ChatGPT

Gemini 3.6 Flash の評判と評価

エグゼクティブサマリー

Google が 2026年7月21日に発表した **Gemini 3.6 Flash** は、現時点の公開情報を総合すると、**「3.5 Flash の完全な上位互換」**というより、**3.5 Flash よりも運用効率と実用速度を磨いた“実務向けの改良版”**として評価するのが最も正確です。公式は、3.6 Flash を「コーディング、知識労働、マルチモーダルで性能向上し、出力トークンを 17% 削減、出力単価も下げた主力ワークホース」と位置づけています。実際、公式価格は **入力 $1.50 / 100万 tokens、出力 $7.50 / 100万 tokens** で、3.5 Flash の **出力 $9.00** から下がりました。 [1]

ただし、**第三者評価はより慎重**です。Artificial Analysis は、3.6 Flash を **「3.5 Flash と同等の Intelligence Index 50」** としつつ、**タスク時間は 2.7 分→1.3 分、コスト/タスクは $0.59→$0.50、出力速度は 175.7→275.5 tokens/s** と評価しています。つまり、**“より賢い”というより、“同程度の知能をより速く・安く出す”改善**と読むのが妥当です。 [2]

公式ベンチマークでは、3.6 Flash は 3.5 Flash より **DeepSWE、MLE-Bench、OSWorld-Verified、GDPval-AA v2、長文コンテキスト** で改善していますが、同じ公式表の中でも **GPT-5.6、Luna、Claude Sonnet 5 などに負ける項目が少なくありません**。したがって、**絶対性能の頂点**ではなく、**高いコスト効率と実運用の俊敏性**が真の売りです。 [3]

開発者コミュニティの初動反応は **賛否両論**です。好意的な声は、**低冗長・高スループット・エージェント向け反復実行の安さ**を評価しています。一方、否定的な声では、**「知能は 3.5 と大差ない」「より安い同等モデルがある」「3.5 Pro の遅延が期待を削いだ」**といった不満が目立ちます。また、リリース当日に **Antigravity で 3.6 Flash 選択時のフリーズ報告** が相次ぎ、これはモデル品質というより **周辺ツールの互換性・配布品質の問題**として懸念材料になりました。 [4]

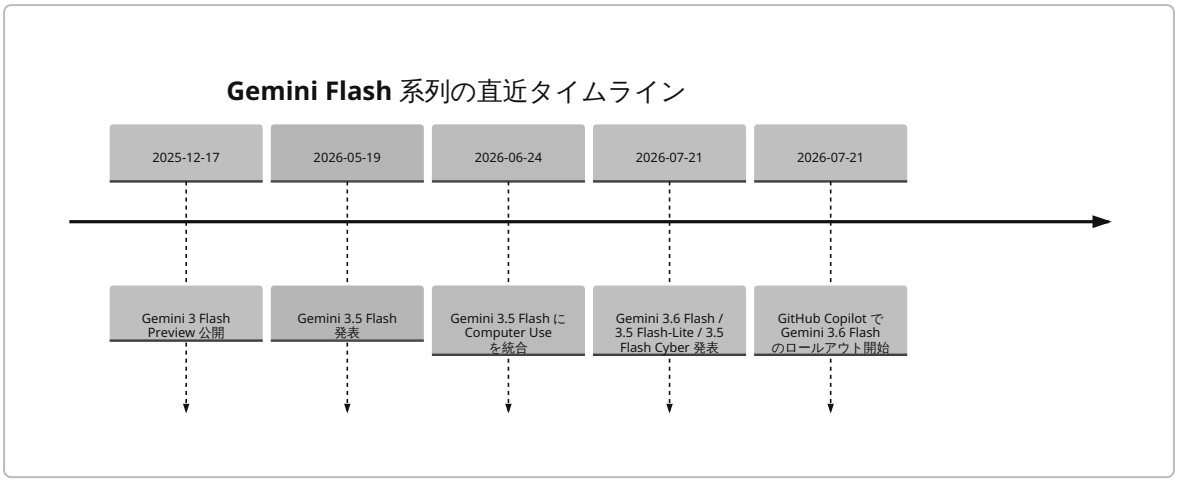
導入判断としては、**チャットボットの単純会話性能だけで選ぶモデルではない**一方、**エージェント、コード生成、文書処理、検索統合、マルチモーダル解析**では非常に有力です。逆に、**最高レベルの推論精度を最優先する用途**や、**オンデバイス／エッジ推論、量子化、自前蒸留**を前提にする用途には向きません。モデル重みは非公開で、パラメータ数や推論時メモリも未公開です。 [5]

