ChatGPT

Anthropicの新型AIモデル「Claude Mythos（クロード・ミトス）」深掘り分析レポート

エグゼクティブサマリー（実務要点）：2026年4月7日（米国時間）、Anthropicは新フロンティアモデル「Claude Mythos Preview」を一般提供せず、限定パートナーと重要インフラ/OSS維持組織に招待制で提供する「Project Glasswing」を同時発表した。公式説明によれば、このモデルは汎用モデルでありながらサイバー攻撃（脆弱性発見・エクスプロイト作成）能力が「全主要OS・主要ブラウザでゼロデイを発見・悪用可能」な水準に達し、短期的には悪用リスクの方が防御利得を上回り得るため“防御目的の囲い込み”が採られている。実務的な含意は、(1)「AIによる脆弱性発見が“専門家の希少資源”から“計算資源でスケールする能力”へ転換した」こと、(2) 90日以内に公的成果報告が予告されており、重要OSS・主要ベンダのパッチサイクル/開示フローが変動し得ること、(3) 高性能モデルの一般解放より前に、利用者側のCVD（協調的脆弱性開示）運用・アクセス管理・監査証跡・開発ライフサイクル（SDL）へAIを組み込む準備が意思決定テーマになること、である。 [1]

発表概要と公式資料一覧

発表日時の整理

Anthropic公式ページおよび開発者向けリリースノートは、いずれも「April 7, 2026」としてProject GlasswingとClaude Mythos Previewの“招待制リサーチプレビュー”提供開始を明記している。 [2]
メディア側では、TechCrunch [3] が記事公開時刻を「April 7, 2026 11:00 AM PDT」と記載しており、東京（JST）換算では概ね4月8日未明に相当する（時差換算は本レポートの補足）。 [4]

「Claude Mitos（ミトス）」表記について

公式英語表記は **“Mythos”**（Mythos Preview）で統一されている。一方、日本語圏では音写として「ミソス／ミトス」等の揺れが起き得るため、本レポートではユーザー指定の「クロード・ミトス」を「Claude Mythos Preview」と同一対象として扱う。 [5]
なお、Project Glasswing本文の注記では “Mythos” の語源を古代ギリシャ語（物語/言説に関する語）と説明している。 [6]

公式資料（一次資料）マップ（URL付き）

下表は、発表当日（4/7）に中核となる一次資料、および本分析で参照した関連の公式・準公式ドキュメントを一覧化したもの。

区分	資料名（発行主体）	公開日	何が書かれているか（要点）	URL
公式（中核）	Project Glasswing: Securing critical software for the AI era（Anthropic）	2026-04-07	限定提供の理由、パートナー、ベンチマーク、価格、90日レポート予告、政府との対話など	https://www.anthropic.com/glasswing

区分	資料名（発行主体）	公開日	何が書かれているか（要点）	URL
公式（技術詳細）	Assessing Claude Mythos Preview's cybersecurity capabilities（Anthropic Frontier Red Team blog）	2026-04-07	ゼロデイ発見・悪用評価の方法と事例、開示制約（99%未パッチ等）	https://red.anthropic.com/2026/mythos-preview/
公式（プラットフォーム）	Claude Platform Release Notes（Claude Platform）	2026-04-07	招待制リサーチプレビューとしての位置づけ（API/Bedrock等の言及を含む）	https://platform.claude.com/docs/en/release-notes/overview
公式（開発者向け）	Models overview（Claude API docs）	随時更新	Mythos PreviewはProject Glasswingの招待制、自己申込不可	https://platform.claude.com/docs/en/about-claude/models/overview
公式（開発者向け）	Context windows（Claude API docs）	随時更新	Mythos Previewのコンテキスト上限（1M tokens）等	https://platform.claude.com/docs/en/build-with-claude/context-windows
公式（開発者向け）	Pricing（Claude API docs）	随時更新	1Mコンテキストの価格扱い、全体の価格体系	https://platform.claude.com/docs/en/about-claude/pricing
公式（法務/データ）	Privacy Policy（Anthropic）	2026-01-12施行	学習データ源、入力/出力の学習利用、法的根拠など	https://www.anthropic.com/legal/privacy
公式（法務/データ）	Commercial Terms of Service（Anthropic）	（最新版）	商用顧客データの学習不使用、入力/出力の権利、競合利用禁止等	https://www.anthropic.com/legal/commercial-terms
公式（行動規範）	Usage Policy（Anthropic）	（最新版）	無許可の脆弱性探索・悪用、マルウェア作成等の禁止	https://www.anthropic.com/legal/aup
公式（ガバナンス）	Responsible Scaling Policy v3.0（Anthropic）	2026-02-24	ASL等の枠組み更新、ロードマップ/リスクレポート導入	https://www.anthropic.com/news/responsible-scaling-policy-v3

