

PERSPECTIVE · NOVEMBER 2026 · 9 MIN READ

Agents are employees without a contract

How companies define the task, permissions and oversight for AI agents



Philipp Pöschke
Founder, Paice

AT A GLANCE

- An agent pursues a goal over several steps and changes something in your systems along the way, so it has to be settled before launch which steps it may take without approval.
- Autonomy is set for each step individually and grows level by level, after a probationary period with metrics agreed in advance.
- A one-page agent profile records what the agent does, what it may do, who is accountable for it and how its work stays traceable.

Agents are taking over work steps in companies that people used to do. They open cases, change records, write messages and trigger orders. The discussion mostly revolves around models and platforms. For use in day-to-day operations, three other questions matter: Which steps may an agent take without approval? Who is accountable for it? How does what it did stay traceable?

We suggest introducing agents like new employees, with a job description, limited permissions, an accountable person and a probationary period after which autonomy grows step by step. In this article we describe four levels of autonomy and the criteria for assigning them, the metrics for the probationary period, the typical risks and a profile that fits on one page.

What makes an agent

The old distinction that an assistant answers and an agent acts no longer holds, now that widely used office and development tools have agent features of their own. A definition based on effect is more precise. An agent pursues a goal over several steps and uses tools that change something along the way. It writes to systems, sends, orders or books. For governance, what matters is therefore not the product but whether a step has an effect outside.

In the AI delta model, agents belong mainly to the Build area. An agent that impresses in a demo has to cope with real permissions, exceptions and costs in operation. The analyst firm Gartner expects that more than 40 per cent of agentic AI projects will be cancelled by the end of 2027, due to rising costs, unclear value or inadequate risk controls. A framework that is in place before launch makes value, costs and risks testable before much money is committed.

Four levels of autonomy

Autonomy can be set in levels. The level applies to a single step. The same agent can check incoming invoices itself and still put every payment forward for approval.

FIGURE 1

Four levels of autonomy for agents

01

Level 1: Suggest

The agent reads and suggests; a human acts. For new tasks, irreversible steps and high potential damage. Example: drafting a reply to a complaint.

02

Level 2: Act after approval

The agent prepares the step completely; a human approves each one. For reversible but consequential steps. Example: an order with a new supplier.

03

Level 3: Handle routine, escalate exceptions

The agent handles defined standard cases itself and escalates everything outside the rules; humans check samples. For frequent, reversible steps with limited damage. Example: orders up to a fixed amount.

04

Level 4: Independent within limits

The agent acts within a tightly defined scope and reports back; humans monitor through metrics and logs. Only if errors are detected automatically and no high-risk area is affected. Example: sorting and routing incoming mail by rules.

Illustration: Paice

Which level a step gets depends on five criteria:

- **Reversibility:** Can the step be undone without harm, like an internal forwarding, or is it final, like a payment or a message to customers?
- **Potential damage:** What does a single error cost, financially and legally?
- **Frequency:** A step that happens a thousand times a month makes human approval expensive and justifies more effort for sampling and monitoring.
- **Detectability of errors:** Does an error show up immediately, for example through a plausibility check, or only weeks later at the customer's end?

- **Legal classification:** Does the step touch a high-risk area under the AI Act, personal data or co-determination rights?

A simplified example from purchasing shows how this works. An agent checks purchase requisitions against framework agreements. It places orders up to a fixed amount with listed suppliers itself, which is level 3. Anything above that, and every order with a new supplier, goes forward for approval, which is level 2. Purchasing sets the amount.

The probationary period

An agent starts at the lowest level that makes sense for its task. Whether it moves up is decided by a probationary period with metrics fixed before launch:

- **Deviations in the sample:** Share of reviewed cases in which a human would have decided differently.
- **Interventions:** How often people had to correct, stop or reverse a step.
- **Severity of errors:** Errors counted by class, from no consequence through correctable to consequences outside the company.
- **Escalations:** Does the agent put forward the cases it should? A missed exception weighs more than one escalated unnecessarily.
- **Cost per case:** Model usage, compute and the time people spend reviewing, added together.
- **Traceability:** Share of cases with a complete log of input, steps, reasoning and result.

The accountable business unit sets the thresholds for each metric before the agent starts, together with the duration and size of the sample. A level only goes up once all metrics have been met over a minimum number of cases agreed in advance. Conversely, a serious error leads straight back to the previous level. After every change of model or major change to the instructions, the probationary period starts again, because in effect it's a different agent. The same applies if a vendor swaps the model behind a platform, so contracts should require such changes to be announced.

The accountable person in the business unit decides on a higher level, after reviewing the probation results with the team that runs the agent. The decision and its basis go into the profile. That way you can still show months later why an agent was allowed to take a step on its own.

