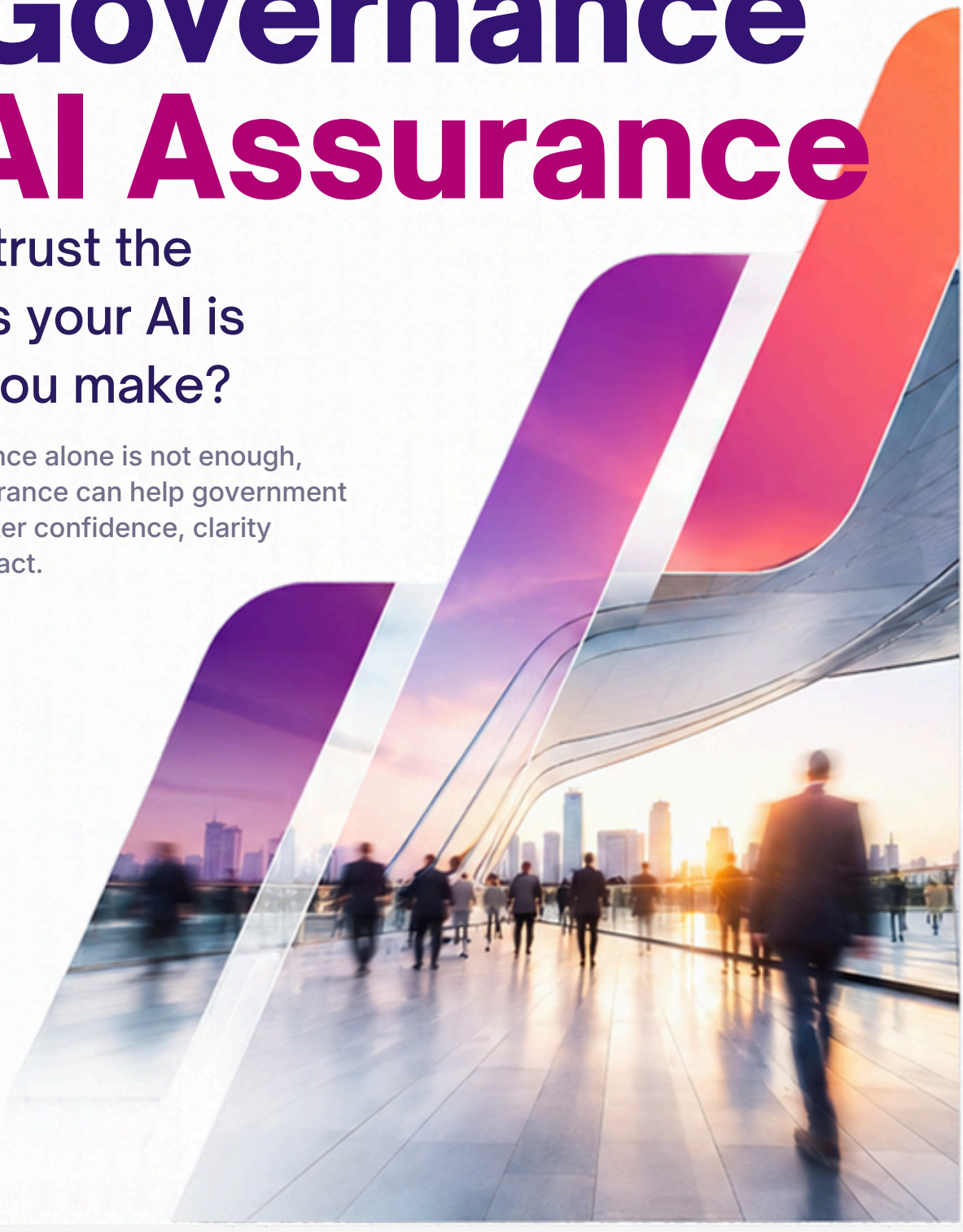
THOUGHT LEADERSHIP ARTICLE

From AI Governance to AI Assurance

Can you trust the decisions your AI is helping you make?

Why AI governance alone is not enough, and how AI assurance can help government use AI with greater confidence, clarity and positive impact.



