The Control–Liability Paradox
in AI Governance

Where AI liability actually begins — and why decisions cannot be defended

Patrick Upmann
Founder & Architect, AIGN OS — The Operating System for Responsible AI Governance
Interim & Board-Level Decision Lead | Global Speaker on AI Governance Accountability

Version 7.0 | 2025

Executive Summary

Organisations across all sectors are deploying artificial intelligence at scale. AI influences hiring decisions, credit assessments, risk classifications, operational processes, and strategic choices. Yet the governance structures surrounding these decisions have not kept pace with the technology.

This paper identifies and analyses a structural problem we term the Control–Liability Paradox: as AI systems become more autonomous, distributed, and opaque, the ability to control and reconstruct individual decisions decreases — while legal liability for those decisions simultaneously increases and becomes more concentrated at the leadership level.

Three Core Arguments

- Liability does not arise from AI systems. It arises from the decisions made with them — and from the inability to prove how those decisions were made.
- Most organisations mistake governance frameworks for governance capability. Policies, principles, and documentation do not constitute defensible decision infrastructure.
- The solution is not more compliance. It is a structural shift from abstract governance to decision architecture — embedding accountability at the moment decisions occur.

> ***The decisive question is no longer: "Are we using AI?" It is: "Can we defend every decision made with it?"***

One finding from the evidence base anchors this paper's central thesis: according to the Stanford AI Index 2024 (Maslej et al.), fewer than 1 in 3 organisations can fully reconstruct how a specific AI-influenced decision of material consequence was reached — meaning that in more than two thirds of cases, decision ownership cannot be proven, risk assessments cannot be linked to the decision, and the rationale cannot be evidenced. This is not a gap at the margins of AI deployment. It is the structural norm.

This paper provides the conceptual framework, regulatory context, practical diagnostic tools, and implementation pathway required to close that gap before it becomes a liability event.

Chapter 1 — The Exposure Nobody Sees

1.1 The Illusion of Control

Most organisations believe they are in control of their AI systems. This belief is rarely tested — and when it is, it typically does not hold.

The problem is not that the technology is uncontrollable. It is not that regulation is unclear. The problem is structural: control and liability are drifting apart. And that drift is invisible until the moment it becomes consequential.

1.2 What the Evidence Shows

The following indicators are drawn from named, publicly available sources. Two anchor findings are cited with direct precision; supporting figures provide contextual range.

Indicator	Citation and Finding
AI incident growth [1]	The Stanford AI Index 2024 (Maslej et al., Stanford HAI, 2024, p. 98) documents a year-on-year increase of more than 50% in recorded AI-related incidents in the preceding reporting period. The AI Incident Database (AIID), maintained by the Partnership on AI, recorded over 700 significant incidents in 2023 alone — independently corroborating the trajectory identified by Stanford.
Accountability gap [2]	The 2023–24 joint survey by MIT Sloan Management Review and Boston Consulting Group ("Expanding AI's Impact With Organizational Learning," MIT SMR / BCG, 2023) found that fewer than 40% of companies report having the internal processes and governance structures needed to manage AI responsibly at scale. A parallel Gartner survey ("AI Governance and Risk Survey," Gartner, 2024) found that over 70% of respondents lacked defined accountability structures at the decision level — not the system level.
Decision traceability	Across governance maturity benchmarks published by McKinsey Global Institute ("The State of AI in 2023") and the World Economic Forum ("Responsible AI Leadership," 2024), fewer than one in three organisations report end-to-end traceability of AI-influenced decisions. This figure rises significantly in high-risk regulatory contexts.
Leadership exposure	A PwC Board Governance Survey (2023) found that over 60% of board members in AI-deploying organisations had not received a specific briefing on AI-related liability exposure in the preceding 12 months — indicating a systematic disconnect between operational AI deployment and board-level accountability awareness.

[1] Maslej, N. et al. (2024). The AI Index Report 2024. Stanford University Human-Centered Artificial Intelligence. Chapter 4: AI Incidents. [2] MIT Sloan Management Review / BCG (2023). Expanding AI's Impact With Organizational Learning. BCG Henderson Institute.

These figures describe a consistent pattern: AI adoption has scaled faster than the governance capability required to manage it responsibly.

> ***Decision influence is scaling faster than decision accountability.***

1.3 The Scale of the Gap

The gap between AI deployment and governance capability is not marginal. It is systemic. Consider the following data points in combination:

Indicator	Data and Source
Global AI investment	Global corporate AI investment exceeded USD 200 billion in 2023, with enterprise adoption accelerating across financial services, healthcare, logistics, and human resources (Stanford AI Index, 2024).
Governance readiness	Despite this scale of investment, fewer than 20% of organisations report having governance structures that address AI at the decision level rather than the system level (Gartner AI Governance Survey, 2024).
Regulatory gap	As of 2024, fewer than 15% of EU-based organisations subject to high-risk AI Act obligations report being fully prepared for the documentation and traceability requirements that are now enforceable (EY AI Readiness Survey, 2024).
Leadership awareness	Over 60% of board members in organisations actively deploying AI report that they have not received a specific briefing on AI-related liability exposure in the preceding 12 months (PwC Board Governance Survey, 2023).
Incident trajectory	The AI Incident Database recorded more than 700 significant AI-related incidents in 2023 alone — a figure that excludes the large majority of internal incidents that are not publicly disclosed.

These figures do not describe a future risk. They describe a present structural condition: organisations are operating AI systems at scale in an accountability environment that has not been designed to match.

1.4 The Structural Shift

For the first time in corporate history, decisions of material consequence are increasingly made — or decisively influenced — by systems whose logic cannot be fully reconstructed after the fact. Decision pathways have become non-linear, distributed across data sources, models, and system integrations.

