April 2026

IT PRACTICE GROUP

The "AI Guidelines for Business (Ver1.2)" for fiscal year 2025 were published on March 31, 2026

The updated "**AI Guidelines for Business**" (Ver1.2) were published by the Ministry of Internal Affairs and Communications, and the Ministry of Economy, Trade and Industry on March 31, 2026.

(Provisional English translation of official release)

[Version 1.2]
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_12.pdf

[Version 1.2 Appendices]
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_14.pdf

The AI Guidelines for Business were first published as Version 1.0 in April 2024, followed by the release of Version 1.1 in March 2025, and are reviewed annually.

[Version 1.1]
https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02ryutsu20_04000019.html

The AI Guidelines for Business are classified as "soft law," meaning that they do not have any legally binding force nor do they carry any penalties. However, they serve as unified guidelines for AI developers, providers, and business users to utilize AI safely and appropriately, and are increasingly functioning as de-facto standards in Japan.
Thus, it is appropriate for AI system developers (**AI developers**), as well as businesses providing services incorporating AI systems (**AI providers**) and businesses utilizing AI in their operations (**AI business users**) to conduct their business activities in accordance with these AI Guidelines for Business. It is highly anticipated that this soft law will help prevent accidents caused by AI and reduce the risk of liability in the event that an accident does occur.

2026 年 4 月

IT プラクティスグループ

2026 年 3 月 31 日に「AI 事業者ガイドライン（第 1.2 版）」が公表されました

総務省と経済産業省は、2026 年 3 月 31 日に「**AI 事業者ガイドライン**」の改訂版（第 1.2 版）を公表しました。

【第 1.2 版】（最終版）
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf

（第 1.1 版からの変更履歴付きの版）
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_10.pdf

【第 1.2 版 別添】（最終版）
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_3.pdf

（第 1.1 版からの変更履歴付きの版）
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_11.pdf

AI 事業者ガイドラインは、2024 年 4 月に第 1.0 版として公表された後、2025 年 3 月に第 1.1 版がリリースされ、年に 1 回のペースで見直されています。

【1.1 版】
https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02ryutsu20_04000019.html

このガイドラインは、あくまでも法的拘束力や罰則をもたないソフトローの位置づけです。もっとも、AI の開発者、提供者、利用者が安全かつ適正に AI を活用するための統一的な指針であり、日本国内における事実上の標準（デファクトスタンダード）として機能しつつあります。
したがって、AI システムの開発者（**AI 開発者**）は当然のこと、AI システムを組み込んだサービスを提供する事業者（**AI 提供者**）や、その事業活動において AI を利用する事業者（**AI 利用者**）は、このガイドラインの内容を踏まえた事業展開を行うことが適切です。これにより、AI に起因する事故を防止することや、万一事故が生じた場合でも責任を負う危険性を低下させることが期待されます。

It appears that in 2025 AI technology underwent a paradigm shift. Until then, the primary role of generative AI had been to respond to user prompts or generate requested content. However, "**AI agents**" — which autonomously perceive their environment, plan, and execute tasks to achieve specific, assigned goals — began to gain widespread adoption. Furthermore, AI has become highly integrated into hardware systems such as military weapons, robots, drones, autonomous vehicles, and medical devices ("**Physical AI**"). While such advancements in AI hold the potential to bring significant benefits to human society, they have also fundamentally altered the risks associated with AI use.
For example, as a result of the autonomous nature of such technology, AI agents might enter into contracts against the user's will or without their knowledge, or engage in actions that infringe upon the interests of others; however, to date there has not been any sufficient discussion regarding which authority should be liable for damage caused by AI agents [1]. Furthermore, with regard to physical AI, the possibility of AI malfunctions directly causing harm to human life and health has increased.

2025 年は、AI 技術が一つのパラダイムシフトを迎えた年であったように思われます。それまで、生成 AI は、ユーザによるプロンプトに回答したり、要求されたコンテンツを生成することが主たる役割でした。しかし、与えられた特定の目標を達成するために、自律的に環境を認識し、計画を立て、行動を起こす「**AI エージェント**」が普及し始めました。また、AI が、軍事兵器、ロボット、ドローン、自動運転車、医療機器などのハードウェアに高度に統合されるようになりました（「**フィジカル AI**」）。このような AI の進化は、人間社会にとって大きな利益を生む可能性がある一方、AI の利用に伴うリスクを明らかに変質させました。
例えば、AI エージェントが自律的に作動した結果、ユーザの意に反して AI が契約をしてしまったり、知らぬ間に他人の利益を侵害する行動をしてしまうことがあり得ますが、この場合、どの主体が責任を負うべきかについて十分な議論が行われていません [1]。また、フィジカル AI については、AI の誤作動が直接人間の生命・身体に危害を加える可能性を増大させています。

