ELASTICITY INSTITUTE

23 September 2026

How to Measure RSI

Eight data asks for frontier AI companies

Cheryl Wu*, Arjun Ramani*, Basil Halperin*
Lukas Althoff, Brian Jabarian, Andrew Koh, Phil Trammell
*Lead authors

INTRODUCTION

Frontier AI companies have indicated willingness to share more information about recursive self-improvement (RSI). This note describes eight data elements that would be highly informative about the pace of RSI, ordered by priority. The requests draw directly from the economic models of RSI in Cunningham et al. (2026).

CAPABILITIES RSI would accelerate capability growth, and the acceleration would show up in internal models before released ones.

1 Internal-model benchmark scores. Results on a diverse set of benchmarks for the latest internally deployed models, including unreleased models. Ideally a comprehensive data, including expenditure curves. Second best is an aggregate such as Epoch's Capabilities Index (ECI). Fig. 1

EFFICIENCY AI-assisted R&D would be reflected in efficiency gains. This is informative over and above capabilities progress because capabilities progress may simply reflect increased inputs.

2 Overall expenditure efficiency. For each major model recipe, capability as a function of total training expenditure. This may not be easy to report, but it is the clearest theoretical representation of progress in R&D. Fig. 2

3 Pre-training compute efficiency. Pretraining loss against pre-training compute for each production run and the small validation experiments that precede it. Fig. 3

4 Post-training compute efficiency. Capability against post-training compute, measured on ECI and on unsaturated AI R&D benchmarks (e.g. test-time scaling curves for real-world AI R&D tasks, especially algorithm optimization). Fig. 3

INPUTS TO AI RESEARCH Gains driven by more human labor are not evidence of RSI; gains driven by more inference compute are. Historical and ongoing series of AI R&D inputs help identify the *source* of progress.

5 Inference compute. Internal model usage for AI R&D by team, in quantity (e.g. tokens) and in expenditure.

6 Experimental compute. Internal usage by team, in quantity (e.g. H100 hours) and in expenditure.

7 Headcount and compensation. Employees and total compensation by team, function, and seniority.

8 How AI generates efficiency gains. Surveys of researcher uplift, and a record of autonomous discoveries.

Figure 1. Hypothetical model capabilities over time. Internal models leave the historical trend first.

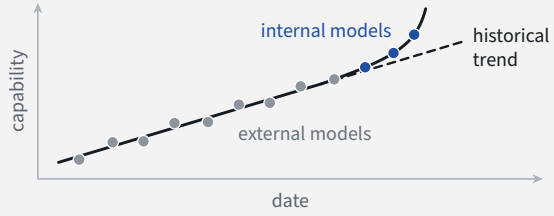


Figure 2. Hypothetical scaling curves of capability against training spend. Arrows show spend saved at equal capability.

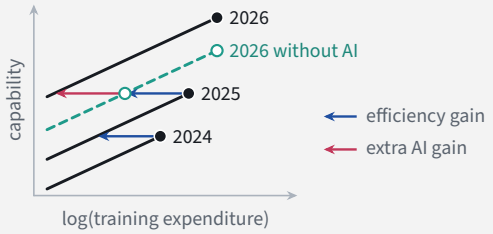
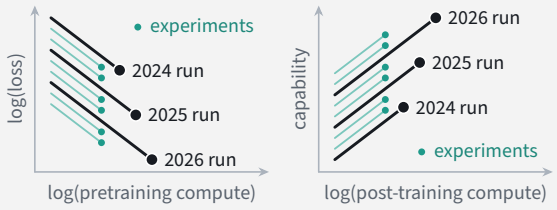


Figure 3. Hypothetical pre-training loss (left) and post-training capabilities (right) against compute. Smaller experiments (teal) precede each production run and run ahead of the latest one.



SCORECARD

OpenAI	2 / 8
Anthropic	1.5 / 8
Google DeepMind	0.5 / 8

We score company disclosures against the eight asks: 1 for a full release, 0.5 for a partial release, 0 for no qualifying data found. We score recently released frontier models from OpenAI, Anthropic, and Google DeepMind in 2026 using the sources below.

released · partial · no credit

Scorecard sources (2026). OpenAI: Research acceleration: The view inside OpenAI; On the Navier–Stokes Millennium Prize Problem (September 8). **Anthropic:** Risk Report: August 2026 (§§1.4, 3.4–3.5); When AI builds itself; Measurements for understanding the pace of AI development inside frontier labs. **Google DeepMind (GDM):** Towards Autonomous Mathematics Research (§2.1, Fig. 2; internal Deep Think, January 2026).