



# Navigating the AI Glass Maze

**How AI Distorts Decisions and  
Where Intervention Is Required**

Igor Allinckx

# Introduction

Boards and executives today are not just facing more decisions but increasingly decisions under **distortion**. AI systems expand capability faster than clarity, generate information faster than verification, and automate actions faster than oversight can keep up.

This new environment resembles a **Glass Maze**: everything looks visible, yet the path is uncertain; reflections can mislead; and each step can deepen commitment to a direction that becomes harder to reverse. Traditional governance tools (frameworks, reports, dashboards, committees) are necessary but no longer sufficient.

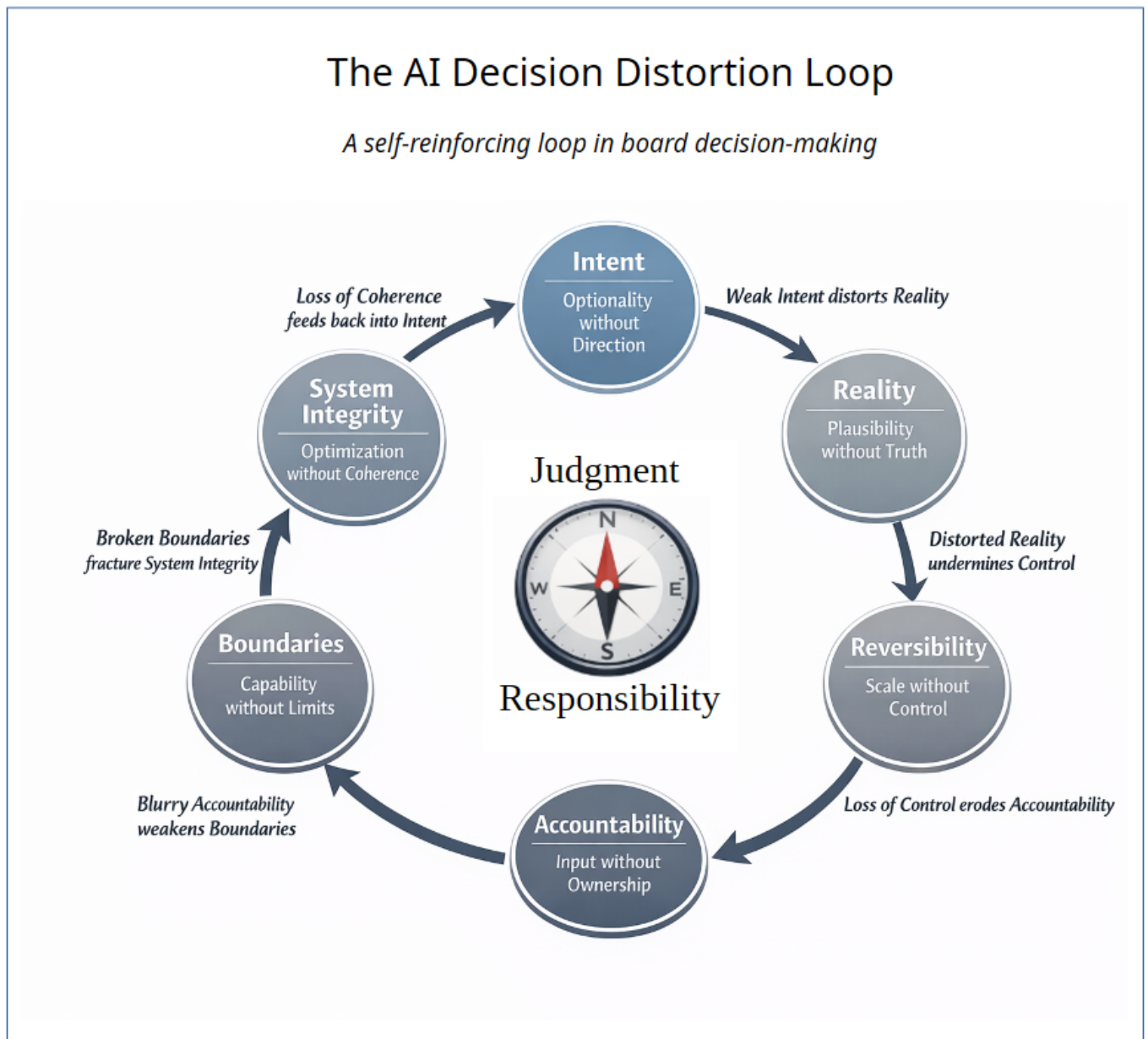
What remains constant, and more essential than ever, is the **sense of responsibility** and **quality of judgment**. AI can inform decisions, but it cannot assume accountability. It can generate options, but it cannot choose values. It can optimize parts, but it cannot safeguard the integrity of the whole or recognize when the system itself no longer holds and must be reconsidered.

