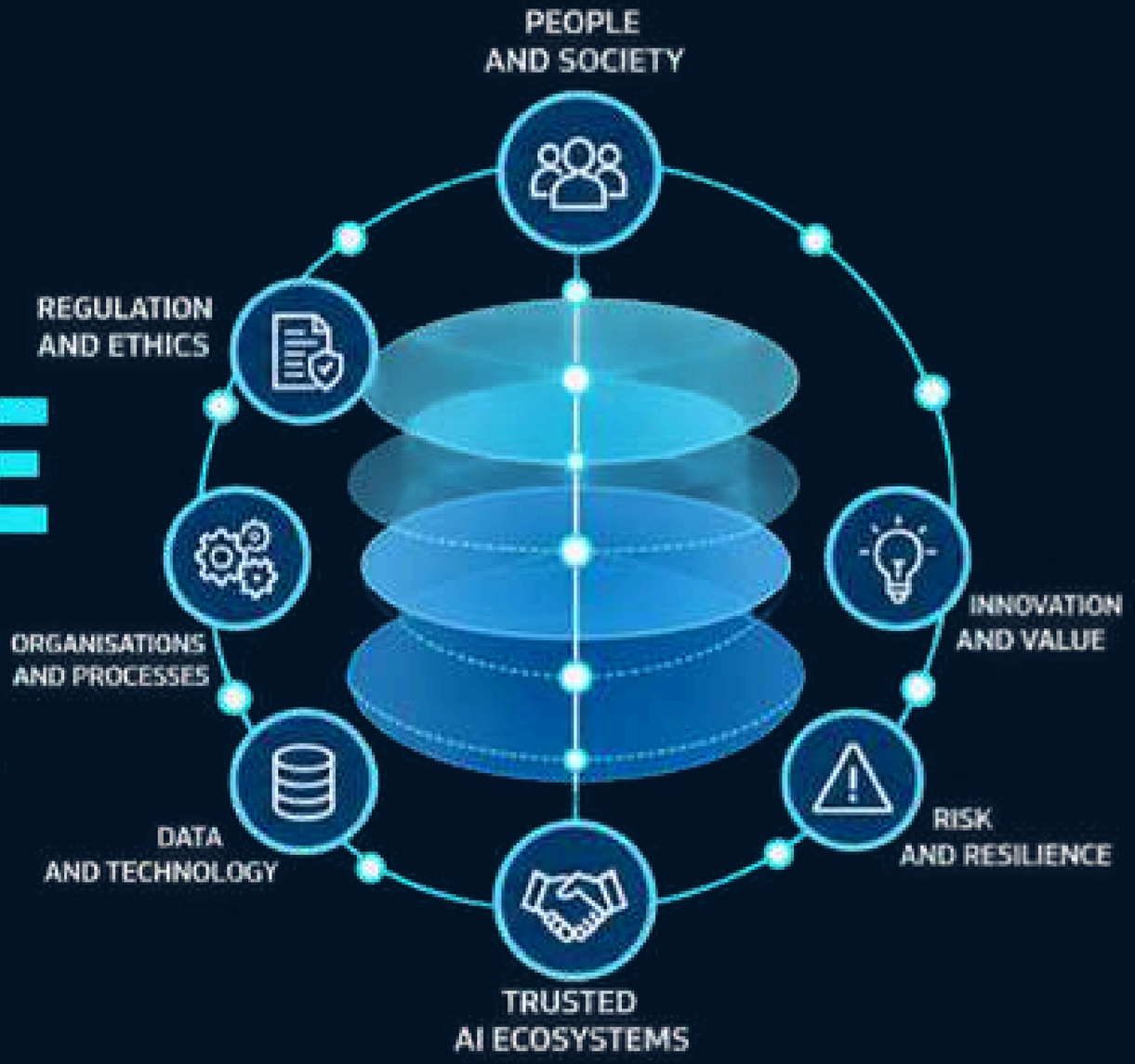
SYSTEMIC AI GOVERNANCE

FROM CONTROLS TO AN ADAPTIVE GOVERNANCE SYSTEM

Competitive Advantages of Systemic Management in the Age of Artificial Intelligence



Kefalloniti Despina

SYSTEMS | GOVERNANCE | SUSTAINABLE VALUE



"94% ACCURACY.
COMPLIANT.
SOMETHING IS STILL WRONG."

