Journal of Interdisciplinary History and Human Societies
ISSN: 3069-3586

PRIME OPEN ACCESS

Volume 2, Issue 2

Research Article

Date of Submission: 11 June, 2026
Date of Acceptance: 10 July, 2026
Date of Publication: 21 July, 2026

From Meaningful Human Control to Decision-Time Accessibility The Architectural Conditions of Responsible Oversight in Automated Systems

Jean Claude Havyarimana*

Independent Researcher

***Corresponding Author:**
Jean Claude Havyarimana Independent Researcher

Citation: Havyarimana, C., J. (2026). From Meaningful Human Control to Decision-Time Accessibility The Architectural Conditions of Responsible Oversight in Automated Systems. *J Interdiscip Hist Hum Soc, 2*(2), 1-19.

Abstract
Meaningful Human Control (MHC) requires that the operation of an automated system be traceable to a human agent in a position to exercise the capacities required for responsibility. Tracing demands more than formal authority; it demands that those capacities remain accessible at the moment of action. This article names that requirement decision-time accessibility (DTA): the architectural condition under which responsibility-relevant capacities reach occurrence under institutional and technological load. DTA is derived directly from MHC's tracing requirement and decomposes into six elements: Truth, Relation, Integration, Doing, Orientation, and Attention & Agency. Where the time required to assemble these elements exceeds the decision window the architecture provides, the system has crossed what the article calls the Judgment Cliff. The article develops the framework through three retrospective cases (Challenger, the Boeing 737 MAX, the 2008 rating-model chain), set against Crew Resource Management as a counter-case, and applied prospectively, with prospectively stated predictions and explicit disconfirmation criteria, to algorithmic clinical decision support under the EU AI Act's Article 14. Where DTA collapses, operator-blame allocation becomes fictive accountability: a misattribution of architectural pathology to individual fault. Meaningful control requires architectural design, not authorizing structure alone.

Keywords: Meaningful Human Control, Decision-Time Accessibility, Tracing Condition, Sociotechnical Systems, Automation, Institutional Architecture, Responsibility, Attention, Condition Ethics, Fictive Accountability

Introduction
Human oversight is often treated as present when a trained and authorized person remains formally responsible for an automated system's action. But responsibility does not become meaningful merely because authority is assigned. It becomes meaningful only when the capacities required for responsible judgment remain accessible at the moment of action. This article argues that the architectural conditions under which such accessibility obtains form a distinct level of analysis, the level at which Meaningful Human Control (MHC) is either operationally satisfied or attestationally simulated. MHC, one of the most influential normative frameworks in the governance of automated systems [1]. requires that any action of an automated system be traceable to a human agent who has the moral and cognitive capacities to be responsible for it, and who, at the time of action, was in a position to exercise those capacities meaningfully. The italicized clause carries decisive normative weight. Without it, tracing collapses into causal traceability. With it, the operational question becomes pressing: under what conditions is the agent actually in such a position?

This article calls that condition decision-time accessibility, or DTA. It names the architectural variable that determines whether the cognitive and moral capacities required for tracing reach occurrence at the moment of action, given the institutional, informational, and temporal architecture of the decision environment. The capacities certified by training are durable; their occurrence is fragile. The framework specifies the architectural conditions on which occurrence

depends.

DTA decomposes into six conditions derived from what MHC's tracing condition already presupposes: Truth, Relation, Integration, Doing, Orientation, and Attention & Agency (Table A, §3.3). Where the time required to assemble these conditions within the available decision window cannot be afforded, the system has crossed what the article calls the Judgment Cliff. Beyond the cliff, action continues but tracing fails: the operator remains formally responsible while substantively unable to exercise the capacities oversight presupposes. The article terms this institutional pathology fictive accountability, a name for the predictable consequence of operationalizing MHC's tracing condition through certification rather than architecture.

The article positions the framework as an operational bridge between existing literatures and MHC's normative requirement, not as a substitute for them. Situation awareness, cognitive load theory, resilience engineering, organizational accident theory, and dynamic risk framework each illuminate components of the problem. None organizes them as the architectural accessibility condition required for MHC tracing at the moment of action. That is the contribution [2-5].

Three clarifications about scope. First, the framework does not argue that human oversight is preferable to automation in general. Many decision classes are demonstrably safer under high automation than under high human discretion. The argument concerns the architectural conditions for meaningful control where control is required, not the prior question of whether control should be required. Second, the framework does not propose to lift responsibility from operators in any general way. Operators retain duties of competence, honesty, reporting, and escalation; the claim is the narrower one that when action occurs beyond the Judgment Cliff, primary blame allocation to the operator misdescribes the architecture of failure. Third, MHC is not the universal standard for technologically mediated action. The framework applies only where a system, regulation, or institution already invokes the tracing requirement. Where MHC is not invoked, DTA is not at issue.

The article proceeds in eight further parts. Section 2 reviews MHC, locates the architectural gap, and positions DTA against five adjacent literatures. Section 3 derives the six conditions, develops the Judgment Cliff formally, and specifies a six-step operational protocol. Section 4 applies the framework to three retrospective cases (Challenger, the Boeing 737 MAX, and the 2008 rating-model chain), set against Crew Resource Management as a counter-case, and adds a cross-case diagnostic matrix. Section 5 develops the prospective application to clinical decision support under the EU AI Act's Article 14, with prospectively stated predictions and disconfirmation criteria. Section 6 develops governance implications including architectural audit and liability shifts. Section 7 addresses scope, testability, and disconfirmation risks. Section 8 concludes.

The article's central claim is summarized in a sentence. Meaningful control requires accessible agency, not merely assigned responsibility. The architectural conditions for that accessibility are the primary site of ethical agency in technologically mediated governance, and the primary object of any oversight regime that aspires to be more than attestational.

MHC and the Architectural Gap

The MHC Framework

Meaningful Human Control was developed initially in the context of autonomous weapons systems [1]. and has since been extended across automated decision-making infrastructures [7-9]. It specifies two conditions.

The tracking condition requires that the system's behavior be responsive to the relevant moral reasons of the human agents in whose name it acts. A system that achieves outcomes orthogonal to its operators' reasons fails to track, even if the outcomes are individually defensible. The tracing condition requires that any action of the system be traceable to a human agent who has the proper moral and cognitive capacities to be responsible for it, and who, at the time of action, was in a position to exercise those capacities meaningfully. A system whose outputs cannot be traced to such an agent fails the tracing condition even if every individual output is defensible.

The framework's productive shift was to move the question from whether a human is present (the "human-in-the-loop" framing) to whether the human's role is substantive. It thereby exposed the inadequacy of governance arrangements that satisfy formal oversight requirements while leaving the operator in a position from which responsible action is not possible.

The Architectural Gap

MHC tells us that the agent must be in a position to exercise the relevant capacities. It does not tell us what conditions constitute being-in-a-position-to-exercise. The clause is doing crucial work; without it, tracing collapses into causal traceability. With it, the operational question becomes pressing: under what institutional, informational, and temporal conditions does the moral and cognitive capacity of a certified, authorized agent actually reach the moment of action? This is the architectural gap. It is not a defect of MHC (the framework was developed at the level of normative specification), but it has practical consequences. Without an operational specification, MHC compliance becomes attestational, and the conditions for substantive tracing are left to the assumption that they hold.

One objection to locating the gap here should be met directly. MHC's tracing condition, in its original formulation, can be satisfied by any agent in the design-and-use chain who possesses the relevant capacities and could foresee the system's behavior; tracing need not run to the front-line operator at the moment of action. If that is so, the objection runs, then upstream designers can already bear the tracing relation, and there is no special problem about the operator's decision-time condition. The reply is twofold. First, where an institution assigns the operator the tracing role, as oversight regimes such as Article 14 of the EU AI Act characteristically do, the operator is precisely the agent to whom action is traced, and whether that agent is in a position to exercise the relevant capacities then becomes unavoidable. Second, and more fundamentally, the framework specifies when tracing should run upstream rather than to the operator: where the architecture has withdrawn decision-time accessibility, tracing to the operator fails, and the responsibility the tracing condition demands belongs to those who configured the foreclosure. Far from conflicting with the upstream reading of tracing, the framework supplies the criterion that reading lacks, namely the architectural condition under which tracing to the operator is and is not available. This is the analytical basis for the liability allocation developed in Section 6.3.

Existing Operationalizations

The literature has begun to provide operational specifications [7]. develop reason-responsiveness conditions for dual-mode vehicles, distinguishing the levels at which tracking and tracing must hold and identifying minimum conditions for each [8]. propose variable-autonomy frameworks in which authority shifts dynamically between human and machine according to context [10]. proposes a generic operationalization applicable across autonomous-system domains, identifying components of the responsibility chain and conditions for their integrity. Related work in automated driving operationalizes tracing through evaluation-cascade tables that map the relationships among system components, human operators, design decisions, and emergent [11].

These efforts share a common move. They identify agents, roles, capacities, system components, and chains of responsibility. They specify who is responsible, for what, with which capacities, and in which institutional and technical roles. What they do not isolate is whether those capacities remain accessible to the responsible agent at the moment of action, under the load the system imposes. The relationship between the responsibility-relevant capacities (which are durable) and their occurrence (which is fragile) remains underspecified.

This article specifies that further condition. The contribution is not that MHC has not been operationalized (it has been, productively, in several adjacent domains) but that the existing operationalizations do not yet isolate the variable that determines whether the capacities they identify actually arrive at the moment of decision.

Adjacent Models and the Question of Distinctiveness

The six conditions developed in Section 3 draw their substantive content from existing literatures. The contribution is integrative, not displacive. Five adjacent models warrant explicit positioning.

