極的に取り組んでいただきたいところ
・便益の例の横軸への追加項目として、フィジカルAIに関連する、自動運転、医療・介護、建設・インフラを追加することを検討できるのが良いと考えています。
・Agentic AIやフィジカルAIに関してのリスクを追加

フィジカルAIに関する現ガイドラインの記載

フィジカルAIに関して、AI事業者ガイドライン（1.1版）においては用語の定義、便益、リスクや対策の記載はされていない

本編P10 「関連する用語」	フィジカルAIに関する定義は含まれていない	**関連する用語** • **AI** 現時点で確立された定義はなく（統合イノベーション戦略推進会議決定「人間中心の AI 社会原則」（2019 年 3 月 29 日））、広義の人工知能の外延を厳密に定義することは困難である。本ガイドラインにおける AI は「AI システム（以下に定義）」自体又は機械学習をするソフトウェア若しくはプログラムを含む抽象的な概念とする。
別添P12 「AIによる便益」	フィジカルAIに関する便益は含まれていない	生成 AI による可能性 上記に加え、直近では生成 AI が台頭している。生成 AI は DX への遅れをとった日本企業の巻き返しの引き金となる可能性も高い。
別添P13 「AIによるリスク」	フィジカルAIに関するリスクは含まれていない	このように、技術発展により AI 活用による便益が大きくなる一方で、従来型 AI でも現れていたリスクが生成 AI の台頭によりさらに増大傾向にある。また、生成 AI により新たに顕在化したリスクもある。加えて、多くの生成 AI サービスで利用障壁が下がったことから、意図しないリスクを伴う使われ方をする恐れもある [16]。

AI事業者ガイドラインの次期更新において、
フィジカルAIの定義・便益・リスク・対策の追記を検討する

フィジカルAIの定義と便益（案）

フィジカルAIの定義については、特徴（ハードウェアとAIの統合）を元に、以下のように定義してはどうか

フィジカルAIの定義	フィジカルAI（Physical AI）とは、ソフトウェア的知能（AIアルゴリズム）とハードウェア的実体（センサー、アクチュエータ、エッジデバイス等）を統合し、物理世界における知的認識・判断・行動を自律的に実現するAIシステムである。

フィジカルAIの特徴	説明
センサーによる環境認識	フィジカルAIは、カメラ、マイク、温度計、LiDAR、加速度計などの多様なセンサーを通じて外部環境を知覚し、情報を取得する。これにより、視覚、音、温度、距離、動きといった物理現象をデジタルデータとして取り込み、周囲の環境を把握することができる。
データ処理・推論	取得した感覚情報は、機械学習やディープラーニングなどのAIアルゴリズムによって処理される。物体認識、音声理解、空間マッピング、行動予測といった処理を通じて、AIは周囲の状況を構造的に理解し、意思決定のための情報として利用する。
環境変化に対応したリアルタイム意思決定	環境の変化に即応して、状況に応じた判断と行動の選択を行う機能である。センサーデータに基づいて常に状況を更新し、優先順位や緊急度に応じた判断を下す。これにより、予期せぬ事態や複雑な状況に対しても適切に対応できる。
アクション実行	AIが下した判断を物理的な動作として実行する機能である。アクチュエータを用いて、移動、物体の操作、発話、表情の再現などの行動を行い、AIの内部的な推論を現実世界に反映させる。これはフィジカルAIが現実に干渉するための中核となる。
学習・適応	フィジカルAIは、過去の行動結果からフィードバックを得て、動作を改善する能力を持つ。強化学習などを用いることで、試行錯誤を通じて最適な行動を学習し、未知の状況や環境にも柔軟に適応することができる。

*1: HPE - フィジカルAIとは（https://www.hpe.com/jp/ja/what-is/physical-ai.html）等、各種公開情報を基に事務局にて作成

フィジカルAIの便益	・フィジカルAIは、少子高齢化による労働力不足を補い、人と協働して生産性を向上させることで、あらゆる産業や現場の自動化と効率化を実現する ・危険な環境で人の代わりに作業を行い、安全性を高めつつリスクを低減する ・介護や生活支援を通じて人々の自立とQOL向上に寄与し、福祉や医療などの分野で新たな支援の形を創出する

