

THE AI INSTITUTE / OPERATING INTELLIGENCE

EXECUTIVE SYNTHESIS · 28 SEP / 2026



Make AI Forecasts Earn Your Commitment

AI can help you explore the future. Before its predictions shape a business plan, require a baseline, a dated evidence record and a reason to change course.

Boards, C-suite and senior management

Research period: September 2026 – September 2026

First published: 28 September 2026

Current to 28 September 2026

PUBLICATION RECORD

An Institute thesis and decision framework

Use AI to widen the possibilities you consider; let tested evidence determine the commitments you make.

WHAT LEADERS SHOULD TAKE FROM THIS PAPER

- | | |
|----|---|
| 01 | Keep scenarios, forecasts and commitments separate. |
| 02 | Compare AI forecasts with a simple, relevant baseline. |
| 03 | Record what changed the prediction, not just its latest answer. |
| 04 | Make local evidence and review triggers part of every consequential decision. |

Research method

Original Institute synthesis of two new research preprints, provider-produced scenario research, an official regional investment statement and the Commission’s implementation guidance. The latest visible arXiv release remains 25 September. Research results are attributed and bounded; the proposed decision discipline is Institute analysis, not an experimentally validated management intervention.

First published 28 September 2026 · Updated 28 September 2026 · Research period September 2026 – September 2026 · Version 1.0 · Suggested citation: The AI Institute, *Make AI Forecasts Earn Your Commitment* (2026).

CONTENTS

Inside this paper

01	A persuasive future is not yet a business case The decision
02	Separate the scenario, the forecast and the commitment Three different objects
03	Make the forecast beat something useful Evidence of improvement
04	Richer customer stories are not necessarily better predictions What remains unproven
05	Keep the reason for changing your mind The learning loop
06	Carry the method across borders, not the forecast Where the implication changes
07	Bring one forecast-linked decision into view This week’s board agenda
08	Faster learning is the near-term opportunity The standard
M	Method and limitations
R	References

How to use this brief

Start with the decision and use the framework to challenge the current plan. Figures retain their original populations and definitions. Institute analysis is labelled, and limitations are recorded before the references.

A persuasive future is not yet a business case

A board can receive three polished versions of the future before its next meeting. AI can gather news, explain competing trends and produce a plausible account of what happens next. That is useful preparation. The risk begins when the coherence of the account substitutes for evidence that its predictions are dependable.

This week's research offers a timely reason to examine that boundary. Forecast-Dojo, a new preprint submitted on 24 September, tests agents that revisit historical events using dated news. Research tools improved the tested models' forecasts, yet all 12 remained behind historical market forecasts on the reported measures. Its evaluation covered 230 held-out events, not business plans or enterprise demand forecasts.¹

The Institute's recommendation is to separate the work AI can make cheaper from the decision it has not yet earned. Use it to search for contrary evidence, articulate alternative explanations and maintain a record of changing assumptions. Before a prediction drives hiring, inventory or expansion, require a comparison with the method the organisation would otherwise use.

That is not a recommendation to wait for certainty. Management already acts under uncertainty. It is a recommendation to make the uncertainty visible before the commitment, when it can still affect the decision rather than become an explanation after the result.

Separate the scenario, the forecast and the commitment

A scenario asks what could follow if a set of assumptions holds. A forecast assigns an expectation to a defined outcome over a defined period. A commitment is an action with consequences: signing a contract, opening a service or changing a staffing plan. A useful scenario can be poor evidence for a specific forecast; a reasonable forecast can still be insufficient justification for an irreversible commitment.

Anthropic's September economic scenario explorer illustrates why the distinction matters. It models possible US outcomes under assumptions about AI capabilities, adoption and labour adjustment. It explicitly omits forces including policy responses, business cycles and financial-market disruptions. This is provider-produced scenario research, not a prediction of every economy or a staffing forecast for an individual employer.²

Ask the author of an AI-assisted board paper to label each important statement: observed fact, model assumption, estimated outcome or proposed action. These labels need not make the paper longer. They can replace the false precision of an unsupported central estimate with a short explanation of what management knows and what it is choosing to believe.

For example, an illustrative service expansion might look attractive if demand grows while delivery costs fall. Those are two assumptions, not one conclusion. The board can authorise a limited customer test while withholding a long property commitment. The scenario has helped design a decision without being mistaken for proof that the favourable future will arrive.

