

Claude Fable 5.1およびMythos 5.1の技術構造と産業的評価

Gemini 3.7 Flash

2026年9月1日、Anthropicは主力フロンティアモデルの最新系統として「Claude Fable 5.1」および「Claude Mythos 5.1」を正式発表した¹。今回のリリースは、単なるパラメータ規模の拡大や学術ベンチマークのスコア競争にとどまらず、数時間から数日規模で自律稼働するAIエージェントの実行基盤、高度な科学研究プロセスの自動化、規制産業におけるデータ主権の確立、そしてプロンプトキャッシュの抜本的なコスト削減を統合した体系的な刷新である⁴。

アーキテクチャ設計と二重展開モデル

Claude Fable 5.1とClaude Mythos 5.1は、同一の基礎アーキテクチャおよびパラメータ構成を共有しながら、適用される安全分類器とアクセスポリシーによって明確に分離された二重展開構造を採用している²。フロンティアモデルの自律的推論能力がサイバーセキュリティや生物医学といったデュアルユース（軍民両用）領域に達する中、Anthropicは一般利用と高リスク専門研究の分離を制度化した⁴。

一般提供モデルであるClaude Fable 5.1は、商用アプリケーションや一般的なソフトウェア開発を対象とし、厳格な製品セーフガード（Production Safeguards）が常時稼働する⁴。このモデルは、Claude API、Amazon Bedrock、Google Cloud、Microsoft Foundry、ならびにClaude.aiやClaude Codeなどのファーストパーティ環境を通じて即日展開された⁶。

これに対してClaude Mythos 5.1は、米国政府との連携のもとで設立された「Life Sciences Verification Program」や「Project Glasswing」などの信頼性アクセスプログラムを通じて、審査済みの研究機関、防衛組織、セキュリティ専門家にのみ限定提供される²。Mythos 5.1では、攻撃的兵器化の防止措置を維持しつつ、防御的サイバーセキュリティや創薬スクリーニングにおける高度な探索を可能にするため、安全ガードレールの緩和が適用されている⁴。

基本仕様として、両モデルは100万トークン（1M tokens）のコンテキストウィンドウと最大128,000トークン（128K tokens）の出力容量を備え、ナレッジカットオフは2026年6月に設定されている⁵。また、推論の深度を動的に制御する適応型思考（Adaptive Thinking）が常時組み込まれている⁴。

項目	Claude Fable 5.1	Claude Mythos 5.1	Claude Opus 5	Claude Fable 5
提供形態	一般提供（GA）	審査制（Project Glasswing等）	一般提供（GA）	旧世代（提供中）
コンテキスト長	1,000,000 tokens	1,000,000 tokens	1,000,000 tokens	1,000,000 tokens
最大出力長	128,000 tokens	128,000 tokens	128,000 tokens	128,000 tokens

入力価格 (1M tokens)	\$10.00	\$10.00 (適格組織)	\$5.00	\$10.00
出力価格 (1M tokens)	\$50.00	\$50.00 (適格組織)	\$25.00	\$50.00
キャッシュ読み取り (1M)	\$0.25	\$0.25	\$1.00相当	\$1.00
セーフガード特性	汎用防御・過誤抑止設計	防衛・研究特化 (緩和型)	汎用標準	汎用標準 (過剰遮断あり)
ナレッジカットオフ	2026年6月	2026年6月	2026年5月	2026年初頭

コスト構造の変革とEnterprise Frontier Safeguards

プロンプトキャッシュ価格の引き下げと運用経済性

Fable 5.1の基本トークン価格は、入力100万トークンあたり10ドル、出力100万トークンあたり50ドルとFable 5の水準を維持しているが、プロンプトキャッシュ読み取り (Cache Read) の料金が従来の1.00ドルから0.25ドルへと75%引き下げられた⁴。

長時間の自律エージェント処理では、過去の対話履歴、ファイル群、実行ログをコンテキストに保持したまま数十から数百ターンの推論を反復するため、累積コストの大部分をキャッシュ読み取りが占める⁴。この価格改定により、一般的なワークロードで約25%、高度に自律化したエージェント処理では最大約45%の総運用コスト削減が実現される²。バッチ処理APIを利用した場合には入力5ドル、出力25ドルが適用され、大規模な非同期バッチ処理における経済性も向上している¹³。