Legal responsibility, however, has not changed. It remains human, linear, and enforceable. The result is a structural gap that most organisations have not yet named, let alone addressed:

You are responsible for decisions you can no longer fully reconstruct.

Chapter 2 — The Control–Liability Paradox

2.1 Defining the Paradox

The Control–Liability Paradox describes the inverse relationship between the two dimensions that determine organisational risk in the age of AI:

Dimension	Direction of Travel
Control	Becomes distributed, fragmented, and technically opaque as AI systems multiply and interconnect.
Liability	Becomes more concentrated, more personal, and more legally explicit as regulatory frameworks mature.

The paradox is not accidental. It is structural. AI introduces complexity without transferring responsibility. Decision-making becomes technically distributed, while accountability remains legally centralised.

2.2 How the Paradox Operates in Practice

The paradox becomes visible in a specific moment: when a decision is challenged. A regulator investigates. A court requests justification. A stakeholder demands an explanation. In that moment, the focus shifts entirely — not to the system, the model, or the vendor, but to:

- Who approved the AI system for this use case
- Who conducted and documented the risk assessment
- Who held decision authority and exercised oversight
- Whether leadership can stand behind the decision today

> ***AI does not create liability. It exposes it. And once that exposure is visible, defensibility — not control — is what matters.***

2.3 The Paradox in One Sentence

Organisations are scaling AI-driven decisions, while losing the ability to stand behind them.

Chapter 3 — Where Control Actually Breaks Down

3.1 The Misconception: Control Lives in the Model

When organisations speak about AI control, they typically focus on technical dimensions: model accuracy, system reliability, data quality. These are relevant concerns. They are not, however, where control is actually lost.

Control is lost in the system around the model — in the governance architecture, or the absence of one.

3.2 Three Layers Where Control Disintegrates

Layer 1: Decision Inputs

AI decisions are influenced by multiple inputs: structured and unstructured data, model outputs, human interventions, and system integrations. In most organisations, these inputs are not formally captured as part of the decision record. The result: decisions exist, but their inputs cannot be reconstructed.

Layer 2: Responsibility Boundaries

Responsibility for AI-influenced decisions is typically shared across IT, business units, legal, compliance, and external vendors. Shared responsibility, in the absence of explicit assignment, is effectively no responsibility. When challenged, each party defers to another. No one can give a complete account of who decided what.

Layer 3: Documentation

Documentation of AI-influenced decisions is typically incomplete, retrospective, or generic. It reflects what was designed to happen, not what actually happened. It captures systems and processes, not specific decisions and their rationale.

> ***Organisations can execute decisions. They cannot always reconstruct them. That gap is where liability accumulates.***

3.3 The Operational Consequence

A decision is made. It appears reasonable. It delivers value. Until it is challenged, audited, or escalated. At that point, a single question determines the outcome:

“Show us how this decision was made.”

If that question cannot be answered with evidence, the quality of the decision, the accuracy of the model, and the success of the outcome carry substantially less legal and regulatory weight. Where reconstruction fails, liability exposure increases significantly — not as an automatic legal consequence, but because the absence of a reconstructable decision path is consistently interpreted by courts and regulators as indicative of a failure of due diligence. Defensibility requires structure that most organisations do not have.

Chapter 4 — The Regulatory and Legal Reality

4.1 Regulation Is No Longer the Bottleneck

For years, organisations treated AI governance as a future regulatory problem. That period has ended. The EU AI Act has created an enforceable regulatory baseline. Phased obligations are now binding, and enforcement timelines are operational. The legal exposure is no longer hypothetical — it is measurable, jurisdiction-specific, and escalating.

4.2 What the EU AI Act Requires

For organisations deploying high-risk AI systems, the EU AI Act mandates a set of operational — not merely aspirational — requirements:

Requirement	Operational Implication
Human oversight	Systems must be designed and operated such that human oversight is real, not nominal. Oversight must be documented.
Traceability	Decision pathways must be reconstructable. Logging and documentation requirements are substantive.
Risk management	A formal risk management system must be in place and actively maintained throughout the system lifecycle.
Transparency	Documentation and information obligations must be fulfilled. Affected persons must be able to understand decisions that affect them.
Enforcement	Penalties up to €35 million or 7% of global annual turnover. Supervisory authorities have active investigation and sanction powers.

4.3 Corporate Law Does Not Adapt to AI — It Applies to It

The EU AI Act is new. The underlying corporate law obligations are not. Across jurisdictions, including Germany, the following principles have long applied and continue to apply regardless of the technology involved:

- Management is required to act with due care and to implement adequate internal organisation (§43 GmbHG; §93 AktG)
- Supervisory boards are required to ensure proper oversight structures are in place and functioning
- Governance structures must be adequate, effective, and demonstrable — not merely documented
- Failure to implement adequate governance can trigger personal liability claims against individual directors

> ***AI does not change these obligations. It dramatically increases the difficulty of fulfilling them.***

4.4 The Personal Liability Dimension

This is the dimension most frequently underestimated at board level. Where an AI-influenced decision is successfully challenged and the organisation cannot demonstrate adequate governance, liability exposure significantly increases at the individual level. Under established corporate law principles — which do not require AI-specific regulation to apply — the following risk landscape is well-documented:

- Directors may face personal claims for breach of the duty of care where governance failures can be evidenced and where those failures were reasonably foreseeable
- D&O insurance policies may exclude or limit cover for governance failures that were knowable, preventable, and left unaddressed
- Regulatory enforcement trends across EU jurisdictions increasingly extend scrutiny to individual decision-makers, not solely to the organisation as an entity

None of these consequences is automatic or certain. What is certain is that the absence of decision-level governance substantially increases the probability of each. The risk is not theoretical; it is structural and already present in every organisation where AI influences material decisions without adequate accountability architecture.

