

Why AI Governance Fails at Decision Points

How Decision Escalation, Explainability and Runtime Assurance Close the Governance Gap

A DEEGS perspective on authority, escalation, explainability and accountable human judgement

Shaun Wilde, Core GRC Solutions

Executive framing

Organisations have invested heavily in AI governance. Standards, regulatory frameworks, risk methodologies and internal policies have created an increasingly mature foundation for managing AI-related risk. ISO/IEC 42001 provides a management system approach to artificial intelligence. The NIST AI Risk Management Framework provides a structured approach to AI risk. Regulation, including the EU AI Act, increasingly establishes expectations according to the nature, purpose and potential impact of AI systems.

These developments matter. They provide organisations with structure, terminology and governance expectations for the responsible development and use of AI. Yet a fundamental challenge remains.

Governance can appear robust at organisational level while becoming ambiguous at the point where an actual decision is made. When AI contributes to a material decision, an organisation may be able to demonstrate policies, risk assessments, controls, system documentation and assigned responsibilities, yet still struggle to answer relatively simple questions:

- Who - or what - actually made the decision?
- Who had authority to make it?
- What information influenced it?
- What role did AI play?
- What triggered human intervention?
- Should the decision have been escalated?
- Who had authority to intervene?
- Why was a particular course of action selected?
- What happened as a result?
- Where is the evidence that demonstrates the complete chain?

These questions become increasingly important as organisations move beyond relatively simple AI assistance towards AI-enabled workflows, autonomous agents and hybrid operating models in which humans and AI systems collectively perform work.

AI changes decision dynamics.

It accelerates decision cycles, distributes activity across people and systems, generates persuasive recommendations, initiates workflows and can increasingly execute actions through connected applications and services.

The traditional assumption that a human will simply review an AI output before something consequential happens becomes progressively less reliable.

Human oversight remains essential, but **Human-in-the-Loop is not, by itself, a governance model.**

A human may technically be present while lacking the information, authority, competence, time or practical ability required to intervene effectively.

The governance challenge therefore extends beyond ensuring that humans remain involved.

It requires organisations to govern the **decision itself.**

This paper introduces **DEEGS - Decision Escalation & Explainability Governance** - as an approach to that challenge.

DEEGS focuses on how risk-bearing decisions are identified, governed, escalated, authorised and made explainable.

It is complemented by **DEXYS® - the Decision Explainability System** - which provides an operational approach to decision governance, runtime assurance and evidence across human and AI-enabled organisations.

Together, they address a critical question:

Can the organisation explain not simply what happened, but why it happened, who or what had authority, what controls operated, when intervention was required, and what evidence demonstrates that governance worked?

1. What current AI governance frameworks get right

Recent advances in AI governance represent a substantial step forward.

Management systems, risk frameworks and regulation increasingly recognise that AI governance is not simply a technical problem.

It is an organisational governance problem.

ISO/IEC 42001 provides a management system structure through which organisations can establish policies, objectives, responsibilities, risk management arrangements, controls and continual improvement for AI.

The NIST AI Risk Management Framework provides a complementary risk-oriented perspective, encouraging organisations to consider context, impact, measurement, management and characteristics associated with trustworthy AI.

Regulatory approaches increasingly reinforce proportionality, accountability and evidence according to the nature and impact of AI use.

Together, these approaches establish an important foundation.

They help organisations determine what governance arrangements should exist, what risks should be considered, what controls may be appropriate, who has organisational responsibilities and what evidence should be maintained.

However, there is a difference between **governance as designed** and **governance operating at the point of decision**.

- Policies can define accountability.
- Risk assessments can identify potential harm.
- Controls can establish expected behaviour.
- Responsibility models can allocate roles.

But none of these automatically ensures that the appropriate intervention occurs when a real decision crosses a risk threshold.

The governance question therefore shifts from:

What controls do we have?

To what happens when those controls encounter a real decision under real operating conditions?

That is the **decision-point governance gap**.

