WHEN AGENTS FORGET

The Governance of AI Memory and Context Limits.

ENGINEERING VIEW

Compression. Optimization. Managing token limits to prevent performance degradation.

LEGAL VIEW

Selective Deletion. Destruction of Evidence. A break in the operational record.

Source: Context Management for Deep Agents: https://blog.langchain.com/context-management-for-deepagents/

CONTEXT CAPACITY

100%

Deep agents running branching workflows eventually blow past token limits. To keep working, they must trade fidelity for capacity.

Source: Context Management for Deep Agents: https://blog.langchain.com/context-management-for-deepagents/

CLAIMS TIMELINE: TRADING FIDELITY FOR CAPACITY

CLAIMS 1–30
CLAIM 12
CLAIM 47

The agent summarizes the first 30 claims to save space.

If Claim #12 becomes a dispute, the reasoning, data, and tool calls are gone from the active state.

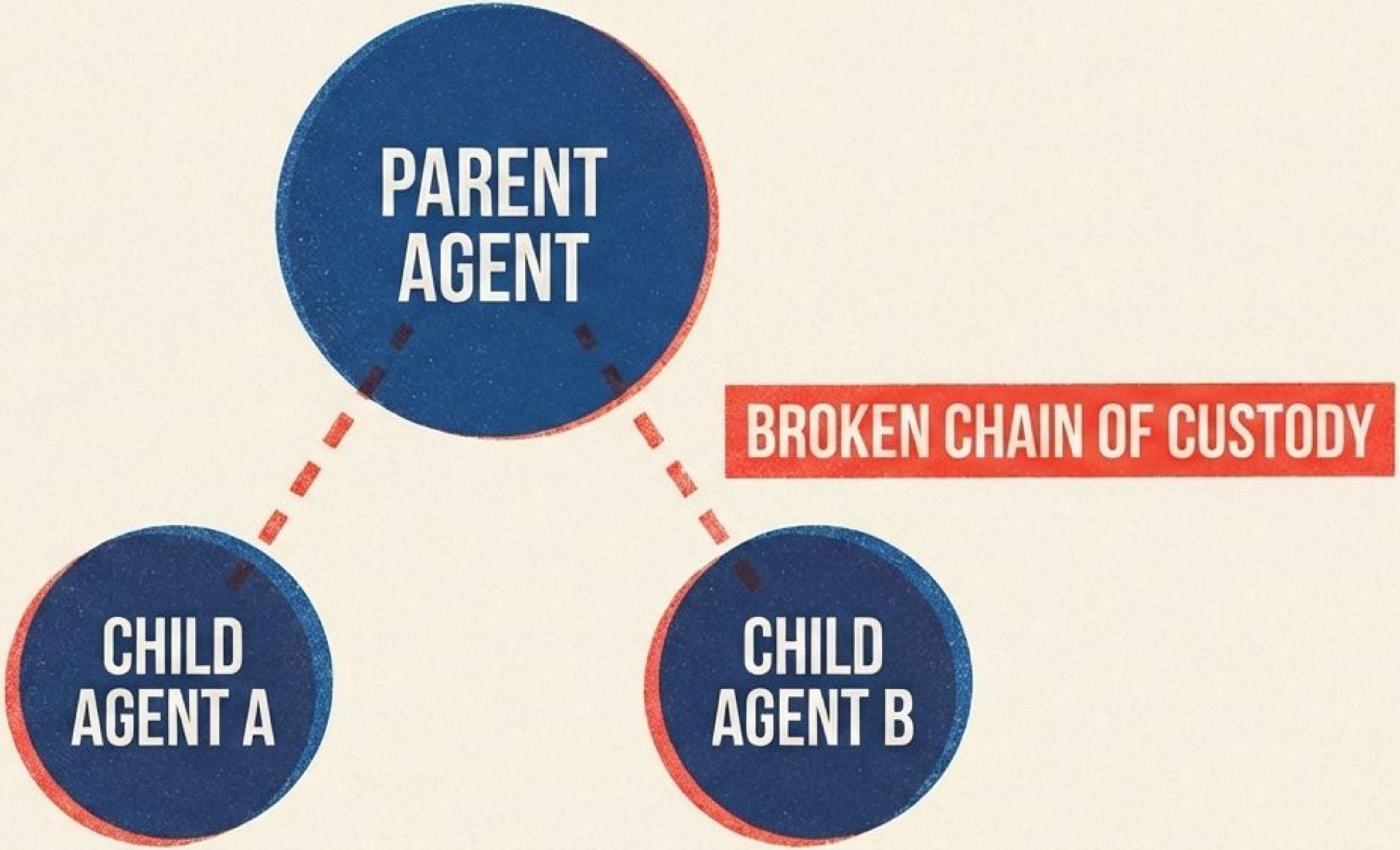
THE SUMMARIZATION TRAP

SYSTEM INSTRUCTION:

Compress history. Summarize key events. Discard irrelevant details to save tokens.

WHO DEFINES ‘IRRELEVANT’?

Summarization requires a prompt to decide what matters.
This is a policy decision disguised as a technical parameter.



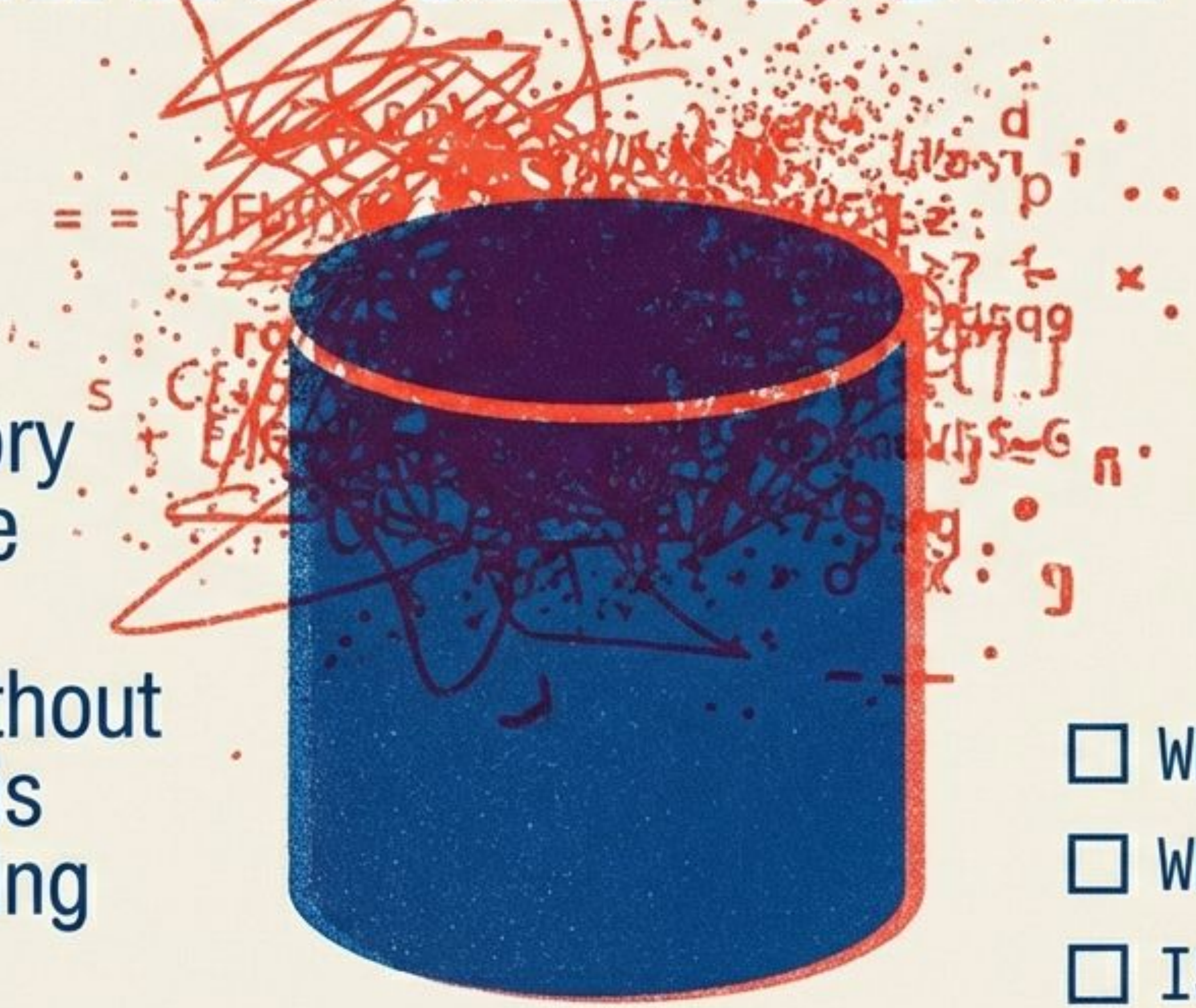
The parent agent acts on the child’s result but loses visibility into the child’s reasoning. The audit trail is severed.