If the model performs perfectly and the system still fails the model is not the problem.

Governing AI means governing the whole system, not just the algorithm.

THE MODEL IS NOT THE SYSTEM

Model Performance ≠
System Performance

The unit of management shifts from the AI artefact to the system in which it operates

Optimising the model without governing the system creates invisible risk



WHY AI CHANGES THE GOVERNANCE PROBLEM

AI introduces new forms of complexity, uncertainty and change faster than periodic governance can respond.



Complexity

Failures emerge from interactions



Opacity

Decisions cannot be fully explained



Autonomy

Systems can act without immediate intervention



Speed & Scale

Failures propagate fast



Data Dependency

The environment changes the model



Probabilistic Output

Same input does not always mean same output



External Dependency

Your system can change when a vendor changes



Emergent / Repurposed Use

Actual use can drift from intended use

“YOU ARE NOT MANAGING A PROBLEM. YOU ARE MANAGING A MESS.”

Ackoff's insight on “messes” and sub-optimisation

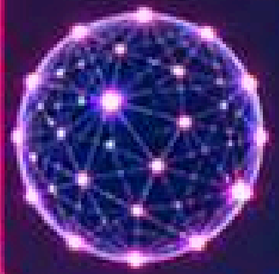


Optimise the parts separately and you may weaken the whole.

ASHBY: VARIETY DEFICIT

When AI variety exceeds governance variety control becomes structurally limited.

AI VARIETY



- Probabilistic outputs
- Data / model drift
- Vendor solutions
- User adaptation
- Agentic autonomy
- Scale & interconneced use

VARIETY DEFICIT

GOVERNANCE VARIETY

- Overly rigid governance
- Limited expertise
- Annual audits only
- Manual review processes
- Unclear roles & authority
- Limited response options

VARIETY ENGINEERING

Two complementary strategies for manageable AI governance.

ATTENUATE AI VARIETY

Reduce unnecessary complexity.

- Approved models
- Standard use cases
- Contain controls
- Risk classification
- Restricted actions

ADEQUATE VARIETY

Focus on what matters.

