



Defining and measuring the success of agentic AI

November 2025

Table of contents

Introduction	03
What success really means for agentic AI	04
The two measurement lenses	05
1. Operational lens: The nine dimensions	05
2. Strategic lens: The four value buckets	06
The operating framework	07
Agentic AI use cases	09
Conclusion	11



Introduction

Agentic AI has moved from being a curiosity to a business priority. According to Deloitte's "The state of generative AI in the enterprise" report, more than half of organisations (52%) are exploring agentic AI for automation, while nearly as many (45%) are exploring multi-agent systems¹. About one-fourth (26%) mention that they are already pursuing autonomous agent development to a great degree. Most teams are now moving from single-turn assistants to agents that can pursue defined goals.

As this transition gathers pace, it becomes crucial to define what success looks like for agentic AI. Unlike generative AI, these systems are not standalone tools. They need to be integrated into business workflows and evolve alongside them.

This requires thoughtful planning, strong cross-functional collaboration and active change management. Agentic AI models must be trained, tested and upgraded continuously, with feedback mechanisms and performance tracking systems built into the implementation process.

Organisations that derive the most value out of it will be those that set clear and measurable goals for their AI initiatives, track process diligently and maintain a steady process for managing risk. This chapter presents simple framework for identifying the right measures of success, monitoring performance at every step and using results to guide decisions and actions consistently across the business.

¹<https://www.deloitte.com/content/dam/assets-zone3/us/en/docs/campaigns/2025/us-state-of-gen-ai-2024-q4.pdf>



What success really means for agentic AI

Agents are systems that follow processes, plan and act across multiple steps, use tools and APIs, handle exceptions, learn from feedback and know when to involve humans. Thus, success for agentic AI cannot be measured in numerical terms. It requires looking at multiple factors for each agent, including outcomes, costs, autonomy, quality, safety, adaptability and adoption.

Agentic AI should be treated as a long-term capability. This means that it should be designed carefully, assigning ownership across teams, managing change thoughtfully and monitoring performance continuously.



Two key measurement lenses

Operational lens: Evaluates whether the agent can reliably complete real business workflows and effectively performs each critical step. It covers nine dimensions: business outcome, cost and efficiency, autonomy, accuracy and roundedness, alignment and safety, adaptivity, flexibility, scalability, and adoption and trust.

Strategic lens: Checks whether the capability is compounding value over quarters. It tracks value realisation into four ROI buckets: direct, indirect, opportunity and efficiency. By capturing performance over time, it gives leaders a clear picture of lasting financial impact rather than just one-off gains.

Operational lens: The nine dimensions

Sr. no.	Success dimension	Description
1	Business outcome	Measures whether the agent reliably delivers the intended result for the workflow (speed, quality, completeness vs. baseline). This is the agent-level go/no-go.
2	Cost and efficiency	Tracks true unit cost per successful run, including compute, licensing and any human effort required. It shows time and effort saved end-to-end. Moreover, it checks whether autonomy is financially viable today, identifies manual tasks that can be automated, and guides choices on model settings, tooling and service levels.
3	Autonomy	Assesses how independently the agent completes multi-step work, including error recovery and smart escalation.
4	Accuracy and roundedness	Monitors decision correctness, data use and evidence to prevent downstream rework and compliance issues.

Sr. no.	Success dimension	Description
5	Alignment and safety (responsible AI)	Ensures that the agent operates as per existing policies covering data use, regulatory compliance, privacy and fairness, with all actions properly approved and traceable.
6	Adaptivity (learning velocity)	Measures how the agent improves its performance over time, learning from feedback, reducing errors and maintaining accuracy as data or tasks evolve.
7	Flexibility (change resilience)	Evaluates the agent's ability to manage variations in tasks, data and environments. It also measures how well the agent can operate across different tools, UIs, platforms and models with minimal rework while continuing to function reliably when conditions change.
8	Scalability	Measures an agent's ability to perform effectively as workload and user engagement increase, and workflows extend across multiple systems.
9	Adoption and trust	Tracks how confidently people rely on the agent through metrics such as increasing usage, fewer overrides, smooth handoffs and satisfaction with collaboration.

