pwc

The AI trust dividend

How trust by design accelerates AI value

July 2026

Table of contents

Introduction

AI opportunities are growing as organisations move from experimentation to enterprise-scale deployment. Those embedding trust by design are building the confidence, speed and resilience needed to turn AI ambition into measurable value.

By Nicola Costello, Ilana Golbin Blumenfeld and Luke Vilain

Artificial intelligence (AI) has entered a new phase. It's no longer a set of isolated use cases or experiments—it's now core infrastructure, built into strategy, operations, and decision-making. In this new phase, organisations are racing to turn AI activity into measurable value, but many are struggling to do so. AI is failing to deliver value for the majority of companies. Just one-in-eight (12%) CEOs say AI has delivered both cost and revenue benefits, while 56% say they have seen no significant financial benefit so far.

The challenge is rarely the technology itself. More often, business leaders are working out where AI can create the most value, how people can work effectively alongside it, and how to build the confidence to deploy and scale at pace.

Sustainable value and organisational transformation come from building the capabilities, governance, and trust to scale AI across the enterprise. To do that, you need to move towards becoming AI-native, integrating AI into your organisation's operating model and value creation processes. This demands more than technology investment. It needs governance, controls, and continuous monitoring to be designed into AI systems from the outset. That is how you build the confidence to deploy AI at scale. PwC's AI performance study shows that the most AI fit companies deliver AI-driven revenues and efficiencies that are 7.2x as high as those of other companies.

This second article in PwC's The AI trust dividend series is for business and risk executives thinking about how to build trust in AI by design. It looks at how you recast AI risk management from a compliance constraint into a scaling advantage and return on investment (ROI) enabler.

Organisations that make this shift can unlock the 'trust dividend': faster product development, better customer engagement, and new types of products and services. Those that do not will find that risk, fragmentation, and lack of confidence become the primary constraints on value creation.

Soon AI adoption will reach such scale inside enterprises that it will be impossible to ignore. Like email before it, the conversation will shift from whether to adopt it to how to make it as safe as possible while moving forward. With email, risk controls took years to catch up. Implementing trust by design can prevent history from repeating itself.

By Luke Vilain

Director, PwC UK

01

The essence of trust by design

Trust is not something that can be added at the end of development. It is the outcome of clear and intentional decisions, such as:

01 **Leadership, strategy, and risk appetite.** Is it clear what your organisation wants AI to achieve and where it will not go? In most organisations there is work to do to improve leadership's AI literacy as part of this process. Clear top-down leadership is needed to bring the cross-section of capabilities—engineering, risk, legal, compliance, cyber, core technology, and business teams—together and help teams understand the strategic imperative behind AI investment. Leadership is also needed to connect AI goals to overall business strategy and transformation goals.

02 **Operating model.** Do you have a clear view of who can access which types of technology, and are appropriate controls in place? For instance, a large language model (LLM) interface will likely have broader enterprise-wide use, while access to AI coding agents or a platform with data connectors or model context protocols to other systems and agents would be more strictly contained.

03 **AI economics and technology enablement.** Is there an enterprise-level approach to which technologies are the right fit for specific AI capabilities, and how scarce resources will be managed as demand scales? The question is broader than choosing the right model. You need a clear view of which capabilities should sit on enterprise platforms, specialist tools, coding agents, data connectors, or bespoke builds, and what level of developer support, infrastructure, compute, and oversight each requires. As adoption grows, token consumption, inference costs, latency, and cloud capacity can become material constraints. Trust by design should therefore include clear usage policies, cost transparency, and monitoring over resource consumption, so AI systems remain scalable, efficient, and commercially sustainable.

04 **Accountability.** Who owns outcomes, risks, controls, and escalation? Establishing accountability is critical. Typically, AI governance spans multiple broad risk areas, each with its own policy owner who needs to be satisfied that the risks they are accountable for are controlled. However, there is often overlap or lack of clarity on where risk ownership lies. AI governance should be orchestrated and coordinated across multiple teams, many of which pre-date AI. We explore this further in our recent article Creating strong foundations, trusted AI, and real returns.

