

2026 年 1 月 28 日 全 7 頁

# AI の社会実装と加速するインフラ投資

## 2025 年の AI 動向と 2026 年の展望

経済調査部 主任研究員 田邊 美穂

### [要約]

- 2025 年は、AI が人間と協働する第一歩を踏み出した年となった。技術面では、AI モデルの高度化およびマルチモーダル化の進展で、より複雑なタスクの処理が可能になり、扱える情報の種類が増えるなど、AI が対応できるタスクの幅が広がった。さらに、軽量でも高性能な AI モデルが登場し、企業導入時の選択肢も拡大している。こうした AI モデルの性能向上を背景に、AI エージェントやフィジカル AI といった実働型の領域でも、社会実装を見据えた実証実験が活発化した。
- 2026 年以降は、この AI エージェントやフィジカル AI の社会実装がさらに進むことが期待される。それぞれ権限管理やセキュリティ、制度整備などに課題はあるが、まずはそれらの影響が比較的小さい定型的な業務や、工場などのクローズドな環境から実装が進むと考えられる。2026 年は、経済全体が“AI を前提とした産業構造”へ移行していく契機の年になると見込まれる。
- AI 企業各社によるデータセンターなどの AI インフラへの投資は年々増加しており、一部では過剰投資との声もあるが、2026 年もこの傾向は続く見込みだ。技術面での効率化は進むものの、それ以上に AI モデルの性能向上が上回っており、計算資源が逼迫している。そのため、直ちに過剰投資とは言い難い状況だ。一方で、収益面では企業による AI 導入が試行段階にあり、収益化まで時間を要する可能性が高く、動向を注視する必要がある。
- 収益化は AI モデルの性能だけでなく、企業での導入がどこまで広がり定着するかにも左右されると考えられる。当面は、導入領域の拡大や利用実績を慎重に見極め、短期的な話題ではなく長期的な視点で評価することが重要である。特化型 AI モデルの活用も進み、用途に応じた最適化との併用が広がることも想定される。技術進展やインフラ整備に加え、企業導入の定着度が持続的な需要につながるかが AI 市場成長の鍵となるだろう。

## 1. はじめに

2025 年は DeepSeek ショックをきっかけに、高い性能を持つオープン型 AI モデルが台頭し、生成 AI モデル市場の多極化が進んだ。また、OpenAI や Google といったフロンティアモデル<sup>1</sup>を有する企業の性能競争は激化しており、同年 12 月に OpenAI が「コードレッド（非常事態宣言）」を発令したと報じられた<sup>2</sup>。

こうした開発競争の加速を背景に、OpenAI の Stargate Project<sup>3</sup>など、主要企業はデータセンターなどの AI インフラへの投資を急速に拡大している。一方で、その投資規模の大きさに対し、収益化の見通しはなお不確実であり、過剰投資を懸念する指摘も強まっている。

本レポートでは、2025 年の AI 動向を振り返るとともに、2026 年の AI 動向および AI 投資の現状とリスクについて整理する。

## 2. 2025 年の AI 動向と 2026 年の見通し

### 2025 年の AI 動向①：AI がより複雑で高度なタスクをこなせる時代に

2025 年は AI が人間と協働する第一歩を踏み出した年となった（図表 1）。生成 AI の技術的進展における主要なキーワードとしては、**AI モデルの高度化**、**マルチモーダル化**、**AI モデルの軽量化**が挙げられる。

従来は質問応答や要約などの静的な処理が中心だったが、推論能力やリアルタイム性の向上により、状況を把握して解決策を提示するなど、より複雑で高度なタスクへと対応範囲が広がっている（**AI モデルの高度化**）。さらに、テキストに加え、画像や動画、音声まで扱えるマルチモーダル AI の性能も大幅に向上した。その結果、AI が扱える情報の種類が広がり、対応できるタスクの範囲も拡大している（**マルチモーダル化**）。

さらに、軽量でも高性能な AI モデルも多く登場した。これにより、応答速度やコスト効率、設置形態（クラウド／オンプレミス／エッジなど<sup>4</sup>）に応じて最適構成を組むことが可能となり、企業にとって導入時の選択肢が大幅に拡大している（**AI モデルの軽量化**）。

### 2025 年の AI 動向②：AI の社会実装に向けた実証実験が活発化

また、先述の AI モデルの進展により、**AI エージェント**や**フィジカル AI** といった”実働型”の

---

<sup>1</sup> フロンティアモデルとは、各社が開発する最高性能クラスの大規模 AI モデルの総称で、AI 性能の上限（フロンティア）を押し広げる基盤モデルを指す。

