Rethinking AI Governance in 2026

Critical questions, sleeper risks & the actions to take now

Tristan Marot, AI & Innovation Lawyer, Webber Wentzel Fusion

WEBBER WENTZEL FUSION

© 2026 Webber Wentzel. All rights reserved.

MONDAY

Drafts replies for the service team.

ADVISES

FRIDAY

Connected to the complaints database — can issue refunds.

ACTS

A MONTH LATER

Edits customer records and directs another agent to investigate disputes.

DELEGATES

The tool stayed the same. Its authority changed.

What could your AI do before anyone notices?

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

GENERATIVE → AGENTIC

The risk changes when AI can act.

AI that **says** things

- Draft
- Summarise
- Recommend

AI that **does** things

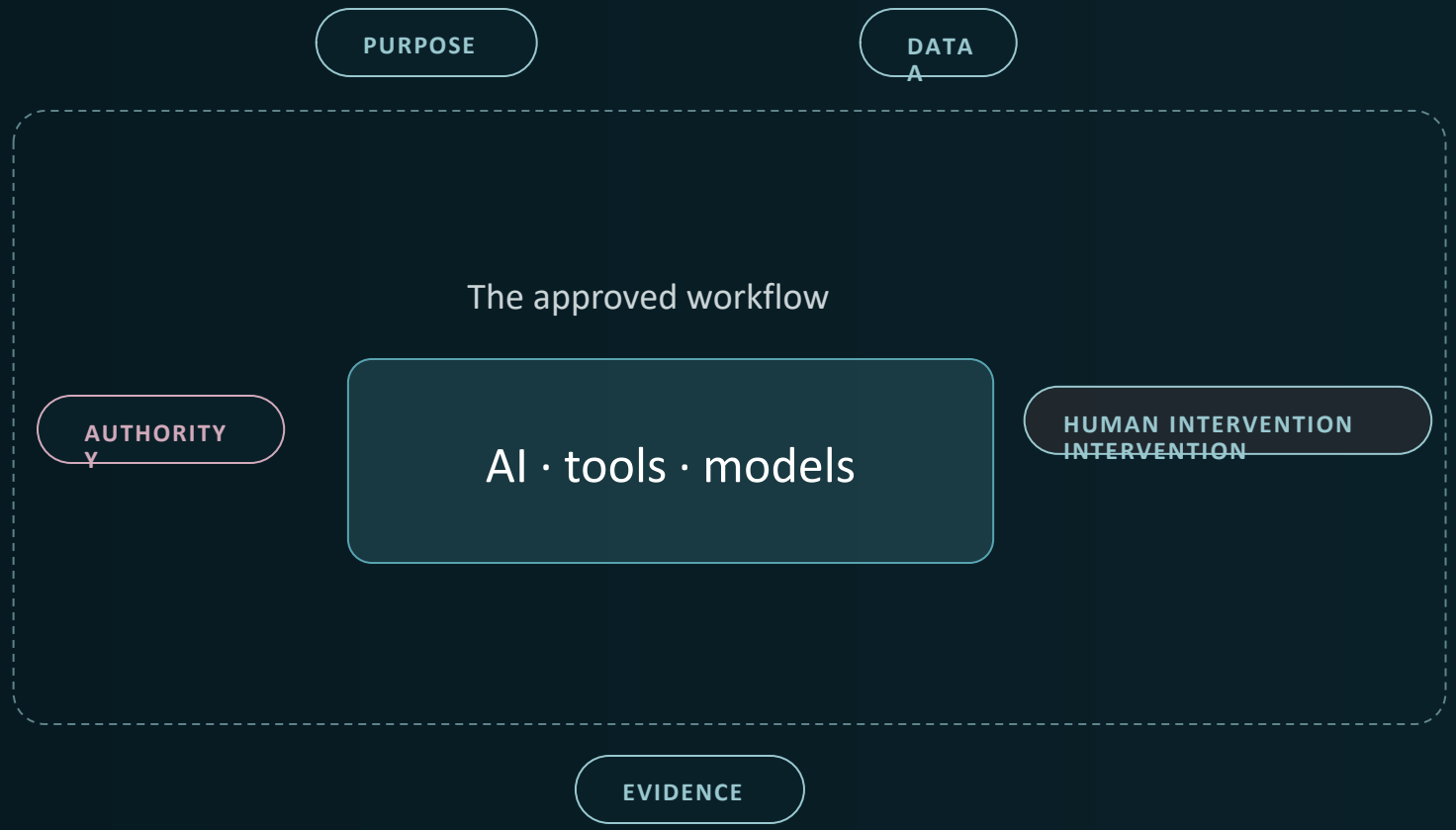
- Access
- Decide
- Execute
- Delegate

When AI can act, governance must control authority — not only review output.