Situation awareness theory [2]. decomposes the perceptual side of operator capacity into three levels (perception of elements, comprehension of meaning, and projection of future states) and has produced operationalizable measurement instruments. Its primary domain is perception under load; it does not address authorial or moral dimensions of agency, and it is calibrated to performance rather than to MHC's normative requirement of tracing.

Naturalistic decision-making [12]. examines how experts decide under time pressure in real environments and identifies recognition-primed decision as the dominant mode. It is descriptive rather than evaluative; it does not provide an architectural diagnostic and does not address responsibility allocation.

Resilience engineering [4]. theorizes functional variability and adaptive capacity in complex systems and has developed sector-specific methods (most notably the Functional Resonance Analysis Method). It is method-heavy and most fully developed in aviation and process industries; its normative orientation is to system performance and adaptive capacity, not to the conditions for occurrent moral responsibility.

Organizational accidents theory [5]. proposes the "Swiss cheese" model of latent and active failures, identifying how organizational, supervisory, preconditional, and unsafe-act layers can align to produce catastrophic outcomes. It is unit-of-analysis-flexible (individual, team, organization) but does not specify temporal locus or the conditions under which capacities reach occurrence.

Risk management framework [13]. develops the dynamic conception of safety as a moving boundary within an envelope bounded by economic pressure, workload pressure, and the boundary of acceptable performance. It is fundamentally architectural but oriented to the systemic dynamics of drift rather than to the moment-of-action conditions for responsibility. Table 1 positions the framework developed here against these five.

Model	Unit of analysis	Temporal locus	Normative orientation	Predictions about MHC tracing
Situation awareness (Endsley)	Individual perception	Moment-of-action	Performance	None directly; partial mapping to Truth
Naturalistic decision-making (Klein)	Expert cognition	Moment-of-action	Descriptive	None directly; concerns disposition more than occurrence
Resilience engineering (Hollnagel et al.)	System	Time-extended	Adaptive capacity	None directly; concerns system performance, not tracing
Organizational accidents (Reason)	Multi-level latent/ active	Time-extended	Failure prevention	Indirect; layer alignment ≠ accessibility collapse
Risk management (Rasmussen)	Sociotechnical system	Drift over time	Boundary preservation	Indirect; envelope drift ≠ moment-of-action accessibility
DTA / six conditions (this article)	Agent–environment dyad	Moment-of-action	MHC tracing	Direct; predicts which condition's collapse forecloses tracing

Table 1: The DTA Framework Positioned Against Five Adjacent Models

The unique contribution claim is now visible and defensible: temporal locus at the moment of action, calibrated to MHC's normative requirement of tracing, with explicit architectural-not-individual responsibility allocation. The framework does not displace situation awareness, NDM, resilience engineering, Reason's organizational accidents model, or Rasmussen's framework. It draws from each at the level of condition-specific theory while organizing the conditions into a single decision-time diagnostic calibrated to MHC's tracing condition.

Why DTA Is Not Merely Situation Awareness, Cognitive Load, or Human Factors

The framework's most likely first-line objection is that it relabels established work. Truth resembles the elements-perception level of situation awareness; Attention & Agency recalls cognitive load theory and the attention-economy literature; Doing recapitulates the active-failure analyses of Reason; Orientation echoes Rasmussen on goal-displacement under economic pressure; Integration is adjacent to the explainability literature in machine learning. The objection deserves a precise answer rather than denial.

The answer is that each adjacent literature provides part of the descriptive or performance-oriented apparatus DTA depends on, but none organizes its components as the architectural accessibility condition required for the normative operation that MHC's tracing condition specifies. The differentiator is not the components but the integrative requirement they jointly satisfy, calibrated to a normative end the source literatures do not pursue.

Situation awareness theory [2]. analyzes how operators perceive elements of the environment, comprehend their meaning, and project future states. It is calibrated to operator performance and produces measurable instruments. Its concern is the perceptual side of capacity; it does not address whether what is perceived can be acted upon within the time the architecture affords, whether the affected stand as moral weight at decision-time, or whether the reasoning is the operator's own. DTA includes the perceptual condition (Truth) as one of six and asks whether all six are jointly accessible for responsibility, a question situation awareness was not designed to ask and does not answer.

Cognitive load theory [2]. explains why attentional resources fail under task demands and informs instructional and interface design. It explains why AA collapses. DTA asks the further question of when AA collapse defeats MHC tracing, the point at which load-induced failure forecloses the conditions for occurrent responsibility rather than merely for task performance. The shift is from performance to tracing, and the practical consequence is the introduction of an evaluative criterion that cognitive load theory does not contain: not the threshold at which performance degrades, but the threshold at which oversight ceases to be substantive.

Organizational accident theory [5]. traces how latent organizational, supervisory, and preconditional failures align with unsafe acts to produce catastrophic outcomes. The Swiss-cheese model is unit-flexible across levels of the institution and has been productive across high-risk industries for three decades. Its concern is failure causation. DTA asks the prior question of whether any responsible agent remained, at the moment of action, in a position to exercise the capacities the tracing condition requires. Reason's framework explains how failures propagate through layers; DTA specifies the architectural state in which the front-line layer can or cannot bear responsibility for its action. The two are complementary, not competitive; the second does not displace the first, but neither substitutes for the other.

Explainability research, the post-hoc and intrinsic-interpretation literature in machine learning, addresses the opacity of automated reasoning and proposes methods for rendering model outputs interpretable to operators. It bears directly on Integration. But Integration is one of six conditions, and explainability, even where successful, does not by itself preserve Truth (the system's representation of reality may remain misleading), Doing (override may remain institutionally penalized regardless of how clearly the operator understands what to override), Orientation (mission may remain displaced by throughput metrics regardless of interpretive access), Relation (the affected may remain abstracted to data

points regardless of how transparent the reasoning), or AA (attention may remain saturated by the cumulative alerting load regardless of explanation availability). An explainable system in a DTA-collapsed architecture satisfies one condition of six. The framework specifies why explainability is necessary for tracing and why it is not sufficient.

Resilience engineering [4]. studies the adaptive capacity of sociotechnical systems and the conditions under which systems sustain functioning under variability. Its normative orientation is to system performance and adaptive capacity over time. DTA's orientation is to responsibility accessibility at the moment of action. A resilient system may sustain performance through architectural arrangements that nonetheless foreclose MHC tracing. High-frequency trading is the clearest example, since its adaptive capacity is preserved at tempos that structurally exclude human co-presence. Conversely, a system whose adaptive capacity is low may preserve DTA where the relevant decisions remain within the architecture's deliberative window. The frameworks evaluate different properties; one is not a substitute for the other. The differentiator across these comparisons is consistent. The adjacent literatures study perception, attention, failure causation, opacity, and adaptation as properties of operators or systems. DTA studies the architectural conditions under which the normative operation of tracing can be performed, the conditions under which an agent can be meaningfully held responsible for an action her institution authorizes her to take. For that reason, the framework is not a competitor to situation awareness, cognitive load theory, organizational accidents research, explainability, or resilience engineering. It is the operational bridge between those literatures and a normative requirement none of them is calibrated to: that responsibility for technologically mediated action remain substantively traceable to a human agent at the moment that action occurs.

The bridge has its own architecture. It selects components from the adjacent literatures, organizes them under a joint-accessibility requirement, locates them at the moment of action, and calibrates them to MHC's tracing condition. The selection, the joint requirement, the temporal locus, and the calibration are the framework's contribution. Each element separately predates the framework; their integration under MHC tracing does not.

Fictive Accountability: The Institutional Pathology

A separate concern arises not from the academic literatures themselves but from their downstream uptake in regulatory and accountability regimes. In their primary research forms, situated cognition, cognitive ergonomics, and distributed cognition explicitly treat environmental architecture as constitutive of cognitive performance. Hutchins's Cognition in the , Reason's Managing the Risks of Organizational, and are among the more uncompromising statements of this view [14-16].

Yet the regulatory uptake of these literatures has frequently been selective. Accountability regimes have absorbed the language of human factors while preserving the institutional convenience of treating the operator as the residual locus of fault. Post-incident analyses commonly identify environmental contributors but assign culpability to individuals whose actions, given the environment, were close to over-determined.

The result is fictive accountability: oversight regimes that are formally robust while operating in environments where the cognitive and moral capacities they presuppose cannot reach the moment of action [17] captures this pattern in her notion of the "moral crumple zone": the structural absorption of liability by the human element in automated systems, designed to protect those further up the architectural chain from the consequences of their design choices. Fictive accountability is the institutional pathology this article diagnoses. It is the predictable outcome of operationalizing MHC's tracing condition through certification and authorization rather than through the architectural conditions for occurrent capacity. Accordingly, the critique here is not directed at the source disciplines but at the institutional appropriation that has flattened them in service of attestational compliance.

Decision-Time Accessibility and the Six Conditions

Judgment as Event, Capacity as Disposition

MHC's tracing condition presupposes that the agent to whom an action is traced is, at the time of action, in a position to exercise responsibility. This presupposition imports an ontological distinction between capacity as disposition and capacity as occurrence.

A trained pilot has the capacity to recognize an unreliable airspeed indication in the sense that, under favorable conditions, she will reliably do so. Whether she has that capacity at the moment of action, that is, whether the recognition actually occurs, depends on the conditions then obtaining. The disposition is durable. Its occurrence is fragile.

This distinction is implicit in MHC's formulation and explicit in much of the literature on expertise, but its architectural implications have not been fully drawn. If tracing requires occurrent capacity, then the conditions for occurrence become governance objects in their own right. They are not background features that may or may not be favorable. They are the architectural layer at which tracing is preserved or withdrawn [12-18].