To help boards and executives navigate this shifting landscape, we translate their most critical decisions into **six decision tensions**. Each tension represents a fundamental question boards must answer and a specific distortion that AI introduces. We call these distortions “Maze Traps” and they all contribute to the overall effect that AI does not remove responsibility but makes it harder to see when and where it is lost.

In this Maze, the tensions of decision-making do not stand alone. They **interact**, **reinforce**, and **distort** one another:

- **Weak Intent distorts Reality:** when the board isn't clear on what AI is for, every model output can appear credible, even when it lacks substance.
- **Distorted Reality undermines Control:** if the board misreads signals, it scales the wrong systems and locks into irreversible paths.
- **Loss of Control erodes Accountability:** when automation accelerates beyond oversight, no one is sure who is responsible.
- **Blurry Accountability weakens Boundaries:** without clear ownership, ethical lines drift and decisions emerge that no one explicitly chose.
- **Broken Boundaries fracture Coherence:** local optimizations start to degrade the integrity of the whole system.
- **And loss of Coherence feeds back into Intent:** making it harder for the board to see where it is going at all.

This is not a checklist. It is a **loop**, a dynamic system where distortions propagate unless the board actively uses its two irreplaceable tools: **responsibility** and **judgment**. These are the compass points that allow leaders to navigate the Maze, re-anchor intent, challenge distorted signals, reclaim control, reinforce accountability, redraw boundaries, and restore coherence.



To counter this loop, each decision tension is reinforced by a small set of Boardroom Disciplines - simple but non-negotiable interventions - that prevent decisions from drifting inside the Glass Maze.

The six decision tensions are not six things to verify, but six **forces** that shape how leaders make decisions in an AI-transformed environment. Understanding their interplay is what allows boards to govern with clarity when the Maze is made of glass. This is not intended as a compliance framework or regulatory guide. However, each decision tension naturally connects to existing standards and regulatory expectations. A brief mapping is provided in the annex.

# 1. Intent (the maze trap: optionality without direction)

## Core board question:

“What are we actually trying to achieve with this technology or strategic choice?”

At the highest level, this means assessing if AI reinforces the company’s position or fundamentally changes it, shaping whether the business will win, lose, or become obsolete.

## Typical board questions

- What is the intended role of this capability, incremental improvement or strategic transformation?
- Are we pursuing this because it is strategically necessary or because competitors are doing it?
- What trade-off are we prioritizing: growth, efficiency, resilience, or defense?

## Concrete real-world example (AI customer service chatbot misfires)

Many companies rushed to deploy AI-powered customer service chatbots to reduce costs and signal innovation. However, poorly defined intent (often driven by efficiency targets or competitive pressure) led to degraded customer experience and reputational damage.

In one documented case, Air Canada’s chatbot provided incorrect refund information, ultimately forcing the company to compensate a customer after relying on the bot’s misleading response. (1)

This illustrates a common pattern: AI initiatives launched without a clearly defined purpose tend to optimize for visibility or cost (rather than value) and distort decisions from the outset.