30日データ保持の撤回とEnterprise Frontier Safeguards

Fable 5の導入時に義務付けられた「悪用検知を目的とする30日間のプロンプトデータ保持ポリシー」は、機密データや規制対象データを扱う金融、医療、防衛関連企業から強い反発を招いていた¹⁰。Anthropicはこの方針を抜本的に改め、ゼロデータ保持 (Zero Data Retention: ZDR) と安全監視を両立する新機構「Enterprise Frontier Safeguards (EFS)」を発表した⁴。

EFS環境下において、対話データや実行履歴はAnthropicのサーバーではなく、顧客が自社で管理するクラウドストレージ (AWS S3、Google Cloud Storage、Azure Blob Storage) 内に顧客管理暗号化キー (CMEK) を用いて排他的に保存される⁴。不正利用の検知はセッション横断型の自動分類器により実行され、Anthropicの従業員による人的閲覧を完全に排除した上で、フラグが立ったアラートは顧客専任のセキュリティ運用チームにのみ通知される⁴。EFSは追加ライセンス料なしで提供され、正式展開までの移行期間中は適格顧客に対してZDR条件でのFable 5.1利用が無償で認可される⁴。

安全分類器の精度改善とフォールバックAPI

従来の安全フィルターが引き起こしていた過剰拒絶 (False Positive) の問題に対し、Fable 5.1では分類アルゴリズムの境界条件が再調整された²。サイバーセキュリティ領域における誤検知は従来比で約60%削減され、生物・医学領域における不要な拒絶も約85%減少した²。これにより、一般提供のFable 5.1であってもソースコード内の脆弱性特定やセキュリティ監査が正式に許可される一方、悪用可能なエクスプロイト生成やサイバー兵器の開発は引き続き遮断される²。さらに、要求がセーフガードに抵触した場合にエラー停止させるのではなく、サイバー関連タスクはClaude Opus 4.8へ、生物医学関連タスクはClaude Opus 5へとシームレスに処理を委任する「Fallback API」が導入された⁴。

ベンチマーク評価と実世界における科学・工学検証

Fable 5.1は、長文脈を要する自律コーディング、複合的な推論、および科学的発見を評価する複数のベンチマークにおいて顕著な向上を示し、OpenAIのGPT-5.6 Solや自社のClaude Opus 5を広範に凌駕している²。

評価指標 / ベンチマーク	Fable 5.1	Mythos 5.1	Opus 5	Fable 5	GPT-5.6 Sol
科学研究自律遂行 (Terminal-Bench-Science 0.1)	52.6%	—	29.0%	24.7%	22.4%
エージェント型コーディング (Terminal-Bench 4.0)	55.8%	60.9%	52.3%	42.0%	37.3%
統合開発環境評価 (CursorBench 3.2.0)	73.4%	—	70.0%	70.5%	67.2%
ナレッジワーク総合 (GDPval-AA v2)	1853	—	1824	1723	1711

OS操作自動化 (OSWorld 2.0 Strict)	41.7%	—	39.6%	36.1%	—
OS操作自動化 (OSWorld 2.0 Partial)	77.9%	—	75.4%	72.9%	—
学術推論 (Humanity's Last Exam - ツールなし)	60.9%	—	56.6%	57.8%	—
学術推論 (Humanity's Last Exam - ツールあり)	65.0%	—	63.6%	63.8%	—
業務プロセス自動化 (Automation Bench)	31.4%	—	26.9%	17.1%	19.6%

科学研究における実験室検証と計算工学の実績

Fable 5.1の能力は、スタンフォード大学やLaude Instituteが主導する自律科学研究ベンチマーク「Terminal-Bench-Science 0.1」において、旧世代Fable 5のスコア(24.7%)から倍増となる52.6%を記録した点に象徴される⁴。