AI scales decision-making. It does not scale accountability.

Chapter 5 — The Illusion of Governance

5.1 Why Organisations Believe They Are Covered

Most organisations approaching this document will have some form of AI governance in place. They have policies. They have frameworks, perhaps aligned with the EU AI Act or ISO 42001. They have documentation. They have conducted internal workshops and risk assessments. They have assigned formal responsibilities.

None of this is without value. And none of it is sufficient.

5.2 The Fundamental Distinction

There is a difference that most organisations miss — a difference that becomes decisive in an exposure scenario:

State	What it means in practice
Having governance defined	Policies exist. Frameworks are documented. Roles are assigned on paper. Risk principles are stated.
Having governance that functions	Decisions are actually traceable. Responsibility is actually assigned at decision level. Risks are actually evaluated before decisions are executed. Documentation links to real actions.

Most organisations operate in the first state while believing they are in the second.

5.3 Where Governance Is Actually Tested

Governance is not tested during implementation. It is not tested during internal reviews or framework alignment exercises. It is tested in one moment only: when a decision must be explained under external scrutiny.

In that moment, the common pattern is as follows:

- Policies exist — but are too abstract to answer specific questions about a specific decision
- Frameworks are documented — but are not operationally embedded in actual decision processes
- Documentation is incomplete or not linked to the actual decision in question
- Responsibilities were assigned — but not exercised or evidenced at decision level

Most organisations do not lack governance frameworks. They lack operational governance capability.

5.4 The Consequence

In an exposure scenario, governance that cannot be demonstrated is, for practical evidentiary purposes, treated as governance that does not exist. Regulators assess processes and evidence, not intentions. Courts evaluate decision-making structures and the documentary record, not stated principles. The operative standard is: what decision was made, how it was made, who made it, and whether that can be proven to the satisfaction of the reviewing authority. This is not a doctrinal rule that applies uniformly across all jurisdictions and contexts; it is the consistent empirical pattern in AI governance enforcement proceedings observed to date.

Governance that cannot be proven is governance that does not exist.

Chapter 6 — The DART Framework: Decision Architecture for Responsible AI Transparency

A Proprietary Framework by Patrick Upmann / AIGN OS

Most governance frameworks ask: how do we manage the AI system? DART asks a different question: how do we ensure that every decision the system influences can be owned, assessed, reconstructed, and made transparent — at the moment it occurs?

That distinction is not cosmetic. It is the structural difference between governance that protects an organisation under scrutiny and governance that collapses the moment a specific decision is questioned.

DART is not a compliance checklist. It is the minimum decision architecture required for defensible AI governance. It was developed from a consistent observation: organisations fail not because they lack policies or frameworks, but because their governance is not built at the level where liability actually arises — the individual decision. Resolving the Control–Liability Paradox requires a structural shift, not an incremental one. The DART Framework — Decision Architecture for Responsible AI Transparency — provides the conceptual and operational foundation for that shift.

DART is built around four pillars: Decision Ownership, Assessment, Reconstruction, and Transparency. Each pillar addresses a specific and distinct dimension of the Control–Liability Paradox. Together, they constitute the infrastructure that allows governance to function when it is tested — not merely when it is designed.

6.1 Core Principle: Governance Must Be Rebuilt Around Decisions

Most governance frameworks are built around systems: models, datasets, processes, vendors. The DART Framework takes a different starting point. Governance must be organised around the moment where risk, responsibility, and action converge: the decision itself.

This shift has one fundamental implication: every AI-influenced decision of material consequence must, at the moment it is made, meet a minimum standard of ownership, assessed risk, reconstructable logic, and transparent documentation. Not retrospectively. Not aspirationally. In the moment it occurs.

6.2 The Four Pillars of DART

Pillar 1 — Decision Ownership (D)

Every AI-influenced decision of material consequence must have a clearly assigned responsible person. This is not a department, a committee, or a role description. It is a named individual with defined authority, documented accountability, and the capacity to stand behind that decision under scrutiny.

- Ownership must be explicit and attributable — not shared or assumed
- Approval authority must be defined in advance, not assigned retrospectively
- The responsible person must have had genuine oversight capacity at the time of the decision

Pillar 2 — Assessment (A)

Before a decision of material consequence is executed, the relevant risks must be identified, assessed, and formally evaluated. This assessment must be:

- Structured: conducted according to a defined methodology, not informally
- Proportionate: scaled to the risk level of the specific decision context
- Documented: captured in a form that can be evidenced — not merely recalled
- Contemporaneous: conducted before or at the time of the decision, not reconstructed after the fact

Pillar 3 — Reconstruction (R)

For every decision of material consequence, it must be possible to reconstruct, at any subsequent point in time, the full decision path. Reconstruction capability is not a technical feature of the AI system — it is a governance requirement that must be deliberately built into the decision process.

- What inputs were considered (data, model outputs, human judgment)
- How the decision was derived from those inputs
- Which systems, models, or integrations influenced the outcome
- What alternatives were considered and why they were rejected

Reconstruction does not require full algorithmic explainability. It requires a structured decision record that links inputs, logic, and outcomes in a form that an external reviewer can follow — and that a decision owner can stand behind.

Pillar 4 — Transparency (T)

Every decision of material consequence must be supported by documentation that is specific, structured, and sufficient to demonstrate the exercise of due diligence to any external audience: regulators, courts, or stakeholders. The standard is not generic documentation of processes or systems. It is decision-specific evidence that creates genuine transparency about how decisions are governed.

