

AIモデルを使い分けると、 月のコストが半分になった

Claude Sonnet 5が変えた「性能とコストの両立」

この資料の要点

- Claude Sonnet 5は上位モデル（Opus 5）の6割超の精度をほぼ半額以下で実現し、「全部Opusで」という使い方を見直す契機になっている。
- 業務を「定型・繰り返し型」と「複雑・判断型」に分類し、前者は低価格モデル・後者は上位モデルと割り当てただけでコストを大幅削減できる。
- ノーコードツール（n8nやDifyなど）でもモデルをノードごとに切り替えられ、技術的なハードルは高くない。

「高いモデル1本」から「用途で使い分け」へ

最上位モデルを全業務に使い続けているために月々のコストが想定外に膨らむ——というケースは珍しくありません。しかし2026年7月に登場したClaude Sonnet 5は、エージェントコーディングのベンチマークでOpus 5の約91%のスコアを半額以下のコストで実現し、使い分けの実益が格段に大きくなりました。

業務タイプ別・モデル選定の早見表

業務タイプ	推奨モデル	代表例
定型・繰り返し型	Sonnet 5 / GPT-4o mini	メール下書き、FAQボット、データ分類、議事録整形
中程度の判断が必要	Sonnet 5 / GPT-4o	提案書の初稿、要件整理、マルチドキュメント要約
複雑・高度な判断型	Opus 5 / GPT-4o (o3)	競合分析レポート、契約リスク洗い出し、長期計画立案

コスト比較試算（Claude API・2026年8月時点）

パターン	月間推定コスト（目安）
Opus 5のみ（入力300万・出力150万トークン）	約 \$157.5 / 月
定型70%をSonnet 5＋残30%をOpus 5で使い分け	約 \$62 / 月（▲60%）

※モデル料金・為替・プランにより変動。Sonnet 5: 入力\$2/出力\$10、Opus 5: 入力\$15/出力\$75（100万トークンあたり目安）。

使い分け設計を定着させる3ステップ

Step 1 — 現状の利用コストを可視化する

直近1〜2ヶ月のAIツールの請求明細・利用ログを確認し、どのモデルにどれだけコストが発生しているかを把握します。「ほとんどを最上位モデルで処理していた」と気づくだけで、改善の優先順位が見えてきます。

Step 2 — 業務タスクを3段階に分類して一覧化する

「低価格モデルで十分」「迷ったら上位モデル」「必ず上位モデル」という3段階の基準を、チームで共有できる一枚の表にまとめます。個人の判断に委ねず、チームで揃えることが長続きの鍵です。

Step 3 — 四半期ごとに品質確認と基準を更新する

モデルは継続的に更新されます。四半期に一度、代表的なタスクをサンプリングして品質確認と基準の見直しを行うことで、コスト効率が長期的に維持されます。

ノーコードツールでの実践例

n8n / Dify：ノードごとにモデルを指定

n8nやDifyではAIノードごとにモデルを指定できます。「問い合わせ一次分類はSonnet 5で、クレーム・緊急と判定された場合のみOpus 5で返答案を生成する」といった分岐が、プログラミングなしで設計できます。

Claude.ai ProプランでもSonnet↔Opus使い分け可

Sonnet 5を日常アシスタントとして、Opus 5を「重要資料のレビュー」「複雑な判断サポート」専用として使い分けると、同じプラン内でより多くの処理を消化できます。

電気代が高いからといってすべての家電を止めるのではなく、消費電力の大きい家電の使い方を見直す——それと同じ発想がAI活用にも当てはまります。

よくある質問

Q. Claude Sonnet 5とOpus 5は何が違うのですか？

コストと推論の深さが主な違いです。Sonnet 5は定型業務に十分な精度を低価格で提供し、Opus 5は複合的な推論・長文コンテキストに本領を発揮します。「どちらが優れているか」ではなく「どの業務に使うか」で選ぶのが正解です。

Q. 中小企業がモデル選定を始めるには、どこから手をつければいいですか？

まずは現在使っているAIツールの「月次利用履歴」か「請求明細」を確認してください。コストが大きい処理から「本当に上位モデルが必要か」を問い直すだけで見直せる箇所が見えてきます。新規導入の場合は、まず低価格モデルで試作し、品質不足の部分だけ上位モデルを使う「下から積み上げる」アプローチが有効です。

Q. AIモデルの使い分けは技術的に難しいですか？

APIを直接使う場合は「model」パラメータを変えるだけです。ノーコードツール（n8nやDifyなど）でも、ノードごとにモデルを指定できる設計が多く、プログラミング知識なしで運用できます。ChatGPTの画面利用なら、会話開始時にモデルを選ぶ習慣を付けるだけで始められます。

Q. ChatGPTのプランにも「使い分け」はできますか？

OpenAIでも同じ発想は使えます。GPT-4o miniはコスト効率に優れ定型処理に向いており、GPT-4oやo3は複雑な分析・推論に強い分、費用も上がります。企業APIプランではモデルをリクエストごとに切り替えられます。

CONTACT

株式会社DnDでは、AI活用コストの見直しや業務タイプに合ったモデル・ツール選定のご支援を承っています。

お問い合わせ：<https://dnd-inc.jp/contact.html>