ChatGPT

Manus 2.0とGPT・Claude・Geminiの技術比較調査

方法論

本調査は**2026年10月7日（日本時間）時点**の公開情報を対象とし、情報源を①各社の公式発表・日本語公式発表、②公式API仕様・モデルカード・システムカード・Trust Center、③公式利用規約・料金表、の順で優先した。Manusについては2026年9月28日のManus 2.0発表だけでなく、翌9月29日のManus Flex、Manus API v2、Trust Centerまで追跡した。GPT、Claude、Geminiについても、それぞれOpenAI、Anthropic、Google/Google DeepMindの一次資料を基準とした。[^M1][^M2][^O1][^A1][^G1]

比較上、最も重要な注意点がある。**Manus 2.0とGPT／Claude／Geminiは、厳密には同じレイヤーの製品ではない。** Manus 2.0で新たに導入された「Cascade」は公式に**agent harness（エージェント・ハーネス）**として説明されており、モデルそのものよりも、モデル選択、コンテキスト管理、ツール、実行環境、成果物作成、継続タスクを統合する上位システムである。一方、GPT、Claude、Geminiは主として基盤モデル／モデルAPIである。Manus自身もManus Flexの説明で、従来Manusが「model selection, harness, agent infrastructure」をまとめて担当してきたと説明し、Flexでは推論プロバイダーとモデルを利用者側で選べるようにしている。したがって、パラメータ数や学習データ量をそのまま横一列に比較することには限界がある。[^M1][^M2] [1]

「最新公開版」の定義についても統一した。OpenAIは**GPT-6.1 Sol（2026年9月29日）**を最新リリースとして扱うが、公式モデル選択ガイド上の最上位flagshipはGPT-6 Astraである。Anthropicは**Claude Sonnet 5.5（9月28日）**が最新の一般モデル。Googleは**Gemini 4 Argon**が最新発表モデルだが、現時点ではFairwind Programを通じた限定展開であり、広く利用できるGemini APIとの実用比較には**Gemini 3.8 Flash**を併記する。[^O1][^O2][^A1][^G1][^G2] [2]

エグゼクティブサマリー

- **Manus 2.0最大の差は「より賢い単一LLM」ではなく「仕事を最後まで実行するエージェント・システム」にある。** Cascade、Manus Studio、Cloud Computer、Remote Control、Automations、Video Editor、Game Dev、Cueが一体化し、モデルの返答を文章で終わらせず、ファイル、Webサイト、コード、動画、ゲーム、外部アプリ操作などの成果物・行動へ変換する設計である。[^M1] [3]

- **したがってManus 2.0の「モデルサイズ・学習データ・コンテキスト長」はGPT、Claude、Geminiのようには公開されていない。** Cascadeは基盤モデルとして仕様化されておらず、パラメータ数、独自モデルの学習データ、RLHF手法、単一のcontext windowは公式資料上未公開である。さらにManus Flexでは推論モデルを外部プロバイダーから選択できるため、Manusの能力は「基盤モデル × Cascade × ツール × 実行環境」の合成能力として見るべきである。[^M2][^M3] [4]

- **モデル単体の透明性と制御性ではGPT／Claude／Geminiの方が明確である。** GPT-6.1 Solは1.05Mコンテキスト、128K最大出力、Claude Sonnet 5.5は1M／128K、Gemini 3.8 Flashは1,048,576／65,536を公式に規定している。一方Manus 2.0は総エージェント文脈長を公開していない。[^O2][^A2][^G2] [5]

- **エージェント型業務の「完成までの時間・コスト」ではManusが独自の最適化軸を持つ。** CascadeについてManusは旧システムに対し、特定テスト構成でトークン23.2%減、タスク完了時間28.2%減、実行コスト32%減を報告している。ただしこれはManus自身が明記するとおり**特定構成における旧**

Manusとの比較であり、GPT／Claude／Geminiに対する外部ベンチマークでも、全タスク平均でもない。[^M1] [3]

- **安全性の思想もレイヤーが異なる。** OpenAIはPreparedness系の能力評価とモデルレベルのalignment、AnthropicはConstitutional AIと行動監査、GoogleはFrontier Safety Frameworkとモデルカードによる危険能力評価を公開している。ManusはSOC 2 Type II、ISO 27001/27701、アクセス制御や実行環境などサービス／インフラ安全性を前面に出す一方、CascadeについてRLHF等のモデルalignmentレシピは公開していない。[^M4][^O3][^A3][^G3] [6]

- **導入判断では「ManusかGPT/Claude/Geminiか」という二者択一にしない方がよい。** 自社でエージェント基盤を構築・統制したいならGPT／Claude／GeminiのモデルAPI、完成成果物まで素早く委譲したいならManus、そしてManus Flexを利用できる環境では**外部モデル＋Manusのエージェント実行層**というハイブリッドが構造的に合理的である。これはManus自身がFlexでモデル選択とagent systemを分離したことから導ける実務上の結論である。[^M2] [7]

詳細比較表

表の読み方:「未公開」は、2026年10月7日までに確認した当該一次資料で数値・仕様が明示されていないものを意味する。「適用不可」はManus 2.0が単一基盤モデルではないため同じ指標を直接適用できない場合を示す。Gemini欄は、最新発表の**Gemini 4 Argon**が限定提供中のため、APIで一般利用できる**Gemini 3.8 Flash**の実装仕様を併記している。[^G1][^G2] [8]

