protiviti®
Global Business Consulting

BOARD PERSPECTIVES

ISSUE 186

エージェント型AI：その正体と取締役会が注目すべき理由

2024年には、生成AIが注目を集めました。そして2025年には、エージェント型AI（自律型AIともいいます）がAI導入の次なる最前線として浮上してきました。エージェント型AIとは何か、経営陣がAIをどのように活用しようとしているのかを取締役が理解することがなぜ重要なのでしょうか。

生成AIは、プロンプトやリクエストに応じテキストや画像など新しく斬新なコンテンツを作り出すことが可能です。生成AIモデルはこの3年間で急速に進化し、限られた推論能力しか持たないテキストのみのモデルから、高度な段階的推論能力と研究能力を備えた複合型モデルへと進歩しました。これらの進歩は、エージェント型AIと呼ばれるAI進化の次の時代を切り開く鍵となります。

「エージェント型AI」とは、大規模言語モデル（LLM）やその他のテクノロジーをエコシステムの一部として採用し、自律的なタスクプランニングを可能にし、人間が定義した目標を達成するための推論をサポートするAIシステムのことです。人間が継続的に介入することなく自律的に動作するエージェント型AIは、データを収集、評価して最善の行動方針を決定し、その目的を達成するために行動します。さらなるアクションのためにコンテンツを生成する生成AIとは異なり、エージェント型AIシステムはLLMを活用して、自律型 AI エージェントが大量のデータを処理・解釈し、計画を立て、割り当てられた目標に沿った出力を生成することを可能にします。

ある意味で、エージェント型AIは、ロボティックプロセスオートメーション（RPA）などの従来の自動化技術からの進化形といえます。ただし、一般的な自動化技術との違いや特徴については、依然として多くの混乱があります。次ページの表は、エージェント型AIと2つの一般的な自動化アプリケーションである RPA とインテリジェントオートメーション（AI による自動化）との違いを比較したものです。

	RPA	インテリジェントオートメーション	エージェント型AI
目的	定義済みのアクションを実行	ステップのリストに従い、少なくとも1つのステップでLLMを使用	目的を達成するために計画を立て、実行
自律性	低い	中程度	高い
適応性	なし	過去の経験から学ぶ	リアルタイムのフィードバックと状況の変化に対応
意思決定	限定的、事前に定義されたルールとスクリプトに従う	高度な機械学習アルゴリズムとデータ分析に基づく	目標を達成するための高度に進化した、動的に調整される計画に基づく
タスク処理能力	繰り返し行われるルーチン作業	複雑で、推論・予測・問題解決を必要とするもの	動的かつ目標志向であり、実行手順が異なる場合があるもの
例	顧客の問い合わせに対し、事前に設定されたテンプレートを用いて自動応答する	自然言語チャットボットを通じて顧客と対話し、一般的な問題は自動で解決し、複雑な問題は人間の担当者へ引き継ぐ	調達プロセスを主体的に支援し、ベンダーの調査、条件交渉、契約書の作成を行い、最終承認には人間を関与させる

要約すると、エージェント型AIシステムとは、強化学習を含む高度なAI技術を活用し、システムが経験から学習し、時間の経過とともに改善することを可能にするものです。これは、人間の監督をほとんど必要とせずに問題を解決するためのテクノロジーとモデルの集合体であり、多くの場合、「オーケストレーターエージェント」によって管理されます。その枠組みの中で、AI エージェントは、適切な自律性をもってタスクやプロセスを処理するように設計されています。オーケストレーターは、システム全体の中でワークフローと意思決定を円滑に最適化するために、特定のタスクを実行するのに最も適したエージェントを決定します。