Strategic lens: The four value buckets

Sr. no.	Value bucket	Description	How to measure for agents?
1	Direct	Incremental revenue or cost impact directly attributable to agent actions in the target workflow	Compare agent instances to a matched control using shadow or canary cohorts. Attribute uplift only when the agent's action caused the improvement.
2	Indirect	Downstream improvements that arise because the agent reasons better, escalates on time and reduces errors	Track lagging indicators and link improvements to agent usage, avoiding double-counting with direct value.
3	Opportunity	Gains from the agent selecting a better alternative among valid options	Log the decision set and the selected option, and compare it with the baseline option and difference in value (revenue/cost) realised.
4	Efficiency	Productivity and throughput gains driven by faster cycles, fewer errors and better use of systems and capacity, enabling more output with the same or lower resources	Track improvements in speed, quality and throughput, and convert those gains into financial value using standard productivity metrics.



The operating framework

The operating framework includes three key parameters - **People, Process and Governance**.

Key components of the framework

People

Pillars	Success dimension
Agents as collaborators	Design workflows so that employees interact with agents as co-workers or co-contributors, not as replacements. Shift the focus from competing with AI to collaborating with it and build confidence in agentic systems.
Co-creation and feedback-driven development	Reward teams for co-creating with AI and integrating agentic support into their work and take constant feedback from employees to test the efficacy of the model.
Promote transparency	Help employees understand how agents work to build trust, transparency and acceptance.

Process

Pillars	Success dimension
Redesign workflows around agents	Do not use agents merely for task automation. Structure workflows to allow agents to make autonomous decisions while continuously gathering data from their environment, including user inputs and feedback loops.
Model maintenance	Periodic retraining is essential to capture model drift and maintain performance over time.
Business decision encoding	Clearly define the core business decision supported by the agent. This includes: <ul style="list-style-type: none"> • A concise success hypothesis • Linking the scenario to a measurable key result • Defining a safety envelope that outlines what the agent is allowed to do and when it should escalate
Operational constraints and metrics	Specify constraints such as SLAs, budget limits and policy boundaries. Identify edge cases and choose relevant KPIs to monitor agent performance.
Responsible scaling	Use pilot evidence to justify scaling. Build systems and guardrails to detect deviations in agent behaviour due to fluctuating data or external factors. Periodically redefine success metrics and agent objectives to align with evolving organisational goals.
Autonomy promotion	Allow agents greater autonomy only when strong evidence supports it. Maintain a visible kill switch and a documented rollback plan detailing when and how to deactivate or revert the agent's actions.

Governance

Pillars	Success dimension
Agent accountability	Monitor the agent's behaviour and provide it with better and frequently updated guardrails to manage any legal or regulatory risks.
Transparency and audit-readiness	Conduct regular internal audits to ensure alignment with business goals and strategy.
Impact validation	Run shadow trials to compare advice versus no advice to capture the impact of the agent.
Decision attribution	Design attribution to ensure that any decision can be tied to a single agent's actions with confidence.



Agentic AI use cases

Case study 1: Automated ticketing solution for a leading automobile manufacturer

We used agentic AI to create an automated ticketing solution for a leading automobile manufacturer, which improved their IT support efficiency and user satisfaction. This involved using interactive agents to engage the user, gather relevant infrastructure issue details and automatically generate IT tickets for issues such as password resets, login assistance and profile summaries.

We applied the below operational levers because they allowed us to test the short-term success criteria for the implemented project.

We used below operational levers to test the short-term success criteria for the organisation.

Operational dimensions (applicable metrics)	KPIs and metrics used (non-exhaustive)
Cost and efficiency	<ul style="list-style-type: none"> • Cost per ticket • Ticket automation rate • IT staff productivity
Autonomy	<ul style="list-style-type: none"> • First contact resolution • Agent engagement success rate
Accuracy and roundedness	<ul style="list-style-type: none"> • Ticket resolution time • Error rate • Rework pace
Alignment and safety	<ul style="list-style-type: none"> • Policy violation rate • Escalation accuracy
Adaptivity (learning velocity)	<ul style="list-style-type: none"> • Feedback loop effectiveness • Drift resistance
Adoption and trust	<ul style="list-style-type: none"> • CSAT • NPS • Feedback sentiment

The overall business impact included improved ROI and reduced cost per ticket, which was tracked with the help of the cost savings and efficiency KPI. There was significant

improvement in systems' performance over time, which was calculated through the adaptivity KPI.