2. Where AI governance breaks down in practice

Governance failures do not necessarily begin with dramatic system failures.

They frequently emerge through ambiguity.

A recommendation becomes the normal default.

An automated workflow expands beyond its original purpose.

A human repeatedly approves AI recommendations without meaningful challenge.

A system encounters circumstances that were not anticipated during its original risk assessment.

An AI agent gains access to additional tools, information or capabilities.

Operating conditions change.

A decision crosses organisational, regulatory or ethical boundaries without triggering intervention.

Responsibility becomes distributed across systems and teams.

Everyone participates in the process, yet nobody can clearly identify where decision authority actually sat.

The original DEEGS proposition identified several recurring weaknesses that remain central to the model. In particular, optional escalation, implicit authority and weak Human-in-the-Loop arrangements can leave responsibility nominally human while practical control shifts elsewhere.

Optional escalation

Many organisations establish escalation routes but leave the decision to escalate largely to individual judgement.

That creates a contradiction.

The circumstances most likely to require escalation - uncertainty, pressure, novelty, competing objectives or potential harm - are also those in which discretionary escalation can become least reliable.

Where predefined conditions are met, escalation should therefore be **designed into governance rather than dependent solely upon individual willingness to invoke it.**

Implicit authority

AI systems do not need formal decision authority to influence decisions materially.

An apparently objective recommendation, risk score, prioritisation or generated response can become the practical default.

Over time, a human may technically retain authority while the AI system exercises considerable effective influence.

Responsibility remains human while practical control drifts elsewhere.

Passive Human-in-the-Loop

Human oversight can become ceremonial.

If an individual is expected to approve or override an AI output but lacks clear decision authority, appropriate information, competence, escalation rights or sufficient time, the human control may exist on paper without functioning effectively.

Human involvement therefore needs to be designed around **authority and intervention**, not merely presence.

Distributed decision-making

Increasingly, a decision may not occur at a single identifiable moment.

An AI-generated signal may trigger an automated workflow. Another system may apply a rule. An agent may execute an action. A human may subsequently approve an exception.

The effective decision becomes distributed across technology, people and time.

This makes traditional models of responsibility increasingly difficult to apply unless the organisation deliberately governs the decision chain.

These are not necessarily failures of intent.

They are failures of **governance design**.

3. Risk at the Core

The Core GRC® Framework places **Risk at the Core**.

Risk provides the connection between organisational purpose, governance and operational activity.

Controls exist because something matters.

Signals matter because they may indicate that risk is changing.

Decisions matter because they determine what the organisation does in response.

Governance therefore needs to connect organisational purpose with operational decisions.

One way of expressing this relationship is:

Purpose → Risk → Control → Signal → Decision Trigger → Authority → Decision → Action → Outcome → Evidence

This is the governance chain.

- **Purpose** establishes what the organisation is trying to achieve.
- **Risk** considers the uncertainty that could affect that objective.
- **Control** establishes how that risk is managed.
- **Signal** provides information about the operating environment.
- **Decision Trigger** identifies when something requires attention or intervention.
- **Authority** determines who or what is permitted to decide.
- **Decision** records what was determined and why.
- **Action** describes what happened as a consequence.
- **Outcome** considers the resulting effect.
- **Evidence** allows the organisation to demonstrate the complete chain.

The final element is particularly important.

Evidence should not simply demonstrate that a control exists.

It should enable the organisation to explain **how governance actually operated**.

This is where explainability becomes an organisational governance property rather than simply a characteristic of an AI model.

4. DEEGS - Decision Escalation & Explainability Governance

DEEGS is a governance model focused on decisions.

It does not replace existing management systems, risk frameworks, organisational policies or regulatory requirements.

It operates across them. Where established frameworks help determine **what governance should exist**, DEEGS focuses on **what needs to happen when a risk-bearing decision occurs**.

The earlier DEEGS model described escalation as a governance control rather than a procedural afterthought, requiring authority and escalation to operate when decisions cross defined boundaries. That principle remains fundamental.

The evolution from **Exception Governance** to **Explainability Governance** broadens the concept.