I will use judgment here as shorthand for the occurrent exercise of the cognitive and moral capacities that tracing requires. The term carries a long philosophical history, but it is used in this article in a deliberately narrow sense, calibrated to the MHC context. Judgment in this sense is not a faculty distinct from the capacities certified by training;

it is the occurrence of those capacities at the time of action [19-21].

Decision-Time Accessibility

Decision-time accessibility (DTA) names the likelihood, given the institutional, informational, and temporal architecture of a decision environment, that the cognitive and moral capacities required for MHC tracing will reach occurrence at the moment of action.

DTA is a relational property of the agent–environment pair, not a personal trait. An agent of unchanged training and disposition may have high DTA in one architecture and low DTA in another. The variable identifies the layer at which architectural choices govern the realization of normative requirements.

DTA is not a performance metric. It does not predict whether the agent's decision will be correct; it describes whether the capacities required for the decision to constitute responsible judgment are available to be exercised. An architecture with high DTA may still produce mistaken decisions. An architecture with low DTA may, by chance, produce correct outcomes. The framework's evaluative claim is that meaningful control requires the former, and that the latter satisfies MHC only by accident.

The Six Conditions

These six conditions are not selected for their evocative power. They are derived from what MHC tracing already presupposes. An agent in a position to be responsible for a technologically mediated action must be able to do six things, each of which the framework labels with the condition that names its accessibility.

MHC tracing requires the agent to…	DTA condition
make reliable contact with the reality the action operates on	Truth
register affected persons or interests as morally weighty rather than as case-objects	Relation
inhabit the reasoning path by which the action is reached	Integration
intervene authorially, not merely assent	Doing
situate the action within the institutional purpose it serves	Orientation
sustain the foregoing jointly under actual decision load	Attention & Agency

Table A: Derivation of the six DTA Conditions from the Tracing Requirement

The conditions are not independent virtues. They are the architectural accessibility requirements that any substantive reading of MHC's tracing condition already entails. Further specification constitutes elaboration of one of the six rather than the introduction of a seventh. Each condition has a distinct collapse mode that the others do not substitute for. The remainder of this subsection specifies each.

Truth (T): Adequacy of Mediation. Truth does not denote unmediated contact with reality. All operator contact with system state is mediated by interfaces, sensors, models, and procedures. The condition names the adequacy of that mediation: whether the representations available to the agent are sufficient for the action's reality to register as it is, or whether they smooth over contingencies, suppress uncertainties, or substitute statistical plausibility for the resistance of the referent [22,23,24]. In algorithmically governed environments, ranked recommendations, risk ratings, and model outputs commonly present synthetic plausibility as evidence. Truth is preserved where mediation makes the referent's resistances reachable; it collapses where mediation closes that channel.

Relation (R): Moral Proximity. Relation names the experiential availability of those affected by a decision, such as patients, passengers, claimants, and citizens, as presences who obligate rather than as cases to be processed [25-28]. Relation does not always require direct emotional encounter with affected persons. In high-abstraction systems, it requires that the affected person, public, patient, passenger, claimant, or protected interest remain institutionally and morally represented at decision-time rather than being reduced entirely to a metric, object, or throughput unit. The condition is not about the operator's general moral commitment, which may be entirely intact, but about whether the architecture permits the affected to register at decision-time as the moral weight they carry.

Integration (I): Inhabitable Reasoning. Integration names the coherence between belief, reasoning, action, and consequence such that the agent can reconstruct, justify, and inhabit the reasoning path of her own conclusions [18-29]. It is the refusal of the black box. Where outputs are produced by opaque systems and the agent's role is to verify rather than to traverse the reasoning, Integration collapses: the action may be correct but is not authored. Integration is distinct from Truth (which concerns access to the referent) and from Doing (which concerns the capacity to intervene); it concerns the reasoning by which the action is reached. It is also distinct from decision-time co-presence (see §3.4): Integration is a logical and procedural property of the reasoning path, whereas co-presence is the spatiotemporal property of the simultaneous accessibility of all six conditions.

Doing (D): Authored Intervention. Doing names the operative capacity to intervene in the system, not merely to authorize its outputs [19-30]. It distinguishes embodied authorship from delegated assent. Real intervention pathways require temporal margin, procedural flexibility, and institutional tolerance for discretionary action. Where override is reframed as risky deviation and conformity is rewarded, Doing collapses into supervision regardless of formal authority. Orientation (O): Purposive Framing. Orientation names the agent's capacity to act toward something, situating the action within a horizon of institutional purpose that exceeds local optimization [31,32]. Automated systems are exquisitely capable of optimizing means; they are indifferent to ends. In accelerated environments, institutional success comes to be operationalized through throughput, accuracy, or compliance metrics that displace the institution's constitutive purposes. Orientation is preserved where the purpose the action serves remains intelligible to the agent at decision-time; it collapses where metric optimization has displaced purpose as the operative criterion.

Attention & Agency (AA). Attention & Agency combines two facets of a single condition: sustained attention to what the decision concerns, and the operative sense of being the one deciding rather than the one being moved through a decision pipeline. The two are treated together because, under decision-time load, they are degraded through the same architectural channels: pacing, alerting, defaults, interruptions, dashboard proliferation, and system-driven temporal control. A person whose attention is captured by the system's tempo may retain formal agency while losing experienced authorship.

AA exercises the highest architectural leverage among the six conditions. It is the condition most directly governed by the decision environment's pacing, alerting, and default structures, and the condition whose preservation most reliably propagates through the others. Without AA, Truth cannot be encountered (verification requires sustained focus), Relation cannot be perceived (moral proximity requires dwelling), Integration cannot be performed (reasoning requires traversal time), Doing cannot be owned (authored intervention requires holding considerations together), and Orientation cannot be maintained (purpose requires contemplative recovery). These dependencies place AA in a different relation to the others than they bear to one another: it has enabling priority, since the remaining five cannot be exercised without it, even though it is not normatively privileged. The distinction dissolves an apparent contradiction. Necessity for tracing is joint and symmetric, because the collapse of any one condition voids tracing as surely as the collapse of any other, and in that normative sense the six are co-equal. Enabling priority is asymmetric, because AA conditions the occurrence of the rest, and in that functional sense it comes first. Holding the two senses apart is what licenses calling AA the highest-leverage condition while still treating the six as co-equal in necessity, and it is why architectural intervention on AA most efficiently preserves or withdraws DTA.

Decision-Time Co-presence

Tracing requires the joint accessibility of all six conditions at the moment of action. I refer to this property as decision-time co-presence, distinguishing it from Integration (the third condition).

Integration is a logical and procedural property: can the operator traverse the reasoning by which the action is reached, reconstruct its steps, and inhabit it as her own? Co-presence is a spatiotemporal property: can the operator assemble Truth, Relation, Integration, Doing, Orientation, and Attention & Agency simultaneously within the time window the decision allows?

The framework's central claim is that co-presence is not a stable developmental achievement but a fragile, recurrent event. It must be reconstituted at each decision point under whatever conditions then obtain. An agent may possess all six conditions in disposition while lacking access to one or more in occurrence.

When co-presence fails, judgment does not erode linearly; it ceases to function as a governance resource. The agent may continue to act, since decisions are still produced, signatures still executed, and authorizations still recorded, yet the action no longer constitutes the occurrent exercise of the capacities tracing requires. I use the language of collapse rather than disappearance: judgment in the relevant sense becomes fragmented, displaced, or attenuated rather than absent in some metaphysical sense. The action persists; tracing does not.

The Judgment Cliff

Decision-time accessibility does not degrade smoothly. As system tempo, informational saturation, and procedural compression increase, the relationship between architectural load and co-presence exhibits a threshold structure: a region of robust co-presence followed by a sharp transition to structural non-co-presence. I call this threshold the Judgment Cliff.

A Judgment Cliff occurs when the minimum time required to assemble the conditions of responsible agency exceeds the decision window the architecture provides. The six conditions do not enter this calculation symmetrically, and their distinct roles are worth making explicit. Truth, Integration, and Doing carry genuine assembly times, since each names a process that consumes part of the decision window: making adequate contact with the referent, traversing the reasoning path, and executing an authored intervention. Let τ_T, τ_I, and τ_D denote these times. They are not summed; they combine in parallel where independent and in serial where dependent, since Doing presupposes some prior Truth and Integration. Let $\tau_{critical}$ denote the longest path through this dependency graph. Relation and

Orientation are not, in the same sense, timed sub-tasks. Relation functions as a standing gating condition: its erosion does not consume the window, but its collapse voids tracing however much time remains. Orientation underwrites the others by fixing the purpose against which the assembled judgment is measured, rather than competing with them for the window. Attention & Agency, denoted $AA \in [0, 1]$, is the scaling factor governing how efficiently the timed conditions can be assembled at all; as AA falls, the effective time each timed condition requires expands. The model, then, times three conditions, gates on two, and scales by one. Then

$$T_{required}(AA) = \tau_{critical} / AA$$

The Judgment Cliff is the inequality

$$T_{available} < T_{required}(AA)$$

The cliff is not metaphor. It is a relation between two specifiable quantities. The architecture imposes a tempo $T_{available}$ (the time between decision-relevant signals and the deadline for response). The decision requires a tempo $T_{required}$ (the time needed to assemble Truth, Integration, and Doing, under preserved Relation and Orientation and sustained Attention & Agency, and then to execute an authored intervention). Where the architecture's tempo falls below the decision's required tempo, co-presence is structurally infeasible.

