自律型AIエージェント「Manus 2.0」の技術構造とフロンティアLLM（GPT・Claude・Gemini）との比較分析

Gemini 3.8 Flash

1. 概念的定義と構造的位置づけ：推論エンジンと自律実行基盤の分離

2026年9月28日、Butterfly Effect社（シンガポール）は汎用自律型AIエージェントの大規模刷新版となる「Manus 2.0」を発表した[1]。Manusは2025年3月の招待制プレビュー公開時に、指示から完了まで自律的に完遂するエージェント機能によって世界的な注目を集めた経緯を持つ[2]。その後、Metaによる買収合意とその撤回、そして2026年9月1日付での独立企業としての運営再開という複雑な経緯を経て[7]、今回のバージョン2.0へと至っている[4]。

技術的観点における最重要の前提は、Manus 2.0がOpenAIのGPTシリーズ、AnthropicのClaudeシリーズ、GoogleのGeminiシリーズのような「新規の基盤大規模言語モデル（Foundation LLM）」ではないという点である[11]。Manus 2.0はこれらフロンティアLLMを自律実行ループの「思考の中枢」として利用し、外部環境への物理的な働きかけ、ツール呼び出し、環境状態の維持を担う「エージェント・オーケストレーション基盤および作業空間」として設計されている[11]。

主要なフロンティアLLMが自然言語やマルチモーダル入力に対して推論を行い、対話形式で最適なテキストやコードを出力する「対話・推論型エンジン」として発展してきたのに対し、Manus 2.0は隔離されたクラウド仮想マシン（Ubuntu Sandbox）やローカルOSを自律操作し、成果物を納品する「行動・実行型エンジン」として機能する[12]。本分析では、Manus 2.0のアーキテクチャ特性を解剖するとともに、GPT、Claude、Geminiとの本質的な相違点を整理する。

2. Manus 2.0のアーキテクチャ特性と主要コンポーネント

Manus 2.0は単一のソフトウェア更新ではなく、実行制御ハーネス、常駐インフラ、制作ワークスペース、およびパーソナルエージェント群が統合された複合システムである[4]。

動的ツールロード基盤「Cascade」

システムの根幹には、自社開発のエージェントハーネス「Cascade」が配置されている[4]。従来のAIエージェントフレームワークでは、タスク初期段階で利用可能なすべてのツール定義やシステムプロンプトをモデルのコンテキストウィンドウに読み込ませるアプローチが一般的であり、コンテキストの圧迫と推論精度の低下を招いていた[4]。Cascadeはプロジェクトを最小限の負荷で維持し、タスクの進捗に応じて必要な専門機能やツールのみを動的に呼び出すオンデマンド型アーキテクチャを採用している[4]。開発元の検証構成（単一の特定ベンチマーク環境）では、旧システムと比較して消費トークン数を23.2%削減し、タスク完了時間を28.2%短縮、実行コストを32%削減したとされる[4]。

永続実行環境「Cloud Computer」とイベント駆動型「Automations」

従来のAIエージェントは、ユーザーのプロンプト入力や指定日時のスケジュール実行（Scheduled

Tasks）を契機として起動していた[15]。Manus 2.0ではこれをイベント駆動型へと進化させ、新着メールの受信、広告パフォーマンスの変動、カレンダーの予定、Slackメッセージ、Notionの更新といった外部SaaSのイベントをトリガーに自律実行を開始する「Automations」を実装した[1]。これらの常時監視タスクや、制作したマルチプレイヤーゲームのホスティング環境として、ローカルPCのシャットダウン後も24時間稼働を継続する専用仮想環境「Cloud Computer」が正式に組み込まれている[1]。

制作ワークスペース「Manus Studio」と人間・AI協調環境

デスクトップアプリケーションは「Manus Studio」（Windows / macOS対応）へと刷新され、AIが生成した成果物を人間が直接引き継いで修正可能なハイブリッド環境へと拡張された[1]。従来の「AIが完成品を一括出力する」一方通行のワークフローを脱し、Video EditorではAIが生成した動画のクリップ、画像、音声、字幕、モーショングラフィックスを個別のタイムライン要素として分解配置し、手動で素材の差し替えやタイミング調整を行える[1]。Video Editor内に組み込まれた「Alchemyモード」では、映像生成モデルとコード生成モデルを動的に連動させ、構図や編集テンポを高精度に制御するクリエイティブディレクション機能が提供される[1]。

