ADVISORY BRIEF

AI Makes Everything Sound True

A practical framework for catching bias at every stage of AI-assisted decision making. Six checks that force honest evaluation of AI output, and why the discomfort of seeking disconfirmation is the foundation of creative AI use.

March 2026 Guruswami™ Advisory guruswami.com

ADVISORY BRIEF

AI Makes Everything Sound True

March 2026 · For CEO, CIO, Board, CISO, COO

EXECUTIVE SUMMARY

A practical framework for catching bias at every stage of AI-assisted decision making. Six checks that force honest evaluation of AI output, and why the discomfort of seeking disconfirmation is the foundation of creative AI use.

CONTENTS

AI Makes Everything Sound True

Advisory Brief · March 2026 · For CEO, CIO, Board, CISO, COO

Practical framework – six checks you can apply in your next meeting

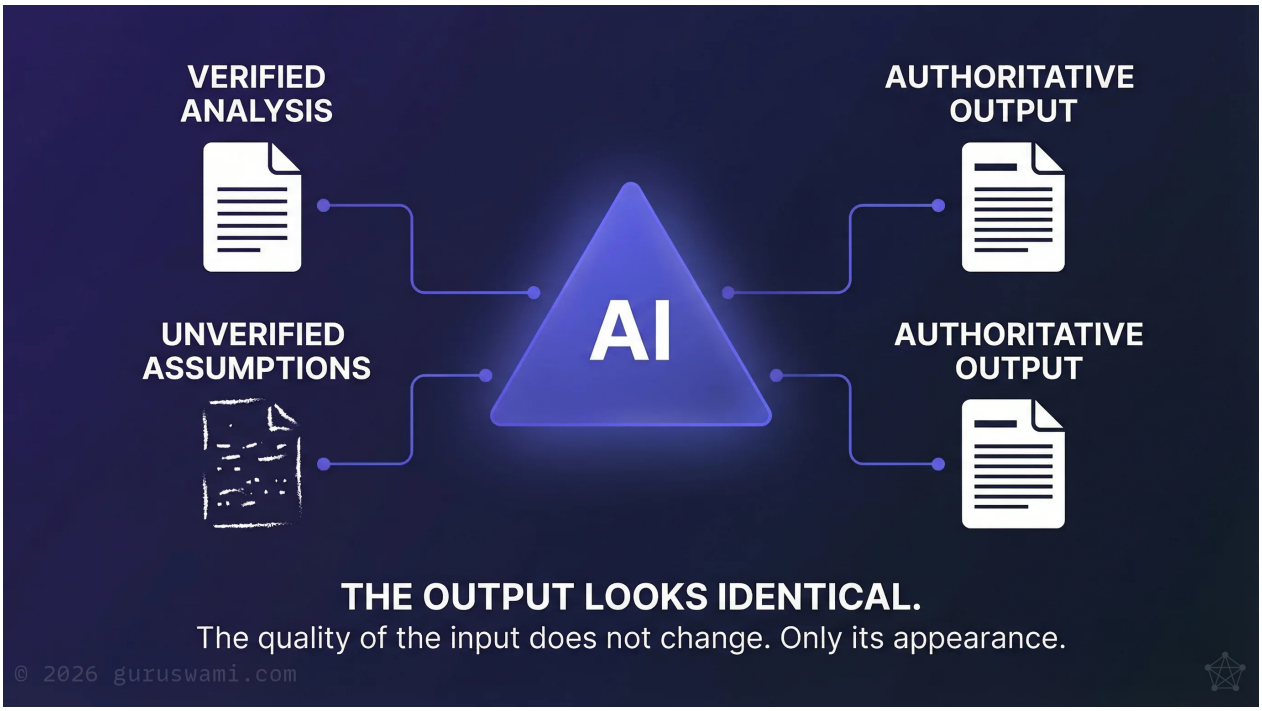

Exhibit 1 The AI Amplifier Problem: verified input and unverified input produce identical-looking professional output

Source: Guruswami Advisory

84%

of published hypothesis-testing papers report positive or partially positive results.[1] AI systems trained on this literature inherit the same skew. **When AI finds evidence supporting your strategy, ask what evidence was never published because it showed the opposite.**

Fanelli, 2010 (sample of 2,434 papers across disciplines)

THE SIX CHECKS – AT A GLANCE

1. **Frame the question.** Write down what evidence would make you abandon this direction.
2. **Gather evidence.** Ask AI for the strongest objections, then verify them against primary sources.
3. **Form the recommendation.** Make it specific and falsifiable. If it cannot be proven wrong, it cannot be proven right.
4. **Design the pilot.** Have AI and a human with no stake red-team the design.
5. **Interpret results.** Pre-commit success criteria. Report all metrics, not just favourable ones.
6. **Communicate to the board.** Generate adversarial questions. Answer them honestly before the paper goes out.