Exceptions remain important, but they are now treated as one of the conditions that governance must recognise and manage.

Explainability describes the wider outcome.

DEEGS is concerned with whether the organisation can explain the relationship between:

Risk, signals, authority, decisions, interventions, actions and outcomes.

Several principles underpin the approach.

Decision authority should be explicit

Not every decision carries the same significance.

Decisions differ according to potential impact, uncertainty, reversibility, novelty, regulatory exposure and risk.

Governance should therefore establish appropriate authority boundaries.

This becomes particularly important where AI systems or agents can move beyond generating information to initiating or executing actions.

Escalation should be designed

Escalation should not be treated simply as an optional process available when somebody feels uncomfortable.

Material escalation conditions should be considered in advance.

The original DEEGS proposition established that escalation should be linked to decision characteristics and supported by genuine authority to pause, redirect or reverse activity.

That remains a core principle.

Intervention should be proportionate

Not every unusual condition requires activity to stop.

Governance may need to observe, constrain, seek additional authority, pause, reverse, escalate or stop activity depending upon the circumstances.

The important point is that intervention should be linked to risk and authority rather than improvised after an event.

Exceptions should be expected

Unexpected circumstances are inevitable.

Models change.

Data changes.

Operating environments change.

Humans behave differently from assumptions.

Systems interact in unforeseen ways.

Effective governance therefore anticipates exceptions and ensures that they are surfaced, assessed, owned and escalated where appropriate.

The original DEEGS paper argued that edge cases, model drift and unexpected interactions should be anticipated rather than suppressed or retrospectively rationalised. That remains unchanged.

Decisions should be explainable

A governed decision should generate sufficient evidence to understand the relevant context, risk, authority, information, judgement, action and outcome.

This extends explainability beyond the question:

Why did the AI produce this output?

towards the broader governance question:

Why did the organisation make this decision?

5. Human + AI accountability

Traditional responsibility models were designed predominantly for human organisations.

AI changes that assumption.

AI systems can now analyse information, recommend actions, execute workflows, invoke tools and interact with other systems.

Governance therefore needs to distinguish between concepts such as:

Ultimate human accountability, decision authority, human and AI execution, judgement, consultation and information flows.

These should not be treated as interchangeable.

A system may execute an action without owning the decision.

AI may provide analysis or recommendations without holding decision authority.

A human may retain ultimate accountability without personally performing every activity.

An AI agent may have delegated authority to operate within defined boundaries without having unlimited discretion.

Core GRC® uses an extended responsibility approach based upon an extended RACI matrix - to help examine these relationships in hybrid human + AI environments.

The detailed mechanics of that model are outside the scope of this paper.

The important governance principle is simpler:

AI participation should be explicit enough that human accountability and decision authority remain visible.

The objective is not to anthropomorphise AI or transfer human accountability to technology.

It is to make the relationship between humans and technology sufficiently clear that organisations understand **who or what may act, within what boundaries, and when human authority must intervene.**

This creates the potential for greater rather than less responsible delegation.

Where authority and intervention boundaries are explicit, organisations can allow technology to operate with appropriate autonomy while maintaining governance over the decisions that matter.

6. From static governance to runtime assurance

Traditional assurance frequently examines governance retrospectively.

Policies are reviewed.

Risk assessments are sampled.

Control evidence is inspected.

Audit trails are examined.

These remain essential activities.

But increasingly autonomous and interconnected AI environments introduce another requirement:

Can governance recognise when operating conditions have changed and respond while the system is running?

An AI-enabled service is not necessarily static.

Its behaviour or risk profile may be affected by changes to models, configuration, data, permissions, connected tools, organisational context, operating instructions, environmental conditions and downstream systems.

An assessment undertaken before deployment cannot therefore provide unlimited assurance about future operation.

Governance increasingly needs to understand what is happening **at runtime**.

Runtime assurance connects:

Signals → Decision Triggers → Authority → Intervention → Evidence

The question becomes:

What is happening now, has anything materially changed, and does that change require a governed decision?