また、「Game Dev」環境では映像・画像・コーディングの各モデルを統合し、プレイ可能なテンプレートからEditパネルを介してアセットやコードを直接改変し、Cloud Computerを経由してマルチプレイヤーゲームを即時公開するパイプラインを確立している[1]。さらに「Computer Use」および「Remote Control」により、スマートフォンからの音声指示を起点に手元のPC内のファイル検索やローカルアプリケーション操作を自律代行させる仕組みも統合された[1]。

独立型パーソナルエージェント「Cue」

Manus本体と並行して早期アクセス提供が開始された新アプリ「Cue」は、完全自律型のパーソナルエージェント環境である[1]。Cue上の各エージェントは、それぞれ独立した専用のメールアドレス、電話番号、デジタルウォレット、仮想マシンを割り当てられる[1]。複数の専門エージェントをグループチャットに招集し、市場調査、外部への通話や連絡、予算枠内での決済処理などを自律分担させるマルチエージェント協調が実現されている[1]。

3. 基盤LLM（GPT・Claude・Gemini）との技術的・実用的な差異

Manus 2.0とGPT（ChatGPT）、Claude、Geminiの相違点は、AI技術スタックのどの階層（レイヤー）を担当しているかという点に集約される。

比較軸	**Manus 2.0 (with Cascade)**	**OpenAI (GPT-5 / ChatGPT)**	**Anthropic (Claude / Claude Code)**	**Google (Gemini / Workspace)**
主要レイヤー	自律実行・オーケストレーション基盤[11]	基盤推論モデル＋汎用対話UI[14]	基盤推論モデル＋開発者特化エージェント[12]	ネイティブマルチモーダル基盤モデル[14]
基盤モデルの	外部モデル動的ルーティング	自社モデル（GPT-5、	自社モデル（Opus 4.7、	自社モデル（Gemini 3 Pro /

利用形態	（Claude, GPT, Gemini等）[12]	o-series等）を直接実行[14]	Sonnet 4.6等）を直接実行[12]	Flash等）を直接実行[21]
実行環境の状態管理	永続的なUbuntu仮想マシン、Cloud Computer、ローカルOS[1]	一時的コンテナ（Advanced Data Analysis等）[14]	開発者のローカルターミナル環境（Claude Code）[12]	Google Cloudサーバーレス実行基盤
ブラウジングと調査機能	Wide Research（複数サブエージェントによる並列分散探索）[12]	Deep Research / 逐次検索[12]	逐次ブラウジング / 外部ツール連携[12]	Google Search統合型リアルタイム検索[22]
タスク起動トリガー	プロンプト指示、定時実行、外部SaaSイベント[4]	ユーザープロンプト入力依存[14]	ユーザーコマンド入力依存（CLI/API）[12]	ユーザープロンプト入力依存（一部Workspace自動化）
成果物の形態	ライブ公開URL、計算式入りXLSX、PPTX、動画タイムライン[1]	テキスト、Markdown、コードブロック、生成画像[13]	コード差分（PR/コミット）、分析テキスト[12]	テキスト、ドキュメント/スプレッドシート下書き[13]
人間の介入・編集方式	Studio上で動画・ゲーム・スライドを視覚的に直接修正可能[1]	チャットを通じた再プロンプト指示[13]	IDE/エディタ上での手動修正および対話修正[12]	Workspace各アプリ上での直接編集

モデルのオーケストレーションと動的ルーティング

GPT、Claude、Geminiは深層学習によって直接事前学習・強化学習された知能モデルそのものである[11]。これに対してManus 2.0は、これら外部モデル群をAPI経由で適材適所に使い分けるシステムである[11]。論理的推論や複雑な分析タスクにはAnthropicのClaudeファミリー（Claude Opus 4.7、Claude 3.5/3.7等）を主軸として割り当て、高速処理や構造化出力、汎用コーディングにはOpenAIのモデル（GPT-5等）やGoogleのGemini 3 Flash、AlibabaのQwenモデルを使い分けるルーティングを行う[12]。Manusは各社モデルの競合相手というよりも、それらを下位モジュールとして統制する上位コントローラーに位置づけられる[12]。

ステートフルな実行環境と環境適応ループ

ChatGPTやGeminiの一般的な利用は、プロンプトの送受信を基本とするステートレスな対話である

[12]。API経由のツール呼び出し機能は備わっているものの、セッションを跨いで永続する環境は限定的である。一方、Manus 2.0はタスクごとに独立した仮想マシン（Sandbox）やローカルファイルシステムを確保し、本物のシェル環境、ヘッドレスブラウザ、ファイルストレージを保持する[12]。ブラウザで外部サイトを遷移してフォームを入力し、取得したデータをスクリプトで集計し、実行エラーが発生した場合は自動で修正して再試行するという一連のフィードバックループが人間の介在なしに完結する[6]。

