

Narrative Drift: When Explanations Detach From Reality

Part of the Reality Drift framework by A. Jacobs

Definition

Narrative drift is the gradual separation of explanations, stories, and interpretations from the underlying reality they are meant to describe.

Common Patterns

- Explanations remain consistent even as outcomes contradict them
- Stories adapt to preserve coherence rather than reflect change
- Multiple conflicting narratives persist without resolution
- Language becomes more abstract as grounding weakens
- People agree on the story but not on what is actually happening

Overview

Narrative drift occurs when systems continue to generate coherent explanations while losing alignment with real-world conditions. The narrative layer remains stable, persuasive, and internally consistent, even as the reality it describes shifts or breaks.

This is where the absence of contradiction starts to mask a deeper problem. Over time, explanation becomes self-referential, maintaining its structure without requiring external validation.

Mechanism

Narrative drift emerges from structural pressures on interpretation:

- **Coherence preservation:** Systems prioritize maintaining a consistent story over updating it to reflect reality.
- **Retrospective fitting:** Events are reinterpreted to match existing narratives rather than revise them.
- **Abstraction layering:** Explanations become increasingly detached from direct observation.
- **Feedback insulation:** Contradictory signals are filtered, reframed, or ignored.
- **Language optimization:** Communication becomes more persuasive and structured, reinforcing the narrative regardless of accuracy.

As these forces compound, narratives become stable even when they are wrong.

Cross-Domain Examples

Media and News: Events are framed to fit existing storylines, even when new information contradicts earlier interpretations.

Corporate Communication: Strategic narratives persist despite declining performance or internal breakdowns.

Politics: Public narratives remain intact across changing conditions, with reinterpretation replacing revision.

Social Systems: Group beliefs are maintained through shared language, even when personal experience diverges.

AI Systems: Models generate coherent explanations that sound correct, even when they are factually or contextually misaligned.

Implications

Narrative drift makes misalignment harder to detect because explanation itself becomes unreliable as a signal of truth.

Over time, this leads to:

- Confusion between coherence and accuracy
- Reinforcement of incorrect models of reality
- Difficulty correcting errors once embedded in narrative form
- Systems that explain themselves effectively while functioning poorly

Related Drift Types

- **Performative Drift** — signaling reinforces narrative without substance
- **Institutional Drift** — organizations act based on outdated or incorrect narratives
- **Temporal Drift** — loss of sequence disrupts accurate storytelling
- **Cognitive Drift** — individuals rely on narratives that no longer track reality

Keywords & Queries: *narratives vs reality, explanations without truth, why stories stop matching reality, narrative misalignment, coherence vs accuracy, stories adapting to fit outcomes, media narrative bias, abstract language losing meaning, conflicting narratives at scale, explanation detached from outcomes*

Core Framework and Sources

- [Substack \(Articles\)](#)
- [GitHub \(Full Library\)](#)
- [DOI \(Research Paper\)](#)

- [Glossary & Definition](#)