

From Operation to Delegation: Why Agentic AI is Fundamentally Changing Work – and What That Means for SMEs

Phronesis AI

2026-02-16

The Phase Transition

In February 2026, 16 autonomous AI agents wrote a complete C compiler in two weeks – 100,000 lines of Rust code that compiles the Linux kernel and passes 99% of a torture test suite. Cost: \$20,000. Just twelve months ago, autonomous agents lost the thread after thirty minutes. Six months ago, it was considered remarkable when an agent lasted seven hours. From thirty minutes to two weeks in one year – that’s not a trend line. That is, as the analyst put it, a *phase transition*.

Such stories seem like news from software development. They are. But the core of what’s happening here doesn’t concern software development. It concerns the question of what happens when AI no longer assists minute by minute, but works independently for days and weeks. And this question concerns every company whose value creation depends on knowledge work.

What Has Actually Changed

Public discussion about AI usually revolves around model sizes, benchmarks, and context windows. These are the wrong metrics. The right metric is one that hardly anyone knows: a model’s ability to retrieve and use information within its context window.

A model that can absorb a million tokens but only retrieves the right information in one out of five cases is like a filing cabinet without an index. The documents are in there, but whether you find what you need is a matter of chance. That was precisely the state of affairs in January 2026: The best models found the needle in the haystack in 18 to 26 percent of cases.

Opus 4.6, released in early February, achieves 76 percent with a million tokens and 93 percent with a quarter of that. This is the real breakthrough: not the amount of information a model can absorb, but the reliability with which it understands and uses it. It’s the difference between a model that sees a file and a model that holds an entire system in its head – every dependency, every interaction, every implication.

This is the capability that distinguishes an experienced employee from an external consultant reading the documents for the first time. The experienced employee knows that a

change in procurement affects the costing, that the complaint rate is connected to supplier selection, that the warranty claim looks different when assembly was performed by the customer themselves. Not because they look it up, but because they've lived in the system long enough to grasp connections intuitively.

Precisely this holistic awareness is what an AI agent can now provide – not through years of experience, but through the ability to hold the entire context simultaneously and reason across it.

From Tool to Counterpart: The Real Revolution

Most companies today use AI as a better search engine or as a text generator. You ask a question, you get an answer. You provide a prompt, you get a draft. This is the paradigm of tool operation: The human formulates the process, the AI executes one step.

What is emerging now is something fundamentally different. Anthropic calls it *outcome orientation* – describing results rather than processes. You don't explain to the AI how to build the spreadsheet. You explain what the spreadsheet needs to show. You don't describe the steps of complaint processing. You delegate: "Handle this complaint."

That sounds like a gradual difference. It isn't. It's a paradigm shift in human-machine interaction as fundamental as the transition from the command line to the graphical user interface in the 1980s. Back then, the computer stopped being a machine you program and became a tool you operate. Now it stops being a tool you operate and becomes a counterpart to which you delegate.

The competency that matters shifts accordingly: away from technical mastery of a tool, toward clarity of one's own intention. Anyone who knows precisely what they need – and can articulate it the way you would tell a competent employee – can now achieve things that previously required entire departments.

The Dissolution of the Boundary Between Technical and Non-Technical

One of the most remarkable aspects of recent developments: At Rakuten, the Japanese e-commerce corporation, non-technical employees use the same AI infrastructure as developers to build features and deploy them to production. Two CNBC reporters – not engineers – built a functioning project management tool in under an hour that replicates the core functionality of a \$5 billion product.

This isn't the democratization of technology in the usual sense, where you make a complicated tool easier to operate. This is the dissolution of the category itself. The distinction between technical and non-technical employees – a distinction that has organized knowledge work, salary structures, and org charts for thirty years – is dissolving in months.

For SMEs, this has a specific meaning. Here there is rarely an IT department with twenty developers. Here there are master craftsmen, sales managers, clerks, engineers – people with deep expertise in their field, but without programming skills. It is precisely these people who are not replaced by Agentic AI, but multiplied. Their expertise – the ability to judge

whether a quote is correct, whether a complaint is justified, whether a standard has been correctly applied – becomes the lever that was previously missing.

Judgment as the New Bottleneck

The common fear is: AI replaces human work. The reality is more differentiated and in some ways more demanding.

What AI replaces is execution. What it does not replace – and what gains dramatically in value through it – is judgment. Domain expertise. What in English is called “taste”: the deep understanding of what constitutes a good result, what a correct quote looks like, which formulation in a complaint response is legally sound and which is not.

The 16 agents who built the C compiler didn’t need anyone to write code for them. They needed someone who could specify precisely enough what a C compiler is. The marketing team no longer needs someone to operate the analytics platform – it needs someone who knows which metrics are relevant and why.

The lever has shifted: from execution to judgment. And this lever multiplies with the number of agents a person can direct. AI-native companies today generate five to seven million dollars in revenue per employee – five to seven times what counts as “excellent” in traditional software companies. Not because they hired better people, but because their people orchestrate agents instead of executing themselves.