成果物の抽象度とデプロイパイプライン

成果物の性質においても大きな隔たりが存在する。例えばWebアプリケーションの構築を依頼した場合、ClaudeやGPTは完成度の高いソースコードをチャットウィンドウに提示し、環境構築やデプロイ作業はユーザーに委ねる[12]。対してManus 2.0は、仮想マシン内でプロジェクトの立ち上げ、依存関係の解決、コードのビルド、テスト実行まで自律的に行い、実際にインターネット経由でアクセス可能なライブURLを発行して納品する[12]。リサーチタスクにおいても、単一のテキスト要約にとどまらず、「Wide Research」によって数十～数百のソースを並列スクレイピングし、検証を経て計算式付きExcelファイル（XLSX）や整形済みプレゼンテーション（PPTX）としてファイル化する[12]。

4. 定量的性能評価、ベンチマーク、および実運用環境の経済性

ベンチマーク評価（**GAIA**および**OSWorld**）

汎用AIアシスタントの複合タスク遂行能力を評価するGAIA（General AI Assistant）ベンチマークにおいて、Manusは初期公開時から92.7%のタスク完了率を記録し、OpenAI Deep Research（74.3%水準）や従来SOTAを上回る結果を示した[6]。特にマルチモーダル処理と高度な推論が要求される最難関のLevel 3タスクにおいても57.7%（対OpenAI 47.6%）を達成している[27]。また、Chrome、VS Code、LibreOfficeなど実OS上のアプリケーション操作を伴うOSWorldベンチマークにおいても、GUI要素の特定やDOMツリー解析を駆使した自律操作能力を実証している[28]。

実運用上の信頼性と「クレジットバーン」

ベンチマーク上の数値が高い一方で、実際の業務環境における安定性には一定の制限が存在する。現場のテスト運用では、複雑なオープンエンドリサーチにおいて高い成果を出す割合は約60%にとどまり、残り40%では平凡または実用に耐えない出力に終わるという指摘がある[29]。特にログイン認証が必要なWebサイト、CAPTCHA認証、厳格なBot排除機構に直面すると、自律ブラウジングが停止する事象が頻発する[5]。

さらに、従量課金体系に起因する「クレジットバーン」も実用上の課題である[5]。単純な質問処理であれば10～50クレジット程度で推移するが、複数エージェントが並行して深掘り調査を行うWide Researchでは、1回の実行で500～900クレジット以上を急速に消費する[17]。実際の消費量が当初の見積もりを2～4倍上回るケースが報告されており、実行途中でクレジット上限に達した際にタスクが中断され、成果物が未完成のまま停止する運用上のリスクを孕んでいる[5]。

提供プラン	月額料金（**USD**）	クレジット付与量・実行制限	主な機能範囲と制約

Free	0ドル[11]	毎日300クレジット（月間消費上限1,500）[11]	Agentモードは Manus 2.0 Liteに限定。同時タスク1件、動画生成・Wide Research利用不可[11]。
Pro (Tier 1)	20ドル[11]	4,000クレジット/月[11]	Manus 2.0 / 2.0 Max利用可能、Wide Research対応、動画・ゲーム制作対応[11]。
Pro (Tier 2/3)	40ドル / 200ドル[11]	8,000クレジット～ / 40,000クレジット[11]	大規模リサーチおよびメディア制作向け拡張クレジット枠[11]。
Team	20ドル/席～（年払17%引）[17]	クレジットチーム内共有、自動購入設定可[18]	顧客データ学習除外、管理ダッシュボード、監査ログ。※30席未満のSSO利用は月額150ドルの追加費用が必要[17]。
Cloud Computer	個別課金（未公表）[2]	24時間常駐インスタンス単位[4]	Automationsの常時実行ハブ、マルチプレイヤーゲームのホスティング環境[4]。

5. 自律化が招くセキュリティ脅威とエンタープライズ統制

Manus 2.0が備える「外部SaaS連携」と「イベント駆動型自動実行」は利便性を大幅に向上させた反面、セキュリティ上の新たな脅威サーフェスを露出させている[17]。