Case study 2: Citizen assistant tool for an international government organisation

We used agentic AI to help an international government organisation enhance citizens' access to critical government service information. This was done by autonomously extracting, analysing and synthesising updated information

using diverse file formats (for example, webpages and PDFs). The solution delivered accurate, consistent, user-friendly and privacy-compliant responses to citizens' queries.

We used the below operational levers to test the short-term success criteria for the organisation.

Operational dimensions (applicable metrics)	KPIs and metrics used (non-exhaustive)
Cost and efficiency	<ul style="list-style-type: none">• Cost per query solution• API compute usage per successful outcome• Reduction in manual content curation time
Autonomy	<ul style="list-style-type: none">• Measured using first contact resolution rate• Agent engagement success rate
Accuracy and roundedness	<ul style="list-style-type: none">• Response time• Error rate of responses
Alignment and safety	<ul style="list-style-type: none">• Percentage of responses that meet privacy, fairness and legal standards
Flexibility	<ul style="list-style-type: none">• Change resilience score• Percentage of successful operations after UI/API/schema changes without intervention
Scalability	<ul style="list-style-type: none">• Load testing
Adoption and trust	<ul style="list-style-type: none">• User reliance rate on the system• Trust users have on the responses generated (Can be assessed using a feedback sentiment analysis)

The business impact was measured majorly against the cost and efficiency KPI, as it resulted in cost savings per query resolved. Another key metric was the percentage reduction in

citizens' time to access critical government service information, which was measured against the scalability KPI.



Conclusion

The dual-lens approach to evaluating agentic AI offers a holistic framework to understand both its immediate effectiveness and long-term impact. The operational lens focuses on real workflows and makes every step observable. It ensures that each agent workflow delivers measurable value in real-world scenarios by assessing performance across the defined set of dimensions. This enables organisations to responsibly integrate agentic AI within existing processes and achieve the desired outcomes.

On the other hand, the strategic lens shifts the focus from short-term results to sustained business value. By aggregating the outcomes into ROI categories (direct, indirect, opportunity and efficiency), it provides a broader view of how benefits of

agentic AI compound over time. The strategic view rolls those results into clear value buckets so leaders can make funding decisions with confidence. The operating framework ties both views into a model for implementing best practices and handling change management within organisations. This way, teams know what to do at each stage.

Together, these lenses help organisations adopt agentic AI in a practical and forward-looking way, delivering tangible results today while also aiming for long-term impact. The message is simple: measure what matters, prove value where the work happens, scale only when the evidence is strong and risks are controlled, and build for portability so progress is not trapped on one platform.

Connect with us

Ashvin Vellody

Partner
Deloitte India
ashvinv@deloitte.com

Moumita Sarker

Partner
Deloitte India
msarker@deloitte.com

Raghu A

Partner
Deloitte India
araghu@deloitte.com

Contributors

Aditi Mittal

Mohammed Riyaz Ahmed

Sachin Suresh O

Nandu R

Sagarika Gupta

Srishti Deoras

Aiswarya Harilal

Acknowledgements

Sonali Lingwal

Ruchira Thakur





Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (“DTTL”), its global network of member firms, and their related entities (collectively, the “Deloitte organization”). DTTL (also referred to as “Deloitte Global”) and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/about to learn more.

Deloitte Asia Pacific Limited is a company limited by guarantee and a member firm of DTTL. Members of Deloitte Asia Pacific Limited and their related entities, each of which is a separate and independent legal entity, provide services from more than 100 cities across the region, including Auckland, Bangkok, Beijing, Bengaluru, Hanoi, Hong Kong, Jakarta, Kuala Lumpur, Manila, Melbourne, Mumbai, New Delhi, Osaka, Seoul, Shanghai, Singapore, Sydney, Taipei and Tokyo.

This communication contains general information only, and none of DTTL, its global network of member firms or their related entities is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication.

© 2025 Deloitte Touche Tohmatsu India LLP. Member of Deloitte Touche Tohmatsu Limited