

Introduction

NEW IN 2026

Harness engineering emerged through 2026 as a distinct discipline, formalized in surveys covering 100+ papers and 23 production systems. The core claim: at fixed model capability, agent-computer interface design materially changes benchmark outcomes.

The Equation:

Agent = Model + Harness

The harness treats the LLM as a frozen reasoning utility and moves the responsibility for safety, execution accuracy, multi-step orchestration, and adaptive memory into the surrounding infrastructure. A production harness is generally described as a three-layer system:

Layer	Responsibility
Information Layer	What data the agent can observe; what tools it may invoke at a given moment; vector storage, memory compilation, tool registries.
Control Layer	Planner / generator / evaluator roles; structured handoff artefacts; verification gates; state machines governing step sequencing.
Governance Layer	Permission boundaries, policy checks, approval gates, and the audit trail that makes autonomy observable.