Part 2 ／ 60分

生成AIの衝撃 2026

原理を知る → 到達点を知る → 責任を知る









2022年11月、世界の前提が変わった

ChatGPT 公開 — 2ヶ月で1億ユーザー



あれから、まだ4年経っていません



皆さんは、AIのない大学生活を知らない最初の世代です

そして2026年、AIは『働く』ようになった



会話するAI → 作業するAI

コーディングエージェント: Claude Code ／ Codex ／ GitHub Copilot (agent mode)



ファイルを読み、コードを書き、実行し、失敗を見て直す ― これを数時間、自律的に

今日の午後、皆さんが使うのはこの系譜の道具です

20年サイクルで起こるデジタル大変革

― 転換期に学生でいることの意味



生成AIとは何か

―― 適用範囲が極めて広い、が今日の主役は言語モデル ――



大量のデータから「特徴」を学習する



学習した特徴をもとに、
まったく新しいデータを「生成」する

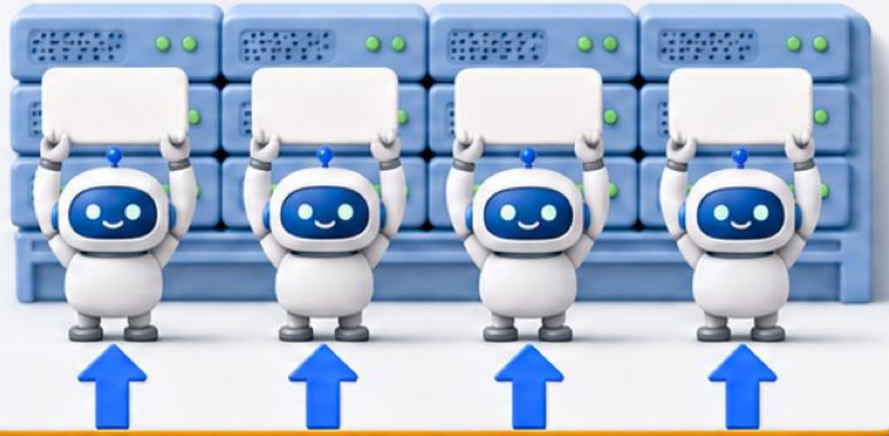
- テキスト
- 画像
- 音声
- 動画
- 3Dモデル
- プログラム

出発点は『Transformer』の発明 (2017)