THOUGHT LEADERSHIP

Opening

As AI becomes embedded in organisational decision-making, the question is no longer whether AI can produce useful outputs. It is whether organisations can explain, verify and defend the decisions those outputs influence.

Consider two increasingly familiar situations:

A government department introduces an AI assistant to help employees interpret organisational policies. Staff can ask questions in plain language and receive immediate, convincing answers, complete with references to internal documents. But some of those documents have been superseded. Others contain conflicting guidance. The AI may accurately summarise what it finds without recognising that the information has never been formally reconciled or approved.

An agency uses machine learning to forecast future demand for public services and inform significant investment decisions. The model performs well against historical data and predicts a substantial increase in demand in a particular region. But decision-makers struggle to understand what is driving the forecast, whether the underlying relationships remain valid, or how proposed interventions might change the outcome.

Both systems may deliver considerable value. Both may also produce outputs that appear more authoritative than the evidence warrants.



The challenge is no longer simply whether AI can generate useful answers. It is whether organisations can explain, verify and defend the decisions those answers influence.

Governance establishes the rules. Assurance demonstrates they work.

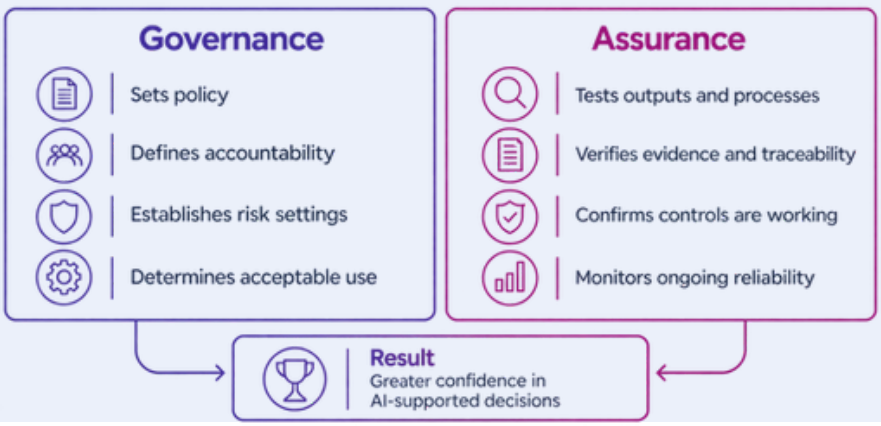
Much of the initial conversation around responsible AI has focused on governance: establishing policies, defining accountability, assessing risks and setting boundaries for appropriate use.

These foundations remain essential. But governance alone does not demonstrate that an individual AI application is producing reliable information or supporting sound decisions.

An organisation may have comprehensive AI policies, documented controls and clearly assigned responsibilities, yet still lack evidence that these arrangements are effective in practice.

Can it demonstrate that an AI recommendation is supported by authoritative information, that appropriate safeguards have operated, or that a previously reliable model continues to perform as expected?

Governance establishes what should happen. Assurance provides the evidence that it is happening.



This is where assurance becomes critical

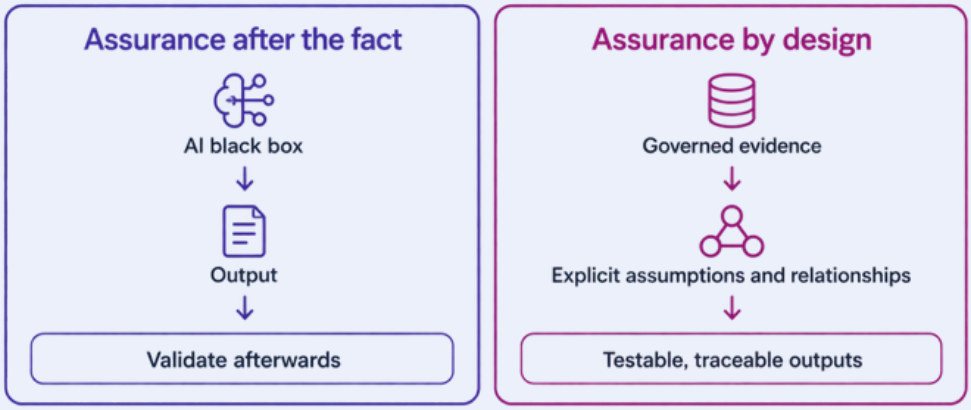
AI assurance involves establishing evidence that systems and their outputs are sufficiently reliable for their intended purpose, supported by appropriate testing, controls, oversight and ongoing monitoring.