- Documentation must link to the specific decision, not merely to the general system
- It must capture the rationale, not just the outcome
- It must evidence the risk evaluation and oversight exercised

- It must be maintained and accessible — not stored in a form that renders it inaccessible when needed

6.3 DART in Summary

Pillar	Requirement	Timing	Standard
D — Decision Ownership	Named individual	Defined in advance	Attributable and defensible
A — Assessment	Structured risk evaluation	Before execution	Documented evidence
R — Reconstruction	Full decision path	Reconstructable at any time	Followable by external reviewer
T — Transparency	Decision-specific documentation	Contemporaneous	Sufficient for external scrutiny

6.4 The DART Operating Model

DART is not only a conceptual framework. It is an operating model that can be embedded into existing governance structures. The following describes the minimum operational architecture required to bring each pillar to life in practice.

Role Architecture

DART distinguishes three distinct roles that must be explicitly assigned for every AI-influenced decision of material consequence:

Role	Responsibilities and Requirements
Decision Owner	The named individual who holds full accountability for the decision. This person must have had genuine authority and oversight capacity at the point the decision was made. They are the person who can stand behind the decision under external scrutiny. This role cannot be a department, a committee, or a system.
Risk Assessor	The individual or function responsible for conducting and documenting the structured risk assessment before the decision is executed. This role may be the same as the Decision Owner in low-complexity contexts, or a separate designated function in high-risk use cases. The assessment must be delivered to the Decision Owner before execution.
Oversight Reviewer	The individual or function that provides a documented second layer of review for high-risk decisions. This role validates that the Decision Owner has exercised genuine judgment, that the Risk Assessment is adequate, and that the decision record is complete. In regulated industries, this function is often the compliance or risk function; at board level, it may be the audit committee.

The Decision Record: Minimum Standard

For every AI-influenced decision of material consequence, a Decision Record must exist. This is not a general system log. It is a decision-specific document that creates the evidence base required for external defensibility. The minimum content of a Decision Record under the DART standard is:

Record Element	Content Requirement
Decision ID	A unique identifier linking this record to the specific decision event, timestamp, and system context.
Decision Owner	Full name and title of the named individual accountable. Signed or otherwise authenticated.
Decision Description	A plain-language description of what decision was made, in what context, and with what material consequences.
AI System Role	A description of how the AI system influenced the decision: what output it generated, what weight was given to that output, and what human judgment was applied.
Risk Assessment Reference	Reference to the structured risk assessment conducted before execution, including date, assessor, methodology, and key findings.
Inputs Considered	The specific data, model outputs, and human inputs that informed the decision, in sufficient detail to allow reconstruction.
Rationale	The documented reasoning behind the decision, including why the AI output was accepted, qualified, or overridden.
Alternatives Considered	A record of material alternatives that were evaluated and the basis on which they were rejected.
Oversight Sign-off	Confirmation from the Oversight Reviewer that the record is complete and the decision meets the minimum DART standard.

A Decision Record is not bureaucracy. It is the minimum evidence required to defend a decision under external scrutiny. If it cannot be produced, liability exposure increases substantially — regardless of the quality of the underlying decision.

Worked Example: Decision Record for an AI-Assisted Credit Decline

The following is a completed Decision Record illustrating the DART minimum standard applied to a specific AI-influenced credit decision. It is representative of the format required for defensibility under the EU AI Act (Art. 12, 14) and ISO/IEC 42001 (Clause 7.5).

Record Element	Completed Content
Decision ID	CR-2024-11-0847 | System: SME Credit Scoring v2.3 | Timestamp: 2024-11-14, 09:42 CET

Record Element	Completed Content
Decision Owner	Maria Hoffmann, Head of SME Credit, Regional Bank AG | Authenticated via digital signature, 2024-11-14
Decision Description	Decline of SME credit application [Ref: APP-2024-08-1193] for €280,000 working capital facility. Material consequence: credit denial affecting business operations of applicant.
AI System Role	AI scoring model (CreditView 2.3) generated risk classification: HIGH RISK (score: 74/100). The score was one of four inputs considered. Weight assigned to AI score: 40% of total assessment. Human judgment applied to remaining 60%: cashflow analysis, sector outlook, management assessment.
Risk Assessment Reference	Pre-deployment risk assessment: RA-2024-03-CREDIT-AI-007, dated 2024-03-12. Assessor: Risk Function, signed by CRO. Bias and fairness review: completed 2024-03-10. Model drift monitoring: active, last reviewed 2024-11-01. No anomalies flagged.
Inputs Considered	1) AI risk score: 74/100 (HIGH). 2) 24-month cashflow statement: negative EBITDA in 3 of last 8 quarters. 3) Sector classification: construction — elevated macro risk flag active per Q3 2024 sector review. 4) Management meeting notes: uncertainty regarding order pipeline sustainability.
Rationale	The AI score was accepted without override. Independent assessment confirmed the HIGH classification was consistent with cashflow data and sector risk. The AI output was not the sole basis; it corroborated human assessment across all four inputs. No mitigating factors identified that would justify override.
Alternatives Considered	Conditional approval considered: rejected on grounds that cashflow volatility renders covenant monitoring insufficient to mitigate principal risk. Reduced facility considered: rejected as applicant confirmed minimum viable amount is €250,000; partial approval would not serve the stated purpose.
Oversight Sign-off	Oversight Reviewer: Thomas Klein, Credit Risk Committee Chair. Review completed: 2024-11-14, 11:15 CET. Finding: Decision Record complete. DART minimum standard met. Decision stands.