1 注意機構 (Attention): 重要な語に焦点を当て、無関係な情報を無視する



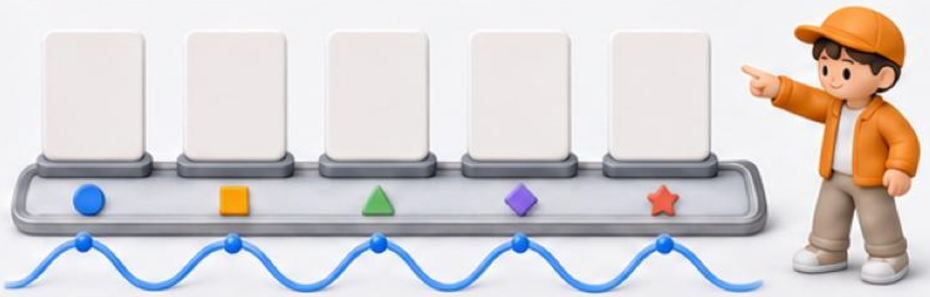
2 並列処理: 多数の語を同時に処理でき、学習を大規模化できる



3 エンコーダ・デコーダ: 応用の広い構造



4 位置エンコーディング: 語の順序を認識する

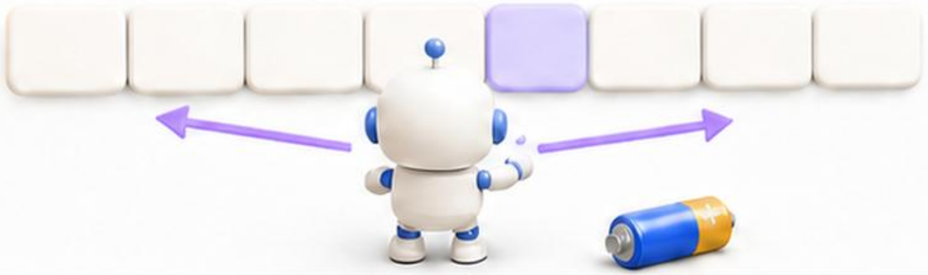




論文『Attention Is All You Need』(2017) — 今のモデルはすべてこの延長線上

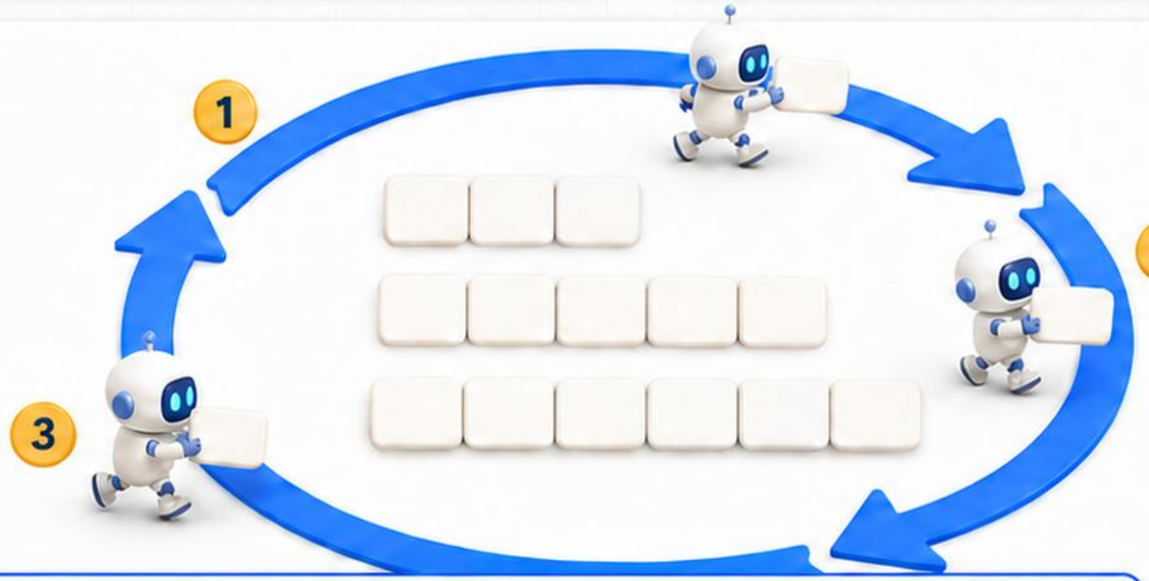
原理は『次の単語を予測する」だけ

穴埋め型:「電池が □ で機能を損なう」→ □ を推定する



継続型:「電池が放電で機能を」 → 次に来る語を予測する



