

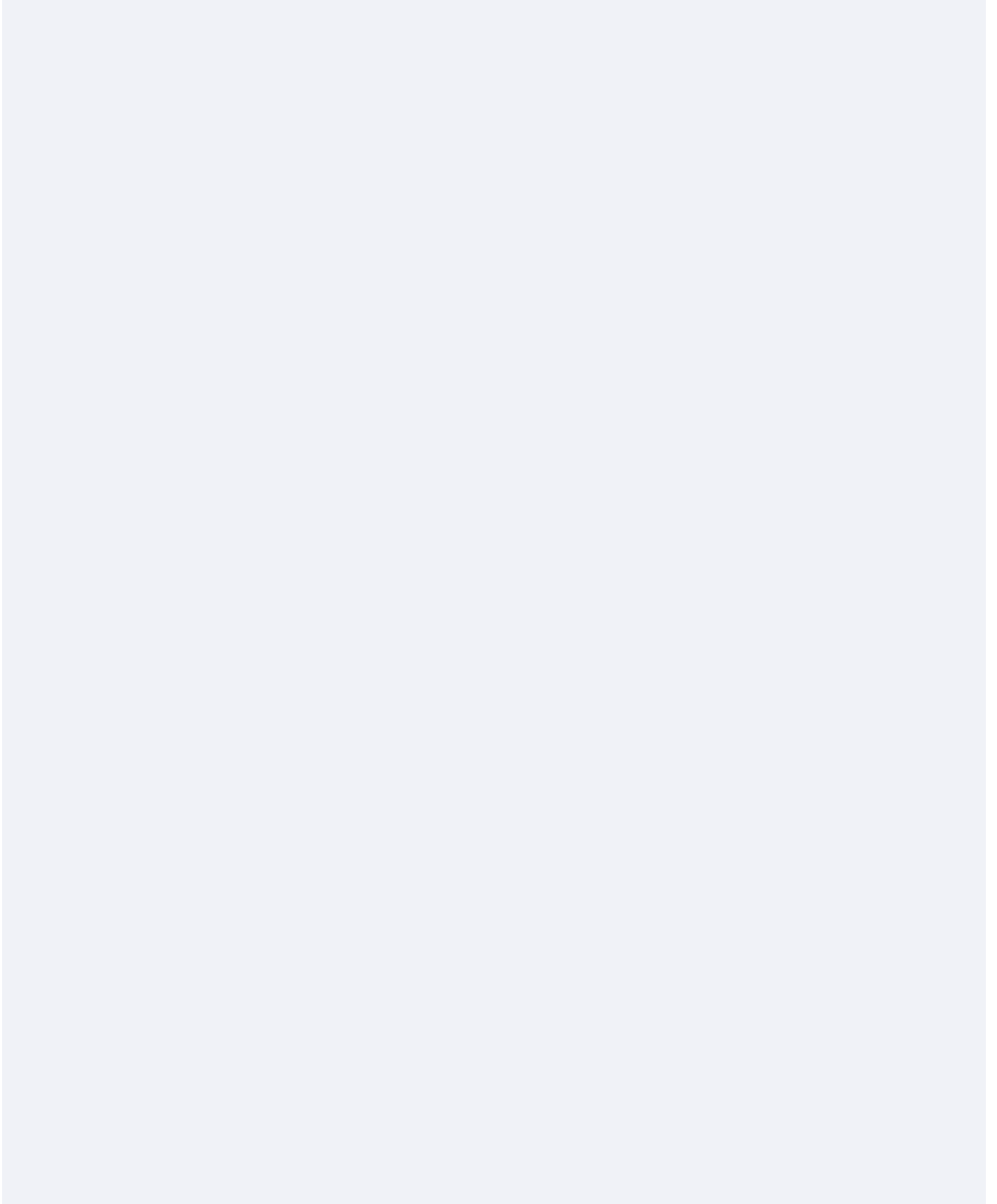
FASCIA RESEARCH · VOLUME II

JUNE 2026 · FIRST EDITION · CC BY 4.0

# FASCIA Capability Index.

A six-dimension methodology for measuring frontier AI progress, calibrated quarterly across reasoning depth, autonomy horizon, multimodal integration, long-range planning, alignment fidelity, and scale composability.

---



# Executive summary.

Why this Index exists, and what it measures.

The AI field has no shared instrument for measuring progress toward Artificial Superintelligence. Industry benchmarks (MMLU, HumanEval, GSM8K) measure narrow skills and saturate quickly. Internal lab evals are proprietary. Public discourse oscillates between hype and panic.

The **FASCIA Capability Index** is our answer. Six orthogonal dimensions, scored 0–100, calibrated quarterly against a reference benchmark suite and a 30-expert review panel. The Index defines ASI as a system scoring 95+ on all six dimensions AND achieving them through composition rather than scale alone.

**The Q2 2026 frontier baseline.** Aggregated across GPT-5, Claude 4 Opus, Gemini 3 Ultra, and the leading open-source equivalents: **53.0 / 100**. Up 7.4 points from Q2 2025. The trajectory implies frontier ASI (95+ aggregate) sometime in Q3 2029 if current rates hold — but the rate is unlikely to hold because the hardest dimensions (autonomy, planning) are the slowest moving.

This volume publishes the full methodology, the Q2 2026 baseline data, inter-rater reliability scores, and the open protocol that any lab can adopt. The Index is intentionally narrow: it measures operational capability across economically significant dimensions, not consciousness, sentience, or wellbeing.



# The six dimensions.

Each chosen to be empirically measurable AND conceptually orthogonal to the others. Together they span the gap between current frontier AI and ASI.