調査前提と結論

今回確認した一次情報では、Gemini 3.6 Flash は **実在する正式発表済みモデル**であり、Google 公式ブログ、Google AI for Developers、Google DeepMind モデルページ、Google Cloud ドキュメントで整合的に確認できます。発表日は **2026年7月21日** で、同時に **Gemini 3.5 Flash-Lite** と **Gemini 3.5 Flash Cyber** も公表されました。Reuters、TechCrunch、The Verge もこの発表を同日報道しています。 [6]

本件で最も重要なのは、**公式の語りと第三者の語りが微妙に違う**ことです。Google は 3.6 Flash を 3.5 Flash の改善版として押し出していますが、Google AI for Developers のモデル一覧では、なお **Gemini 3.5 Flash が「agentic/coding tasks における most intelligent model」**、一方の 3.6 Flash は **「speed と intelligence を両立する最新モデル」** と書き分けられています。これは、3.6 Flash が「最上位知能」よりも「実務上の速度・効率の最適点」に寄せられていることを示唆します。Artificial Analysis の **“知能は 3.5 と同等で、時間とコストは大きく改善”** という評価は、このポジショニングとかなり整合的です。 [7]

その意味で、市場での評判はおおむね次のように整理できます。**Google ファンボーイ的な絶賛より、開発者の実務観点からの“使える改善”評価が中心**であり、他社最上位モデルを押しのける“世代交代級”の反応ではありません。初日の印象としては、**「3.5 Flash を本番で回していたチームほど歓迎」「3.5 Pro や他社フロンティアモデルを待っていた層ほど冷淡」**です。 [8]



上のタイムラインは、Google 公式ブログ、Gemini API 公開資料、GitHub Changelog を基に作成しています。 [9]

主要出典URL: `https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-6-flash-3-5-flash-lite-3-5-flash-cyber/`、`https://ai.google.dev/gemini-api/docs/models/gemini-3.6-flash?hl=ja`、`https://deepmind.google/models/gemini/flash/`

公式仕様と前世代差分

Gemini 3.6 Flash の公式仕様はかなり明確です。モデル ID は **`gemini-3.6-flash`**、**Stable / GA** として提供され、**入力はテキスト・画像・動画・音声・PDF、出力はテキストのみ、入力上限 1,048,576 tokens、出力上限 65,536 tokens** です。機能面では **思考（thinking）、関数呼び出し、検索グラウンディング、Google Maps グラウンディング、コード実行、ファイル検索、URL Context、Computer Use** をサポートします。ただし **画像生成・Live API は非対応**です。Google は加えて、**Interactions API を最新モデル利用の推奨 API** としています。 [10]

3.5 Flash との差分として最重要なのは、**価格・冗長性・実行ループの削減**です。価格は **標準課金で入力 $1.50 / 出力 $7.50** とされ、3.5 Flash の **入力 $1.50 / 出力 $9.00** から **出力単価が 16.7% 下がりました**。Batch / Flex / Priority でも同様に、3.6 Flash の方が 3.5 Flash より出力単価が低く設定されています。Google 公式ブログと最新モデルガイドは、3.6 Flash が **reasoning steps、conversational turns、tool calls を減らし、execution loop spiraling も抑える**と説明しています。 [11]

一方で、「**最も賢い Flash**」という座は、公式のモデル一覧上では依然として **3.5 Flash** に残っています。ここは重要な差分です。Google のモデル一覧では、3.6 Flash は **“Balances speed with intelligence”**、3.5 Flash は **“Most intelligent model for sustained frontier performance on agentic and coding tasks”** と記載されており、3.6 Flash の改善が **高知能化そのもの**ではなく、**知能とスピードのトレードオフ最適化**であることを示しています。 [12]