The discontinuity is in the outcome, not the underlying function. $T_{required}$ is smooth and hyperbolic in AA; as AA collapses toward zero, $T_{required}$ diverges. But the outcome, co-presence vs. non-co-presence, is binary: the decision either assembles within $T_{available}$ or it does not. The asymmetry between smooth degradation of inputs and binary collapse of the output is the formal content of the framework's claim that judgment does not erode linearly but ceases, once a threshold is crossed, to function as a governance resource. Two cautions bound the strength of this claim. First, on the model as stated the threshold is the crossing point of a smooth, monotone function rather than a dynamical singularity; the word cliff names the consequence for tracing, not a discontinuity in the underlying variable. Second, a genuinely bistable structure, in which two stable equilibria are separated by an unstable region so that small load increases near the threshold produce disproportionate and hysteretic collapse, would follow only if the conditions interact through specified reinforcing couplings, for instance if Integration accelerates the acquisition of Truth while sustained AA stabilizes Orientation strongly enough to generate positive feedback. That fold-catastrophe structure is a conjecture this article does not establish; it would require the coupling dynamics to be specified and estimated, and it is offered as a hypothesis for the critical-transitions literature to test rather than as a result [33,34]. The weaker threshold claim, on which the rest of the argument depends, does not require it.

The location of the cliff is sector-specific and varies with the moral weight and complexity of the decision: routine decisions have higher cliffs, while high-consequence, morally weighted decisions have lower ones. The Judgment Cliff is the institutional analogue of cognitive thresholds documented in human factors research [2-35], but it operates at the temporal scale of institutional action, minutes, hours, and days, rather than milliseconds. Two regimes should be distinguished, because the model applies to them differently. In acute compression, such as the 737 MAX cockpit, $T_{available}$ is a literal reaction-time window and the inequality is meant straightforwardly. In chronic erosion, such as the Challenger teleconference or the months-long rating chain of 2008, the operative failure is less a reaction-time deadline than a standing disjunction of the conditions from the locus of authority; there $T_{available}$ denotes the institutional interval within which the conditions could have co-assembled, and the inequality should be read as an analogy whose force comes from the same joint-accessibility requirement rather than from a stopwatch. The quantitative form is sharpest in the acute regime; in the chronic regime the work is done by the diagnostic profile of which conditions were accessible to whom, as the case analyses in Section 4 show.

Three zones follow. In the deliberative zone, $T_{available} \geq T_{required}(AA)$ and system tempo remains compatible with co-presence. At the cliff, $T_{available}$ approaches $T_{required}(AA)$ and the relationship between load and convergence becomes discontinuous: action persists but shifts from authored judgment to pattern-completion and system-confirming reaction. In the compliance zone beyond the cliff, $T_{available} < T_{required}(AA)$ and operators remain nominally responsible but are functionally absent from the decisions to which they are formally traced. From outside the institution, this state is often indistinguishable from preserved oversight, because the formal markers remain in place; the collapse becomes visible only when something goes wrong.

Per-condition assembly times are, in principle, estimable from cognitive task analysis in the relevant domain. The framework is not currently calibrated for any specific domain; calibration is the empirical work the framework makes possible.

A worked illustration, with deliberately hypothetical values, shows how the inequality is meant to be applied and why calibration matters. Consider the 737 MAX case of Section 4.2. Between successive MCAS activations the architecture afforded an interval on the order of $T_{available}$ of five to ten seconds. An authored recovery required recognizing an undisclosed failure mode against conflicting cues, retrieving the reasoning trail of an opaque system, and executing a non-routine trim procedure, a critical path that cognitive task analysis of comparable upset-recovery sequences would

plausibly place on the order of τ_critical of twenty to forty seconds before any degradation of Attention & Agency is counted. With AA depressed under repeated startle and alert saturation to, say, 0.4, T_required = τ_critical / AA rises to roughly fifty to one hundred seconds, an order of magnitude beyond the window the architecture allowed; the action falls deep in the compliance zone, which is what the framework means in saying the cliff was crossed within seconds. The same arithmetic for a structured clinical handover or a Crew Resource Management challenge, where T_available runs to tens of seconds or minutes and AA is protected by design, places the action in the deliberative zone. These numbers are stipulated, not measured; their only purpose is to make the inequality concrete and to show that its terms are the kind of quantities cognitive task analysis can estimate. Supplying measured values for a defined decision class is the empirical work the framework invites, and a framework that produced no such numbers, or whose numbers never tracked observed collapse, would be failing on its own terms.

Diagnostic Indicators

To render the framework operational, each of the six conditions can be specified across three states corresponding to degrees of accessibility preservation. The states are not points on a continuum but qualitative regions. The boundary between eroded and collapsed marks the local threshold beyond which the condition no longer functions as an enabling condition for tracing.

Table 2 provides indicators in governance-specific form, with the qualitative state description and illustrative measurable proxies for each condition–state combination.

Condition	Intact	Eroded	Collapsed
Truth (T)	Direct, unaggregated feedback with explicit margin-of-error ranges. Proxies: uncertainty-display coverage; independent verification channels; ratio of ground-truth to model-output checks.	Aggregated indices; delayed data streams; hidden telemetry variances. Proxies: latency between event and display; suppressed-uncertainty rate.	Purely synthetic outputs; model outputs as the sole proxy for reality. Proxies: zero independent verification; uncertainty hidden by design.
Relation (R)	Immediate, human-scale feedback connecting action to affected party. Proxies: narrative case representation; minimal abstraction distance.	Highly abstracted representations ("units", "risk bands"); affected parties present only in summary form. Proxies: distance metrics between operator and affected; absence of narrative.	Human impact externalized or invisible at decision time. Proxies: zero direct exposure to affected; cases present only as statistics.
Integration (I)	Clear, step-by-step logic trails the operator can traverse. Proxies: reasoning-path reconstructability tests; documented traversal time.	Explainable-AI summaries that require deliberate manual extraction. Proxies: explanation-on-demand only; partial logic traces.	Opaque black box; binary validation only. Proxies: zero reconstructability; verify-only interface.
Doing (D)	Physical, unpenalized override with adequate temporal margin. Proxies: override frequency; symmetric liability for override and deference.	Bureaucratic friction; override triggers mandatory retrospective review. Proxies: asymmetric override penalties; required authorization chains.	Discretionary path locked; override requires multi-party keys beyond the decision window. Proxies: zero unilateral override; deadline-incompatible authorization.
Orientation (O)	System incentives directly match the institutional mission. Proxies: metric–mission alignment audit; presence of telos in performance review.	Token mission compliance paired with intensive throughput/volume tracking. Proxies: metric–mission drift; mission language in policy but throughput in incentives.	Proxy metrics completely supplant original mission. Proxies: zero mission reference in incentives; throughput is sole performance criterion.
Attention & Agency (AA)	Mindful, continuous engagement; operator dictates the system's temporal cycle. Proxies: alert frequency within tolerance; deliberative-interval availability; low task-switching rate.	Frequent notifications; alert fatigue; reactive task-switching. Proxies: alert frequency exceeds threshold; deliberative intervals frequently interrupted.	Cognitive saturation; system-driven pacing producing reactive paralysis. Proxies: alert frequency saturates attention budget; deliberative intervals structurally absent.

Table 2: Diagnostic indicators for the six conditions.

The proxies are illustrative rather than exhaustive. For any given domain, operationalization will require domain-specific instrumentation. The architecture of the table, a qualitative state paired with measurable proxies across three states per condition, is the methodological contribution.

A Six-Step Operational Protocol

DTA is domain-adaptive rather than metric-universal. The six conditions apply across domains; their indicators differ between aviation, clinical medicine, finance, military systems, and algorithmic governance. The following six-step protocol structures DTA analysis in any domain.

- **Identify the decision.** What action is being traced under MHC?
- **Locate the responsible agent or locus.** Who is the formally responsible party, and at what level of the institution?
- **Specify the decision window.** What is T_available between the decision-relevant signal and the deadline for action?
- **Specify the conditions for responsible judgment.** What must be accessible for this decision class, in the sense of adequate Truth, present Relation, traversable Integration, available Doing, intact Orientation, and sustained AA?
- **Assess each condition.** Score each as intact, eroded, or collapsed using domain-specific indicators on the Table 2 schema.
- **Locate the action relative to the Judgment Cliff.** Did T_required for the assessed DTA profile fall within T_available, or did the action occur beyond the cliff?

The protocol is the methodological counterpart to the diagnostic schema of Table 2 and supports both retrospective audit and prospective architectural design.

Cases: Collapse and Preservation of DTA

If joint accessibility of the six conditions is the architectural condition for MHC tracing, the central operational question is: through what mechanisms do contemporary decision environments degrade accessibility? Four mechanisms recur: attentional capture, responsibility diffusion, temporal compression, and automation bias. This section develops them through three structurally varied retrospective cases (Challenger, the Boeing 737 MAX, the 2008 rating-model chain) and a counter-case (Crew Resource Management), followed by a cross-case diagnostic matrix.

The cases vary by sector, temporality, and institutional scale while sharing a common oversight structure: trained human agents formally responsible within technologically mediated systems. They are theory-elaborating plausibility probes, not a representative sample. The prospective application in §5 is the stronger test of the framework. The cases are not selected to prove that every sociotechnical failure is a DTA failure; they are selected because each contains a publicly reconstructed decision architecture in which the relation between formal responsibility and accessible judgment can be examined.

Challenger: Severance of Authority from Concern

The Challenger launch decision has been extensively reconstructed [36,37]. The framework specifies what was severed where. Truth was largely intact among the engineers, who had reliable contact with O-ring vulnerability under cold temperatures and communicated it to managers. At the managerial level, Truth was eroded: engineering data was filtered through schedule pressure and recoded as acceptable variance. Integration was intact among the engineers and partially intact among the managers who reviewed the data. Relation was thinned throughout: astronauts, public, and downstream stakeholders were not phenomenologically present in the decision room as obligating others. They appeared as program participants and schedule beneficiaries.