[1] If liability arises due to the use of an AI technology, under conventional frameworks (in terms of civil liability: contractual liability, tort liability, product liability, etc.), there is a possibility that one or more of the respective business operators will bear civil or criminal liability depending on the specific cause. However, at present, there is no established theory even on fundamental issues, such as whether a contract executed without authorization by an AI agent binds the user. Moreover, concepts of granting a pseudo-corporate personality to AI technology itself in order to separate the risks of each business operator have been proposed. However, since a consensus has not yet been reached, and it is considered that large-scale legislative measures would be necessary, it would be unwise to conclude that such proposition would be adopted and put into practice in the near future.

[1] AI の利用に起因して責任が生じた場合、従来の枠組み（民事責任で言えば、契約責任、不法行為責任、製造物責任等）によっても、具体的な原因次第では、各事業者のいずれか（またはそのうちの複数）が民事・刑事上の責任を負う可能性があります。しかし、例えばエージェント AI が無断で締結した契約は利用者を拘束するのかといった基本的な論点すら現時点では定説はありません。さらに、AI に疑似的な法人格を付与することで、各事業者のリスクを切り離すといった構想も提唱されていますが、コンセンサスが得られているものではなく、さらに大規模な立法措置が必要と考えられますので、すぐに実現するものではないといえます。

The guidelines were updated following this background and involved not only mere technical updates, but also significant changes that may involve corporate responsibility and governance structures.
We will focus on the key points of these updated guidelines: **AI agents** and **Physical AI**.

今回の改正は、このような背景を踏まえて行われたものであり、単なる技術的なアップデートにとどまらず、企業の責任やガバナンス体制に関連し得る重要な変更が盛り込まれています。
以下では、本改正の内容のうち、主要な論点である **AI エージェント**と**フィジカル AI** にフォーカスをあてて解説いたします。

Note: The page numbers cited below refer to Ver1.2 (provisional English translation).

注：なお、以下において引用する頁数は、本ガイドライン（第 1.2 版・日本語版）の最終版のページ数です。変更履歴付きの版のページ数とは異なる場合がありますので、ご注意ください。

First, in light of recent advances in the social implementation of AI agents, a new **definition of an AI agent** has been introduced (p. 10). Specifically, an AI agent is defined therein as "an AI system that perceives its environment and acts autonomously to achieve a

はじめに、近年の AI エージェントの社会実装の進展を踏まえ、新たに **AI エージェントの定義**が導入されました（11 頁）。すなわち、AI エージェントは「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」と定義

specific goal." Additionally, as a more comprehensive and evolutionary concept than that of an AI agent, a goal-driven AI system which autonomously makes decisions and takes actions through the use of multiple AI agents is defined as "**Agentic AI**." Here, the concept of "autonomy" does not only refer to a highly autonomous state but also includes other systems with their own degree of autonomy.

されています。あわせて、AI エージェントよりも包括的かつ進化的な概念として、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 システムを「**エージェンティック AI**」と定義しています。
ここでいう自律的とは、高度な自律状態だけを指しているのではなく、ある程度の自律性を持つものも含むとされています。

Similarly, a **definition of Physical AI** has also been added (p. 11). Specifically, Physical AI is defined as "systems which perceive their environment through sensors, etc., process that information using an AI model, autonomously infer and determine strategies for achieving objectives given by humans, prompting physical actions via actuators (drive systems), etc., which are characterized by taking direct actions (moving, operating, processing, etc.) in the real world, rather than being limited to processing in cyberspace."

同様に、**フィジカル AI の定義**も追加されました（11 頁）。すなわち、フィジカル AI は、「センサ等によるセンシングを通じて物理環境の情報を取り込み、AI モデルによる処理を経て、設定された目的を達成するための最適な方策を自律的に推論・判断し、アクチュエータ（駆動系）等を介して物理的な行動へとつなげるシステムであり、サイバー空間での処理に留まらず、現実世界に対して直接的な働きかけ（移動、操作、加工など）を行うことを特徴とするもの」と定義されています。

Note: The definition in the published provisional translation is simpler as cited below. However, since there is a discrepancy between this and the official Japanese version, we have translated it as shown above.
"*In these Guidelines, physical AI refers to a system that takes in environmental information through sensors, processes that information using an AI model, autonomously infers and determines strategies for achieving objectives given by humans, and furthermore acts on those strategies without human intervention*."