取締役が注目すべき理由：可能性

AIエージェントは、職場を変革し、人間が行う仕事の定義を再構築し、顧客対応やバックオフィスのプロセスを一新する柔軟性を持っているため、取締役会は、企業戦略全体と整合性のある責任ある方法で経営陣が、これらのエージェントをどのようにトレーニングし、導入し、評価し、監視するつもりなのかを理解しておく必要があります。この点において、AIエージェントは 業績に対する期待が与えられ、訓練され、評価および監視される人間の従業員と似ています。AIエージェントを労働力の延長としてとらえ、人間の従業員と同じように業務に統合し、そのパフォーマンスを監視する企業は、この次世代技術を最適化する上で有利な立場にあると考えられます。

上表に示されているように、エージェント型AIは、高度な意思決定能力によって自動化を強化し、予測不可能な環境においても、継続的な人間の監視なしに適応し、学習することができます。特に、AIエージェントは、規則的なAIソリューションでは対応できない状況において、価値創造の機会をもたらします。その一例として、「リード（見込み顧客）の発掘と育成」といったケースでは、エージェントが個人とつながりアプローチするために創造的な手法を考える必要がありますが、実際にどのような手順を踏むかはそれほど重要ではありません。

一方で、「取引の照合作業」のように、タスクをどう実行するかの手順そのものが重視される業務では、より従来型の自動化やインテリジェントオートメーションの方が適していると考えられます。

この重要な違いを念頭に置きながら、エージェント型AIの価値創造の機会は、業界を超えて広がっています。その例としては、次のようなものがあります。

- **生産性と効率を高める**　製造、物流、サービスのいずれにおいても、AIエージェントは自律的に動作し、リアルタイムのデータと学習された経験に基づいて意思決定と行動を行い、新しい状況に適応し、時間の経過とともに継続的にパフォーマンスを向上させるように設計されています。

- **タスクやデータ量の増加に合わせて動的に拡張**　ビジネスが成長するにつれて、AIエージェントはさまざまな作業負荷に対応できるように業務を拡張し、人間の介入なしに新しいタスクに適応することができます。この柔軟性により、リソースを大幅に増やすことなく、多様で広範な方法で AI エージェントを導入することができます。

エージェント型AIは、高度な意思決定機能によって自動化を強化し、予測不可能な環境においても、人間の継続的な監視なしに適応し、学習することができます。

- **顧客サービスのスピード向上**　AIエージェントは、人間よりも迅速かつ正確に問い合わせに対応し、個別化された体験を大規模に提供することができます。この変革は、消費者向け産業における対応時間と解決率を向上させることができます。

- **意思決定の強化**　金融、医療、戦略策定においてエージェント型AIは、大量のデータをリアルタイムで分析し、隠れたパターンを発見し、重要な洞察を浮き彫りにすることで、よりスマートな情報に基づいた意思決定を促します。

- **製品とサービスの変革**　例えば、パーソナライズされたアドバイスや遠隔治療によって患者のケアに革命を起こすことは、医療業界に大きな変化をもたらすでしょう。

- **労働に新たな次元が加わる**　AIエージェントは、統合された労働力の一部として、正社員や派遣社員、請負業者とともに業務を遂行することで、コストを削減し、人的ミスを最小限に抑えるという変革をもたらします。

エージェント型AIによってビジネスプロセスを変革し、効率を高め、認識と反応に要する時間を短縮する可能性は基本的に無限です。例えば、サイバーセキュリティの分野では、AIエージェントが自律的にインシデントを監視し、対応します。ITサポートでは、自動でソフトウェアのアップデートを実行します。人事では、採用プロセスを効率化します。サプライチェーンマネジメントでは、AIによる物流の効率化が図られます。マーケティング戦略では、AIによる消費者行動の分析により、戦略の精度が向上します。

取締役が注目すべき理由：リスク

エージェント型AIの導入には、組織が検証すべき特定のリスクも伴います[1]。最新のトップリスク調査では、世界的な懸念事項のトップ10のひとつが、AIの導入による新たなリスクの出現でした。