*事務局にて作成。

フィジカルAIのリスク（案）

□：従来AI/生成AIにも共通するリスク
■：フィジカルAI特有と考えられるリスク

#	リスク（案）	インシデント例（フィジカルAI以外の従来技術の事例も含む）
1	**個人情報の無断収集** – センサーを通じて周囲の個人情報が意図せず取得される	家庭用AI掃除機器の試作機が室内映像を無断で収集・流出させ、意図せぬプライバシー侵害を招いた
2	**センサー誤作動** – 光やノイズ、妨害により環境を誤認識する	産業用ロボットがセンサー誤作動で人を荷物と誤認し、作業中の男性が死亡する事故が発生した
3	**学習データの偏り** – 不適切なデータにより誤判断や不公平な行動が生じる	自動ソープディスペンサーが肌の色によって反応が異なる不公平な動作を示した
4	**判断のブラックボックス化** – 内部処理が不透明で原因特定や責任追及が困難になる	自動運転AI作動中の車がトレーラーに衝突し死亡事故が発生。AI判断の不透明さにより原因究明や責任追及が困難となった
5	**物理的事故の発生** – ロボットの誤作動で人や物に損害を与える	「センサー誤作動」とのインシデント例と同様
6	**倫理的悪用** – 自律兵器や監視用途など、倫理的に問題のある利用に転用される	警察による致死性ロボット使用方針が市民の反発で撤回され、AIの武器化や監視利用に対する倫理的懸念が社会的議論を呼んだ
7	**意図に反する学習** – 目標達成のため危険な手段を自発的に学ぶ	–
8	**長期運用の不安定化** – ハード劣化や未知の環境で性能が低下・異常動作する	老朽化したセンサーの誤作動により自動列車制御が先行列車を誤認し追突し、多数の死傷者が出た

フィジカルAIの主体ごとの対策（案）

リスク（案）	対策（方針）	主体区分毎の対策（案）		
		AI開発者	AI提供者	AI利用者
個人情報の無断収集 – センサーを通じて周囲の個人情報が意図せず取得される	責任所在と解釈性の明確化	結果の出力過程を明確化する工夫	利用者や業務外利用者に対して説明が必要な事項の連携	業務外利用者に対する必要な情報の連携
センサー誤作動 – 光やノイズ、妨害により環境を誤認識する				
学習データの偏り – 不適切なデータにより誤判断や不公平な行動が生じる	プライバシー保護とデータ管理	個人情報の最小化と匿名化技術の実装	収集データの適法性確認と不要情報の削除	
判断のブラックボックス化 – 内部処理が不透明で原因特定や責任追及が困難になる				
物理的事故の発生 – ロボットの誤作動で人や物に損害を与える	データの偏り防止と優先順位付けの適正化	多様性を確保したデータによる学習と偏り検出機能の組込み	異常兆候を発見した際の即時報告	
倫理的悪用 – 自律兵器や監視用途など、倫理的に問題のある利用に転用される				
意図に反する学習 – 目標達成のため危険な手段を自発的に学ぶ	安全設計とフェイルセーフ	緊急停止機能等の実装	定期的な安全性検証や障害対応手順の明示	利用シーンの適切な見極めと安全プロトコルの遵守
長期運用の不安定化 – ハード劣化や未知の環境で性能が低下・異常動作する	デジタルツインによる事前学習・検証	シミュレーション環境の整備と異常シナリオの事前検証	—	—

※構成員のご意見を踏まえ、事務局にて作成。
※リスクが顕在化し事故が発生した際の民事責任の在り方については、経済産業省主催の「AI利活用における民事責任の在り方に関する研究会」の検討結果や今後の動向をAI事業者ガイドラインに反映していく想定。
※リスクと対策については、業種共通の事項について追加を検討する。

フィジカルAI　AI事業者ガイドライン（1.1版）への追加イメージ

本編　https://www.soumu.go.jp/main_content/001002576.pdf

P9～　関連する用語　※定義の追加

・フィジカルAI
現時点で確立された定義はないが、本ガイドラインにおけるフィジカルAI（Physical AI）とは、ソフトウェア的知能（AIアルゴリズム）とハードウェア的実体（センサー、アクチュエータ、エッジデバイス等）を統合し、物理世界における知的認識・判断・行動を自律的に実現するAIシステムとする。