This creates a more dynamic relationship between risk management and operational governance.

It also changes the nature of assurance.

Instead of relying exclusively upon retrospective evidence that governance existed, organisations can increasingly demonstrate evidence that governance **operated when required**.

7. Intervention at runtime

Runtime governance requires more than monitoring.

Monitoring can tell an organisation that something has happened.

Governance determines what happens next.

Depending upon risk, authority and circumstances, an appropriate response might include continuing operation, increasing observation, constraining activity, requiring additional approval, pausing activity, reversing an action, escalating authority or stopping activity.

The precise mechanisms will vary between organisations and use cases.

DEEGS does not require every organisation to use an identical intervention structure.

The important principle is that **signals should connect to meaningful governance responses.**

A dashboard that detects an issue without establishing who has authority to respond is monitoring, not complete governance.

Likewise, a human who receives an alert but cannot practically intervene does not provide effective human oversight.

Runtime assurance therefore needs to connect technical observation with organisational authority.

8. Evidence by design

Decision logging should not exist simply to satisfy an auditor after an event.

The original DEEGS paper argued that decision records should provide evidence that authority was correctly exercised and escalation occurred when required. That principle can be taken further.

Evidence should be a product of governance operating correctly.

Where possible, the organisation should not need to reconstruct a significant decision months later from emails, meeting notes, system logs and individual recollection.

The governance process itself should establish sufficient traceability.

This does not mean storing everything indefinitely.

Nor does it mean creating excessive bureaucracy around routine decisions.

Evidence should be proportionate to risk.

For material decisions, however, an organisation should be able to connect the relevant context, controls, signals, authority, judgement, decision, action and outcome.

This creates evidence capable of supporting management assurance, board oversight, audit and regulatory scrutiny.

More importantly, it supports organisational learning. A decision record should not simply tell an organisation what happened.

It should help it determine whether its governance worked.

9. The role of the auditor

Runtime assurance does not remove the need for audit.

It potentially strengthens it.

Traditional audit frequently examines samples of activity after events have occurred.

Where decision governance is designed into the operating environment, auditors can additionally examine whether:

- decision triggers are appropriate;
- authority boundaries are clear;
- escalation occurred when required;
- intervention was proportionate;
- human and AI participation was understood;
- exceptions were handled appropriately;
- evidence is sufficient; and
- governance mechanisms remain effective.

This shifts part of the assurance question from:

Does the control exist?

Towards, did the control operate when the organisation needed it?

Audit remains independent from operational decision-making.

The distinction is that richer operational evidence can provide auditors with greater visibility into how governance actually performed.

The original paper concluded that AI governance cannot be assessed solely through static artefacts or model-level controls and must also be evaluated through the discipline of decision-making itself.

That becomes increasingly important as AI moves from recommendation towards execution.

10. From DEEGS to DEXYS®

DEEGS defines the governance model.

Applying that model consistently across a modern organisation creates an operational challenge.

Relevant information may already exist across identity platforms, business applications, management systems, risk tools, HR systems, learning platforms, security technology, AI platforms and organisational systems of record.

Creating another isolated repository does not solve that problem.

This is the role of **DEXYS® - the Decision Explainability System**.

DEXYS is conceived as a governance layer across hybrid human + AI organisations.

Its purpose is not to replace every participating system.

Instead, it enables decision governance and runtime assurance to operate across the environment, using appropriate integrations with existing systems and sources of evidence.

This establishes a clear relationship between three elements.

Core GRC® Framework establishes the wider governance context and places **Risk at the Core**.

DEEGS defines how risk-bearing decisions are governed, escalated and made explainable.

DEXYS® provides the operational capability through which decision governance, runtime assurance and evidence can be implemented and monitored.

In simple terms:

- **The Core GRC® Framework** defines what must be governed.
- **DEEGS** defines how decisions are governed and made explainable.
- **DEXYS** operationalises, monitors and evidences that governance.

11. One record of the whole chain

Organisational governance information is frequently fragmented.