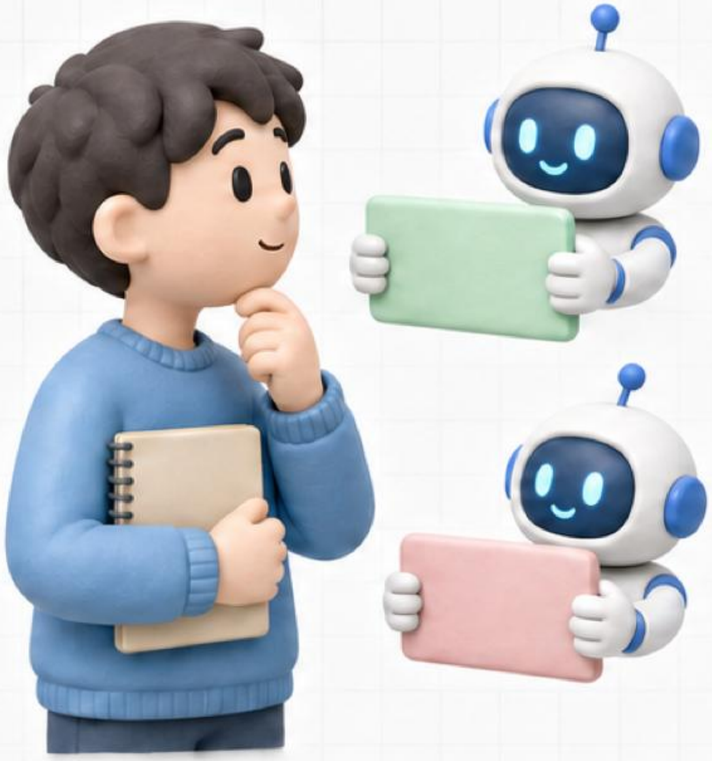


予測 → 出力 → その出力を入力に足して、また予測

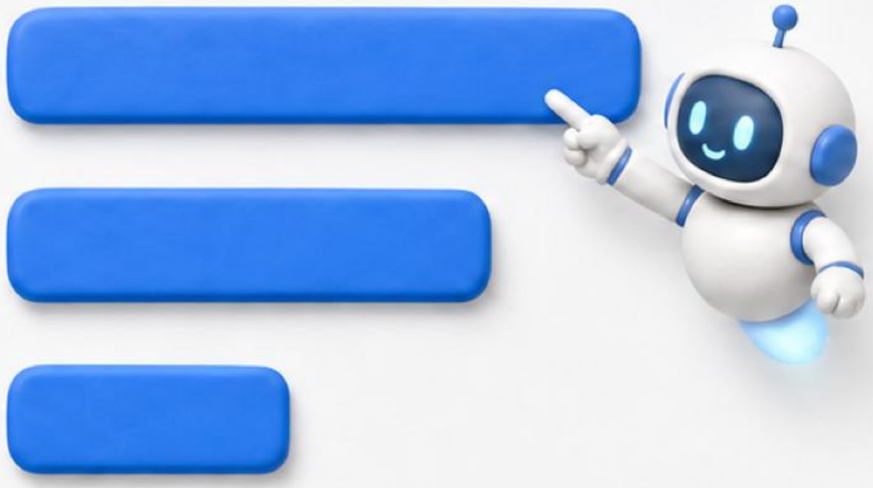


要点: これを繰り返すだけ。それでも出力には「知性」が感じられる

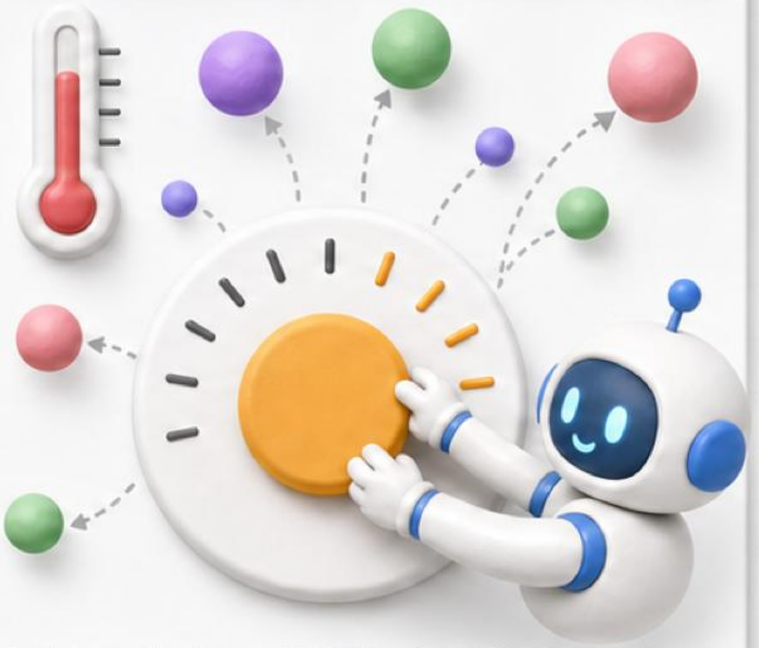
予測を確率で選ぶ — だから毎回違う答えになる



『電池が放電で機能を損なう』の次:『とき』40% ／『場合』28% ／『ので』12% …



統計力学のボルツマン分布の温度と同じ概念



温度 T を上げる = 低確率の語も選ばれやすくする

同じ質問に毎回同じ答えが返らないのは、故障ではなく設計

スケーリング則 — 賢さは投資で買えることが分かった