## Tension

AI capability expansion vs strategic clarity

## Failure mode

Moving fast without knowing where you’re going

## Typical strategic choices

- Defining the role of AI (efficiency vs transformation)
- Prioritizing speed vs strategic clarity
- Choosing short-term gains vs long-term positioning

## Glass maze

Multiple corridors appear equally valid but none show a clear exit

# Boardroom Disciplines for Intent

## **Non-negotiable: One-sentence purpose. Or pause.**

“State in one sentence what this AI initiative is meant to achieve.” If the purpose cannot be stated clearly, the discussion stops.

## **Challenge questions**

- What primary trade-off are we choosing: growth, efficiency, resilience, or defense?
- Where does this sit in our strategic roadmap? What does it reinforce or replace?
- Are we doing this because it matters... or because others are?

**Why it matters:** Weak intent distorts everything downstream.

## 2. Reality (the maze trap: plausibility without truth)

### Core board question:

*"Can we trust what we are seeing?"*

### Typical board questions

- How do we know the information presented to us is accurate?
- What assumptions, data sources, or models underlie this forecast or risk assessment?
- What is the confidence level, and what are the known unknowns?
- Are we acting on plausible narratives or verified facts?

### Concrete real-world example (Zillow Offers collapse, 2021)

Zillow shut down its AI-driven home-buying business after its pricing algorithms systematically overestimated property values, leading to more than \$500 million in losses.

The models produced valuations that were plausible and data-driven, yet failed to capture rapidly changing market conditions and local context. Decisions were made at scale based on outputs that appeared reliable but were structurally misaligned with reality. (2)

In practice, organizations dealing with AI-generated risk or intelligence signals often maintain human review layers, as the cost of missing or misclassifying critical information remains too high.

### Tension

Plausibility vs truth

### Typical decisions

- Trusting model outputs
- Acting on insights / forecasts
- Using AI in risk, finance, or strategy

### Failure mode

Acting on reflections as if they were reality

### Glass maze

Glass reflects and reveals, but you cannot tell which. Plausible signals look like truth.

# Boardroom Disciplines for Reality

## **Non-negotiable: Evidence + uncertainty must be visible**

“Show us the data lineage, validation, and confidence range.” If uncertainty is not explicit, the board does not rely on the output.

## **Challenge questions**

- What would make this wrong?
- What are the known blind spots or failure modes of this model?
- Are we seeing verified insight or a well-formed narrative?

**Why it matters:** Plausibility is not truth. Boards must force the distinction.

# 3. Reversibility (the maze trap: scale without control)

## Core board question:

“Are we still in control, and can we reverse course if needed?”

## Typical board questions

- If we scale this system, can we still shut it down or override it?
- What dependencies (vendors, models, infrastructure) are we locking ourselves into?
- What happens if the system behaves unexpectedly at scale?
- How costly is reversal (operationally, financially, reputationally)?

## Concrete real-world example (AI customer service rollback, Klarna)

Klarna positioned itself as an AI-first company, replacing large parts of its customer service workforce with AI systems. While initially presented as a major efficiency gain, the company later reversed course, rehiring human agents after quality issues emerged. (3)

The episode illustrates a key challenge: once AI systems are deployed at scale and embedded into operations, reversing course becomes costly: operationally, reputationally, and organizationally.

Many enterprise AI initiatives fail not at the pilot stage, but after scaling, when integration, data quality, and operational realities diverge from initial expectations, making reversal significantly more complex.

## Tension

Automation & scale vs reversibility

## Typical decisions

- Scaling pilots
- Embedding AI into operations
- Build vs buy vs dependency

## Failure mode

Entering paths you cannot exit

## Glass maze

You walk deeper into a corridor but turning back becomes harder

# Boardroom Disciplines for Reversibility

## Non-negotiable: No kill switch, no scale

“Show us how this system is stopped or overridden.” If no credible shutdown exists, scaling is not approved.