Appendix 1 provides examples of AI agents, such as the "Travel Destination Suggestion and Booking AI Agent" (an AI agent service provided by an airline that proposes optimal travel destinations and flight options according to the user's requirements, and links to external booking systems for the purpose of making actual flight reservations) and the "Sales and Customer Support Assistance AI Agent" (a business support AI agent that autonomously provides 24-hour assistance in sales and customer support departments, covering the task of responding to prospective customers to making a sale through to handling inquiries) (Appendices, p. 12).
Appendix 1 also highlights the potential of AI agents, stating that understanding user intentions and autonomously executing tasks can streamline complex business processes and significantly reduce human workloads; rather than merely executing instructions, such technology collaborates with multiple systems and applications to make independent decisions and optimizations according to each individual situation enabling the automation of

別添 1 では、AI エージェントの実用事例として、「旅行先提案・予約 AI エージェント」（ある航空会社が提供している、ユーザの希望条件に応じて最適な旅行先とフライトオプションを提案し、外部予約システムと連携して実際の便予約まで行う AI エージェント型サービス）や、「営業・CS 支援 AI エージェント」（営業やカスタマーサポート部門において、見込み客への対応から商談進捗の管理、お問い合わせ対応までを 24 時間自律的に支援する業務支援型 AI エージェント）が紹介されています（別添 12 頁）。
このような AI エージェントのポテンシャルとして、ユーザの意図を理解し、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ほか、単なる指示実行にとどまらず、複数のシステムやアプリケーションと連携し、状況に応じた判断や最適化を行うことで、従来は人手に依存していた調整・分析・意思決定を自動化することが可能となることが紹介されています（別添 17 頁）。

adjustments, analyses, and decision-making power that traditionally relied on human intervention (Appendices, p. 17).

Similarly, as practical examples of Physical AI, the "Autonomous Driving System" (a vehicle control system equipped with autonomous driving technology that, when a user specifies a destination, analyzes the surrounding environment, autonomously plans and executes the driving route, and performs safe and efficient driving by detecting the actual traffic conditions, road signs, and obstacles in real time) and the "Autonomous Mobile Robot" (a warehouse transportation service performed by an autonomous mobile robot characterized by its ability to flexibly perform route planning and respond to environmental changes, where the robot analyzes the floorplan and environment of the warehouse, autonomously selects the optimal route to transport packages, and significantly improves work efficiency by avoiding obstacles and performing collaborative operations among multiple robots) are highlighted (Appendices, p. 13).
Regarding the potential of such Physical AI, Appendix 1 notes that it can supplement the labor shortage caused by the declining birthrate and aging population, reduce risks while enhancing safety by operating in hazardous environments, and contribute to people's independence and improved quality of life (QOL) through nursing care and daily life support (Appendices, p. 17).

同様に、フィジカル AI の実用事例として、「自動運転システム」（ユーザが目的地を指定すると、車両は周囲環境を認識し、走行経路を自律的に計画・実行したり、交通状況や道路標識、障害物をリアルタイムで検知し、安全かつ効率的な運転を実現する、自動運転技術を搭載した車両制御システム）や、「自律移動ロボット」（ロボットが倉庫内の地図を認識し、最適なルートを自律的に選択して荷物を搬送したり、障害物回避や複数ロボット間の協調動作を行うことで作業効率を大幅に向上させる、経路計画や環境変化への対応を柔軟に行える点を特徴とする自律移動ロボットによる倉庫内搬送サービス）が紹介されています（別添 13 頁）。
このようなフィジカル AI のポテンシャルとして、少子高齢化による労働力不足への補充、危険な環境において稼働することで安全性を高めつつリスクを低減すること、介護や生活支援を通じて人々の自立と QOL 向上に寄与することなどが紹介されています（別添 17 頁）。

On the other hand, the Appendices warns of the risks posed by AI agents and Physical AI, which are summarized as follows.

● Attacks on AI systems, such as data poisoning attacks
In systems handling complex and diverse information, such as AI agents and Physical AI, the number of diverse input routes and external linkages increases. This widens the scope of attacks, raising concerns that the risk of data poisoning, etc., will further increase (Appendices, p. 18).

● Impact of incorrect outputs due to hallucinations, etc.
In the case of AI agents, the risk of incorrect outputs by AI agents is not limited to traditional risks such as the spread of disinformation but poses a risk of actions such as the unintended ordering of goods or deletion of files being performed (Appendices, p. 19).

● Problems associated with black-boxing
Because AI agents and Physical AI have more complex configurations and mechanisms compared to ordinary AI systems, they are more significantly affected by the "black box" nature of AI decision-making, which may increase the difficulty of maintenance and troubleshooting (Appendices, p. 19).