比較軸	Manus 2.0	GPT（最新公開版）	Claude（最新公開版）	Gemini（最新公開版）
比較対象の正体	**エージェント・システム／実行プラットフォーム。** 中核はCascade agent harness。Studio、Cloud Computer、Remote Control、Automations、Cue等を統合。[^M1] [9]	**基盤モデル。** 最新リリースはGPT-6.1 Sol。OpenAI公式ではGPT-6 Astraが最上位flagship、Solは知能・コストのバランス型。[^O1][^O2] [10]	**基盤モデル。** Claude Sonnet 5.5。Anthropicは速度と知能のバランス型として位置付け、より複雑・開放的な判断にはOpus 5.5を上位に置く。[^A1][^A2] [11]	最新発表は**Gemini 4 Argon**だが限定展開。一般APIの主要最新モデルは**Gemini 3.8 Flash**。[^G1][^G2] [8]
公開日・バージョン履歴	Manus 2.0: **2026-09-28**。Manus Flex: **09-29**。[^M1][^M2] [1]	GPT-6 Astra: 09-03、GPT-6.1 Sol: **09-29**。[^O2] [12]	Opus 5.5: 09-22、Sonnet 5.5: **09-28**。Sonnet 5.5はActive (latest)。[^A1][^A2] [13]	Gemini 3.8 Flash: **09-02**。Gemini 4 Argonは英語発表後、日本公式記事が**10-01**公開。[^G1][^G2] [14]

比較軸	Manus 2.0	GPT（最新公開版）	Claude（最新公開版）	Gemini（最新公開版）
アーキテクチャ	Cascadeは「agent harness」。必要な専門能力を必要な時点で呼び出し、タスク・ツール・実行環境を編成。基盤モデルの具体構成は**未公開**。[^M1][^M2] [1]	GPT-6.1 Solの詳細ネットワーク構造は**未公開**。OpenAIはモデルAPIとして提供。[^O1] [15]	Sonnet 5.5の詳細な層構成等は**未公開**。公式仕様は機能・context・thinking等を公開するが内部構造は開示していない。[^A2] [16]	3.8 Flashは3.7 Flashを基礎とするとModel Cardに記載。3.8固有の詳細アーキテクチャは3.7カードを参照する構成で、3.8ページ自体には詳細を再掲していない。[^G3] [17]
モデルサイズ／パラメータ数	**適用不可／未公開。** Cascadeをパラメータ数付き基盤モデルとして発表していない。[^M1] [9]	**未公開。** [^O1] [15]	**未公開。** [^A2] [16]	**未公開。** 3.8 Model Cardにもパラメータ数の記載なし。[^G3] [17]
トレーニング手法	CascadeについてRLHF、SFT、事前学習等の具体的レシピは**未公開**。Flexでは外部推論モデルを選べるため、そもそも推論モデルとManusのagent harnessを分けて考える必要がある。[^M2] [7]	GPT系reasoning modelでは強化学習を用いたreasoning/alignmentをOpenAIが説明しているが、GPT-6.1 Sol固有の完全な学習レシピは**未公開**。[^O3] [18]	Constitutional AIが現在もClaude trainingの中心的役割を持つ。Anthropicは2026年の研究でchat RLHFに加え、constitutionally aligned documents、chat data、agentic environments等によるalignmentを説明。ただしSonnet 5.5の完全な学習レシピは**未公開**。[^A3] [^A4] [19]	3.8発表では長時間agentic loopsによる反復評価・改善やcybersecurity領域での厳格なtrainingを説明。完全な学習レシピは**未公開**。[^G4] [20]

比較軸	Manus 2.0	GPT（最新公開版）	Claude（最新公開版）	Gemini（最新公開版）
学習データソース	Cascade独自の学習データは**未公開**。[^M1] [9]	OpenAIの基盤モデル一般では、公開インターネット、第三者パートナーからアクセスした情報、ユーザー・人間トレーナー・研究者が提供／生成した情報を主要カテゴリとして説明。GPT-6.1固有配分は**未公開**。[^O4] [21]	Sonnet 5.5はtraining data cutoffが**2026年6月**。Anthropic Trust CenterにはSonnet 5.5 Training Data Summaryの存在が示されるが、モデル概要では詳細配合は公開されていない。[^A2] [22]	3.8 Flash Model Cardはtraining dataset／processingについて3.7 Flashカードを参照。**3.8固有の追加データ内訳は未公開**。[^G3] [17]
マルチモーダル	文書、スプレッドシート、PDF、スライド、Web、コード、動画、ゲーム等を**製品／ツールとして横断処理**。ただし「単一Cascadeモデルのnative modalities」としては仕様化されていない。[^M1] [9]	GPT-6.1 Solは**text＋image入力→text出力**。当該モデルはaudio非対応。[^O1] [15]	Sonnet 5.5は**text＋image→text**。[^A2] [16]	3.8 Flashは**text、image、video、audio、PDF入力→text出力**。画像・音声生成は当該モデル自体では非対応。[^G2] [23]
長文コンテキスト	統一されたCascadeのcontext windowは**未公開**。Manus API `task.create`の直接promptは約5,000推定token上限だが、file/task references等を別途扱えるため、これは基盤モデルcontext windowとは別概念。[^M5] [24]	**1,050,000 tokens** context、最大出力**128K**。[^O1] [15]	**1M** context、最大出力**128K**。Batch API betaは最大300K output。[^A2] [16]	3.8 Flash: input **1,048,576**、output **65,536**。[^G2] [23]