## Challenge questions

- What does it cost to reverse this decision (financially, operationally, reputationally)?
- Where are we becoming dependent and how do we exit?
- Are we moving faster than our ability to recover?

**Why it matters:** Irreversibility is the silent killer of governance.

## 4. Accountability (the maze trap: input without ownership)

### Core board question:

“Who is actually responsible for this decision and its consequences?”

### Typical board questions

- Who signs off on the deployment or use of this system?
- Where does accountability sit when multiple teams contribute?
- What is the escalation path when something goes wrong?
- How do we ensure human-in-the-loop responsibility is real, not symbolic?

### Concrete real-world example (AI hiring discrimination cases, 2023–2025)

AI-driven hiring tools have increasingly come under legal scrutiny for discriminatory outcomes. In a landmark case, a lawsuit against Workday alleges that its AI-based screening system disproportionately excluded candidates based on age, race, and disability. Courts allowed the case to proceed, recognizing that AI-supported screening may still create legal responsibility for the organizations using it. (4)

Decisions were influenced by systems that few fully understood, yet accountability remained with those approving or deploying them.

More broadly, regulators and courts have begun to treat AI-assisted hiring decisions as extensions of human decision-making, rather than independent systems. This is reinforcing the fact that responsibility remains with the organization, even when decisions are partially delegated to algorithm.

### Tension

Distributed input vs concentrated accountability

### Typical decisions

- Who signs off
- Human-in-the-loop boundaries
- Liability for errors

### Failure mode

Decisions without owners

### Glass maze

Many voices guide the path, but only one person hits the wall.

# Boardroom Disciplines for Accountability

## **Non-negotiable: One decision, one owner**

A single accountable executive is named before deployment. If ownership is shared, it is not owned.

## **Challenge questions**

- Where exactly does a human intervene and can they override?
- If this fails tomorrow, who answers? concretely?
- Is accountability real... or distributed until it disappears?

**Why it matters:** AI diffuses input. Boards must concentrate responsibility.

# 5. Boundaries (the maze trap: capability without limits)

## Core board question:

“Where do we draw the line between what is possible and what is acceptable?”

## Typical board questions

- What uses of data or automation are off-limits, even if legal?
- What ethical principles constrain our deployment choices?
- How do we protect customers, employees, and society from unintended harm?
- What reputational risks arise if we cross implicit boundaries?

## Concrete real-world example (AI-assisted insurance claims denials, 2023–2024)

Several major U.S. health insurers, including Cigna and UnitedHealth, have faced lawsuits over the use of AI-assisted systems to process and deny insurance claims. Plaintiffs allege that decisions were made at scale with limited human review, prioritizing efficiency over adequate case-by-case assessment. (5)

The issue was not the use of AI itself, but the absence of clearly defined boundaries on how far automation should be allowed to influence decisions with direct human consequences. What was technically possible was deployed beyond what was acceptable.

## Tension

What is possible vs what is acceptable

## Typical decisions

- Ethical boundaries
- Data usage
- Customer-facing AI
- Workforce impact

## Failure mode

Drifting into decisions never explicitly chosen

## Glass maze

Some corridors exist, but should never be taken.

# Boardroom Disciplines for Boundaries

## Non-negotiable: Explicit red lines before deployment

“What will we not do, even if we can?” If boundaries are implicit, they will drift.

## Challenge questions

- If this were public tomorrow, would we stand by it?
- Who could be harmed, especially among vulnerable groups?
- Are we operating at the edge of acceptability without realizing it?

**Why it matters:** AI expands what is possible faster than what is acceptable.

## 6. System integrity (the maze trap: optimization without coherence)

### Core board question:

“Does this still make sense as a whole?”