分子標的薬設計の実証実験において、Mythos 5.1はオープンソースのタンパク質設計・構造予測ツールを自律的に操作し、生成された設計物を外部のウェットラボ2機関で実験検証に供した⁴。その結果、Adaptyv Bioのコンペティションにおける最高水準と比較して10倍高い結合親和性を達成し、12の標的タンパク質に対する有効バインダーのヒット率は、一般的な業界平均である10～15%を大幅に上回る約50%に達した⁴。

地球外惑星の計算地質モデリングにおいては、NASAのマゼラン探査機が取得したレーダー画像データを基に、Fable 5.1が自律的にニューラルネットワークを構築・学習させ、金星表面の3分の1に及ぶ高解像度標高マップを生成した⁴。従来の解像度(10～20km)から2～3kmへと微細化され、高度精度が最大25%改善した成果物は、将来の惑星探査ミッション(NASA VERITAS、ESA EnVision)を支援する目的でクリエイティブ・コモンズライセンスのもと一般公開された⁴。

さらに計算生物学の高速化検証では、Mythos 5.1が独自のカスタムGPUカーネルを実装し、中間計算のキャッシングパイプラインを構築することで、出力を完全に維持したまま主要な7つの深層学習モデルの推論・学習速度を最大2.5倍に引き上げた⁴。

産業界パートナーによる長時間自律運用の評価

早期アクセス企業による検証では、単一プロンプトへの応答を超えた、数時間から数十時間におよぶ自律セッションでの堅牢性が確認されている⁴。

ヘッジファンドのMillenniumにおける検証では、人間のシニアエンジニアや他社の主要モデルが4～5年間にわたり解明できなかった内部システムの極めて低頻度なクラッシュ（100万回に1回発生する不具合）の原因究明を実施した⁴。Fable 5.1は外部ベンダー製ライブラリを逆アセンブルし、コアダンブと照合することによってライブラリ内部の潜伏バグを特定した⁴。

経費管理プラットフォームのRampでは、38時間にわたる無人運用において、モデルが機械学習モデルの過去の実験データからラベルの不整合を発見し、自動修正を加えた上で6つの並列実験を立ち上げ、結果を総括した⁴。また、Jane Street Capitalは複数ファイルにまたがる大規模リファクタリングにおいて長大なコンテキスト下でも論理の破綻が起きない点を評価し、Cognitionは自社AIエンジニア「Devin」のバックエンドをOpus 5からFable 5.1へと切り替えている⁴。

API仕様・アーキテクチャの変更点と制限事項

Fable 5.1の実装に伴い、自律エージェントの構築を支援する複数のAPI機能が拡充された一方で、移行時の障害となり得る破壊的変更や固有の振る舞いが導入されている¹⁰。

新機能と運用制御の拡張

推論プロセスの柔軟な制御を実現するため、会話の途中でプロンプトキャッシュを無効化することなく思考深度を変更できる動的Effort制御機能が追加された¹⁵。これにより、定型的な処理ではEffortを下げてレイテンシとコストを抑え、難度の高い検証段階でのみEffortを引き上げる運用が可能となる¹⁵。

また、単一ターン限定のシステムメッセージ機能（clear_at: "next_user_message"）が提供され、次のユーザー入力時に自動破棄されることで、履歴の手動編集に伴うキャッシュ不一致や思考コンテキストの破損を防止できるようになった¹⁵。さらに、非公開の思考ブロックから抽出された簡潔な進捗テキストをツール実行の合間にストリーミングする機能（thinking.display: "updates"）や、生成コンテンツの真正性を保証する統計的ウォーターマークおよびC2PAクレデンシャルの付与が標準化された²。

破壊的変更と競合蒸留防止策

Fable 5.1への移行に際しては、3つの仕様変更に対する修正が必要となる¹⁵。

第1に、tool_choiceパラメータにおける特定ツールの強制指定（{"type": "any"}や特定ツール名）が非推奨となり、指定時は400 invalid_request_errorが返される¹⁵。ツール呼び出しを強制する場合は、{"type": "auto"}を指定した上で厳格な構造化出力スキーマを適用しなければならない¹⁵。

