



FIELD NOTE / AI AT WORK

Stuck in Pilot

~70%

OF SURVEYED SMBs REMAIN IN EARLY AI STAGES

SAS / IDC | n>1,600 | 28 COUNTRIES

A useful workflow can save two hours. That does not mean the business has changed how work gets done.

Don Long · Founder, SummitMark · Helping people and organizations become capable with AI



A polished output can move quickly past the moment when judgment matters.

The first AI workflow that works can feel like proof that you've figured it out. The harder question is whether the improvement compounds - or just makes one step faster.



01 / THE FAST WIN

Confidence is high. Gains stay modest.

There's a specific feeling that comes with the first AI workflow that actually works. Something that used to take two hours takes twenty minutes. The output is good enough. You tell a colleague, do it three more times, and begin to think: we've figured this out.

That conviction can be misleading. A workflow that saves time is real progress. It doesn't tell you whether the improvement is spreading, whether the work itself is better, or whether the people around it are still exercising the judgment that catches mistakes.

62%

of small-business owners and managers were very or extremely confident handing high-stakes tasks to AI agents.

74%

said AI improved their productivity. Most reported gains were below 25%.

Upwork, Q1 2026. U.S. SMB subgroup: 195 leaders at firms with 10-99 employees.

High confidence. Modest returns. Still in the pilot phase.

WHAT IS KEEPING THE WORK FROM CHANGING?

The SAS/IDC report points to fragmented data and tools, isolated initiatives, limited readiness, and gaps in governance and ROI measurement. These are also design questions: which decisions should AI inform, which may it own, what happens when it is wrong, and who checks?

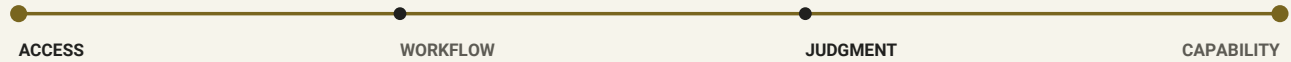
SAS / IDC research was commissioned. Survey populations and questions differ; the findings do not establish cause and effect.



02 / THE HUMAN REVIEW

The loop can become a rubber stamp.

CONCEPTUAL PATHWAY / NOT A MATURITY SCALE

*The design work happens between having tools and changing how decisions get made.*

From inside an experimental phase, the tools are working. Teams have chat assistants and automations. People have found prompts that save time. It feels like progress because it is - just not necessarily the kind that compounds. A tool added to a broken process can make its measurable surface more efficient while the underlying problem stays exactly where it was. An AI-generated customer response is useful when a person checks tone, accuracy, and what was left unsaid.

BCG's global survey of 70 executives found that half report distributed de-skilling; more than 60% see it as a material threat in three to five years. Leaders named judgment and problem-framing among the skills exposed to risk. The survey is self-reported and directional, not proof that a specific AI workflow caused skills to erode. The finding does sharpen the question: are people approving outputs, or evaluating them?

People stay in the loop. They may stop doing the part that mattered.

That is the gap between having AI and having AI-enabled operations: deciding what the system is optimizing for, where to escalate when it fails, and whether the benefits improve the whole process.

THE VISIBLE RESULT CAN BE MISLEADING

Tools in place. Time saved. Polished outputs. Yet no one may have decided what the AI is optimizing for, how failures escalate, or whether the gains improve the full process.

SOURCE: BCG GLOBAL SURVEY OF 70 C-SUITE LEADERS AND SENIOR EXECUTIVES; RESULTS ARE DIRECTIONAL.

**03 / FROM PILOT TO CAPABILITY**

Make the review real.

The businesses that move beyond experimentation do quieter work: they decide which decisions AI should inform, which it may own, and which should stay human. They build review that tests judgment, define what happens when the system is wrong, and measure whether the whole process improves.

The transformation is in the work around the tool.

TRY THIS ON ONE RECURRING TASK

Is review still real?

- 01** Who reviews the output before it acts?
- 02** What specifically do they check?
- 03** When did they last push back and change something?

If the team can't name what the reviewer checks, the review step needs redesign. Start with one process. Make the judgment visible.

REPLY WITH ONE EXAMPLE

The last time someone changed an AI output - and what they caught.

M.R.

SOURCES: SAS / IDC (COMMISSIONED SMB STUDY); UPWORK RESEARCH INSTITUTE (Q1 2026); BCG (GLOBAL SURVEY, n=70).