Glossary Card

DeployCos.com cheat sheet. Updated 11-08-2026. Sources at deploycos.com/sources/.

Term	Meaning
Agent	A software system in which a language model decides which actions to take - calling tools, querying systems, drafting outputs - toward a goal, usually across multiple steps, rather than producing a single response. In enterprise deployment, agents are typically constrained by approval gates and evaluated against a quality bar before they act autonomously. See How DeployCos Work.
Agent cannibalization	The risk that AI agents deployed into services businesses automate the billable work those businesses sell, shrinking their revenue. It is a particular concern for private equity sponsors whose portfolios include B2B services firms and who also back DeployCos. The usual response is a shift to outcome-based pricing. See Risks and Debates.
Agent harness	The software built around a model to make it useful in a specific workflow: retrieval over the customer's documents, tool definitions that let the model read or write systems of record, orchestration logic, memory, guardrails and logging. Jaipuria describes DeployCos building "an agent harness that wires the model into a company's data and other systems". The harness, not the model, is usually the durable asset in an engagement, which is why ownership of it at exit matters. See Evaluating a DeployCo.
AI deployment gap	The difference between how many organizations use AI somewhere and how many have scaled it across the enterprise. McKinsey's 2025 survey put the figures at 88% and 7% respectively. See Why DeployCos Exist.
AI-native rollup	An investment vehicle that acquires operating businesses outright in order to deploy AI from inside and capture the resulting margin improvement directly. 2026 examples: Long Lake's $6.3 billion take-private of Amex GBT and Sequence Holdings' $7.7 billion agreement to take The Baldwin Group private. Distinct from a DeployCo, which serves businesses it does not own. See Market and Economics.
Capability overhang	OpenAI's term for the gap between what its models can do and how typical users use them. OpenAI's January 2026 report found the typical power user (95th percentile) uses about 7x more reasoning capacity than the median user. See Why DeployCos Exist.
Claude-first	Ode with Anthropic's stated model policy: use Anthropic's models where possible and rival models where requirements demand it. Contrast with the OpenAI Deployment Company, whose announcement describes building on OpenAI models. See Ode with Anthropic.
Delta	Palantir's internal name for a forward deployed software engineer. Early Palantir business-development teams were named after NATO alphabet letters; "Delta" became the customer-facing engineering role and "Dev" the platform engineering role. See Palantir.
DeployCo	Short for "deployment company." Used two ways: (1) as OpenAI's short name for the OpenAI Deployment Company; (2) as a category label for AI services firms that embed forward deployed engineers inside enterprises to take AI from pilot to production, popularized by Tanay Jaipuria's October 2026 essay. A separate, unrelated company also operates under the name at deployco.co. See What Is a DeployCo?.
Deployment specialist / deployment strategist	Non-engineering roles on DeployCo teams that handle workflow discovery, change management and customer coordination alongside forward deployed engineers. OpenAI's Tomoro acquisition was described as bringing "Forward Deployed Engineers and Deployment Specialists".
Diagnostic	The short, scoped first phase of a DeployCo engagement in which engineers identify and rank candidate workflows. OpenAI's Deployment Company starts with a diagnostic and then focuses on a small number of priority workflows; independents recommend a paid diagnostic of two to four weeks. See How DeployCos Work.
Diffusion gap	Sequoia Capital partner Pat Grady's term for the lag between the creation of AI capability and its spread into enterprise use; capabilities "diffuse into the market far slower than they're created". See Why DeployCos Exist.

Evals (evaluations)	A test set of real inputs with expected or expert-graded outputs, run against an AI system before and after changes to measure quality. DeployCos "build evals as needed, and iterate on and deploy the agents once they clear some quality bar". The eval suite is the main evidence that a deployment is production-grade rather than a demo. See How DeployCos Work.
Forward deployed engineer (FDE)	A software engineer who works inside a customer's organization to build, integrate and ship production systems on a vendor's platform or models, and who is measured on whether the problem is solved. Created at Palantir; adopted by OpenAI, Anthropic, AWS, Google Cloud and many startups. Also forward deployed software engineer (FDSE). See Forward Deployed Engineers.
Fractional FDE	A forward deployed engineer engaged part-time - days per week or per month - on a project basis, bringing the DeployCo playbook to companies that cannot fund a full-time embedded team. An expected 2027 development rather than an established offering. See 2027 Outlook.
Frontier Alliance	OpenAI's multi-year partnership program, announced 02-23-2026, with Accenture, BCG, Capgemini and McKinsey to deploy its agent platform in enterprises, with OpenAI forward deployed engineers working alongside consulting teams. The OpenAI Deployment Company also works with Frontier Alliance partners.
Function-specific DeployCo	An AI deployment company focused on one business function across many industries, such as customer support (Sierra, Decagon) or procurement. See Types of DeployCos.
GDPval	OpenAI's September 2025 benchmark of model performance on real-world work products across 44 occupations in nine industries (1,320 tasks; 220 public), graded by experienced professionals. The top model matched or beat experts on close to half of tasks in the initial results. See Why DeployCos Exist.
Horizontal DeployCo	An AI deployment company that deploys AI across many functions and industries, usually with a shared internal platform. The lab-backed ventures, Distyl and Brain Co. are the usual examples. See Types of DeployCos.
Human-in-the-loop (HITL) / approval gate	A control in which a person must approve an agent's consequential actions - payments, customer communications, record changes - before they take effect, with autonomy expanded only as the agent demonstrates verified results within limits the customer sets. Standard in production DeployCo work; described explicitly by several firms. See How DeployCos Work.