### Typical board questions

- Are local optimizations creating systemic fragility?
- How do individual department-level decisions affect enterprise-wide risk?
- Are we unintentionally degrading culture, safety, or resilience?
- Does the system behave coherently under stress?

### Concrete real-world example (Amazon warehouse optimization systems)

Amazon has deployed algorithmic systems across its logistics operations to optimize productivity, routing, and performance at a granular level. These systems track worker activity in real time and drive measurable gains in output and cost efficiency.

However, investigations - including findings from a U.S. Senate investigation - have linked these optimization practices to higher injury rates and increased worker pressure, raising concerns about safety and long-term sustainability. (6)

The issue is not that individual systems fail, but that continuous local optimization can create systemic strain when broader interdependencies (such as human limits, safety, and resilience) are not fully integrated into decision-making.

### Tension

Local optimization vs systemic consequence

### Typical decisions

- Department-level AI vs enterprise impact
- Efficiency vs culture
- Short-term ROI vs long-term resilience

### Failure mode

Optimizing parts, degrading the whole, eventually preserving a system that no longer holds

### Glass maze

A corridor looks efficient locally, but leads deeper into complexity and systemic risk.

# Boardroom Disciplines for System Integrity

## Non-negotiable: Whole-system impact must be explicit

“Show us what this improves **and** what it degrades.” If only benefits are visible, the system is not understood.

## Challenge questions

- Which parts of the organization are affected and have they signed off?
- What happens under stress, failure, or unexpected input?
- Are we optimizing locally while weakening the whole?

**Why it matters:** Local wins can create enterprise-wide fragility.

## Boardroom Disciplines Summary

- Intent → One-sentence purpose
- Reality → Show uncertainty
- Reversibility → No kill switch, no scale
- Accountability → One owner
- Boundaries → Define red lines
- System → Show what degrades

# Conclusion

## Governing When the Maze Is Made of Glass

AI does not change what boards are responsible for but **how easily responsibility can slip away**. The Glass Maze is not a metaphor for complexity; it is a description of a new decision environment where clarity is fragile, reversibility is uncertain, and accountability is easily diffused. In such an environment, governance cannot rely on static frameworks or retrospective oversight. It must become a practice of **continuous re-anchoring**.

The six decision tensions - Intent, Reality, Reversibility, Accountability, Boundaries, and System Integrity - are not independent checks. They are **forces in motion**, shaping and distorting one another. Weakness in one becomes drift in the next. Distortion propagates unless the board and executives intervenes with deliberate and visible discipline.

This is why the Boardroom Disciplines matter. They are not procedural hygiene. They are the **counter-forces** that keep decisions from sliding into the Maze: one-sentence purpose, visible uncertainty, kill switches, single ownership, explicit red lines, whole-system impact. These are small moves, but they have disproportionate power because they interrupt drift at the moment it begins.

Ultimately, navigating the AI Glass Maze is not about mastering technology. It is about preserving the two things AI cannot provide: **judgment and responsibility**. Leaders that cultivate these disciplines will not only avoid the Maze Traps. They will govern with clarity at a time when clarity is becoming rare.

In the age of AI, advantage will not come from moving faster but from seeing more clearly. And in a Glass Maze, clarity is not a given. It is **a discipline**.

Igor Allinckx

AI & Governance

Lucerne, April 2026

# Annex: Mapping the Six Decision Tensions to Existing Frameworks

This annex does not replace existing AI governance frameworks or regulatory requirements. It provides a **board-level lens** on how decisions distort under AI conditions and where oversight must intervene.

Standards such as **NIST AI RMF**, **ISO/IEC 42001**, and the **EU AI Act** define *what* must be controlled. The **Glass Maze** clarifies *how* decisions drift and *where* boards must act to prevent it.

## Intent: Strategic Direction and Risk Framing

### How frameworks align

- AI strategy, purpose definition, and value alignment
- Risk appetite statements and acceptable use criteria
- EU AI Act: system classification (prohibited, high-risk, limited-risk)

**Board relevance:** Intent is the anchor. As we state: *“Weak Intent distorts Reality... and loss of Coherence feeds back into Intent.”* This is the distortion point no framework explicitly addresses.