この議論は、取締役にとって重要な意味があります。なぜなら、従来のオートメーションからエージェント型AIへの移行は、テクノロジーと人間の創意工夫が協力して目覚ましい成果を達成する未来への道を開く、重要かつ変革的な技術進化を意味するからです。しかし、その一方で課題も生じます。

以下は、取締役が考慮すべき、エージェント型AIの導入に関連する主なリスクです。一般的には他の形態のAIと類似していますが、エージェント型AIの特性や用途によって重要性が増す可能性があるものについては、以下が考えられます。

1　2025年のトップリスクに関するエグゼクティブの視点、プロティビティとNC州立大学のERMイニシアチブ、2025年2月：https://www.protiviti.com/jp-jp/survey/executive-perspectives-top-risks-2025

- **説明責任**　AIエージェントが危害や損害につながる決定を下したり、行為を行ったりした場合、その責任は誰にあるのでしょうか。ベンダー、開発者と連携して作業していた従業員であるトレーナー、もしくはユーザー、あるいはユーザーのマネージャーでしょうか。問題が発生した場合、その状況を分析し、何が問題だったのか、どこで問題が発生したのかを特定するためのプロセスはどのようなものになるのでしょうか。

- **バイアスの可能性**　雇用や融資を決定したり、法律や規制を執行したりする際、AIエージェントを訓練するために使用されるデータにバイアスがあると、不公平で差別的、さらには違法な結果につながる可能性があります。この問題は、AIエージェントがバイアスのある判断によって報酬を得るような設計になっていると、急速に拡大する恐れがあります[2]。バイアスのリスクは、コンプライアンスの問題も引き起こします。例えば、ニューヨーク市地方法144条は、雇用主がバイアスについて評価する年次独立監査を受けない限り、自動化された雇用決定ツールを使用することを禁じています。

- **データプライバシーとセキュリティ**　これらのリスクは、エージェント型AIシステムの動作方法、AIシステムがトレーニングに使用され依拠するデータ、およびユーザーや他のシステムとの相互作用から生じます。これらは、データ侵害によって機密情報が悪用されるリスク、またデータ漏洩のリスクをさらに深刻にします。

- **人材獲得競争**　エージェント型AIおよびその他の新興テクノロジーの採用により、AI、機械学習、データサイエンス、および関連分野の専門知識を持つプロフェッショナルに対する需要が高まっています。組織がエージェント型AIの可能性を認識するにつれ、こうしたシステムの開発と管理に長けた有能な人材を確保し定着させる必要があります。

これらのリスクに加え、エージェント型AIには、主にこれらのシステムの自律的な意思決定能力から生じる特有のリスクがいくつかあります。

- **制御の喪失**　エージェント型AIシステムがより複雑で自律的になる中、本来の目標とは矛盾した行動をとる可能性があります。操作の制限として定義されていない状況に直面したとき、AIエージェントはどのように行動するのでしょうか。 これらのシステムは適応と学習を繰り返しながら、人間の直接的な介入なしに動作します。信頼性と透明性が重要な状況では、無謀な行為のリスクが大きな問題となる可能性があります。

- **これまでとは異なるコラボレーション**　AIエージェントが従来は人間が担当していた業務を引き継ぐことで、新しい職務が生まれ、一部の職務は廃止されます。この移行には、AI エージェントとの効果的な連携と監督、そして新しく変革した職場環境への円滑な移行を確保するため、従業員のスキルアップと再教育に多大な努力が必要となります。エージェント型AI導入のユニークな点は、人間のエージェントとAIのエージェントが協働する職場を生み出し、これまでとはまったく異なる文化の管理を必要とすることです。セールスフォース社のCEOが最近ダボス会議で「これからは……人間の労働者だけでなく、デジタルワーカーも管理することになります[3]」と述べたように。