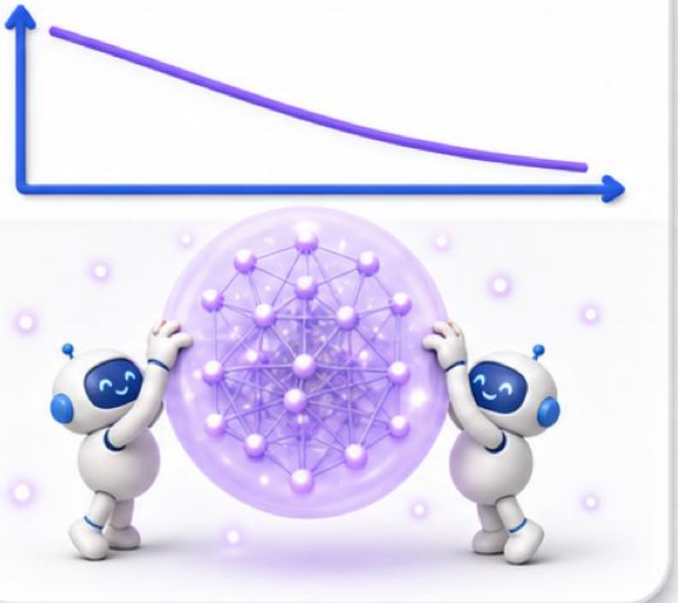
1 投入計算量



2 学習データ量



3 モデルの規模（パラメータ数）



この3つと予測誤差の間に、べき乗則 (Scaling Law) が成り立つ

= 投資対効果が事前に見積もれる

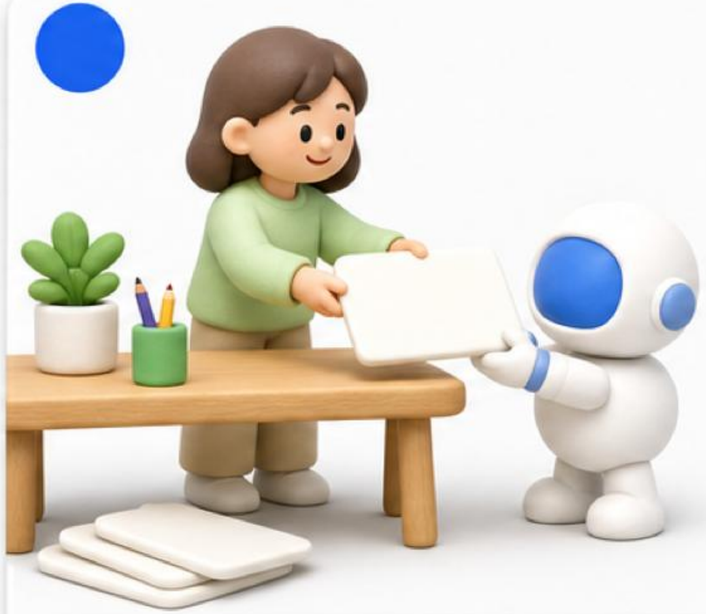




要点:
だから世界中のデータセンター投資が加速した

RLHF — 人間の好みを教える

次単語予測だけでは、ただ『続きを書く機械』にすぎない



① 良い応答の例で微調整する



② どちらの応答が良いか人間が順位づけし、報酬モデルを学ぶ



③ その報酬を最大化するよう強化学習する

要点: これで『指示に従う助手』になった (RLHF: Reinforcement Learning from Human Feedback)