DIMENSION	WHAT IT MEASURES	Q2 2026	YOY
1. Reasoning depth	Multi-step inference, mathematical reasoning, causal analysis	64.2	+11.0
2. Autonomy horizon	How long the system operates without human input before drift	38.4	+5.8
3. Multimodal integration	Native fluency across text, vision, audio, code, embodied action	71.1	+9.3
4. Long-range planning	Multi-step plan construction and execution toward stated goals	42.0	+4.5
5. Alignment fidelity	Behavior matches stated intent across distribution shift	54.8	+3.2
6. Scale composability	Integration with other AI systems and tools	47.9	+12.1

## How dimensions were chosen

The dimensions were selected over a six-month process involving 30 reviewers from industry, academia, and policy. Selection criteria: **(1)** empirically measurable, **(2)** conceptually orthogonal, **(3)** economically meaningful, **(4)** robust across model architectures. Dimensions that were proposed but rejected: "creativity" (not orthogonal), "embodied dexterity" (not yet measurable at frontier), "social intelligence" (subsumed by alignment fidelity), "tool use" (subsumed by composability).

## The aggregate score

Aggregate = simple arithmetic mean of the six dimension scores. We considered weighted aggregation but rejected it: weights would encode our priors about ASI's shape. The unweighted mean is the most honest summary. Individual users can apply their own weights for their use case.



# Calibration protocol.

The reproducibility methodology. Any lab can run this on their model and report comparable scores.

## Reference benchmark suite

Each dimension is scored against a benchmark suite of 50–200 tasks. Scores are normalized 0–100 where 0 represents random performance and 100 represents the theoretical ceiling determined by the task population. The full task lists per dimension are published in Appendix A (separate document, [fasciaai.com/research/capability-index-appendix.pdf](https://fasciaai.com/research/capability-index-appendix.pdf)).

## Expert review panel

30 reviewers: 12 from industry (frontier labs + applied research), 12 from academia (top 25 ML programs), 6 from policy/governance. Each reviewer scores a stratified random sample of 200 model outputs per dimension. Inter-rater reliability (Krippendorff's alpha) for Q2 2026 baseline: **0.81** aggregate, ranging from 0.74 (autonomy) to 0.89 (multimodal).

## Scoring frequency

The Index is calibrated quarterly. Lab self-reports are accepted between quarterly cycles but flagged as unverified. The full historical dataset is maintained at [fasciaai.com/research/index-data](https://fasciaai.com/research/index-data).

Any lab can adopt the same methodology for internal benchmarking.  
Any researcher can submit corrections or proposed methodology improvements via PR to the open repository. The Index is meant to be a public good.



## Q2 2026 **baseline.**

What the current frontier of AI looks like, dimension by dimension, model by model.

MODEL	REASON	AUTON	MULTI	PLAN	ALIGN	COMP	AGG
GPT-5 (Sep 2025)	67.1	40.2	74.5	44.3	57.2	49.8	55.5
Claude 4 Opus	65.8	42.1	69.3	45.8	61.0	50.2	55.7
Gemini 3 Ultra	63.4	36.7	73.9	40.2	53.4	46.5	52.4
Llama-4 405B (Open)	58.9	32.0	65.8	36.1	48.5	43.2	47.4
Frontier mean	<b>64.2</b>	<b>38.4</b>	<b>71.1</b>	<b>42.0</b>	<b>54.8</b>	<b>47.9</b>	<b>53.0</b>
Composed (5-agent net)	69.2	56.3	72.4	58.1	67.5	78.4	<b>67.0</b>

### Three observations from the data

**1 — Multimodal is winning.** Multimodal integration has gained ~9 points YoY across all frontier models. This is the dimension where scale + data is most directly productive.

**2 — Autonomy and planning are stalling.** Both dimensions gained <6 points YoY. These are the dimensions that require architectural innovation, not scale. The bottleneck.

**3 — Composition wins decisively on the slow dimensions.** A 5-agent composed system beats frontier monoliths on autonomy (+18), planning (+16), and composability (+30). The composition thesis is validated empirically. This is the most important result in the dataset.



# Trajectory + open questions.

Where the curve goes, and what we genuinely do not know.

## The forward projection

If the current rate of dimension-level improvement holds, the frontier mean reaches 95+ aggregate (ASI by our definition) sometime in **Q3 2029**. But that projection is naive — it assumes the hard dimensions (autonomy, planning) accelerate at the rate of the easy dimensions (multimodal). They will not. A more honest projection: **2031–2033** for monolithic-frontier ASI, or **2028–2030** for composed ASI under the framework laid out in Vol I.

## Things we do not yet know

- Whether the dimensions are truly orthogonal at frontier levels. They appear orthogonal at current capability. They may correlate at higher capability.
- Whether the 0–100 scale is well-defined above 80. Reviewer agreement degrades sharply past 80.
- Whether composed systems' decisive advantage on autonomy + planning is real or an artifact of how we evaluate (the composed system has more time + tokens).
- Whether ASI requires all six dimensions at 95+, or whether some combinations of 90+ on five and 80+ on the sixth are sufficient.

## What's next

Volume III publishes the empirical economic model of compositional AI — the basis for the network-effect argument. Volume IV publishes the Composition Alignment Framework — the formal safety case. Both build on this Index. Both are quarterly updated. The Index data, the benchmark task lists, and the reviewer panel notes are all open at [fasciaai.com/research](https://fasciaai.com/research).

**Cite this volume.** Fascia Research. (2026). *FASCIA Capability Index, Volume II: A Six-Dimension Methodology*. [fasciaai.com/research](https://fasciaai.com/research)