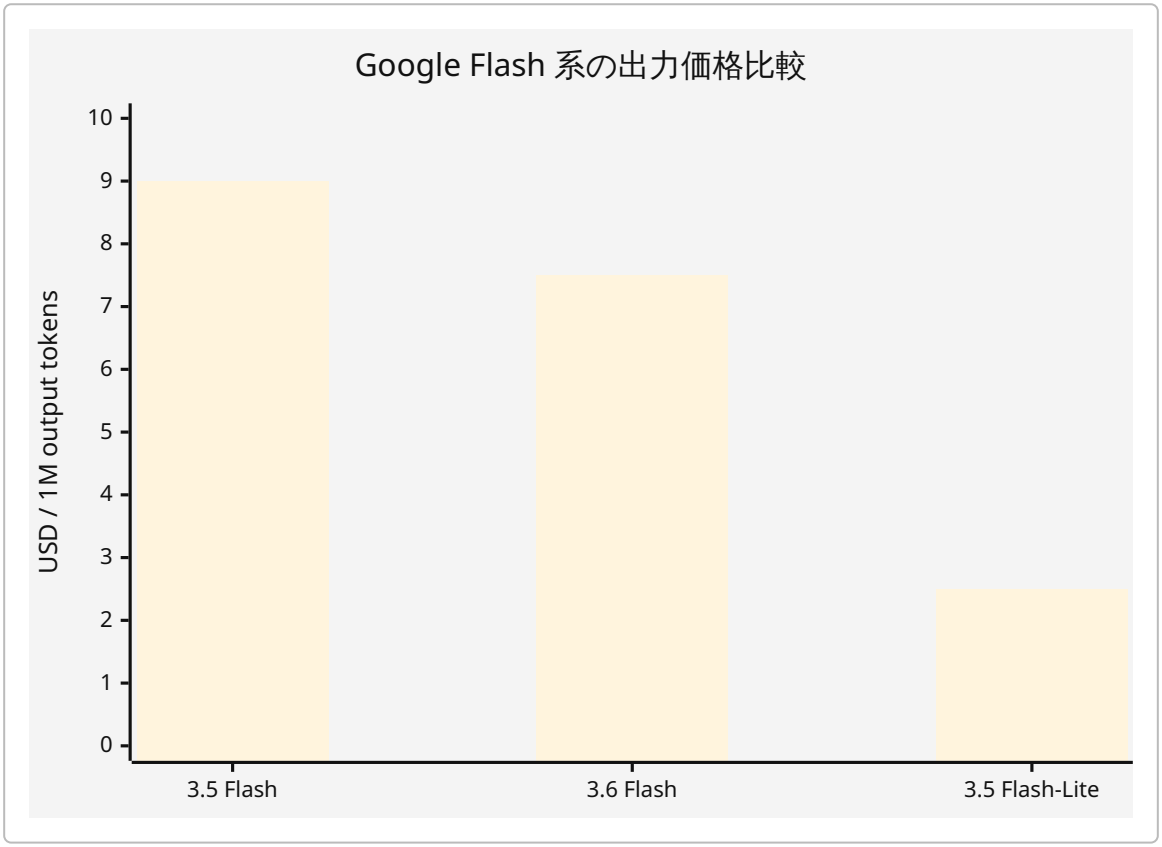
加えて、**API 移行面の変化**もあります。Gemini 3.6 Flash と 3.5 Flash-Lite からは、`temperature`、`top_p`、`top_k` のような旧来のサンプリングパラメータが非推奨となり、`thinking_level` ベースへの移行

が促されています。Qiita の国内開発者解説でも、この変更は **静かな破壊的変更の入り口**として注意喚起されており、運用コードの棚卸しが推奨されています。 [13]