比較軸	Manus 2.0	GPT（最新公開版）	Claude（最新公開版）	Gemini（最新公開版）
推論／生成品質	標準化されたGPT/Claude/Gemini対抗のモデル品質ベンチマークは**未公開**。Manus 2.0の公式数値は主に旧Manusに対するシステム効率改善。[^M1] [25]	OpenAIはSolをAstraに近い知能をより低コストで提供するモデルとして位置付け、agentic coding、computer use、professional workを主な強化点とする。[^O2] [26]	Sonnet 5.5はcoding、長時間タスク、image understanding、knowledge workを強化。ただしAnthropic自身、複雑でopen-endedな持続判断ではOpus 5.5が上と明記。[^A1] [27]	3.8 Flashはlong-horizon coding、autonomous agent、enterprise workflow向け。GoogleはHLE-Verified 54.9%等を公表するが、ベンチ構成が各社で異なるため単純順位化は不可。[^G3][^G4] [28]
速度／レイテンシ	特定テスト構成で旧システム比**タスク完了時間28.2%減**。絶対レイテンシ／SLAは**未指定**。[^M1] [3]	GPT-6.1 Sol標準APIの一律の絶対レイテンシ値は**未指定**。reasoning effort等で速度／品質を調整。[^O1] [15]	公式分類は**Fast**、Sonnet 5比で**30%以上高速**。ms単位等の絶対値は未指定。[^A1][^A2] [11]	Googleは3.7 Flashと「同じspeed」を維持すると説明。絶対値は**未指定**。高effortでは多くのreasoning steps/tokenを使い、Model Cardは時にslowness／timeoutがあり得るとする。[^G3] [^G4] [29]
安全性・alignment	CascadeのRLHF等は**未公開**。サービス側ではSOC 2 Type I/II、ISO 27001、ISO 27701等を公表。Remote Control等は許可された環境を操作する設計。Flex利用時は選択した基盤モデル側の安全性も全体の挙動に影響すると考えるべき。[^M4] [30]	GPT-6.1 SolはOpenAIのDeployment Safety評価対象で、GPT-6 Astraと同系統の防御スタックを利用。OpenAIはpretraining data composition、RL grading、model policies、安全評価を組み合わせる。[^O3] [31]	Constitutional AI、約1,850シナリオのautomated behavioral audit、cyber safeguards、biology safeguards、reasoning-extraction対策classifier、preserved thinkingを採用。[^A1][^A3] [32]	3.8 FlashはCBRN／cyber misuse safeguards、Frontier Safety Framework、prompt-injection robustness、専門red teamingを採用。[^G3][^G4] [29]

比較軸	Manus 2.0	GPT（最新公開版）	Claude（最新公開版）	Gemini（最新公開版）
API	API v2、base URL `https://api.manus.ai`。Tasks、Projects、Files、Webhooks、Skills、Agents。`task.create`は非同期実行。API keyまたはOAuth2。[^M5][^M6] [33]	Responses APIがtool利用の中心。Chat Completionsも利用可能だがGPT-6.1 Solではtool calling上の制約あり。[^O1] [15]	`claude-sonnet-5-5`としてClaude APIで提供。Batch API対応。AWS Bedrock、Google Cloud、Microsoft Foundry、Claude Platform on AWSでも提供。[^A2] [16]	`gemini-3.8-flash`。Batch、Flex、Priority inference対応。function calling、code execution、file search、Computer Use preview等。[^G2] [23]
外部インテグレーション／ツール	Manusの強み。Slack、Notion、Zapier、Asana、monday.com、Make、Linear、GitHub、Gmail、Calendar、Drive、Vercel、Stripe等、多数のOAuth connectorを公式API側で列挙。[^M7] [34]	web search、file search、code interpreter／hosted shell、computer use、MCP、skills等をResponses API側から利用可能。[^O1] [15]	Tool use、computer use、Claude Code、クラウドプラットフォーム統合。Sonnet 5.5はagentic/tool workloadsを主要用途に含む。[^A1] [^A2] [11]	Search／Maps grounding、function calling、code execution、computer use、file search、URL context等をモデルAPIで標準サポート。[^G2] [23]
API料金	Manus 2.0/API v2の**明確なtoken単価は現在のAPI資料では未指定**。Flexではモデル推論分は選択したproviderから課金され、Manus側のサービス／infraはManus creditsを消費。[^M2] [7]	GPT-6.1 Sol標準: **input $2／1M、output $10／1M**。cached input等の別価格あり。長大promptやBatch/Flex等で価格体系が変化。[^O1] [15]	Sonnet 5.5: **input $2／1M、output $10／1M、cache read $0.20／1M**。Batchはinput/output 50%割引。[^A1][^A5] [35]	3.8 Flashは2026-12-31まで**input $0.75／1M、output $3.75／1M**、2027-01-01から$1.50／$7.50予定。Batch/Flexは半額水準。Argonの導入価格は$2／$10だが現状限定提供。[^G1][^G5] [36]