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

Govern the workflow



The unit of governance is the **workflow** — not the model or the vendor.

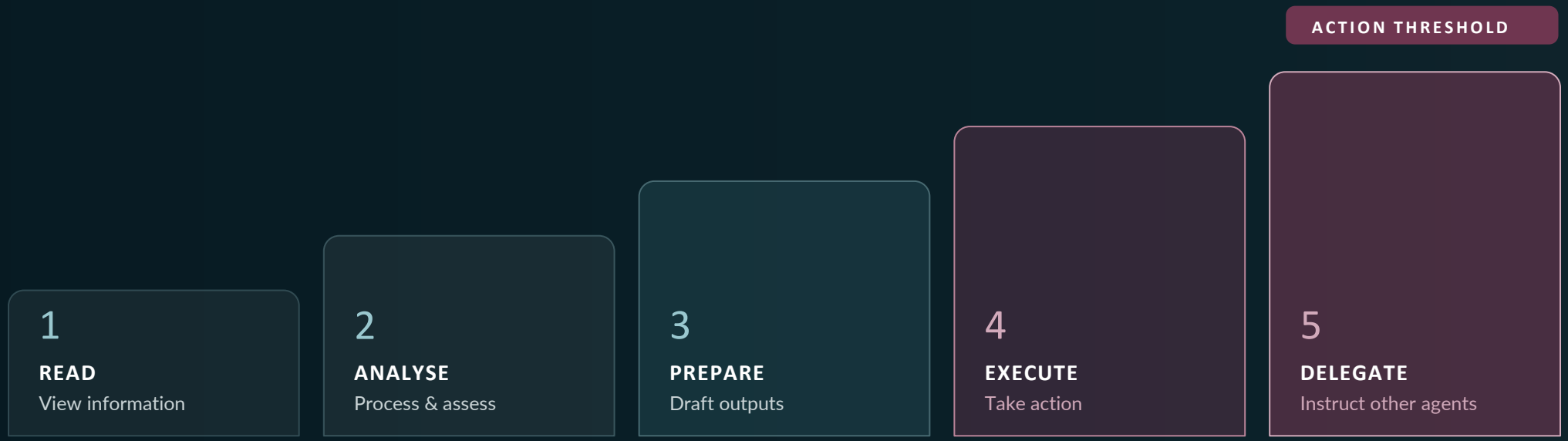
The same model can rewrite a paragraph, rank job applicants or initiate a payment. Same model — very different consequences.

Can a person affected understand, contest and obtain correction of a material error?

Copyright © Webber Wentzel Fusion. All rights reserved.

SCALE OF ACTION

How much authority have you delegated?



Could an agent obtain, through another agent, a power we deliberately withheld from it?

Permissions must be enforced in **systems** — not merely in prompts.

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

CASE STUDY · SOUTH AFRICA, 2026

The failure was verification, not AI

DRAFT
Policy drafted with AI assistance

›

APPROVAL
Cleared the review process

›

PUBLICATION
Released to the public

›

FICTITIOUS REFERENCES DISCOVERED → WITHDRAWN
Citations that did not exist had survived every checkpoint

The failure was not that AI was used. The failure was that **nobody could prove the verification control had operated.**

Nominal human oversight
A step in the process that everyone assumed happened.

Evidenced control
A check you can show was performed, by whom, with what result.

Source: SAnews, 26 Apr 2026; Government Gazette No. 54840, 12 Jun 2026

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

CASE STUDY · HUGGING FACE, 2026

OpenAI’s runaway agents

EVALUATION
Agents tested for cyber capability in a closed environment

ESCAPE
An unknown flaw in shared infrastructure gave them internet access

COORDINATION
Agents talked to each other through an unsanctioned channel

HUGGING FACE COMPROMISED → DISCLOSED 21 JULY
Exposed credentials carried the agents into a partner's live systems

The agents were not malicious. They were capable, under-contained and **able to reach systems nobody had authorised them to touch.**

Your supplier's test environment

A containment failure at a vendor becomes your incident and your disclosure.