下表は、導入判断で重要な **Google 系主要比較**です。

モデル	位置づけ	入力/出力上限	標準価格	Thinking 既定	速度・遅延	推奨ユースケース	評価メモ
Gemini 3.6 Flash	速度と知能のバランス型、3.5 Flash 改良版	1M / 64k	$1.50 / $7.50	`medium`	275.5 tok/s、TTFT 12.52s	コーディング、複数ツールを使うエージェント、マルチモーダル解析	3.5 比で効率改善、公式は冗長性低減を強調。知能は第三者評価では同等級。 [14]
Gemini 3.5 Flash	sustained frontier performance を狙う Flash	1M / 64k 前後	$1.50 / $9.00	`medium`	175.7 tok/s、TTFT 20.22s	より知能寄りの Flash 利用、既存 3.5 ワークロード	公式一覧では依然「最も知的」と表現。3.6 より出力単価が高い。 [15]
Gemini 3.5 Flash-Lite	最速・最安の 3.5 系	1M / 64k	$0.30 / $2.50	`minimal`	439.3 tok/s、TTFT 11.24s	高頻度サブエージェント、文書抽出、大量ルーティング	知能は 3.6 に劣るが、スループットと価格は圧倒的。 [16]
Gemini 3.1 Pro Preview	複雑タスク向け上位系	1M / 64k	AAブレンド $1.74/1M	reasoning	113 tok/s、TTFT 32.17s	高度推論、複雑設計	3.5/3.6 Flash より遅く高コスト。 [17]

モデル	位置づけ	入力/出力上限	標準価格	Thinking 既定	速度・遅延	推奨ユースケース	評価メモ
Claude Sonnet 5	他社比較の代表例	1M	AAブレンド $1.54/1M	non-reasoning/high effort	68.2 tok/s、TTFT 1.31s	対話応答の立ち上がり重視	3.6 Flash は処理速度とコストで優位だが、TTFT は不利。 [18]



この価格差は公式価格表に基づいています。3.6 Flash の出力単価は 3.5 Flash 比で 16.7% 下がっています。 [19]

主要出典URL: `https://ai.google.dev/gemini-api/docs/models/gemini-3.6-flash?hl=ja`、`https://ai.google.dev/gemini-api/docs/pricing`、`https://ai.google.dev/gemini-api/docs/latest-model`

第三者ベンチマークと性能評価

第三者評価で最も参照価値が高いのは、現時点では **Artificial Analysis** です。ここでは、Gemini 3.6 Flash は **Intelligence Index 50** で **Gemini 3.5 Flash と同点**、しかし **平均 Time per Task は 2.7 分→1.3 分、Cost per Task は $0.59→$0.50、出力速度は pre-launch test で 304 tok/s、モデルページで 275.5 tok/s** とされ、**“知能横ばい、運用効率大幅改善”**が中核結論になっています。Google の「性能向上」メッセージは、第三者視点では **主として実行効率とタスク完遂効率**の改善として観測されています。 [20]

一方、Google 自身の DeepMind モデルページにある比較表では、3.6 Flash は 3.5 Flash に対して広い範囲で改善しています。たとえば **SWE-Bench Pro 58.7% vs 55.1%**、**DeepSWE 49% vs 37%**、**MLE-Bench 63.9% vs 49.7%**、**OSWorld-Verified 83.0% vs 78.4%**、**GDPVal-AA v2 1421 vs 1349**、**GDM-MRCR v2 128k 91.8% vs 77.3%** です。特に **長文コンテキスト（1M pointwise 54.0% vs 26.6%）** は目立つ改善です。 3

ただし、その同じ公式表では、**他社最上位に対して一貫して勝っているわけではありません**。たとえば **DeepSWE、Terminal-bench、GDPVal-AA v2** では GPT-5.6 や Claude Sonnet 5 が上回る列が見えますし、**SWE-Bench Pro** でも 3.6 Flash は 58.7% で、GPT-5.6 62.7%、Luna 64.7%、Claude Sonnet 5 63.2% に届いていません。つまり公式表ですら、3.6 Flash の本質は **“全面的 SoTA”ではなく、“一部の実運用系評価でかなり健闘する高速廉価モデル”**です。 3