区分	資料名（発行主体）	公開日	何が書かれているか（要点）	URL
パートナー公式	Claude Mythos Preview on Vertex AI（Google Cloud）	2026-04-07	Vertex AIでのPrivate Preview提供（Glasswingの一環）	https://cloud.google.com/blog/products/ai-machine-learning/claude-mythos-preview-on-vertex-ai
パートナー公式	Amazon Bedrock now offers Claude Mythos Preview（AWS）	2026-04-07	Bedrockでのgated research preview提供	https://aws.amazon.com/about-aws/whats-new/2026/04/amazon-bedrock-claude-mythos/
パートナー公式	MSRC blog（Microsoft）	2026-04-07	Microsoft側の評価姿勢、Foundry経由の研究プレビュー言及	https://www.microsoft.com/en-us/msrc/blog/2026/04/strengthening-secure-software-global-scale-how-msrc-is-evolving-with-ai

上表の中核一次資料は「Project Glasswing」と「Frontier Red Team技術記事」であり、性能・運用方針・公開制約の多くはこの2本から復元できる。[7]
なお、Anthropic公式の「Model System Cards」ページには“Mythos Preview (April 2026)”のシステムカードが掲載されているが、**本レポート作成環境では当該PDF本文の一次確認が完了していないため、当該カード由来の詳細は『未確認』として扱う。**[8]

技術仕様と提供形態

技術仕様（公開情報ベース）

公開一次資料から確実に言える仕様は、**(a) コンテキスト長、(b) 一部ツールの提供可否と提供面（Surface）、(c) ベンチマーク群（Anthropic自己報告）**が中心である。[9]

以下、ユーザー要望の項目に沿って「公開済／不明」を区分する。

項目	公開状況	内容（要約）	根拠
アーキテクチャ	不明（未公開）	公式一次資料（Project Glasswing / Red Team記事）上、アーキテクチャ詳細は明示されない。※システムカードに記載の可能性はあるが『未確認』。	[10]
パラメータ数	不明（未公開）	同上。	[10]
トレーニングデータ	一部公開（一般方針）	Anthropic全体のプライバシーポリシーとして、公開Web情報、第三者との契約データ、ユーザー/クラウドワーカー提供データ（オプトアウト除く）、フィードバック、安全審査フラグ、社内生成データ等を学習に用いると明記。Mythos固有の配合は不明（未公開）。	[11]

項目	公開状況	内容（要約）	根拠
トレーニング手法（RLHF等）	不明（未公開）	Mythos固有のRLHF/RLAIF等の実装詳細は一次資料に明示なし（システムカードに詳細の可能性＝『未確認』）。	12
コンテキストウィンドウ	公開	Mythos Previewは1Mトークンのコンテキストを持つ。	13
推論/コーディング性能（主要指標）	公開（自己報告）	SWE-bench Verified/Pro、Terminal-Bench、GPQA Diamond、HLE等でOpus 4.6を大きく上回る数値が提示されている。	6
API/製品形態	公開	招待制のResearch Previewとして、Claude API / Amazon Bedrock / Google Cloud Vertex AI / Microsoft Foundry で提供される（一般公開なし）。	14

