



AR-004 Confidence: 72%

The AI Agent Maturity Model

A Framework for Measuring How Ready Your Organization Actually Is

February 2026

v1.0

Florian Ziesche · Ainary Ventures

CONTENTS**FOUNDATION**

1	How to Read This Report	4
2	Executive Summary	5
3	Methodology	6

ANALYSIS

4	The Maturity Illusion	7
5	Why Existing Models Fail for Agents	9
6	The AGENT Framework: 5 Dimensions	12
7	The 5 Levels: From Playground to Organism	14
8	The 5-Minute Self-Assessment	20

ACTION

9	The Level 1 to Level 3 Playbook	21
10	Why Level 3 Is the Real Goal for 2026	23
11	Predictions	24
12	Transparency Note	25
13	Claim Register	26

1. How to Read This Report

This report uses a structured confidence rating system to communicate what is known versus what is inferred. Every quantitative claim carries its source and confidence level.

RATING	MEANING	EXAMPLE
High	3 plus independent sources, peer-reviewed or primary data	62% experimentation rate (McKinsey n equals 1,993)
Medium	1 to 2 sources, plausible but not independently confirmed	Greater than 40% project cancellation (Gartner forecast)
Low	Single secondary source, methodology unclear	Compliance cost estimates (vendor-sourced)

This report was produced using a **multi-agent research pipeline** with structured cross-referencing and gap research. Full methodology details are provided in the Transparency Note (Section 12).

2. Executive Summary

No existing AI maturity model accounts for what makes agents different from traditional AI. Organizations think they're further along than they are because they're measuring the wrong thing.

- 62% of enterprises experiment with AI agents, but fewer than 10% deploy them enterprise-wide, and only 6% see meaningful EBIT impact.^{[1][2]} The gap between experimentation and production is where most organizations live — and die.
- The AGENT framework introduces 5 measurable dimensions — Autonomy, Governance, Error Handling, Networked Trust, and Team Integration — across 5 maturity levels.
- Level 3 ("Calibrated") is the survival threshold for 2026. EU AI Act enforcement begins August 2026. Organizations below Level 3 face regulatory exposure, and the compliance cost is \$2 to 5M — but the cost of a single uncalibrated agent catastrophe will exceed \$100M.^[13]
- The model is a hypothesis, not gospel. It's built on patterns from CMMI and DORA applied to agent-specific research across 22 sources. It has not been empirically validated across enterprises. Use it as a starting diagnostic, not a certification.

62%

experimenting with agents

Source: McKinsey 2025 (n equals 1,993) |

Confidence: High

Less than 10%

enterprise-wide deployment

Source: McKinsey 2025 | Confidence: High

6%

AI High Performers (greater than or equal to
5% EBIT)

Source: McKinsey 2025 | Confidence: High

Keywords: *AI Maturity, Agent Governance, Calibration, CMMI, DORA, Autonomy Levels, Trust Infrastructure, Organizational Readiness*

3. Methodology

This framework synthesizes two categories of input: (1) a systematic review of 6 existing AI maturity models (Gartner, McKinsey, Deloitte, Microsoft, Google Cloud, IBM) to identify what they measure and what they miss, and (2) 15 research briefs on agent-specific phenomena — overconfidence calibration, adversarial attacks on multi-agent systems, memory poisoning, human-in-the-loop failure modes, non-human identity management, and regulatory convergence — totaling 22 primary and secondary sources.

Limitations: The maturity model structure draws on design principles from two proven precedents: CMMI (Carnegie Mellon, 1987 to present) and DORA (Google, 2014 to present), specifically their emphasis on outcome-based measurement, prescriptive levels, and self-assessability. The model itself is a proposed framework — not an empirically validated assessment tool. It should be treated as a structured hypothesis about what agent readiness looks like, to be tested against real organizational data.

Full methodology details, including confidence calibration and known weaknesses, are provided in the Transparency Note (Section 12).