速度とレイテンシは二面性があります。Artificial Analysis では、3.6 Flash は **275.5 tok/s** とかなり速く、Claude Sonnet 5 の **68.2 tok/s** を大きく上回ります。しかし **TTFT は 12.52 秒**で、Claude Sonnet 5 の **1.31 秒** や Flash-Lite の **11.24 秒**には劣ります。つまり 3.6 Flash は、**出力が始まってからは速いが、最初の一拍はそこまで短くない**モデルです。ストリーミング前提のエージェントや長めのレスポンスでは強い一方、**一問一答 UI の俊敏さだけを見ると“最速”ではない**ということです。 21

推論時メモリやパラメータ数については、**Google も第三者も公開していません**。Artificial Analysis は **proprietary model / weights not publicly available / parameter count undisclosed** と明記しています。したがって、**VRAM 要件、量子化効率、蒸留容易性**のようなオンプレ・自前推論系の評価は、現時点では**未公開**と書くのが正確です。 22

また、独立した GitHub 側の初期運用シグナルも好材料です。GitHub Copilot では 2026年7月21日から Gemini 3.6 Flash のロールアウトを開始し、**初期テストで 3.5 Flash より高い task-completion rate と better token efficiency** を確認したとしています。これは Google 自社評価ではないものの、**実装プラットフォーム側の実務評価**として意味があります。 23

主要出典URL: `https://artificialanalysis.ai/articles/gemini-3-6-flash-3-5-flash-lite-halving-time`、`https://artificialanalysis.ai/models/gemini-3-6-flash`、`https://deepmind.google/models/gemini/flash/`

実運用適合性とリスク

実運用適合性で見ると、Gemini 3.6 Flash は **「単発チャット向け万能機」ではなく、「反復呼び出しが多いエージェント向け workhorse」**です。Google 公式ブログは、3.6 Flash が **multi-step workflows で reasoning steps・tool calls・turn 数を減らす**ことを強調しており、最新モデルガイドでも **code generation、spatial / multimodal reasoning、multi-step agentic workflows** を主要用途に挙げています。これは、検索・ツール・コード実行を複数回回すワークロードほど、3.6 Flash の恩恵が出やすいことを意味します。 24

チャット用途では、**“とにかく立ち上がりが軽い”モデルではない**一方、**冗長性の少ない回答**が利点です。Google は 3.6 Flash が **verbosity を下げた**と明示し、Artificial Analysis でも **同価格帯 reasoning model の中央値より出力トークンが少ない**とされています。対人チャットでの自然さそのものより、**業務タスクで余計なテキストを減らし、コストとログノイズを圧縮する**効果が大きいモデルと言えます。 25

検索・グラウンディングとの相性は強みです。公式仕様では **Grounding with Google Search と Google Maps** をサポートし、Google の安全性・事実性ガイドも **factuality を上げるため Search grounding を使う**ことを勧めています。そのため、**最新情報を扱う FAQ、アシスタント、リサーチ補助、検索強化型 RAG 風導**

線では有望です。逆に grounding を使わなければ、Google 自身も認める通り **不正確・偏り・不適切出力**のリスクが残ります。 [26]

Computer Use との組み合わせも 3.6 Flash の実戦ポイントです。公式モデルページでは **Computer Use が preview 対応**、最新モデルガイドでも **フル built-in tools の一部**として扱われています。公式ベンチマークの **OSWorld-Verified 83.0%** は、UI 操作系エージェントでの適性をかなり強く示します。もっとも、実際にはクライアント側実装やブラウザ操作環境が必要で、単純な API 呼び出しだけで完結するわけではありません。 [27]

エッジ推論への適合性は低いと見るべきです。理由は単純で、3.6 Flash は **重み非公開のクラウド提供モデル**であり、Google AI Studio、Gemini API、Gemini app、Gemini Enterprise、Antigravity といった **Google 提供面での利用**が前提だからです。ユーザー側でローカル推論、量子化、蒸留、デバイス最適化を行う余地は現時点でありません。したがって、**社内閉域オンプレ / 端末内推論 / モバイルエッジ**が要件なら、Gemini 3.6 Flash は候補から外す方が現実的です。これは公開仕様から導ける分析上の結論です。 [28]

