

Anthropic and Accenture each expect to invest \$1bn in embedded evaluation of frontier AI

Source copy kept by wheresthe.ai for the story "Anthropic and Accenture each expect to invest \$1bn in embedded evaluation of frontier AI" (<https://wheresthe.ai/d/anthropic-accenture-embedded-evaluation-2026-09/>).

Source: <https://www.anthropic.com/news/accenture-embedded-evaluation>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 47cec97aaaa6da2ee3515f5204f47dcc9d9ff28aa9ac0d5a80fddfe70df5c1a0

Documents the source offers: <https://www-cdn.anthropic.com/files/4zrzovbb/website/0a58d567024a8b448ff15158ebc3625328dfcc1f.pdf>

Announcements Partnering with Accenture on embedded evaluation Sep 18, 2026

We're partnering with Accenture on independent evaluation of frontier AI. This is an important step toward the commitment, made in our CEO's essay "We Must Pace the Frontier," to embed evaluators within Anthropic. The partnership will be led by Faculty, Accenture's specialist AI business, and will include evaluating and red-teaming models, conducting alignment assessments, and testing model safeguards. Accenture helps businesses and governments deploy AI across many industries. Their understanding of how enterprises use AI in practice informs their safety approach, and they will bring that perspective to evaluating our models. Anthropic and Accenture each expect to invest at least \$1 billion in building capacity in this area over the next five years. Embedded evaluation is new, and many of the details about how it will operate are still being worked out. Unlike today's external evaluators, embedded evaluators will work inside AI companies, with access comparable to an employee's. That access allows them to watch models take shape in training, follow the decisions that govern how those models are built and deployed, and speak directly to employees. From this vantage point, embedded evaluators can assess how a company operates, verify that it is keeping its safety commitments, and identify blind spots. They can also report incidents and give the public a more informed account of benefits and risks.

To be clear, independent embedded evaluators do not reduce our accountability, but help to make it more verifiable. The safety of our models remains our responsibility.

There are, as yet, no standards for what information embedded evaluators should have access to, or how they should report what they find. There is also no settled system for funding independent evaluation. Long-term, we think funding should come from pooled or government sources, as we called for in our Advanced AI Framework in June. As neither exists today, we plan to work with different evaluators under different funding arrangements.

Given the importance and urgency of this work, Anthropic will fund Accenture's work directly. We are also in dialogue with METR and other nonprofit evaluators to pilot elements of embedded evaluation using their own funding. Ultimately, we believe frontier AI needs an ecosystem of evaluators operating with shared standards.

We expect frontier labs to work with several organizations at once. Our partnership is non-exclusive; Anthropic will work with other evaluators to be announced in the coming weeks, and Accenture will work with other AI developers in similar capacities.

We'll continue to train and release frontier models, and we want independent evaluators working alongside us as we do. We're sharing these early efforts now so people and other AI developers can see our process. We expect our approach to evolve as the field matures, and we'll share more as our work begins and as we bring on additional evaluators.

Related content

An update on independent evaluators To evaluate the safety of our systems, Anthropic is giving employee-like access to a team from METR, a nonprofit that evaluates frontier AI models, and we're in early discussions with several other organizations, including Hugging Face. Read more Preparing for the 2026 US midterms Earlier this year, we shared an update on our election safeguards. Here, we outline how we've continued to strengthen these safeguards ahead of this year's US midterms. Read more Anthropic opens Singapore office and signs MOU with the Government of Singapore Anthropic has opened an office in Singapore and signed an MOU with the country's government to advance safe, responsible AI adoption. Read more