05 Control architecture. Are reusable controls, guardrails, and testing patterns in place and embedded in workflows? Are there automations and tactical guidance to support execution? We will explore this further in a future article in this series.

06 AI lifecycle governance. Does this provide clear and complete coverage of intake, risk tiering, approval, deployment, monitoring, and decommissioning? Are lifecycles adjusted for building predictive AI, generative AI, or agentic AI? In some organisations, fundamental reform of oversight and governance processes may be needed, as they were built for a small number of models going live each month and staying stable in production.

07 Observability and assurance. Is there continuous visibility, evidence, auditability, and intervention mechanisms? Have you defined performance thresholds and the instances where escalations to humans will occur? We will also explore this topic further in a future article in this series.

Trust by design means moving from late-stage review and challenge to a more proactive model, where the conditions for confidence are built into how AI is conceived, developed, deployed, and improved over time. Trust lets you build control into the end-to-end design and deployment of AI systems. This helps you move from isolated pilots to enterprise-wide adoption more efficiently and institutionalises governance from the beginning.

Conversely, where trust is treated as an afterthought, speed to market can slow dramatically, risks can multiply, and you may face customer loss, regulatory intervention, talent attrition, system shutdowns, or strategic disadvantage from an inability to deliver on your AI ambition. The leadership test: if every AI use case still needs a bespoke governance debate, trust has not been designed in.

> “As AI becomes developer- and citizen-led, tracking every use case individually becomes impractical—as it was with spreadsheets before it. Organisations need embedded controls in approved platforms that generate an inventory and apply safeguards automatically. Governance must enable AI at scale, not constrain it.

By Ilana Golbin Blumenfeld

Partner, PwC US

02

Building trust into each use case

A **trust by design** approach integrates into every stage of the AI lifecycle, ensuring controls are woven into how AI is designed, deployed, and evolved. At a high level, this means:

- Understanding the risks that the use case or agent—and its components—trigger, and each risk's impact level.
- Determining what controls are available that mitigate the risks identified and, through testing, whether the control keeps the risk within your organisation's risk appetite.
- Monitoring the use case while in production to see if new risks or risks outside of your organisation's risk appetite are materialising, and act accordingly.

In each stage, there's an automation opportunity that can turbocharge speed to value while controlling risk.

01 Understand risks and their level of impact

Every AI initiative should start with a clear outcome the use case or agent needs to deliver. You should define this outcome—and how you will measure it—before you commit funding. As development progresses, this outcome serves as your north star. It helps you check that the use case is solving the right problems and can deliver the anticipated value.

Business and development teams are often clear on the outcomes they want to achieve but may be less aware of the risks that come with them. That's why it's essential to involve risk teams early. They can provide insight into the risk impacts of each use case or agent, as well as the risks linked to any model components you build in. This will determine the controls required.

Decision rights, stop or go gates, and 'kill switch criteria' (who decides, when to proceed, and when to stop) are agreed up front based on risk understanding. Automation can accelerate this process by identifying associated risks based on a use case description.

A risk tiering approach that matches the use case risks and associated impacts to pre-defined tiers (such as low, medium, high) that are aligned to enterprise governance, helps you apply proportionate controls. This way, low-risk use cases are not over controlled and unnecessarily slowed, and high-risk use cases are not under controlled, which could lead to loss of trust in the market or regulatory action.

02 Establish controls that keep risk within appetite

Well-designed, effectively operating controls help to build confidence and, ultimately, trust. Some are code-based. For example, automated triggers can detect when sensitive data is accessed, shared, or used in prompts outside approved parameters, while runtime guardrails can block harmful outputs or halt activity when predefined thresholds are exceeded. Some controls are enforced through the development and deployment platforms themselves, for example requiring certain summary data to be provided so an accurate inventory can be maintained. Some are policy-based but can be enforced via code, such as identity and access management controls that restrict who can access particular models, agents, data sources, or development environments.