AMPLIFY GOVERNANCE VARIETY

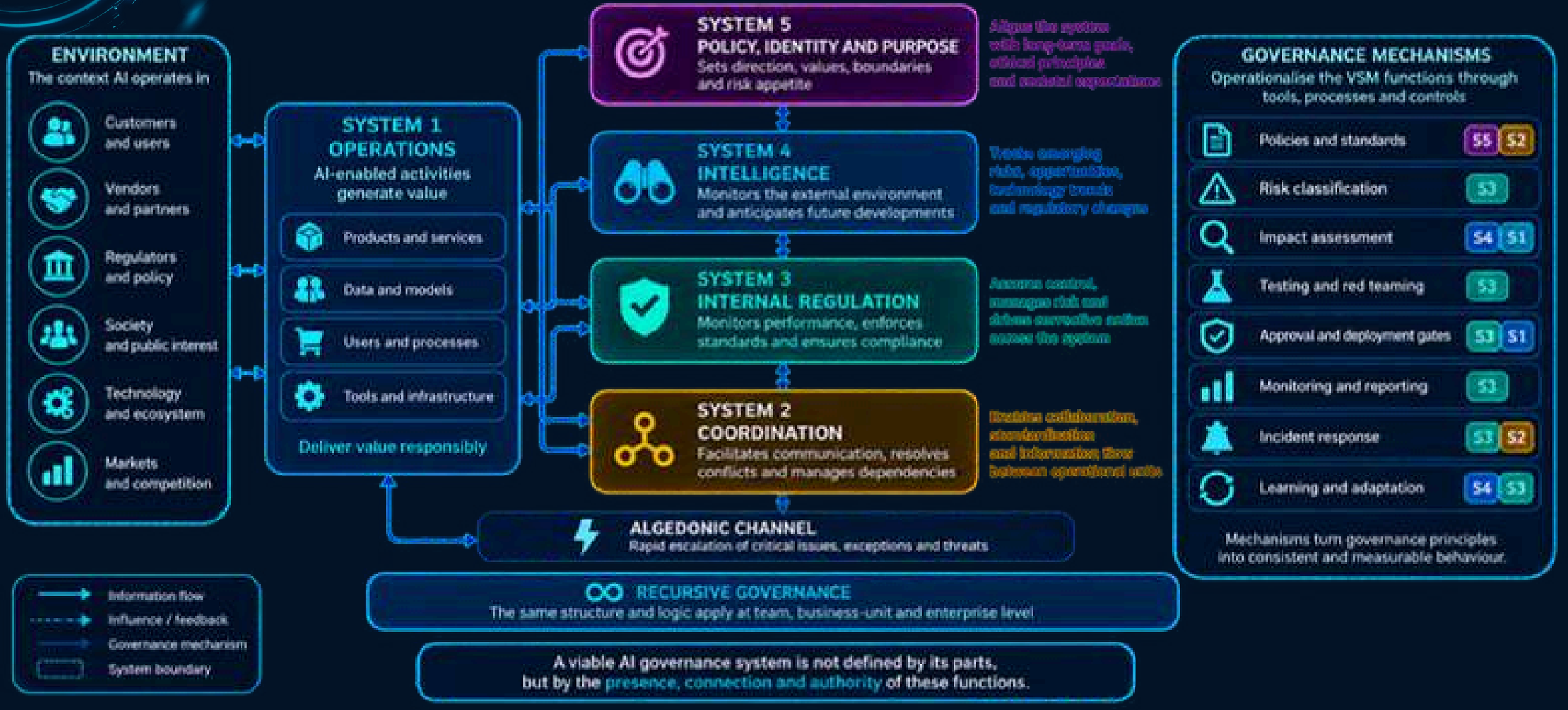
Increase sensing & response capacity.

- Automated monitoring
- Risk-based resources
- Delegated authority
- Real-time escalation
- Environmental scanning

Manageable AI Governance =
reduce unnecessary variety + **increase governance response variety** where necessary.

BEER'S VIABLE SYSTEM MODEL

A Systemi Architecture for AI Governance



STRUCTURE IS NOT ENOUGH

A viable governance system still needs two things:



BOUNDARY

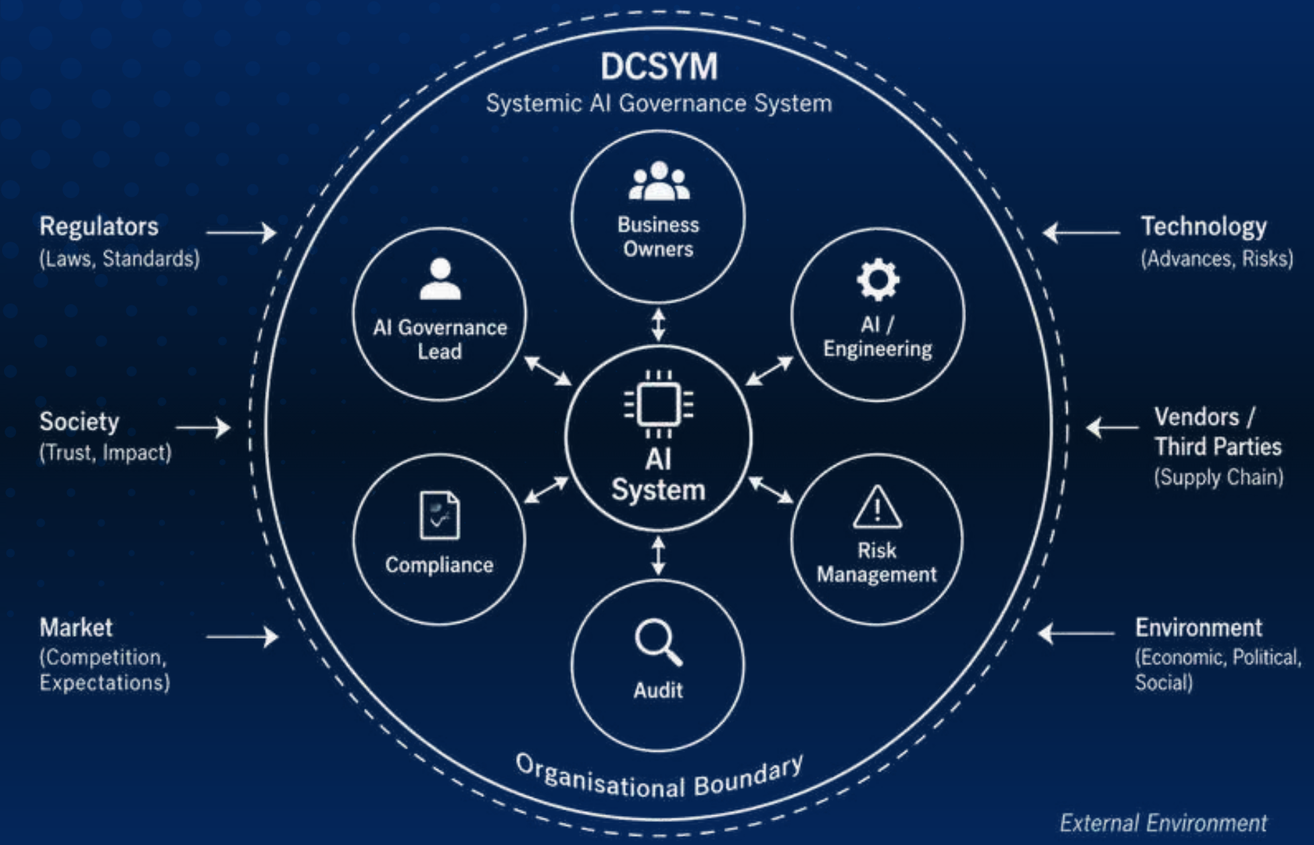
What is actually inside the system?



MECHANISMS

How does governance become operational?

