Reality Drift Markers:

A Diagnostic Framework for Synthetic Systems

Prof. Daniel Rook
Center for Systems Legibility
Working Paper Series – CSL-DR-004

Abstract

Reality Drift Markers (RDM) is a diagnostic framework for identifying when organizational, algorithmic, or institutional systems have transitioned from reality-tracking behavior into self-referential simulation. The framework operationalizes *reality drift* as a set of observable system properties and provides a standardized rubric for screening epistemic degradation in synthetic environments, including AI-mediated platforms, performance management systems, and high-reflexivity institutions.

Scope

The Reality Drift Markers framework applies to:

- AI-enabled organizations
- Algorithmic governance systems
- Performance management platforms
- Institutional knowledge infrastructures
- Policy modeling and simulation environments

The framework is intended for use in:

- Organizational audits
- Risk management assessments
- AI governance reviews
- Institutional diagnostics
- Knowledge system evaluations

RDM does not assess system effectiveness. It assesses epistemic integrity under reflexive conditions.

Core Concept

Reality Drift is defined as a system condition in which internal representations, metrics, and symbolic processes increasingly replace external reality as the primary reference point for decision-making.

In high-drift systems:

- Symbolic coherence increases
- Operational legitimacy decreases
- Feedback loses corrective force

Reality Drift Markers translate this condition into screenable system properties.

Diagnostic Structure

Each marker includes:

- Description
- Observable signals
- Risk level
- Remediation suggestions

Markers may co-occur and reinforce one another.

Reality Drift Markers

<u>Ornamental Constraint</u>

Description
Constraints exist formally but lack operational force. Rules, policies, and limits function primarily as symbolic indicators of control rather than as effective boundaries on behavior.

Observable Signals

- Policies that are frequently bypassed without consequence
- Compliance rituals with no enforcement impact
- Controls that exist only at the documentation layer

Risk Level
High in mature bureaucracies and AI-mediated institutions.

Remediation

- Introduce enforceable constraints tied to real-world costs

- Reconnect policy mechanisms to material outcomes
- Remove non-functional governance artifacts

Confidence Without Cost

Description
Systems generate high-certainty outputs without exposure to failure, loss, or corrective feedback.

Observable Signals

- Predictive systems with no penalty for error
- Executive dashboards insulated from downstream consequences
- AI systems evaluated only on internal metrics

Risk Level
Critical in recursive decision environments.

Remediation

- Introduce consequence-sensitive feedback loops
- Tie system outputs to accountable outcomes
- Implement external validation channels

Invalidation Failure

Description
The system lacks mechanisms for detecting when its internal models are no longer aligned with reality.

Observable Signals

- Persistent use of outdated assumptions
- Models that cannot be formally challenged
- Decision frameworks that resist disconfirmation

Risk Level
Severe in closed or self-referential systems.

Remediation

- Embed adversarial review processes
- Introduce independent reality checks
- Enforce periodic model invalidation cycles

<u>Non-Stoppability</u>

Description
Systems continue operating despite loss of legitimacy, accuracy, or relevance.

Observable Signals

- Projects that cannot be terminated
- Metrics that persist despite known distortions
- AI systems that self-perpetuate without performance review

Risk Level
High in large-scale automated infrastructures.

Remediation

- Establish formal shutdown criteria
- Introduce termination authority
- Implement system-level “stop conditions”

<u>Performance Without Consequence</u>

Description
Outputs are optimized for appearance of effectiveness rather than real-world impact.

Observable Signals

- KPI achievement without material improvement
- Success metrics disconnected from lived experience
- Systems optimized for reporting, not outcomes

Risk Level
Universal in metric-driven institutions.

Remediation

- Replace proxy metrics with outcome-linked indicators
- Audit symbolic performance layers
- Re-anchor evaluation to external reality

Risk Classification

Systems may be classified according to cumulative marker presence:

Marker Density	System State
0–1	Reality-tracking
2–3	Drift-prone
4–5	High simulation risk
5+	Fully synthetic operational regime

Implementation Notes

Organizational Use

RDM may be applied through:

- Structured interviews
- Policy audits
- Knowledge system reviews
- AI governance assessments

AI Governance Use

RDM supports evaluation of:

- Model training environments
- Recursive content pipelines
- Decision-support systems
- Automated compliance tools

Implications

High-functioning systems may exhibit strong internal coherence while becoming increasingly detached from external reality. In such conditions, symbolic stability replaces epistemic accuracy, and institutional behavior shifts from problem-solving to simulation maintenance.

Reality Drift Markers provide a minimal technical vocabulary for diagnosing this condition using observable system properties rather than subjective interpretation.

References

Rook, D. (2019). *Symbolic Control and Institutional Legibility*. Systems & Society Review.
Chen, M. (2020). *Epistemic Governance in Automated Organizations*. Knowledge Infrastructure Quarterly.

Voss, A. (2021). *Recursive Systems and Model Degradation*. Semantic Systems Proceedings.
Moravec, E. (2018). *Semantic Fidelity and Synthetic Knowledge*. Journal of Computational Epistemology.