- **トレーニングにおける潜在的な利益相反の可能性**　AIエージェントによって業務が行われる人々は、エージェントのトレーニングを支援することが求められるかもしれません。このプロセスは、企業にとって管理が難しい課題を提示する可能性があります。具体的には、職の不安定化、変化への抵抗、責任の問題、トレーニングデータの偏りの可能性、そして士気の低下などが挙げられます。

- **経験的学習の喪失**　AIエージェントがより多くの仕事を引き継ぐようになると、人間が問題解決や意思決定の根底にある重要な思考に関与することで得られる経験的学習や本質的な知識が損なわれるリスクがあります。これは、長期的に見て労働力のレジリエンスにどのような影響を与えるのでしょうか。一方で、AIシステムは意思決定のプロセスを可視化し、記録として残すことができるため、その点では人間にとって有益なサポートとなる可能性もあります。

2　AIエージェントは、その目標に有益な行動に対して肯定的なフィードバックを受けると、強化学習を通じてその経験から学習し、その行動を繰り返すようになります。例えば、顧客サービスの対応では、人間の介入を必要とせずに顧客満足につながる対応が成功するたびに、「+1」のフィードバックが与えられます。肯定的なフィードバックが与えられると、AI エージェントは将来も同様の決定を行うよう指導されます。

3　"Today's CEOs are the last to manage all-human workforces, says Marc Benioff," by Anna Cooban, CNN Business, January 23, 2025: www.cnn.com/2025/01/23/business/davos-marc-benioff-salesforce-ai-prediction-intl/index.html

AIエージェントは自律的に動作し、人間の指示や継続的な介入なしに機能しますが、それでも人間の仕事と同じように、そのパフォーマンスに対する監督と、必要に応じた是正措置は人間が行わなければなりません。

このようなリスクを軽減し、エージェント型AIの責任ある導入を確実にするためには、強固なデータガバナンスの実践と、AIエージェントとその判断の定期的な監査とレビューが不可欠です。AIエージェントは、人間の従業員と同様に監督される「デジタル従業員」と捉えることが望ましいでしょう。

つまり、

- AIエージェントの職務とパフォーマンスの期待値を定義する際に、顧客重視の姿勢を採用する。
- エージェントの行動の中核となる価値観とガイドラインを明確にしたポリシーを設定する。
- AIエージェントの期待値に対するパフォーマンスの監視を容易にするために、測定基準と測定方法を導入する。
- 必要に応じて是正措置を取り、AIエージェントが継続的に学習し改善できるようにする。

AIエージェントは自律的に動作し、人間の指示や継続的な介入なしに機能しますが、それでも人間の仕事と同じように、そのパフォーマンスに対する監督と、必要に応じた是正措置は人間が行わなければなりません。したがって、AI エージェントは労働力の一部として不可欠なものとなります。

上述の議論は取締役にとって重要です。なぜなら、従来の自動化からエージェント型AIへの移行は、重大かつ変革的な技術進化を意味するからです。エージェント型AIは、より高度で効率的かつ自律的なビジネスプロセスを実現し、技術と人間の創意工夫が協力して驚くべき成果を達成する未来への道を切り開きます。しかし同時に、倫理的な利用の確保、データプライバシーの維持、従業員の移行管理といった課題も伴います。企業はこれらの課題を乗り越えて、エージェント型AIの持つ潜在能力を最大限に引き出さなければなりません。