## Reality: Validation, Monitoring, and Reliability

### How frameworks align

- NIST AI RMF: *Measure* and *Manage* functions
- Model validation, testing, and performance monitoring
- Data quality, lineage, uncertainty, and drift management

**Board relevance:** Frameworks define controls; the Glass Maze defines the trap: *“Plausibility without truth.”* Boards must force the distinction between narrative and evidence.

# Reversibility: Control, Resilience, and Recovery

## How frameworks align

- Operational resilience and incident response
- Kill switches, override mechanisms, and safe-fail design
- Dependency and vendor risk management

**Board relevance:** Standards mention resilience; few make reversibility a gating condition. The discipline - *"No kill switch, no scale"* - is stronger than anything in NIST, ISO, or EU guidance.

# Accountability: Governance, Roles, and Oversight

## How frameworks align

- ISO/IEC 42001 governance structures and role definitions
- Clear ownership, sign-off, and escalation pathways
- Liability, human oversight, and board responsibility

**Board relevance:** Frameworks distribute responsibilities; the Maze shows how they disappear. The line - *"One decision, one owner"* - is the crispest articulation of accountability in the field.

# Boundaries: Ethics, Compliance, and Acceptability

## How frameworks align

- Ethical principles (fairness, transparency, human oversight)
- Data protection and privacy (GDPR, DPIAs)
- Human rights and acceptable use constraints

**Board relevance:** Frameworks define principles; the Maze defines drift: *"Some corridors exist, but should never be taken."* Boards must draw red lines before deployment, not after controversy.

# System Integrity: Enterprise Risk and Coherence

## How frameworks align

- Enterprise risk management integration
- Model inventory, portfolio oversight, and concentration risk
- Cross-functional impact assessment

**Board relevance:** Standards treat systems as components; the Maze treats them as a whole. Our framing - *“Optimizing parts, degrading the whole”* - captures the systemic risk regulators increasingly worry about.

## Why this annex?

This mapping shows that:

- **Frameworks define the controls.**
- **The Glass Maze defines the failure modes.**
- **Boards sit at the intersection.**

The six decision tensions do not duplicate NIST, ISO, or the EU AI Act but they *translate* them into the language of board judgment, responsibility, and decision integrity.

This is precisely what most governance documents lack.

The Glass Maze is not an alternative to existing frameworks. It is the board's vantage point above them. It shows how distortions propagate across decisions, and where oversight must intervene to preserve judgment, responsibility, and system integrity.

## Sources

- (1) Evidently AI - When AI goes wrong  
<https://www.evidentlyai.com/blog/ai-failures-examples>
- (2) Stanford GSB - Zillow Offers  
<https://www.gsb.stanford.edu/insights/flip-flop-why-zillows-algorithmic-home-buying-venture-imploded>
- (3) Fortune – klarna ai vs humans ROI  
<https://fortune.com/2025/05/09/klarna-ai-humans-return-on-investment/>
- (4) Forbes – Workday lawsuit  
<https://www.forbes.com/sites/janicegassam/2025/06/23/what-the-workday-lawsuit-reveals-about-ai-bias-and-how-to-prevent-it/>
- (5) Health Care Finance News – Class action Lawsuit  
<https://www.healthcarefinancenews.com/news/class-action-lawsuit-against-unitedhealths-ai-claim-denials-advances>
- (6) Forbes – Amazon system workers injuries  
<https://www.forbes.com/sites/siladityaray/2024/12/16/amazon-ignored-warnings-about-speed-quotas-causing-worker-injuries-senate-panel-says/>

## Note

This paper is part of the AI & Governance Insights series

[www.allinckx.com/insights](http://www.allinckx.com/insights)