レベル 3 AI エージェントにおける LLM 層とアプリケーション層の役割：Genspark 、Manus、OpenAI Operator 、Deep Research の事例分析

Gemini Deep Resear

1. はじめに：レベル 3 AI エージェントとそのコアアーキテクチャの定義

レベル 3「自律型 AI」エージェントの定義

AI エージェントの進化は目覚ましく、特に「レベル 3」に分類される自律型 AI エージェントは、ユーザーの代理として長期間（場合によっては数日間）にわたり自律的にタスクを実行する能力によって特徴づけられる [1]。これには、人間による常時介入なしに、独立した意思決定、変化する状況への適応、そして定義された目標の持続的な追求が含まれる [1]。OpenAI のフレームワークでは、「エージェント」（レベル 3）は「リーズナー」（レベル 2）の次に位置づけられており、これは効果的な自律的行動のためには堅牢な推論能力が前提条件であることを示唆している [5]。この階層的な進展は、レベル 3 エージェントが初期レベルの対話能力や推論能力の上に構築されることを示している。これらのエージェントは単に受動的に反応するのではなく、目標指向の行動、計画、実行を示す [10]。

基本的なアーキテクチャの二分法：LLM 層とアプリケーション層

レベル 3 を含む全ての先進的な AI エージェントは、大規模言語モデル（LLM）層が認知機能を提供し、アプリケーション層が相互作用と実行の手段を提供するという階層型アーキテクチャを通じて理解することができる [11]。LLM は「脳」または「心」として機能し、アプリケーション層はエージェントが環境を認識し、その中で行動することを可能にする「身体」または「手」として機能する [17]。これら 2 つの層間の相乗効果は、レベル 3 エージェントに特徴的な自律性とタスクの複雑性を達成するために不可欠である。

LLM の進化とアプリケーション層の高度化の相互依存性（レベル 3 における考察）

レベル 3 エージェントに記述される能力（例えば、「数日間」にわたる操作 [1]）は、推論と計画のための強力な LLM だけでなく、持続的な状態管理、信頼性の高いツール使用、エラー回復、および長期的な目標の一貫性のための高度に洗練されたアプリケーション層をも必要とする。この点を深く考察すると、レベル 3 エージェントの真の能力

は、これら二つの層が互いに依存し、共に進化することによってのみ実現可能であることが明らかになる。

まず、レベル 3 エージェントは長期間自律的に動作する必要がある。長期間の動作は、中断への対処、セッションを跨いだコンテキストの維持、変化する状況への適応を意味する。LLM は推論能力に優れているものの、その生の形態ではしばしばステートレスであるか、限定的なコンテキストウィンドウしか持たない[19]。したがって、アプリケーション層は、LLM の計画能力を補完するために、長期記憶、状態管理、および持続的な目標追跡のための堅牢なメカニズムを提供しなければならない。さらに、LLM はこれらのアプリケーション層コンポーネントを効果的に指示し制御できる必要があり、アプリケーション層はこれらの指示を確実に実行し、フィードバックを提供する必要がある。

このことから、真のレベル 3 エージェントの開発は、LLM 開発の課題であると同時に、アプリケーションエンジニアリング（堅牢で、ステートフルで、インタラクティブなシステムの構築）の課題でもあると言える。LLM 層の高度化（例：高度な推論、ツール使用能力）とアプリケーション層の高度化（例：複雑なワークフローオーケストレーション、信頼性の高いマルチツール統合、永続メモリ）は、共進化しなければならないのである。

2. レベル 3 AI エージェントにおける LLM 層の役割：認知エンジン

LLM 層は、レベル 3 AI エージェントの「認知エンジン」として機能し、人間のような理解、推論、計画、意思決定能力を提供する。この層がエージェントの知的行動の核心を担う。

中核的認知機能

- **自然言語理解（NLU）：** ユーザーの複雑でしばしば曖昧な目標や指示を解釈する[20]。これは、潜在的に長期間にわたるタスクに対するユーザーの意図を理解するための基礎となる。
- **推論と問題分解：** 高レベルの目標をより小さく管理可能なサブタスクに分解し、全体的な目標を達成するための段階的な計画を策定する[11]。レベル 3 エージェントの場合、この計画は数日間にわたる実行の可能性を考慮しなければならない。
- **意思決定：** 現在のコンテキスト、利用可能な情報、および全体的な目標に基づいて、計画の各ステップで適切な戦略、ツール、またはアクションを選択する[29]。これには、明確化を求めるか、人間の介入を求めるかを決定することも含まれる。
- **知識統合と統合：** 多様な情報源（内部知識、検索された文書、ウェブ検索結果）

から情報にアクセスし統合して、計画と実行に役立てる[20]。

ツール呼び出しとオーケストレーションロジック

- 外部ツール（例：ウェブブラウザ、コードインタプリタ、API）の必要性を特定し、それらの使用に適したパラメータを策定する[10]。
- レベル 3 エージェントの LLM は、しばしばツール呼び出しのシーケンスを管理し、ワークフローを効果的にオーケストレーションする[32]。OpenAI の o3 や o4-mini のような例は、ツールをいつ、どのようにエージェント的に使用するかについて推論するように訓練されている[24]。

適応性と学習（現在のシステムでは主にファインチューニングとプロンプトエンジニアリングによる）

- 真の継続的学習は進行中の研究分野であるが[37]、現在の LLM は特定のタスクに関するファインチューニングや、動的な状況での行動を導く洗練されたプロンプトエンジニアリングによって適応させることができる。
- 「遭遇した情報に応じて必要に応じて反応し、方向転換する」能力[24]は、レベル 3 エージェントにとって重要な LLM 駆動の能力である。

LLM の「動的スクリプター」としての役割（レベル 3 エージェントにおける考察）

数日間にわたる複雑なタスク向けに設計されたレベル 3 エージェントにおいて、LLM は単一のコマンドを発行するだけでなく、アプリケーション層が実行するための一連の動的な指示、すなわち「スクリプト」を生成することが多い。このスクリプトは、中間結果に基づいて修正され得る。この関係性を考察すると、LLM の役割が単なるプランナーを超えたものであることが見えてくる。

レベル 3 のタスクは、単一の事前定義されたワークフローでは処理するには複雑すぎることが多い。LLM の強みは、新しい状況に対応して推論し計画する能力にある。アプリケーション層は、実行可能な一連の能力（ツール、API）を提供する。複雑で長期的な目標を達成するために、LLM は、*どの*アプリケーション層の能力を、*どのような順序*で、そして*どのようなパラメータ*で使用するかを継続的に決定しなければならない。この一連の決定とツールの呼び出しは、LLM がリアルタイムまたは複数のステップで生成および改良する動的なスクリプトを効果的に形成する。アプリケーション層は、これらの「スクリプト化された」アクションを実行し、結果を返し、LLM はそれを使用して計画を更新し、スクリプトの次の部分を生成する。

したがって、レベル 3 エージェントの LLM 層は、単なるプランナーではなく、タスクの期間中、アプリケーション層のアクションを導く継続的な「ディレクター」または

「動的スクリプター」であると言える。これには、2 つの層間の堅牢な通信とフィードバックループが必要となる。

3. レベル 3 AI エージェントにおけるアプリケーション層の役割：実行と対話のフレームワーク

アプリケーション層は、LLM 層によって策定された計画と指示を現実世界のアクションに変換し、エージェントの自律的な運用を支える実行基盤および対話インターフェースを提供する。

LLM 駆動アクションのためのインターフェース

- LLM 層が外部世界と対話するために呼び出すことができる具体的なツールと機能を提供する [11]。これには、ウェブ検索、データベースアクセス、コード実行、ソフトウェア GUI との対話、あるいはロボットエージェントの場合は物理的なアクチュエータのための API が含まれる。
- LLM の抽象的な計画やツール使用の決定を実行可能なコマンドに変換する [13]。

ツール実行と環境対話

- LLM によって選択されたツールの実際の実行を管理する（例：Python コードの実行、API 呼び出し、ブラウザの制御）[10]。
- 様々なデジタル環境（ウェブサイト、オペレーティングシステム、ソフトウェアアプリケーション）および潜在的に物理的環境との対話を処理する。

状態管理と長期記憶

- 数日間にわたって動作するレベル 3 エージェントにとって極めて重要であり、アプリケーション層はしばしばタスクの状態、会話履歴、および蓄積された知識を維持する責任を負う [17]。これには、データベース、ファイルシステム、または特殊なメモリアーキテクチャが含まれる場合がある。
- Retrieval Augmented Generation（RAG）のようなメカニズムはアプリケーション層の一部であり得、LLM に長期記憶からの関連コンテキストを提供する [35]。
- 永続メモリとコンテキストの一貫性は、拡張操作にとって鍵となる [52]。