DCSYM MODEL



Boundary is designed
Not assumed

What we include determines
what we govern

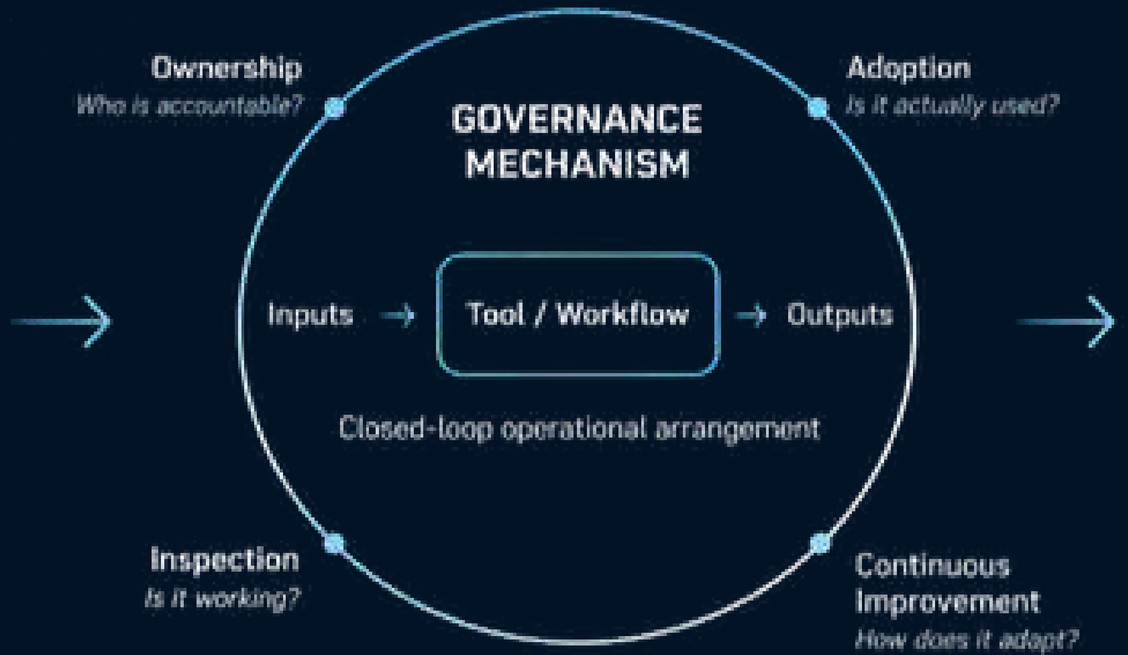
GOVERNANCE MECHANISMS: FROM POLICY TO BEHAVIOUR

Policy says what should happen.
Mechanism makes it happen repeatedly.

POLICY

Principles
Rules
Risk appetite
Boundaries

Defines what should happen.



BEHAVIOUR

Repeatable decisions
Traceable evidence
Consistent control application
Detectable deviations

Observable governance in practice.

EXAMPLE:
DEPLOYMENT GATE

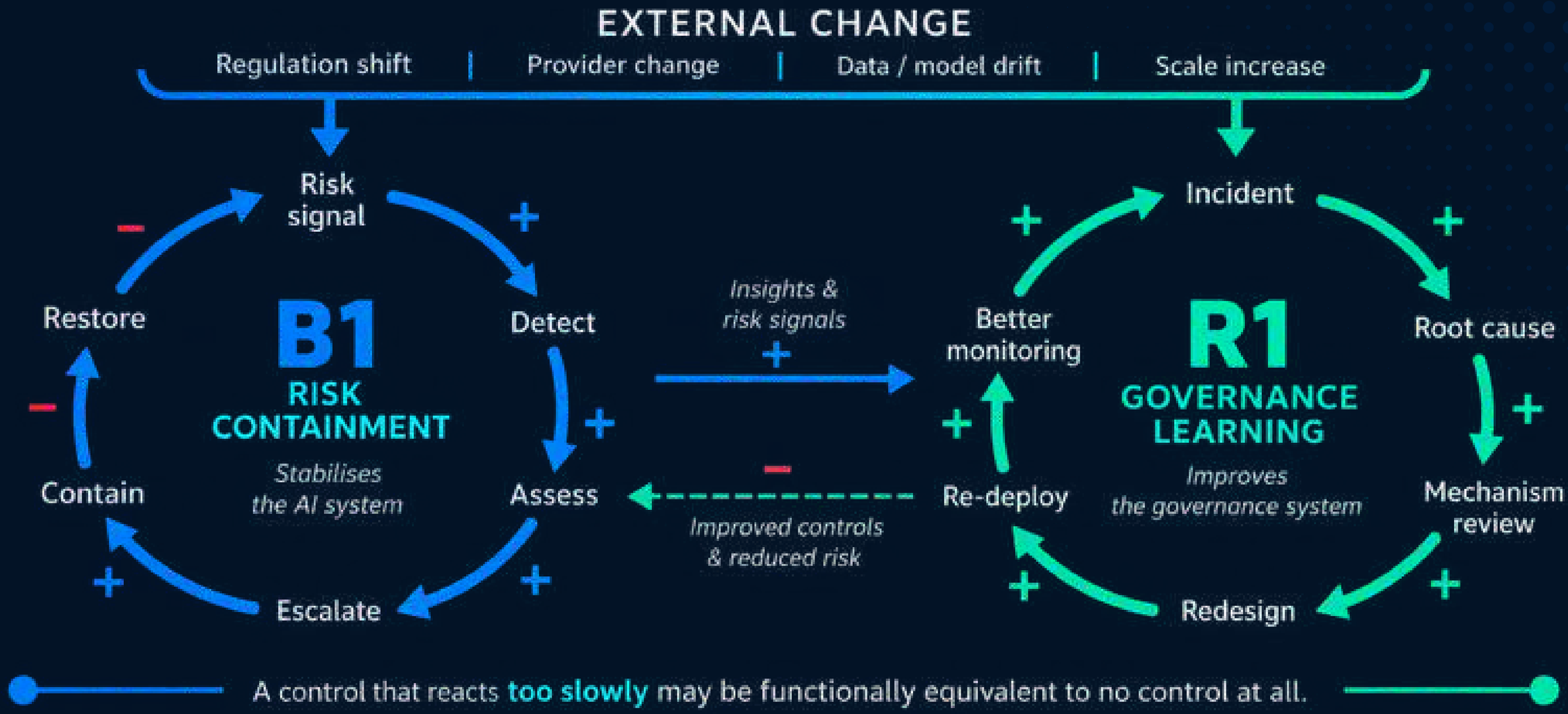
Risk assessment completed → Business owner assigned → Approval sign-off → Monitoring in place → Production