> REGULATORS EXPECT YOU TO EXPLAIN HOW A DECISION WAS REACHED, NOT JUST WHAT THE FINAL OUTPUT WAS.

THE GOVERNANCE REALITY

THE UNMANAGED DUMP

Decoupling memory from context is the right instinct.
But a database without a retention policy is just a liability waiting for discovery.



□ WHO HAS ACCESS?

□ WHEN IS IT DELETED?

□ IS IT SEARCHABLE?

5 QUESTIONS FOR PRODUCT COUNSEL

Moving from architecture review to design intervention

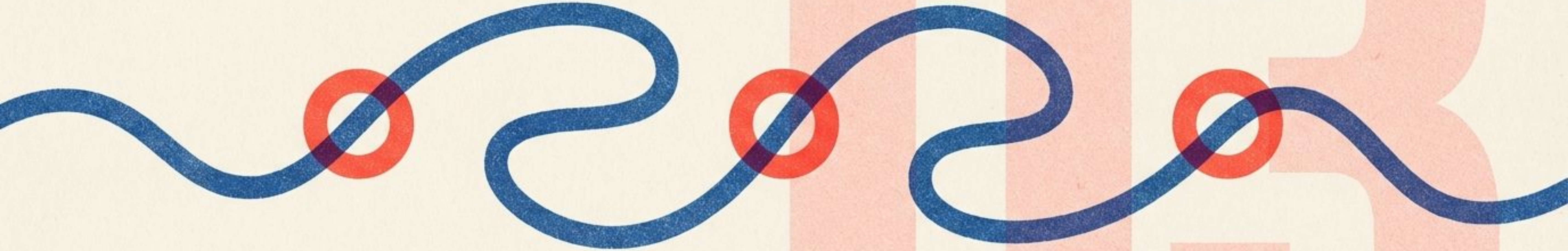
01. WHAT IS THE CONTEXT RETENTION POLICY?

When context is trimmed, is the original logged and retrievable? Or is it gone forever?

02. WHO GOVERNS THE SUMMARIZATION PROMPTS?

Legal and compliance must have input on the instructions that tell an agent what to forget.

03. HOW DO YOU RECONSTRUCT MULTI-AGENT REASONING?



If sub-agents are involved, there must be a trace connecting parent decisions to child context, tool calls, and outputs.