第2に、思考ブロックの参照が一方向性に制限され、Fable 5.1は過去モデルの思考履歴を読み込めるものの、旧モデル側はFable 5.1の思考ブロックを読み取ることができず破棄される¹⁰。

第3に、対話履歴やシステムプロンプトの途中変更が検出された場合、以降の思考ブロックが無効化されてエラーとなる¹⁰。これは競合企業による合成データ蒸留（Distillation）攻撃を防ぐための保護機構であり、新規作成されたAPIアカウントに対して厳格に適用されている¹³。

ネイティブな振る舞いの変化

Fable 5からの移行に伴い、いくつかの実行特性の変化が確認されている¹⁰。長大なエージェントループ内において、複数ツールの並列バッチ実行を行わず1ターンにつき1つのツールを逐次実行する傾向が見られ、APIの往復回数と所要時間が増加する場合がある¹⁰。また、コード修正時に差分編集ではなくファイル全体の再生成を選択しやすく、出力トークンの消費が肥大化しやすい傾向や、低Effort設定時に外部検索ツールを呼び出さず内蔵知識から即答しようとする挙動が報告されている¹⁰。

開発者コミュニティおよび産業界における評価

技術コミュニティ(Hacker News、Reddit、X等)および実務開発者の間では、推論能力と経済性への高い評価と、出力文体や安全機能の過剰作動に対する不満が並存している¹⁵。

肯定的な評価と強み

開発者コミュニティでは、長時間にわたるセッションでの方針維持能力と、テストコードの自己記述による自律的デバッグ能力が極めて高く評価されている⁴。特にプロンプトキャッシュの75%値下げは、これまでコスト面から躊躇されていた大規模コードベースの常時走査やマルチエージェントオーケストレーションを商業的に成立させる要因として歓迎されている⁴。また、科学研究分野における自律探索の成功事例は、LLMが単なるコーディング支援ツールから科学的発見の協働者へと進化しつつある証拠として受け止められている⁴。

批判的評価・不満点・課題

一方で、出力されるテキストのトーンに関しては批判的な意見が多い¹⁵。過度に追従的な言い回し(Sycophancy)や、マーケティング文書のように装飾過多な文体、不要な専門用語の多用が指摘されており、プロンプトで簡潔さを指示しても長文対話の後半で指示が形骸化する問題が報告されている¹⁵。

また、安全分類器の誤検知は改善されたものの、Win32 APIを直接操作する低レベルなRustコードやOSカーネルに近いコードを検査させようとすると、依然として安全フィルターに抵触してOpusモデルへ強制フォールバックされる事例が存在する¹⁵。この挙動により、正規のシステムプログラマーやサイバー防衛担当者から開発効率が阻害されるとの不満が上がっている¹⁵。加えて、短編的な問題解決や一部の競技的プログラミングベンチマーク(DeepSWE 1.1等)において、旧世代からの明確なスコア向上が見られない点も指摘されている²³。

結論

Claude Fable 5.1およびClaude Mythos 5.1の投入は、先端AIモデルの競争軸が「単一ターンの知識応答」から「環境と相互作用しながら長時間タスクを完遂する自律性、ガバナンス、運用経済性の総合力」へと移行したことを示している⁷。

一般向けのFableと審査制のMythosという二重ブランド構造、顧客インフラに主権を残すEnterprise Frontier Safeguards、そしてプロンプトキャッシュを基軸とした大幅な実効コスト削減は、今後のエンタープライズAI導入における標準アーキテクチャを提示している²。導入組織においては、キャッシュを前提としたシステム設計を構築し、逐次ツール呼び出しやファイル再書き込みを抑制するプロンプト制御を適切に適用することが、モデルの性能と費用対効果を最大化する鍵となる⁶。