The crucial collapse was in the spatial disjunction between where the conditions were accessible and where authority resided. Engineers retained Relation and Truth but were severed from Doing: they could articulate the risk but could not halt the launch. Managers held the authority required for Doing but operated under an Orientation collapse in which NASA's constitutive purposes (safe spaceflight, scientific advancement, public trust) were eclipsed by operative metrics: schedule adherence, program continuity, and congressional optics. The decision was locally rational against the metrics and globally aimless against the mission.

Attention & Agency was strained but not fully collapsed; the decision unfolded over hours, not seconds. But the attentional capacity of those with authority was bent toward schedule metrics rather than risk integration.

The mechanism most active in Challenger was responsibility diffusion: the distribution of authority across actors, procedures, and processes such that no participant encountered the outcome as something for which she was existentially responsible [27-38,39]. Severing Relation and Doing from Truth and Integration was the architectural form of this mechanism. Managerial-level attentional capture by schedule metrics eroded AA for the actors with decision authority. On this reading, Challenger is a governance failure produced not by ethical or competence deficit but by the spatial disjunction of accessible conditions from authority, combined with Orientation collapse among those who held the authority.

Boeing 737 MAX: Truth Foreclosed by Design

The 737 MAX crashes are often attributed to automation malfunction or inadequate training [40,41]. The framework specifies a deeper structural problem: the decision environment foreclosed tracing even in the presence of professional competence.

Truth collapsed at the operator level not under load but by design. The MCAS system was not disclosed in pilot training; its existence as a control authority was not part of the pilots' working model of the aircraft. MCAS interventions presented through fragmented instrument anomalies rather than as system-level events. Pilots had to infer system-level causality from partial cockpit symptoms such as the stick shaker, runaway trim, and conflicting airspeed indications, all under

rapidly escalating conditions. The formal availability of manual trim or checklist response did not guarantee accessibility, because effective Doing required prior interpretive recognition of an opaque automation pathway.

Integration was structurally constrained. Even pilots who recognized the anomaly could not reconstruct the system's reasoning path because the system's reasoning was not exposed.

Attention & Agency was overwhelmed by temporal compression: MCAS interventions occurred rapidly, repeatedly, and without transparent disclosure, compressing the decision window below the threshold required for co-presence. The Judgment Cliff was crossed within seconds. T_available, the time between successive MCAS activations and the deadline for response, fell below the T_required for pilots to assemble Truth, locate the reasoning trail of Integration, and execute an authored intervention.

Doing was nominally available, since pilots retained physical control inputs, but effective intervention required recognizing an opaque failure mode, diagnosing it correctly against incorrect system feedback, and executing exceptional recovery procedures under extreme time pressure. The intervention pathway existed in principle but was structurally inaccessible in practice. Some 737 MAX crews did recover from MCAS events; the framework predicts that recovery was possible where the recognition occurred quickly enough to leave deliberative margin, and infeasible where it did not.

Orientation collapse occurred at the architectural rather than the operator level. The Boeing certification process privileged commercial purposes (minimizing pilot retraining costs, accelerating certification, defending market share) over the institutional purpose of aviation safety. Operator-level failures were downstream of an Orientation collapse located upstream in the certification architecture.

The active mechanisms at the operator level were temporal compression and automation bias: the rational adaptation of operators to systems that operate faster than human deliberation and carry institutional authority [30-35. At the corporate-architectural level, the mechanisms were Orientation eclipse and responsibility diffusion across the certification chain.

The 2008 Financial Crisis: Rating Models and the Collapse of Inhabitable Risk Judgment

Rather than attempt to analyze the 2008 crisis as a whole, a project beyond any single framework, this section examines one structurally diagnostic chain: mortgage origination → securitization → rating-model output → AAA designation → institutional reliance → risk blindness [42,43]. At each link, distinct DTA conditions collapsed.

At origination, Relation collapsed: borrowers were absorbed into standardized application packets feeding throughput targets and disappeared as obligating others. At securitization, Integration collapsed: the reasoning path from individual loan to packaged security was traversable in principle but inhabited by no analyst. At rating, Truth collapsed in a particular form: synthetic plausibility was treated as evidence of substance, and the model's AAA output became the operative reality against which institutional decisions were calibrated. At institutional reliance, Doing was eroded by incentive asymmetry, since refusal to participate in AAA-rated instruments imposed a competitive cost while participation imposed none until the chain failed. Orientation collapsed across the chain into market metrics, namely deal volume, market share, and returns, which displaced financial intermediation's constitutive purpose. AA fragmented under the pace of issuance and the proliferation of dashboards.

The chain demonstrates the framework's central claim at one site rather than at the whole crisis: no locus in the chain retained co-presence of the six conditions; the rating output substituted for the judgment the chain was supposed to assemble. The crisis's broader features, including funding-market dynamics, regulatory capture, and monetary policy, are real but beyond the framework's scope. The DTA contribution is the diagnosis of how, in the one chain examined, the conditions for responsible judgment had been architecturally withdrawn well before the crisis revealed them as absent.

Counter-Case: Crew Resource Management

Not all technological mediation degrades accessibility. The framework's evaluative claim is that some architectures preserve co-presence while others consume it; the question is design, not technology in general. Crew Resource Management (CRM), developed in commercial aviation following the late-1970s investigation of accidents in which technically sound crews failed to coordinate under stress, illustrates a DTA-preserving architecture [44]. CRM protocols explicitly preserve each of the six conditions.

Standardized callouts and cross-checking protect Truth by requiring verification of system state independent of any single channel. Crew briefings and explicit role assignments preserve Relation by maintaining the social structure of authority and accountability. The two-challenge rule, in which any crew member must be heard if a concern is raised twice, protects Integration and Doing simultaneously: reasoning must be made explicit, and intervention pathways are protected against deference cascades. Sterile-cockpit rules during critical phases of flight protect AA by limiting non-essential communication during periods when attentional sovereignty is most fragile.

CRM does not eliminate automation; it operates alongside extensive cockpit automation. Its contribution is architectural: it inserts deliberative structure into a high-tempo automated environment in ways that preserve co-presence at the moment of action. Similar architectural patterns appear in surgical timeouts, ATC handoff protocols, and structured clinical handovers. The framework predicts that such architectures should raise the Judgment Cliff for their domains, and the empirical record of dramatic safety improvements in commercial aviation since the 1980s is consistent with that prediction without proving it.

Cross-Case Diagnostic Matrix

Condition	Challenger	Boeing 737 MAX	2008 rating chain	CRM (counter-case)
Truth	Intact (engineers); eroded (managers)	Collapsed by design	Collapsed through model substitution	Preserved by cross-checks
Relation	Eroded	Indirect	Collapsed through abstraction	Preserved by role structure
Integration	Partial	Collapsed	Collapsed across chain	Preserved by explicit challenge
Doing	Collapsed for those with knowledge	Nominal but inaccessible	Eroded by incentives	Preserved by intervention norms
Orientation	Collapsed at managerial level	Collapsed upstream in certification	Collapsed into market metrics	Preserved by safety telos
Attention & Agency	Strained	Collapsed under temporal compression	Fragmented across the chain	Preserved by sterile-cockpit discipline
Primary collapse	Authority severed from concern	Opaque automation; compressed window	Distributed absence of responsible locus	(none) DTA-preserving architecture

Table B: DTA profile across the four cases.

Three features of the matrix bear on the framework's general claims. First, no two cases share an identical collapse profile, which weighs against the objection that DTA is a single-cause theory dressed in six labels. Second, every failure case shows collapse of at least three conditions and AA strain or collapse, consistent with the joint-necessity claim. Third, the counter-case shows that the same architectural variables, deliberately configured, preserve rather than consume DTA. The framework predicts collapse patterns, not single-condition failures.

Across the failure cases, three further features recur. Each exhibits collapse of at least three of the six conditions at the moment of action. Each involves both condition collapse at the operator level and Orientation eclipse at the architectural level; operator failure is downstream of architectural failure. And each shows AA strain or collapse as part of the failure structure. The framework does not predict that all failures share an identical collapse pattern; it predicts that catastrophic failures involve a configuration of inaccessibility that prevents tracing.

Prospective Test: Clinical AI under Article 14

The retrospective cases establish plausibility. The prospective application is the stronger test. This section applies the framework to one currently emerging configuration: algorithmic clinical decision support systems (CDSS) operating under the EU AI Act's Article 14 human-oversight requirements. It then states a set of prospective predictions about the implementation period, each paired with explicit disconfirmation criteria.

The Article 14 Application

Article 14 of the EU AI Act requires that high-risk AI systems be designed and developed in such a way that they can be effectively overseen by natural persons during the period in which the AI system is in use. It enumerates measures the provider must enable: the human must be able to understand the system's capacities and limitations, monitor its operation, remain aware of automation bias, correctly interpret outputs, decide not to use the system or to override it, and intervene or interrupt operations [45].

Applied to a typical CDSS deployment, such as a triage support system that recommends acuity levels in an emergency department or a diagnostic support system that ranks differential diagnoses, the framework predicts the following accessibility profile. The claim is not that all CDSS deployments collapse DTA, but that deployments placing system outputs inside time-pressured clinical decision windows are especially vulnerable unless Article 14 oversight is architecturally specified.