Management as an Emergent Property

A fascinating result of recent developments: When you set multiple AI agents on a complex task, they organize themselves independently into hierarchical structures. A lead agent breaks the project into subtasks, assigns them to specialists, tracks dependencies, resolves blockers. The specialists communicate not only through the lead but also directly with each other – peer-to-peer coordination.

This is not an imposed structure. It is convergent evolution. Hierarchy is not a cultural convention that humans impose on AI systems. It is an emergent property of coordinating multiple intelligent actors on complex tasks. Humans invented management because management is what intelligence does when it needs to scale. AI agents discovered the same thing – for the same structural reasons.

For the argument in favor of a platform like Phronesis, this is central: The platform doesn’t simply digitally replicate existing workflows. It provides the infrastructure from which agents organize themselves – with skills as defined workflows, tools as individual capabilities, and contexts as department-specific knowledge. The platform is what a good company offers its employees: clear structures, available knowledge, defined processes. The agent uses all of this – but it decides for itself what it needs for the respective task.

The Pace and Its Consequences

The phase transition happening here is remarkable not only in its direction but above all in its speed. The tools that were state of the art in January are a different generation in February. The researcher at Anthropic who was involved in the C compiler project put it this way: “I did not expect this to be anywhere near possible so early in 2026.”

This speed has a paradoxical consequence: Anyone who commits to a specific AI tool today and masters it must expect that their knowledge will be obsolete in a few months. This applies to ChatGPT as much as to Copilot. Anyone who has optimized their workflow around a particular prompt pattern or model version experiences a devaluation of their expertise with every update.

The answer to this is not to learn individual tools faster. The answer is a layer of abstraction: a platform that decouples the company’s expertise from the specific AI technology. Skills that define what should be done remain stable even when the underlying model changes every three months. Contexts that specify which knowledge is relevant in which department survive every model change. Company knowledge – product data, price lists, standards, guidelines – remains independent of whether Opus 4.6, Opus 5, or something entirely different is working under the hood.

This is Phronesis’s core architectural idea: decoupling company knowledge and workflows from rapidly changing AI technology. The platform absorbs the technological change so that the company can focus on what remains stable: its expertise, its processes, its judgment.

Why SMEs Are Not Waiting But Acting

The numbers from Silicon Valley – Cursor with \$5 million in revenue per employee, McKinsey with the goal of agent-human parity by the end of 2026, Amazon teams reorganizing to “two people plus agent fleet” – that sounds like a different world than the kitchen studio in Lower Bavaria or the machinery manufacturer in the Bergisches Land.

But the core of the argument actually hits SMEs harder than large corporations. Because:

SMEs have what AI does not have: deep, specific expertise. The ability to judge whether a kitchen quote is correctly calculated. The knowledge of which DIN standard applies to a particular construction type. The experience of how to answer a complaint in a way that satisfies the customer while legally protecting the company. This knowledge resides in the minds of employees who have often been with the company for decades – and who are increasingly difficult to replace.

What SMEs do not have: infinitely scalable labor. Skilled workers are missing, and they will continue to be missing. Every master craftsman, every clerk, every sales manager spends a significant portion of their working time on tasks that require expertise but are essentially repetitive: writing quotes, creating reports, looking up standards, processing complaints. Not because these tasks are trivial – they aren’t – but because they follow a pattern that an agent can learn.

Agentic AI multiplies precisely this combination. The employee's expertise becomes the lever, the agent infrastructure becomes the multiplier. The master craftsman no longer writes every quote themselves – they delegate it and review the result. The clerk no longer processes every complaint from scratch – they delegate the standard cases and concentrate on those that require genuine judgment. The engineer no longer searches through standards for hours – they delegate the research and evaluate the result.

This is not automation in the industrial sense, where a robot replaces the human. It is delegation in the true sense: A competent employee hands over a task to a competent counterpart that knows the procedures, has the knowledge, and delivers the result in the right form.

The Question at Hand

McKinsey is advising its own partners to bring the number of AI agents to parity with the number of human employees by the end of 2026. The question for SMEs is not whether this development is coming. It is whether you accompany it with generic tools like ChatGPT – tools that don't know company knowledge, that have no skills, that don't know what a quote in this specific company must look like – or with an infrastructure tailored to your own expertise, your own processes, and your own quality standards.

The question is not: "Should we use AI?" The question is: "What is our ratio of agents to employees – and what must each employee excel at for this ratio to work?"

Phronesis is the infrastructure that makes this question answerable. Not as a promise, but as a productive system: 39 skills in use, over 40 tools available, company knowledge fully integrated, GDPR-compliant on dedicated infrastructure. Not someday. Now.

Based on an analysis of recent developments in Agentic AI, particularly the results of Anthropic's Opus 4.6 (February 2026), Rakuten's productive deployment of agent teams, and the emerging reorganization of knowledge work toward human-agent teams.