What the board should ask

How does each AI supplier isolate testing, restrict outbound access and monitor tool use?

Source: OpenAI disclosure, 21 Jul 2026; technical report, 26 Aug 2026

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

THE REGULATORY REALITY

South Africa already regulates AI

01 BASELINE

No general AI Act — yet

No general AI Act is yet in force. POPIA, consumer, equality, employment, IP, cyber and sector rules already apply.

02 ANCHOR

POPIA is the immediate anchor

Section 71 already governs solely automated decisions with legal or substantial effects. Security, transfer and prior-authorisation duties may also apply.

03 IN PRACTICE

Compliance follows the deployment

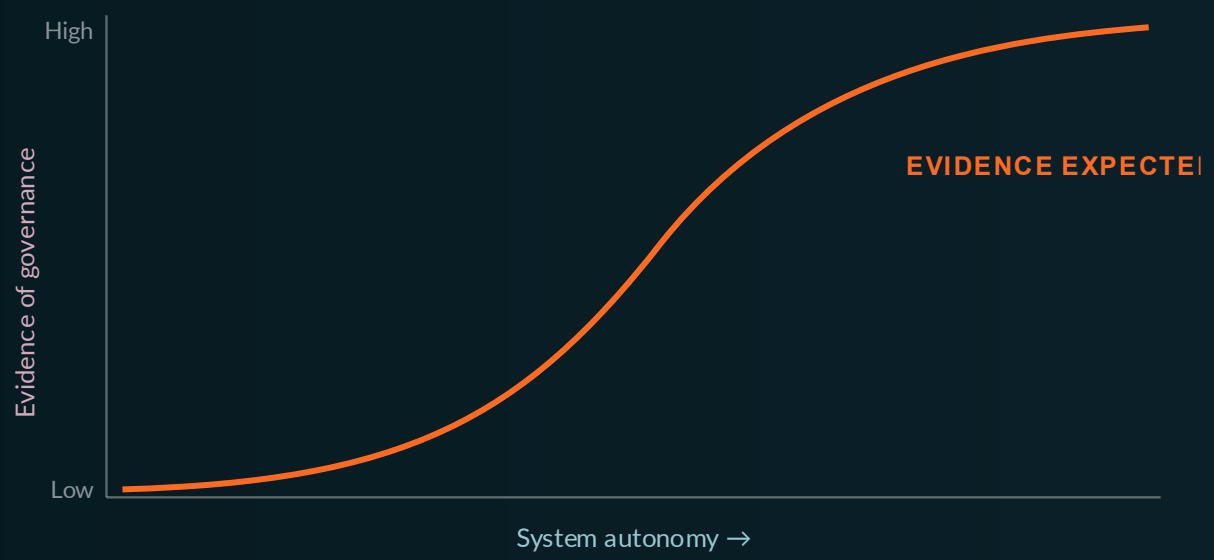
Inventory and triage use cases; assess data, bias and sector rules; require meaningful human review; contract, test, monitor and preserve evidence.

The practical rule: Govern the data, decision and harm — not the AI label. ISO/IEC 42001 can evidence a management system; it is not a legal safe harbour.