性能（ベンチマーク）と“サイバー能力”の特徴

Project Glasswingは、Mythos Previewが「主要OS・主要ブラウザすべてでゼロデイを見つけ、指示されれば悪用できる」水準に到達したと述べる。さらに、OpenBSDの27年物バグやFFmpegの16年物バグ等の例を挙げ、既存の自動テストが見逃してきた欠陥を発見したとしている。 7
Frontier Red Team記事は、開示可能なバグが全体の1%に留まる（99%は未パッチで開示不能）と説明し、CVDを前提に情報公開を抑制している。 15

Mythos PreviewとOpus 4.6の差分（公式提示の数値）

以下はProject Glasswing本文の比較値（Anthropic自己報告）を要点化したもの。導入検討では「絶対値」よりも、**(1) 攻撃側・防御側双方の作業フローが“自律エージェント化”でどれだけ短縮されるか、(2) 誤用時の被害半径がどこまで拡大するか**に着目すべきである。 7

指標（Anthropic提示）	Mythos Preview	Claude Opus 4.6	差分の意味（実務観点）
CyberGym（脆弱性再現）	83.1%	66.6%	既知系タスクの再現度上昇＝“武器化の速度”に直結し得る
SWE-bench Verified	93.9%	80.8%	実コード修正の成功率向上＝防御側のパッチ作成も加速
SWE-bench Pro	77.8%	53.4%	難度が高い実務課題での伸びが大きい
Terminal-Bench 2.0	82.0%	65.4%	端末操作型エージェントとしての自律遂行力上昇
GPQA Diamond	94.6%	91.3%	高難度推論でも上積み（差分は小さめ）
Humanity’s Last Exam（with tools）	64.7%	53.1%	ツール併用での総合推論・探索能力上昇
OSWorld-Verified	79.6%	72.7%	“実環境操作”寄りの能力上昇（オペレーション自動化と表裏）

出典：Project Glasswing本文の比較表（Anthropic）。[6]

提供面（Surface）とツール可用性の重要点

Mythos Previewは「どこから呼び出せるか」だけでなく「どのツールが使えるか」が安全性と運用価値を左右する。Claude API docs上、Mythos Previewについては次が明記されている。[16]

- **Code execution（サンドボックス実行）**：Claude APIとMicrosoft Foundryでサポート、BedrockとVertex AIでは非対応。[17]
- **Web search**：Claude API、Microsoft Foundry、Vertex AIでサポート。Bedrockでは非対応。[18]

これらは「攻撃的な探索・検証の自動化」を強める一方、企業内の監査・承認フローが弱いと事故リスクも上げるため、提供面ごとに**許可されたツールセット**を前提に利用設計が必要となる。[19]

```
graph TD
A[Claude Mythos Preview] --> B{提供形態}
B -->|一般公開なし| C[招待制 Research Preview]
C --> D[Project Glasswing パートナー]
C --> E[重要インフラ/OSS維持組織 40+]
D --> F[防御的スキャン/修正/検証]
E --> F
F --> G[90日以内に学びを公表]
G --> H[将来のより広い展開に向けセーフガード開発]
```

（上図は公式方針の要素を業務フローとして再構成したもの。出典はProject Glasswing本文。）[6]

セーフティ・ガバナンス

“一般公開しない”というリリース判断そのものが安全策

Anthropicは、Mythos Previewを「汎用モデル」だとしつつも、観測されたサイバー能力の跳躍により、まずは防御目的の限定プログラム（Project Glasswing）で運用し、**一般提供はしない**方針を明記している。[20]
この判断は、従来の「利用規約・コンテンツフィルタ・レート制限」中心の抑止に加え、**“アクセス権そのものを絞る”**という強いガバナンス選択であり、特にエクスプロイト生成のようなデュアルユース領域では合理的（＝外部化リスクを抑える）と評価できる。[21]

