

3分でわかる「AI事業者ガイドライン」第1.2版改定のポイント

2026年3月公開予定のガイドライン改定案における、
新技術への対応指針と実務上の重要変更点。

総務省が提示した最新改定案の分析。

新たに定義された2つの重要カテゴリとリスク

AIエージェントとフィジカルAIの定義



目標に向け自律行動する
「AIエージェント」と、
物理世界で動作する
「フィジカルAI」が明文化。



人間に
よる関与
(HITL)

事業者求められる実務対応とガバナンス

「人間による承認 (HITL)」の必須化



高リスクな操作について
は、AIの判断だけで
完結させず人間の承認
を介在させる設計が
必須。

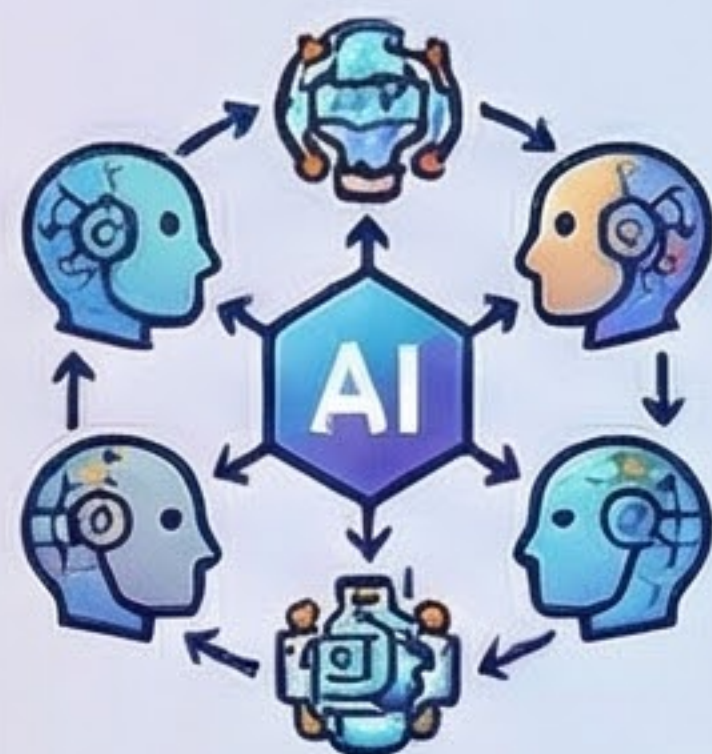
特有のリスク種類の整理



ツールの濫用や物理的
事故など、これまでの
AIにはなかった新技術
固有のリスクが特定さ
れた。

エージェントAI (マルチエージェント)

複数のAIが協働・競争
して意思決定を下す、
より高度な概念も4類
型で整理。



AIエージェントの安全性を確保するための、主体別の主な対策

AIエージェント安全対策：主体別アクション



AI開発者
(AI Developer)

対策方針



自律行動の制御：
ガードレール設計



権限管理：
最小権限設計の徹底



信頼性確保：
信頼度評価の付与



AI利用者
(AI User)

対策方針



自律行動の制御：
高リスク操作の人間承認(HITL)



権限管理：
操作履歴の定期確認



信頼性確保：
AIの限界を理解した運用

リスクベースアプローチの導入

影響度や発生確率に基
づく2つの評価手法が提
示され、定量的・定性的
な判断が求められる。



RAG(検索拡張生成)の役割の明確化

モデルのパラメータを更
新しないRAGは、開発者
ではなく「利用者・提供
者」の責任範囲となる。