

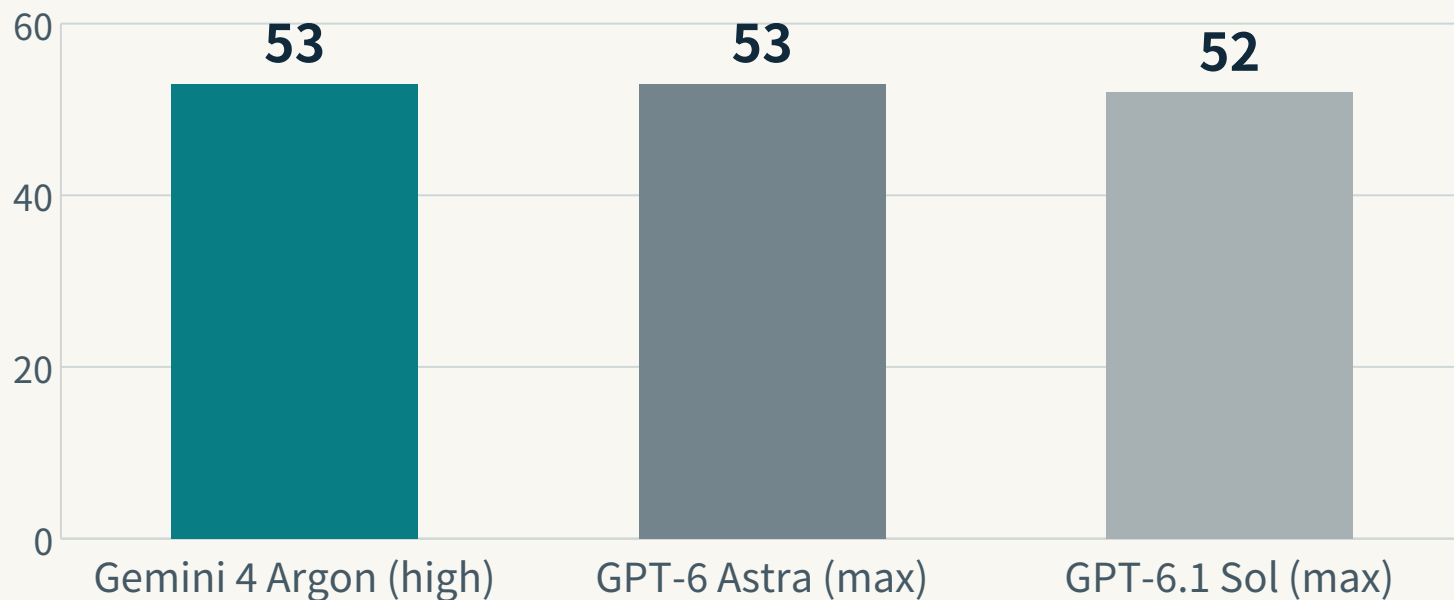
Gemini 4 Argon と パーソナル AI の現在地

AI の性能と、
仕事を任せられる条件を読む

TBS CROSS DIG 「AI QUEST」
今井翔太氏の解説と公式資料をもとに
2026 年 10 月 2 日収録

Google の新モデルは、総合評価で先頭集団に並ぶ

Artificial Analysis Intelligence Index



53 は「総合指標」

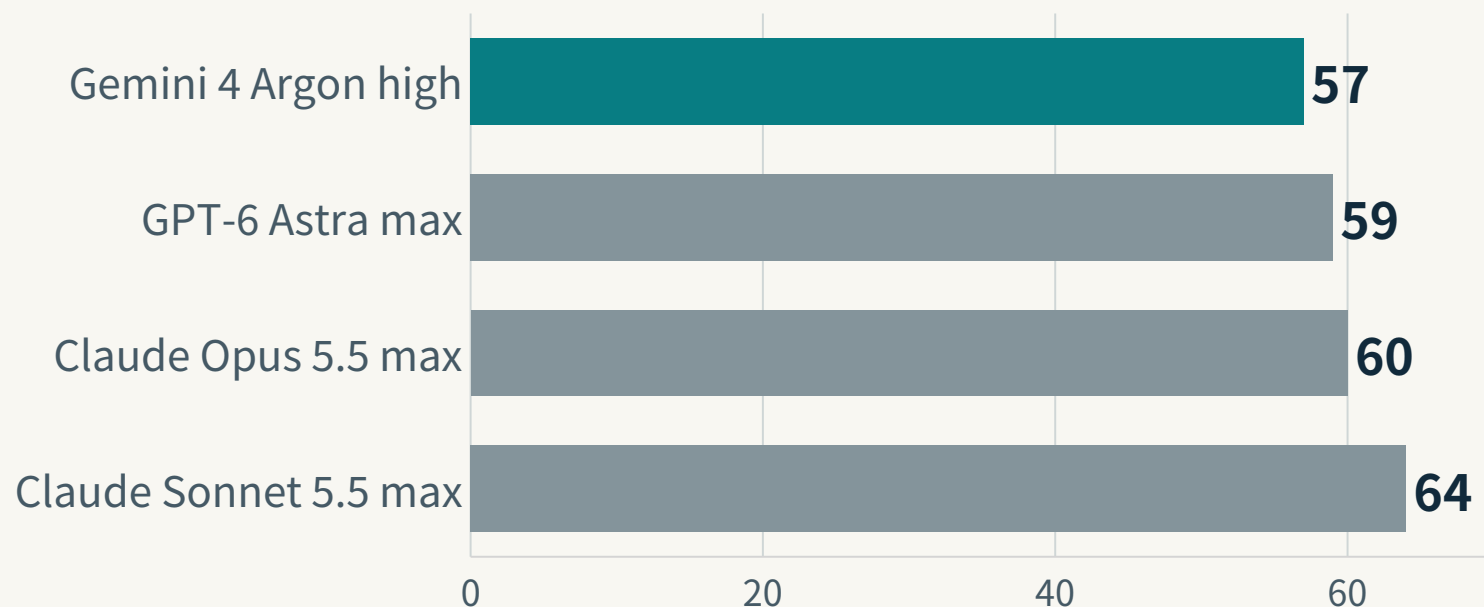
複数の試験をまとめた値。
普段の仕事で 53% 正解する、という意味ではない。

同点でも、得意な仕事まで
同じとは限らない

確認済み事実 / 評価時点と推論設定をそろえた比較

コード操作の課題では、モデル間の得意差が残る

Terminal Bench 4： 端末を使って作業する課題の成功率（％）



**この課題で Argon は
首位ではない**

「総合評価で並ぶ」と
「すべての仕事で勝つ」は
分けて読む必要がある。

確認済み事実 / 同じ試験でも得点の順序は総合指標と異なる

幻覚率 15% でも、正答率は 50%

AI がもっともらしく誤る「幻覚」と、正しく答える力を分けて見る

15%

幻覚率

不正解・部分正解・未回答のうち、
不正解が占める割合

50%

正答率

評価の全質問のうち、
正しく答えた割合

確認済み事実 / 分母が違うため、「 $100 - 15 = 85\%$ 正しい」とは読めない

動画理解も進むが、利用は段階提供

91.7%

動画理解の試験 LVBench

動画を見て質問に答える評価。
Google が公表した自社測定値。
テキスト以外の能力も持つ。

9 月 30 日時点の提供範囲