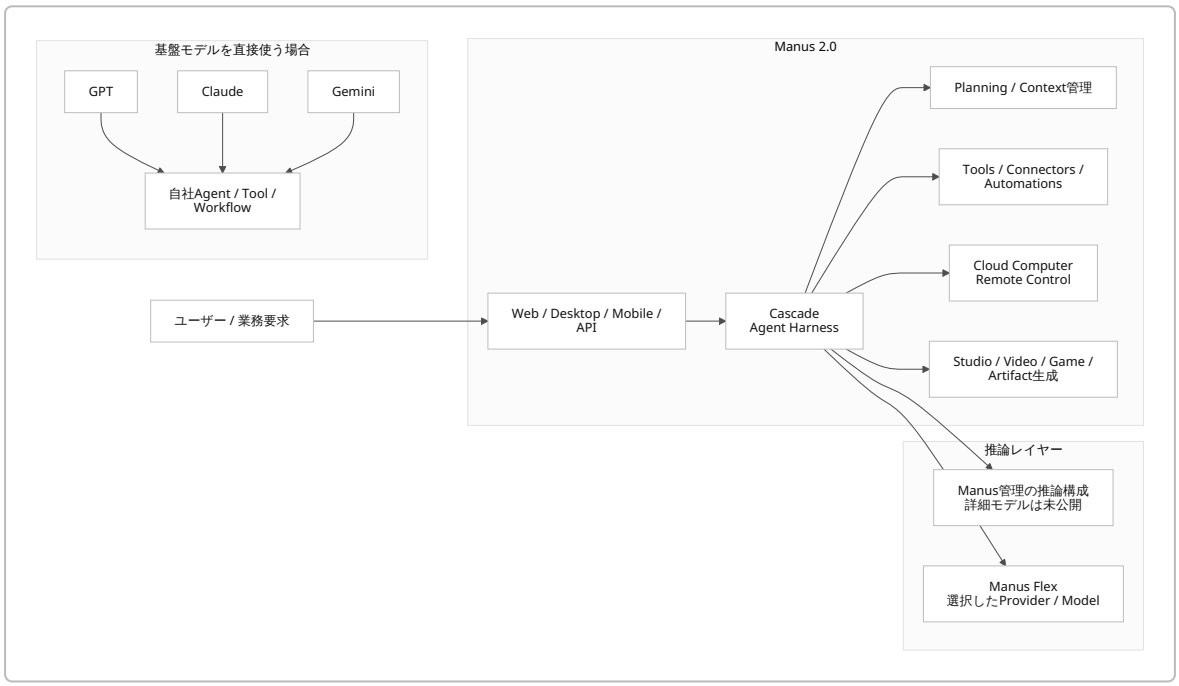
比較軸	Manus 2.0	GPT（最新公開版）	Claude（最新公開版）	Gemini（最新公開版）
ライセンス・利用規約	Proprietary SaaS/API。Manus 2.0/Cascadeのモデル重みやopen-weight licenseは発表されていない。利用はManus Termsに従う。[^M8] [37]	Proprietary service/API。OpenAI利用規約は基盤モデル等のreverse engineering等を制限。個人向け規約では、法令上可能な範囲でInputの権利を保持しOutputをユーザーに帰属させる。APIはService/Business Termsが適用される。[^O5] [38]	Proprietary service/API。現行仕様ではモデルはAnthropic APIおよび各クラウドを通じて配布され、open-weight licenseとしての提供は記載されていない。[^A2] [16]	Gemini API Additional Termsが適用。underlying model/weightsのreverse engineering・抽出を禁止。Googleは生成したoriginal contentについて所有権を主張しない。Paid Servicesではprompt/responseを製品改善に利用しない一方、無料サービスは条件が異なる。[^G6] [39]
主要ユースケース	「依頼→調査→ツール操作→成果物→公開／継続運用」までのend-to-end業務。資料、Web、コード、動画、ゲーム、automations、desktop操作、Cue agents。[^M1] [9]	自社アプリへの推論、coding、computer use、tool orchestration、professional knowledge work。より高難度ではAstraを選択可能。[^O1][^O2] [40]	coding、knowledge work、bug fixing、文書・スライド・表計算、長時間agent task。複雑なopen-ended判断はOpus 5.5がより適する。[^A1] [27]	3.8 Flashはsoftware engineering、autonomous agents、enterprise workflowsを低コストで大量処理。Argonはより複雑なsoftware/legal/finance/cybersecurity workflowを狙う。[^G1][^G4] [14]
主な制限	基盤モデル、総context、学習データ、parameter数、absolute latency、2.0 APIの明確なtoken料金、外部標準benchmarkが**未公開／未指定**。自律操作では権限設計・監督が重要。[^M1][^M2] [41]	モデル内部構造／parameter数非公開。Solは当該モデル単体ではaudio非対応、fine-tuning非対応。[^O1] [15]	parameter数非公開。高リスクcyberではfallbackが生じ得る。一部microbiology/virology要求が誤検知される可能性をAnthropic自身が明記。[^A1] [27]	3.8は高effortでtoken利用が増え、slowness/timeoutの可能性。Argonはまだ一般公開ではなく、現時点で通常のGemini API代替として扱えない。[^G1][^G3] [42]

Manusの「コンテキスト5,000 tokens」についての注釈

Manus API v2の `task.create` にある約5,000推定tokenという制限を、「Manus 2.0のcontext windowが5,000」と解釈するのは誤りである。これは**APIで一回に送る直接prompt部分の制約**であり、Manusは Files、Projects、過去Task referenceなどを別のコンテキスト源として扱う。またCascadeそのものの内部context windowは公式発表では公開されていない。したがってGPTの1.05M、Claudeの1M、Geminiの約1Mと5,000を直接比較すべきではない。[^M5] [43]