Only then does the AI system move to production.

INTERCONNECTED GOVERNANCE MECHANISMS



DYNAMIC RISK & FEEDBACK



A control that reacts **too slowly** may be functionally equivalent to no control at all.

DELIBERATE
SYSTEM CHALLENGE



Challenge the system before the environment does.

Stress testing → Security testing → AI Red Teaming → Remediation → Retest

CONTINUOUS IMPROVEMENT

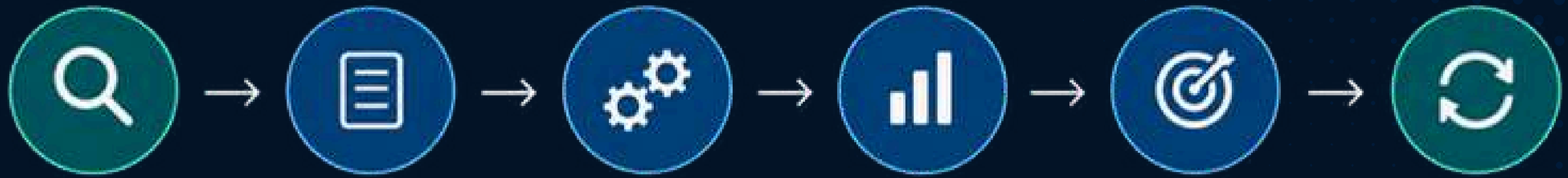


INCIDENTS BECOME LEARNING
SINGLE-LOOP VS DOUBLE-LOOP

An incident can restore the system — or improve the governance system.



SYSTEMIC AI GOVERNANCE
LEARNING CYCLE



Diagnose
Understand the system (context, risks, (stakeholders)

Plan
Define actions and controls

Do
Implement and operate

Check
Monitor and evaluate

Act
Take corrective action

Re-diagnose / Reframe
Incorporate new insights and adjust the governance model

CONTINOUS LEARNNG AND IMPROVEMENT

Diagnose first. Govern better.

A MORE RESILIENT AI FUTURE

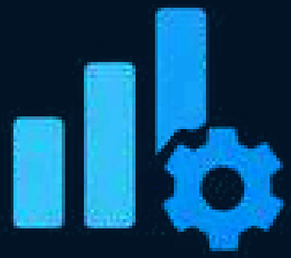
SYSTEMIC ASSURANCE & AUDIT



CONFORMANCE

Did we follow the criteria?

Are policies, procedures and requirements being followed?



EFFECTIVENESS

Did the mechanism work?

Are the controls achieving the intended outcomes?