For consequential decisions, simply reviewing the final AI-generated answer is not enough. Organisations need visibility into the foundations on which that answer was produced.

Assurance should be designed in, not added afterwards

Traditional approaches to technology assurance often emphasise testing outputs, measuring performance and monitoring for errors. These activities remain important, but AI presents an opportunity to build greater transparency and control into the systems themselves.

Two areas illustrate how assurance can be designed into AI-enabled systems from the outset: governing the organisational knowledge that informs AI-generated responses, and making the assumptions and causal relationships behind predictive models explicit and testable. Both approaches are explored below.



Can we trust what AI knows?

Returning to our policy assistant, the central problem may not be the AI's ability to retrieve or summarise documents. It may be the organisation's inability to establish which information should be considered authoritative.

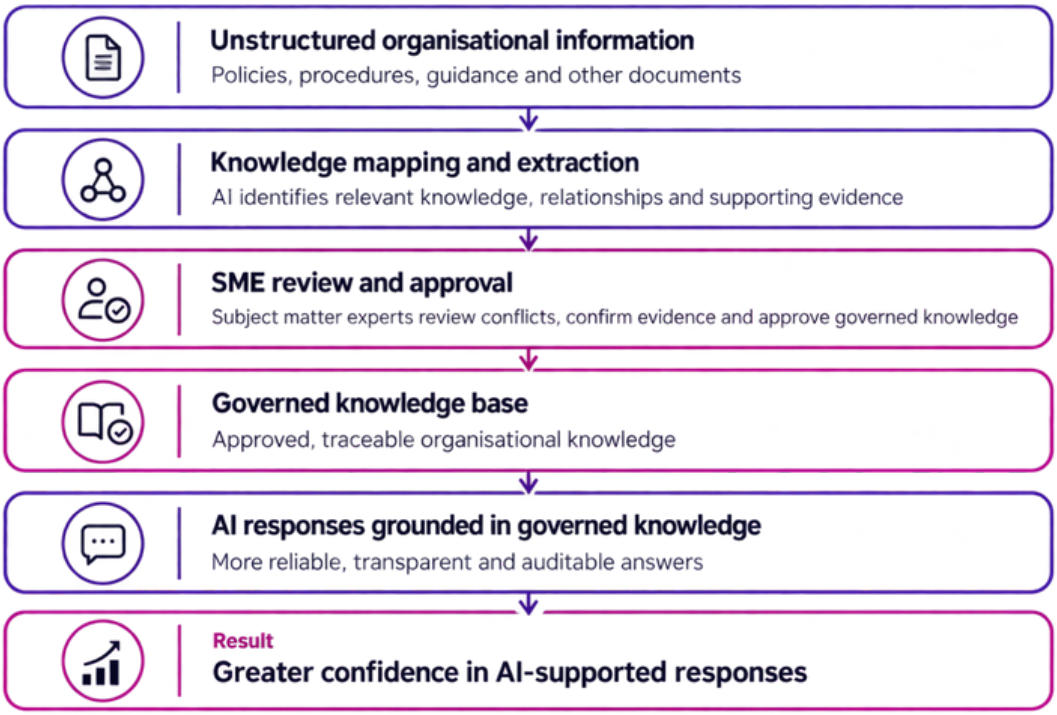
Many enterprise AI assistants use retrieval-augmented generation, drawing on internal documents to inform their responses. While this improves access to organisational information, it introduces an important assurance challenge: the AI may retrieve a document perfectly without knowing whether its contents remain valid, whether another document supersedes it, or whether the organisation has formally approved the interpretation being presented.