<sup>2</sup> 日本経済新聞「[OpenAI、対 Google『非常事態』宣言と米報道 3 年前と形勢逆転](#)」（2025 年 12 月 2 日）

<sup>3</sup> OpenAI「[Stargate Project を発表](#)」（2025 年 1 月 21 日）

<sup>4</sup> オンプレミスとは、企業内に設置したサーバーで処理を行う方式を指し、セキュリティやカスタマイズ性に強みがある。エッジとは、データ発生場所に近い機器や端末側で処理する方式を指し、低遅延でリアルタイム処理に適する。

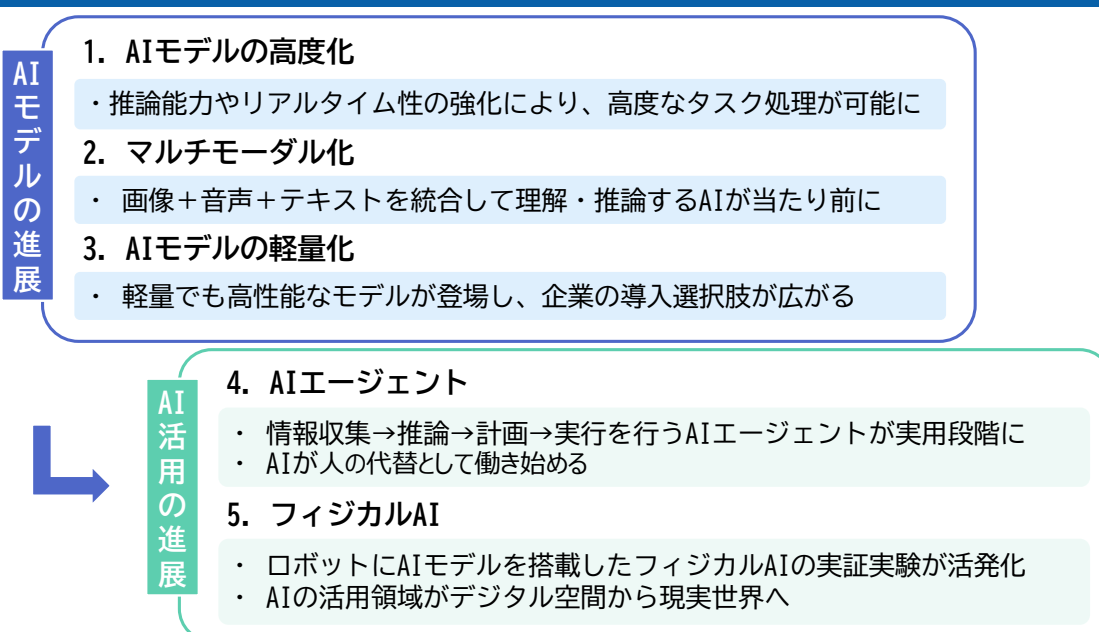
---

技術領域においても、社会実装を見据えた動きが加速した。

**AI エージェント**は、情報収集・推論・判断・実行を連続して行うことで、従来は人間が担ってきた一連の作業をAI のみで完結させることを可能にする。これらは、推論精度やリアルタイム性の向上により、一連の業務プロセスをAI が自律的にこなす環境が整いつつあり、バックオフィスやカスタマー対応といった定型的な業務領域から導入が広がりつつある。

また、ロボットに AI モデルを組み込む**フィジカル AI** の領域でも、自動運転や人型ロボット等を中心に安全性評価や実環境でのデータ取得を目的とした実証実験が多く行われた<sup>5</sup>。これらは、モバイル環境でも動作する軽量型 AI モデルの普及や、センサー・ロボティクス技術の進展が後押ししている。AI がデジタル空間にとどまらず、現実世界のタスクに踏み込むための基盤が整ってきたと言えるだろう。

図表 1 2025 年の AI 動向



(出所) 各種資料より大和総研作成

## 2026 年の AI 動向：一部領域・特定業務から AI の社会実装が進む

2026 年以降は、この AI エージェントやフィジカル AI の社会実装がさらに進むことが期待される。しかし、これらの技術にはそれぞれ固有の課題も残されている。

AI エージェントでは、AI 自体が企業の業務プロセスに深く入り込み、複数のシステムにまたがって業務を実行することが求められる。そのため、AI に与える権限の精査および管理に加え、誤作動や情報漏洩を防ぎ、異常を検知するためのセキュリティ対策や、責任追跡に不可欠な監査ログの確保が今まで以上に重要となる。また、どこまでの判断・実行をAI に任せるのかとい

