

生成AIの基礎

Chapter 1 Section 1 | AIEージェント・ストラテジスト試験対策

今日学ぶこと

1 **AIの3フェーズ**
作るAI / 使うAI / 任せるAI

2 **生成AIの正体**
非構造化データ活用が最大の変化

3 **LLMの仕組みと限界**
ハルシネーション・トークン・コンテキストウィンドウ

4 **業務シナリオ：ヘルプデスクAI企画**
モデルルーティングとAPI費用試算

5 **性能を決めるのはモデルではない**
データ整備が競争優位の源泉

なぜストラテジストが
生成AIの中身を
知る必要があるのか？

AIの進化を捉える3つのフェーズ

フェーズ1

作るAI（～2022）

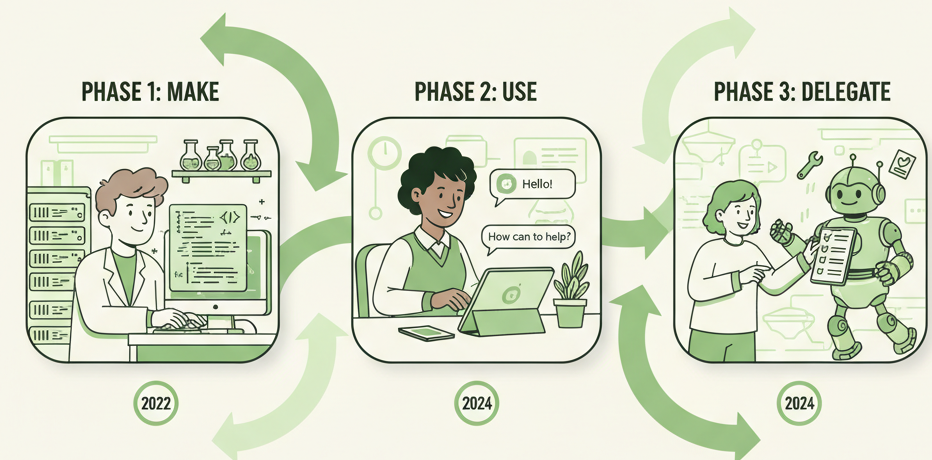
フェーズ2

使うAI（2022～）

フェーズ3

任せるAI（2024～）

ストラテジストに求められるのは個々のトレンドを追うのではなく、AIの本質的な変化の方向を見極めること。



作るAI

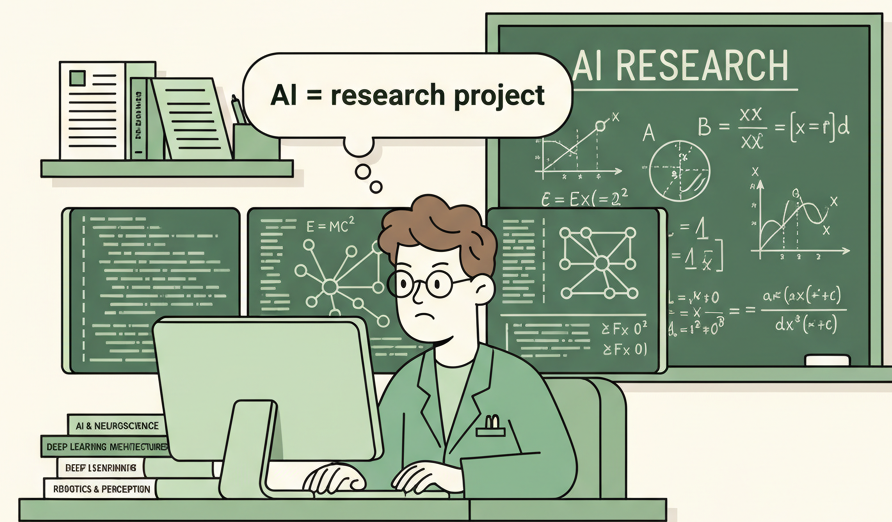
使うAI

任せるAI

フェーズ1：作るAI（～2022年ごろ）

- AIは「使う」ものではなく「依頼する」もの
- 主役：研究者・データサイエンティスト・AIエンジニア
- AIプロジェクトは数ヶ月～数年単位の大規模投資
- 経営層はAIを「IT投資の一種」として位置づけ