Simply retrieving information from an internal system does not establish that it is authoritative. Even references to source documents provide limited assurance if the information itself has not been governed.

A more assured approach begins by governing the knowledge itself.

At OCM, we have developed GRAIL (Governance Retrieval Augmented Intelligence Layer) to address this challenge. GRAIL uses AI to identify knowledge and relationships within organisational information, while retaining traceability to supporting evidence. Subject matter experts review and approve the knowledge before it becomes part of a governed knowledge base.

Changes and conflicting evidence can be identified and referred back for review.



This introduces an important distinction: rather than allowing an AI assistant to independently interpret everything it retrieves, organisations can establish a controlled foundation of approved knowledge on which those assistants can draw.

By establishing which organisational knowledge is approved, preserving its supporting evidence and managing changes through defined governance processes, GRAIL is designed to provide a more reliable and auditable foundation for AI-generated responses.

Importantly, organisations gain greater control over the knowledge their AI systems rely upon, rather than leaving those judgments entirely to the AI itself.

Can we explain what AI predicts?

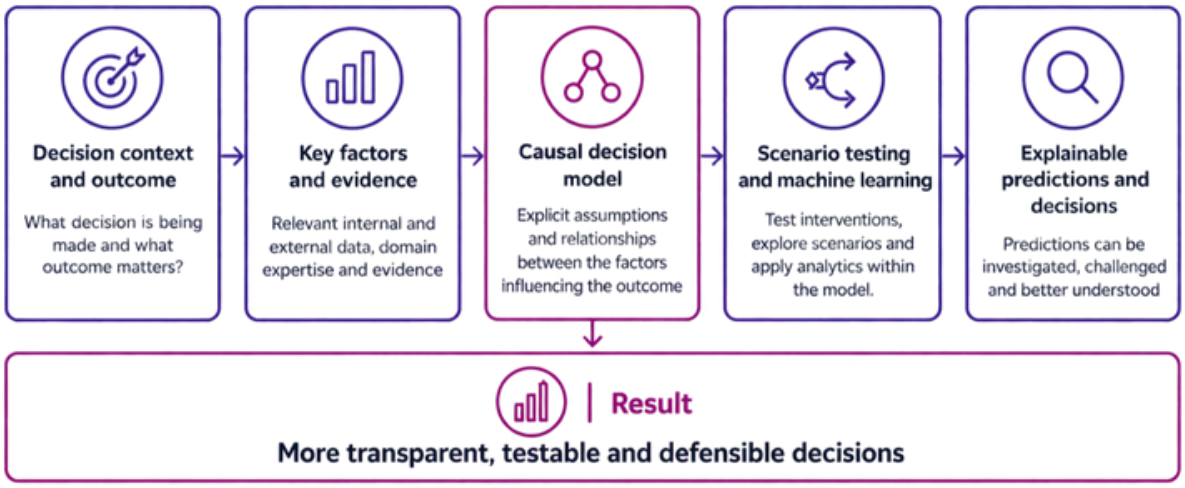
A different assurance challenge arises when machine learning informs forecasting, planning and investment decisions.

Traditional machine learning can be highly effective at identifying patterns and predicting outcomes. However, complex black-box models can offer limited insight into the underlying relationships driving those predictions.

This becomes especially problematic when circumstances change. A model that previously forecast service demand accurately may begin producing unexpected results, leaving decision-makers struggling to explain what has changed or whether an intervention will deliver the intended outcome.

At OCM, our Decision Intelligence approach addresses this challenge through causal modelling.