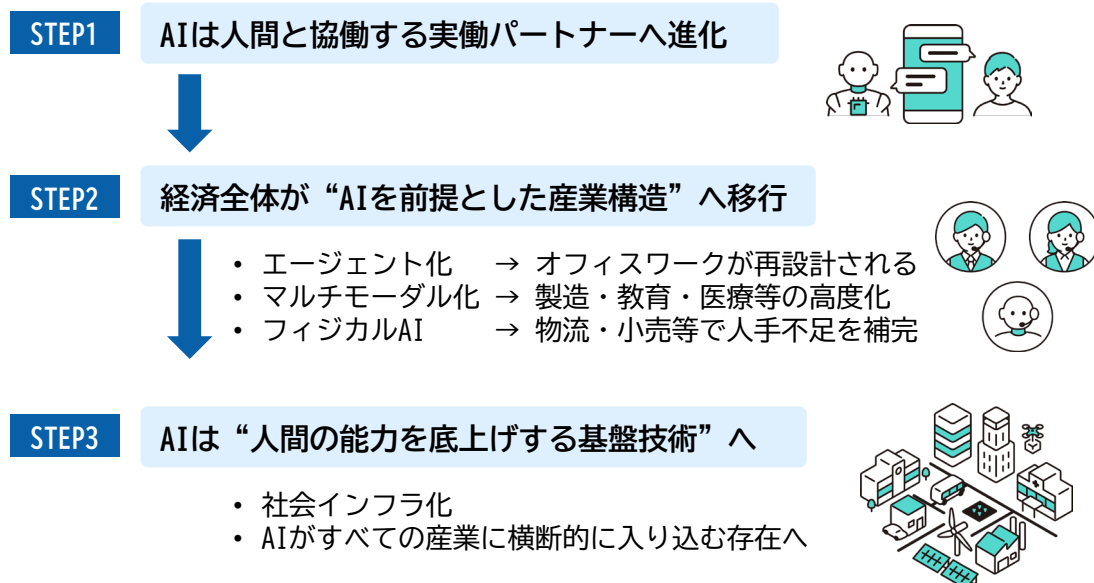
<sup>5</sup> フィジカル AI の 2025 年の動向についての詳細は、田邊美穂「[フィジカル AI の進展で注目の人型ロボット：日本はロボット技術の優位性を武器に AI 分野で存在感を示せるか](#)」大和総研レポート（2025 年 9 月 1 日）を参考にされたい。

う業務プロセスの切り分けだけでなく、AI が利用しやすいようにシステムや業務フローを再設計する動きも出てくるだろう。こうした取り組みは、バックオフィスの定型処理や情報整理など、既に AI 活用が広がりつつある領域から進むと考えられる。

フィジカル AI では、現実世界での稼働を前提とするため、法規制、安全基準、責任の所在といった技術面以外の課題も多い。そのため、社会実装を進めるにあたり、まず政府主導で関連法制度や安全基準、責任分担の枠組みを整備し、実証環境を拡充していくことが重要となるだろう。加えて、各国の基準や規制が異なることから、相互運用性や越境での運用を見据えた国際的な調整も求められる。こうした取り組みは早急に進める必要があるが、当面は工場や倉庫、構内搬送など、制約の影響が比較的小さく管理しやすいクローズドな環境での実装が先行すると考えられる。

総じて、課題は明確化されつつあり、AI モデルの性能や業務設計の面で実装環境が整うにつれて、社会実装は段階的に前に進む可能性が高い。**図表 2** で示したように、STEP1 で求められる実働パートナーとして、AI は実務運用に堪える水準へ接近しており、今後も向上が期待される。加えて、2026 年はこの基盤を踏まえ、STEP2 にあたる経済全体が“AI を前提とした産業構造”へ移行する契機の年になり、まずは限られた範囲から徐々に進むと見込まれる。

図表 2 2026 年は“AI を前提とした産業構造”へ移行する契機の年に



(出所) 各種資料より大和総研作成 (イラストはソコスト (<https://soco-st.com/>))