Others are people- and process-based and enacted through culture and training. Non-code controls are often the strongest options yet thought about the least. The sharing of confidential data, for example, is largely controlled because employees have been made aware they should not do it, not by technical controls alone—a lesson collectively learned over the years rather thanconsidered from the outset. There is an opportunity with AI to take these hard-learned lessons forward by considering how employees will be made aware of what should and should not be done with AI.

Having a common controls taxonomy, or way of thinking and communicating about controls, helps all parties understand what a control is for and how it should be tested. It confirms that a uniform approach to applying controls is used across use cases. It also helps in controls automation.

Risk teams, including Risk, Compliance, Legal, Quality Management, Chief Information Security Officer (CISO), and Financial Control functions, can play a valuable role in helping to identify methods and technology enablement to control risk. For example, development teams may use automatically generated test plans aligned to the risk profile of a use case, with higher-risk AI systems subject to more extensive evaluation, red teaming, and validation requirements than lower-risk deployments.

When risk teams build a central controls repository or other method of sharing common controls, development teams can avoid one-off control builds and take advantage of easily configurable, pre-built controls and observability tools.

Once your list of desired controls is established, the build process is iterative: look at the technology capabilities at your disposal, and the outcomes you're trying to get to, and then build, test, iterate; build, test, iterate. As controls are built and tested, keeping a trail of control evidence will facilitate the independent assessments and audits that will likely be needed as the process progresses.

By looking across how controls are being deployed, you can identify opportunities to apply a control at the platform level rather than the use case and agent level. In doing so, you can, by design, embed that platform-level control into any use case being developed on the platform, creating both scale efficiencies and improvements in the control environment.

> “As agentic AI becomes more complex, you’ll need clearer accountabilities and new roles to manage risk. The lesson from earlier technology shifts is simple: accountability must evolve before incidents expose gaps.

By Nicola Costello

Partner, PwC Australia

03 Monitor whether controls continue to keep risk within appetite once AI is live

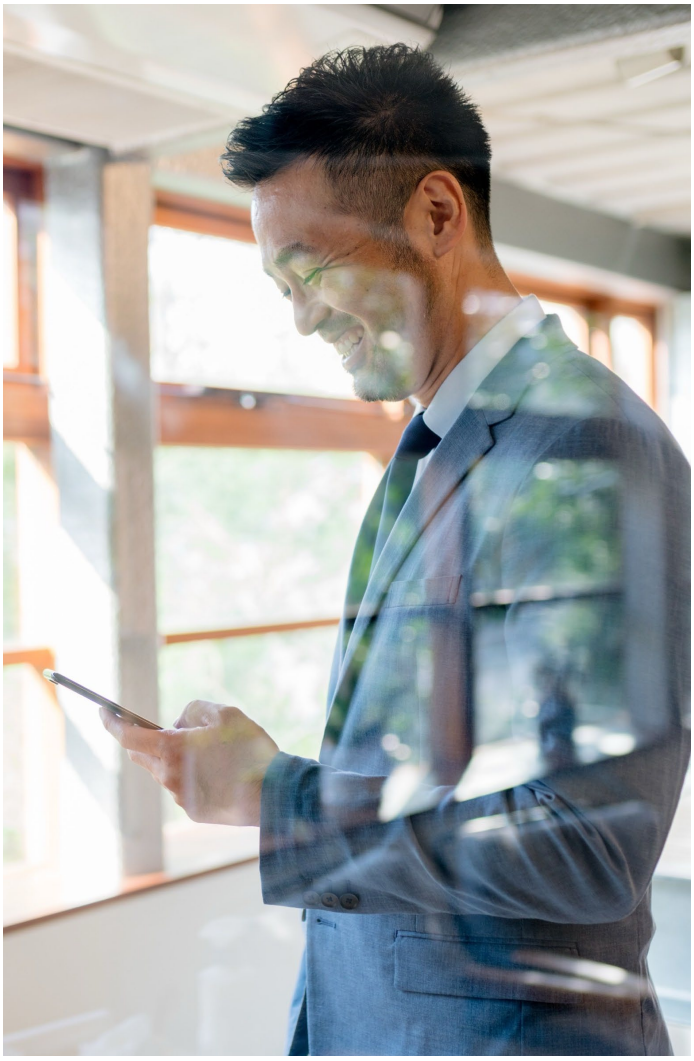
Significant scaling advantages come when trust capabilities are embedded once and reused many times. Monitoring is therefore not separate from the control architecture. It's the way controls continue to operate, generate evidence, and trigger intervention once AI systems are live. We explore this further in our recent PwC US article, Observability: A critical ingredient in making AI and agents work for you.