まず、サイバー防御に関わる
信頼済みの組織へ限定提供。

有料 API 顧客・ Google AI Ultra は
今後の提供対象として案内。

確認済み事実 / 試験の高得点と、誰でもすぐ使えることは別の情報

Google の競争力は、日常の仕事との接点にもある

今井氏の見立て

日常的な文章のやり取りに強い AI と、Google が持つサービスや利用者との接点の組み合わせを評価する。

会議資料を読む

メールの下書きを作る

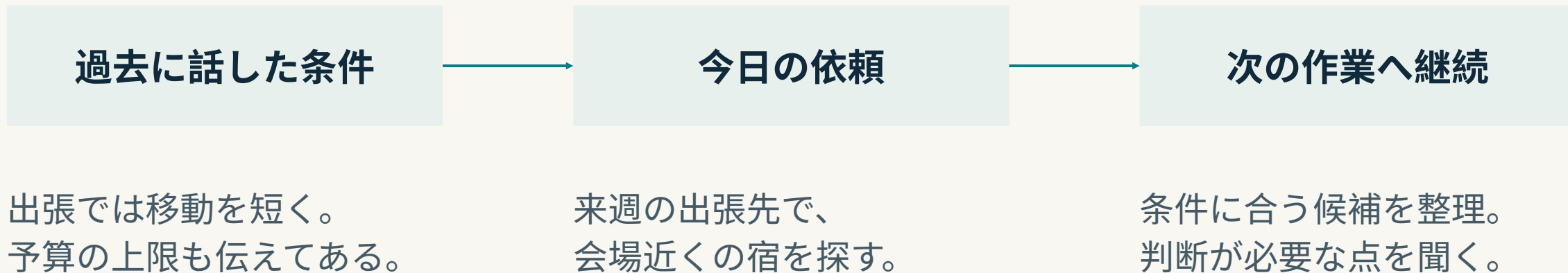
普段の仕事で使う

仮想例：回答の質に加え、必要な資料にたどり着く手間や、いつもの仕事の中で使えることが便利さにつながる。

見立て・仮想例 ／ 下段は使い方の説明であり、Argon の一般向け機能の保証ではない

パーソナル AI は、前の話を踏まえて仕事を続ける

会話のたびに説明し直す負担を減らし、背景や好みを次の作業に生かす。
個人向け AI の価値は、この「文脈の継続」にある。



公式に説明された方向性 ／ 下段の出張手配は、使い方を示す仮想例

dots と Meta Muse は、専用環境で作業を続ける

OpenAI dots

2026 年 9 月 29 日発表

専用のクラウド PC とブラウザを持つ。
連携したアプリを使って作業を進め、
判断が必要になると本人に聞く。
会話間の関連する記憶も活用する。

Meta Muse

2026 年 9 月 8 日発表

専用のクラウド環境とブラウザを持つ。
アプリを閉じた後も仕事を続ける。
好みを覚え、購入やメール送信の前に
本人の承認を求める設計。

各社の公表内容 / クラウド環境＝インターネット上にある作業用コンピュータ

似たサービスの同時登場で、利用が分散する可能性

確認できる事実： Meta Muse と dots は同じ月に登場