安全性設計の枠組み：RSP（Responsible Scaling Policy）との接続

Anthropicは2026年2月にResponsible Scaling Policy（RSP）v3.0を公開し、ASL（AI Safety Level）に基づく安全策強化、ロードマップ公開、リスク報告の仕組みを説明している。[22]
Mythos Preview関連では、少なくともProject Glasswing本文が「最も危険な出力を検出・遮断するサイバーセーフガードを開発し、将来より広い展開を可能にする」という方向性を明記しており、RSPの“能力上昇に伴う必要な緩和策強化”という思想と整合する。[23]

セーフガードの中身：現時点で確実に言えること／言えないこと

現時点で一次資料から確実に言えるのは、次の「運用面」の安全策である。

- **防御目的の利用に限定**（Project Glasswing）[5]

- **CVD（協調的脆弱性開示）を前提に詳細開示を抑制**（未パッチが多数のため） [15]
- **今後、危険出力を検出・遮断するセーフガードを段階的に導入**（“次期Opus”でまず展開し改善する旨） [7]

一方、ユーザーが求める「RLHF/その他の安全手法」「監査・透明性施策」の詳細については、**システムカードに記載されている可能性が高いが『未確認』**であるため、本レポートでは断定しない。 [8]

制限（Usage Policy）との噛み合わせ：攻撃能力を“合法な防御”へ寄せる条件

AnthropicのUsage Policyは、**無許可の脆弱性発見・悪用、無許可アクセス、マルウェア作成、DoSツール作成**等を明確に禁止している。 [24]
このため、Mythos Previewの正当な利用は、(1) 企業内/顧客との契約に基づいたセキュリティ評価、(2) OSSメンテナとの合意、(3) バグバウンティのルール遵守、(4) CVDのプロセス運用、など「許諾と証跡」を満たす設計が前提になる。 [25]

プライバシー・データ利用

収集元・学習利用（Anthropic全体方針）

Anthropicのプライバシーポリシーは、モデル学習に用いるデータ源として「公開インターネット情報」「第三者との商用契約で得たデータセット」「ユーザーやクラウドワーカー提供データ（オプトアウトを除く）」「ユーザーのフィードバック」「安全・セキュリティ・ポリシー審査でフラグされた素材」「社内生成データ」を列挙している。 [11]
これはMythos固有の訓練仕様ではないが、「どの種類のデータが学習に入り得るか」を判断する最低限のベースラインとして重要である（特に社内情報・個人情報を入力する運用のリスク評価に直結）。 [26]

ユーザーデータの扱い：商用と消費者向けの差

商用（Commercial Terms）では、**AnthropicはCustomer Content（入力＋出力）でモデルを学習しない**と明記している（商用顧客視点では大きい）。 [27]
一方、消費者向けでは、プライバシーセンターの説明として、ユーザー設定によるオプトイン、セーフティレビューにフラグされた会話、Trusted Tester等の明示参加などの条件で、会話/コーディングセッションがモデル改善に用いられる旨が整理されている。 [28]

データ保持（Retention）と削除

消費者向けのデータ保持方針では、ユーザーが会話を削除すると履歴から即時消え、バックエンドからは概ね30日以内に削除されると説明される。加えて、モデル改善を許可した場合は、学習パイプライン内で最長5年保持し得る旨が記載されている。 [29]
企業実装では「どの提供形態（Claude API / Foundry / Bedrock / Vertex等）を使うか」「ZDR（Zero Data Retention）契約の可否」「ログ/監査の保持要件」を組み合わせ、**機密度別に入力ポリシーを分ける**のが現実的となる。 [30]

比較分析（Claude系列・OpenAI・Google・Meta）

“性能比較”の注意点

Mythos Previewは一般提供されていないため、第三者が同一条件で広範に測定した“横並びベンチ”は限定的になりやすい。現時点で公開一次資料にあるのは、**Anthropic自己報告のベンチマーク（Mythos vs Opus**