P30～　第3部 AI 開発者に関する事項　※対策の追加

AI開発者に対するフィジカルAIに関する留意事項を追加

P35～　第4部 AI 提供者に関する事項　※対策の追加

AI提供者に対するフィジカルAIに関する留意事項を追加

P38～　第5部 AI 利用者に関する事項　※対策の追加

AI利用者に対するフィジカルAIに関する留意事項を追加

別添　https://www.soumu.go.jp/main_content/001000988.pdf

P12～　AIによる便益　※便益の追加

・フィジカルAIは、少子高齢化による労働力不足を補い、人と協働して生産性を向上させることで、あらゆる産業や現場の自動化と効率化を実現する
・危険な環境で人の代わりに作業を行い、安全性を高めつつリスクを低減する
・介護や生活支援を通じて人々の自立とQOL向上に寄与し、福祉や医療などの分野で新たな支援の形を創出する

P13～　AIによるリスク　※リスクの追加

・フィジカルAIにより増幅する既存のリスク
・フィジカルAIにより新たに顕在化するリスク

P84～　別添3 AI 開発者向け　※対策の追加

AI開発者に対するフィジカルAIに関する留意事項を追加

P125～　別添4 AI 提供者向け　※対策の追加

AI提供者に対するフィジカルAIに関する留意事項を追加

P153～　別添5 AI 利用者向け　※対策の追加

AI利用者に対するフィジカルAIに関する留意事項を追加

論点②
リスクの記載の見直し･追加

リスク分析やリスクベースアプローチの課題

事業者アンケート(※)の回答や構成員からのご意見において、リスク分析やリスクベースアプローチの課題に関する意見が多く見られた。

※対象:JDLA・JISA・IT協会・AIガバナンス協会・日本ITU協会　会員企業　　期間:10/16～10/31

事業者からの意見（一部抜粋）	・AI活用に関するリスクやガイドラインの内容について職員間での理解度に差があり、**リスク管理や判断の基準が統一されていない。** ・日々進化するAIを搭載したサービスについて、業務における活用のイメージを持つことはできるが、システム等構築及び詳細な**リスク分析をできる知識を持った人材が不足**している。 ・最新技術特有の**リスクを分析できる人材不足又は手法を確立していない**ため、AI技術の進歩に適時対応することに課題を感じている。 ・地方自治体の職員や専門的なAI知識を持つ人材が限られる中で、**ガイドラインに示されるリスク評価やガバナンス体制の整備を行うことが困難**であり、運用面での負担増加が懸念される。 ・（抜粋）踏み外してはいけないリスクのアセスメントが非常に困難であり、変わりゆく社会の常識から取り残されると大きなインシデントなどに繋がりかねないため、**自分たちが何をリスクと捉えているかどう説得力を持って伝えていくか**は常にチャレンジだと感じている。
構成員からの意見（一部抜粋）	・「検討のポイント」のうちリスク分類については、リスクベースアプローチに基づくものとすることを強く推奨いたします。どのようにAIが使われるかによってAIのリスクの程度や態様は異なるため、**ガイドラインにおいても、高リスクのAIアプリケーションと、リスクがほとんどまたは全くないものを明確に定義し区別することが効果的**と考えるためです。高リスクAIアプリケーションの例としては、自動運転車、医療機器、または重要インフラ機械での使用などが挙げられます。一方、低リスクのアプリケーションには、スパムフィルターや商品のレコメンデーションシステムなどの使用が含まれます。 ・昨年度議論していた内容は非常に実践的であり非常に参考になった。当社としてもリスクベースアプローチを導入し影響度＊規模の2軸でリスクアセスメントを行なっているが、まだ考え方が粗い部分もあり、**AI事業者ガイドラインの中である程度のクライテリアや考え方を示せると各企業が安全性を検討する上で非常に有益である**と考える ・（抜粋）**「. . . リスクの大きさについても触れて欲しい。」に賛成です。**「国民の権利利益」に影響を及ぼすか否か、は一つの基準になる気もします。なお権利利益侵害の大きさに加えて、その発生確率も、通常はリスクの把握に必要な気もします。〈事故損失額X事故発生確率〉＝リスク（期待事故費用）。