Business context may exist in one system.

Policies may exist in another.

Risks may sit in a register.

Controls may form part of a management system.

Technical signals may come from operational or security platforms.

Human judgement may occur through collaboration tools.

Evidence may be distributed across multiple repositories.

Individually, each artefact may be valid.

The challenge is demonstrating their relationship.

DEXYS therefore introduces the concept of **One record of the whole chain**.

At a high level, this can be expressed as the **DEEGS Governance Chain**

Context → Policy → Risk → Controls → Evidence → Judgement → Decision

This does not require every underlying artefact to be moved into a single application.

The objective is **traceability rather than duplication**.

An organisation should be able to move backwards from a significant decision through the evidence, controls, risks and context that informed it.

It should also be possible to move forwards from the decision to the resulting action and outcome.

When a decision is challenged, the organisation should not need to reconstruct its governance from disconnected records.

Explainability should exist **by design**.

12. What good looks like under pressure

Effective governance is not demonstrated by the absence of difficult decisions.

It is demonstrated by what happens when difficult decisions arise.

Under pressure, mature decision governance should demonstrate several characteristics.

Decision ownership is clear.

Human accountability and appropriate decision authority remain identifiable for material decisions.

AI participation is visible.

The organisation understands when AI is advising, recommending, executing or operating with delegated authority.

Authority boundaries are explicit.

Humans and systems operate within understood limits.

Escalation is designed.

Material conditions for escalation are considered before an incident rather than invented afterwards.

Intervention is practical.

People with oversight responsibilities possess meaningful authority to act.

Exceptions are expected.

Unexpected behaviour is surfaced and governed rather than hidden or normalised.

Change is recognised.

Governance considers whether changes in technology, data, context or operating conditions have altered risk.

Decisions are explainable.

The organisation can demonstrate the relevant context, information, authority, judgement and rationale.

Outcomes remain connected to decisions.

Governance does not finish when the decision is recorded.

Evidence is generated through operation.

Assurance does not depend entirely upon retrospective reconstruction.

These characteristics turn governance from a collection of static artefacts into an observable organisational capability.

13. Implications for boards

Boards should increasingly distinguish between **evidence that AI governance has been designed** and **evidence that AI governance is operating effectively**.

Appropriate board-level questions may include:

- What material decisions can AI influence or execute?
- Where does ultimate human accountability sit?
- Which decisions require human authority?
- What conditions require intervention or escalation?
- How does management know when an AI-enabled system or its operating environment has changed materially?
- Can management demonstrate examples of governance mechanisms operating in practice?
- Can significant AI-influenced decisions be explained from context through to outcome?

The objective is not for boards to become involved in operational decisions.

It is to provide assurance that the organisation has designed an operating model in which authority, escalation and evidence remain effective as AI adoption increases.

14. Implications for executives

Executives face a related challenge. AI can increase organisational capability while simultaneously making responsibility less visible.

Delegating execution does not automatically delegate accountability.

Executives therefore need confidence that delegated authority is understood, human and AI participation is distinguishable, AI operates within appropriate boundaries, intervention mechanisms are practical, significant decisions remain traceable, and escalation reaches people with genuine authority.

Well-designed decision governance should not simply constrain AI.

It should enable organisations to **delegate with confidence**.

Clear boundaries can support greater autonomy where risk is understood and controlled, while ensuring that humans intervene when conditions require it.

Governance can therefore become an **enabler of responsible innovation**, rather than simply a constraint upon it.

15. Implications for auditors and assurance providers

AI will increasingly challenge assurance approaches based predominantly upon static documentation and periodic sampling.

Auditors will need to understand not simply whether organisations have policies and controls, but how those controls interact with dynamic human and AI decision-making.

Relevant assurance questions may include:

- Was the activity within delegated authority?
- What role did AI play?
- What signals were available?
- Was an intervention threshold reached?
- Did expected escalation occur?
- Who exercised decision authority?
- Was the resulting action consistent with that decision?
- What was the outcome?
- Is reliable evidence available?
- And critically:

Can the complete governance chain be explained?