ルールベース、機械学習、ディープラーニング。いずれも高度な数学的知識とプログラミングスキルが前提だった時代。

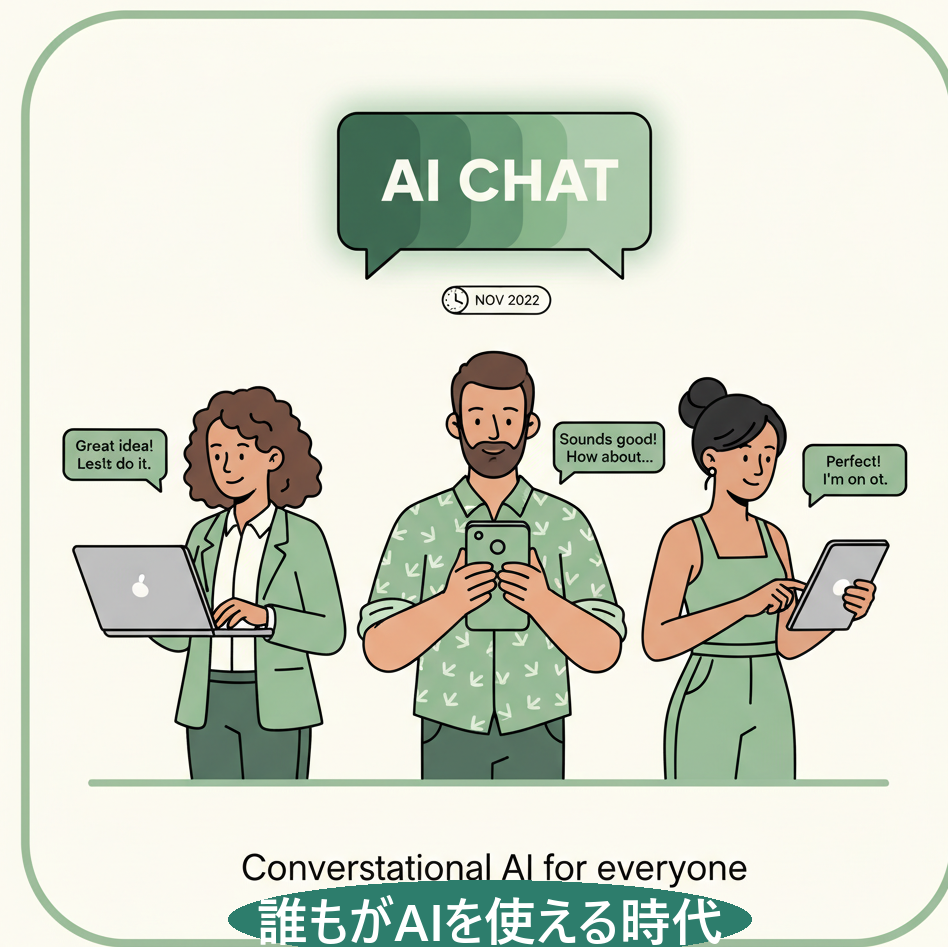


専門家だけがAIを作る時代

フェーズ2：使うAI（2022年～）

- 2022年11月のChatGPT公開が転換点
- 自然言語で誰でもAIに指示できるように
- 主役はビジネスパーソン・一般ユーザー
- 競争優位は「いかに上手く使いこなすか」

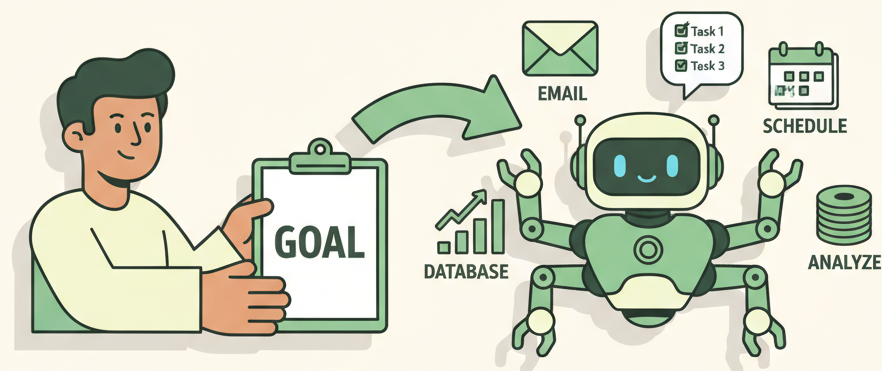
多くの組織はいまもこのフェーズの途上にある。プロンプト設計・ワークフロー改善・AI人材育成が経営課題に。