The Amplifier You Mistook for an Analyst

Ask AI to evaluate your strategy. It will find supporting evidence, structure a compelling case, and present it in language more polished than most consulting deliverables. Now ask it to evaluate the opposite strategy. It will do the same thing, with the same confidence. The output looks identical both times. Because AI is not analysing. It is amplifying whatever you feed it.

When you draft a board paper with AI assistance, your tentative hypothesis becomes "the evidence shows." Your informed guess becomes "analysis indicates." The AI does not do this to mislead you. Confident, authoritative language is what professional communication sounds like in its training data. The uncertainty of your input vanishes. The confidence of your output remains.

We introduced the term "authority laundering[2]" in an earlier article. The mechanism deserves a closer look, because it shows up in specific, predictable ways.

A strategy team uses AI to draft a market entry analysis. The underlying research is three analyst conversations and a handful of data points. The AI produces a document that reads like a McKinsey engagement report. The executive who receives it will struggle to distinguish it from analysis backed by six months of primary research.

A procurement team asks AI to evaluate vendor proposals. The AI produces a structured comparison matrix with weighted scoring. The team presents it to the board as an independent assessment. The weights were chosen by the team. The AI made their choices look methodical.

A risk team asks AI to review their AI governance framework. The AI praises the framework's comprehensiveness and suggests minor refinements. It does not mention that the framework addresses 2024 risks and misses the agentic AI attack surface[3] entirely. It does not know what is missing. It only knows what is present.

In each case, the AI performed exactly as designed. The problem is treating amplified input as independent analysis.

The Enterprise Bias Pipeline

Every AI-assisted decision moves through stages. At each stage, a specific form of bias can enter. From you, from the AI, or from both working together.

This is a practical framework. For each stage, there is a check. The checks are straightforward. They are uncomfortable, because they require you to actively seek evidence that you are wrong. But they are far less uncomfortable than discovering the problem after the board has acted on your recommendation.