他方で、別添は、AI エージェントやフィジカル AI の持つリスクについて、概要以下の通り警告しています。

●データ汚染攻撃等の AI システムへの攻撃
AI エージェントやフィジカル AI のような複雑で多様な情報を扱うシステムでは、より多様な入力経路や外部連携が増えるため、被攻撃対象が拡大し、データ汚染攻撃等のリスクがさらに高まる懸念がある（別添 18 頁）。

●ハルシネーション等による誤った出力の影響
AI エージェントの場合、AI による誤った出力は、従来の偽情報の拡散といったリスクにとどまらず、人間の意図しない商品の注文やファイル削除等の動作が行われるリスクがある（別添 19 頁）。

●ブラックボックス化に伴う問題
AI エージェント、フィジカル AI は、通常の AI システムに比して複雑な構成や機構を持つため、AI の判断が「ブラックボックス」である影響をより大きく受けることとなり、メンテナンスやトラブルシューティングの難易度が上がるおそれがある（別添 19 頁）。

● Serious cybersecurity issues caused by misuse
Advances in AI agents have made code generation easier, thereby making it easier to create malware and generate code for infiltrating networks. Thus, the misuse of AI agent technology increases risks such as the leakage of personal information, the shutdown of systems related to critical infrastructure, and the falsification of information (Appendices, p. 20).

● Leakage of confidential information
There have been an increasing number of incidents where AI agents autonomously collaborate with external systems and cloud services to execute various tasks. In doing so, there is a risk that confidential data may be unintentionally leaked due to attacks exploiting vulnerabilities, etc. (Appendices, p. 21).

● Risks of privacy violations
Physical AI has the possibility of acquiring personal information by itself, and there is a risk of privacy violations where such information is retained on a device which can lead to the identification of individuals or inappropriate use (therefore, it is necessary to establish a system that does not acquire or retain unnecessary information). In addition, there is a risk of data leakage when discarding hardware (therefore, it is desirable to appropriately delete data stored in the storage medium of equipment or the cloud) (Appendices, pp. 140 [footnote], 163 [footnote]).

● Risks of AI networking due to Agentic AI
Furthermore, regarding Agentic AI, it is well known that the potential "networking" of AI could induce and amplify risks where the AI in use becomes uncontrollable by connecting and interacting with other AI technologies through the Internet, etc. (Appendices, p. 27 footnote).

The AI Guidelines for Business emphasize the use of a risk-based approach (a method of determining the priority of countermeasures based on the purpose of AI use, stakeholders, the magnitude of the impact/probability of potential risks, etc.) for mitigating risks posed by AI. This risk-based approach has also been incorporated in the OECD AI Principles, the NIST AI Risk Management Framework, the EU AI Act, etc., and is a mainstream concept globally.
Each business operator is required to analyze risks for each individual case, service, or product, and take necessary levels of protective measures according to the results.

This also applies to AI agents and Physical AI. Particularly for AI agents, it is important to prioritize matters requiring human judgement according to their level of importance and to

●悪用による深刻なサイバーセキュリティ上の問題
AI エージェントの進歩により、コードの生成が容易になることで、マルウェアの作成や、ネットワークに侵入するためのコードの生成がより容易になっている。すなわち、AI エージェント技術が悪用されることで、個人情報の漏洩や重要インフラに関わるシステムの停止・情報の改ざん等のリスクが増大することになる（別添 20 頁）。

●機密情報の流出
AI エージェントでは、外部システムやクラウドサービスと自律的に連携して各種タスクを実行するケースが増えており、その過程で、脆弱性を突かれた攻撃等により、内部データが意図せず外部に送信されるおそれがある（別添 21 頁）。

●プライバシー侵害のリスク
フィジカル AI は、それ自体で個人情報を取得する可能性があり、こうした情報がデバイス上に残存し、解析や不適切な利用に結び付くことでプライバシー侵害のおそれがある（そのため、不要な情報を取得・保持しない仕組みを設ける必要がある）。また、ハードウェア廃棄時にデータが流出するリスクがある（そのため、機器やクラウドの記憶媒体に保存されたデータも適切に削除することが望ましい）（別添 145 頁脚注、171 頁脚注）。

●エージェンティック AI による AI のネットワーク化リスク
さらに、エージェンティック AIについては、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指摘されている（別添 27 頁脚注）。

このガイドラインは、AI によるリスクのコントロールについて、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、対策の優先順位を決定する手法）を用いることを重視しています。リスクベースアプローチは、OECD AI 原則、NIST AI Risk Management Framework、EU AI Act 等でも採用されており、世界的にも主流の考え方です。
各事業者は、ユースケース、サービス又は製品ごとにリスクを分析し、その結果に応じて必要なレベルの保護措置を講じることが求められます。