### 3. 増加する AI インフラへの投資と過剰投資の懸念

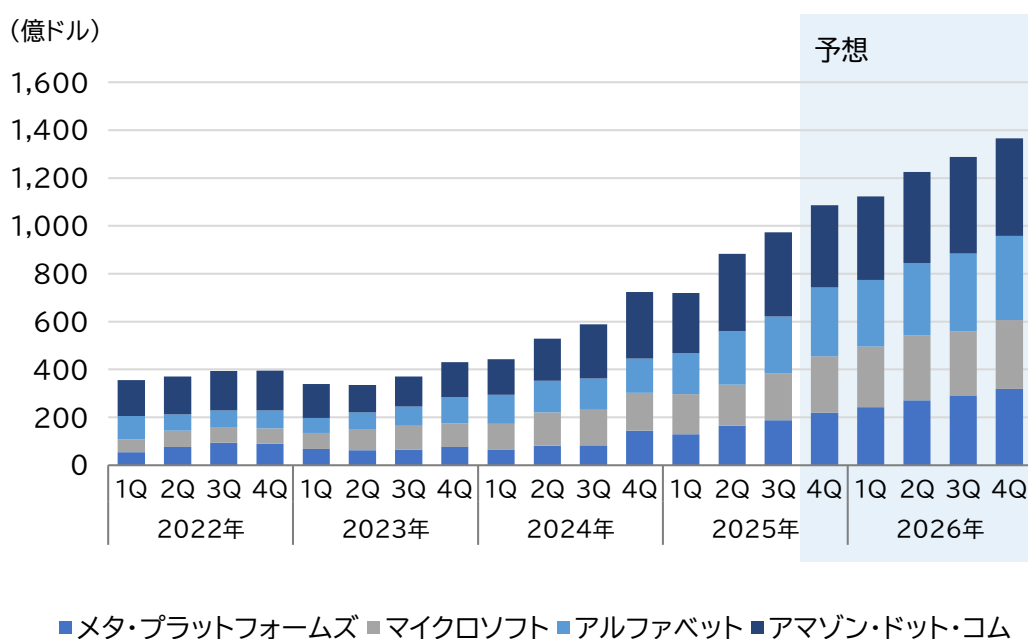
#### 2026 年も AI 関連企業各社のインフラへの投資は続く

2025 年は、AI 関連企業各社がデータセンターの建設を発表するなど、AI インフラへの大規模な投資が相次いだ。四半期ベースで見ても設備投資額は右肩上がりとなっており、2026 年もこの傾向は続くと思われる (**図表 3**)。こうした AI インフラへの投資姿勢に対しては、収益化までの時間差や資金負担を懸念する声があり、市場では過剰投資ではないかという議論が徐々

に高まっている。

こうした見方が生まれる背景には、AI インフラが長期的な利益を見据えた先行投資になりやすい構造がある。AI モデルの開発や運用には膨大な計算量が不可欠であり、その確保にはデータセンターなどのインフラ整備が前提となる。データセンターには着工から稼働までのリードタイムがあり、立地・電力・水資源の確保といった物理的制約も多い。そのため、需要の立ち上がりを待たずに基盤を整える投資が先行しやすい。

図表 3 AI 関連企業各社の設備投資額の推移



(注) 2025 年 4Q 以降は Bloomberg の予測値

(出所) Bloomberg より大和総研作成

### 足元の投資は必要性が勝り、過剰投資との判断は難しい

この前提を踏まえ、AI インフラへの投資状況について、過剰投資に関する見方も含めて、供給・技術・収益の三つの側面から整理する（図表 4）。

現状、**供給面（必要量に対して投資が過剰な可能性）**に関しては、現状、必要な計算量が不足しているため、各社はその不足を補うべくデータセンターの増強を進めており、投資の先行は当面の運用上も一定の合理性があると考えられる。AI 需要が急増する中、主要クラウド事業者では計算資源<sup>6</sup>の逼迫が課題として浮上している。また、AI モデルのスケーリング則<sup>7</sup>は一部で収穫逦減の指摘はあるものの依然として有効であり、AI モデルの学習に要する計算量の拡大も続いている。さらに AI の推論や、マルチモーダル化による動画・画像生成は、テキスト生成

<sup>6</sup> 計算資源とは、AI の学習・推論に必要な計算処理（GPU 等）・ストレージ・ネットワークといった基盤技術を指し、データセンターはこれらの計算資源を稼働させるための物理的な基盤（稼働基盤）にあたる。

<sup>7</sup> スケーリング則は、学習データの量・計算リソース・モデルのパラメーター数を増やすことで AI モデルの性能が向上するという経験則。



と比べてはるかに多くの計算量を要し、計算資源の逼迫を一段と強めている。

**技術面（技術の進化が投資に追いつかない可能性）**に関しては、技術の進化と効率化の両面から捉える必要がある。まず進化の側面では、フィジカルAIといった推論・計画・行動できるAIが目されるなど、AIの活動範囲は拡大を続けている。この結果、AIが扱う情報は一段と多様化・増加しており、研究開発から社会実装に至るまでに必要となる計算量は今後も増加が見込まれる<sup>8</sup>。