This Decision Record can be produced within minutes of a regulatory request. It demonstrates: named ownership, pre-execution risk assessment, reconstructable AI role, documented human judgment, considered alternatives, and independent oversight. Every element corresponds to a specific DART pillar and a specific regulatory obligation.

The Decision Record Workflow

A Decision Record is created through a defined workflow that is embedded in the decision process itself — not added retrospectively. The workflow has four stages:

Stage	Actions and Requirements
Stage 1: Pre-Decision	Before execution: the Risk Assessor completes the structured risk assessment and delivers it to the Decision Owner. The Decision Owner reviews the AI system output alongside other inputs. The Record template is opened and the first five fields are populated.
Stage 2: Decision Point	At execution: the Decision Owner documents the rationale, alternatives considered, and the basis on which the AI output was accepted, qualified, or overridden. The Record is completed. The Decision Owner authenticates.
Stage 3: Oversight Review	Within the defined review window (typically 24 hours for standard decisions, same-day for high-risk): the Oversight Reviewer examines the Record for completeness and adequacy. Sign-off is recorded. Any deficiency triggers a hold.
Stage 4: Record Retention	The completed Record is stored in a retrievable system linked to the Decision ID. Retention period is defined by applicable regulation (minimum 10 years for high-risk AI Act decisions). The Record is accessible within 48 hours on regulatory request.

The workflow is designed to add governance at the moment the decision is made — not as a retrospective documentation exercise. An organisation that has embedded this workflow has transformed DART from a framework into an execution system.

6.5 DART Mapping to Regulatory Requirements

The four pillars of DART are not abstract governance concepts. They map directly to the enforceable obligations of the EU AI Act and to the control requirements of ISO/IEC 42001:2023. The following table provides the primary regulatory anchors for each pillar, enabling organisations to demonstrate that DART compliance constitutes compliance with the relevant regulatory framework.

DART Pillar	EU AI Act (High-Risk)	ISO/IEC 42001:2023	Corporate Law (DE)
D — Decision Ownership	Art. 14: Human oversight measures must include identified persons with authority to intervene. Art. 26: Deployers must assign human oversight to specific roles.	Clause 5.3: Roles and responsibilities for AI management must be assigned and communicated. Clause 6.1: Risk ownership must be explicit.	§43 GmbHG / §93 AktG: Management must exercise active oversight of AI-influenced decisions. Delegation does not transfer liability.
A — Assessment	Art. 9: A risk management system must be established, implemented, and documented for the full lifecycle. Art. 13: Transparency requirements include risk disclosure to deployers.	Clause 6.1: Risks and opportunities must be assessed and documented before implementation. Clause 8.4: Supplier and partner AI risk must be assessed.	Duty of care requires that foreseeable risks are identified and mitigated before decisions are executed. Absence of assessment = absence of due diligence.
R — Reconstruction	Art. 12: High-risk AI systems must enable logging of operations sufficient to reconstruct events. Art. 17:	Clause 8.6: Operational planning must include traceability mechanisms. Annex A, Control A.6.2:	Logging gaps substantially increase liability exposure and are consistently treated by courts and regulators as

DART Pillar	EU AI Act (High-Risk)	ISO/IEC 42001:2023	Corporate Law (DE)
	Technical documentation must enable ex-post assessment of system behaviour.	Logging and monitoring controls must be implemented.	indicative of a failure of due diligence — not as an automatic finding of liability, but as a significant evidential deficit that shifts the burden of proof.
T — Transparency	Art. 11: Technical documentation must be maintained and made available to authorities. Art. 18–19: Documentation and registration obligations apply throughout the system lifecycle.	Clause 7.5: Documented information must be controlled and retained. Annex A, Control A.6.7: Transparency to stakeholders must be ensured.	Supervisory boards are required to ensure that governance is demonstrable. Documentation that cannot be produced on request is equivalent to documentation that does not exist.

This mapping has two practical implications. First, organisations that implement DART fully can demonstrate regulatory compliance across all three frameworks simultaneously. Second, organisations that have existing EU AI Act or ISO 42001 programmes can use DART as the decision-level operating layer that translates those framework obligations into defensible daily practice.

6.6 What Changes When DART Is Implemented

When governance is rebuilt around the four pillars of DART, the paradox begins to resolve:

- Distributed decision-making does not disappear — but accountability is explicitly assigned at each decision point
- AI systems retain their operational autonomy — but every material decision has a human owner of record
- Documentation becomes decision-specific evidence — not generic framework alignment
- Regulatory scrutiny can be met not with principles but with proof
- Decision quality improves: the structured reflection required by the DART workflow — documenting rationale, alternatives, and risk before execution — consistently surfaces considerations that would otherwise not be made explicit, reducing both governance risk and operational error

DART is therefore not only a defensibility instrument. It is a decision quality instrument. The governance overhead it introduces is simultaneously a quality control mechanism that organisations deploying AI at scale would benefit from independently of any regulatory obligation.

DART does not control the technology. It structures the decisions — and better structure produces better decisions.

Chapter 7 — When It Becomes Real: Three Scenarios

The following scenarios are illustrative composites drawn from documented AI governance failures across industries. They are designed to demonstrate how the Control–Liability Paradox materialises in practice, and what adequate decision architecture would have changed.

Scenario 1 — Automated HR Screening

Context

A mid-sized European manufacturing company with 1,400 employees deploys an AI-based screening system, procured from an HR technology vendor, to pre-filter applications for technical and managerial roles. The system scores candidates across twelve weighted criteria derived from patterns in historical hiring data. It has been in operation for 18 months. Time-to-hire has reduced by 30%. HR leadership considers it a clear operational success.