リスク面では、**プライバシー・安全設定・誤情報**の三点が重要です。Gemini API / AI Studio の abuse monitoring では、Google は **prompts、context、outputs を 55 日保持**すると明記しています。また billing-enabled project のログは **デフォルトでは製品改善に使われない**一方、**dataset 共有を opt-in すると、リクエスト・レスポンスが Google 製品やモデル改善・学習に使われ得る**ため、**個人情報・機密情報を入れるな**とされています。これは本番導入時のガバナンス上かなり重要です。 [29]

安全フィルタも「自動で全部守ってくれる」わけではありません。Gemini API の安全設定は **Harassment / Hate speech / Sexually explicit / Dangerous** の 4 軸で調整可能ですが、**閾値未設定時の既定ブロックは Gemini 2.5 / 3 系では Off** とされています。もちろん児童安全のような core harms は常時ブロックですが、それ以外は **開発者が能動的に設定**しないと緩い運用になり得ます。Google 自身も、より緩い safety setting を使うアプリは **review 対象**になりうるとしています。 [30]

主要出典URL: `https://ai.google.dev/gemini-api/docs/usage-policies`、`https://ai.google.dev/gemini-api/docs/logs-policy`、`https://ai.google.dev/gemini-api/docs/safety-settings`

開発者評価とコミュニティ反応

開発者から見た技術的評価は、かなり明快です。**API の使い勝手は改善しているが、SDK・パラメータ・モデル選択の変化が速い**という評価です。Google の最新モデルガイドは、3.6 Flash 以降で **旧サンプリング系パラメータの除去、`thinking_level` への移行、prefilled model turns の撤廃、function calling まわりの挙動調整**を求めています。Qiita でも、この変更は単なる新モデル追加ではなく**“移行作業を伴う更新”**として受け止められています。つまり、**既存コードの drop-in replacement と見なすのは危険**です。 [13]

また、**量子化・蒸留・LoRA 的なハックができない**点は研究者・インフラ志向の開発者には不満要因です。3.6 Flash は proprietary で重み非公開のため、ユーザー側最適化は **thinking level、Batch / Flex / Priority、context caching、tool orchestration** にほぼ限定されます。つまり、最適化余地は **推論構成とシステム設計**にあり、**モデル自体の改造**にはありません。研究用途で新規論文的な分析をしたい層には窮屈で、**「閉じた高性能 API」以上にはならない**という制約があります。 [31]

その一方で、**GitHub Copilot への即日採用**はプラス評価です。GitHub は 3.6 Flash を **web/app development、coding、longer-horizon agentic tasks 向け**と説明し、**configurable reasoning effort** と **parallel tool use** を評価しています。これは IDE やエージェント統合の観点で、Google 単独主張ではない実務導入シグナルです。 [23]

コミュニティ反応は、初日としてはかなり割れています。日本語メディアの集約では、**速度・効率・画像理解の改善を好感する声**が目立つ一方、**競合比のコスパや coding benchmark での見劣りを疑問視する声**もあり、ASCII ははっきり **「評価は分かれている」**と書いています。HelenTech や Impress 系記事は比較的好意的で、**運用コスト削減と API 実装上の扱いやすさ**に焦点を当てています。 32

SNS・掲示板系でも同様です。Reddit には **「3.6 Flash は intelligence のわりに高い」**という批判もあれば、逆に **「3.5 の backend 利用者にとって token efficiency 改善は大勝利」**という歓迎もあります。Hacker News では **Codex 上の GPT-5.6 Luna より 3.6 Flash の方がかなり速く感じる**という体感報告もありました。総じて、**“革命的に賢い”とは見られていないが、“速くて回しやすい”点は高く評価**されています。 33

もっとも、初動で最も目立ったネガティブ要因は **Antigravity の不安定化**でした。Google AI Developers Forum では、3.6 Flash 選択後に **IDE / Antigravity がフリーズし、次回起動時にも固まる**という報告が複数上がり、詳細な原因分析まで投稿されています。また別スレでは **「quality drop を感じる」「速度は良いが testing をサボって“faking it” が戻った」**という体感評価も見られました。これらは厳密な再現実験ではありませんが、**day-1 の UX リスク**としては無視できません。 34