通信プロトコルとオーケストレーション

- LLM と様々なツール/サービス間の通信を管理する [14]。Model Context Protocol (MCP) [19] や Agent-to-Agent (A2A) [62] のようなプロトコルは、これらの対話を標準化することができる。

- マルチエージェントシステム（Genspark や Manus など）では、アプリケーション層には、異なる特化型エージェントや LLM 間のワークフローと通信を管理するオーケストレーションコンポーネントが含まれる場合がある [17]。

ユーザーインターフェース（UI）とユーザーインタラクション（UX）

- ユーザーがタスクを開始し、進捗を監視し、フィードバックを提供し、必要に応じて介入する手段を提供する [39]。
- Operator のようなエージェントの場合、UI にはブラウザでのエージェントのアクションの観察や「テイクオーバーモード」が含まれることがある [45]。

アプリケーション層の「現実アンカー」および「永続エンジン」としての役割（レベル 3 エージェントにおける考察）

LLM が推論と計画を提供する一方で、アプリケーション層はこれらの認知プロセスを現実世界（またはデジタル世界）に接地させ、数日間にわたるタスクに必要な継続性を保証する。堅牢なアプリケーション層がなければ、LLM の計画は抽象的なままであり、その状態は一時的なものとなる。この関係性を深く掘り下げると、アプリケーション層がエージェントの自律性と持続性にとって不可欠な役割を担っていることがわかる。

LLM は、その訓練データと現在のプロンプトに基づいて計画と指示を生成する。これらの計画は、特定の環境（例：ウェブブラウザ、ファイルシステム、API）で実行される必要がある。アプリケーション層は、そのためのツールとインターフェースを提供する。これが「現実アンカー」機能である。レベル 3 のタスクは長期間に及び、エージェントは過去のアクション、中間結果、および複数のセッションや中断を跨いでの全体的な目標を記憶する必要がある。LLM 自体は、システムの再起動や長時間の休止を跨いで、この長期的で永続的な記憶を本質的に持っていない場合がある。アプリケーション層は、データベース、ファイルストレージ、または特殊なメモリモジュール（例：RAG 用のベクトルストア、MCP におけるステートフルセッション管理 [54]）を通じて、この「永続エンジン」を提供する。また、長期間実行されるタスクの実務（エラー回復（例：失敗した API 呼び出しの再試行）、ロギング、潜在的なリソース管理）も処理する [10]。

したがって、レベル 3 エージェントの信頼性、堅牢性、および真の自律性は、そのアプリケーション層の設計と能力に大きく依存する。LLM が「賢い」だけでは不十分であり、アプリケーション層は複雑な環境での持続的でステートフルかつ回復力のある運用をサポートするように設計されなければならない。LLM 層での「幻覚」のような課題は、アプリケーション層が検証や安全チェックなしに盲目的に実行する場合、特に危

険となる[42]。

4. ケーススタディ：レベル 3 自律運用のための LLM とアプリケーション層の連携

このセクションでは、Genspark Super Agent、Manus AI Agent、OpenAI Operator、および OpenAI Deep Research の 4 つの具体的な AI エージェントを取り上げ、それぞれの LLM 層とアプリケーション層のコンポーネント、そしてそれらがどのように相互作用してレベル 3 の特性、特に長期間の自律的なタスク完了能力を実現しているかを詳細に分析する。

4.1 Genspark Super Agent

Genspark Super Agent は、その「Mixture-of-Agents」（MoA）アーキテクチャと広範なツールセットにより、複雑なタスクの自律的処理を目指すレベル 3 エージェントの一例である。

- **LLM 層（認知エンジン）：**
  - **Mixture-of-Agents (MoA) アーキテクチャ：** 9 つの異なる LLM を利用し、タスクやサブタスクを複雑性、速度、精度要件に基づいて最適な LLM にインテリジェントにルーティングする[63]。これにより、複雑で長期的な目標のさまざまな側面に対して特化した推論が可能となる。
  - **自律的計画とタスク分解：** Genspark は人間のアシスタントのように「考え、計画し、行動し、ツールを使用する」ことができ、高レベルの指示を実行可能な計画に分解する[87]。この計画は、特化した LLM の 1 つ以上で行われる可能性が高い。
  - **自然言語理解：** 複雑で潜在的に数日間にわたるプロジェクトを開始するために、自然言語で与えられたユーザーの目標を処理する。
- **アプリケーション層（実行と対話のフレームワーク）：**
  - **広範なツールセット：** マルチメディア作成（ビデオ、ウェブサイト、プレゼンテーション）、データ分析、現実世界の対話など、多様な操作のための 80 以上の内製ツールを備える[63]。
  - **直接 API 統合：** ブラウザ専用エージェントとは異なり、Genspark は構造化された迅速なデータ検索のために API を直接呼び出すことができ、これは長期間実行されるタスクの効率にとって不可欠である[63]。
  - **リアルタイム音声自動化：** 予約や在庫確認などのタスクのために、AI 生成音声を使用して実際の電話をかけることができ、純粋にデジタルな領域を超えて対話能力を拡張する[81]。

    - **Model Context Protocol (MCP)** ：複数の LLM と多様なツール間のコンテキストと通信を管理するために、MCP または類似のプロトコルを使用している可能性が高い[19]。MCP は、エージェントとワークフロー間でステートフルで永続的なメモリを可能にし、これは長期間のタスクにとって不可欠である[54]。MCP のコンテキストオブジェクトは、静的知識、動的入力、タスク状態、およびエージェントの対話を保存し、タスクの再開やエージェントが以前の作業に基づいて構築することを可能にする[54]。
    - **マルチタスクと潜在的な非同期操作：** ユーザーは複数の AI スーパーエージェントプロジェクトを異なるタブで一度に実行できる[81]。この記述では明示的に「数日間」とは言及されていないが、複数の複雑なタスクを同時に処理するためのアーキテクチャは、長期的な単一タスクに対応可能な堅牢なアプリケーション層を示唆している。
    - **ファクトチェック機能：** 信頼できる情報源に対してデータを検証するアプリケーション層の機能で、信頼性の層を追加する[81]。
- **レベル 3 能力のための相乗効果（数日間の運用に焦点）：**
    - MoA アーキテクチャ（LLM 層）は、長期タスクのサブ問題を最も適切な LLM にルーティングすることにより、持続的な推論と動的な戦略調整を可能にする。
    - アプリケーション層の豊富なツールセットと直接 API アクセスにより、数日間にわたって必要な多様なアクションの実行が可能になる。
    - MCP（アプリケーション層インフラストラクチャ）は、複数の LLM とツール間で長期間にわたり目標の一貫性と状態を維持するために不可欠である。これにより、システムは進捗、中間結果、および変化するコンテキストを「記憶」することができ、これは「新製品の包括的な市場分析と発売計画を作成する」といったタスクが数日間にわたる場合に不可欠である。
    - Genspark が単一のタスクを数日間にわたって人間の介入なしに管理するという直接的な証拠は提供された資料からは限定的であるが[83]、そのアーキテクチャ（MoA、MCP、広範なツール）は、そのような拡張された自律性を目指した設計であることを強く示唆している。「自律的に考え、計画し、行動し、ツールを使用する」能力[87]と MCP を介した永続的なコンテキスト[54]の組み合わせは、この方向性を示している。
- Genspark の MoA と MCP がスケーラブルで長期間の自律性を可能にするという考察：

  Genspark が Mixture-of-Agents（LLM 層）と MCP のようなプロトコル（アプリケーション層）を明示的に使用していることは、単一のモノリシックな LLM では対応が困難な、非常に複雑で潜在的に長期間にわたるタスクを処理するための意図

的なアーキテクチャ上の選択である。この構造は、レベル 3 エージェントの能力を拡張する上で重要な意味を持つ。
複雑な数日間のタスクには、多様な推論能力（例：創造的なコンテンツ生成、論理分析、データ検索）が必要である。単一の LLM は、非常に大規模なものであっても、これらすべての多様なサブタスクにわたって最適に熟達しているか、効率的であるとは限らない。Genspark の MoA は、サブタスクを特化した LLM に委任することを可能にし、より大きな目標の各部分についてパフォーマンスとリソース使用を最適化する [82]。このような分散型認知システムが長期的な目標に対して一貫して機能するためには、堅牢なコンテキスト管理とエージェント間通信が最も重要である。MCP は、異なる LLM とアプリケーションコンポーネント間でコンテキスト、ツール、および状態を共有するためのこの標準化されたフレームワークを提供する [54]。これにより、「永続メモリ」と「ステートフルワークフロー」が可能になる [54]。
したがって、MoA と MCP の組み合わせは、長期間にわたって一貫性を維持し、適応し、確実に実行できるエージェントを構築するための戦略的アプローチであり、これはレベル 3 の特性である。このアーキテクチャは、よりスケーラブルで堅牢なレベル 3 以上のエージェントへの道筋を示唆しており、アプリケーション層は多数の特化型 LLM 駆動認知サービスのための洗練されたオーケストレーションおよびメモリプラットフォームとなる。その場合の課題は、MCP の効果的な設計とアプリケーション層内のオーケストレーションロジックに移る。