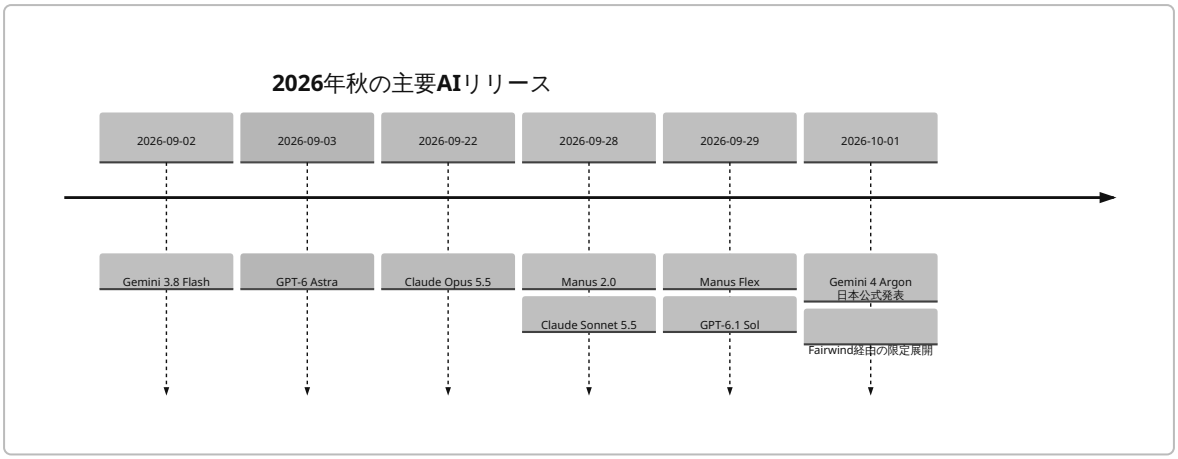
アーキテクチャと公開履歴

Manus 2.0を理解する最も有効な方法は、LLMそのものではなく、**LLMを仕事へ変換するオーケストレーション層**として描くことである。Manus Flexの公式説明では、Manusが担うものとしてmodel selection、harness、agent infrastructureが明示され、Flexではモデル／inference providerをユーザー側が選択する一方、Manusはplanning、tool use、execution environments、completed tasksへの変換を引き続き担当すると説明される。[^M2] [7]



この図で重要なのは、**GPT／Claude／Geminiを直接利用する場合には、右下の「自社Agent / Tool / Workflow」に相当する部分を導入企業が設計するのに対し、Manusはその部分を製品としてかなり大きく引き受ける**点である。逆に言えば、モデル単体を細かく制御したい企業から見ると、Manusは一段高い抽象化層であり、内部選択や挙動の透明性とのトレードオフが生じる。Manus Flexはそのトレードオフを緩める方向の機能である。[^M2] [7]

主要リリースを時系列にすると、Manus 2.0は単独の出来事ではなく、主要AIベンダーがほぼ同時期に「モデル単体」から「長時間エージェント／computer use／professional workflow」へ重点を移していたタイミングに登場したことが分かる。[^M1][^O2][^A1][^G1] [44]



Geminiだけは「最新」と「実際に誰でもAPI採用可能な最新」を分ける必要がある。Google日本公式の2026年10月1日記事ではGemini 4 Argonを新しいfrontier modelとして発表しているが、まずFairwind Programを通じて信頼できるcybersecurity担当者へ段階的に展開し、開発者・企業・一般消費者への広範なアクセスに向けてguardrailを改良中としている。したがって**2026年10月7日時点でArgonを一般的な本番API候補としてManusやGPT-6.1 Sol、Claude Sonnet 5.5と同列に扱うのは早い**。[^G1] 45

技術的考察

Manus 2.0の本質は「モデル性能」より「コンテキストと実行の経済性」

CascadeでManusが強調している23.2%のtoken削減、28.2%のtask completion time削減、32%のexecution cost削減は、典型的なLLMの「正答率を何ポイント上げた」という発表と方向が違う。これは**必要な能力だけを必要な時にロードし、タスク中の文脈やツールを過剰に抱えないagent harness設計**の改善として提示されている。ManusのHelp Centerも、この数値が特定のテスト構成に基づき、全体平均ではないことを明記している。[^M1] 3

これは長時間エージェントでは重要である。基盤モデルの1回の推論品質だけでなく、「何回モデルを呼ぶか」「不要なcontextを何度送り直すか」「tool callを何回失敗するか」「ブラウザ・コード・ファイル間をどう状態遷移させるか」が、最終的な料金と完了時間を大きく左右するからである。Manus 2.0は、この**task-level efficiency**を製品競争力の中心に置いていると解釈できる。これは公式アーキテクチャ説明と効率改善値からの分析的推論である。[^M1] 9

一方で、その結果として、モデル研究の観点からは透明性が低い。GPT-6.1 Sol、Sonnet 5.5、Gemini 3.8 Flashならcontext、output limit、modality、API料金、model IDをかなり厳密に固定できるが、Manusの場合「どのモデルが、どのstepで、何token使ったか」という制御面はCascadeに抽象化される。Flexはまさにこの問題に対して、モデル／providerをユーザーが選べるようにした機能と読める。[^M2][^O1][^A2][^G2] 46

GPT・Claude・Geminiはモデルレイヤーでもエージェント化している

ただし「Manusだけがagent、GPT/Claude/Geminiはchat」という区分も既に古い。GPT-6.1 SolはResponses APIでweb search、file search、computer use、MCP等を利用でき、Claude Sonnet 5.5はagentic codingやcomputer-use系評価を主要指標とし、Gemini 3.8 Flashもfunction calling、code execution、computer use、search groundingを備える。[^O1][^A1][^G2] 47

差は**どこまでを完成品としてベンダーが引き受けるか**である。OpenAI／Anthropic／GoogleのモデルAPIは、開発者がエージェント実装を組むためのbuilding blockとして強い。一方Manus 2.0は、Cloud Computer、

Studio、Video Editor、Game Dev、Remote Control、Automationsまでを最初から統合し、「モデル呼び出し」ではなく「完成した仕事」をAPI／UIの抽象単位にしようとしている。Manus API v2で最上位リソースが `Tasks` であり、タスクが非同期実行されることもこの設計思想を反映している。[^M1][^M5] 48

マルチモーダルという言葉の意味が異なる

Gemini 3.8 Flashの「マルチモーダル」は、モデル自身がtext、image、video、audio、PDFを入力tokenとして処理するという**model-native multimodality**である。GPT-6.1 SolとClaude Sonnet 5.5は現在の当該APIモデルについてtext＋image入力が中心である。[^G2][^O1][^A2] 49

Manusがいう文書・スライド・表計算・動画・ゲーム対応は、それより広い**application-level multimodality / artifact orchestration**に近い。たとえばVideo Editorでは動画をclips、images、text、motion、audioといった編集可能な要素として扱い、Game Devではcoding/image/video系能力を組み合わせて実行可能成果物を生成する。したがって「Manusは動画まで扱えるからGeminiよりnative multimodalである」といった比較は技術的には不正確である。[^M1] 9