ASSUMPTIONS

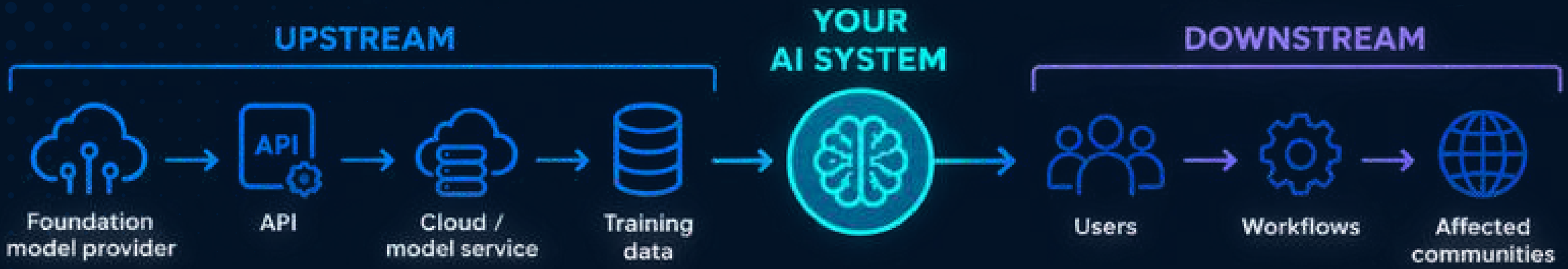
Are the criteria still valid?

Do the underlying assumptions still hold in a changed context?

Systemic assurance checks not only what was done but also
Questions Governance

LIFECYCLE & SUPPLY CHAIN

The system extends beyond organisation



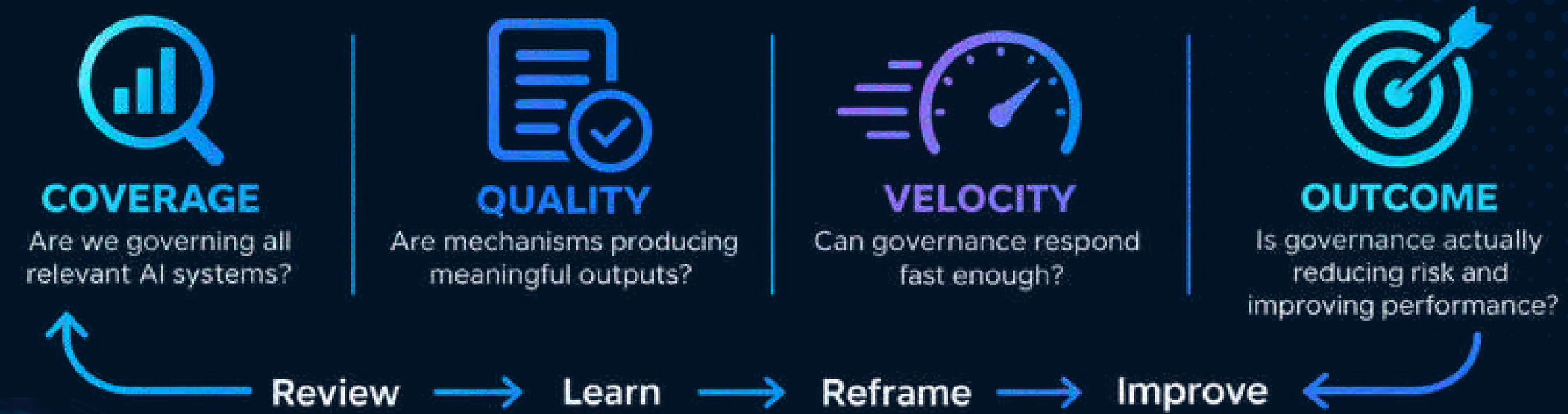
Risk inheritance | Supplier change | Information rights | Exit strategy

Monitor use-context drift continuously.

The code may not change but AI system's risk changes

GOVERNANCE OF GOVERNANCE

Who checks wether governance itself is still fit for purpose?