フェーズ3：任せるAI（2024年～）

- AIが**目標**を受け取り、タスク分解・ツール呼び出し・実行を**自律的に**こなす
- 「回答ツール」から「**行動する存在**」へ
- 業務設計・組織設計・意思決定の構造ごと変わる

本資格が扱うのは、まさにこのフェーズです



目標を伝えればAIが実行

補足：AGI / ASI をどう扱うか

AGI

汎用人工知能（Artificial General Intelligence）

人間と同じように「なんでもこなせる」AI。
身体性問題もあり、実現時期は不確定

ASI

人工超知性（Artificial Superintelligence）

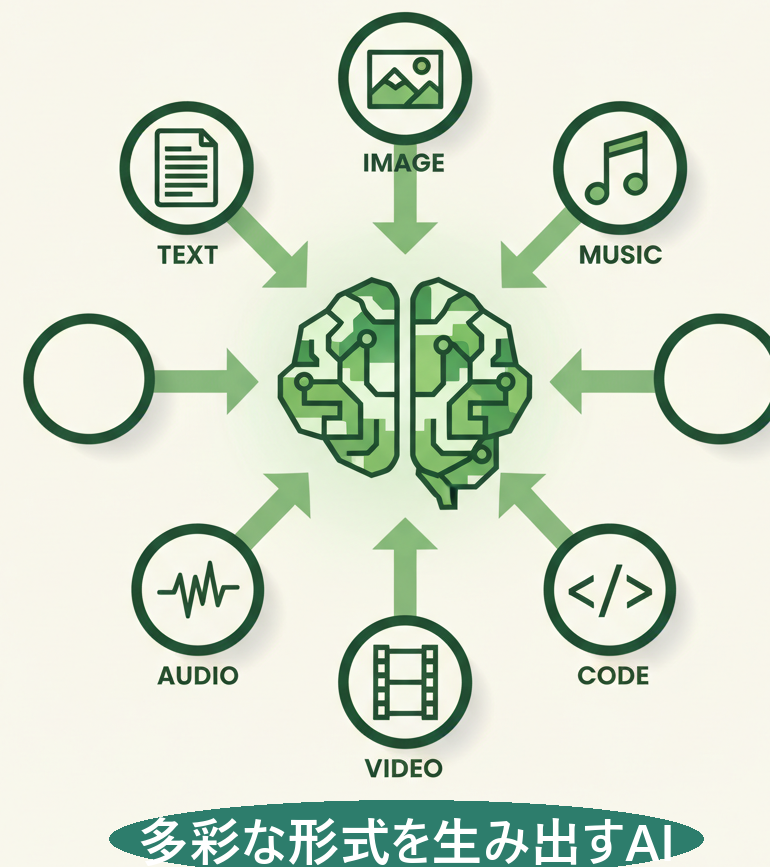
あらゆる面で人間をはるかに超える知能。
AGIのさらに先の仮説の領域

ストラテジストの姿勢：単なる夢物語として退けず、確定した近未来として語らず、「不確実性の高い長期シナリオ」として冷静に参照する。

生成AIとは

- 入力された内容をもとに、テキスト・画像・音声・動画・コードなどを生み出すAI
- 従来のAIは「予測」「識別」が中心だった
- データのパターンを学習し、新しい出力を生成できる

英語で**Generative AI**。テキスト生成、画像生成、音声生成、音楽生成、プログラム生成、動画生成と多様な形式に対応。



真のインパクトは「非構造化データ活用」

- 従来のDXで扱えたのは行と列に整理された構造化データ中心
- 生成AIで議事録・企画書・チャットなど「文章そのもの」も処理可能に
- 企業に眠るデータの8割以上は非構造化

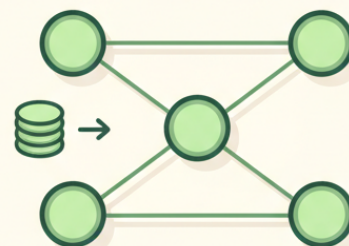
この領域に手が届くようになったことが、生成AIの最大の変化

機械学習とディープラーニング

- **機械学習**：データからパターンを自動学習。手法は教師あり/教師なし/強化
- **ディープラーニング**：人間の脳を模した多層ニューラルネットワーク
- ディープラーニングの強み = **特徴量の自動抽出**