Truth is at risk of erosion. CDSS outputs typically present as confidence-ranked recommendations; uncertainty is often hidden behind point estimates or rendered through opaque probabilistic outputs. The intact state requires explicit uncertainty display, independent verification channels (the clinician's own examination, not only system inputs), and a ratio of clinical assessment to model output that preserves the clinician's contact with the patient as referent rather than as data subject. Article 14 requires that the clinician "correctly interpret the output"; the framework predicts this is

impossible without operational specification of uncertainty display, which Article 14 does not provide.

Relation is at high risk of collapse. Where the patient is encountered primarily as a record on a screen and the CDSS output is the primary representation of the case, Relation collapses. The intact state requires that the patient register at decision-time as the obligating other, not as the case-object. Article 14 contains no provision for this condition. Workflow designs that route the clinician's attention through the CDSS interface rather than through the patient's body and account are particularly at risk.

Integration is at risk of collapse. The black-box character of most CDSS systems produces precisely the verify-only interface that collapses Integration. Article 14 requires that the human "understand the capacities and limitations" of the system; if the system's reasoning is not exposed, this is unsatisfiable. The framework predicts that Article 14 compliance will be attestational rather than substantive for CDSS systems whose reasoning is not reconstructable by the clinician.
Doing is at risk of erosion. The clinician retains formal override authority. Whether override is practically accessible depends on whether the architecture imposes asymmetric liability, penalizing an override that proves wrong more severely than a deference that proves wrong. Article 14 does not address asymmetric liability. The framework predicts that Doing will collapse de facto where institutional liability structures penalize override.

Orientation is at systemic risk. CDSS deployment is frequently justified through throughput metrics such as reduced wait times, increased capacity, and improved coding accuracy, none of which is the institutional purpose of clinical medicine. Where these metrics displace clinical purpose as the operative criterion of CDSS deployment, Orientation collapses at the architectural level.

Attention & Agency is at high risk of collapse. The pacing of emergency department triage, in particular, is typically set by the rate of patient arrival, not by the clinician. The CDSS introduces an additional alerting system; the cumulative alert frequency may exceed the clinician's attention budget. Article 14 requires that the human "monitor the operation" of the system; this is unsatisfiable where attention is saturated.

The prediction is that CDSS systems deployed under Article 14 without architectural attention to the six conditions will produce fictive accountability: the clinician will be formally responsible while substantively unable to exercise the capacities Article 14 presupposes.

Prospective Predictions for the Article 14 Implementation Period (2026–2030)
The retrospective cases of §4 establish the framework's plausibility but cannot test it. Cases analyzed after their outcomes are known are vulnerable to coding bias; the analyst can specify conditions whose collapse predicts the failure already observed. The stronger test is forward-facing. The following predictions are stated prospectively as of the article's date of publication. They concern the period of EU AI Act Article 14 implementation across high-risk CDSS deployments, 2026–2030. Each prediction is paired with a disconfirmation criterion.

Prediction 1 (Incident framing). Of CDSS-related incidents that produce regulatory or judicial review during 2026–2030, more than 70% will be analyzed in their primary published findings through operator-level variables (training adequacy, procedural adherence, communication failures, individual judgment) rather than through architectural variables (uncertainty display, override-liability asymmetry, alert-frequency saturation, reasoning-path opacity). Disconfirmation: fewer than 50% of incident analyses framed in operator-level terms.

Prediction 2 (Remedy pattern). Of remedies imposed by regulators or hospitals following such incidents, more than 70% will fall under training, procedure, documentation, or attestation requirements; fewer than 20% will modify architectural conditions identifiable with one of the six DTA conditions (interface uncertainty display, override-liability rebalancing, alert-budget caps, reasoning reconstructability requirements). Disconfirmation: architectural remedies exceed 30% of the imposed remedy mix.

Prediction 3 (Liability allocation). In civil and disciplinary proceedings arising from CDSS-mediated misdiagnosis or mistriage during the period, primary liability will be allocated to the clinician of record in more than 80% of cases that reach judgment, irrespective of audit-level evidence that the deployment foreclosed one or more of the six conditions. Disconfirmation: primary architectural liability allocation in more than 30% of judgments.

Prediction 4 (Audit gap). No more than 15% of high-risk CDSS deployments certified as Article 14–compliant during the period will, on independent architectural audit, exhibit intact status across all six conditions. The most frequent collapsed condition will be Integration (reasoning-path opacity); the most frequent eroded condition will be Attention & Agency (alert-budget saturation). Disconfirmation: either more than 30% of deployments show all-intact status, or the most frequent collapsed condition is not Integration.

Prediction 5 (Discourse pattern). The published regulatory commentary on Article 14 implementation will continue to frame human oversight as a question of operator capability, training, and procedural adherence rather than as a question of architectural design. Frequency analysis of EU AI Office and member-state competent authority publications during

the period will show architectural framings in fewer than 25% of substantive references to Article 14. Disconfirmation: architectural framings exceed 40%.

Prediction 6 (Dose-response). Predictions 1 through 3 and 5 concern blame and discourse, and would be confirmed even if the architectural mechanism were false but institutions blamed operators for unrelated reasons; they test the fictive-accountability claim rather than the causal core. Prediction 6 tests the core. Where independent architectural audits of CDSS deployments are available, the incidence of tracing-relevant adverse events will rise monotonically with the number and severity of conditions assessed as eroded or collapsed at the time of the event, after adjustment for case mix and operator experience; a deployment with two or more collapsed conditions will show a materially higher tracing-relevant incident rate than an otherwise comparable deployment with all six intact. Disconfirmation: no monotonic association between audited condition collapse and tracing-relevant incidents, or an association that disappears once case mix and experience are controlled. This prediction depends on the architectural audits that Prediction 4 also requires, and those instruments do not yet exist at scale; the framework itself proposes them in Section 6.3. Until they are deployed, Prediction 6 is conditionally rather than immediately testable, and the framework's strongest empirical claim is hostage to the construction of the very instruments that would test it. The dependence is stated rather than concealed.

Joint disconfirmation. Sustained failure of three or more of these predictions over the observation period would constitute serious disconfirmation of the framework, indicating either that the institutional pathology it diagnoses is less robust than claimed or that it misidentifies the mechanisms through which fictive accountability is maintained. Failure of the dose-response prediction (Prediction 6) would weigh more heavily than failure of the discourse predictions, since it alone tests the causal core rather than the pattern of blame; confirmation of the discourse predictions alongside failure of Prediction 6 would indicate that fictive accountability is real while its architectural explanation is mistaken.

Predictions are testable against (i) published incident reports from national health regulators and patient safety organizations, (ii) EU AI Office implementation publications and reports, (iii) court records in member-state jurisdictions, (iv) independent architectural audits conducted under emerging Article 14 conformity assessment regimes, and (v) systematic content analysis of regulatory publications. Prospective specification is not a claim to scientific status but a discipline against motivated coding: the framework is philosophical in its concepts and empirical in its predictions, and the explicit statement of disconfirmation criteria is what makes the second possible without contaminating the first.

Governance Implications: From Oversight to Architecture

The framework's governance implications are not a derivative of the analysis but its operational stake. If DTA names the architectural condition under which MHC tracing is satisfied, governance is the apparatus by which that condition is preserved or withdrawn. This section develops the implication along five lines: the six architectural commitments that follow directly from the framework, a design methodology that operationalizes them across the system lifecycle, architectural audit and liability allocation as institutional instruments, mapping onto existing regulatory regimes, and the honest position about domains where DTA is structurally unavailable. A closing subsection treats condition ethics, the normative orientation toward responsibility that DTA entails, as the framework's principled consequence rather than as a competing theory.

Six Architectural Commitments

Each condition implies a design directive. The directives are governance variables to be configured for the decision class, not universal rules. Protect Attention & Agency. Limit alert saturation and dashboard proliferation. Build in mandatory deliberative intervals for decisions above stipulated consequence thresholds. The architectural leverage of AA, which propagates through the other five conditions, makes this the highest-priority commitment in any DTA-preserving design. Preserve Truth. Distinguish representation from referent in interface design. Display model uncertainty alongside outputs. Provide independent verification channels for high-consequence judgments. Resist the tendency of optimization systems to present synthetic plausibility as evidence of substance.

Sustain Relation. Maintain institutional arrangements that keep affected parties present at decision-time above pure abstraction. The condition is fragile and easily designed out; its preservation requires deliberate counter-pressure against throughput optimization. Relation does not require direct emotional encounter; it requires that the patient, passenger, claimant, or protected interest remain institutionally and morally represented at decision-time rather than collapsed into a throughput unit. Defend Integration. Require reconstructable reasoning paths for high-consequence automated outputs. Resist black-box deployment in domains where tracing is normatively required. Build explanation requirements into procurement and certification with substantive rather than attestational standards.

Secure Doing. Establish protected intervention pathways with adequate temporal margin. Modify liability structures so override is not penalized asymmetrically relative to deference. This commitment, more than any other, requires modification of institutional culture and liability law, not only interface design.

Anchor Orientation. Track institutional purpose alongside throughput. Make explicit the purposes the institution claims to serve and audit performance against them. Orientation collapse is the architectural failure most easily concealed by metric success; its detection requires audit instruments calibrated to mission rather than to throughput.

DTA-by-Design
The six commitments become a methodology when mapped to the development lifecycle, on the model of Privacy-by-Design [46]. and Value-Sensitive Design [47]. At the requirements stage, the design team identifies which decisions mediated by the system invoke MHC tracing, the prior question of whether oversight is required at all. For those decisions, the team estimates τ_critical through cognitive task analysis and specifies T_available as a binding design constraint: the system's tempo must afford co-presence within the relevant decision window. Where it cannot, the trade-off must be surfaced explicitly rather than tacitly resolved through deployment.