安全性では「モデルalignment」と「エージェント権限」を分けるべき

GPT、Claude、Geminiにはモデルが**何を回答すべきか／拒否すべきか**というalignmentがあり、それぞれOpenAIのPreparedness／deployment safety、AnthropicのConstitutional AI、GoogleのFrontier Safety Frameworkといった公開枠組みがある。Claude Sonnet 5.5では約1,850シナリオのbehavioral audit、高リスクcyberでのfallback、biology safeguard、reasoning extraction対策まで公表されている。Gemini 3.8 FlashもCBRN・cyber misuseとprompt injectionを明示的評価対象とする。[^O3][^A1][^A3][^G3] 50

Manusでこれに加えて重要になるのは、**「モデルが何を言うか」ではなく「エージェントに何を実行させる権限を与えるか」**である。Remote Controlがローカルのbrowser/apps/filesを動かし、Automationsがメールやカレンダー等のevent triggerで動作し、Cueではagentがemail、phone、wallet、computerを持ち、設定されたbudget内で支払い等を扱えるという設計は、従来のchatbotより権限リスクが大きい。[^M1] 9

このため企業導入では、SOC 2やISO認証だけでなく、「送信・削除・購入・公開・本番deployといった不可逆操作にhuman approvalを置くか」「OAuth scopeをどこまで絞るか」「Cloud Computerとlocal Remote Controlを分離するか」が重要になる。これはManusの公開機能構成から導かれる運用上の推奨であり、認証取得そのものがagent actionの安全性を保証するわけではない。ManusはSOC 2 Type II、ISO 27001、ISO 27701等をTrust Centerで公表している。[^M4] 51

ベンチマークで「最強」を決めるのは現時点では不適切

Claude Sonnet 5.5、Gemini 3.8 Flash、GPT-6系はいずれもcoding、agent、knowledge work等の多数のbenchmarkを公表しているが、effort設定、tool使用、agent harness、token budget、評価モデルのversionが異なる。Anthropic自身もSonnet 5.5発表で「benchmarkは能力の一側面」と注意し、さらにGPT-6 Solの画像理解に関する修正によって外部スコアが最新状態を反映していない可能性まで注記している。[^A1] 27

Googleの3.8 Flash Model CardもDeepSWE、GDPVal、Terminal-Bench、HLE、OSWorld等を掲載する一方、評価方法やagentic/static条件を個別に記載している。したがって、同名ベンチであってもスコアだけ抜き出して「A > B」とするのは危険である。[^G3] 17

特にManusはさらに比較困難である。Manus 2.0の公式性能数字は**旧Manusに対するシステム内比較**であり、GPT/Claude/Geminiと同一benchmarkを同一harnessで評価した数字ではない。現時点で「Manus 2.0はGPT-6.1よりX%高性能」といった主張を一次情報から正当化することはできない。[^M1] 25

実用的影響と導入判断

技術選択では「どのAIが一番賢いか」より、**AIスタックのどの層を自社で所有したいか**を先に決める方が合理的である。

導入要件	推奨方向	理由
社員が「この調査をして、資料とWebサイトまで完成させて」と委譲したい	**Manus 2.0優位**	Studio、Cloud Computer、artifact creation、browser/tool操作を一つのtaskとして束ねるため、agent infrastructureを自社開発する必要が小さい。[^M1] [9]
自社SaaS／業務システムの内部にAIを深く埋め込みたい	**GPT／Claude／Gemini API優位**	context、model ID、token、tool definition、pricing等をモデル単位で制御しやすく、自社agent runtimeを自由に設計できる。[^O1][^A2][^G2] [5]
モデルを固定せず、上位agent layerだけ利用したい	**Manus Flexを評価**	provider/model選択を利用者側に残しながらManusのplanning/tools/execution環境を利用する構造。推論料金とManus infra料金は分かれる。[^M2] [7]
長大文書を一回のモデルcallで厳密に処理したい	**GPT／Claude／Geminiを優先検証**	1M級context windowが明文化されている。Manus Cascadeの総context windowは未公開。[^O1][^A2][^G2] [5]
動画・音声・長い映像をモデルに直接理解させたい	**Gemini 3.8 Flashが明確**	text/image/video/audio/PDFを単一モデルの入力modalitiesとして公式仕様化している。[^G2] [23]
coding＋knowledge workを高品質・比較的低コストで回したい	**Claude Sonnet 5.5を重点評価**	AnthropicはSonnet 5.5をspeed/intelligenceのバランスモデルと位置付け、$2/$10、1M context、30%以上の速度改善を公表。[^A1][^A2] [11]
Responses API／MCP／computer useを中心に自社agentを構築したい	**GPT-6.1 Solを重点評価**	Responses APIと多様なhosted toolsを統合し、1.05M context、$2/$10のモデルとして提供。[^O1] [15]
大量処理でtoken単価を最優先	**Gemini 3.8 Flashが価格面で有力**	2026年末までの導入価格は$0.75/$3.75で、Batch/Flexはさらに低価格。ただし2027年1月から価格改定予定。[^G5] [52]
最高難度のGemini frontier modelを今すぐ一般本番利用したい	**現時点ではArgonを前提にしない**	Gemini 4 Argonは最新だが、2026年10月7日時点では段階的・限定展開。[^G1] [45]

企業のPoCで特に見るべきKPIは、単なる「1M tokenあたり価格」ではない。Manus型agentでは、**task completion rate、human intervention回数、tool-call失敗率、成果物修正時間、end-to-end task cost、不可逆操作の誤実行率**を測る必要がある。一方、基盤モデルAPIを直接採用する場合には、これに加えてprompt/cache token、reasoning effort、first-token latency、output throughput、長文context精度を測るべきである。各社がeffort、caching、Batch/Flexなど異なる料金・推論モードを持つ以上、token単価だけでは総コストを正確に比較できない。[^O1][^A5][^G5] [53]