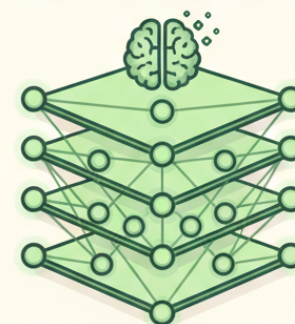
人間が「これは猫の耳だ」と特徴量を設計する必要がなくなり、**非構造化データから直接学べる**ようになった。LLMはこの技術の延長線上にある。

MACHINE LEARNING



EVOLUTION

DEEP LEARNING



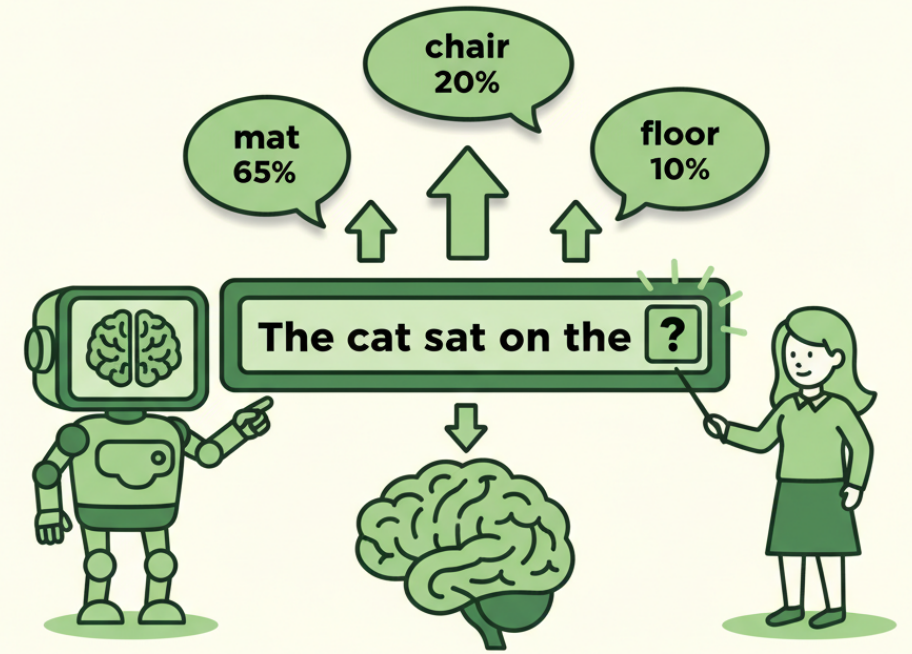
機械学習

ディープラーニング

LLMの仕組み

- Large Language Model（大規模言語モデル）＝生成AIの中核技術
- インターネット上の膨大なテキストから言語のパターンを学習
- 動作原理は「次に来る最も自然な言葉は何か」を確率予測すること

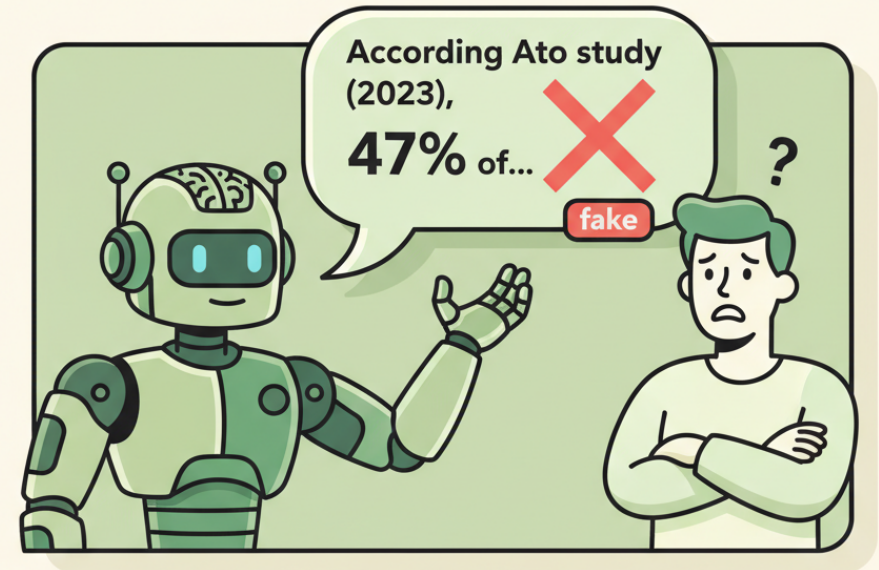
「予測を繰り返すだけ」と聞くと頼りなく感じるが、数千億の文章パターンを学習した結果、文脈理解・推論・要約・翻訳・コード生成までできるように。