4.2 Manus AI Agent

Manus AI Agent は、そのマルチエージェントシステムと長期記憶（LTM）メカニズムにより、特に長期間の自律運用を目指したレベル 3 エージェントとして注目される。

- **LLM 層（認知エンジン）：**
  - **コア LLM：** 広範なテキストおよびマルチモーダルデータで訓練されたトランスフォーマーベースの LLM（複数または単数）に基づいており、一般的な知能、言語理解、および推論能力を提供する [17]。一部の報告では、Anthropic の Claude 3.7 Sonnet との統合が示唆されている [66]。
  - **マルチエージェントシステム（認知的専門化）：** 異なる認知機能のために特化したサブエージェントを採用している [17]。
    - **プランナーエージェント：** ユーザーの要求/目標を管理可能なサブタスクに分解し、段階的な戦略を策定する [18]。これは長期的なタスクにとって不可欠である。
    - **検証エージェント：** 品質管理として機能し、実行エージェントの結果を確認し、正確性と完全性をチェックし、必要に応じて再計画をトリガーする

ことができる[18]。
    - **コンテキスト認識型意思決定：** コンテキストと中間結果の内部メモリを維持し、タスクの進化する状態とユーザーの好みを考慮して意思決定を行う[17]。論理的な次のステップのためにシーケンス・トゥ・シーケンス予測を使用する。
    - **強化学習（RL）と RLHF：** オープンエンドのシナリオでの効果的な運用と望ましい結果との整合のために、RL と RLHF を通じて改良された可能性が高い[17]。
- **アプリケーション層（実行と対話のフレームワーク）：**
    - **実行エージェント：** プランナーの計画を受け取り、必要な操作やツールを呼び出すことによって実行する[18]。
    - **ツール統合：** ウェブブラウザ、コードエディタ、データベース管理システム、シェルコマンド、ファイル管理、およびその他の外部アプリケーション/API とシームレスに統合する[17]。
    - **マルチモーダル機能：** テキスト、画像、コードを含む様々なデータタイプを処理および生成できる[17]。
    - **長期記憶（LTM）：** 短期および履歴データから学習し、将来のパフォーマンスを向上させるために過去の対話から情報を保持するために、LTM メカニズム（例：階層的メモリネットワーク、アテンションベースの検索、メモリオグメントニューラルネットワーク - MANN）を採用している[17]。進捗状況の追跡と情報の保存にはファイルベースのメモリが使用される[92]。
    - **非同期クラウド運用と継続性：** クラウドで動作し、ユーザーが切断してもタスクを続行できる[41]。これは数日間にわたるタスクにとって不可欠である。
    - **サンドボックス化された実行環境：** コードの実行やツールとの対話時にセキュリティを確保するため、制御されたランタイム環境（Linux サンドボックス）で実行される[52]。
    - **透明性：** サイドパネルを通じてプロセスを表示し、ユーザーがワークフローを確認できるようにする[86]。
- **レベル 3 能力のための相乗効果（数日間の運用に焦点）：**
    - プランナーエージェント（LLM 層）は、複雑で潜在的に長期的な目標の分解のために明示的に設計されている。検証エージェント（LLM 層）は、長期間の運用におけるエラー修正と再計画のためのフィードバックループを提供する。
    - 実行エージェント（アプリケーション層）は、これらの計画を実行するために広範なツールを使用する。
    - 重要なのは、LTM メカニズムとクラウド継続性（アプリケーション層）により、Manus は状態を維持し、長期タスク内の過去の対話から学習し、数日間にわたって自律的に運用を継続できることである[17]。「中断されたタスクを完全なコンテキスト保持で再開する」能力[52]は、数日間の運用の直接的な実現要

因である。
    - 分析→計画→実行→観察の反復ループ[92]と、コンテキスト認識型の意思決定[17]の組み合わせにより、Manus は数日間のタスク中に変化する要件や予期せぬ問題に適応できる。
    - 一部のユーザーレポートでは、信頼性に関する初期の課題（例：ループエラー[92]）が指摘されているが、アーキテクチャ設計は明示的に長期的な自律運用を対象としている。段階的なプロンプトなしで旅行全体の旅程を計画するような例[17]は、この能力を示している。
- Manus の明示的なサブエージェントアーキテクチャと LTM が堅牢な長期自律エージェンシーの青写真となるという考察：
  Manus のプランナー、実行、検証という明確なサブエージェントを持つアーキテクチャと、専用の LTM の組み合わせは、信頼性の高い長期間の自律的タスク完了を達成するための構造化されたアプローチを示している。この設計は、レベル 3 エージェントが直面する持続性と適応性の課題に対処する上で重要な示唆を与える。
  長期的な（数日間にわたる）タスクは、本質的に複雑さ、不確実性、およびエラーや初期計画からの逸脱の可能性を伴う。モノリシックなエージェントは、そのような期間にわたって計画、実行、検証のすべての側面を効果的に管理するのに苦労する可能性がある。Manus のこれらの懸念を専門のサブエージェント（計画/検証のための LLM 層、ツールを介した実行のためのアプリケーション層）に分離するアプローチは、各段階でより集中的かつ潜在的に堅牢な処理を可能にする[18]。プランナーは高レベルの戦略に集中でき、実行者はツールの使用に、検証者は品質管理に集中できる。
  このようなシステムが数日間にわたって単一の目標に対して一貫性を維持するためには、メモリが最も重要である。Manus の LTM とコンテキスト認識型の意思決定[17]は、中間結果、学習した経験、およびタスクの進化する状態が保存され、活用されることを保証する。検証エージェントが再計画をトリガーする能力[18]は、長期間実行されるタスク中のエラー回復と適応のための重要なメカニズムを提供し、エージェントが行き詰まったり目標から逸脱したりするのを防ぐ。クラウド継続性[85]は、プロセスが単一のユーザーセッションに縛られないことを保証し、真の非同期の数日間の運用を可能にする。
  したがって、このアーキテクチャは、長期間のタスクに対する堅牢なレベル 3 エージェンシーが、認知負荷が分散され、検証と再計画のための明示的なメカニズムが組み込まれたモジュラー設計から恩恵を受けることを示唆している。永続的な LTM を提供し、運用上の継続性を確保する上でのアプリケーション層の役割は、LLM の推論能力と同じくらい重要である。

4.3 OpenAI Operator

OpenAI Operator は、Computer-Using Agent (CUA) モデルを活用し、主にウェブブラウザを介したタスク自動化を目指すエージェントである。その LLM 層とアプリケーション層の連携は、GUI 操作の汎用性と自律性のバランスを示している。

- **LLM 層（認知エンジン）：**
  - **Computer-Using Agent (CUA) モデル：** GPT-4o の視覚能力と強化学習による高度な推論を組み合わせた特化モデル [30]。
  - **GUI 理解：** スクリーンショット/ピクセルからグラフィカルユーザーインターフェース（ボタン、メニュー、テキストフィールド）を解釈するように訓練されている [30]。
  - **思考連鎖推論：** 複雑なタスクを管理可能なステップに分解し、「内なる独り言」を使って観察結果を評価し、中間ステップを追跡し、動的に適応する [30]。
  - **自己修正：** 課題に直面したり間違いを犯したりした場合に、推論能力を活用して自己修正することができる [30]。
- **アプリケーション層（実行と対話のフレームワーク）：**
  - **ブラウザ自動化：** マウスのクリック、スクロール、タイピングをシミュレートしてウェブブラウザと対話する [30]。ウェブサイト用のカスタム API 統合は必要ない。
  - **スクリーンショット分析：** 視覚能力を使用して画面を「見て」、GUI 要素を理解する [30]。
  - **タスク実行：** フォーム入力、旅行予約、食料品注文などの反復的なブラウザタスクを自動化する [44]。
  - **ユーザー連携とハンドオフ：** ログイン、支払い詳細、または CAPTCHA 解決が必要なタスクについては、ユーザーに引き継ぎを積極的に要求する。ユーザーはいつでも制御を引き継ぐことができる [30]。重要なアクションにはユーザーの確認が必要である。
  - **並列タスク実行：** 新しい会話を作成することで、複数のタスクを同時に実行できる [45]。
  - **保存済みタスク/プロンプト：** ユーザーは繰り返しタスク用にプロンプトを保存できる [45]。
  - **Responses API とコンピュータ使用ツール：** 開発者は、CUA を利用した Responses API のコンピュータ使用ツールを使用して、モデルが生成したマウス/キーボードアクションを実行可能なコマンドに直接変換することで、コンピュータタスクを自動化できる [49]。
- **レベル 3 能力のための相乗効果（数日間の運用に焦点）：**