これは、エージェント AI やフィジカル AI も同様で、とりわけエージェント AI については、人の判断を介在させるべき事項を重要度に応じて整理し、適切に対象を選定することが重

appropriately select the applicable scope (Appendices, p. 134). This does not mean that "human decision-making is always necessary and AI agents should not be used," but rather that while effectively utilizing the autonomy of AI agents, it is necessary to appropriately systematize the situations requiring human decision-making according to the impact of the potential risks.
If a business operator neglects to perform appropriate risk analysis, or fails to implement steps requiring human decision-making despite acknowledging the potential of significant risks occurring, and an accident is subsequently caused by an AI agent, such circumstances may establish the legal liability of the business operator (e.g., default on contractual obligations, breach of the duty of care in tort liability, defects under product liability). Furthermore, examples of criteria for determining the necessity of human decision-making are given on page 155 of the Appendices.

The AI Guidelines for Business detail specific risk management methods to be implemented by companies, with a focus on specific target audiences. We recommend referring to these sections first prior to putting risk management into practice.

- For Management: Appendix 2
- For AI Developers: Appendix 3
- For AI Providers: Appendix 4
- For AI Business Users: Appendix 5

Final Thoughts

This newsletter has outlined certain fundamental concepts and points to note with respect to AI agents and Physical AI as updates in the "AI Guidelines for Business (Ver 1.2)".
As previously stated, official discussions as to the civil and criminal legal liability in cases where liability arises due to AI technologies have not yet taken place, and the judicial precedents that can be relied upon are extremely limited. Therefore, for the time being, we take the position that while utilizing AI technologies stakeholders should acknowledge the opacity of such legal liability as a risk. For each business operator, conducting appropriate risk management based on the AI Guidelines for Business (as well as various guidelines in each individual industry) must be an indispensable requirement for the development of AI.

At Kitahama Partners, we handle consultations on AI risk management, backed by our extensive track record in supporting new businesses, litigation and negotiations related to system development and product liability, etc., and crisis management cases.
If you have any specific matters you would like to discuss, please feel free to contact our IT Practice Group.

要です（別添 138 頁）。これは、「常に人の判断が必要でありエージェント AI を使うべきではない」という趣旨ではなく、エージェント AI の自律性は上手く活用しつつも、リスクの濃淡に応じて、人の判断が必要な局面を適切に仕組み化しておく必要があるということを意味しています。
仮に、適切なリスクの分析を怠ったり、重大なリスクを認識しながらも人の判断を必要とする仕様にしていなかったことで、結果的にエージェント AI により何らかの事故が生じた場合、かかる事情は、当該事業者の法的責任（契約上の債務不履行、不法行為責任における注意義務違反、製造物責任上の欠陥等）を基礎づける可能性があります。
なお、人の判断の要否を決定する基準については、別添 161 頁に例が挙げられています。

AI 事業者ガイドラインは、各企業がとるべきリスクマネジメントの具体的手法については、対象者ごとに分類したうえで詳細に定めているため、実際にリスクマネジメントを行う際は、まずはそちらをご参照ください。

- 経営層：別添 2
- AI 開発者：別添 3
- AI 提供者：別添 4
- AI 利用者：別添 5

【まとめ】

この解説では、「AI 事業者ガイドライン」（第 1.2 版）のアップデートのうち、AI エージェントとフィジカル AI に関する内容を解説いたしました。
ご説明した通り、AI に起因して責任が生じた場合の民事・刑事上の法的責任については議論が成熟しておらず、依拠し得る判例等もごく限られています。したがって、当面の間は、かかる法的責任の不透明さをリスクとして甘受しながら AI を利用していくほかないと思われます。各事業者におかれましては、このガイドライン（及び個別分野における各種ガイドライン）を踏まえた適切なリスクマネジメントを行うことが AI の発展のために必要不可欠な要求であるといえます。

北浜法律事務所では、これまでの豊富な新規事業支援、システム開発・製造物責任等の訴訟・交渉、危機管理案件の実績を背景に、AI のリスクマネジメントのご相談に対応しております。
具体的なご相談がございましたら、是非 IT プラクティスグループの弁護士までご相談ください。

*****

KITAHAMA PARTNERS

Ryosuke Naka (Partner)
RNaka@kitahama.or.jp

Minami Hosoi (Associate)
MHosoi@kitahama.or.jp

*****

弁護士法人北浜法律事務所

中 亮介（パートナー）
RNaka@kitahama.or.jp

細井 南見（アソシエイト）
MHosoi@kitahama.or.jp