必ず押さえる：ハルシネーション

- AIが事実でない内容を、もっともらしく生成する現象
- 原因：「正しいことを言う」のではなく「次に来る言葉として自然なものを選ぶ」仕組み
- 自信たっぷりに出力されるため、専門外の内容では見抜きにくい

ハルシネーションは欠陥ではなく言語モデルの構造的特性。完全になくすことは難しい。一次ソースで確認する習慣が不可欠。

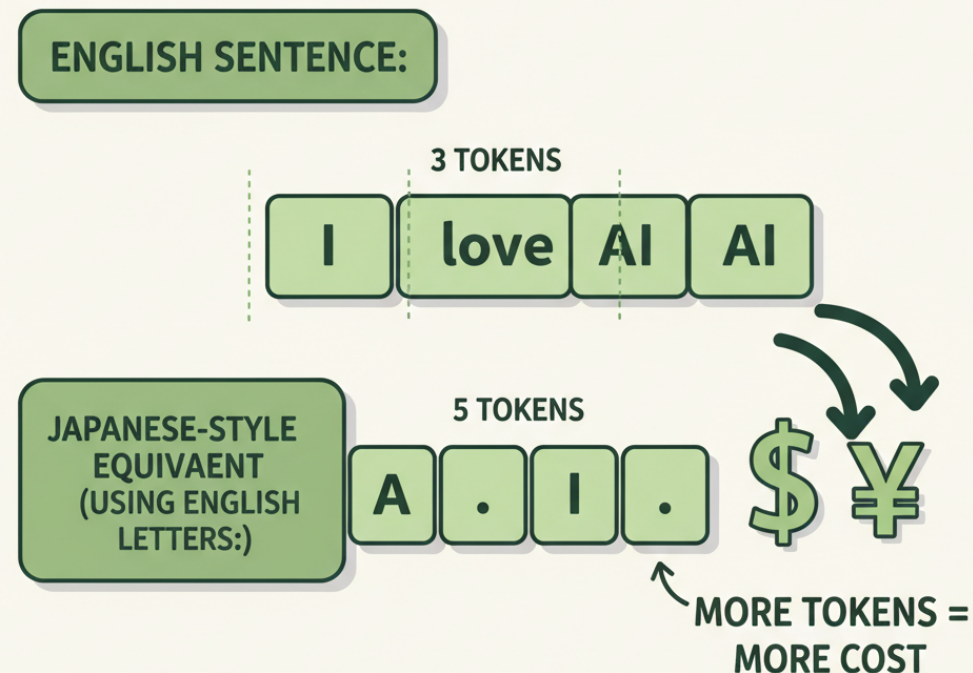


自信たっぷりに事実と違うことを言う

トークン：コスト構造の出発点

- LLMはテキストを**トークン**という最小単位に分割して処理
- 英語は1単語=約1トークン、日本語は1文字～数文字で1トークン
- APIはトークン数に応じた従量課金