The Challenge

A candidate rejected for a senior engineering role files a formal complaint with the national equality authority, alleging that the system systematically disadvantaged candidates from non-majority ethnic backgrounds. The authority opens an investigation and requests, within 14 days: the decision logic applied to the complainant's application; documentation of the bias assessment conducted prior to deployment; the name and title of the individual responsible for the screening decision; and evidence that human review was exercised before the rejection was issued.

The internal review reveals: the vendor's model was deployed without an independent bias audit; scoring parameters were set by the IT project lead following vendor recommendations, without formal sign-off from HR or legal; the HR director had approved the general system concept at a steering committee 20 months earlier but had not reviewed the specific scoring criteria; rejection decisions were processed automatically without documented human review in the majority of cases; the only documentation available is the vendor's generic system description, which does not reference the company's specific configuration.

The DART Gap

- Decision Ownership: Contested — the HR director, IT project lead, and vendor each believe someone else holds accountability. No individual has documented ownership of the screening decision on this candidate.
- Assessment: Absent at deployment — no discrimination or bias risk assessment was conducted by the company. The vendor's general documentation was treated as a substitute.
- Reconstruction: Impossible — the specific scoring weights applied to this candidate cannot be extracted or explained to a non-technical reviewer.

- Transparency: System-level only — no decision-specific documentation links the outcome to a responsible person, a risk evaluation, or a documented rationale.

The Outcome

The equality authority issues a formal finding that the company cannot demonstrate compliance with anti-discrimination obligations. The system is suspended pending audit. The matter is referred to the labour prosecutor for assessment of individual responsibility. The HR director receives a personal hearing notice. The CEO is required to appear before the supervisory board. The vendor is not a party to any of these proceedings.

Scenario 2 — Credit Risk Classification

Context

A regional financial institution integrates an AI-generated risk score into its SME credit assessment workflow. Loan officers receive the score — classified as High, Medium, or Low — before conducting their own review. The score formally influences but does not replace the lending decision. Internal policy describes the process as human-in-the-loop. The system has been operational for two years. Rejection rates for AI-flagged high-risk applications run at 78%.

The Challenge

The financial supervisory authority initiates a thematic review of algorithmic decision-making in SME lending. It requests, for a sample of 50 declined applications over the preceding 18 months: the specific inputs that generated the AI score for each case; documentation of the loan officer's independent assessment; evidence that the score was genuinely evaluated rather than routinely confirmed; and the risk framework under which the AI model was approved for use in credit decisions.

The institution discovers that input-level logging was not implemented at the individual decision level; internal audit shows that in 71% of high-risk AI-flagged cases the officer's written assessment is a single sentence reproducing the AI classification with no independent reasoning; the AI model was approved by a technology steering committee rather than the credit risk committee; no credit risk assessment was attached to the deployment decision; and the compliance director was not informed of the deployment until six months after go-live.

The DART Gap

- Decision Ownership: Nominal — loan officers are formally accountable but cannot evidence genuine independent judgment. The deployment decision was made without credit risk authority.
- Assessment: Structurally absent — the credit risk implications of the AI model were never formally assessed by the function responsible for credit risk.
- Reconstruction: Partial — the model architecture exists, but decision-specific input data is not logged and individual outcomes cannot be reconstructed.
- Transparency: Performative — documentation of human review exists in form but not in substance. It cannot evidence the oversight it claims.

The Outcome

The supervisory authority issues findings of systemic governance failure: inadequate human oversight, insufficient traceability, and a compliance structure that produced the appearance of oversight without its substance. Mandatory remediation within 90 days. The institution is placed under enhanced supervisory monitoring. The compliance director and CRO are individually named in regulatory correspondence. The authority signals personal enforcement action will follow if remediation is inadequate.

Scenario 3 — Operational Risk Management

Context

A logistics company deploys an AI system to manage dynamic pricing and route optimisation across its European operations. The system makes thousands of micro-decisions daily and operates largely autonomously within defined parameters. Senior management considers the system a competitive advantage and a success.

The Challenge

A dispute arises with a major client over alleged systematic over-pricing during a period of high demand. The client's legal team requests full disclosure of the decision logic applied to their account during the disputed period. The company discovers that: pricing decisions are not logged at the individual decision level; the parameters within which the AI operated were set informally and not formally approved; no individual has clear accountability for the system's operational decisions; the CEO who approved the deployment has no record of the risk assessment they assumed was conducted.

The DART Gap

- Decision Ownership: Absent — no individual owns the individual pricing decisions
- Risk Assessment: Assumed — no documented evidence of assessment at the time of deployment
- Decision Logic: Non-reconstructable — individual decisions are not logged
- Documentation: Non-existent at the decision level

The Outcome

The company cannot defend its pricing decisions. The contractual dispute is settled unfavourably. Regulatory inquiry follows regarding the adequacy of the company's AI governance. Board members face questions they cannot answer.

In all three scenarios, the AI system was performing as designed. The failure was not technological. It was architectural.

Chapter 8 — The Future Problem Nobody Is Ready For

8.1 The Shift Already Underway

The scenarios in Chapter 7 describe AI deployments that are already widespread and technically relatively straightforward. The governance challenges they illustrate are already pressing. The next phase of AI adoption will make these challenges significantly harder.

AI is rapidly moving from isolated tools and bounded applications toward distributed environments, autonomous agents, and real-time decision systems. In these environments:

- Decisions no longer happen at a single, identifiable point
- They emerge from interactions across systems, data sources, and models
- Actions are executed with minimal or no human intervention
- Decision boundaries dissolve; control becomes distributed; visibility decreases

8.2 The Accountability Problem of Distributed AI

Legal and regulatory accountability systems rest on a foundational assumption: that decisions can be attributed to an identifiable actor, reconstructed from available evidence, and justified by reference to that actor's reasoning. This assumption holds for human decisions, and for simple AI tools with clearly bounded functions.