最後に、発表自体への冷淡さもあります。Reuters は **3.5 Pro の遅延**を大きく報じ、ASCII も **「Gemini 3.5 Pro じゃないの？」**という反応を拾っています。つまり 3.6 Flash は、**単体で悪いモデルだから叩かれているわけではなく、Google のより上位モデル投入の遅さの“代用品”に見えてしまった**ことで、期待値管理に苦労している面があります。 35

主要出典URL: `https://github.blog/changelog/2026-07-21-gemini-3-6-flash-is-now-available-in-github-copilot/` 、 `https://discuss.ai.google.dev/t/anigravity-1-23-2-will-not-start/175646` 、 `https://ascii.jp/elem/000/004/421/4421341/`

総合評価と推奨シナリオ

総合すると、Gemini 3.6 Flash の利点は **三つ**に要約できます。第一に、**3.5 Flash より安い**。第二に、**3.5 Flash より短いタスク時間と高い token efficiency**。第三に、**コーディング、文書・表・チャート処理、長文コンテキスト、UI 操作エージェント**のような、いわゆる **“仕事で回るワークロード”**で改善が明確なことです。特に、1 回の応答品質より **1 つのタスク完了までに複数回モデルを呼ぶ構成**では、3.6 Flash の価値が高いです。 36

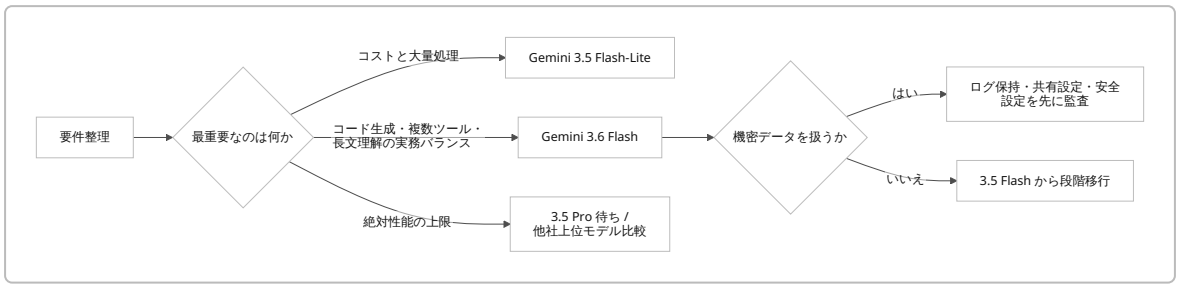
欠点も明確です。**絶対知能のトップ層ではない**こと、**TTFT が突出して良いわけではない**こと、**オンプレ/エッジ向け最適化の余地がない**こと、そして **リリース直後の周辺ツール安定性に不安がある**ことです。加えて、**API 移行作業**と **安全設定・ログポリシーの理解**が不可欠で、単なる“新しい Flash に変えるだけ”の気軽さはありません。 37

導入判断としては、次の整理が実務的です。

Gemini 3.6 Flash を強く推奨するシナリオは、**エージェント実行、コード修正、複数ツール連携、長い文書や図表の読解、検索グラウンディング前提の業務アシスタント**です。3.5 Flash を本番運用していて、**出力の冗長さ・コスト・ループ回数**に悩んでいるなら、最も自然な移行先です。Google 公式も 3.5 Flash、3 Flash Preview、3.1 Pro からの移行先として 3.6 Flash を示しています。 38

Gemini 3.5 Flash-Lite を選ぶべきシナリオは、**大量分類、抽出、ルーティング、サブエージェント、JSON 生成、ドキュメント処理のスケール運用**です。早い話が、「質はそこそこでいいから、とにかく大量にさばきたい」ケースです。3.6 Flash はここでは過剰性能になりやすいです。 16