Manus 2.0を評価する際の最大の判断ポイントは、「最高性能のモデルを買う」のではなく、「AIに業務遂行レイヤーまでアウトソースする」ことを許容できるかである。 モデル選択、context engineering、tool orchestration、sandbox/cloud computer、成果物生成を自社で構築する組織にとってはGPT／Claude／Geminiの直接利用の方が透明性と制御性を得やすい。反対に、そのエージェント基盤自体を構築・保守する

コストを避け、社員が自然言語から完成成果物まで短距離で到達することを優先する組織にはManus 2.0の価値が大きい。Flexの登場によって、両者は排他的な選択肢ではなくなった。[^M2] [7]

また、日本企業で機密データを扱う場合、**無料版・有料版でデータ利用条件が変わるサービスを同一視しないこと**が重要である。たとえばGemini APIでは、公式Additional Terms上、無料サービスでは入力／生成レスポンスを製品改善に利用する場合があるのに対し、Paid Servicesではprompt／responseを製品改善に使用しないと明記されている。Google Search grounding等には追加の保持条件もある。[^G6] [39] Manus、OpenAI、Anthropicについても、本番導入時には一般消費者向け規約ではなく、契約するTeam／Enterprise／APIプランの当該時点のデータ処理条項を確認すべきである。OpenAIでもindividual TermsとAPI／business向け契約体系は分離されている。[^O5] [38]

最終的な推奨を一文でまとめると、**「AIに回答してほしい」の延長ならGPT／Claude／Gemini、「AIに仕事を完了してほしい」ことを第一目的にするならManus 2.0を優先的に評価し、モデル品質を自分で選びたい場合はManus Flexを含むハイブリッド構成をPoCする**、というのが2026年10月時点で最も技術的に整合した選び方である。[^M1][^M2] [1]

出典注記

以下は本レポートで優先的に使用した一次資料である。日本語公式ページが存在する場合は可能な限り日本語版を優先した。

[^M1]: **Manus,「Introducing Manus 2.0」日本語公式発表, 2026-09-28.**
https://manus.im/ja/blog/introducing-manus-2-0
Manus 2.0、Cascade、Studio、Cloud Computer、Remote Control、Automations、Video Editor、Game Dev、Cue、旧システム比の効率改善値の主要一次資料。 [9]

[^M2]: **Manus,「Introducing Manus Flex」日本語公式発表, 2026-09-29.**
https://manus.im/ja/blog/introducing-manus-flex
モデル選択、harness、agent infrastructureの分離、およびBYOK／provider選択を確認する重要資料。 [7]

[^M3]: **Manus Open API v2 — Introduction.**
https://open.manus.im/docs/v2/introduction
REST API v2、Tasks、Projects、Files、Webhooks、Skills、Agents等のAPI構造。 [54]

[^M4]: **Manus Trust Center.**
https://trust.manus.im/
SOC 2 Type I/II、ISO 27001、ISO 27701等のsecurity/compliance情報。 [51]

[^M5]: **Manus Open API v2 — Create Task.**
https://open.manus.im/docs/v2/task.create
非同期Task API、約5,000推定tokenのprompt制限、file/task references、agent profiles等。 [55]

[^M6]: **Manus Open API v2 — Authentication.**
https://open.manus.im/docs/v2/authentication
API keyおよびOAuth2認証。 [56]

[^M7]: **Manus Open API v2 — Connectors.**
https://open.manus.im/docs/v2/connectors
Slack、Notion、Google系、GitHub、Zapier等のconnector一覧とOAuth連携。 [34]

[^M8]: **Manus Terms.**
https://manus.im/terms
Manusサービス利用規約。 37

[^O1]: **OpenAI Developer Documentation — Models / GPT-6.1 Sol.**
https://developers.openai.com/api/docs/models
https://developers.openai.com/api/docs/models/gpt-6.1-sol
GPT-6 Astra／GPT-6.1 Solの位置付け、1.05M context、128K output、modalities、tools、API料金等。 10

[^O2]: **OpenAI,「Introducing GPT-6.1 Sol」日本語公式発表, 2026-09-29.**
https://openai.com/ja-JP/index/introducing-gpt-6-1-sol/
GPT-6.1 Solの公開日、位置付け、agentic coding、computer use、価格等。 26

[^O3]: **OpenAI Deployment Safety Hub — GPT-6.1 Sol / GPT-6 Astra Safety Overview.**
https://deploymentsafety.openai.com/gpt-6-1-sol
https://openai.com/index/safety-overview-gpt-6-astra/
能力リスク分類、alignment・safeguardに関する一次資料。 31

[^O4]: **OpenAI, "How ChatGPT and our foundation models are developed."**
https://openai.com/policies/how-chatgpt-and-our-foundation-models-are-developed/
公開Web、partners、人間／ユーザー／研究者由来データ等、OpenAI基盤モデル一般のデータソース説明。 21

[^O5]: **OpenAI Service Terms / Terms of Use.**
https://openai.com/policies/service-terms/
https://openai.com/policies/terms-of-use/
API、Output、IP、reverse engineering等に関する公式規約。Service Termsは2026-09-29更新。 38

[^A1]: **Anthropic, "Introducing Claude Sonnet 5.5," 2026-09-28.**
https://www.anthropic.com/claude-sonnet-5-5
速度、料金、主要benchmark、alignment audit、cyber/biology safeguards、用途・制限。 27

[^A2]: **Claude Platform Docs — Claude Sonnet 5.5.**
https://platform.claude.com/docs/en/models/sonnet-5-5/overview
1M context、128K output、training cutoff、model IDs、availability、platforms等の公式仕様。 16