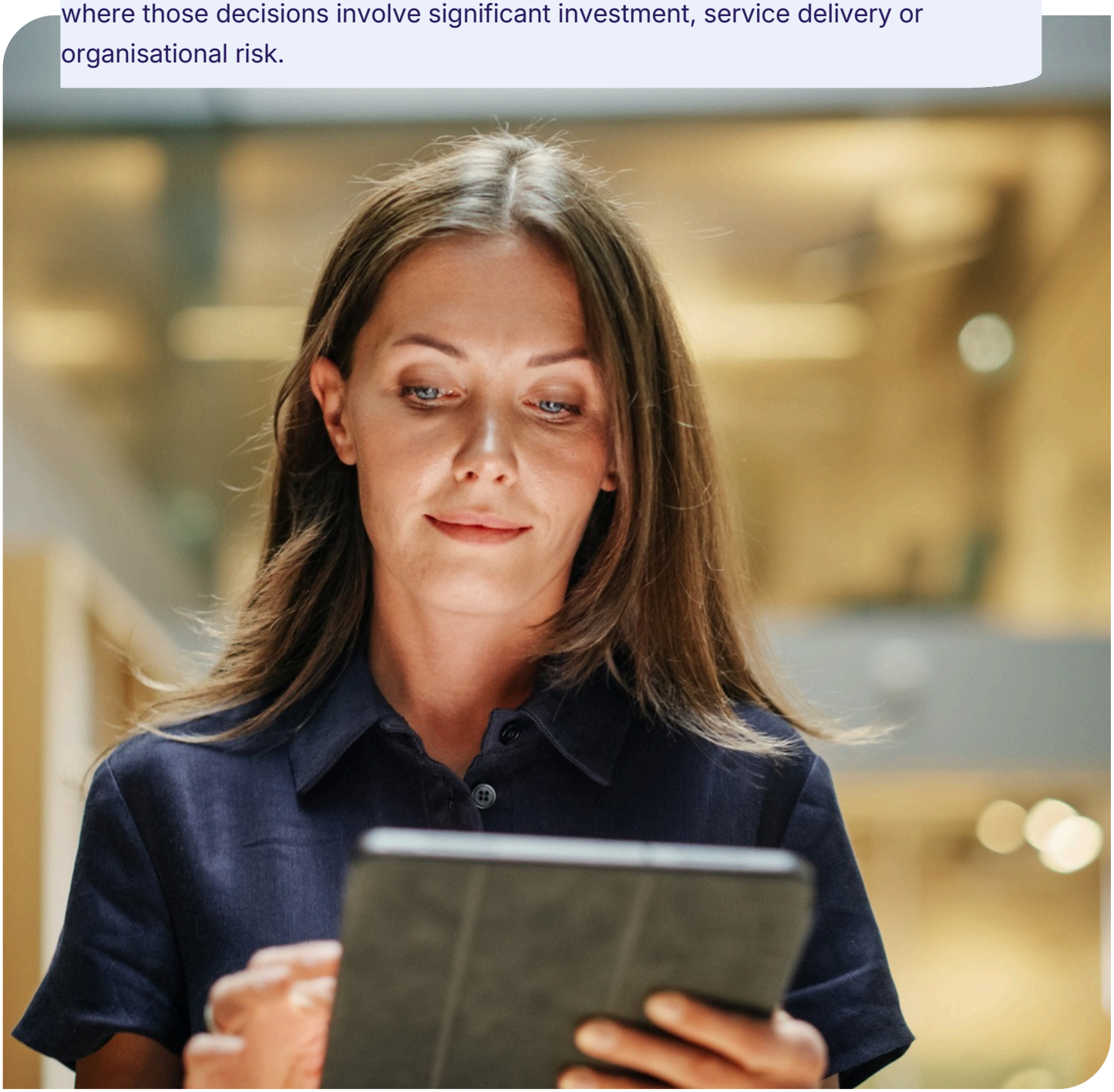
Rather than relying solely on statistical associations, we work with subject matter experts to explicitly model the factors influencing an outcome and the causal relationships between them. This establishes a transparent decision model in which assumptions can be examined, relationships tested against evidence, and potential interventions explored.



Machine learning can still be applied within this framework, but with a clearer understanding of the factors and relationships underpinning the decision.

From an assurance perspective, this represents a significant advantage over approaches where the relationships informing predictions are difficult to interpret. Decision-makers can examine the assumptions behind a forecast, test the evidence supporting key relationships and investigate why predictions change as circumstances evolve.

While causal modelling does not eliminate uncertainty, it provides a substantially more transparent and auditable foundation for AI-supported decisions, particularly where those decisions involve significant investment, service delivery or organisational risk.