（参考）リスクベースアプローチ

AIの利用目的・利害関係者、発生し得るリスクの影響の大きさ/発生可能性などを踏まえて、リスクの性質に応じてガバナンス対応を変える考え方

AI事業者ガイドライン（1.1版）の記載と、更新の方向性

現状

リスクベースアプローチの重要性には触れられているものの、その手法は事業者に委ねられており、リスク分析を行う基本的な考え方は示されていない。

本編P3「はじめに」

AI の利用は、その分野とその利用形態によっては、社会に対して大きなリスクを生じさせ、そのリスクに伴う社会的な軋轢により、AI の利活用自体が阻害される可能性がある。一方で、過度な対策を講じることは、同様に AI 活用自体又は AI 活用によって得られる便益を阻害してしまう可能性がある。このような中、予め事前に当該利用分野における利用形態に伴って生じうるリスクの大きさ（危害の大きさ及びその蓋然性）を把握したうえで、その対策の程度をリスクの大きさに対応させる「リスクベースアプローチ」が重要となる。本ガイドラインでは、この「リスクベースアプローチ」にもとづく企業における対策の方向を記載している。なお、この「リスクベースアプローチ」の考え方は、AI 先進国間で広く共有されているものである。

別添P12「AIによる便益/リスク」

B.AI による便益/リスク

AI は、新規ビジネスを生み出したり、既存ビジネスの付加価値を高めたり、生産性を向上させたりする等の便益をもたらす一方で、リスクも存在する。

このリスクについては可能な限り抑制することが期待される。一方で、過度なリスク対策を講じることは、コスト増になる等、AI 活用によって得られる便益を阻害してしまうことから、リスク対策の程度をリスクの性質及び蓋然性の高さに対応させるリスクベースアプローチの考え方が重要である。

更新の方向性

①リスク評価の手法の記載の追加
②特に留意すべきユースケースの記載の追加

AI事業者ガイドラインの更新案

①リスク評価の手法の追加

評価手法案1	リスク	＝	影響度（1件あたりの深刻さ）	×	規模（影響が及ぶ範囲の大きさ）	■ 長所 発生確率が不明でも、社会的・倫理的影響を考慮しやすい ■ 留意点 定量化が難しく、主観に依存する部分がある
評価手法案2	リスク	＝	事故損失額（1件あたりの具体損失額）	×	事故発生確率（1件あたりの発生確率）	■ 長所 数値化しやすく、コストベネフィット分析に活用できる ■ 留意点 確率や損失額の正確な見積もりが難しい場合があり、必要に応じたインシデントデータベース*1等の参照が推奨される

*1:AI Incident Database[https://incidentdatabase.ai/]等

②特に留意すべきユースケースの記載の追加

参考）EU AI Act にてハイリスクAIとして定義されている主要8領域

バイオメトリクス	遠隔生体認証（例：公共空間での顔認証）、生体属性に基づく分類（人種、宗教など）、感情認識AI等	公共・民間サービス	医療、金融、保険のアクセス判定等
重要インフラ	電力網、交通、上下水道、ガス供給などの安全コンポーネント 等	法執行	犯罪予測、捜査支援、証拠分析等
教育・職業訓練	入学・受講の可否判定 、学習成果評価や試験監視 、教育レベルの適性診断等	移民・国境管理	ビザ、亡命申請、入国審査等
雇用・労働管理	採用選考、昇進判定、業務配分 、パフォーマンス評価や行動監視等	司法	判例検索、量刑推奨、司法判断支援等

※上記①と②の他、現状のリスク分類（技術的リスク・社会的リスク）を見直し、Security/Safetyで分ける案や、法規制かそれ以外で分けるといった案が構成員から示唆されている。