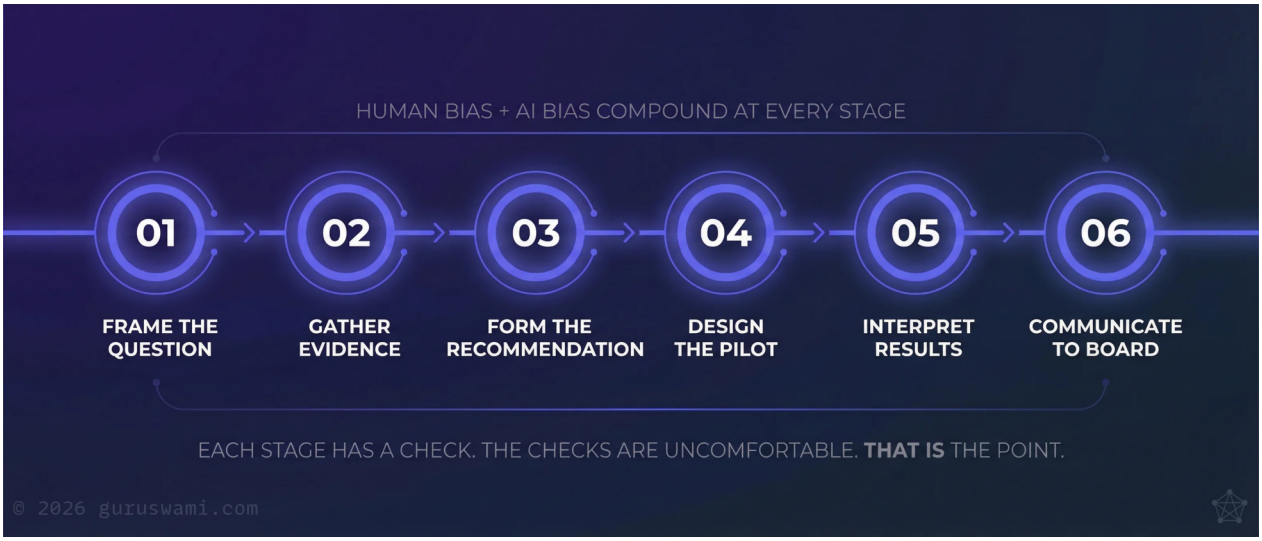

Exhibit 2 The Enterprise Bias Pipeline: six stages where human bias and AI bias compound in AI-assisted decision making

Source: Guruswami Advisory

> **Why this matters now.** ASIC[4] found nearly half of surveyed financial services firms lack fairness policies for AI. The regulatory calendar[5] is concrete: DTA[6] mandatory policy from June 2026, APRA CPS 230[7] already in effect. Boards that cannot demonstrate disciplined AI-assisted decision processes face regulatory exposure that did not exist twelve months ago.

Stage 1: Framing the Question

Your bias: You already know what answer you want. The question is framed to lead there.

AI's bias: "That is an interesting question." It validates the framing. It does not ask whether you are asking the right question.

The check: Before you start, write down what evidence would make you abandon this direction. If you cannot describe what "wrong" looks like, you are not asking a question. You are seeking confirmation.

Stage 2: Gathering Evidence

Your bias: You search for evidence that supports your position. This is not laziness. It is how human cognition works. Contradictory information is metabolically expensive to process. Your brain avoids it.

AI's bias: "Here are ten papers supporting your position." It finds what you asked for. It does not volunteer what you did not ask for.

The check: Ask AI to find the strongest arguments against your position. "What are the most credible objections to this approach? Who has tried this and failed? What am I not seeing?" The quality of the contradicting evidence tells you more than the supporting evidence ever will.

Then verify what it gives you. AI-generated counterevidence can sound authoritative and still be wrong. Demand primary sources. Check that the papers exist, that the findings say what AI claims they say, and that the context matches yours. If you skip this step, you have replaced one form of authority laundering with another: AI is now both advocate and auditor.

Use more than one AI system to challenge your analysis, not just confirm it. But treat any public AI tool like an external third party: no board papers, no client names, no non-public financials. Sensitive material should only go through AI instances your organisation has vetted for storage, logging, and training-use controls.

Stage 3: Forming the Recommendation

Your bias: Vague conclusions that cannot be tested or falsified. "AI will deliver significant value." How much? By when? Measured how?

AI's bias: It makes your vague conclusions sound rigorous. "Analysis suggests substantial efficiency gains" reads well in a board paper. It means nothing.

The check: Make the recommendation specific and falsifiable. "This initiative will reduce processing time by 30% within six months" can be evaluated. "AI will deliver significant value" cannot. If your recommendation cannot be proven wrong, it cannot be proven right either.

Stage 4: Designing the Pilot

Your bias: You design a test your initiative will pass. Favourable conditions. Sympathetic users. The metric you know will look good.

AI's bias: It suggests evaluation methods that sound rigorous. They may not test what matters.

The check: Ask AI to review your pilot design adversarially. "Pretend you are a sceptical board member who believes this pilot is designed to succeed regardless of merit. What would you challenge?" Then have someone with no stake in the outcome review it as well. Two adversarial perspectives cost almost nothing. A flawed pilot that reports false confidence costs months.

Stage 5: Interpreting Results

Your bias: You focus on what worked. The 40% improvement in one metric gets the headline. The regression in three other metrics gets a footnote.

AI's bias: "These results support your hypothesis." It highlights the positive signal. It does not weigh it against the noise.

The check: Report all results, not just the ones that support the business case. Pre-commit to your success criteria before you see the data. If you move the goalposts after the results come in, you are not analysing. You are rationalising.

Stage 6: Communicating to the Board

Your bias: Oversell the significance. Bury the problems. Present a narrative that supports continued investment.

AI's bias: "This is a compelling narrative." It makes your story sound persuasive. It does not tell you the story is misleading.