Make the forecast beat something useful

Choose the comparison before reviewing the AI result. Depending on the decision, that could be the previous planning method, a seasonal forecast, an unchanged-demand assumption or a qualified analyst's estimate. The comparison must face the same question and information cutoff. Otherwise, a system with fresher inputs may appear smarter when it simply knows more.

The simplest method deserves a fair trial. If a stable seasonal pattern already answers the question, a complex agent should justify its additional cost and variability. If the decision depends on rapidly changing public evidence, research assistance may be more useful. These are proposed test conditions, not findings about which method will win.

The operating test should be prospective where possible. Record the prediction before the outcome is known and preserve unsuccessful predictions alongside successful ones. If a historical replay is used first, check whether later information could have entered the data or the model. A realistic-looking backtest is still a backtest.

Do not measure only whether the most likely outcome occurred. Also examine how confidence behaved and what decisions followed. A forecast that repeatedly assigns high confidence to wrong outcomes may be more dangerous than a cautious one with similar headline accuracy. Finance and operations should specify the cost of acting too early, too late and on the wrong signal.

Richer customer stories are not necessarily better predictions

AI may also make it easier to collect and simulate customer views. A new pre-registered study involving 317 participants and three industry partners compared AI-moderated, human-moderated and static interviews. Digital twins built from AI interviews predicted held-out responses better than demographics-only personas, but not better quantitatively than twins built from static interviews. The human-moderated arm was small, with 24 participants.³

The implication is narrower than either a promise to replace customer research or a reason to reject it. Management should distinguish discovering a need from predicting a purchase. A richer interview can help generate a product hypothesis while leaving the commercial forecast unproven.

In an illustrative product launch, use AI-assisted interviews to identify objections and design alternatives. Then test those alternatives with the actual intended customers. Keep the research population, language and purchasing conditions attached to the result. Do not let a simulated customer's persuasive explanation become evidence of willingness to pay.

This also changes procurement. A supplier demonstrating convincing simulated conversations should be asked for held-out predictive tests relevant to the intended use. Buying a tool for exploratory research is a different decision from buying one to determine a revenue plan. The contract and acceptance test should reflect which job the organisation is actually commissioning.

Keep the reason for changing your mind

A forecast should have a history, not just a latest version. Record the outcome being predicted, the evidence date, the current estimate, the previous estimate and the reason for any material change. Distinguish new external evidence from a model update, a revised prompt or a management judgement.

Forecast-Dojo's persistent notebook reduced research cost more consistently than it improved forecast quality. Its time-bounded design is useful research infrastructure, but selected prediction-market events and reported model cutoffs do not prove transfer to an organisation's own decisions.¹ The management opportunity is to make revisions inspectable, not to assume that adding memory makes a system more trustworthy.

Before a test starts, decide what evidence would prompt a review. A change in customer orders might matter; another commentary article repeating an existing claim might not. Treat review triggers as requests to reconsider, not automatic instructions to invest or withdraw. The business owner remains responsible for the action.

A useful record should preserve disagreement. If sales expects a surge and operations expects a capacity constraint, do not ask the model to average the narratives into artificial consensus. Ask which observable fact would distinguish them, who can obtain it and when. AI can help organise that inquiry without resolving it by assertion.

Carry the method across borders, not the forecast

The discipline travels more readily than the numbers. US economic scenarios do not establish labour adjustment elsewhere. Prediction-market questions do not represent every firm's customers. A forecast built from abundant English-language reporting may miss local procurement cycles, informal activity or changes that are visible only in another language.

Regional infrastructure announcements also need a status label. Microsoft's 23 September Middle East statement describes planned investment and resilience commitments across named markets. It provides evidence of supplier intent, not proof that every promised capability is already available to every customer. A plan dependent on future capacity should retain that dependency.⁴

Legal assumptions deserve the same treatment. The European Commission distinguishes the AI Act's requirements from voluntary compliance tools and phased application dates. Its current implementation page is a starting point for an EU-specific review, not permission to assume that one policy applies worldwide.⁵ This article does not offer a legal classification of forecasting systems.

For operations in Asia-Pacific, Africa or Latin America, ask local teams to identify missing evidence before importing a central forecast. Which customer groups are absent? Which data cannot be used? How would connectivity, financing or institutional capacity change the action? These are diligence questions, not claims that every organisation in a region faces the same constraint.