メール起点の間接的プロンプトインジェクション

2026年9月下旬、セキュリティ調査機関Salt Labsによって公表された脆弱性は、自律型エージェント特有の構造的弱点を浮き彫りにした[17]。攻撃者が悪意ある命令を難読化されたJavaScriptコードとしてGmail等に送信し、ユーザーがManusにメールの確認・要約を依頼するか、Automationsが新着メールを検知して自動処理を開始した際にエクスプロイトが発生した[17]。Manusは受信したメールを受動的なテキストデータとしてではなく、実行可能なシステム命令として誤認し、難読化技術によってガードレールをすり抜けてコードを実行した[30]。

これにより攻撃者はクラウドSandbox内にリバースシェルを確立し、ユーザーが連携していたGmail、

Google Drive、GitHubなどのOAuthトークンや認証情報を取得可能な状態に陥った[30]。この脆弱性はMetaのバグバウンティプログラム経由で報告され、Manus 2.0正式公開前に修正されたが[17]、「不審な挙動を検知してアラートを発した時点では、自律エージェントがすでに悪意ある外部通信を完了している」という実行速度の非対称性が根本的な課題として顕在化した[30]。

情報システム部門における統制フレームワーク

エンタープライズ環境でManus 2.0を安全に運用するためには、利便性とセキュリティリスクを天秤にかけた厳格なポリシー設計が求められる。不特定多数の外部送信者から届くメールをトリガーとする自動化は原則として禁止し、社内イントラネットや承認済みデータソースからのイベントのみに限定する運用が推奨される[17]。また、Google WorkspaceやSlackを連携させる際も、書き込みや送信の権限を付与せず、読み取り専用の最小特権スコープに絞る必要がある[17]。

さらに、Cloud Computer上で動き続けるバックグラウンド自動化スクリプトや連携認証は、作成した従業員が退職・異動しても自動的に停止しないため、ID管理と連動したアカウントおよび接続の強制遮断プロセスを整備しなければならない[18]。また、エージェントが独自の電話番号やウォレットを保持して自律的な外部通話や決済を行う「Cue」については、企業のガバナンスと予算統制が確立されるまでの間、公式な業務利用を制限し、検証目的の個人試用にとどめる判断が現実的である[15]。

6. 結論

Manus 2.0は、AI業界の競争軸が「基盤モデルの純粋な知能・推論性能」から「知能を行動に変換するオーケストレーション基盤の完成度」へと移行したことを象徴するプロダクトである。

GPT、Claude、Geminiが人間の思考を補完・拡張する「対話型の知能パートナー」であるのに対し、Manus 2.0はそれらのフロンティアLLMを手足となるツールとして統率し、クラウド仮想マシンやローカル環境を駆使して具体的な成果物を納品する「自律型のデジタル労働者」として機能する[11]。

したがって、実務における最適解は単一ツールの選択ではなく、多層的なハイブリッド運用にある[22]。戦略立案や高度な壁打ち、緻密なリファクタリングにはClaudeやGPTを直接活用し、複数ソースからの並行リサーチ、スライドやWebアプリの自動構築、イベント駆動型の反復処理にはManus 2.0を適用するアプローチが合理的である[12]。その際、間接的プロンプトインジェクションへの防御策やクレジット消費の予算上限管理を講じることが、自律型エージェントを安全に定着させるための前提条件となる[17]。

引用文献

1. 汎用AIエージェント「Manus 2.0」が発表 ～デスクトップアプリは, https://forest.watch.impress.co.jp/docs/news/2144171.html
2. Cascade・Manus Studio・自動化・Cloud Computer・Cueを解説, https://techgym.jp/column/manus-2-0/
3. Manus 2.0、アーキテクチャ級アップデート：自社開発「Cascade, https://finance.biggo.jp/news/f4b75310-5221-4d6a-a3f9-755047e52065
4. Manus 2.0 のご紹介, https://manus.im/ja/blog/introducing-manus-2-0
5. What Is Manus AI? How to Use It, Pricing, and an Honest Review, https://penchan.co/en/ai/agent/manus/
6. 【2026年8月最新】AIエージェント「Manus（マナス）」完全ガイド, https://genai-ai.co.jp/ai-kanri/blog/cc-manus-reading-guide/
7. AIエージェント「Manus 2.0」リリース - GIGAZINE,