月間でいくらかかるかを見積もれるのがストラテジ
ストの基本スキル

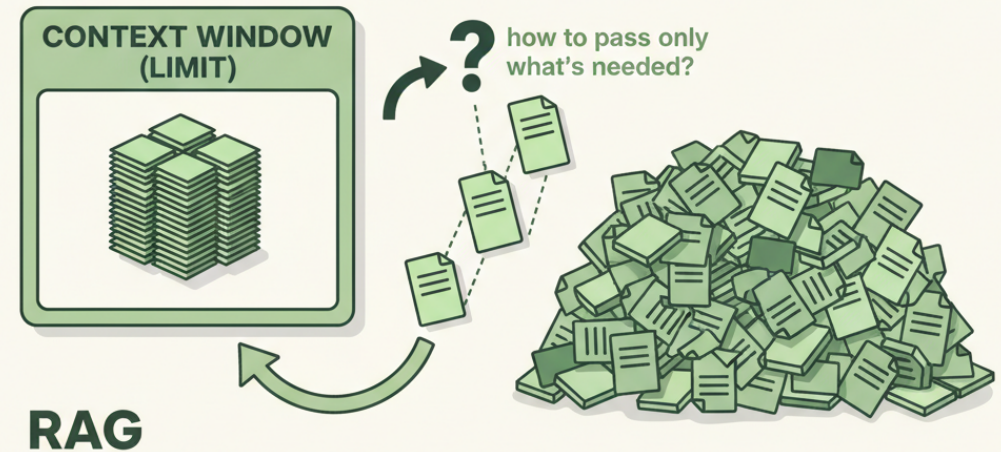


トークンが多いほどコスト増

コンテキストウィンドウとRAG

- 1度にLLMに渡せる**トークン数の上限**
- 現在の主要モデルは大幅拡大、技術的制約はほぼ解消
- ただし「**入る**」と「**入れるべき**」は別

毎回マニュアル全文を入力すると月間コストが数十倍に膨らむ。**必要な部分だけ検索して渡す**仕組みがRAG（Section 14で詳説）。



必要な部分だけ渡す = RAG

シナリオ：ITヘルプデスクAIの現状業務

体制と数字

全社**3,000名**のIT部門ヘルプデスク
月**500件**の問い合わせを**3名**で対応
FAQ該当：月300件×15分 = **月75時間**
非FAQ：月200件×45分 = **月150時間**
合計**月225時間**消費中

日々の作業の流れ

メール受信→カテゴリ判断
FAQフォルダのExcelをキーワード検索
FAQ該当：手入力で返信／非FAQ：個別調査
対応記録もExcelに手入力
月次レポートは管理者が手作業集計

今ある課題：担当者の暗黙知に依存（属人化）／Excelキーワード検索だと該当検出が漏れる／対応記録漏れ月**20件**／月次集計に**8時間**

業務シナリオ：ITヘルプデスクAIの企画

現状をふまえて いまの月225時間を、3ステップで企画に落とし込む

ステップ1

件数と内容から
タスクの性質を見極める

ステップ2

性質に応じて
モデルルーティング

ステップ3

処理件数から
API費用を概算



シナリオ：月500件のヘルプデスク

ステップ1：タスクの性質を見極める

- ヘルプデスク業務は2種類の処理でできている
- ① **分類処理**：パスワード系・ネットワーク系・ソフトウェア系に仕分け→**識別**
- ② **回答生成**：PCの環境に合わせた対処法を文章に→**生成**