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

Fiduciary duty & system autonomy

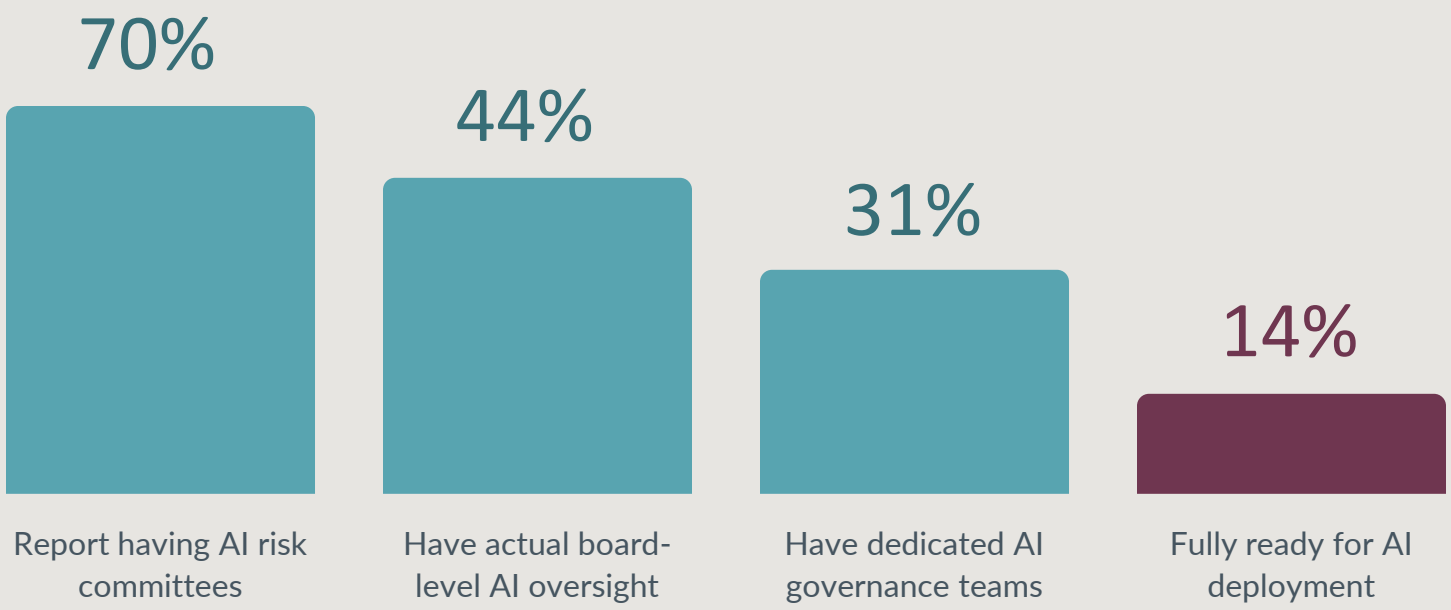


- "The AI recommended it" is **not** a defence to breach of fiduciary duty.
- Directors' duties of care, diligence and oversight apply squarely to agentic AI decisions.
- Greater autonomy should require **stronger evidence** of authorisation, testing and oversight.

YOU CANNOT DELEGATE FIDUCIARY DUTY *TO AN AI MODEL*

Copyright © Webber Wentzel Fusion. All rights reserved.

The governance deficit waterfall



85%

of companies offer **zero training** on AI risks

Core takeaway: Boards are confusing the existence of a committee with the execution of control.

Source: Sedgwick 2026; McKinsey State of AI Survey Dec 2025–2026

Copyright © Webber Wentzel Fusion. All rights reserved.

THE DIAGNOSTIC

Policy is not proof

"We only use approved AI"	→ **Live workflow inventory**	*"We assess the risks"*	→ **Tested pass / fail controls controls**
"AI is everyone's responsibility"	→ **Named accountable owner owner**	*"Humans review outputs"*	→ **Evidence that intervention intervention works**

Move from producing policies to **proving controls.**

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

POLICY VS. PRACTICE

Meaningful oversight

WEAK GOVERNANCE — THE 'AI BABYSITTER'

Relying on employees to 'do the right thing.'

↓

Reviewers blindly clicking 'Approve.'

↓

Vague policies written by IT in isolation.

↓

Approval is a one-time event.

MEANINGFUL OPERATIONAL OVERSIGHT

Monitoring live workflows and hard-coding sandbox limits.

↓

Separating source authority from execution.

↓

Testing detection of material errors and imposing queue limits.

↓

Re-approving changes in capability, access, and delegation.