4.6）が中心であり、OpenAI/Google/Metaとの直接スコア比較は本レポートでは「不明（未公開）」扱いとする。 [5]

主要モデルの比較（用途・安全性・提供形態・価格）

下表は、意思決定者がまず比べるべき「提供形態」「価格」「文書化（透明性）」「用途の制約」を中心に整理した。

観点	Claude Mythos Preview（Anthropic）	Claude Opus 4.6（Anthropic）	OpenAI [31] GPT-5.4（API）	Google [32] Gemini（Developer API）	Meta [33] Llama 4（代表例）
提供形態	招待制Research Preview。防御的サイバー用途を優先（一般公開なし）。	広く提供（Claude.ai / API等）。	APIで提供（価格表に掲載）。	Gemini APIで提供（モデル別に価格・ティア）。	オープンウェイトモデルとして公開（系列は2025年に発表）。
主な用途（公式文書から）	防御的脆弱性発見・修正、ペンテスト/検証の準備、重要ソフトウェア保護（Project Glasswing）。	エージェント・コーディング等の一般用途。	“professional work”向け最上位（OpenAIの位置づけ）。	高ボリューム/推論/マルチモーダル等、モデル階層で使い分け。	オープンな基盤モデル活用（研究・企業内運用など）。
コンテキスト	1M tokens。	1M tokens。	価格表は「270K未満の標準処理」言及（最大長は別ページ参照が必要＝本レポートでは不明）。	価格表はプロンプト長（<=200k/>200k）で価格が変わる前提。	不明（本レポートでは一次資料未確認）。
価格（Input/Output）	$25 / $125（1M tokensあたり、クレジット期間後の参加者価格）。	$5 / $25（1M tokens）。	$2.50 / $15.00（1M tokens）。	例：Gemini 3.1 Proは$2.00/$12.00（<=200k）等。	不明（モデル公開＝商用API価格とは別軸）。

観点	Claude Mythos Preview（Anthropic）	Claude Opus 4.6（Anthropic）	OpenAI [31] GPT-5.4（API）	Google [32] Gemini（Developer API）	Meta [33] Llama 4（代表例）
安全性の基本方針	高能力ゆえ一般公開停止＋限定プログラム。危険出力遮断セーフガード開発を予告。	RSP/Usage Policy等に基づく一般提供。	安全性はOpenAI側の別文書体系（本レポートは価格表中心で確認）。	価格表上に「Used to improve our products」等のデータ利用表示。	オープン性ゆえ、利用者側のガバナンス負担が相対的に大きい。
透明性ドキュメント	Project Glasswing＋Red Team技術記事＋システムカード（PDFは『未確認』）。	モデル/システムカードが体系的に公開されている。	価格表は明確。安全/評価の詳細は別途（本表では未調査）。	価格・ティア・データ利用表示が詳細。	公式ブログ/論文等（本表では最小限）。

出典（表の根拠）：Mythos/OpusはAnthropic公式（Project Glasswing、Claude API docs）。OpenAIはAPI価格表。GoogleはGemini API価格表。Metaは公式ブログおよび報道。 [34]

商業化・規制/倫理的懸念・今後の展望

商業化・ライセンス・価格戦略

Project Glasswingでは、(1) 研究プレビュー期間中に最大$100Mの利用クレジット、(2) OSSセキュリティ団体へ$4Mの直接寄付、(3) 研究プレビュー後は$25/$125で継続提供、(4) 提供面はClaude API/Bedrock/Vertex AI/Foundry、と明記される。これは「モデル一般提供によるスケール」ではなく、「限定顧客に高価格で提供しつつ、成果（脆弱性修正・ベストプラクティス）を公表して市場全体の安全水準を引き上げる」戦略と読める。 [35]
また、Microsoft [36] は、研究プレビューモデルへのアクセスを評価・リスク低減に使う姿勢を明記しており、Foundry上での提供が「ガバナンスを含む統合基盤」として説明されている。 [37]