推論時スケーリング ―『考える時間』を買う

学習時の計算量だけでなく、答えるときの計算量を増やしても性能が伸びる



要点: ＝ 難しい問題には時間とコストをかけて解かせる、という選択ができる

チャットからエージェントへ — ツールを持ったLLM

LLM は文章を出力するだけ。では、なぜ作業ができるのか

出力の一部を『ツールの呼び出し』として解釈する仕組みを外側に置いた



考える

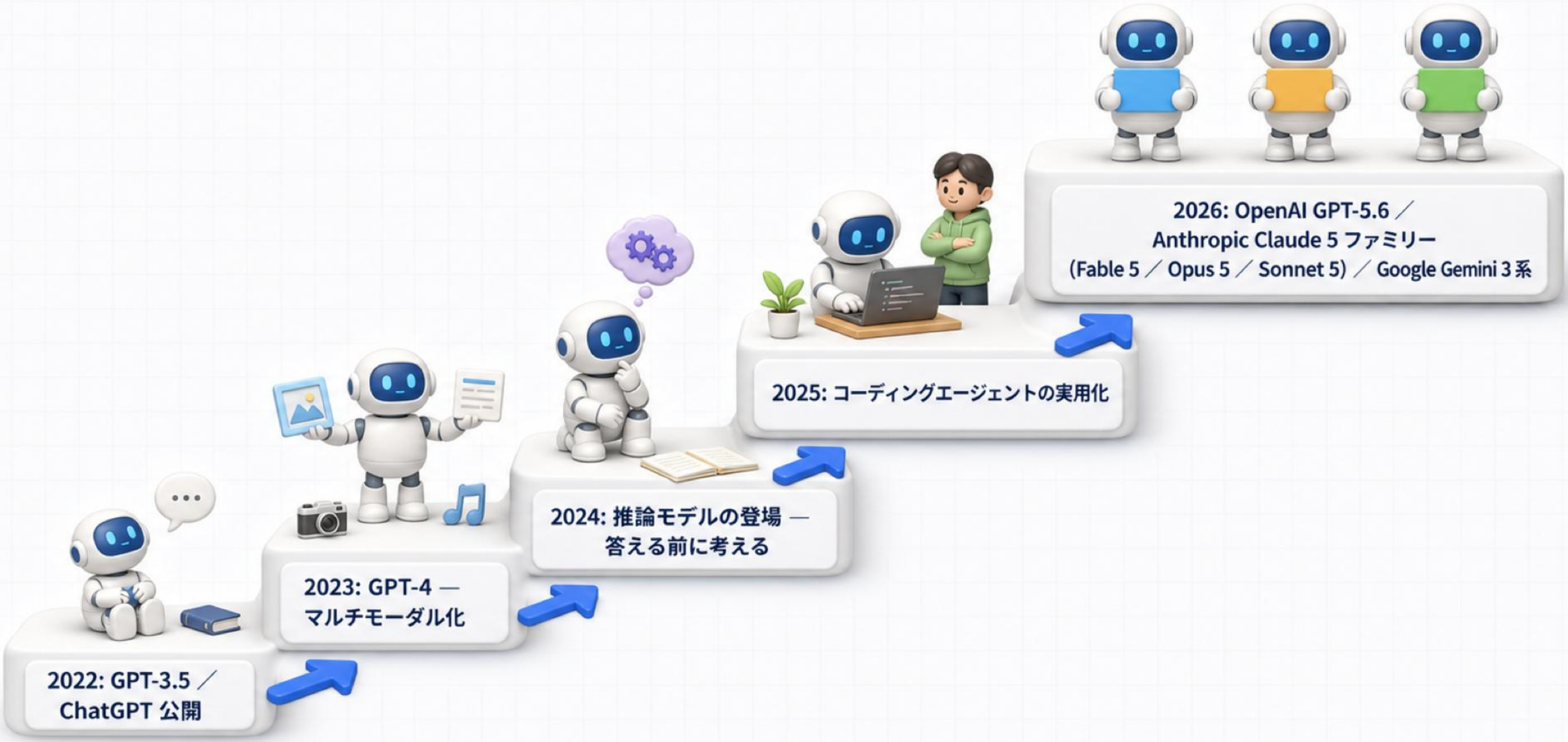
ツールを呼ぶ（読む・書く・実行する）

結果を読む

また考える

要点： このループを人間の確認なしで回せる範囲が、この2年で大きく伸びた

モデル進化の年表 (2022 → 2026)