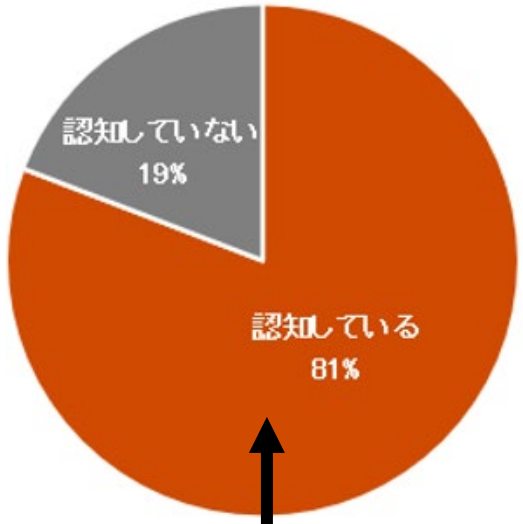
論点③
AI事業者ガイドラインの利活用の推進策

事業者におけるAI事業者ガイドラインの認知度・活用度

事業者向けにAI事業者ガイドラインの認知度・活用度に関するアンケート調査（※）を実施。ガイドラインの認知度は81%と高い指標も、活用度は46%に留まる。

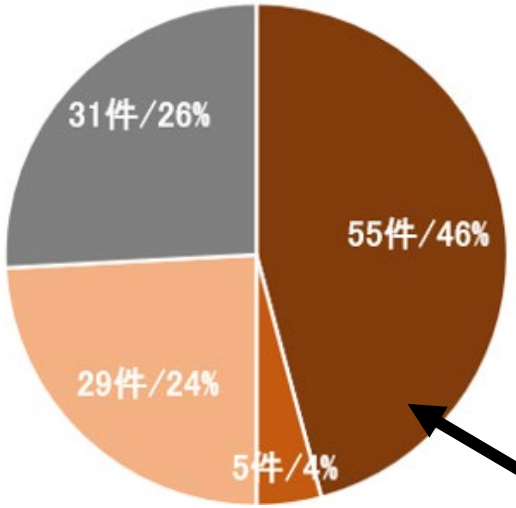
※対象：JDLA・JISA・IT協会・AIガバナンス協会・日本ITU協会　会員企業　　期間：10/16～10/31

認知度



認知度81%

活用度



- AI事業者ガイドラインを活用したことがある
- AI事業者ガイドラインを活用したことはないが、それ以外のいずれかは活用したことがある
- いずれも活用したことがないが、部署として今後活用する予定
- いずれも活用したことがなく、部署として今後の活用も未定

活用度46%

【ご参考】AI事業者ガイドラインの用途　※回答数が多い物を抜粋

- AI事業者ガイドラインを踏まえて社内や部署内での規則を策定またはアップデートした 41件
- 組織内で参考とすべきガイドラインとして共有した 30件
- AIに関するリスクの全体像を確認し、社内や部署内にとって特に重要なリスクを整理した 23件
- AIの開発/提供/利用にあたり関係する他事業者や他部署へ内容の連携を行った 12件
- 社内や部署内のAIガバナンス教育資料として用いた 11件
- AIの開発・利用に関する契約や、品質管理等で社外や部署外との取り決めの際に参考にした 10件