規制・倫理的懸念（デュアルユース）

Mythos Previewの核心的リスクは、脆弱性探索・悪用がスケールすることで「攻撃のコストと専門性の閾値が低下」し、攻撃サイドが優位を取る移行期が生じ得る点である（Anthropic自身が“産業として緊急対応が必要”と表現）。 [38]
Anthropicは、Mythos Previewの攻撃的・防御的サイバー能力について米国政府関係者と継続協議している旨を明記しているが、現時点で「規制当局の公式声明（具体的措置）」が一次資料として確認できたわけではないため、各国の具体反応は**不明（未公開）**とする。 [35]

研究・産業への影響と展望（短期・中期・長期）

短期（〜90日）は、Anthropicが公表を予告している「学び・修正された脆弱性・推奨プラクティス」の報告が最大のイベントとなる。重要OSSやクラウド/OSベンダの修正・開示が増える可能性があり、企業側は脆弱性管理とパッチ適用の“受け側能力”が問われる。 [7]

中期（数か月〜1年）は、「危険出力を検知・遮断するセーフガード」を次期Opus系モデルで先に展開して改善するとされており、Mythos級能力の広域展開に向けた“安全な提供形態”が具体化する段階になる。[39] 長期（1年〜）は、脆弱性発見・修正がAIで高速化し、ソフトウェア工学の標準プロセス（SDL、CVD、監査、証跡、依存関係管理）が再設計される可能性が高い。MicrosoftはMSRC/SDL/SFIの文脈で、AIによる発見量増大と、それを捌くための自動化＋人間の関与を同時に強める方針を述べている。[40]

未解決の疑問点と追加調査が必要な項目

本レポート時点で、意思決定に直結するにもかかわらず一次資料で確定できない（あるいは本環境で一次確認できない）論点を「未確認」または「不明（未公開）」として列挙する。

論点	現状	なぜ重要か	追加調査の当たり先（優先）
モデルアーキテクチャ/パラメータ数	不明（未公開）	推論コスト、脆弱性探索のスループット、運用コスト見積に直結	システムカード（PDF）※要一次確認、Anthropic公式補足
Mythos固有の学習手法（RLHF/RLAIF等）	不明（未公開）	“危険能力”がどの学習段階で増幅されたかの特定は安全策設計に重要	システムカード（PDF）※未確認、RSP関連文書
学習データの範囲（Mythos固有）・カットオフ	不明（未公開）	ゼロデイ評価の「記憶」疑義、監査の前提	システムカード、Anthropicの透明性ハブ更新
外部監査/第三者評価の範囲	不明（未公開）	限定提供モデルほど外部性が大きく、第三者評価が信頼形成に重要	Project Glasswingの90日報告、パートナー各社ブログ
具体的な出力遮断セーフガード（検知方式、誤検知、適用範囲）	不明（未公開）	正当なセキュリティ業務を阻害せず悪用のみ阻止できるかが鍵	Claude API docs（更新）、Cyber Verification Programの詳細
日本企業/日本の規制当局の反応	不明（未公開）	国内基幹インフラ・サプライチェーン防御への影響	日本語一次情報（総務省/経産省/IPA等の公開資料）※今後

このうち最優先は「システムカード一次確認」と「90日報告のフォロー」である。公開一次資料だけでも戦略意図は読める一方、仕様・安全策の“実装詳細”が未確定のため、導入や連携を検討する場合は契約条項・監査要件・データ保持設定を含めた精査が不可欠となる。[41]