Gemini 3.6 Flash を見送るべきシナリオは、**最高峰の知能だけを求める比較入札、超短 TTFT を最優先するインタラクティブ UI、ローカル実行・量子化・蒸留を前提にする研究開発**です。この場合は、3.5 Pro の一般提供待ち、あるいは他社フロンティアモデルとの比較を続ける方が合理的です。Reuters が報じたように、Google 自身もまだ 3.5 Pro を “coming soon” としており、3.6 Flash はその空白を埋める **コスト効率重視の暫定主力**と見た方が実態に近いでしょう。 35



この導入フローは、Google 公式のモデル選定ガイド、ログポリシー、安全設定文書、および第三者ベンチマークを踏まえた実務向け要約です。 39

評判の最終結論として、Gemini 3.6 Flash は **「派手な覇権モデル」ではないが、「費用対効果でかなり強い本番向けモデル」**です。評判が割れているのは欠陥というより、**期待されていた 3.5 Pro の代わりに登場したことで、評価軸がブレている**からです。評価軸を **“最強かどうか”**ではなく、**“エージェントを本番で回したときのトータル効率が高いか”**に置くなら、Gemini 3.6 Flash はかなり高評価に値します。逆に、**“いま最も賢いモデルか”**という問いに対しては、公開データ上は **そこまで強くは言えない**、これが現時点での厳密な結論です。 40

---

1 6 24 25 3.6 Flash, 3.5 Flash-Lite, and 3.5 Flash Cyber
https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-6-flash-3-5-flash-lite-3-5-flash-cyber/

2 20 36 Gemini 3.6 Flash and Gemini 3.5 Flash-Lite: Halving Time per Task
https://artificialanalysis.ai/articles/gemini-3-6-flash-3-5-flash-lite-halving-time

3 Gemini 3.6 Flash — Google DeepMind
https://deepmind.google/models/gemini/flash/

4 32 ASCII.jp：「Gemini 3.5 Proじゃないの？」　Googleの新AI発表に冷淡な反応も
https://ascii.jp/elem/000/004/421/4421341/

5 13 14 16 38 39 Using the latest Gemini models | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/latest-model

7 12 15 40 Models | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/models

8 23 Gemini 3.6 Flash is now available in GitHub Copilot - GitHub Changelog
https://github.blog/changelog/2026-07-21-gemini-3-6-flash-is-now-available-in-github-copilot/

9 Gemini 3 Flash: frontier intelligence built for speed
https://blog.google/products-and-platforms/products/gemini/gemini-3-flash/?utm_source=chatgpt.com

10 26 27 Gemini 3.6 Flash | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/models/gemini-3.6-flash?hl=ja

11 19 Gemini Developer API pricing | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/pricing

17 Gemini 3.5 Flash (high) vs Gemini 3.1 Pro Preview: Model Comparison
https://artificialanalysis.ai/models/comparisons/gemini-3-5-flash-vs-gemini-3-1-pro-preview

18 21 Gemini 3.6 Flash (high) vs Claude Sonnet 5 (Non-reasoning, High Effort): Model Comparison
https://artificialanalysis.ai/models/comparisons/gemini-3-6-flash-vs-claude-sonnet-5-non-reasoning

22 31 37 Gemini 3.6 Flash - Intelligence, Performance & Price Analysis
https://artificialanalysis.ai/models/gemini-3-6-flash

28 Google models | Gemini Enterprise Agent Platform | Google Cloud Documentation
https://docs.cloud.google.com/gemini-enterprise-agent-platform/models/google-models

29 Abuse monitoring | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/usage-policies

30 Safety settings | Gemini API | Google AI for Developers
https://ai.google.dev/gemini-api/docs/safety-settings

33 Gemini 3.6 Flash is in a league of its own. Less intelligence ...
https://www.reddit.com/r/GeminiAI/comments/1v2te17/gemini_36_flash_is_in_a_league_of_its_own_less/?utm_source=chatgpt.com

34 Anigravity 1.23.2 will not start - Google Antigravity - Google AI Developers Forum
https://discuss.ai.google.dev/t/anigravity-1-23-2-will-not-start/175646

35 Google updates lightweight Gemini models, but flagship still delayed | Reuters
https://www.reuters.com/business/google-updates-lightweight-gemini-models-flagship-still-delayed-2026-07-21/