効率化の側面では、AIモデルの学習や推論にかかる計算量などの効率化に対する取り組みが、ハード・ソフトの両面で進められている。これにより、以前と比較して少ない計算量で学習や運用を行える環境が整いつつある。ただし現状は、AIモデルの高度化や活動範囲の拡大に伴い必要となる計算量の増加のスピードが効率化の効果を上回っており、効率化は投資判断を転換させるほどの抑制要因とは現時点では言い難い。このため、当面は必要計算量の増勢に対応する“量の確保”を優先しつつ、効率化の進展を見極める局面にある。

一方で、**収益面（収益に対して投資額が多すぎる可能性）**に関しては、動向を注意深く見守る必要がある。現状、企業のAI導入は試行が中心であり、本格的な需要の顕在化には一定の時間が見込まれる。そのため、収益化までには数年単位の期間を要する可能性が高いだろう。AIEージェントやフィジカルAIは実用的な使用例を想定しやすいものの、普及時期の見通しは流動的である。もっとも、収益化が見えてくれば過剰投資への懸念が落ち着く可能性はある。当面はその移行の足並みを慎重に見定めたい。逆に、収益化の立ち上がりが想定より遅れる場合には、OpenAIやNVIDIAなどのAI関連企業間で資金が行き来する投資の相互依存（循環投資）が資本効率を一時的に圧迫する可能性にも留意が必要である。

図表4 AIインフラ投資の過剰投資リスク



（出所）各種資料より大和総研作成（イラストはソコスト（<https://soco-st.com/>））

<sup>8</sup> もっとも、現状は実証段階が中心であり、今後技術的課題に直面する可能性には留意が必要である。

まとめると、足元の投資は必要性が勝っていると考えられ、直ちに過剰と判断するのは難しい。一方で、収益化の立ち上がりには注意が求められる。ただし、これらの前提は投資の対象となっている物理インフラが確保されることが条件となる。データセンターには、立地・電力・水資源、系統接続などの物理的制約があり、とりわけ電力・系統接続についてはボトルネックになりやすい。実際、データセンター竣工後に電力確保が間に合わず稼働が遅延する事例も見られており、こちらも注意が必要である。

#### 4. AI の企業導入がどこまで定着するかが鍵に

2026 年は、AI が本格的な社会実装に向けて着実に基盤を固める移行期となる。AI モデルの高度化やマルチモーダル化に加え、AI エージェントやフィジカル AI といった実働領域でも環境整備が進み、産業構造そのものが AI を前提とした姿へと変化していくことが期待される。ただし、こうした構造変化は一朝一夕に進むものではなく、本格的な需要の立ち上がりには時間を要するとみられる。

OpenAI や Google による汎用型 AI モデルの性能競争は当面続くとみられ、性能向上には継続的な計算資源やデータセンターなどへのインフラ投資が不可欠となる。もっとも、AI は単一の技術革新で状況が大きく変化する可能性がある点は、引き続き認識しておく必要がある。

一方で、収益化は AI モデルの性能だけで決まるわけではなく、企業での AI 導入がどこまで広がり定着するかにも左右されると考えられる。当面は、導入領域の拡大や利用実績の積み上げを丁寧に検証し、短期的な話題性に過度に影響されることなく、長期視点から動向を見定めることが重要になるだろう。

導入領域の広がりを見ていくうえでは、用途・業界に特化した AI モデルの動向にも注目したい。医療や金融分野など、専門性と精度および説明責任が求められる領域、ならびにソフトウェア開発分野におけるコーディングスキル、研究開発分野での数学的解析やシミュレーション技術など、高度な専門能力が要求される分野では、特化型 AI モデルの採用が進んでいくことも想定される。今後は、汎用型 AI モデルを基盤としつつ、用途別に最適化された AI モデルの活用が並行して進む可能性は十分に考えられる。

このように、当面は AI 開発企業の動向を注視しつつ、短期の流行に過度に影響されず、長期的な視点から AI 産業の進化とリスクを適切に評価していくことが重要になる。最終的には、技術進展やインフラ整備の進捗だけでなく、企業導入がどこまで定着し、持続的な需要へとつながるかが、AI 市場の成長の鍵を握るだろう。

以上