9 月 8 日 Meta Muse



9 月 29 日 dots

今井氏の見立て

似た個人 AI が同時に競争すると、利用者が複数に分かれ、一つのサービスが広く定着しにくくなる可能性がある。

見立て / 発表の近さは事実。利用分散が普及を妨げるという因果は予測

予約や購入は、一度のミスが大きな負担になる

「減点型」の仕事：一度の失敗が、ほかの成功の価値を大きく下げる仕事

文章の下書き

表現が少し違ってても、
読んで直してから使える。
やり直しの負担を抑えやすい。

予約日・購入数量の確定

日付や数量を誤ると、
予定の変更や追加費用につながる。
後から簡単に戻せない場合がある。

任せる範囲は、正確さと「失敗したときの影響」を一緒に見て決める

見立て・仮想例 ／ 「何回成功したか」に加え、失敗の重大さも考える

外部サービスの許可が、AI に任せられる範囲を決める



報道例： Amazon が Meta Muse の代理購入を遮断

GeekWire の取材によると、Amazon 側は事前合意や自己識別、顧客情報の安全性を問題に挙げた。

AI が操作できる能力と、相手が代理操作を認めることは、別々の条件

報道・公式説明 / すべての外部 AI や、すべてのサービスが遮断されるという話ではない

任せる範囲と確認点を決めると、AIを試しやすい

例：出張手配を、準備する作業と確定する判断に分ける

AIに任せる準備

条件に合う候補を調べる
料金や場所を比較する
旅程の下書きを作る

人が確認して確定する判断

日付・金額・相手先は合っているか
変更・取消の条件を受け入れるか
この内容で予約や支払いを進めるか

選ぶ基準は、性能に加えて「確認・修正を含めても、役に立つか」

提案・仮想例 ／ まずは低リスクで、結果を確かめやすい仕事から試す