    - Operator の CUA モデル（LLM 層）は、ウェブベースのタスクを分解し、さまざまなウェブサイト UI に適応するための推論を提供する。
    - アプリケーション層は、CUA の決定に基づいてブラウザを直接操作する。
    - 数日間の運用に関して、Operator の現在の設計（リサーチプレビューとして）は課題を抱えている。繰り返しタスク用にプロンプトを保存したり[45]、タスクを並行して実行したりすることはできるが、人間の介入や再初期化なしに*単一の進化するタスクを数日間*自律的に管理するという明示的な言及はない[44]。
    - ログイン、支払い、または行き詰まった場合の人間による引き継ぎへの依存[45]は、そのようなハードルに頻繁に遭遇する可能性のある複雑な目標に対する完全な自律的長期運用における制限を示唆している。
    - 「リサーチプレビュー」というステータスと、述べられている制限（例：スライドショーやカレンダーのような複雑なインターフェースでの苦労[45]）は、単一の複雑な目標に対する堅牢な数日間の自律性が、現在の完全に実現された能力というよりは、将来の願望である可能性が高いことを示している。現在の焦点は、セッション内で完了できる一連の反復的なウェブタスクの自動化にあり、複数のそのようなタスクを並行して実行できるとしても、そうであるように見える。
- Operator の GUI 中心アプローチが汎用性と深い長期自律性の間のトレードオフを浮き彫りにするという考察：
  Operator の強みは、GUI を介してほぼすべてのウェブサイトと対話できる能力にあり、API を必要とせずに幅広い適用性を提供する。しかし、このアプローチ自体が、視覚的解釈とエミュレートされた人間のアクションに依存するため、単一の進化する目標に対する持続的で完全な自律的な数日間の運用に課題をもたらす可能性のある複雑さとエラーポイントを導入する可能性がある。
  Operator の CUA モデル（LLM 層）は画面を「見て」、どこをクリック/タイプするかを決定する[30]。アプリケーション層はこれらの GUI インタラクションを実行する。これは、ウェブサイトが API を持っている必要がないため、大きな柔軟性を提供する[47]。しかし、GUI は一貫性がなく、頻繁に変更されたり、AI がピクセルだけから確実に解釈するのが難しい複雑な JavaScript 要素を持っていたりする可能性があり、特にウェブサイトが更新されたり動的コンテンツを提示したりする数日間ではそうである[47]。これにより、エージェントがより頻繁に「行き詰まる」可能性があり、OpenAI が組み込んだ人間による引き継ぎメカニズムが必要となる[45]。
  数日間にわたって進化する単一のタスクの場合、回復不可能な GUI 関連のエラーや人間の介入が必要な状況（繰り返される CAPTCHA やログインが必要なセッシ

ョンタイムアウトなど）に遭遇する累積確率は大幅に増加する。Operator は自己修正できるが[46]、非常に長く、複雑で、進化する GUI ベースのインタラクションにおいて、API レベルの制御なしでのこの信頼性は、*単一の継続的な目標*に対する真の数日間の自律性にとって大きなハードルである。
したがって、Operator は、現在の形態では、数日間にわたって 1 つの継続的で進化するプロジェクトを自律的に管理するのではなく、ユーザーがセッションで行う可能性のある一連の個別のウェブタスクを自動化することに重点を置いているように見える。後者を達成するには、おそらくさらに高度な推論、視覚的理解、エラー回復、そして重要な状態情報のための API インタラクションまたはより構造化されたデータ交換がある程度必要となり、長期的な永続性と信頼性のために純粋な GUI インタラクションを超える必要があるだろう。現在の安全策と人間によるハンドオフポイント[45]は安全性にとって不可欠であるが、本質的に完全な無人の数日間の自律性を制限する。

4.4 OpenAI Deep Research

OpenAI Deep Research は、特定の複雑な調査タスクを自律的に実行するエージェントであり、その LLM 層とアプリケーション層は、高度な情報収集と分析に特化している。

- **LLM 層（認知エンジン）：**
  - **最適化された OpenAI o3 モデル：** ウェブブラウジングとデータ分析に特化して最適化された、今後登場する o3 モデルのバージョンを搭載[97]。このモデルが中核的な推論能力を提供する。
  - **複数ステップの調査計画：** 推論を活用して情報を検索、解釈、分析し、遭遇した情報に反応して必要に応じて方向転換する。複雑なタスクに対してインターネット上で複数ステップの調査を実施できる[97]。
  - **情報統合：** 何百ものオンラインソースを見つけ、分析し、統合して、リサーチアナリストのレベルで包括的なレポートを作成する[97]。
  - **実世界のツール使用に関するトレーニング：** OpenAI o1 と同様の強化学習手法を使用して、ブラウザと Python ツールの使用を必要とする実世界のタスクでトレーニングされている[97]。
- **アプリケーション層（実行と対話のフレームワーク）：**
  - **ウェブブラウジングツール：** テキスト、画像、PDF から情報を収集するためにインターネットを積極的に閲覧する[97]。
  - **データ分析用 Python ：** 収集した情報を処理するために Python をデータ分析に利用できる[97]。
  - **レポート生成：** 明確な引用と考察の要約を含む包括的で文書化されたレポート

に洞察を統合する[97]。
    - **ファイルアップロード：** アップロードされたファイルやスプレッドシートを使用して、調査クエリにコンテキストを追加できる[97]。
    - **非同期操作（短期間）：** クエリが送信された後、バックグラウンドで動作し、作業完了までに 5 分から 30 分かかる[97]。
- **レベル 3 能力のための相乗効果（数日間の運用に焦点）：**
    - Deep Research の LLM（o3 バリアント）は、複雑な複数ステップの調査戦略を計画し実行する。アプリケーション層は、この戦略を実行するためのブラウジングツールと Python ツールを提供する。
    - その出力は、人間のリサーチアナリストが何時間もかけて作成するような包括的なレポートであり[97]、特定の種類の目標（詳細な調査）に対する高レベルの自律的なタスク完了を示している。
    - しかし、記述されている操作は、単一の、ただし複雑な調査クエリに対してであり、「数十分」以内に完了する[97]。提供された資料には、Deep Research が単一の調査目標を*数日間にわたって自律的に*追求し、各フェーズで新しいユーザープロンプトなしに継続的に更新、適応、構築するという兆候はない。
    - 「何百ものオンラインソース」を分析できるが[97]、これは 1 つのクエリによって開始された単一の調査実行のコンテキスト内であるように見える。これは「独立して作業を行うことができるエージェント」ではあるが[97]、「数日間」と比較して比較的短い完了ウィンドウを持つ定義された調査タスクの範囲内である。
    - 現在の設計は、「この複雑な調査を今すぐ行ってください」という非常に強力なツールであり、「来週にかけてこの進行中の調査プロジェクトを管理してください」というエージェントではないように見える。
- Deep Research が特化型「バースト活動」レベル 3 エージェントであり、自律性のスペクトルを浮き彫りにするという考察：

  Deep Research は、その自律性とタスク実行の複雑性（複数ステップの調査、ツール使用、統合）においてレベル 3 の特性を示している。しかし、その運用時間枠（クエリあたり数十分）は、単一の進化する目的に対する継続的な数日間の目標追求ではなく、集中的な「バースト」活動のためのエージェントとして位置づけられる。この事実は、レベル 3 エージェントの能力範囲に関する重要な理解を促す。

  レベル 3 エージェントは、ユーザーの代理として、潜在的に長期間にわたってタスクを自律的に実行することによって定義される[1]。Deep Research は、ツール（ブラウジング、Python）を使用し、何百ものソースから情報を統合して、複雑な複数ステップの調査を自律的に実行する[97]。これは「自律的実行」の側面に合致

する。典型的なタスク期間は 5〜30 分である[97]。これは同じ深さの調査を行う人間よりもはるかに高速であるが、単一の継続的で進化する目標に適用される場合、「数日間ユーザーの代理として行動する」という一部のレベル 3 の解釈には適合しない。

Deep Research は、複雑な調査*クエリ*を受け取り、そのクエリに対する包括的な*レポート*を返すように設計されているように見える。広範な調査*プロジェクト*を受け取り、数日間にわたって反復的に作業し、新しい明確なクエリなしにプロジェクトの途中で新しい発見やユーザーのフィードバックに適応するとは記述されていない。

したがって、Deep Research は、セッション内で完了する特定の詳細なタスクに対して高い自律性と複雑なツール使用を示すタイプのレベル 3 エージェントを例示している。これは、「レベル 3 AI エージェント」がモノリシックなカテゴリではないことを示唆している。レベル 3 内には、自律性の期間と性質が異なるエージェントのスペクトルが存在し得る。Deep Research は、集中的で限定的な調査サイクルに最適化された LLM 層とアプリケーション層を持っている。一方、Manus や Genspark に記述されている理想的な他のレベル 3 エージェントは、はるかに長い時間スケールでのより永続的で進化する目標管理を目指す可能性がある。