This means thinking about controls across the lifecycle, not treating monitoring as a separate afterthought. Some controls mitigate risks before the use case, agent, or system goes live and are tested through red teaming, evaluation, and validation. Others operate in production in real time, with the capability to pause, block, or escalate activity if thresholds are breached. Still others provide ongoing monitoring and after-the-fact review to confirm that the AI application is working as anticipated, surface changes in risk profile, and alert the need for human intervention.

In AI-native organisations, observability becomes the enterprise capability that makes this continuous control environment possible. It's not limited to monitoring technical performance or model drift. It provides visibility into whether AI systems continue to behave as intended, whether outcomes remain aligned with business objectives and risk appetite, and whether intervention is needed as models, agents, and workflows evolve. This becomes increasingly important as AI systems move from discrete tools to interconnected agents operating across business processes.

The need for observability rises with the pace of technology innovation. As organisations move forward with agentic AI, how will you monitor the possibility that unanticipated risks are developing as the agent learns over time? With multiple agents working together, how will you identify and control unexpected outcomes that are not controlled at the individual agent level?

Gartner forecasts that large language model observability investment will cover 50% of GenAI deployments by 2028, up from 15% today.[1]



This approach shifts trust by design from a use-case-by-use-case review activity to an enterprise operating model capability. Done well, it gives business leaders clearer ownership of the value, outcomes, and risks of AI. It also gives boards, regulators, and other stakeholders greater confidence that AI is being scaled with appropriate evidence, controls, and oversight.

[1] Source: Gartner, "Gartner Predicts by 2028, Explainable AI Will Drive LLM Observability Investments to 50% for Secure GenAI Deployment," published 30 March 2026, accessed 14 July 2026, https://www.gartner.com/en/newsroom/press-releases/2026-03-30-gartner-predicts-by-2028-explainable-ai-will-drive-llm-observability-investments-to-50-percent-for-secure-genai-deployment.

Snapshot—Same capability, different outcomes.

Consider two financial services institutions building the same capability: automated electronic trading for complex derivatives.

Company A treats controls as a final approval step. Each new trading algorithm is reviewed on its own, with risk, technology, and business teams debating what safeguards to add late in development. Controls are implemented inconsistently across teams, and monitoring relies on periodic manual checks. As trading volumes increase, confidence becomes the constraint. Leaders are reluctant to scale because they can't easily see whether every algorithm is operating within agreed parameters.

Company B builds the same capability, but designs trust in from the start. It sets risk-based procedures, where capabilities that impact customers or operate critical company decisions are routed to pre-agreed teams that discuss and confirm the right controls. It is built upon a platform that guides the development team and use case owner to consider potential shortcomings and improvements throughout the development process. The company creates reusable platform-level controls and configures them as each use case develops. One control is an automated stop mechanism that checks each proposed trade against an industry database of mid-market prices for complex derivatives. If the trading price moves more than two standard deviations from that reference point, the algorithm pauses and sends the trade for human review. This control is embedded into the code base for every new trading algorithm, with production monitoring to identify any algorithm where the control has been disabled.

The difference is not the trading ambition or the underlying technology—it's the operating model around it. Company B can scale with greater confidence because its controls are reusable, produce clear evidence, and are embedded in the process by default. What would otherwise be a bespoke governance debate for every new algorithm becomes a repeatable enterprise capability. That's the trust dividend in practice—stronger control, faster deployment, and greater confidence to increase automated trading volumes safely.