Confidence must be proportionate to the decision

Not every use of AI warrants the same level of assurance. The controls appropriate for an assistant drafting routine correspondence are very different from those required for AI informing public funding allocations, procurement assessments or decisions affecting access to essential services.

As the consequences of an incorrect recommendation increase, so too should the depth of assurance. Higher-impact applications may require more rigorous testing, stronger evidence of reliability, greater traceability, independent review and clearly defined decision-making responsibilities.

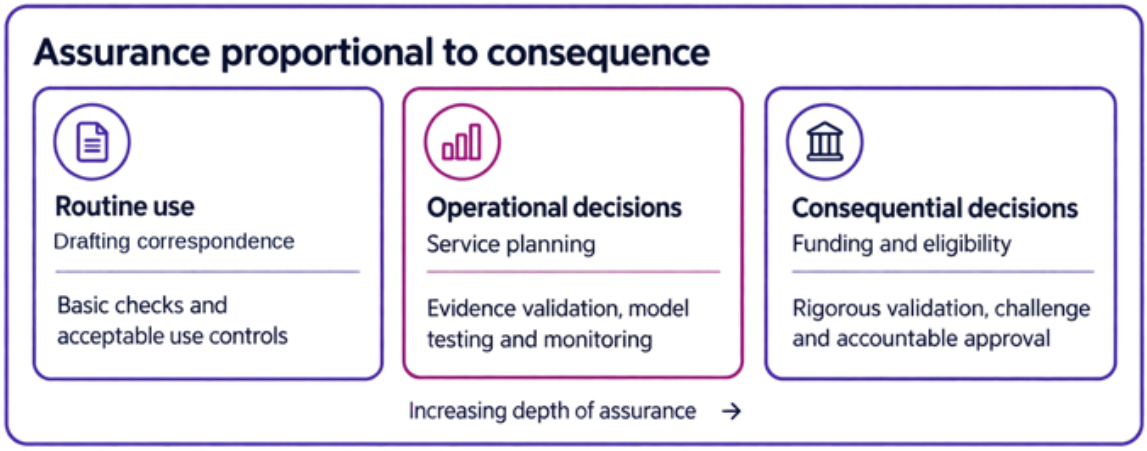
Human oversight is particularly important, but it must be meaningful. A person simply approving an AI recommendation does not provide effective assurance if they lack the information, expertise, time or authority to challenge it. Effective oversight means decision-makers can understand the supporting evidence, question the recommendation and reject or escalate it where necessary.

Assurance should enable innovation, not become another obstacle to it.

A successful AI pilot may demonstrate potential, but scaling that application requires confidence that it will remain reliable as information, models and operating conditions change.

By building assurance into the design of AI-enabled systems, organisations can establish clearer boundaries for experimentation, identify problems earlier and expand successful applications with greater confidence.

The objective is not to eliminate risk or demand certainty before proceeding. It is to make informed decisions about where AI can be trusted, where additional oversight is required and where the consequences warrant greater caution.



Five questions every leader should ask

01 **Evidence**
Do we know what information the AI relies on, and how its authority and accuracy are established?

02 **Explainability**
Can we understand and challenge the assumptions and relationships informing recommendations?

03 **Validation**
Has performance been tested against realistic scenarios, including unexpected conditions and failures?

04 **Accountability**
Who reviews consequential recommendations and makes the final decision?

05 **Ongoing assurance**
How are changes in information, assumptions, or performance identified and acted upon?

Digital confidence does not come from simply trusting AI. It comes from being able to demonstrate why that trust is justified.

Continue the conversation...

OCM works with organisations to strengthen AI governance, build assurance into AI-enabled systems and improve confidence in the decisions they support.

How confident is your organisation in the information, assumptions and evidence underpinning its AI-supported decisions?

Contact our team to explore what assurance by design could mean for your organisation.

Talk to us



Jonathan Cherian
Partner
Performance & Transformation
0412 116 474
jcherian@ocmau.com



Bryn Liepins
Executive Director
Digital & Data
0430 001 621
bliepins@ocmau.com

1300 882 633
talkwithus@ocmau.com
weareocm.com