Typical risks

Agents inherit the risks of language models and bring their own. In December 2025 the OWASP GenAI Security Project summarised them in a separate list for agentic applications. Four of them are particularly important for companies:

- **Manipulation through content it reads.** An agent reads emails, documents or web pages. If they contain hidden instructions, it can switch goals without anyone noticing. Specialists call this indirect prompt injection; OWASP calls it "Agent Goal Hijack".
- **Over-broad permissions.** It's tempting to give an agent more functions, rights or freedom than its task requires. OWASP calls this "Excessive Agency". An agent that only sorts mail doesn't need the right to

send it.

- **No identity of its own.** If an agent works with a person's account or a shared service account, you can't later separate who did what, and you can't revoke its rights in a targeted way.
- **Chains of errors.** A wrong intermediate step becomes the basis for the next ones, and with several agents working together, across system boundaries too. Then there are loops in which an agent tries the same action again and again, running up costs.

The countermeasures are well known. Permissions belong in the target systems, because a model can be talked round by content. Each agent gets its own identity with exactly the rights its task needs. Consequential actions need approval, budget and time limits stop loops, and every log can be traced to an agent and a task.

What the law adds

As of 17 November 2026. If an agent is used in a high-risk area under Annex III of the AI Act, for example when selecting job applicants, the obligations for deployers under Article 26 apply. Among other things, they require human oversight by people with the necessary competence and authority, monitoring of operation and keeping the automatically generated logs for at least six months. Employers must inform workers' representatives and the workers affected before the system is used. Following the Omnibus (Regulation (EU) 2026/1744), these obligations apply to Annex III systems from 2 December 2027.

In Germany, the Works Constitution Act (BetrVG) adds to this. Under Section 90(1) no. 3, the works council must be informed in good time about the planning of working methods and processes, explicitly including the use of artificial intelligence. Section 87(1) no. 6 gives it a co-determination right over technical equipment designed to monitor conduct or performance. According to established case law of the Federal Labour Court, it's enough that the equipment is objectively capable of monitoring. An agent that records who handled which case and how can meet that test. If the works council has to assess the use of AI, calling in an expert is deemed necessary (Section 80(3) BetrVG).

For companies that build agents into their own products, the new Product Liability Directive (EU) 2024/2853 matters. It explicitly covers software and must be transposed by 9 December 2026. The German Bundestag passed the new Product Liability Act on 8 October 2026; it is due to enter into force on 9 December 2026.

This section is a general overview and does not constitute legal advice.

The agent profile

Everything that has to be settled before launch fits on one page. The profile serves as the approval document, as the basis for the probationary period and as an entry in your register of AI systems. It is maintained like a job description and updated whenever the level changes.

FIGURE 2

The agent profile: what is settled before launch

01

Task

What the agent does, for whom, and where its task ends. Plus the cases it explicitly doesn't handle.

02

Permissions

Systems and data it accesses, read or write in each case, under its own identity. Budget and time limits per case.

03

Autonomy level

The level for each step, justified by reversibility, potential damage, frequency, detectability of errors and legal classification.

04

Accountability

An accountable person in the business unit, a team for technical operation and the people allowed to approve and intervene.

05

Probation

Metrics, thresholds, duration and sample. Who decides on a higher level and what triggers a step down.

06

Records and emergency

What is logged and for how long. How the agent is stopped and who is allowed to do it.

Illustration: Paice

What changes for the teams

With agents, the work of the people alongside them shifts. They handle fewer standard cases themselves and more exceptions, check samples and decide on the cases the agent escalates. These tasks need their own description and time in the working day, otherwise oversight becomes a box ticked on a form. A solution that works technically and is bypassed in practice hasn't been embedded.

> Treat the introduction of an agent like a new hire, with a clear task, limited permissions, an accountable person and a probationary period.

Sources

- OWASP Top 10 for Agentic Applications for 2026, OWASP GenAI Security Project, 9 Dec 2025
- LLM06:2025 Excessive Agency, OWASP GenAI Security Project, 2025
- Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027, Gartner, 25 Jun 2025

- Article 26: Obligations of deployers of high-risk AI systems, AI Act Service Desk, European Commission, version of 27 Jul 2026
- Regulation (EU) 2026/1744 (Digital Omnibus on AI), Official Journal of the EU, 24 Jul 2026
- Section 90 BetrVG, Federal Ministry of Justice, gesetze-im-internet.de
- Section 87 BetrVG, Federal Ministry of Justice, gesetze-im-internet.de
- Section 80 BetrVG, Federal Ministry of Justice, gesetze-im-internet.de
- Directive (EU) 2024/2853 on liability for defective products, Official Journal of the EU, 18 Nov 2024
- Bundestag passes modernisation of product liability law (Modernisierung des Produkthaftungsrechts), German Bundestag, 8 Oct 2026



AUTHOR

Philipp Pöschke

Founder, Paice

Philipp Pöschke founded Paice in Bremen in 2026. He is personally accountable for every engagement and works hands-on on analyses, prototypes and solutions. He is the author of the articles in Insights.