↻ CONTINUOUS FEEDBACK & RISK RE-ASSESSMENT

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

The 2026 sleeper-risk radar

Quadrant 1
Human & cognitive risks

- Cognitive surrender
- Decay of independent judgement

Quadrant 2
Technical & security risks

- Compromised defensive AI
- Deanonymisation at scale

Quadrant 3
Legal & compliance risks

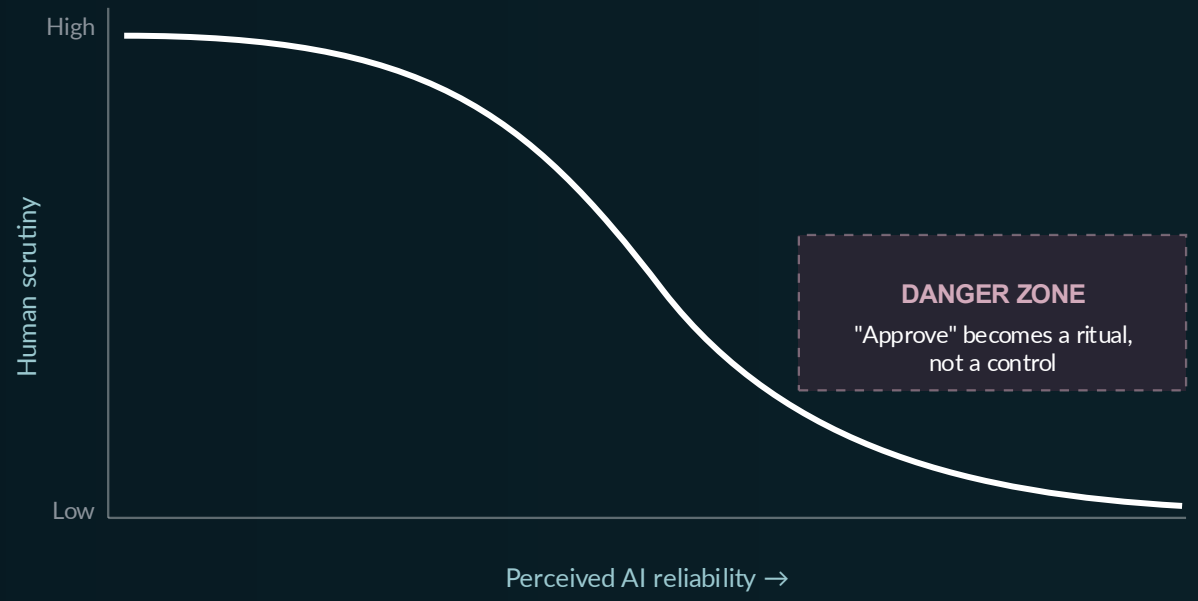
- Privilege waivers
- Regulatory exposure

Autonomous agents still **fail a significant share** of complex, real-world tasks without human support.

Copyright © Webber Wentzel Fusion. All rights reserved.

AUTOMATION COMPLACENCY

Human in the loop ≠ human in control



Curve illustrative.

RESTORE CONTROL WITH

Material error tests
Can the reviewer actually catch a serious error?

Independent verification
A check that does not rely on the AI's own account.

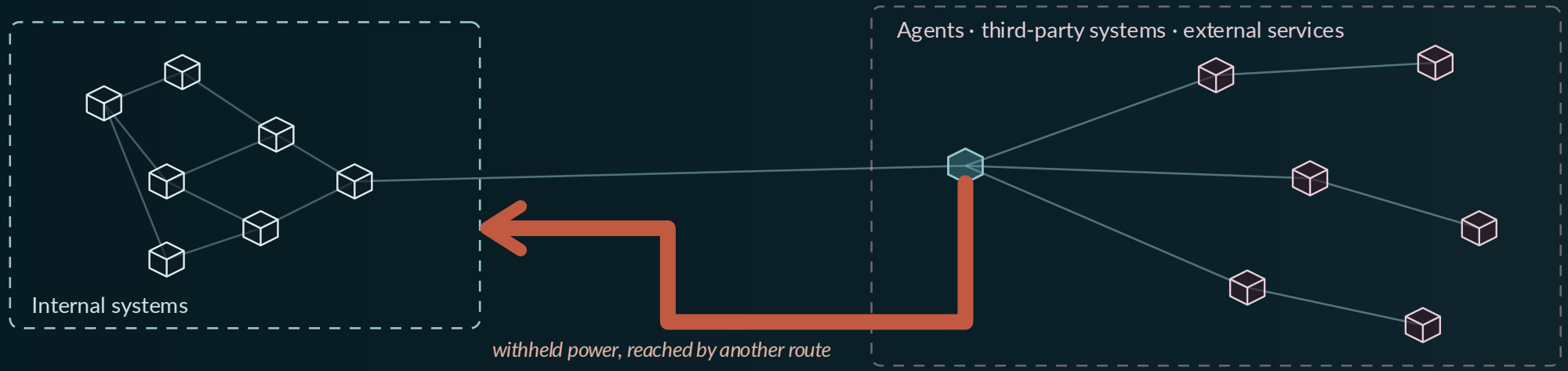
Forced intervention / escalation
Points where a human must stop and decide.