At the architecture stage, each of the six conditions is configured as a design variable: Truth via uncertainty schema and verification channels; Relation via workflow placement of affected-party representation; Integration via the reasoning-reconstructability standard the system must meet; Doing via override pathway and its liability structure; Orientation via mission-aligned metrics tracked alongside throughput; AA via the alert budget and protected deliberative intervals. At the implementation stage, the diagnostic indicators of Table 2 are instrumented as system telemetry. The deployed system produces, as part of its operating record, data sufficient to score each condition's state under production load. At the pre-deployment stage, an architectural audit assesses the DTA profile against expected operating load. Any condition rated eroded triggers documented mitigation; any condition rated collapsed halts deployment for the decision class until remediated. The audit is signed by architectural decision-makers, establishing the locus of architectural responsibility ex ante rather than reconstructing it ex post.

At the operational stage, continuous DTA monitoring through the implementation-stage instrumentation. Where the production profile diverges from the pre-deployment audit, typically through AA erosion under load growth or Orientation drift under metric pressure, the divergence triggers architectural review rather than operator retraining. At the incident stage, architectural review precedes operator review. The standing question is whether DTA was preserved at the moment of action; if not, the analysis proceeds at the architectural level, and operator conduct is examined only against the conditions DTA actually afforded.

Architectural Audit and Liability Allocation
DTA-by-Design has institutional preconditions that exceed design practice. The most consequential is the locus of liability. In many technologically mediated failures, corporate and regulatory liability structures tend to preserve formal operator responsibility even where the architecture has withdrawn the conditions that make such responsibility meaningful. The current regulatory default in most jurisdictions is that compliance with procedural requirements establishes a presumption that human oversight was substantive; when something goes wrong, the operator who acted within the architecture is the first locus of blame.

The framework's analysis suggests reversing this default for high-stakes deployments. The burden of demonstrating that the architectural conditions for occurrent capacity were preserved should fall on the deploying institution, not on the operator subsequently blamed for their absence. Three institutional moves follow. Architectural audit requirements. The diagnostic schema of Table 2 serves as the basis for systematic audit of high-risk systems against the six conditions, with intact/eroded/collapsed assessments documented at deployment and reviewed under operational load.

Liability allocation that tracks the audit. Where architectural audit identifies a collapsed condition at the moment of an incident, primary liability should accrue to the architectural decision-makers who configured the foreclosure, rather than to the operator who acted within it. This is consistent with established legal principles in adjacent domains (product liability, certain forms of constructive defense) but would require regulatory implementation to become routine. Architectural visibility for affected parties. Patients, claimants, citizens subjected to algorithmically mediated decisions should have access to information about the architectural conditions under which decisions affecting them were made. This is the institutional counterpart to any meaningful right of contestation.

Mapping to Existing Regulatory Regimes
A useful way to read the framework is as a substantive specification of what existing oversight regimes already require. Article 14 of the EU AI Act enumerates outcomes that human oversight must achieve, among them understanding capacities and limitations, awareness of automation bias, correct interpretation of outputs, and the capacity to override and intervene, yet it does not specify the architectural conditions for those outcomes. The six conditions are precisely such a specification. The Article 14 requirement that the human "understand the capacities and limitations" of the system is unsatisfiable where Truth and Integration are not preserved; the requirement to "remain aware of automation bias" is unsatisfiable where AA is saturated; the requirement to "decide not to use or to override" is unsatisfiable where Doing has been institutionally penalized through asymmetric liability.

NIST's AI Risk Management Framework [48,49]. on AI concepts and terminology, and [50]. on human-system interaction can be read similarly. Each enumerates outcomes oversight must achieve; each leaves the architectural specification of the conditions under which those outcomes can occur underdeveloped. Here, the contribution is not to displace these regimes but to convert their attestational compliance requirements into substantive ones.

Where DTA Is Structurally Unavailable
There are decision classes in which the speed or volume that makes the system valuable is incompatible with the time required for co-presence. Air defense against high-velocity threats, high-frequency trading, large-scale content moderation, and certain forms of public-health emergency response impose tempos that exceed the available window for occurrent capacity.

The honest position is not that DTA must always be preservable, but that where it is not, the resulting oversight regime is attestational rather than meaningful and must be acknowledged as such. Two implications follow. First, operator-blame allocation in such domains is structurally misallocated by the framework's lights; moral responsibility belongs further upstream, to the architectural decision to deploy a system whose tempo forecloses MHC. Second, nominal human-in-the-loop arrangements that do not preserve DTA constitute the mechanism by which fictive accountability is established. They should be resisted as compromise positions rather than adopted as such.

Condition Ethics: The Normative Implication
The governance program above presupposes a normative orientation the framework makes explicit. Condition ethics names the orientation: that substantive moral evaluation of technologically mediated action presupposes the architectural conditions under which agency could have been exercised at all. Where those conditions were absent, the operator was not failing to exercise responsibility; responsibility was structurally unavailable for exercise.

Condition ethics is not a substantive normative theory in competition with virtue ethics, deontology, consequentialism, or contractualism. It is prior to them, in the way that the question of whether an utterance was made under coercion is prior to the evaluation of its truth. Three predecessors locate it [28]. social connection model of responsibility argues that responsibility for structural injustice cannot be analyzed through individual liability and requires attention to background social conditions; condition ethics shares this critical orientation toward individualized blame while differing in being architectural rather than political-structural. distributed moral responsibility distributes moral weight across networks of human and artificial agents; condition ethics differs in being pre-evaluative rather than evaluatively distributive, since it asks whether the conditions for such distribution to be coherent obtained at all latent-condition analysis anticipates the architectural orientation but operates within a descriptive-causal register; condition ethics extends Reason's insight into a pre-evaluative ethics.

Its institutional content is fully specified by the architectural commitments of §6.1 and the audit-and-liability program of §6.3. Its philosophical content is the recognition that responsibility-bearing agency has architectural conditions, and that where those conditions are foreclosed, the moral evaluation that proceeds as if they obtained is misdescriptive.
The framework's analysis converges on a single normative claim. The architecture under which an agent is asked to act is not a neutral context within which her exercise of responsibility is evaluated; it is the condition of possibility for that exercise. Where the architecture preserves the conditions for occurrent capacity, MHC tracing is satisfiable and moral evaluation has a coherent object. Where it does not, the formal apparatus of oversight, certification, and authorization produces fictive accountability, the appearance of governance over conditions the governance regime itself made ungovernable.

Scope, Testability, and Disconfirmation Risks
The framework's claim is operational and bounded. This section makes its scope explicit, identifies the empirical and methodological discipline through which the framework can be tested, and specifies the forms of evidence that would weaken or disconfirm it.

Scope
DTA is operational where MHC's tracing condition is invoked. It is silent where the governance regime does not require human responsibility for the system's action, for instance where automation is judged categorically appropriate and no residual human authorship is claimed. The framework does not adjudicate whether MHC should be invoked for a given decision class. It adjudicates what MHC's invocation requires architecturally.

Two further scope clarifications. First, the framework concerns decision environments in which a human agent is asked to bear responsibility for an action mediated by a technological system. It does not extend straightforwardly to fully autonomous systems lacking any operator role, nor to decisions made entirely outside technological mediation. Second, the framework's resolution is the moment of action, the discrete decision point at which the system's action is taken under human authorization. Decision processes that unfold across extended periods may exhibit multiple decision points, each of which the framework can address separately; cumulative effects across many such points are an empirical generalization the framework supports but does not by itself establish.

Testability and Methodological Discipline
DTA is domain-adaptive rather than metric-universal. The six conditions apply across domains; their indicators differ between aviation, clinical medicine, finance, military systems, and algorithmic governance. The proxies suggested in Table 2 are illustrative; domain-specific operationalization requires calibration by practitioners familiar with the relevant decision class.

This domain adaptation is methodological discipline, not weakness. It is a feature of any framework that aspires to apply across sociotechnical sectors with substantively different operating tempos, signal structures, and institutional cultures. The six conditions, the three-state diagnostic schema, the joint-accessibility requirement, and the Judgment Cliff threshold are general; their measurement instruments are domain-specific. Consequently, the framework is testable only through domain-calibrated audit, not through a universal metric, a constraint shared by frameworks such as situation awareness theory, organizational accident theory, and resilience engineering, each of which requires sector-specific instrumentation to be empirically deployed.

The strongest test is prospective. The prospective predictions of §5 specify what the framework expects to observe in CDSS deployments under Article 14 over the implementation period; sustained failure of those predictions would weigh against the framework. Retrospective analyses, including those in §4, establish plausibility but cannot test the framework against its own predictions, since the analyst already knows the outcome and can code accordingly.

Disconfirmation Risks

The framework would be weakened by three forms of evidence.

First, cases of systemic governance failure despite independently documented preservation of all six DTA conditions. This would weigh against the necessity claim, namely that the six conditions are jointly required for MHC tracing. The relevant standard is independent documentation rather than the analyst's own coding: disconfirmation requires that the six conditions have been assessed as intact through audit conducted ex ante and survived under operational load that the framework expected to test them.

Second, cases in which all six conditions are preserved and MHC tracing nonetheless fails through a mechanism the framework does not capture. This would weigh against the sufficiency claim, that the six conditions are jointly enough for tracing. Here a closure principle is required, since the option of absorbing every counterexample as a fresh sub-condition would make the decomposition unfalsifiable. The principle is this: a candidate requirement counts as an elaboration of an existing condition only when it can be derived from that condition's already-stated collapse mode without enlarging it; otherwise it is a genuine addition, and a genuine addition is not a free repair. The first well-documented case in which all six conditions are independently assessed as intact and tracing still fails counts against the decomposition, and a pattern of such cases, each demanding a new condition, would disconfirm the claim that tracing decomposes into a small fixed set of accessibility requirements at all. The framework is committed to the six as jointly sufficient and treats the need to add a seventh as a cost to be paid openly rather than a move to be made at will.