1つの業務に識別的な処理と生成的な処理が混在するのは珍しくない。改善プランは、まずタスクの性質を分解することから始まる。

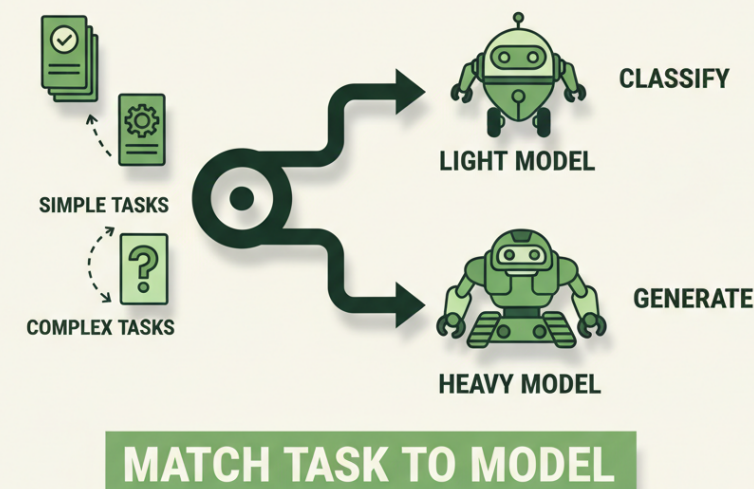


シナリオ：月500件のヘルプデスク

ステップ2：モデルルーティング

- ステップ1で分けた2つのタスクに、それぞれ違うモデルを当てる
- 分類処理 = 軽量・低コストLLM
- 回答生成 = 高性能LLM

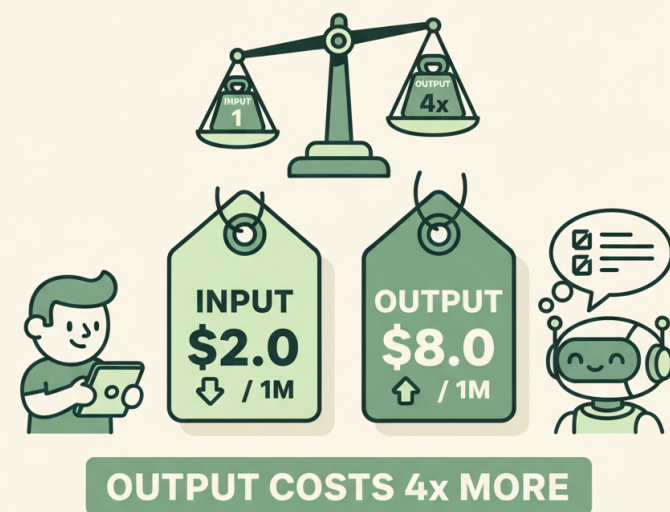
コストと品質を両立させる基本パターン



タスクに合わせてモデルを使い分け

試算の前提値を押さえる

項目	想定値
入力トークン 単価	約 \$2.0 / 100万トークン（GPT-4.1 クラス）
出力トークン 単価	約 \$8.0 / 100万トークン（同上）
1件あたり入 力	約 2,500トークン（質問 + FAQ参 照含む）
1件あたり出 力	約 500トークン（回答文）

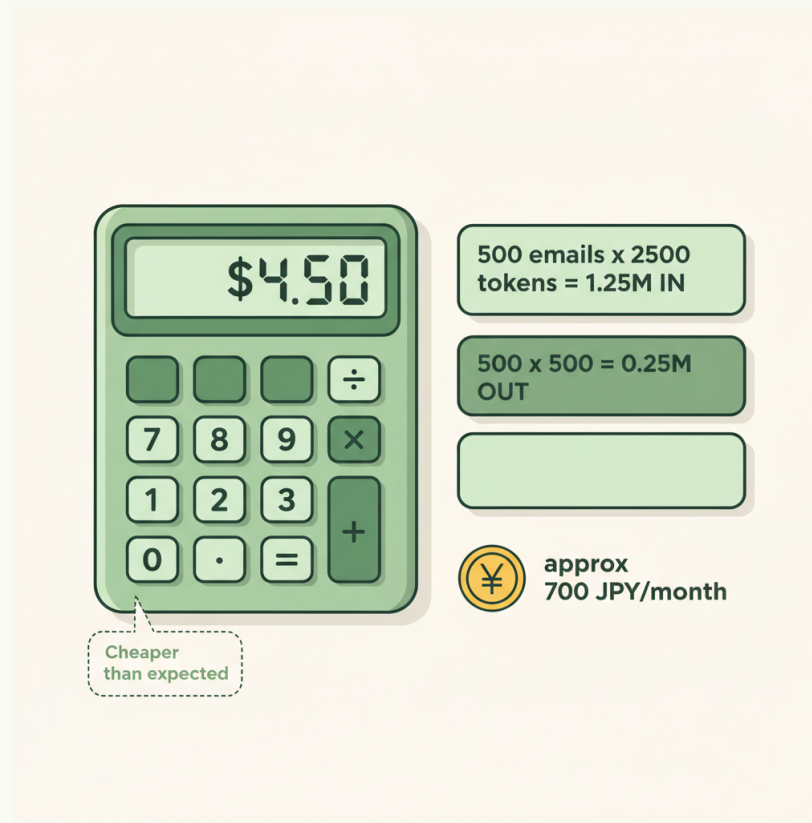


シナリオ：月500件のヘルプデスク

ステップ3：API費用を概算する

項目	量	コスト
入力（500件 ×2,500トークン）	月125万 トークン	約\$2.50
出力（500件 ×500トークン）	月25万ト ークン	約\$2.00
合計	-	約\$4.50/月 （≒700円）

「意外と安い」が正しい感覚。API費用より構築費用とランニングコストのほうがはるかに大きいのが実態。Section 6のTCOで詳述。



性能を決めるのはモデルではなくデータ

- どのLLMを選ぶかより、**自社データの質と量**が決定的
- OpenAIやGoogleが持っていない**自社固有の情報**こそ差別化要素
- 議事録や業務資料を蓄積してきた企業ほど**活用できる質に差**が出る

AIエージェントの性能はモデルではなくデータ整備で決まる

試験で問われる5つの論点

- 1 AIの3フェーズと本資格の位置づけ（任せるAI）
- 2 LLMの動作原理とハルシネーションの本質
- 3 トークン数とコストの見積もり方
- 4 モデルルーティングの考え方（軽量×高性能の使い分け）
- 5 コンテキストウィンドウとRAG（必要部分だけ渡す設計）