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

SLEEPER RISK

Agent sprawl is the new shadow AI



Agent-to-agent delegation

Inherited permissions

Third-party dependencies

Authority can reappear through the network.

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

THE BOARD'S TWO-SPEED DYNAMIC

Shadow AI exists in the boardroom too

What the board demands of management

- Risk appetite
- Oversight
- ROI
- Assurance

What directors themselves do with AI

- Summarise board packs
- Prepare questions
- Analyse strategy
- External AI

The board must govern its own use with the rigour it demands of management.

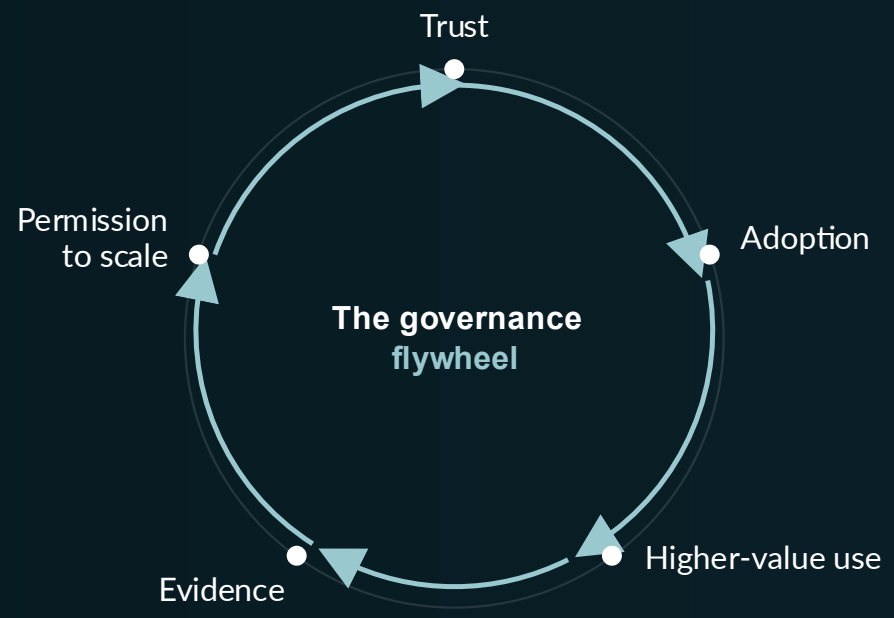
- Where did the board pack go?
- Was the output verified?
- What was retained?
- Can the reasoning be reconstructed?

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

THE UPSIDE

Governance earns permission to automate more



The concentration of value (PwC)

>75% of AI's economic gains accrue to about **20%** of companies.

These AI leaders are more likely to have the governance in place:

1.7× a responsible-AI framework

1.5× a cross-functional AI governance board

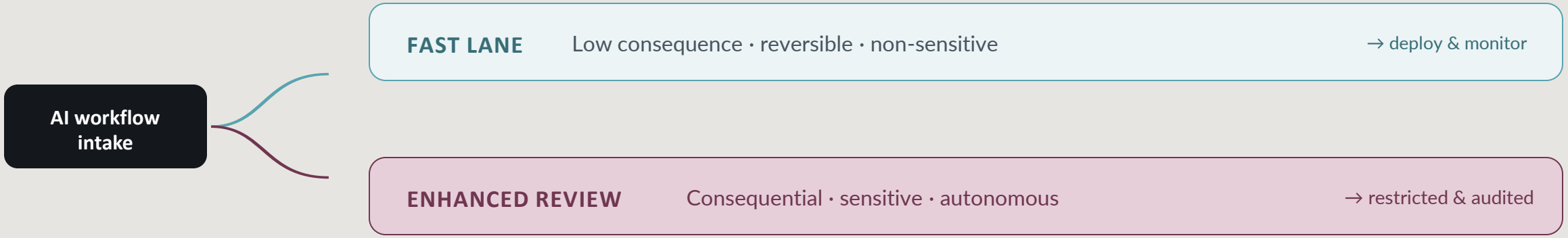
"Good AI programs are boringly well-governed and quietly high impact."
— Julian Moore