[^A3]: **Anthropic, "Claude's new constitution," 2026-01-22.**
https://www.anthropic.com/news/claude-new-constitution
Constitutionをtraining各段階およびsynthetic data生成に利用する現在のalignment方針。 57

[^A4]: **Anthropic Research, "Teaching Claude why," 2026-05-08.**
https://www.anthropic.com/research/teaching-claude-why
constitutionally aligned documents、chat data、agentic environments、従来RLHFとの関係を扱う研究。 58

[^A5]: **Claude Platform Docs — Pricing.**
https://platform.claude.com/docs/en/about-claude/pricing
Sonnet 5.5 $2/$10、prompt caching、Batch、regional pricing等。 59

[^G1]: **Google Japan Blog,「Gemini 4 Argon：フロンティア知能の新時代」, 2026-10-01.**
https://blog.google/intl/ja-jp/company-news/technology/gemini4argon/
Gemini 4 Argonの日本語一次発表、限定展開、段階的安全公開、導入価格$2/$10。 45

[^G2]: **Google AI for Developers — Gemini 3.8 Flash.**
https://ai.google.dev/gemini-api/docs/models/gemini-3.8-flash
1,048,576 input token、65,536 output、text/image/video/audio/PDF、computer use、function calling等の公式API仕様。 23

[^G3]: **Google DeepMind — Gemini 3.8 Flash Model Card, 2026-09-02.**
https://deepmind.google/models/model-cards/gemini-3-8-flash/
model dependency、training-data参照関係、benchmark、known limitations、red teaming、Frontier Safety評価。 17

[^G4]: **Google, “Introducing Gemini 3.8 Flash and 3.8 Flash Cyber,” 2026-09-02.**
https://blog.google/innovation-and-ai/models-and-research/gemini-models/3-8-flash-and-3-8-flash-cyber/
agentic workflow、long-horizon coding、training、safety、価格と展開先。 20

[^G5]: **Google AI for Developers — Gemini Developer API Pricing.**
https://ai.google.dev/gemini-api/docs/pricing
3.8 FlashのStandard／Batch／Flex／Priority料金と2027年1月の価格変更予定。 52

[^G6]: **Google AI for Developers — Gemini API Additional Terms of Service, effective 2026-03-23.**
https://ai.google.dev/gemini-api/terms
generated content、model extraction/reverse engineering、無料／有料サービスのデータ利用差、agentic servicesの責任等。 39

---

1 3 9 44 Manus 2.0 のご紹介
https://manus.im/ja/blog/introducing-manus-2-0

2 10 40 Models | OpenAI API
https://developers.openai.com/api/docs/models

4 7 46 Manus Flex のご紹介：Manus ワークスペースで独自の API キーを使用
https://manus.im/ja/blog/introducing-manus-flex

5 15 47 53 GPT-6.1 Sol Model | OpenAI API
https://developers.openai.com/api/docs/models/gpt-6.1-sol

6 30 51 manus.ai Trust Center
https://trust.manus.im/

8 45 Gemini 4 Argon
https://blog.google/intl/ja-jp/company-news/technology/gemini4argon/

11 27 32 35 Introducing Claude Sonnet 5.5 \ Anthropic
https://www.anthropic.com/claude-sonnet-5-5

12 26 GPT-6.1 Sol のご紹介
https://openai.com/ja-JP/index/introducing-gpt-6-1-sol/

13 Newsroom \ Anthropic
https://www.anthropic.com/news

14 20 29 Introducing Gemini 3.8 Flash and 3.8 Flash Cyber
https://blog.google/innovation-and-ai/models-and-research/gemini-models/3-8-flash-and-3-8-flash-cyber/

16 22 Claude Sonnet 5.5 - Claude Platform Docs
https://platform.claude.com/docs/en/models/sonnet-5-5/overview

17 28 42 Gemini 3.8 Flash - Model Card — Google DeepMind
https://deepmind.google/models/model-cards/gemini-3-8-flash/

18 GPT-6 Astra System Card - OpenAI Deployment Safety Hub
https://deploymentsafety.openai.com/gpt-6-astra/sec%3Aappendix-sol-luna

19 57 Claude's new constitution
https://www.anthropic.com/news/claude-new-constitution

21 How ChatGPT and our foundation models are developed | OpenAI
https://openai.com/policies/how-chatgpt-and-our-foundation-models-are-developed/

23 49 Gemini 3.8 Flash | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/models/gemini-3.8-flash

24 43 55 task.create - Manus API
https://open.manus.im/docs/v2/task.create

25 41 Manus 2.0 の新機能 | Manus Help Center
https://help.manus.im/ja/articles/17190150-manus-2-0-%E3%81%AE%E6%96%B0%E6%A9%9F%E8%83%BD

31 50 Addendum to GPT-6 Astra System Card: GPT-6.1 Sol
https://deploymentsafety.openai.com/gpt-6-1-sol

33 48 54 Introduction - Manus API
https://open.manus.im/docs/v2/introduction

34 Connectors - Manus API
https://open.manus.im/docs/v2/connectors

36 52 Gemini Developer API pricing | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/pricing.md

37 Terms
https://manus.im/terms

38 Service terms | OpenAI
https://openai.com/policies/service-terms/

39 Gemini API Additional Terms of Service | Google AI for Developers
https://ai.google.dev/gemini-api/terms.md

56 Authentication - Manus API
https://open.manus.im/docs/v2/authentication

58 Teaching Claude why
https://www.anthropic.com/research/teaching-claude-why?gsid=6fc91da3-d143-4370-a306-086708cb85f3

59 Pricing - Claude Platform Docs
https://platform.claude.com/docs/en/about-claude/pricing