※ 2026年7月時点。この表は数ヶ月で古くなります

2025年の予測は当たったか





当たった：エージェント化は
予測どおり進んだ — AIは道具を使い、
自律的に作業するようになった



判定できない：そもそも
AGI の定義が定まっていない

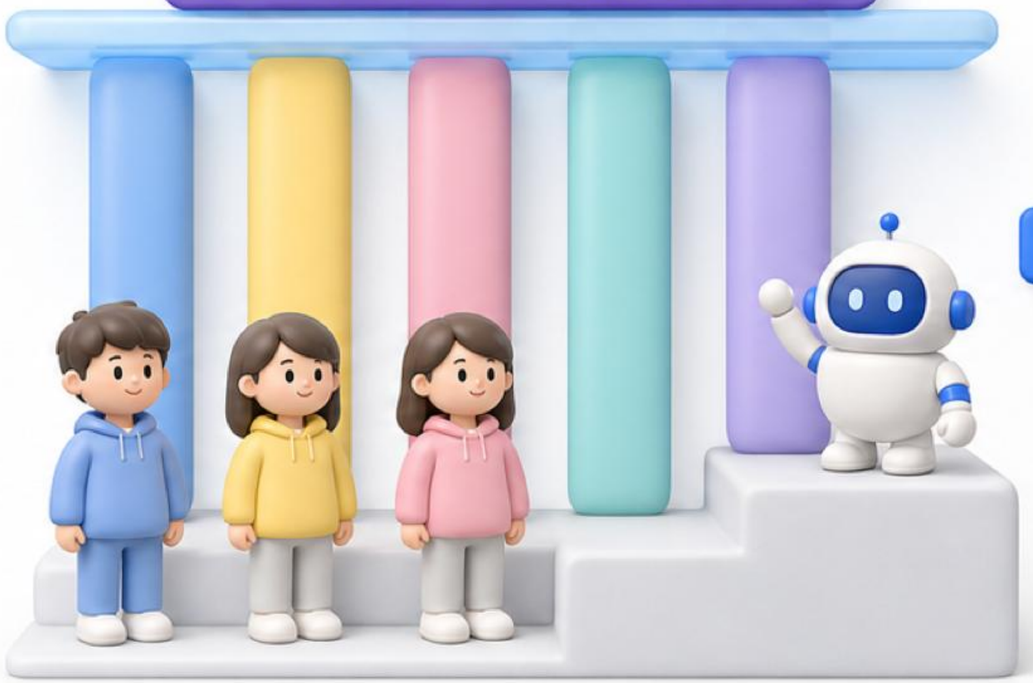
ヒントン：AGIは技術用語ではなく
『重要だが明確に定義されていない概念』

学び：方向はかなり当たる。時期と形はよく外れる

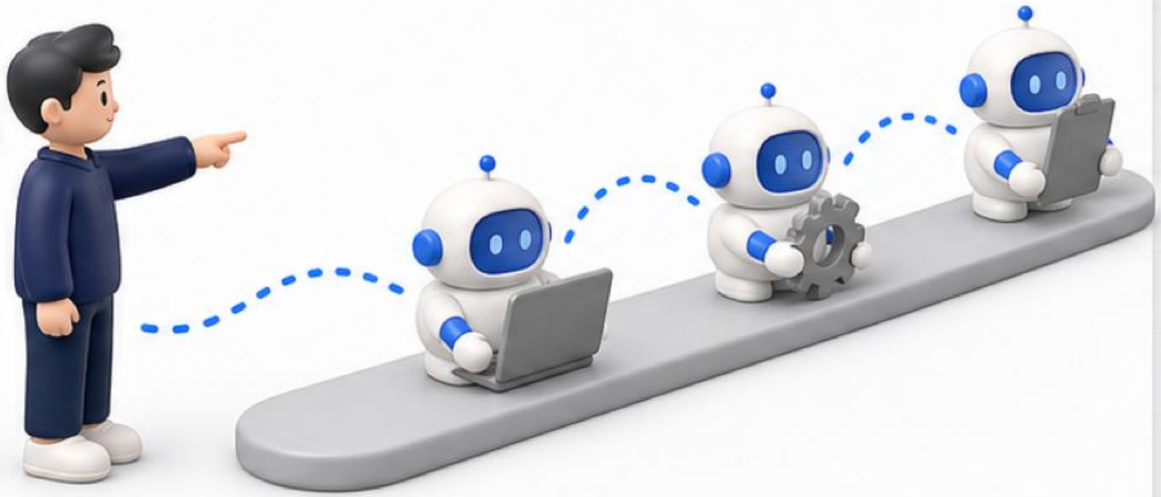
主要ベンチマークは飽和しつつある

かつての指標: 大学入試、資格試験、競技プログラミング
— 一部評価では人間の上位層を超えた

飽和した指標では、もう能力差を測れない



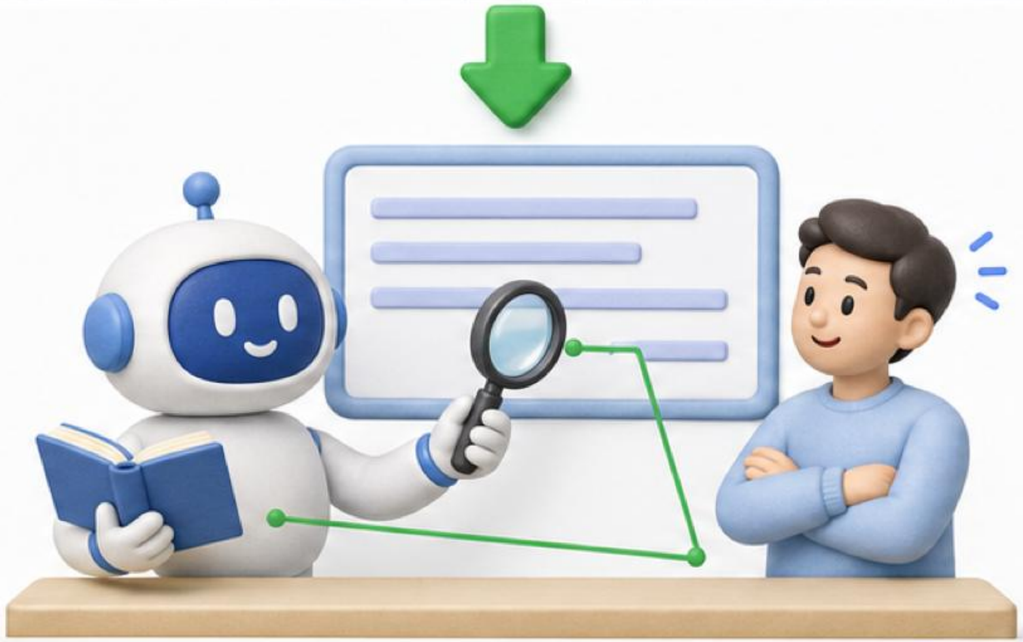
今の関心: 実務の課題を、どこまで自律的に、
どれだけ長く進められるか



評価は『点数』から『任せられる時間と範囲』へ

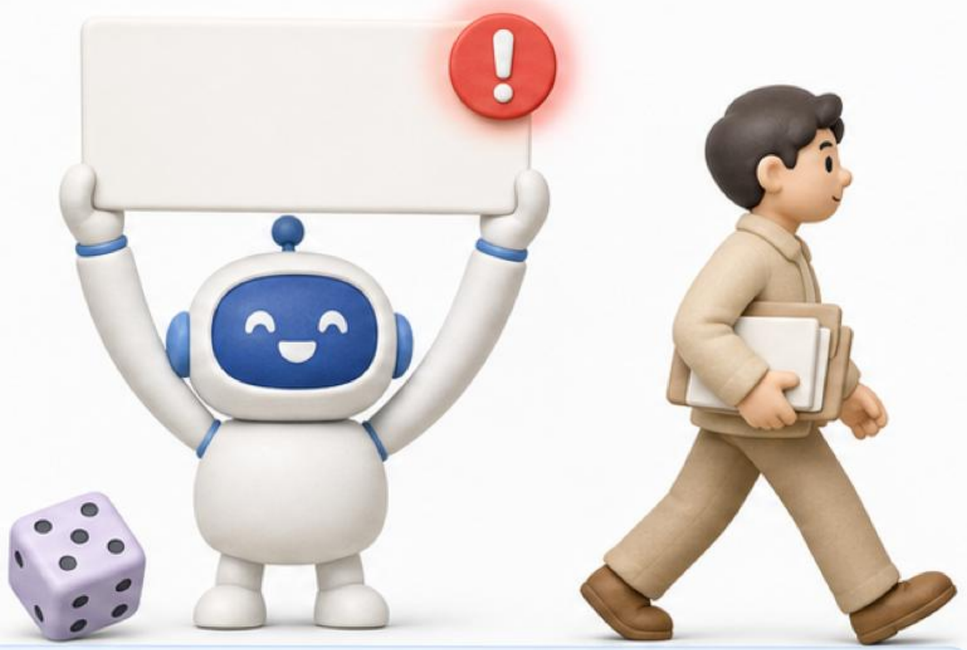
ハルシネーションは減った、しかしゼロではない

ハルシネーション ＝ もっともらしい誤りを、自信を持って出す現象



一部モデル・一部の事実性評価では、
発生率は大きく下がった

効いた手段：検索や資料の参照、
根拠の提示、推論時に長く考えること



それでも原理上ゼロにはならない ―
確率的に語を選ぶ仕組みだから

効いた対策：『分からない』と言える設計 ／