Keep any resulting difference in the recommendation. A global board may reasonably authorise a test in one market and defer it in another because the evidence differs. Consistency means applying the same standard of justification, not forcing identical commitments.

Bring one forecast-linked decision into view

Ask management to select one consequential decision already in progress. The strategy lead should state the outcome, time horizon and plausible alternatives. The finance lead should identify the commitment at stake and the loss the organisation can tolerate. The operating owner should explain what evidence can be collected before the next irreversible step.

The data or research lead should then propose the comparison and the measurement plan. Record the source cutoff and keep the forecast version that informed the decision. Where customer or employee information is involved, the relevant privacy and legal owners should review its permitted use. A board does not need to choose a model to insist on these conditions.

Use the next review to examine a decision record, not another demonstration: what did we expect, what changed, what did we do and what happened? If the outcome has not resolved, say so. An attractive intermediate signal should not be substituted for the result that originally justified the test.

Build the ability to retain assumptions and revisions. Buy research assistance where it improves the work at an acceptable cost. Test predictions on the actual decision. Watch whether performance survives new periods and markets. Ignore claims that fluent explanations, more tool calls or more elaborate simulations alone demonstrate forecasting value.

EXHIBIT / FROM A POSSIBLE FUTURE TO AN ACCOUNTABLE DECISION

FRAME	TEST	REVISE	ACT
Define the outcome State the horizon and assumptions.	Use a baseline Compare with the current method.	Keep the evidence Record why the estimate changed.	Limit commitment Name the owner and review trigger.
Institute proposed discipline	Institute proposed discipline	Institute proposed discipline	Institute proposed discipline

Faster learning is the near-term opportunity

The research does not justify handing a business plan to an autonomous forecaster. It does justify testing whether AI can make the organisation better at finding evidence, exposing assumptions and revisiting decisions. Those benefits can be valuable even when the final forecast remains imperfect.

The board's standard should be simple: show the comparison, preserve the uncertainty and explain what would change the action. AI can help a leadership team consider more futures. It should not make an untested future feel inevitable.

RESEARCH RECORD

Method and limitations

Method

Original Institute synthesis of two new research preprints, provider-produced scenario research, an official regional investment statement and the Commission's implementation guidance. The latest visible arXiv release remains 25 September. Research results are attributed and bounded; the proposed decision discipline is Institute analysis, not an experimentally validated management intervention.

Limitations

No enterprise forecasting return, global productivity rate or representative regional ranking is established. Historical prediction-market evaluation differs from live commercial forecasting. Customer-research results are sample-specific. Vendor publications carry commercial interests. Legal obligations require use-specific advice. No guest participation or partnership is implied.



Sources and limitations

Source facts link to the original publication. Institute analysis reconciles definitions and turns the material into a management proposition. Frameworks are decision aids, not claims of universal causality. Read every important statistic with its population, date, definition and caveat.

Research cutoff: 28 September 2026. Review cycle: six months, or earlier after a material change in the underlying facts, standards or policy.

REFERENCES

Evidence behind the thesis

01

Ye et al.: Forecast-Dojo

New preprint submitted 24 September 2026, listed 25 September. Evaluation: 230 held-out events, March–May 2026; 12 models. Methods and cutoff checks: <https://arxiv.org/html/2609.28876v1>. Not peer-reviewed deployment evidence.

02

Anthropic: Scenarios for our Economic Future

Version 1.0, September 2026; checked 28 September. US scenario model, not a global forecast. Provider-produced; omitted mechanisms explicitly acknowledged.

03

Deng et al.: AI-Moderated Interviews for Market Research and Digital Twins Calibration

New preprint submitted 24 September 2026. Pre-registered study: 317 participants, three partners; AI 139, human 24, static 154. Held-out responses to six marketing stimuli, not realised purchases.

04

Microsoft: Middle East technology, resilience and people commitments

Brad Smith, 23 September 2026. Supplier statement covering Kuwait, Qatar, Saudi Arabia and UAE; distinguish future commitments from deployed capacity.

05

European Commission: AI Act implementation

Current official implementation page, checked 28 September 2026. Differentiates legal requirements, voluntary tools and phased application; no use-specific legal advice.

THE AI INSTITUTE

Make the decision.
Then measure what changes.

Research · Readiness · Capability

theai.institute