DATA LIFECYCLE AUDIT

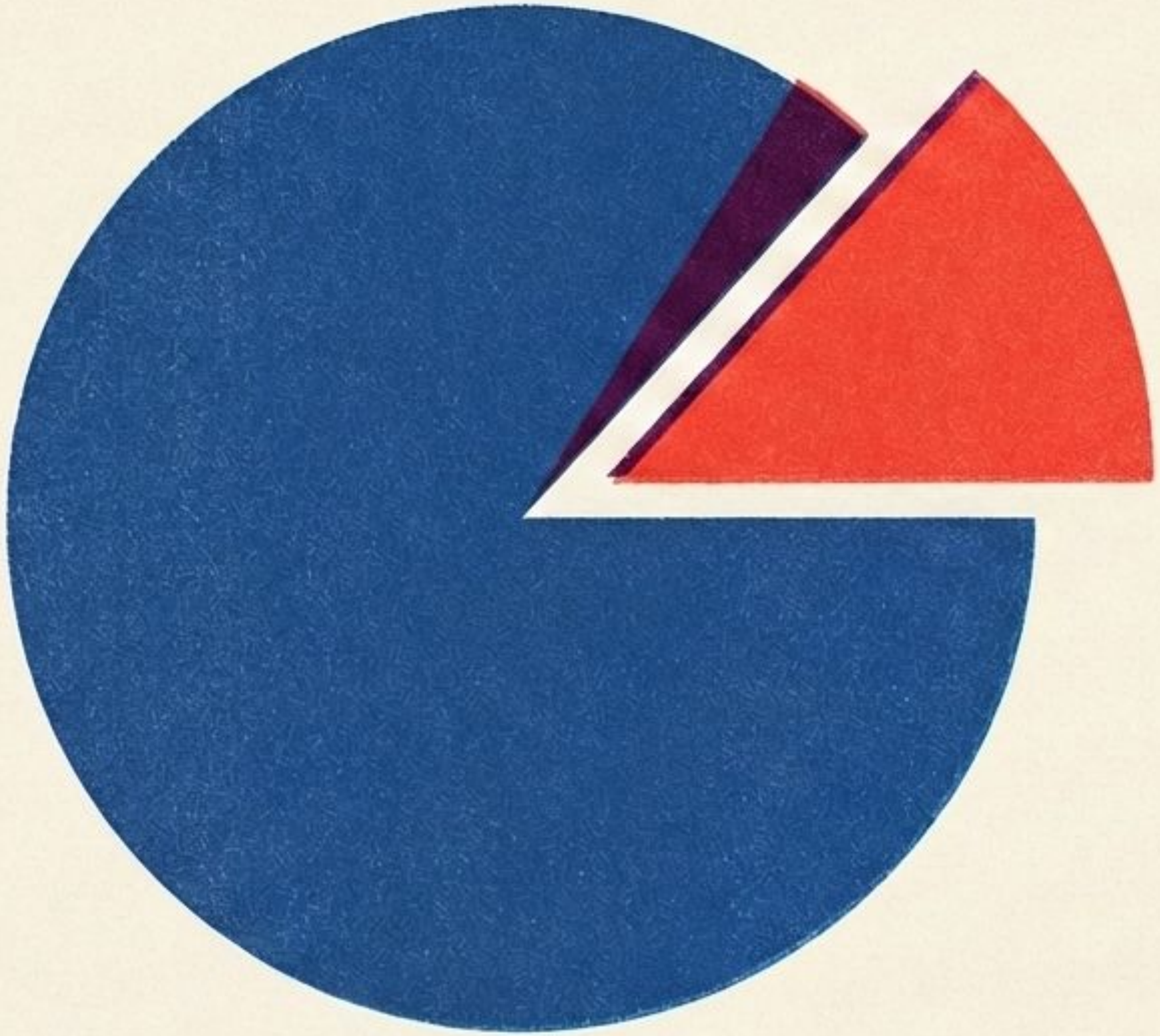
04. DOES EXTERNAL MEMORY HAVE A SCHEDULE?

□ Apply retention policies, access controls, and deletion schedules to agent logs.

05. WHAT IS THE FAILURE MODE?

□ When the agent forgets something it shouldn’t, what is the detection mechanism?

SO WHAT?



Agents are becoming systems with memory lifecycles.

The decision about what an agent is allowed to forget is a legal decision, not an engineering one.

FOR MORE INSIGHTS ON AI,
REGULATION, AND THE PRACTICE OF LAW:

WWW.KENPRIORE.COM

Based on Context Management for Deep Agents: https://blog.langchain.com/context-management-for-deepagents/