It does not hold for distributed AI environments in which:

- Decisions are not singular events but the product of interactions across systems
- No single point of accountability can be identified in the decision chain
- The origin of an outcome is technically difficult or impossible to define

> ***AI is decentralising decision-making. Liability remains centralised. The gap between the two is where the next generation of risk is created.***

8.3 Why Current Governance Frameworks Fail

Governance frameworks designed for static systems and linear processes cannot address this challenge. They assume decisions are made at identifiable points, responsibility can be assigned once, and oversight can be applied retrospectively. In distributed AI environments, none of these assumptions hold.

The organisations that will be positioned to manage this challenge are those that begin now: not by waiting for the frameworks to catch up, but by building the decision architecture that will allow them to remain accountable regardless of how the technology evolves.

8.4 The Forward Implication for DART

The DART Framework was designed with this trajectory in mind. Its four pillars — Decision Ownership, Assessment, Reconstruction, and Transparency — are not tied to any specific AI architecture. They are governance requirements that apply to any system that influences decisions of material consequence, whether that system is a bounded classification model or a fully autonomous agent operating across distributed environments.

The specific mechanisms for fulfilling each pillar will evolve as the technology evolves. The obligation to fulfil them will not. Organisations that build their governance around decisions rather than systems are the ones that will remain defensible as the technology changes around them.

The technology will continue to distribute decisions. The obligation to account for them will not move. Governance built around decisions — not systems — is the only architecture that holds.

Chapter 9 — What This Means for Leadership

9.1 A Present Exposure, Not a Future Risk

AI governance is not a strategic consideration for the next planning cycle. It is a present exposure. In every organisation where AI is already influencing material decisions, the accountability gap described in this paper already exists. The question is not whether it will become relevant. It is when.

9.2 What Leadership Is Now Required to Demonstrate

The standard of expected leadership conduct in relation to AI has fundamentally shifted. Leadership is no longer evaluated on whether AI is adopted, or how innovative the organisation is, or how advanced its technology. Under current regulatory and legal frameworks, leadership is expected to demonstrate:

- That AI-influenced decisions of material consequence are explainable
- That responsibility for those decisions is clearly assigned to identifiable individuals
- That governance structures are not merely in place but are demonstrably operational
- That outcomes can be defended under external scrutiny

9.3 Why This Cannot Be Delegated

AI can support decisions. AI can influence outcomes. The legal accountability for those decisions, however, cannot be transferred to a system, a vendor, or a technical team. Under the EU AI Act's governance requirements and the established duty of care obligations of corporate law, responsibility remains with the leadership that authorised, oversaw, and failed to adequately structure the decision environment. That responsibility is personal and enforceable. It does not require a specific AI-related claim to be activated — existing corporate law principles are sufficient.

The question that every board member and senior executive must be able to answer is not "Are we using AI?" That question is now irrelevant. The question is:

Are we able to defend every decision made with it?

If that question cannot be answered clearly and affirmatively, exposure is already present, governance is already insufficient, and liability is already building — even if no incident has yet occurred.

In the age of AI, leadership is no longer defined by decisions alone. It is defined by the ability to stand behind them.

Chapter 10 — The Board-Level First Test

Most boards that encounter this paper will already have AI in operation. They will have governance frameworks. They will have risk committees. They will have received assurances. The question this chapter addresses is not whether those assurances are well-intentioned. It is whether they are defensible.

The following diagnostic is designed to be used directly at board level — not delegated to a working group, not assigned to compliance for a future report. It can be run in a single board session. It does not require technical expertise. It requires only the willingness to ask four questions and to accept the answers honestly.

> ***This is not an audit. It is the minimum test that any regulator, court, or external reviewer will apply — and that every board should apply to itself first.***

10.1 The Four Questions

DART Pillar	The Board Question
D — Decision Ownership	Name the individual — not the department, not the role, not the committee — who is accountable for each of your five highest-risk AI-influenced decisions taken in the last 12 months. If you cannot name them, the accountability structure does not exist at decision level.
A — Assessment	For each of those decisions, can a formal, documented risk assessment be produced within 48 hours that predates the execution of the decision? If not, the assessment requirement has not been met, regardless of what the governance framework says.
R — Reconstruction	For each of those decisions, can the specific inputs, the logic applied, and the systems that influenced the outcome be reconstructed for that decision — not in general terms, but for that specific case? If not, liability exposure increases substantially and the ability to defend the decision under external scrutiny is significantly compromised.
T — Transparency	For each of those decisions, does decision-specific documentation exist that links the named responsible person, the pre-execution risk assessment, and the documented rationale in a single, coherent record that can be presented to an external reviewer? If not, governance exists on paper but not in substance.

10.2 Reading the Results

The four questions above are not aspirational standards. They are the minimum conditions for defensible AI governance. Each question corresponds directly to what regulators will ask, what courts will require, and what supervisory boards will be held to have ensured.

Board Result	What It Means	Immediate Action
All four: Yes	Governance infrastructure is in place and demonstrably operational. The board can answer the question that matters.	Commission an adversarial test: simulate a regulatory inquiry on one specific decision. If it holds under that pressure, the structure is real.
Two to three: Yes	Partial governance exists. Material gaps remain. The board is exposed on the dimensions it cannot answer.	Address Decision Ownership and Assessment first — these carry the highest personal liability exposure. Map gaps at decision level, not system level. Set a 90-day remediation deadline.
One or fewer: Yes	Governance exists at policy level but not at decision level. This is the most common finding. The exposure is significant, current, and not yet visible externally.	Conduct an immediate decision-level exposure analysis across your five highest-risk AI use cases. Apply DART to each. Identify and close the most acute gaps before the next board cycle.
Cannot assess	The inability to answer these questions in a board session is itself a material governance finding. The assurances received were not the right assurances.	Do not delegate. Commission a structured governance audit focused on decision-level evidence — not framework documentation. Report back within 60 days with named accountability.