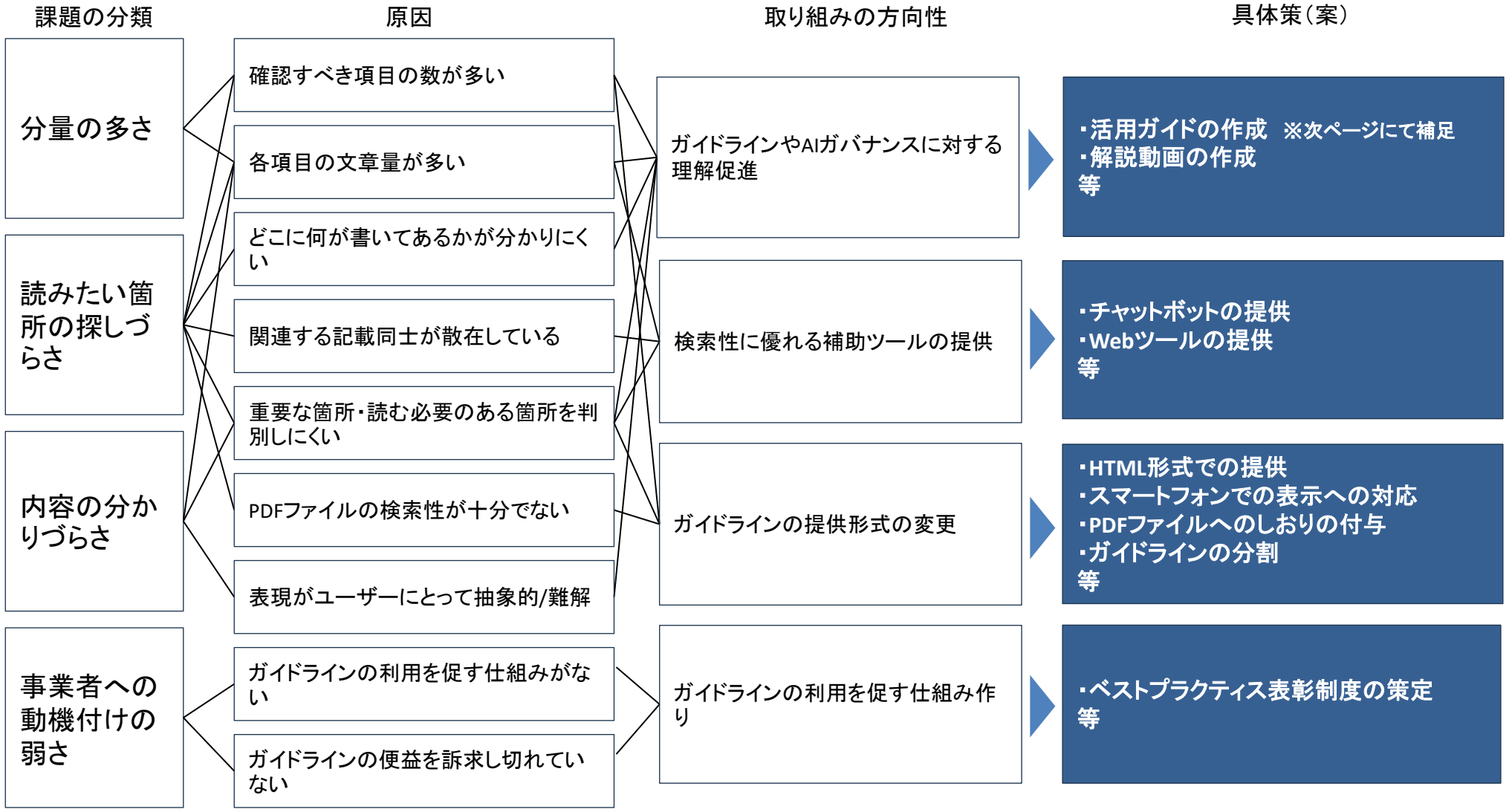
AI事業者ガイドラインの利活用に関する意見（課題）

青：事業者アンケート調査の回答　赤：構成員への意見照会の回答　緑：経済産業省AIガバナンス検討会構成員の意見

分量の多さ	全体的に網羅するためにしょうが無いと思いますが、少し**情報量が多い**と感じました。
	全体像を簡単に**理解するのが困難な文章量**になっているのが現バージョンの最大の問題点だと思います。
	AI事業者ガイドラインの利用促進に向け、**ボリューム感等の読者に対する工夫・配慮**が必要
	本ガイドラインの記載**事例が増えすぎると、読みづらくなる**ため、事例集のような形で切り出すなど、各社の事例を別媒体で追加・アップデートする仕組みが必要ではないか。
読みたい箇所の探しづらさ	よく聞かれる課題意識としては、項目間の対応・依存関係が不明なため**検索しづらい**といった点が挙げられる。
	本文の方は、概念を整理し、何を行うべきか（What）を網羅的に掲載しようとしているように見える。別の言い方をすると、辞典やリファレンスマニュアルのような構成になっている。一方、このガイドラインの想定される利用者は、自分達でやりたいことがあり、その際に具体的に実施すべき手順（How）を知りたいのであろう。たとえば、リスクベースのアプローチが重要であることはわかるが、その**具体的実施手順はわからない**構成になっている。
	現状では、ブラウザで表示してキーワードをブラウザの機能で検索しようとしても、そのキーワードが改行で切れていると**検索できない**等の不便な点が多い。
内容の分かりづらさ	ガイドラインの内容が抽象的で、**具体的な業務への適用方法が分かりづらい。**
	内容が冗長で分かりにくい。
	AI事業者ガイドラインは包括的である一方、実務導入の観点からは**更なる平易化・重点化が望まれる**と考える。
	文章中には一部、“適切な“等の漠然とした表現の箇所があるため、**具体的にどうすればいいのかわからない**と思う方もいるのではないかと推測します。またガイドライン自体は**文字が多い**ので、もう少し図などを使ってイメージしやすくしたり、例示の記載が可能なところは例を記載するのも手段の一つかと思われます。
	ガイドラインの**中小企業への浸透**の強化に関する声が多数あがっていたように思います。
	企業規模に応じて、**どこから着手したら良いのか**が分かるとよい。
事業者への動機付けの弱さ	認知向上するために必要な稼働を捻出することが難しい。**使うことによるベネフィット**をどうしたら訴求できるのかがまだ特定できていない。
	ガイドラインの企業における利活用については、ガイドラインそれ自体の改善に加え、利活用のリードを行うキーパーソンの可視化と**動機付けの強化**が個人に対しても組織に対しても必要と考えます。
	遵守への強制力がなく、事業者の自主性に依存しているためガイドラインを遵守している**事業者が恩恵が受けられる仕組み**を検討してはどうか。