4. The Maturity Illusion

High

(Confidence: High)

Here's the picture: 62% of enterprises are experimenting with AI agents.^[2] Gartner projects 40% of enterprise applications will incorporate agentic AI by end of 2026.^[18] The agent market is forecast to grow from \$7.8B to \$52B by 2030 — a 45.8% CAGR.^[21]

Now here's the other picture: fewer than 10% of those experimenting organizations have deployed agents enterprise-wide.^[2] Only 6% of enterprises qualify as "AI High Performers" with measurable EBIT impact, according to McKinsey's survey of 1,993 organizations.^[1] Only 54% of AI projects make it from pilot to production.^[4] And Gartner predicts more than 40% of agentic AI projects will be abandoned by 2027.^[3]

Evidence

These numbers come from large-sample surveys (McKinsey n equals 1,993, Gartner enterprise data). The 6% figure is particularly robust — McKinsey defines "High Performer" as organizations attributing greater than or equal to 5% of EBIT to AI, which is a measurable threshold, not self-assessment.^[1]

Interpretation

The gap between experimentation rates (62%) and production deployment (less than 10%) suggests a structural problem, not a timing problem. Organizations aren't slowly moving up a maturity curve — they're stuck.

I believe the core issue is that every existing AI maturity model measures the wrong thing. They ask: "How well do you USE AI?" The right question for agents is: "How well do you GOVERN actors that make decisions on your behalf?"

That distinction — tool versus actor — is why organizations think they're further along than they are. If you measure yourself against a tool-use framework, having

ChatGPT Enterprise and a few LangChain workflows puts you at Level 3. If you measure yourself against an actor-governance framework, those same deployments are Level 1.

WHAT WOULD INVALIDATE THIS?

If the 62% experimentation rate includes organizations with robust governance frameworks that simply haven't scaled yet (i.e., the bottleneck is business case, not maturity), then the "stuck at Level 1" thesis overstates the problem. I don't see evidence of this — McKinsey's data shows high performers are differentiated by workflow redesign (55% vs 20%), not by governance maturity — but it's possible.^[1]

SO WHAT?

If you're a CTO reading this, the question isn't whether you're "doing AI agents." It's whether you could answer, right now: How many agents does your organization run? What was their error rate last month? What happens when one fails? If you can't answer those questions, you're at Level 1 — regardless of your AI budget.

14. References

- [1] McKinsey & Company, "The State of AI in 2025," McKinsey Global Survey (n equals 1,993), 2025.
- [2] McKinsey & Company, "The State of AI in 2025" — agent experimentation and deployment data, 2025.
- [3] Gartner, "Predicts 2025: AI Agents" — projection on agentic project cancellation rates, 2025.
- [4] Gartner, "4 Levels of AI Maturity and How to Achieve Them" — pilot-to-production conversion rate, 2024.
- [5] PMC/12249208, "Overconfidence in Large Language Models" — study of 9 LLMs across 351 scenarios.
- [6] arXiv:2602.00279, "Verbalized Confidence Expressions in LLMs: Calibration and Reliability," January 2026.
- [7] Cost estimate based on API pricing for consistency-based confidence calibration (3 SLM calls per check). See also: Vashurin et al., "CoCoA: A Minimum Bayes Risk Framework," ICLR 2026.
- [8] Vectra AI, "2023 State of Threat Detection" — survey of 2,000 SOC analysts, 2023.
- [9] arXiv:2503.12188, "Hijacking Attacks on Multi-Agent Systems" — 45 to 64% success rates.
- [10] arXiv:2503.03704, "MINJA: Memory Injection Attacks on Multi-Agent Systems" — greater than 95% success rates.
- [11] arXiv:2510.09023, Meta, "The Rule of Two: Adversarial Prompt Injection" — 12 out of 12 defenses broken.
- [12] Volkswagen AG public filings — Cariad software unit cumulative losses of \$7.5B.
- [13] European Parliament, "Regulation (EU) 2024/1689 — Artificial Intelligence Act" — Articles 9, 14; penalty Article 99.

Citation: Ainary Research (2026). *The AI Agent Maturity Model: A Framework for Measuring How Ready Your Organization Actually Is.* AR-004.

About the Author

Florian Ziesche is the founder of Ainary Ventures, where AI does 80% of the research and humans do the 20% that matters. Before Ainary, he was CEO of 36ZERO Vision and advised startups and SMEs on AI strategy and due diligence. His conviction: HUMAN times AI equals LEVERAGE. This report is the proof.

ainaryventures.com



AI Strategy · Published Research · Daily Intelligence

Contact · Feedback

ainaryventures.com
florian@ainaryventures.com

© 2026 Ainary Ventures