[^1]: Anthropic「Project Glasswing」 https://www.anthropic.com/glasswing
[^2]: Anthropic Frontier Red Team「Assessing Claude Mythos Preview’s cybersecurity capabilities」 https://red.anthropic.com/2026/mythos-preview/
[^3]: Claude Platform Release Notes（2026-04-07項目あり） https://platform.claude.com/docs/en/release-notes/overview
[^4]: Claude API docs「Models overview」 https://platform.claude.com/docs/en/about-claude/models/overview
[^5]: Claude API docs「Context windows」 https://platform.claude.com/docs/en/build-with-claude/context-windows
[^6]: Anthropic Privacy Policy（2026-01-12施行） https://www.anthropic.com/legal/privacy

[^7]: Anthropic Commercial Terms of Service https://www.anthropic.com/legal/commercial-terms
[^8]: Anthropic Usage Policy https://www.anthropic.com/legal/aup
[^9]: Anthropic「Responsible Scaling Policy v3.0」 https://www.anthropic.com/news/responsible-scaling-policy-v3
[^10]: OpenAI API Pricing https://openai.com/api/pricing/
[^11]: Google Gemini Developer API pricing https://ai.google.dev/gemini-api/docs/pricing
[^12]: Microsoft MSRC Blog（Project Glasswing/Foundry言及） https://www.microsoft.com/en-us/msrc/blog/2026/04/strengthening-secure-software-global-scale-how-msrc-is-evolving-with-ai
[^13]: Google Cloud Blog（Vertex AIでのPrivate Preview） https://cloud.google.com/blog/products/ai-machine-learning/claude-mythos-preview-on-vertex-ai
[^14]: AWS「What's New」（Bedrockでのgated research preview） https://aws.amazon.com/about-aws/whats-new/2026/04/amazon-bedrock-claude-mythos/

---

1 2 5 6 7 9 10 14 20 21 23 32 34 35 36 38 39 41 Project Glasswing: Securing critical software for the AI era \ Anthropic
https://www.anthropic.com/glasswing

3 29 How long do you store my data? | Anthropic Privacy Center
https://privacy.anthropic.com/en/articles/10023548-how-long-do-you-store-personal-data

4 Anthropic debuts preview of powerful new AI model Mythos in new cybersecurity initiative | TechCrunch
https://techcrunch.com/2026/04/07/anthropic-mythos-ai-model-preview-security/

8 12 Model system cards \ Anthropic
https://www.anthropic.com/system-cards

11 26 33 Privacy Policy \ Anthropic
https://www.anthropic.com/privacy

13 Context windows - Claude API Docs
https://platform.claude.com/docs/en/build-with-claude/context-windows?utm_source=chatgpt.com

15 31 Claude Mythos Preview \ red.anthropic.com
https://red.anthropic.com/2026/mythos-preview/

16 17 19 Code execution tool - Claude API Docs
https://platform.claude.com/docs/en/agents-and-tools/tool-use/code-execution-tool?utm_source=chatgpt.com

18 Web search tool - Claude API Docs
https://platform.claude.com/docs/en/agents-and-tools/tool-use/web-search-tool?utm_source=chatgpt.com

22 Responsible Scaling Policy Version 3.0 \ Anthropic
https://www.anthropic.com/news/responsible-scaling-policy-v3

24 Usage Policy \ Anthropic
https://www.anthropic.com/aup

25 Responsible Disclosure Policy \ Anthropic
https://www.anthropic.com/responsible-disclosure-policy

27 Commercial Terms of Service \ Anthropic
https://www.anthropic.com/legal/commercial-terms

28 Is my data used for model training? | Anthropic Privacy Center
https://privacy.anthropic.com/en/articles/10023580-i-want-to-opt-out-of-my-prompts-and-results-being-used-for-training-models

30 I have a zero data retention agreement with Anthropic. What products does it apply to? | Anthropic Privacy Center
https://privacy.anthropic.com/en/articles/8956058-i-have-a-zero-retention-agreement-with-anthropic-what-products-does-it-apply-to

37 40 Strengthening secure software at global scale: How MSRC is evolving with AI
https://www.microsoft.com/en-us/msrc/blog/2026/04/strengthening-secure-software-global-scale-how-msrc-is-evolving-with-ai