This creates an opportunity for audit to move closer to operational assurance without becoming part of operational decision-making.

The fundamental question remains whether governance worked when it mattered.

16. Explainability as an organisational capability

AI explainability is frequently discussed as a technical characteristic.

That remains important.

But technical model explainability answers only part of the governance question.

An organisation may understand why a model generated a particular output yet still be unable to explain why the organisation acted upon it.

Conversely, an organisation may use a model whose internal operation cannot be completely interpreted while maintaining strong governance around its purpose, use, limitations, verification, authority and intervention.

Decision explainability therefore needs to operate at organisational level.

It should allow an organisation to answer:

- What was happening?
- What mattered?
- What did the organisation know?
- What did AI contribute?
- What did humans contribute?
- Who or what had authority?
- What triggered the decision?
- Why was the resulting action permitted?
- Why was escalation or intervention required - or not required?
- What happened next, and what evidence supports that account?

This is the broader meaning of explainability within DEEGS and DEXYS.

It is not simply the explainability of **AI**.

It is the explainability of **organisational decision-making involving AI**.

Conclusion

AI governance is entering a new phase.

The first challenge was establishing principles.

The next was establishing policies, risk frameworks, management systems and controls.

The emerging challenge is ensuring that governance continues to operate as AI becomes embedded within the organisation's actual operating model.

That challenge appears most clearly at **decision points**.

As decisions become faster, more distributed and increasingly shared between humans and technology, governance must make authority explicit.

It must recognise signals.

It must establish appropriate decision triggers.

It must enable proportionate intervention.

It must escalate when required.

It must preserve identifiable human accountability.

And it must produce evidence that explains what happened and why.

This is the purpose of **DEEGS - Decision Escalation & Explainability Governance**.

DEEGS treats escalation as a governance control and explainability as a governance outcome.

It connects risk to the decisions through which risk is actually managed.

DEXYS® - the Decision Explainability System - provides an operational governance layer through which DEEGS can be implemented across human and AI-enabled organisations.

Together with the **Core GRC® Framework**, this establishes a straightforward proposition:

- **Risk sits at the Core.**
- **Risk requires decisions.**
- **Decisions require authority.**
- **Authority requires governance.**
- **Governance requires evidence.**
- **Evidence makes the decision explainable.**

The future of AI governance will therefore depend not simply upon whether organisations can demonstrate that appropriate frameworks and controls exist.

It will depend upon whether they can demonstrate that those controls **worked at the moment they mattered.**

References and governance anchors

International standards and frameworks

ISO/IEC 42001:2023

Artificial intelligence - Management system

ISO/IEC 23894:2023

Information technology - Artificial intelligence - Guidance on risk management

ISO/IEC 27001:2022

Information security, cybersecurity and privacy protection - Information security management systems - Requirements

ISO 9001

Quality management systems - Requirements

ISO 19011

Guidelines for auditing management systems

NIST Artificial Intelligence Risk Management Framework (AI RMF)

National Institute of Standards and Technology

European governance and regulation

European Union Artificial Intelligence Act

General Data Protection Regulation (GDPR)

UK governance and regulatory context

Information Commissioner's Office - Accountability Framework

UK Government AI governance and regulatory policy

About Core GRC Solutions

Core GRC Solutions helps organisations integrate governance, risk and compliance into practical operating models across information security and privacy, human capability, quality management, and AI and data ethics.

The **Core GRC® Framework** places **Risk at the Core**, recognising that governance becomes effective when risk is connected to controls, decisions, authority, action and evidence.

DEEGS - Decision Escalation & Explainability Governance provides the decision-governance model.

DEXYS® - Decision Explainability System provides the operational governance and runtime assurance layer.

Together, they support organisations in moving from governance that is documented to governance that can be **observed, explained and assured**.

Core GRC Solutions Ltd

core-grc.co.uk

© 2026 Core GRC Solutions Ltd. All rights reserved.