AI training should follow the work people do

Assign role-specific exercises and measure output quality after review.

RESEARCH BRIEFING · WORKFORCE · 11 NOVEMBER 2026

Employers should train staff against specific tasks and train managers to assess the resulting work. Licence activation, attendance and prompt-writing confidence do not show whether an organisation can produce better accepted output.

Key points

• Map the task and its acceptance owner before choosing a course.

• Teach workers and supervisors different responsibilities.

• Assess performance on representative cases, including failures.

Reported benefits exceed reported revenue gains

The OECD’s late-2024 survey covered more than 5,000 SMEs in seven countries. Among generative-AI users, 65% reported improved employee performance and 26% reported increased revenue. The results are self-reported benefits, not measured percentage improvements or causal effects. [1]

A 2023 writing experiment with 453 professionals found a 40% reduction in task time and an 18% increase in assessed quality. It tested specified assignments, not an entire job or a company training programme. [2]

Train to a defined responsibility

For an analyst, use a briefing exercise: locate evidence, verify a numerical claim, challenge a generated interpretation and correct the final document. For a service employee, practise classifying a request and escalating an exception. For a developer, assess accepted changes rather than code volume.

Supervisors need a separate curriculum: defining acceptable quality, allocating review time, identifying systematic errors and deciding when to stop using a tool. A team can draft more material while creating a bottleneck for the person authorised to approve it.

Score the final work

Record an unassisted baseline, then compare assisted tasks of similar difficulty. Measure material errors, review minutes, completion time and the number of accepted outputs. Retain failed and abandoned cases.

METR’s early-2025 developer trial recorded a 19% slowdown despite access to AI. Its February 2026 update cautioned that follow-up selection problems obscured the current effect. The practical lesson is to test the work locally rather than teach a historical productivity percentage as a promise. [3][4]

A proposed 30/60/90-day sequence

By day 30, select roles, tasks and assessors; agree permitted data and establish the baseline. By day 60, run supervised exercises and investigate corrections. By day 90, review accepted-work quality and decide which uses can expand.

The schedule is an illustrative management plan, not a validated training dose. Publish the number of people assessed, tasks reviewed and errors found. Course completion should remain separate from authorisation to use AI in consequential work.

Sources & research notes

01 / OECD / Generative AI and the SME Workforce (2025)

https://www.oecd.org/en/publications/generative-ai-and-the-sme-workforce_2d08b99d-en.html
Survey of more than 5,000 SMEs in seven countries in late 2024. Reported benefits are respondents’ assessments, not causal estimates.

02 / Noy and Zhang / Science (2023)

https://pubmed.ncbi.nlm.nih.gov/37440646/
Randomised experiment: 453 professionals; writing tasks, not diplomatic performance.

03 / METR / Experienced developer productivity study (10 July 2025)

https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/
16 developers, 246 repository tasks; early-2025 tools and experienced maintainers.

04 / METR / Experiment design update (24 February 2026)

https://metr.org/blog/2026-02-24-uplift-update/
Selection and measurement problems make the follow-up an unreliable estimate of current productivity.

Independent TACITUS analysis. Research checked 11 November 2026. Published findings retain their dates and limits. Proposed curricula, tests, schedules and worked calculations are labelled analytical examples, not client results.