出典を出させる ／ 人が突き合わせる

コンテキストウィンドウの拡大

コンテキストウィンドウ = 一度に読み込める情報量の上限



2023年ごろの主力モデル: 数千〜数万トークン



例: Claude 5 ファミリー（Fable 5 ／ Opus 5 ／ Sonnet 5）
はいずれも公称100万トークン

2026年、一部の主力モデルの公称値: 100万トークン規模





ただし『入る』ことと『同じ精度で扱える』ことは別問題

料金の民主化 — 学生でも最前線に触れられる



2023年、最前線のモデルは高価で、試すこと自体にコストがかかった



2026年：同等以上の性能が、当時の何分の一かの単価で使える



学生向けの無料・優待プランが整った
(GitHub Student Developer Pack など)

※ 価格・提供条件は2026年7月時点。
頻繁に変わります

今日、皆さん全員が有料相当の環境を持っています

それでも残る限界

1



知識には時間の境界がある

学習時点より後のことは知らない
（検索で補えるが、補えたかは別問題）

2



出力は確率的に選ばれる

同じ問いに同じ答えが返る保証はない

3



もっともらしい誤りを、自信ありげに出す

誤りに「自信がない」という印がつかない

これらは性能向上では消えない。原理に由来する限界だから

品質保証の概念は不変

出力の正しさを保証する責任は、常に使う人間の側にある







ハルシネーションが減っても、
検証責任は減らない

1



実行して確かめる