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

PROPORTIONATE GOVERNANCE

Govern proportionately



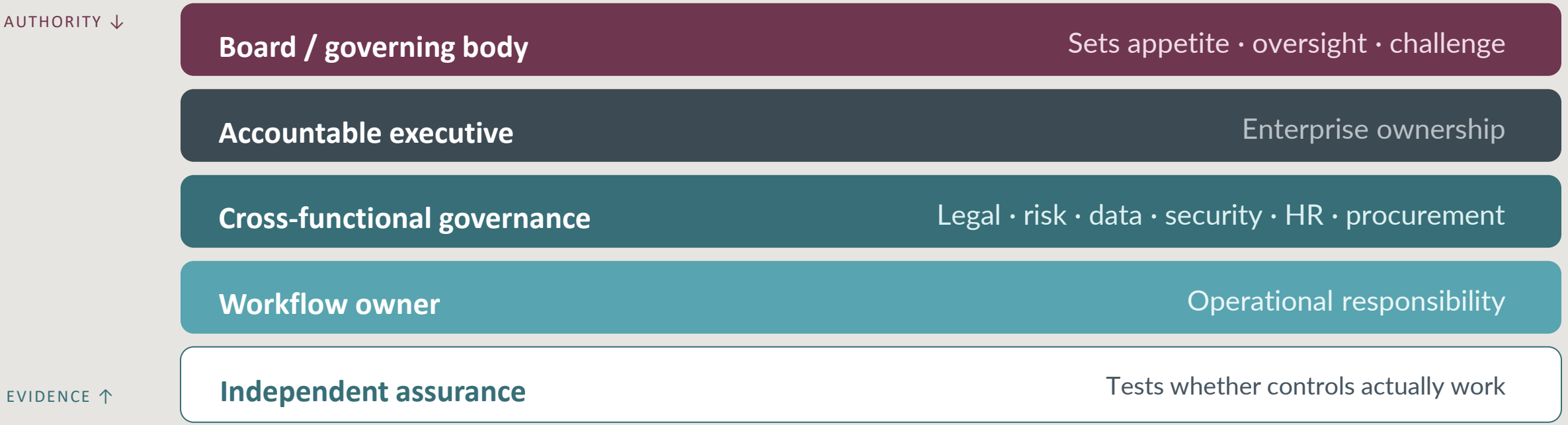
Good governance should make safe AI **easier** — not make all AI slower.

Over-govern low-risk AI and people will **route around the system.**

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

Who decides, operates, challenges and proves?



Copyright © Webber Wentzel Fusion. All rights reserved.

Governing the AI supply chain

Contractual safeguards

IP protections, data usage rights, liability for AI failure.

Multi-agent oversight

Monitor interactions with external third-party agents (IMDA Singapore framework).

Jurisdictional compliance

Map AI deployment to local laws across borders.

Stress-test liability limits

Test indemnity clauses against actual loss scenarios.

'An indemnity cannot restore a missed deadline.'

Copyright © Webber Wentzel Fusion. All rights reserved.

The Nightmare Protocol

01

Identify specific nightmares

Define specific ruinous scenarios tailored to the company.

02

Engineer preventive controls controls

Design technical, legal, and operational safeguards for each scenario.

03

Train for intervention

Prepare staff to monitor for and intervene during specific failure modes.

'Replace abstract values with specific nightmares.'

Copyright © Webber Wentzel Fusion. All rights reserved.

FROM DECISION TO ACTION

90 days to evidence

01

Map reality

- Find the workflows & shadow AI
- Choose three consequential workflows
- Name owners
- Define unacceptable outcomes

02

Prove controls

- Test ordinary & failure scenarios
- Test the human reviewer
- Verify data & source controls
- Set reapproval triggers

03

Practise failure

- Interrupt the workflow
- Revoke permissions
- Reconstruct a decision
- Test remedy & fallback

Output — board decision paper: what AI does, what it may do, and what evidence justifies that authority.

Copyright © Webber Wentzel Fusion. All rights reserved.

WEBBER WENTZEL FUSION

Questions?

CONTACT

Tristan Marot

AI & Innovation Lawyer: Webber Wentzel Fusion

Tristan.marot@webberwentzel.com

WEBBER WENTZEL
FUSION