5. 比較分析と統合：レベル 3 エージェンシーのためのアーキテクチャ

Genspark、Manus、OpenAI Operator、Deep Research の事例は、レベル 3 の自律性を達成するための多様なアーキテクチャ戦略を浮き彫りにする。これらのエージェントは、LLM 層の認知能力とアプリケーション層の実行能力を組み合わせるという共通の原則に基づいて構築されているが、その具体的な実装と重点は異なり、結果として異なる種類のレベル 3 能力が生まれている。

LLM 層機能における共通原則

- **高度な LLM の活用：** 4 つのエージェントすべてが、中核的な推論、計画、自然言語理解のために高度な LLM（OpenAI の o3/CUA、Genspark の MoA、Manus のコア LLM）を利用している。
- **タスク分解：** LLM の思考連鎖に暗黙的に含まれるか、マルチエージェント計画（Manus）で明示的に行われるかにかかわらず、タスク分解は共通の戦略である。
- **意思決定とツール選択：** LLM 層は一貫して、*何を行うか*、そして*どのツールを使用するか*を決定する責任を負っている。

アプリケーション層実装のバリエーション

- **ツールの多様性と対話モード：** Genspark は 80 以上の広範なツールと直接 API アクセスを誇る。Manus もブラウザやコードエディタなどの多様なツールを統合している。Operator はブラウザ自動化による GUI 対話に焦点を当てている。Deep Research はブラウジングと Python を使用する。
- **マルチエージェント vs. 単一高度 LLM：** Genspark（MoA）と Manus（プランナー/実行/検証）は、認知負荷と実行を分散させるために、LLM 層および/またはアプリケーション層でマルチエージェントアーキテクチャを採用している。Operator と Deep Research は、それぞれのコア推論のために単一の非常に有能な LLM（それぞれ CUA と o3 バリアント）に依存しているように見える。
- **拡張タスクのための状態とメモリ管理：**
  - Genspark（MCP 経由）と Manus（LTM、クラウド継続性経由）は、数日間の運用に不可欠な永続的な状態と長期記憶のために設計されたアプリケーション層に明示的なアーキテクチャコンポーネントを持っている [17]。
  - Operator の現在の公開情報は、単一の継続的な目標に対する自律的な数日間の状態永続性に関する詳細が少なく、人間によるハンドオフを伴うセッションベースのタスク自動化に重点を置いている [45]。
  - Deep Research は集中的で限定的な調査タスク用に設計されており、進化する調査プロジェクトに関する数日間の継続的な目標追求のためのメカニズムを明示的に記述していない [97]。

アーキテクチャ選択がレベル 3 能力に与える影響

- **タスクの複雑性と期間：** 明示的なマルチエージェントシステムと専用の長期記憶を持つアーキテクチャ（Genspark、Manus）は、理論的には数日間にわたる非常に複雑で進化する目標により適している。これらのシステムのアプリケーション層は、永続性と調整のために設計されている。
- **信頼性とエラー処理：** Manus の検証エージェントは、明示的な LLM 駆動のチェックを提供する。Operator は CUA の自己修正と人間による引き継ぎに依存する。Genspark の MoA は冗長性を提供する可能性がある。Deep Research は、o3 モデルの堅牢性とユーザーによる検証のための明確な引用に依存する。数日間のタスクの場合、LLM によって導かれるアプリケーション層内の堅牢なエラー回復と適応的再計画が不可欠である [68]。
- **対話パラダイム：** Operator の GUI 対話（アプリケーション層）は幅広い適用性を提供するが、UI の変更により長期的なタスクでは脆弱性が増す可能性がある。Genspark と Manus の API/ツール統合（アプリケーション層）は、API が安定していればより堅牢であり得る。

進化する共生関係

- LLM がより有能になるにつれて（例：より長いコンテキストウィンドウ、より優れた推論、o3/o4-mini のようなより信頼性の高いツール使用[24]）、アプリケーション層への要求は変化する可能性がある。しかし、真のレベル 3 の数日間の自律性のためには、状態管理、プロセスオーケストレーション、および信頼性の高い実行インターフェースを提供するというアプリケーション層の役割は依然として基本的である。
- アプリケーション層におけるエージェントネイティブインターフェースと標準化された通信プロトコル（MCP など[54]）への傾向は、より相互運用可能でスケーラブルなレベル 3 エージェントを構築するための鍵となるだろう。

主要表：レベル 3 AI エージェント例における LLM およびアプリケーション層実装の比較概要

ユーザーが異なるレベル 3 エージェントにおける LLM 層とアプリケーション層の役割を理解する必要があるため、各エージェントのアーキテクチャはこれらの層に対して独自のものである。テキストによる説明だけでは、これらの多様なアプローチを直接比較し、共通のパターンや主要な差別化要因、特に長期間タスクへの適合性に関して特定することが困難な場合がある。中核的な LLM 機能、主要なアプリケーション層コンポーネント、複雑/長期タスクへのアプローチ、および特定のレベル 3 属性（特に数日間の自律性）をまとめた構造化された表は、明確で簡潔かつ直接比較可能な概要を提供する。これにより、読者はアーキテクチャ戦略と、それらがどのようにレベル 3 能力に貢献または制限するかを迅速に把握でき、本文の詳細な分析を補強する。これは「レベル 3」実装のスペクトルを強調する。

エージェント名	コア LLM と主要な推論/計画能力（LLM 層）	主要なアプリケーション層の機能、ツール、プロトコル	タスク分解と長期間実行へのアプローチ	主要な対話/出力モード	明記された/示唆されるレベル 3 特性（数日間の自律性に焦点）
Genspark Super Agent	Mixture-of-Agents（9 つの LLM）、タス	80 以上の内製ツール（マルチメディア作成、データ	高レベルの指示を分解し、複数の LLM とツールを	ウェブインターフェース、生成されたコンテンツ（ウ	MoA と MCP により、理論上は複雑で長期間のタスク

	クに応じた最適な LLM へのインテリジェントルーティング、自律的計画とタスク分解 [81]	分析等）、直接 API 統合、リアルタイム音声自動化、Model Context Protocol (MCP)による状態・メモリ管理 [54]	MCP で連携させて計画を実行。マルチタスク対応 [81]	ェブサイト、レポート、ビデオ）、音声通話	に対応可能。MCP は永続メモリとステートフルワークフローをサポート [54]。ただし、単一タスクでの数日間の完全自律運用に関する直接的な事例は限定的。
Manus AI Agent	トランスフォーマーベース LLM（Claude 3.7 Sonnet 統合の可能性）、マルチエージェントシステム（プランナー、検証エージェント）、コンテキスト認識型意思決定、RLHF による改良 [18]	実行エージェント、ツール統合（ブラウザ、コードエディタ、DB 等）、マルチモーダル処理、長期記憶（LTM、MANN）、非同期クラウド運用、サンドボックス実行 [41]	プランナーが目標をサブタスクに分解し、実行エージェントが LTM とツールを活用して実行。検証エージェントが品質管理と再計画をトリガー。クラウド継続性により中断なしの長期運用をサポート [18]	ウェブインターフェース（プロセス表示あり）、生成されたレポート、コード、計画	LTM とクラウド継続性により、数日間にわたる自律運用と中断タスクのコンテキスト付き再開が可能 [51]。アーキテクチャは明示的に長期自律運用を対象。
OpenAI Operator	Computer-Using Agent (CUA)モデル（GPT-4o 視覚＋RL）、GUI 理解、思考連鎖推論、自己修正 [30]	ブラウザ自動化（クリック、スクロール、タイピング）、スクリーンショット分析、ユーザー連携（ハンドオフ、確認）、並列タスク実行、保	思考連鎖でタスクをステップに分解し、GUI を介してブラウザで実行。複雑な状況や機密情報入力時には人間にハンドオフ [30]	ウェブインターフェース（ブラウザ操作の観察）、フォーム入力、オンライン注文	主にセッションベースの反復的ウェブタスクの自動化に焦点。単一の継続的な目標に対する数日間の完全自律運用能力は、現在のリサーチプレビ

		存済みプロンプト、Responses API[44]			ュー段階では限定的[46]。
OpenAI Deep Research	最適化された OpenAI o3 モデル、複数ステップの調査計画、情報統合、実世界のツール使用に関する RL トレーニング[97]	ウェブブラウジングツール、データ分析用 Python、レポート生成、ファイルアップロード、非同期操作（5-30 分）[97]	単一の複雑な調査クエリに対し、複数のオンラインソースを探索・分析し、包括的なレポートを生成[97]	ChatGPT インターフェース経由でのクエリ入力、生成された詳細レポート（引用付き）	高度な自律性と複雑なツール使用を示すが、運用は単一クエリに対する「バースト活動」（数十分）であり、数日間にわたる継続的な目標追求は記述されていない[97]。