Third, and most decisively, sustained prospective failure. If DTA audits applied to systems under design repeatedly misidentify where tracing will succeed or collapse under load, the framework loses its predictive function and becomes only a retrospective interpretive scheme. This would be the most consequential form of disconfirmation, because the framework's stronger claim is prospective.

None of these forms of evidence would be decisive on a single instance. Frameworks of this scope are weakened or strengthened by patterns across cases, not by individual data points. But the framework should be applied in a manner that allows such patterns to accumulate, which requires that its predictions be specific enough to be checked and its diagnostic instruments precise enough to be applied consistently. The prospective predictions of §5 and the six-step protocol of §3 are designed to support exactly this kind of accumulation.

Conclusion

Meaningful Human Control correctly identifies that oversight must be substantive rather than nominal. Its tracing condition correctly requires that the responsible agent be in a position to exercise the capacities for which she is held responsible. This article has argued that the architectural conditions under which being-in-a-position-to-exercise can be operationally satisfied form a distinct level of analysis, the level at which MHC compliance is either substantive or attestational, and the level that existing operationalizations of MHC have not yet fully isolated.

Decision-time accessibility names the variable at that level. Its six conditions are derived from what tracing requires rather than selected for their evocative power: Truth, Relation, Integration, Doing, Orientation, and Attention & Agency. The Judgment Cliff names the threshold beyond which the joint assembly of these conditions becomes architecturally infeasible. Where action occurs beyond the cliff, the operator may remain formally responsible while the conditions for substantive responsibility have been withdrawn; the resulting blame allocation is fictive accountability, and the institutional regime that produces it is attestational rather than meaningful.

The framework's contribution is operational and normative. Operationally, it specifies the architectural variable at which MHC's tracing condition is satisfied or fails, decomposes the variable into six conditions amenable to design, develops the Judgment Cliff threshold beyond which the conditions cannot assemble, and provides a six-step diagnostic protocol applicable across sectors. Normatively, it identifies an orientation, condition ethics, that takes the architectural conditions for responsibility as the primary site of ethical agency in technologically mediated governance, and it diagnoses fictive accountability as the institutional pathology that emerges when the architectural and the normative are decoupled.

What the framework does not do is settle the prior question. It does not decide whether human control should be required for any given decision class; it specifies what MHC's invocation requires when invoked. It does not absolve operators of duties of competence, honesty, reporting, and escalation; it specifies the narrower claim that when action occurs beyond the Judgment Cliff, primary blame allocation to the operator misdescribes the architecture of failure. It is not another theory of human factors. It is a theory of when human factors matter normatively for the tracing condition of Meaningful Human Control.

The practical question for governance is no longer whether humans remain in the loop. It is whether the loop preserves the architectural conditions under which human presence can count as meaningful control. Where those conditions are preserved, MHC tracing remains possible. Where they are withdrawn, oversight becomes attestational, and accountability becomes fictive, regardless of how robust the formal structure appears from outside the institution. The architectural conditions are not a residual feature of governance. They are its primary site.

Meaningful control requires accessible agency, not merely assigned responsibility.

References

1. Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. Frontiers in Robotics and AI, 5, 323836.
2. Endsley, M. R. (2017). Toward a theory of situation awareness in dynamic systems. In Situational awareness (pp. 9-42). Routledge.
3. Sweller, J., Ayres, P., & Kalyuga, S. (2011). Altering element interactivity and intrinsic cognitive load. In Cognitive load theory (pp. 203-218). New York, NY: Springer New York.
4. Hollnagel, E., Woods, D. D., & Leveson, N. (Eds.). (2006). Resilience engineering: Concepts and precepts. Ashgate Publishing, Ltd..
5. Reason, J. (1990). Human error. Cambridge university press.
6. Roff, H. M., & Moyes, R. (2016, April). Meaningful human control, artificial intelligence and autonomous weapons. In Briefing paper prepared for the informal meeting of experts on lethal Au-Tonomous weapons systems, UN Convention on certain conventional weapons.
7. Mecacci, G., & Santoni de Sio, F. (2020). Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. Ethics and Information Technology, 22(2), 103-115.
8. Methnani, L., Aler Tubella, A., Dignum, V., & Theodorou, A. (2021). Let me take over: Variable autonomy for meaningful human control. Frontiers in Artificial Intelligence, 4, 737072.
9. Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. Philosophy & technology, 34(4), 1057-1084.
10. Calvert, S. C. (2025). Principles and Framework for the Operationalisation of Meaningful Human Control Over Autonomous Systems. Science and Engineering Ethics, 31(5), 27.
11. Heikoop, D. D., Hagenzieker, M., Mecacci, G., Calvert, S., Santoni De Sio, F., & Van Arem, B. (2019). Human behaviour with automated driving systems: a quantitative framework for meaningful human control. Theoretical issues in ergonomics science, 20(6), 711-730.
12. Klein, G. A. (2017). Sources of power: How people make decisions. MIT press.
13. Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. Safety science, 27(2-3), 183-213.
14. Hutchins, E. (1995). Cognition in the Wild. MIT press.
15. Reason, J. (2016). Managing the risks of organizational accidents. Routledge.
16. Dekker, S. (2016). Drift into failure: From hunting broken components to understanding complex systems. CRC press.
17. Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction (pre-print). Engaging Science, Technology, and Society (pre-print).
18. Koschmann, T. D. (1987). Mind over machine: The power of human intuition and expertise in the era of the computer: Hubert L. Dreyfus and Stuart E. Dreyfus (Basil Blackwell, Oxford, 1986); 223 pages,£ 15.00.
19. Aristotle. (2000). Nicomachean ethics (R. Crisp, Trans.).
20. Arendt, H. (1981). The life of the mind: The groundbreaking investigation on how we think. HMH.
21. Beiner, R. (2013). Political judgement. Routledge.
22. Polanyi, M., & Sen, A. (2009). The tacit dimension: University of Chicago press. Chicago, IL.
23. Suchman, L. A. (2007). Human-machine reconfigurations: Plans and situated actions. Cambridge university press.
24. Norman Donald, A. (2013). The design of everyday things. MIT Press.
25. Levinas, E. (1979). Totality and infinity: An essay on exteriority (Vol. 1). Springer Science & Business Media.
26. Murdoch, I., & Midgley, M. (2013). The sovereignty of good. Routledge.
27. Thompson, D. F. (1980). Moral responsibility of public officials: The problem of many hands. American Political Science Review, 74(4), 905-916.
28. Young, I. M. (2011). Responsibility for justice. Oxford University Press.
29. Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253.
30. Weber, M. (1978). Economy and society: An outline of interpretive sociology (Vol. 2). University of California press.

31. Rosa, H. (2013). Social acceleration: A new theory of modernity. Columbia University Press.
32. Zeeman, E. C., & Barrett, T. W. (1979). Catastrophe theory, selected papers 1972-1977. IEEE Transactions on Systems, Man, and Cybernetics, 9(9), 609-610.
33. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., ... & Sugihara, G. (2009). Early-warning signals for critical transitions. Nature, 461(7260), 53-59.
34. Bainbridge, L. (1983). Ironies of automation. In Analysis, design and evaluation of man–machine systems (pp. 129-135). Pergamon.
35. Vaughan, D. (1996). The Challenger launch decision: Risky technology, culture, and deviance at NASA. University of Chicago press.
36. United States. Presidential Commission on the Space Shuttle Challenger Accident. (1986). Report to the President (Vol. 1). Presidential Commission on the Space Shuttle Challenger Accident.
37. Bovens, M. A. P. (1998). The quest for responsibility: Accountability and citizenship in complex organisations. Cambridge university press.
38. Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford university press.
39. DeFazio, P. A., & Larsen, R. (2020). Final committee report—The design, development & certification of the Boeing 737 MAX. The US House Committee on Transportation and Infrastructure, Subcommittee on Aviation. Washington, DC: US House Committee on Transportation and Infrastructure.
40. Demirci, S. (2022). The requirements for automation systems based on Boeing 737 MAX crashes. Aircraft engineering and aerospace technology, 94(2), 140-153.
41. Gorton, G. B. (2010). Slapped by the invisible hand: The panic of 2007. Oxford University Press.
42. MacKenzie, D. (2011). The credit crisis as a problem in the sociology of knowledge. American Journal of Sociology, 116(6), 1778-1841.
43. Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (2017). The evolution of crew resource management training in commercial aviation. In Human error in aviation (pp. 275-288). Routledge.
44. Smuha, N. A. (2025). Regulation 2024/1689 of the Eur. Parl. & Council of June 13, 2024 (eu artificial intelligence act). International Legal Materials, 64(5), 1234-1381.
45. Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. Information and privacy commissioner of Ontario, Canada, 5(2009), 12.
46. Friedman, B., Kahn Jr, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. In Early engagement and new technologies: Opening up the laboratory (pp. 55-95). Dordrecht: Springer Netherlands.
47. RISK, C. (2021). Risk management framework. COPRPORATE REPORT.
48. ISO/IEC. (2022). ISO/IEC 22989: Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. International Organization for Standardization.
49. ISO. (2020). ISO 9241-810: Ergonomics of human-system interaction — Part 810: Robotic, intelligent and autonomous systems. International Organization for Standardization.
50. Floridi, L. (2020). Distributed morality in an information society. In The Ethics of Information Technologies (pp. 63-79). Routledge.