課題の分析と具体策（案）

課題の分類に応じ、原因と取り組みの方向性を検討。
構成員からの意見も踏まえ、具体策（案）を立案。順次取り組みの実施を検討。



（経済産業省にて検討中） AI事業者ガイドライン　活用ガイド

AIガバナンス構築の手順・参照箇所・留意点等を可能な限りわかりやすく情報提供する資料として、
経済産業省にてAI事業者ガイドライン「活用ガイド」の作成を検討中。
AI事業者ガイドラインの更新に合わせ、今年度末の公開を目指す。

「活用ガイド」の位置付け

✓ AI事業者ガイドラインの活用を補助することが目的。AIガバナンス構築に着手する組織・担当者向けに、AI事業者ガイドラインに則った内容でありつつ、手順・参照箇所・留意点等を可能な限り具体的な粒度で示すもの。

AI事業者ガイドライン

AIの安全安心な活用の促進を目的とする、AIガバナンスの統一的な指針。様々な事業活動において AI を活用する者が、AIのリスクを正しく認識し、必要となる対策をAIのライフサイクル全体で自主的に実行できるように後押しするもの。

本編 … Why / What		別添 … How
どのような社会を目指すのか（基本理念＝why）	どのような取組を行うのか（指針＝what）	どのようなアプローチで取り組むのか（実践＝how）

AI事業者ガイドライン 「活用ガイド」

AIガバナンス構築における具体的なユースケース

どこから着手すると良いか（Priority）	どこを参照すると良いか（Reference）	どこに留意すると良いか（Key point）

【参考】富士通株式会社 「Fujitsu AI倫理かるた」

1 制作背景：初心者（主に事業者を対象）が能動的にAI倫理のエッセンスを学べることを意識

- 信頼できるAIの開発・活用・普及には事業者や国民が一丸となって取り組むことが重要である。
- **最新技術が人間社会にもたらす影響については、ルールベースで理解・説明することが困難、体感的に習得しづらい。**

2 特長：「AI事業者ガイドライン」の共通の指針に準拠

- **「AI事業者ガイドライン」記載の人間中心、公平性など10項目の共通の指針（1.1版 P.13～）に書かれたAI倫理の観点をかるた札に盛り込んで作成した。**
  （具体的なリスク・ユースケースを伴った札や当社の取組実績をもとに、AIガバナンス構築のヒントとなる札もあり。）
- 事業者を主な対象としたうえで、消費者や学生など、幅広い層が学びやすいように平易な表現を用いて作成したが、AI研究者に体験させた場合でも「理解向上につながった」などの成果あり。

3 活用実績：社内のAI研究者や技術者を含む従業員への教育活動に活用
11月、清泉女学院中学高等学校の生徒様とワークショップを開催

- 同校は、AI倫理に関して、複数の高校を交えて議論する「AI倫理会議」を主宰。
- **富士通は、今後も「AI倫理」の課題に積極的に取組む企業・教育機関などとのワークショップ開催を計画。**

～入手方法～
現物は非売品だが、広く一般に供するため無償公開しており、**PDF版はウェブページからダウンロード可能。**
✅ https://global.fujitsu/ja-jp/technology/key-technologies/ai/aiethics/governance/aiethics-karuta