レベル 3 は「システム統合」の課題であるという考察

堅牢なレベル 3 エージェンシー、特に数日間にわたるタスクを達成することは、単一の画期的な LLM というよりも、永続性、ツーリング、およびプロセスマネジメントを提供する適切に設計されたアプリケーション層との LLM（複数または単数）の洗練された統合にかかっている。この視点は、AI エージェント開発の現状と将来の方向性を理解する上で重要である。

これらの事例は、多様な LLM アプローチ（MoA、専門化されたサブエージェント、強力な単一モデル）を示している。しかし、数日間にわたって自律的に行動する能力は、アプリケーション層の機能、すなわちメモリ（Manus LTM、Genspark MCP 状態）、ツールの信頼性、エラー処理、およびワークフローオーケストレーションに一貫して依存している。最も高度な LLM でさえ、アプリケーション層がセッション間で状態を維持したり、2 日目に失敗した API 呼び出しから回復したり、正しいツールが利用可能で機能していることを保証したりできなければ、数日間のタスクを実行することはできない。

エージェント開発で強調される課題は、しばしばこれらのアプリケーション層の懸念に関連している。信頼性の高い実行、状態の永続性、マルチエージェントの調整、ツール使用におけるセキュリティの脆弱性などである[10]。したがって、レベル 3 は、LLM 層

によってオーケストレーションされる、アプリケーション層における重要なシステム統合およびエンジニアリングの課題である。

このことから、より有能なレベル 3 エージェントへの今後の進歩は、LLM の拡張だけでなく、アプリケーション層アーキテクチャ、プロトコル（MCP など）、および堅牢でステートフルかつ安全なエージェントシステムを構築するためのエンジニアリングプラクティスの進歩に大きく依存するだろう。

6. 結論：AI エージェントの進化における LLM 層とアプリケーション層の相互依存的かつ共進化的な役割

本レポートで詳述したように、Genspark、Manus、OpenAI Operator、Deep Research といったレベル 3 AI エージェントの能力は、LLM 層の認知機能とアプリケーション層の実行・対話フレームワークとの間の緊密な連携によって実現されている。これらの層は独立して機能するのではなく、相互に依存し、共に進化することで、より高度な自律性とタスク遂行能力を AI エージェントにもたらす。

重要な機能の要約

- **LLM 層：** 知能を提供し、理解、推論、計画、動的な意思決定を担う。エージェントの行動の「指揮者」である。
- **アプリケーション層：** 具体化と運用能力を提供し、ツール、環境との対話、状態管理、メモリ、通信インフラストラクチャを担う。LLM の指示を現実に移す「役者と舞台」である。

レベル 3 の自律性に不可欠な相乗効果

- どちらの層も単独ではレベル 3 のエージェンシーを達成できない。LLM の計画は、それを実行するアプリケーション層がなければ無力である。アプリケーション層のツールは、LLM のガイダンスがなければ方向性を見失う。
- 長期間にわたる数日間のタスクの場合、この相乗効果はさらに重要になり、両層にわたる緊密な結合、堅牢なフィードバックループ、および洗練されたコンテキスト管理が必要となる。

今後の展望

- **継続的な共進化：** LLM の推論能力の進歩（例：より長いコンテキストの処理、より複雑な計画、より優れたツール学習）は、より洗練されたアプリケーション層の能力（例：より複雑なワークフローオーケストレーション、より多様なツールセット、より堅牢な長期記憶ソリューション）の開発を促進するだろう。

- **逆に、アプリケーション層アーキテクチャの革新**（例：MCP/A2A のような標準化されたエージェント通信プロトコル、ツール使用のための安全なサンドボックス化、新しいメモリシステム）は、LLM がより複雑で自律的な方法で適用される新たな可能性を切り開くだろう。
- AI エージェンシーのより高いレベル（レベル 4 および 5）への道は、信頼性、安全性、および真の理解における現在の課題に対処しながら、LLM 層とアプリケーション層の両方におけるさらなる深い統合とより高度な能力に間違いなく依存するだろう[1]。

「アプリケーション層」は「エージェントオペレーティングシステム」へと進化するという最終的な考察

エージェントがより複雑になり、長期間にわたる多面的なタスクを実行することが期待されるにつれて、「アプリケーション層」は伝統的にオペレーティングシステムに関連付けられてきた特性を帯び始めている。これには、リソース管理、プロセススケジューリング（サブタスクまたはサブエージェント用）、メモリ管理（長期状態）、プロセス間通信（エージェントコンポーネントまたはツール間）、およびハードウェア抽象化（さまざまなデジタルツールおよび環境とのインターフェース接続）が含まれる。この進化は、高度な AI エージェントの将来の設計と機能にとって重要な意味を持つ。

レベル 3 エージェント、特に Genspark（MoA）や Manus（マルチエージェント）のようなものは、協調して動作する複数の認知的および実行的コンポーネントを伴う。これらのコンポーネントは管理され、調整され、リソースを提供される必要がある。アプリケーション層は、MCP、オーケストレーションフレームワーク、およびツール統合レイヤーのようなメカニズムを通じて、これらの管理および調整機能を実行している。これは、OS がコンピュータ上のアプリケーション、プロセス、およびハードウェアリソースを管理する方法に類似している。

エージェントが数日間にわたって堅牢に動作し、状態を管理し、エラーを処理し、一貫した目標追求を保証するためには、単なるツールの集合ではなく、体系的なアプローチが必要である。したがって、高度に自律的なエージェント（レベル 3 以上）の将来の開発は、アプリケーション層の重要な部分として、より形式化された「エージェントオペレーティングシステム」または「エージェント実行プラットフォーム」の出現を見る可能性が高い。これらのプラットフォームは、LLM 駆動の知能が世界で効果的、持続的、かつ安全に動作するために必要な基盤サービスを提供するだろう。これは、アプリケーション層を単なるツールセットから、洗練された管理および実行環境へと格上げするものである。

引用文献

1. OpenAI's 5 phases toward AGI | Next Level Impact, 5 月 18, 2025 にアクセス、 https://nxtli.com/en/openai-agi/
2. curriculumredesign.org, 5 月 18, 2025 にアクセス、 https://curriculumredesign.org/wp-content/uploads/The-Frustrating-Quest-to-Define-AGI.pdf
3. The path to AGI: A deep dive into openAI's 5-level framework - Twimbit, 5 月 18, 2025 にアクセス、 https://twimbit.com/about/blogs/the-path-to-agi-a-deep-dive-into-openais-5-level-framework
4. The Path to AGI: How Do We Know When We're There? - Lumenova AI, 5 月 18, 2025 にアクセス、 https://www.lumenova.ai/blog/artificial-general-intelligence-measuring-agi/
5. Understanding OpenAI's Five Levels of AI Progress Towards AGI ..., 5月 18, 2025 にアクセス、 https://quinteft.com/understanding-openais-five-levels-of-ai-progress-towards-agi
6. OpenAI's 5 Steps to AGI: Level 2 Reasoners Officially Here - First Movers, 5 月 18, 2025 にアクセス、 https://firstmovers.ai/openai-agi-levels/
7. OpenAI's 5 Steps to AGI - Perplexity, 5 月 18, 2025 にアクセス、 https://www.perplexity.ai/page/openai-s-5-steps-to-agi-STzkIF5SSQ6JOiBTaV.cfA
8. OpenAI's Five-Step Journey from AI to AGI - HulkApps, 5 月 18, 2025 にアクセス、 https://www.hulkapps.com/blogs/ecommerce-hub/openais-five-step-journey-from-ai-to-agi
9. Understanding OpenAI's Five Levels of AI Progress Towards AGI, 5月 18, 2025 にアクセス、 https://quinteft.com/understanding-openais-five-levels-of-ai-progress-towards-agi/
10. Autonomous AI Agents: Exploring Their Role - Neontri, 5 月 18, 2025 にアクセス、 https://neontri.com/blog/autonomous-ai-agents/
11. AI Agents: Evolution, Architecture, and Real-World Applications - arXiv, 5 月 18, 2025 にアクセス、 https://arxiv.org/html/2503.12687v1
12. AI Agents: Evolution, Architecture, and Real-World Applications - arXiv, 5 月 18, 2025 にアクセス、 https://arxiv.org/pdf/2503.12687
13. The role of AI agents in artificial intelligence and LLMs - ActivDev, 5 月 18, 2025 にアクセス、 https://www.activdev.com/en/the-role-of-ia-agents-in-artificial-intelligence-and-llms/
14. arxiv.org, 5 月 18, 2025 にアクセス、 https://arxiv.org/pdf/2503.04596
15. The Next Frontier of LLM Applications: Open Ecosystems and Hardware Synergy - arXiv, 5 月 18, 2025 にアクセス、 https://arxiv.org/pdf/2503.04596?
16. OpenAI's Deep Research on Training AI Agents End-to-End - Sequoia Capital, 5 月 18, 2025 にアクセス、 https://www.sequoiacap.com/podcast/training-data-deep-research/