引用文献

1. Claude Fable 5.1 と Mythos 5.1 の概要 | npaka - note,
<https://note.com/npaka/n/n14013b319725>
2. Anthropic Launches Claude Fable 5.1 With Lower Costs and Fewer,
<https://www.macrumors.com/2026/09/01/anthropic-claude-fable-5-1/>
3. Claude Fable 5.1 Launch: Pricing, Benchmarks, and What's New,
<https://alphacorp.ai/blog/claude-fable-5-1-launch-pricing-benchmarks-and-whats-new>
4. Introducing Claude Fable 5.1 and Claude Mythos 5.1 \ Anthropic,
<https://www.anthropic.com/claude-fable-and-mythos-5-1>
5. Claude Fable 5.1 - Claude Platform Docs,
<https://platform.claude.com/docs/en/models/fable-5-1/overview>
6. Claude Fable - Anthropic, <https://www.anthropic.com/claude/fable>
7. Anthropic's Claude Fable 5.1 and Mythos 5.1 arrive with a 75% cost,
<https://venturebeat.com/technology/anthropics-claude-fable-5-1-and-mythos-5-1-arrive-with-a-75-cost-reduction-for-fable-cache-reads>
8. Introducing Claude Fable 5.1 and Claude Mythos 5.1 - daily.dev,
<https://daily.dev/posts/pex5p6uxm>
9. Claude Fable 5.1 & Claude Mythos 5.1 System Card - Anthropic,
<https://www-cdn.anthropic.com/0339e6a7c5c7b87f5c07798616dc32c215d14235/Claude%20Fable%205.1%20&%20Claude%20Mythos%205.1%20System%20Card.pdf>
10. What's new in Claude Fable 5.1,
<https://platform.claude.com/docs/en/models/fable-5-1/whats-new-fable-5-1>
11. Claude Fable 5.1 - API Pricing & Providers | OpenRouter,
<https://openrouter.ai/anthropic/claude-fable-5.1>
12. Claude Fable 5.1, Anthropic's new frontier model is now ... - AWS,
<https://aws.amazon.com/about-aws/whats-new/2026/09/claude-fable-5-1-aws/>
13. Anthropic Drops Claude Fable 5.1 and Mythos 5.1, Pricing Down,
<https://glitchwire.com/news/anthropic-drops-claude-fable-51-and-mythos-51-pricing-down-and-distillation-prot/>
14. Claude Mythos - Anthropic, <https://www.anthropic.com/claude/mythos>
15. Claude Fable 5.1 and Claude Mythos 5.1 - Hacker News,
<https://news.ycombinator.com/item?id=49525378>
16. Anthropic Revises Enterprise Data Retention Policy After Push,
<https://www.pymnts.com/news/artificial-intelligence/2026/anthropic-revises-enterprise-data-retention-policy-after-customer-pushback/>
17. Anthropic Launches Enterprise Frontier Safeguards With Customer,
<https://aiweekly.co/node/11423>
18. Anthropic Walks Back Data Retention Policy After Enterprise,
<https://www.binance.com/en/square/post/362065882978422>
19. Anthropic launches Enterprise Frontier Safeguards for Claude,
<https://www.investing.com/news/stock-market-news/anthropic-launches-enterprise-frontier-safeguards-for-claude-93CH-4884473>

20. Anthropic Releases Claude Fable 5.1 and Claude Mythos 5.1: 52.6,
<https://www.marktechpost.com/2026/09/01/anthropic-releases-claude-fable-5-1-and-claude-mythos-5-1-52-6-on-terminal-bench-science-and-75-cheaper-cache-reads/>
21. 「AIが科学を知っている」から「AIが研究をする」へ。Claude,
<https://note.com/allay0224/n/n63993835bfd5>
22. 「Fable 5.1」のYahoo!リアルタイム検索 - X(旧Twitter)を,
https://search.yahoo.co.jp/realtime/search?rkf=1&p=Fable%205.1&btid=MjA5NDg2NDA2NDk5NjYxNDQ3Ng%3D%3D&ifr=tp_bz
23. Introducing Claude Fable 5.1 and Claude Mythos 5.1 \ Anthropic,
https://www.reddit.com/r/ClaudeAI/comments/1w4juj2/introducing_claude_fable_5_1_and_claude_mythos_51/
24. Claude Fable 5.1が出た。Pro月3000円のわたしには - note,
https://note.com/moyuchi_aistu/n/nbef15156550a