2



差分を読む

3



テストを書かせる

4



出典に突き合わせる

作る速さと、確かめる速さは別物 — 速く作れる時代ほど、確かめる設計が要る

全員が「管理者」として働くようになる



指示する — 目的と前提を言葉にし、必要な情報を渡す



判断する — 誤りがないか、方法が妥当か、目的に沿っているかを見る



責任を取る — 問題を認識し、原因を調べ、再発を防ぐ

知識へのアクセスコストはほぼゼロになった。残るのは、何を問い、何を選び、誰が責を負うか

AI利用の作法



① 入れてはいけないもの

個人情報、未公開の研究データ、機密情報。
無料サービスでは入力が学習に使われる場合がある



② 事実は自分で確かめる

事実性が争点になる用途では、必ず出典に当たる



③ 引用と出典を明示する

誰の成果かを曖昧にしない。
学術倫理はAIでも変わらない



④ 責任は利用者にある

「AIが言った」は理由にならない

※ 所属機関・学会・投稿先の規定が最優先。必ず確認する

生成AIはゲームチェンジャー『である』



午後は、皆さんが『作る側』に回ります



12:30 — AI駆動開発の体験: ブラウザだけで、動くアプリを作る

14:15 — 開発環境の構築: uv ／ VS Code + Copilot ／ Git & GitHub

16:30 — 開発実践: 仕様書を書く → 生成させる → 検証する → コミットする

明日 — 自作アプリを GitHub に push して、発表する

原理は分かった。次は、動かします