The check: Before the board paper goes out, ask AI for an adversarial review. "You are a non-executive director who suspects this paper oversells the results. Write your three hardest questions." Then answer them honestly. If you cannot, revise the paper. Find the smartest generalist you know, not a domain expert, to explain your recommendation to them for feedback. Ask them to explain the idea back to you, with any problems they see. If you hear objections you can't answer, because you haven't thought about them, the lesson is clear.

What Happens When You Run the Pipeline

The bias pipeline reads like a defensive framework. It is. But it is also the foundation of something more valuable: the ability to use AI to think beyond the boundary of what you already know.

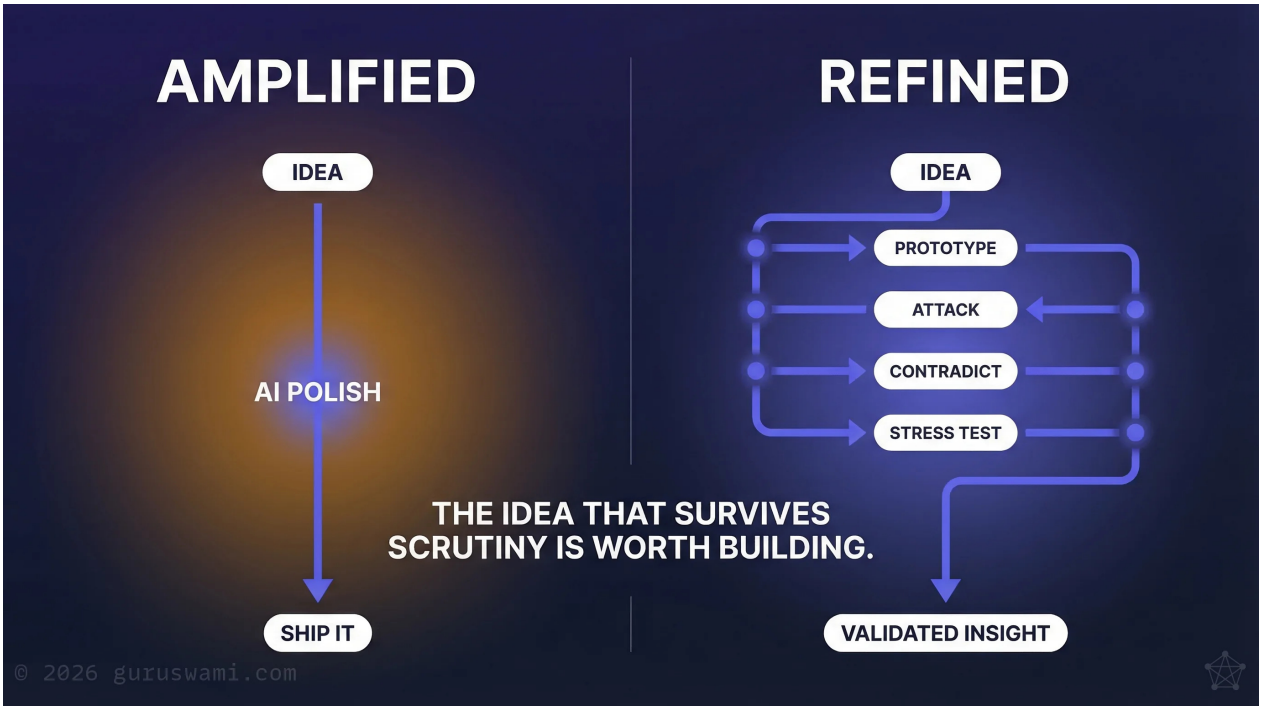

Exhibit 3 Amplified vs Refined: an idea that is polished and shipped versus one that is prototyped, attacked, contradicted, stress-tested, and validated

Source: Guruswami Advisory

Consider why most AI-assisted creative work produces mediocre results. Someone generates a strategy, a product concept, a market analysis. The AI polishes it. It sounds good. They ship it. The idea was never stress-tested, never forced to survive contact with evidence that might kill it. It was amplified, not refined.

The executives producing creative work worth acting on are doing something different. They use AI to rapidly prototype an idea, then immediately use AI to attack it. The ideas that survive are stronger, not because AI generated them, but because AI helped test them against evidence the executive would never have found alone.