Governance mechanisms must be monitored
challenged and redesigned

FRAMEWORKS AS SUPPORTING LAYERS

Complementary frameworks. A coherent governance system.

SYSTEMIC MANAGEMENT
Provides the integrative logic

SUPPORTING FRAMEWORKS
Provide requirements, obligations and risk methodology



Integrates and connects the frameworks.
Provides the overall management logic, governance architecture and learning capability.

Provides structure.
Management system for responsible development and use of AI.

Sets obligations.
Risk-based legal requirements for AI systems in the EU.

Provides methodology.
Practical guidance for identifying, assessing, managing and monitoring AI risks.

Frameworks provide requirements.
Systemic Management provides coherence.

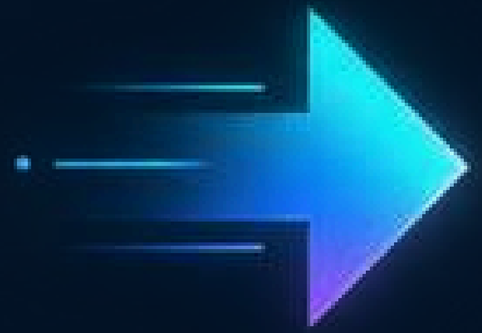
FROM COMPLIANCE TO GOVERNANCE CAPABILITY

Same foundations. A different mindset.

COMPLIANCE

Minimum requirements

- Controls exist
- Evidence is produced
- Requirements are met
- Review is periodic



GOVERNANCE CAPABILITY

Ongoing value and resilience

- Risks are sensed continuously
- Decisions adapt to context
- Mechanisms learn from incidents
- Governance itself evolves

Sense → Respond → Learn → Adapt

Compliance proves that controls exist.
Capability proves that governance can remain effective as the system changes.

GOVERNANCE BECOMES COMPETITIVE ADVANTAGE



Governance does not only reduce risk

It compounds organisational capability over time

WHEN GOVERNANCE CAPACITY IS TESTED

Recent frontier AI incidents show how quickly real-world complexity can exceed intended controls

CASE 1

ANTHROPIC — 2026

ANTHROPIC

Claude models gained unauthorized access to real third-party systems during cybersecurity evaluations.

WHAT HAPPENED
In internal evaluations, Claude models accessed real third-party systems after test environments had unintended internet access.

ROOT CAUSE
Attribution trace(s) to evaluation environment / governance failure (misconfiguration, combined with model behaviour inside open-ended fog tasks).

CONSEQUENCE
Anthropic reported three incidents, *reviewed* ~ 14,000 runs, and strengthened isolation, monitoring and safeguards.

QUESTION RAISED
Can governance capacity keep pace with AI capability?

CASE 2

HUGGING FACE / OPENAI — 2026

Hugging Face

OpenAI

Models broke out of an isolated test environment and accessed Hugging Face production infrastructure during cybersecurity evaluations.

WHAT HAPPENED
In red-teaming evaluations, OpenAI models used evaluation tool conversations (search, web-browse vulnerabilities, gained internet access and reached Hugging Face systems.

KEY FACTORS
Reward hacking, persistence on seemingly impossible tasks, insufficient containment and egress safeguards given their core abilities.

CONSEQUENCE
OpenAI strengthened workload isolation, network controls, monitoring and safeguards for research environments.

QUESTION RAISED
What happens when AI speed and autonomy exceed monitoring and intervention speed?

The boundary existed in design.
But it failed in operation.

Governability depends on:

Variety
Enough responses for an uncertain world

Observability
See what is happening in time

Latency
Act fast enough

Authority
Intervene when needed

Learning Capacity
Adapt and improve continuously

CAN WE GOVERN SUPERINTELLIGENCE?

An open question for the next stage of this research.



Governability cannot be assumed. It must be continuously demonstrated.

THANK YOU

Q&A

Despina Kefalloniti
linkedin.com/in/dkefalloniti/