

State of AI, April 2026: Three Decisions to Make Carefully

Juan Beltran | Revised 2 October 2026

An archived April 2026 executive memo on provider choice, longer delegated work and capacity commitments, with evidence and reversal conditions for each.

Archived executive memo, April 2026. Revised in October 2026 to focus on three decisions; use current service evidence for present procurement.

A model announcement can change what is worth testing without changing what is worth buying. That distinction is easy to lose during a busy release cycle. The demonstration is immediate; the consequences of a supplier commitment, a redesigned process or a new infrastructure budget arrive later.

For an executive, the useful output of a state-of-AI review is a small number of decisions with conditions attached. Which work should we re-evaluate? What evidence would justify changing its operating model? Which commitments would be expensive to reverse if our assumptions were wrong?

This memo concentrates on three decisions: how to choose providers, how to test longer delegated work, and how much capacity to commit. They belong together. A capable model can still be the wrong service for a workflow, and an attractive workflow can still be uneconomic under its delivery conditions.

Decision one: select a service for the work, then decide how much flexibility to buy

The attraction of a leaderboard is that it gives procurement a clear winner. The difficulty is that the company does not buy the benchmark task. It buys a service that must handle its documents, access rules, languages, exceptions and response times.

Start with a representative work sample. Include routine cases, awkward inputs, incomplete information and situations in which declining to answer is the correct result. Establish what an acceptable output means before comparing providers. Otherwise, the most fluent response can become the implicit quality standard.

For a hypothetical engineering-support workflow, correctness might require a statement to match an approved specification and identify its source. An unsupported but plausible answer fails that requirement. A slower answer that exposes a missing specification may be more useful.

Compare the whole delivered task: output quality, source traceability, latency, integration effort, service availability, data handling and total cost. A model's public benchmark score can inform the shortlist. It cannot replace that comparison.

Provider flexibility also has a price. Supporting several models can mean more evaluation, version management and operational complexity. A simple application with a reliable supplier may not justify a routing layer. A consequential workflow with material continuity exposure may justify a tested alternative and an exportable evaluation set.

The conditional decision is therefore specific: choose the service that meets the workflow's requirements now, and purchase additional flexibility where the cost of dependence warrants it. Keep the evidence needed to repeat the comparison. A generic instruction to become “model agnostic” leaves both the expense and the reason for paying it undefined.

Decision two: delegate a longer task only when its checkpoints can detect failure

Longer task capability matters because more work may become possible between human interventions. But a task lasting several hours for a person is not necessarily an agent running safely for several hours in a business system.

[METR's January 2026 time-horizon update](#) estimates task horizons using human reference durations and specified success levels on its task suite. Those measures concern evaluated tasks, with sensitivity to task selection and methodology. They do not establish a production runtime limit, a universal rate of enterprise progress or a deadline for removing human supervision.

The operating question is what happens between the start and the final answer. Does the system retain the right goal? Can it recognise missing information? Does it know which actions require permission? Can an operator reconstruct the sequence when something goes wrong?

Consider a hypothetical workflow for assembling a tender evidence pack. The system may retrieve approved documents, identify unanswered requirements and prepare a draft matrix. That is different from authorising a commercial promise or submitting the tender. A longer run can be useful while those consequential decisions remain with their owners.

Choose checkpoints according to failure consequences. A source retrieval failure needs a visible exception. An unsupported certification claim needs to block the affected output. An external submission needs an authorised release step. Checking only whether the final document looks complete would miss all three distinctions.

The evidence for expanding delegation should include failures and recovery, not just successful demonstrations. Measure omission rates, unsupported claims, review effort and the effect on the complete process. Include people who understand the work in the evaluation, and test whether they can operate the fallback.

If the system saves drafting time but creates a larger verification burden, expand neither the task nor the budget on the strength of the draft alone. If it improves the process under representative conditions, extend one boundary and evaluate again. Progress can be staged without making every stage a permanent pilot.

Decision three: commit capacity against demand you can explain

Infrastructure headlines can make a modest company-level decision feel like a wager on the future of computing. It rarely needs to be. The immediate budget should connect expected workload to a delivered service, with enough uncertainty visible to change course.

Estimate ordinary demand, peaks and failure recovery. Include long contexts, retries, evaluation runs and review costs. Clarify which workloads require low latency and which can wait. These distinctions affect capacity more directly than an undifferentiated forecast of “AI usage”.

An illustrative service might prepare internal summaries during the working day and process large document collections overnight. Buying the same response-time commitment for both could waste money. Equally, a

low-cost service that cannot meet a customer-facing deadline could undermine the process it was meant to improve.

Use a bounded commitment where demand is uncertain. A larger commitment becomes reasonable when usage is persistent, service requirements are stable and the economics survive less favourable assumptions. Reserved capacity, internal infrastructure and provider contracts should be compared on a common horizon, including implementation and exit costs.

There may be sound reasons to pay more: continuity, controlled data handling, support or predictable response times. Record which requirement the extra cost serves. “Strategic” should explain a trade-off, not exempt it from one.

What would change the recommendation?

A useful executive memo names its reversal conditions. A provider decision changes when another service meets the requirements at a better total cost, or when the current supplier's reliability, data terms or support become unacceptable. A delegation decision changes when evaluation reveals an undetected consequential error, or when the process demonstrates that an additional action can be controlled. A capacity decision changes when sustained demand differs materially from the assumption behind the commitment.

Assign someone to each evidence stream and a date to the next decision. The dates are management commitments to review the evidence, not predictions that technology or adoption will be ready on schedule.

That is the enduring value of an April 2026 review. Keep learning from capability changes while making company decisions at the scale of the work. The [AI Opportunity Map](#) can structure the workflow question; the [Build, Buy or Partner tool](#) can help make the sourcing trade-off explicit.

[Download the archived April decision memo.](#)

Reading worksheet

Use these questions with the essay. Record your own evidence and the next decision; the examples in the essay do not supply a result for your organization.

1. Which business decision is this archived memo helping us make now?

2. Which assumptions need a fresh check before we commit?

3. What representative workflow test would change our view?

4. Which commitment can remain reversible while we learn?

Sources and scope

Revised executive interpretation of the archived period, using a dated METR source and newly introduced illustrative workflows. It is not a current market survey, leaderboard or empirical business study.

The source concerns its evaluated task suite. The tender and service examples have no claimed results. Current provider, capacity and legal decisions require current evidence.

METR. [Time Horizon 1.1](#). Published 2026-01-29. Task-horizon measurement using human reference durations, with scope and methodology limits; not enterprise runtime or a deployment deadline.

[Read the current article](#)