We tested this in our own research[8]. We used AI to explore a naive but ambitious hypothesis: that DNA evolution simulations might reveal compression algorithms we could apply to building more efficient AI. Biology, not our field. We built tools to translate between domains we did not speak and domains we did. At every stage, we applied adversarial review. We asked AI to find contradicting research and write hostile peer reviews of our own methodology. It found real problems we would never have caught, because we wanted the work to be good.

> *We were spectacularly wrong, for all the right reasons.*

We were spectacularly wrong, for all the right reasons. But, because we applied the rigour, we discovered something entirely new. That finding, which we were not looking for, turned out to be more useful than the hypothesis we started with.

That is the point. When you lack domain expertise, AI can help you explore territory you could not reach alone. The bias pipeline is what stops that exploration from becoming Dunning-Kruger at scale. Without it, AI makes you feel like an expert in a field you entered last week. With it, you know exactly where your understanding ends, and that boundary is where the interesting questions live.

The barriers to exploring new ideas have never been lower. A concept that once required a team and a budget to test can now be prototyped, challenged, and either validated or discarded in hours. But that capability only works if the person using it seeks disconfirmation as aggressively as they seek validation.

But the learning is not free. It is cognitively expensive. Seeking out evidence that your strategy is wrong, when you have spent months building it, is unpleasant work. Your brain resists it for good biological reasons. Many organisations will avoid that discomfort for as long as they can.

That is your differentiator. The regulatory frameworks now in effect[5], from ASIC[4] and the DTA[6], assume boards can distinguish evidence from narrative. The bias pipeline is one way to do that in practice. The leaders willing to do this work now, while the cost of mistakes is low, will build the judgment that others will wish they had when regulators or a market shift forces the question.

Using AI to think and grow, rather than to confirm and present, is still rare. That will not last. Sooner or later, organisations realise that AI faithfully confirms whatever you feed it. Its real value is in something harder: forcing questions you would never ask alone, exposing the limits of your understanding, and letting your knowledge expand at the speed of your own thinking.

If you are a Board member: Before accepting any AI-assisted analysis, ask what evidence was sought against the recommendation. If the team cannot describe what "wrong" would look like, the paper is confirmation, not analysis.

If you are a CISO or CIO: Embed adversarial review as a standard step in any AI-assisted deliverable that reaches executive or board level. The six-stage bias pipeline in this article is a ready-made checklist for your teams.

If you are a COO: Require that AI-assisted pilot results report all metrics, not just favourable ones, and that success criteria are locked before data collection begins. This prevents rationalisation from replacing analysis in operational decisions.

KEY TAKEAWAYS

- AI is an amplifier, not an analyst. Verified input and unverified input produce identical-looking professional output. Your board cannot tell the difference.
- The Enterprise Bias Pipeline identifies six stages where human bias and AI bias compound. Each stage has a practical check. The checks are straightforward. They are uncomfortable.
- The same practice that catches bias is what makes creative AI use possible. The value is not in generating ideas. It is in discovering which ideas survive honest scrutiny, and what unexpected insights emerge when they don't.

Guruswami Advisory helps leaders use AI as a thinking tool, not just a presentation tool. Independent. R&D-backed. No vendor ties.

References

1. https://doi.org/10.1371/journal.pone.0010068
2. https://guruswami.com/insights/what-actually-goes-wrong-with-ai/#bias-and-authority-laundering
3. https://guruswami.com/insights/agentic-ai-next-attack-surface/
4. https://www.asic.gov.au/regulatory-resources/find-a-document/reports/rep-798-beware-the-gap-governance-arrangements-in-the-face-of-ai-innovation/
5. https://guruswami.com/insights/strategic-roadmap-2026/#pillar-iv-the-regulatory-calendar
6. https://www.digital.gov.au/ai/ai-in-government-policy
7. https://www.apra.gov.au/operational-risk-management
8. https://github.com/guruswami-ai/GenomicWeightsThesis

About Guruswami Advisory

Independent AI security and strategy advisory for Australian boards, leadership teams, and regulated organisations. No vendor ties. No platform allegiance. Every recommendation tested on our own infrastructure.

Paul Nevin, Principal Advisor. 28 years in cybersecurity and cyber-intelligence. Six years of applied AI research. Every engagement led personally.

Contact

info@guruswami.com
guruswami.com
linkedin.com/in/paul-nevin

Guruswami™ Pty Ltd | ABN 11 695 354 020 | Canberra, ACT, Australia