17. From Mind to Machine: The Rise of Manus AI as a Fully Autonomous Digital Agent - arXiv, 5 月 18, 2025 にアクセス、 https://arxiv.org/html/2505.02024v1
18. (PDF) From Mind to Machine: The Rise of Manus AI as a Fully ..., 5 月 18, 2025 にアクセス、 https://www.researchgate.net/publication/391461097_From_Mind_to_Machine_The_Rise_of_Manus_AI_as_a_Fully_Autonomous_Digital_Agent
19. Scaling AI Agents in the Enterprise: The Hard Problems and How to Solve Them, 5 月 18, 2025 にアクセス、 https://thenewstack.io/scaling-ai-agents-in-the-enterprise-the-hard-problems-and-how-to-solve-them/
20. What is a large language model (LLM)? | SAP, 5 月 18, 2025 にアクセス、 https://www.sap.com/resources/what-is-large-language-model
21. Guide to Large Language Models (LLMs) Explained | Balbix, 5 月 18, 2025 にアクセス、 https://www.balbix.com/insights/what-is-large-language-model-llm/
22. What Are Large Language Models (LLMs)? - IBM, 5 月 18, 2025 にアクセス、 https://www.ibm.com/think/topics/large-language-models
23. What is a large language model (LLM)? - SAP, 5 月 18, 2025 にアクセス、 https://www.sap.com/swiss/resources/what-is-large-language-model
24. Introducing OpenAI o3 and o4-mini | OpenAI, 5 月 18, 2025 にアクセス、 https://openai.com/index/introducing-o3-and-o4-mini/
25. [The AI Show Episode 145]: OpenAI Releases o3 and o4-mini, AI Is ..., 5 月 18, 2025 にアクセス、 https://www.marketingaiinstitute.com/blog/the-ai-show-episode-145
26. OpenAI outlines plan for AGI—5 steps to reach superintelligence | Tom's Guide, 5 月 18, 2025 にアクセス、 https://www.tomsguide.com/ai/chatgpt/openai-has-5-steps-to-agi-and-were-only-a-third-of-the-way-there
27. 7 reasons why AI reasoning models accelerate time to value ..., 5 月 18, 2025 にアクセス、 https://lumenalta.com/insights/7-reasons-why-ai-reasoning-models-accelerate-time-to-value
28. o3 and o4-mini: Unlock enterprise agent workflows with next-level ..., 5 月 18, 2025 にアクセス、 https://azure.microsoft.com/en-us/blog/o3-and-o4-mini-unlock-enterprise-agent-workflows-with-next-level-reasoning-ai-with-azure-ai-foundry-and-github/
29. Large Action Models, toward operational artificial intelligence - Tech4Future, 5 月 18, 2025 にアクセス、 https://tech4future.info/en/large-action-models-operational-ai/
30. Computer-Using Agent | OpenAI, 5 月 18, 2025 にアクセス、 https://openai.com/index/computer-using-agent/
31. cdn.openai.com, 5 月 18, 2025 にアクセス、 https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf
32. What is LLM Orchestration? | IBM, 5 月 18, 2025 にアクセス、 https://www.ibm.com/think/topics/llm-orchestration
33. Embracing Foundation Models for Advancing Scientific Discovery ..., 5 月 18,

2025 にアクセス、 https://pmc.ncbi.nlm.nih.gov/articles/PMC11923747/
34. GPT-4.1: OpenAI unveils its latest model - Swiftask, 5 月 18, 2025 にアクセス、 https://www.swiftask.ai/blog/gpt-4-1
35. What are AI agents? | Speakeasy, 5 月 18, 2025 にアクセス、 https://www.speakeasy.com/mcp/ai-agents-intro
36. Multi-Agent AI Systems: Orchestrating AI Workflows - V7 Labs, 5 月 18, 2025 にアクセス、 https://www.v7labs.com/blog/multi-agent-ai
37. Building stateful AI agents: why you need to leverage long-term ..., 5 月 18, 2025 にアクセス、 https://hypermode.com/blog/building-stateful-ai-agents-long-term-memory
38. Anatomy of an AI Agent – How AI Processes Information and Learns | C1M, 5 月 18, 2025 にアクセス、 https://c1m.ai/anatomy-of-an-ai-agent-how-ai-processes-information-and-learns/
39. What is the Application Layer? A Deep Dive into Network Communication - OurCrowd, 5 月 18, 2025 にアクセス、 https://www.ourcrowd.com/learn/what-is-the-application-layer
40. What is the Application Layer? - AI Glossary, 5 月 18, 2025 にアクセス、 https://www.theainavigator.com/blog/what-is-the-application-layer
41. Manus AI - China's Fully Autonomous AI Agent - OpenCV, 5 月 18, 2025 にアクセス、 https://opencv.org/blog/manus-ai/
42. AI Agents are NOT coming for your job. My experience with OpenAI's Operator : r/csMajors, 5 月 18, 2025 にアクセス、 https://www.reddit.com/r/csMajors/comments/1i8joi4/ai_agents_are_not_coming_for_your_job_my/
43. I am among the first people to gain access to OpenAI's "Operator" Agent. Here are my thoughts. : r/ChatGPTPro - Reddit, 5 月 18, 2025 にアクセス、 https://www.reddit.com/r/ChatGPTPro/comments/1i8jln3/i_am_among_the_first_people_to_gain_access_to/
44. OpenAI Redefines Business Productivity with Launch ... - DesignRush, 5 月 18, 2025 にアクセス、 https://www.designrush.com/news/openai-redefines-business-productivity-with-launch-of-first-ai-agent-operator
45. Operator - OpenAI Help Center, 5 月 18, 2025 にアクセス、 https://help.openai.com/en/articles/10421097-operator
46. Introducing Operator | OpenAI, 5 月 18, 2025 にアクセス、 https://openai.com/index/introducing-operator/
47. What is OpenAI Operator? Complete Guide to OpenAI's AI Agent - Leanware, 5 月 18, 2025 にアクセス、 https://www.leanware.co/insights/openai-operator-guide
48. OpenAI's Operator: Examples, Use Cases, Competition & More - DataCamp, 5 月 18, 2025 にアクセス、 https://www.datacamp.com/blog/operator
49. New tools for building agents | OpenAI, 5 月 18, 2025 にアクセス、 https://openai.com/index/new-tools-for-building-agents/

50. Computer-using agent (CUA) models - ZBrain, 5 月 18, 2025 にアクセス、 https://zbrain.ai/cua-models/
51. Manus AI: A Technical Deep Dive into China's First Autonomous AI ..., 5 月 18, 2025 にアクセス、 https://dev.to/sayed_ali_alkamel/manus-ai-a-technical-deep-dive-into-chinas-first-autonomous-ai-agent-30d3
52. Manus AI: The Ultimate AI Agent With Its Own Computer in 2025 ..., 5 月 18, 2025 にアクセス、 https://www.cursor-ide.com/blog/manus-ai-agentic-automation-guide-2025
53. The BEST AI Agent I've Ever Used (Genspark AI) - YouTube, 5 月 18, 2025 にアクセス、 https://www.youtube.com/watch?v=4_gH1PCow1g
54. MCP in Action: Why Your Next AI Workflow Won't Work Without ..., 5 月 18, 2025 にアクセス、 https://www.fluid.ai/blog/mcp-in-action-why-your-next-ai-workflow-wont-work-without-model-context-protocol
55. agent-architectures-with-mcp - AIXplore - Tech Articles - Obsidian Publish, 5 月 18, 2025 にアクセス、 https://publish.obsidian.md/aixplore/AI+Systems+%26+Architecture/agent-architectures-with-mcp
56. Multi-Agent Communication Protocols in Generative AI and Agentic AI: MCP and A2A Protocols - Architecture &Governance Magazine, 5 月 18, 2025 にアクセス、 https://www.architectureandgovernance.com/uncategorized/multi-agent-communication-protocols-in-generative-ai-and-agentic-ai-mcp-and-a2a-protocols/
57. Framework, Use Cases and Requirements for AI Agent Protocols - IETF, 5 月 18, 2025 にアクセス、 https://www.ietf.org/id/draft-rosenberg-ai-protocols-00.html
58. Internet of Agents: Fundamentals, Applications, and Challenges - arXiv, 5 月 18, 2025 にアクセス、 https://arxiv.org/html/2505.07176v1
59. What is Model Context Protocol? | A Practical Guide by K2view, 5 月 18, 2025 にアクセス、 https://www.k2view.com/model-context-protocol/
60. Model Context Protocol (MCP) - A Deep Dive - WWT, 5 月 18, 2025 にアクセス、 https://www.wwt.com/blog/model-context-protocol-mcp-a-deep-dive?utm_source=social&utm_medium=linkedin&utm_campaign=platform_share
61. Model Context Protocol (MCP): Solution to AI Integration Bottlenecks - Addepto, 5 月 18, 2025 にアクセス、 https://addepto.com/blog/model-context-protocol-mcp-solution-to-ai-integration-bottlenecks/
62. Building A Secure Agentic AI Application Leveraging Google's A2A Protocol - arXiv, 5 月 18, 2025 にアクセス、 https://arxiv.org/html/2504.16902
63. AI Agent Comparison: GenSpark.ai, Manus.im, and Similar Platforms · AI for LinkedIn - evyAI.com - Skool, 5 月 18, 2025 にアクセス、 https://www.skool.com/evyai/ai-agent-comparison-gensparkai-manusim-and-similar-platforms
64. (PDF) From Mind to Machine: The Rise of Manus AI as a Fully Autonomous Digital