取締役が検討すべき質問

エージェント型AIの導入に関して経営陣と戦略的な対話を行う場合、取締役会は以下の質問を検討する必要があります。

- AI技術の将来的な発展の可能性に備え、戦略を進化させながらビジネスにどのように適応させていくのか。
- エージェント型AIを組織に導入するための長期的なビジョンと、その導入が組織全体のAI戦略にどのように適合するのか。
  - エージェント型AIを使って、具体的にどのようなビジネス上の問題や機会に取り組んでいるか。
  - 逆に、私たちがエージェント型AIで対処しようとしているビジネス上の問題や機会は、そのテクノロジーに適しているか。
- エージェント型AIの導入は、カスタマーエクスペリエンスにどのような影響を与えるのか。これらのシステムによって、近い将来に何を達成しようとしているのか。また、これらのシステムおよびそれをサポートするベンダーは、既存のプロセスやテクノロジーとどのように統合され、スムースな適合性、拡張性、使いやすさを実現するのか。
- エージェント型AIシステムを効果的に設計、開発、管理するために必要な人材や専門知識が組織内に備わっているか。AI投資を支える価値提案を確実に実現できるよう、従業員の再教育とスキルアップのために、どのような研修および開発プログラムを実施しているか。
- 組織内におけるエージェント型AIの責任ある展開と利用を監督するために、どのようなガバナンス構造があるのか。
  - AIエージェントは、導入前に潜在的な問題を特定し、対処するための徹底的なトレーニングを受けており、

顧客が示した目標に沿った顧客の期待を確実に満たすことができるか。AIエージェントのトレーニングおよび運用には、どのような堅牢なデータ収集および管理システムを用いているのか。

- AIエージェントに起因する侵害や悪用から機密データを保護するために、どのような対策が講じられているか。データ保護規制および要求される機密データの保護措置を遵守していると確信できるか。その判断はどのように下しているか。

- 潜在的なバイアスや差別のリスク、その他の業務上、倫理上、風評リスクなど、エージェント型AIの導入に伴う潜在的なリスクを認識しているか。これらのリスクを軽減するための計画はどのようなものか。また、AIのパフォーマンスと行動を監視するために、どのようなセーフガードが設けられているのか。

- AIエージェントのパフォーマンスと学習、変化する顧客ニーズとの継続的な対応、社内ポリシーや適用される法規制の遵守について誰が責任を負うのか。AIエージェントの急速な進化と、影響が広範囲に及ぶことを考えると、有害な影響を最小限に抑えるために、逸脱した行動を迅速に特定するにはどうすればいいのか。

- どうやって成功を評価するのか。つまり、AIエージェントが与えられた役割を果たすうえでの成果や有効性、さらにはビジネスに与える影響を評価する上で、どのような指標を用いるのか。

- より複雑で機密性の高いやりとりに対して、人間の従業員が「セーフティネット」として介入すべきタイミングと方法に関する明確なガイドラインがあるか。

- エージェント型AIの導入について、その戦略的目的や期待される影響も含め、顧客、従業員、規制当局、その他の利害関係者にどのように伝えているか。どのようなフィードバックメカニズムがあるのか。

これらの質問は、取締役がエージェント型AI導入の影響をより明確に理解するための手助けとなります。また、経営陣がAIエージェントの導入について戦略的かつ顧客重視の責任あるアプローチを行っていることを取締役会が見守る上で役立ちます。

プロティビティの支援

AIはビジネスのやり方を急速に変えています。テクノロジー産業からヘルスケア、金融サービス、消費財まで、あらゆる業界において、企業はAI、インテリジェントオートメーション、そして高度な分析を採用し、プロセスを改善し、新たなビジネスチャンスを創出し、競争優位性の強化を図っています。

組織がこのような転換期を迎えるにあたり、私たちは、取締役会や経営幹部のニーズに合わせて、以下のようなさまざまな支援を提供しています。

- 取締役会教育セッションの運営
- AI活用機会の優先順位付けを支援するファシリテーションおよびデザインシンキングセッションの主導
- AI戦略の全体的な策定、およびその戦略に基づく実行と構築
- AIガバナンスおよびポリシーのレビューとアドバイス
- カスタマーエクスペリエンスの向上、業務効率、スピード、信頼性を高めるデータ駆動型ソリューションの導入支援
- サードパーティリスクおよびコンプライアンス要件の特定と管理の支援

また、大規模な組織が生成AIプラットフォームをエンドツーエンド（テクノロジー、人材、プロセス）で評価し、安心して導入できるよう支援しています。

protiviti®