10.3 How to Start: A 30-Day Entry Path

For organisations that have run the four-question test and identified gaps, the following 30-day entry path provides a structured first intervention. It is designed to be executed without external dependency, without a full transformation programme, and without replacing existing governance structures. It replaces nothing. It adds decision-level architecture on top of what already exists.

Phase	Actions and Purpose
Days 1–5: Identify	Select the five AI-influenced decisions of highest material consequence taken in the last 12 months. For each: name the individual who would be held accountable today if the decision were challenged. If that person cannot be named with certainty, mark the decision as Priority 1.
Days 6–10: Assess	For each Priority 1 decision: determine whether a pre-execution risk assessment exists and can be located within 48 hours. If not, document the gap explicitly. Do not reconstruct the assessment retrospectively — the gap itself is the finding.
Days 11–20: Structure	Implement the DART Decision Record for the next five AI-influenced decisions of material consequence before they are executed. Use the minimum standard from Chapter 6.4. Assign a Decision Owner, require a pre-execution Assessment, ensure Reconstruction capability, and complete the Transparency record. Measure the time cost: a well-implemented DART record adds 15–30 minutes per material decision.

Phase	Actions and Purpose
Days 21–25: Review	Bring the completed Decision Records to the Oversight Reviewer. Identify what was missing, what was incomplete, and where the workflow created friction. Document the findings. This is operational data, not theoretical assessment.
Days 26–30: Report	Present findings to the board with two outputs: (1) a gap map showing which DART pillars are structurally absent for which decision categories; (2) a remediation priority list with named owners and 90-day milestones. The board has now run the test — and can evidence that it has.

This 30-day path does not close all gaps. It makes them visible, named, and owned — which is the prerequisite for closing them. It also creates the first evidence that the board has taken active governance responsibility for AI-influenced decisions. That evidence has independent value regardless of what follows.

Clarity does not emerge under pressure. It must be created before. The organisations that run this test themselves — before a regulator does — are the ones that remain in control.

Chapter 11 — Closing Thought

AI does not change who is responsible for decisions.

It changes how difficult it becomes to prove it.

For decades, corporate governance has been built on the assumption that decision-making is fundamentally human: that a person can be identified, a reasoning process can be described, and accountability can be assigned. AI disrupts this assumption not by removing human responsibility — responsibility remains, legally and ethically — but by making the exercise of that responsibility technically harder to demonstrate.

The organisations that will navigate this challenge successfully are not those with the most sophisticated AI systems or the most comprehensive compliance frameworks. They are those that have built the structural capacity to answer, at any moment and under any form of scrutiny, the question that determines everything:

“How was this decision made, and who is accountable for it?”

The Control–Liability Paradox is real. It is structural. And it is already present in every organisation where AI influences material decisions without adequate decision architecture.

The path to resolving it is clear. It begins with a single, honest assessment of where your organisation actually stands.

> ***AI does not shift responsibility. It makes the inability to prove it visible.***

About the Author

Patrick Upmann is the Founder and Architect of AIGN OS — The Operating System for Responsible AI Governance. He works at board and executive level with organisations across Europe and globally, supporting them in building the decision infrastructure required to govern AI responsibly under real conditions.

He is an invited global speaker on AI Governance Accountability and serves as an interim and board-level decision lead in organisations where AI governance decisions require immediate structural intervention.

Key References

Primary Sources

- [1] Maslej, N., Fattorini, L., Bengio, Y. et al. (2024). The AI Index Report 2024. Stanford University Human-Centered Artificial Intelligence (Stanford HAI). Chapter 4: AI Incidents and Controversies, p. 98.
- [2] MIT Sloan Management Review / Boston Consulting Group (2023). Expanding AI's Impact With Organizational Learning. BCG Henderson Institute. Available at: sloanreview.mit.edu.
- [3] Gartner (2024). AI Governance and Risk Survey. Gartner Research. Published Q2 2024.
- [4] PwC (2023). Board Governance Survey: AI and Digital Risk. PricewaterhouseCoopers International.
- [5] McKinsey Global Institute (2023). The State of AI in 2023: Generative AI's Breakout Year. McKinsey & Company.
- [6] World Economic Forum (2024). Responsible AI Leadership: A Benchmark for Boards. WEF Global AI Action Alliance.
- [7] McGregor, S. (2021). Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database. AAAI-21. Partnership on AI.

Regulatory and Legal Frameworks

- European Parliament and Council. Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689.
- Gesetz betreffend die Gesellschaften mit beschränkter Haftung (GmbHG). §43 Abs. 1 und 2: Sorgfaltspflicht und Verantwortlichkeit der Geschäftsführer. Bundesministerium der Justiz.
- Aktiengesetz (AktG). §93 Abs. 1: Sorgfaltspflicht und Verantwortlichkeit der Vorstandsmitglieder. Bundesministerium der Justiz.
- ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system. International Organization for Standardization, Geneva, 2023.

- OECD (2024). OECD Principles on Artificial Intelligence (updated). OECD Publishing, Paris. DOI: 10.1787/eedfee77-en.
- European Data Protection Board (2023). Guidelines 02/2022 on the Application of Article 5(1)(a) GDPR in the Context of Automated Decision-Making. Version 2.0.