Agent, 5 月 18, 2025 にアクセス、 https://www.researchgate.net/publication/391461097_From_Mind_to_Machine_The_Rise_of_Manus_AI_as_a_Fully_Autonomous_Digital_Agent/download
65. The General-Purpose AI Agent Product, Manus - Kaggle, 5 月 18, 2025 にアクセス、 https://www.kaggle.com/discussions/general/566681
66. Manus AI: The Revolutionary Multi-Agent System Reshaping the Future of AI Assistants, 5 月 18, 2025 にアクセス、 https://dev.to/aniruddhaadak/manus-ai-the-revolutionary-multi-agent-system-reshaping-the-future-of-ai-assistants-2i35
67. The Benefits And Challenges Of Implementing AI Agents - 66degrees, 5 月 18, 2025 にアクセス、 https://66degrees.com/benefits-and-challenges-of-ai-agents/
68. The Rise of Agentic AI: From Conversation to Action | Insights ..., 5 月 18, 2025 にアクセス、 https://www.goodwinlaw.com/en/insights/publications/2025/05/insights-technology-aiml-the-rise-of-agentic-ai-from-conversation
69. 3 Things We Need To Fix Before AI Agents Go Mainstream | Built In, 5 月 18, 2025 にアクセス、 https://builtin.com/articles/prepare-ai-agents-mainstream
70. How agentic AI enables an autonomous SOC with minimal human involvement - IBM, 5 月 18, 2025 にアクセス、 https://www.ibm.com/think/insights/agentic-ai-enables-autonomous-soc
71. Understanding AI Agents: From Fundamentals to Implementation - AI Academy, 5 月 18, 2025 にアクセス、 https://www.ai-academy.com/blog/understanding-ai-agents-from-fundamentals-to-implementation
72. Measuring AI Ability to Complete Long Tasks - METR, 5 月 18, 2025 にアクセス、 https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/
73. The Rise of AI Agents: A New Paradigm - Credo AI Company Blog, 5 月 18, 2025 にアクセス、 https://www.credo.ai/blog/the-rise-of-ai-agents-a-new-paradigm
74. A general motion control architecture for an autonomous underwater vehicle with actuator faults and unknown disturbances through deep reinforcement learning - ResearchGate, 5 月 18, 2025 にアクセス、 https://www.researchgate.net/publication/363318016_A_general_motion_control_architecture_for_an_autonomous_underwater_vehicle_with_actuator_faults_and_unknown_disturbances_through_deep_reinforcement_learning
75. Daily Papers - Hugging Face, 5 月 18, 2025 にアクセス、 https://huggingface.co/papers?q=virtual%20agents
76. Daily Papers - Hugging Face, 5 月 18, 2025 にアクセス、 https://huggingface.co/papers?q=defenses%20for%20generalist%20web%20agents
77. Adapting to Environmental Dynamics with an Artificial Circadian System - Georgia Tech, 5 月 18, 2025 にアクセス、 https://sites.cc.gatech.edu/ai/robot-lab/online-publications/MJO_Revised_PrePrint.pdf

78. Advancing Multi-Agent Systems Through Model Context Protocol: Architecture, Implementation, and Applications - arXiv, 5 月 18, 2025 にアクセス、 https://arxiv.org/html/2504.21030v1
79. The true LLM Agent - LongPort, 5 月 18, 2025 にアクセス、 https://longportapp.com/en/news/232841515
80. Agents - Blockchain Council, 5 月 18, 2025 にアクセス、 https://www.blockchain-council.org/blogs/agents/
81. AI Super Agent Explained: 5 Ways GenSpark's New Update Is ..., 5 月 18, 2025 にアクセス、 https://www.reddit.com/r/AISEOInsider/comments/1kl3spk/ai_super_agent_explained_5_ways_gensparks_new/
82. INSANE One-click MCP AI Agent Hits the Market | HackerNoon, 5 月 18, 2025 にアクセス、 https://hackernoon.com/insane-one-click-mcp-ai-agent-hits-the-market
83. AI Automation and Large Language Models - Artificial Intelligence ..., 5 月 18, 2025 にアクセス、 https://www.artificialintelligencezone.com/ai-automation/large-language-models/
84. Genspark AI Super Agent: Automate Tasks with Next-Gen AI - Toolify.ai, 5 月 18, 2025 にアクセス、 https://www.toolify.ai/ai-news/genspark-ai-super-agent-automate-tasks-with-nextgen-ai-3331995
85. Top 5 AI Agent Platforms in 2025 - Labellerr, 5 月 18, 2025 にアクセス、 https://www.labellerr.com/blog/top-ai-agent-platforms-revolutionizing-automation/
86. Top 7 Computer Use Agents - Analytics Vidhya, 5 月 18, 2025 にアクセス、 https://www.analyticsvidhya.com/blog/2025/05/computer-use-agents/
87. Meet GenSpark Super Agent: The All-in-One AI Agent that ..., 5 月 18, 2025 にアクセス、 https://www.marktechpost.com/2025/04/05/meet-genspark-super-agent-the-all-in-one-ai-agent-that-autonomously-think-plan-act-and-use-tools-to-handle-all-your-everyday-tasks/
88. Genspark AI: Revolutionizing search and automation with an intelligent super agent, 5 月 18, 2025 にアクセス、 https://anthemcreation.com/en/artificial-intelligence/genspark-ai-revolution-search-and-automate-grace-a-super-agent-intelligent/
89. Manus AI vs GenSpark AI: The Battle of Next-Gen Super Agents - MPG ONE, 5 月 18, 2025 にアクセス、 https://mpgone.com/manus-ai-vs-genspark-ai-the-battle-of-next-gen-super-agents/
90. How The GenSpark AI Agent Creates Entire Apps With ONE Click - Reddit, 5 月 18, 2025 にアクセス、 https://www.reddit.com/r/AISEOInsider/comments/1jv6yjz/how_the_genspark_ai_agent_creates_entire_apps/
91. Manus (AI agent) - Wikipedia, 5 月 18, 2025 にアクセス、 https://en.wikipedia.org/wiki/Manus_(AI_agent)

92. In-depth technical investigation into the Manus AI agent, focusing on its architecture, tool orchestration, and autonomous capabilities. - GitHub Gist, 5 月 18, 2025 にアクセス、 https://gist.github.com/renschni/4fbc70b31bad8dd57f3370239dccd58f
93. Manus AI Takes the Industry's First Step Into Full Autonomy - AI-Pro.org, 5 月 18, 2025 にアクセス、 https://ai-pro.org/learn-ai/articles/manus-ai-the-industrys-first-step-into-full-autonomy/
94. Case Study: "Manus" The First Autonomous AI Agent - ستيدي بيس, 5 月 18, 2025 にアクセス、 https://www.steadypace.sa/en/post/case-study-manus-the-first-autonomous-ai-agent
95. My Experience Building AI Agents With Manus and Genspark - Labellerr, 5 月 18, 2025 にアクセス、 https://www.labellerr.com/blog/my-experience-building-ai-agents-with-manus-and-genspark/
96. OpenAI launches responses API for AI agent development | Digital ..., 5 月 18, 2025 にアクセス、 https://dig.watch/updates/openai-launches-responses-api-for-ai-agent-development
97. Introducing deep research | OpenAI, 5 月 18, 2025 にアクセス、 https://openai.com/index/introducing-deep-research/
98. Deep Research FAQ | OpenAI Help Center, 5 月 18, 2025 にアクセス、 https://help.openai.com/en/articles/10500283-deep-research-faq
99. Manus AI: The NEW Autonomous AI Agent that is INSANE - Appy Pie Automate, 5 月 18, 2025 にアクセス、 https://www.appypieautomate.ai/blog/manus-ai-new-autonomous-agent
100. Before AI Agents Act, We Need Answers | TechPolicy.Press, 5 月 18, 2025 にアクセス、 https://www.techpolicy.press/before-ai-agents-act-we-need-answers/