03

Actions for leaders to turn trust into AI ROI

The organisations that capture the greatest return from AI will not be those with the largest number of pilots. They will be those that can repeatedly identify high-value use cases, move them through delivery with confidence, and scale them safely across the enterprise.

This is where AI governance becomes an enabler of ROI. Effective governance helps leaders make better investment decisions, focus resources on the use cases most aligned to strategy, reduce delays caused by late-stage risk concerns, and create reusable trust capabilities that lower the cost and complexity of scaling AI over time.

Trust by design therefore needs to be treated as a strategic growth capability, not a compliance exercise. Leaders should be asking whether their organisation has the governance, accountability, and operating model needed to convert AI ambition into measurable business value, while giving boards, regulators, and other stakeholders greater confidence that AI is being scaled with appropriate evidence, controls, and oversight.

Key questions include

Are you directing AI investment towards the outcomes that matter most?

AI governance should help you prioritise use cases based on strategic value, feasibility, risk, and expected return. Without this alignment, you risk spreading investment across disconnected pilots that consume resources without creating enterprise-scale impact.

Do you understand the confidence threshold required to scale each use case?

Different AI use cases require varying levels of trust. Internal productivity tools, customer-facing assistants, decision-support models, and autonomous agents each carry different implications for customers, employees, regulators, and the business. Leaders need to understand what level of evidence, control, and assurance is required before each type of use case can scale.

Are you reducing friction in the path from idea to value?

Trust by design should make delivery faster, not slower, and should not stop agility and innovation. When risk appetite, decision rights, control expectations, and approval pathways are clear from the outset, teams can avoid late-stage rework and move more quickly from experimentation to production.

Are you building reusable trust capabilities rather than reinventing governance for every use case?

The ROI from AI improves when common controls, guardrails, monitoring capabilities, and approval patterns can be reused across multiple use cases. This shifts governance from a bespoke review process to a scalable enterprise capability. Platform-level controls, common taxonomies and reusable assurance evidence help reduce duplication, create consistency, and make each subsequent deployment faster and more reliable.

Who owns the value, outcomes, and risks of the AI they deploy?

AI ROI depends on adoption, process change, and measurable business impact. Business leaders need to own not only the expected benefits, but also the decisions, behaviours, and risks created by AI-enabled processes. This shifts accountability from a centralised review model to a first-line ownership model, where business leaders remain accountable for the value, outcomes, and risks of the AI they deploy, supported by risk, technology, legal, and compliance teams that are involved early enough to shape delivery decisions.

Can you see whether deployed AI is continuing to deliver value within appetite?

ROI is not secured at launch. Organisations need visibility into whether AI systems are performing as intended, being adopted by users, producing reliable outcomes, and remaining within agreed risk appetite as models, data, business processes, and user behaviours change.

This visibility also creates the evidence base needed for internal challenge, independent assurance, and external stakeholder confidence. As AI systems become more dynamic and agentic, leaders will need assurance over not just whether AI is performing technically, but whether it continues to deliver intended outcomes within appetite.

Are you learning from each deployment to make the next one faster, safer, and more valuable?

The trust dividend compounds when organisations capture lessons, evidence, controls, and patterns from each use case and feed them back into the AI operating model. This creates a cycle of faster deployment, stronger confidence, and better returns over time.

In this way, trust is not an input to AI delivery or a final approval step. It’s an outcome that compounds as organisations reuse what works, strengthen what doesn’t, and continually improve the operating model that supports AI at scale.

Authors and contributors

Nicola Costello

Partner, Digital and AI Trust Leader, PwC Australia

nicola.costello@au.pwc.com

Ilana Golbin Blumenfeld

Partner, Trust AI Co-Leader, PwC US

ilana.a.golbin@pwc.com

Luke Vilain

Director, AI Trust and Governance, PwC UK

luke.vilain@pwc.com

Thank you

Access the online version