https://gigazine.net/news/20260929-manus-2-0/
8. Manus AIの危険性｜安全性と企業利用の可否【2026年9月】, https://uravation.com/media/manus-ai-safety-risks-2026/
9. Manus AI Review (2026): Strengths, Failures & the Meta Split - Carly AI, https://www.usecarly.com/blog/manus-ai-review/
10. Manus Leaves Meta: Back Up Your Data Before Aug 23, 2026, https://explainx.ai/blog/manus-meta-separation-data-deletion-backup-guide-august-2026
11. Manus 2.0とは？LLMとの違いや新機能、無料・有料プランで, https://note.com/kazu_t/n/n70df5c012cd3
12. Manus vs Claude Code: Which Tool is Right For You?, https://manus.im/compare/manus-vs-claude-code
13. Manus Pro AI Agent 2026: Advanced Capabilities vs Gemini & Claude, https://integrated.social/blog/manus-pro-ai-agent-advanced-capabilities-vs-gemini-claude-2026/
14. Manus AI is better than ChatGPT, Gemini and Claude. Here ... - Reddit, https://www.reddit.com/r/ThinkingDeeplyAI/comments/1rby6ah/manus_ai_is_better_than_chatgpt_gemini_and_claude/
15. Manus 2.0とは｜Cascade・Cue・変更点と注意点【2026年9月】, https://aigentlab.tech/articles/manus-2-0-cascade-cloud-computer-studio-cue-guide-2026/
16. Manus 2.0 Isn't About the First Result. It's About Everything After., https://gnana70.medium.com/manus-2-0-isnt-about-the-first-result-it-s-about-everything-after-c6f6ff2910b2
17. Manus 2.0 Lets Email Trigger Agents Days After an Email Hijack Bug, https://www.beri.net/article/manus-2-0-email-triggered-automations-cue-agents-prompt-injection-team-plan-sso
18. 「Manus 2.0」公開｜新機能・料金や新アプリ「Cue」と情シスの, https://admina.moneyforward.com/jp/blog/manus-2-0-enterprise-guide
19. Manus 2.0発表、AIが作った動画やゲームを人が直接編集可能に, https://xenospectrum.com/manus-2-0-editable-studio-automation/
20. AI競争の最前線「常時動作型AIエージェント」とはなにか Dots, https://www.watch.impress.co.jp/docs/series/nishida/2145835.html
21. Built-in AI Capabilities - Manus Documentation, https://manus.im/docs/website-builder/ai-capabilities
22. Manus・Claude・ChatGPT・Gemini、ビジネスで使うならどれ？ 4, https://note.com/nahouemura/n/ne55c1386d19e
23. In-depth technical investigation into the Manus AI agent, focusing on, https://gist.github.com/renschni/4fbc70b31bad8dd57f3370239dccd58f
24. Manus AI: The Dawn of Autonomous Agents and What It Means for, https://aiwhisperer.org/blog/manus-ai-the-dawn-of-autonomous-agents-and-what-it-means-for-business
25. "Best AI Agents in 2026: 14 Tools for Autonomous Work" - Rework, https://resources.rework.com/tools/ai-tools/best-ai-agents-2026
26. Manus AI 完全解説 - 機能一覧・スライド生成機能・教育現場での,

https://xnjnaute.manus.space/
27. Manusとは？できること・強み・実際の活用法 | willB Blog, https://willb-inc.com/blog/manus-ai-agent-guide/
28. うさぎでもわかる NeurIPS 2024採択！OSWorldでAIエージェントの, https://note.com/taku_sid/n/n1de18326b34b
29. Claude vs ChatGPT vs Gemini vs Perplexity (2026) - OneWave AI, https://www.onewave-ai.com/blog/ai-showdown-claude-chatgpt-gemini-perplexity-manus
30. Salt Labs Research: A Single Email Could Hijack an AI Agent and, https://cyberdefensewire.com/salt-labs-research-a-single-email-could-hijack-an-ai-agent-and-reach-a-users-connected-accounts/
31. How We Hijacked an AI Agent With a Single Email, https://securityboulevard.com/2026/10/how-we-hijacked-an-ai-agent-with-a-single-email/
32. Prompt-Injection Bug Hits $4B Agentic AI App 'Manus' - Dark Reading, https://www.darkreading.com/application-security/prompt-injection-bug-agentic-ai-app-manus
33. 【2026年8月最新】Claude Code vs Manus 徹底比較, https://genai-ai.co.jp/ai-kanri/blog/cc-vs-manus-08/
34. AIツール比較 - AI鬼管理 - 株式会社GENAI, https://genai-ai.co.jp/ai-kanri/blog/category/ai-tools-comparison/page/101/