

Context Is the Differentiator

HOW LARGE LAW FIRMS TURN AI INTO STRUCTURAL ADVANTAGE

Contents

INTRODUCTION..... 3

WHEN FASTER DOES NOT MEAN BETTER..... 4

THE UNDERLYING CONSTRAINT: COHERENCE 5

AI AS AN ACCELERANT OF STRUCTURAL PRESSURE..... 7

CONTEXT AS INFRASTRUCTURE 8

FROM FRAGMENTED SYSTEMS TO A COHERENT OPERATING MODEL 10

WHAT CHANGES: CAPACITY, CLARITY, AND CONTROLLED SCALE 14

ASSESSING CONTEXT MATURITY 16

Introduction

Large law firms have moved quickly to deploy AI across drafting, research, review, and analysis. In many firms, adoption remains concentrated in pilots and practice-specific rollouts rather than embedded across the operating environment. Even where usage has expanded, it is often uneven across teams, workflows, and systems.

Those investments are producing meaningful gains at the level of individual tasks. Lawyers can generate first drafts more quickly, surface relevant information faster, and complete certain forms of analysis with less manual effort.

What remains less clear is how reliably those gains translate into firm-level performance. In many large firms, there is still a visible gap between faster task execution and broader operational improvement. Matters do not consistently move from initiation to completion with greater speed or clarity. Time saved in one part of the workflow often reappears elsewhere, in review, verification, coordination, or rework. AI can improve the speed of production without changing the conditions that determine whether work actually advances.

That gap points to a deeper issue. In large law firms, legal work moves across multiple systems, teams, and decision points. Documents, communications, financial data, and institutional knowledge each capture part of the matter, but the state of the work is rarely maintained in a continuous, shared form. The burden of assembling the full picture falls back on lawyers and staff. As AI increases the speed and volume of output, that burden becomes more visible and more consequential.

1. When faster does not mean better

AI has compressed the early stages of legal work. Drafting cycles are shorter. Research is more immediately accessible. Analysis that previously required sustained effort can now be generated in compressed timeframes. At the level of individual tasks, the improvement is observable and measurable.

At the level of the firm, the picture is less consistent. In many large firms, these gains do not translate cleanly into improvements in realization, margin, or throughput. Matters may start faster, but they do not reliably finish faster. Time saved in one stage of work often reappears in another: drafts are produced quickly but require extended review; analysis is generated rapidly but must be validated, contextualized, and reconciled before it can be used.

This is not a limitation in the tools. It reflects how work is absorbed by the organization. When the cost of producing output decreases, the volume of output entering the system increases. Each output still needs to be evaluated, aligned with the broader matter, and integrated into downstream work. That effort does not disappear so much as redistributes.

Gains at the point of production are real. But they often redistribute effort rather than eliminate it.

Three patterns tend to emerge. First, production introduces sustained verification pressure. AI-generated drafts, summaries, and analyses must be checked against source materials, validated for accuracy, and assessed for completeness within the context of the matter. The faster work is produced, the more frequently this validation cycle occurs. Time previously spent drafting shifts toward confirming whether the output can be relied upon.

Second, acceleration increases coordination frequency. More drafts, iterations, and intermediate artifacts introduce more moments where alignment is required. Each participant must understand how their work connects to the current state of the matter. Where that state is not continuously maintained, alignment must be re-established repeatedly through messages, meetings, and review cycles.

Third, expanded output produces parallel representations of the same work. A document may reflect one set of assumptions; an email thread may reflect

another; a financial model may incorporate changes not yet reflected in the legal documents. Each representation can be internally consistent. Across the matter, they may not align. Without a unifying structure that maintains the current state, the system carries multiple partial truths simultaneously.

These patterns are not unique to firms that have adopted AI aggressively. They appear wherever speed and volume have increased without a corresponding change in how work is connected across the firm. Activity increases, but the effort required to move that work into a usable, aligned, and trusted state remains substantial. From the perspective of the firm, this can create the appearance of progress without a corresponding shift in outcomes.

2. The underlying constraint: Coherence

Most senior leaders in large firms already know where the friction lives. The issue is the absence of a reliable mechanism for bringing systems together into a consistent, usable understanding of the work. Firms have the systems and the data. What they lack is alignment.

Documents reside in document management platforms. Financial data lives in billing and ERP systems. Communication is dispersed across email, messaging platforms, and informal channels. Knowledge sits in precedents, playbooks, and personal archives. Each system captures a valid representation of the matter.

The state of the matter, in most firms, is not maintained. It is reconstructed. To understand what is actually happening in a matter at a given moment, participants assemble context across systems (pulling documents, reviewing communications, cross-checking financials, reconciling assumptions) before they can act. This reconstruction is not occasional. It is continuous, embedded in the normal flow of work.

A firm can have full access to all relevant information and still lack a consistent understanding of the matter. While access improves retrieval, coherence determines whether work can move without reconstruction.

How fragmentation shapes the operating model

In practice, each participant builds a working view of the matter based on the systems they access, the documents they review, the communications they are part of, and the assumptions they carry forward. These views are often aligned in direction, but they are rarely identical in detail. As work progresses, they diverge. Alignment is re-established through coordination (meetings, message threads, document reviews) which creates a shared understanding that quickly fades as new work is created and changes propagate unevenly.

In transactional practices, the legal, financial, and commercial dimensions of a deal evolve across parallel systems. Maintaining alignment requires continuous reconciliation: confirming that documents reflect the latest negotiated position, checking that financial assumptions remain consistent with legal terms, and ensuring that precedent fits the current deal structure. In litigation, the pattern appears differently but produces the same effect. The current posture of the case must be reconstructed from filings, discovery platforms, strategic discussions, and client communications that are not structurally integrated.

The consequences are consistent in that a significant portion of effort is spent locating information, confirming its relevance, reconciling differences, and re-establishing context before any substantive action is taken. Handoffs take longer than expected. Review cycles expand. Reporting requires reconciliation. Decisions are deferred pending alignment.

Why fragmentation has become a governing constraint

Large firms have operated with fragmented context for years, with human expertise historically compensating for the gaps. Skilled professionals carried the state of the matter mentally, reconstructed it across systems, and absorbed the cost as part of delivering the work.

What has changed is the pace. As speed, volume, and frequency of legal work increase, fragmentation is no longer a background inefficiency that can be absorbed through effort. It becomes the primary limiter on performance. The firm's ability to generate work begins to exceed its ability to maintain a shared understanding of that work. Context fragmentation now matters at a strategic level. The cost of compensating for it is scaling faster than the capacity to absorb it.

3. AI as an accelerant of structural pressure

As output accelerates, the mechanisms that firms rely on to maintain coherence are required to operate more frequently, across more artifacts, and with less time to reconcile differences.

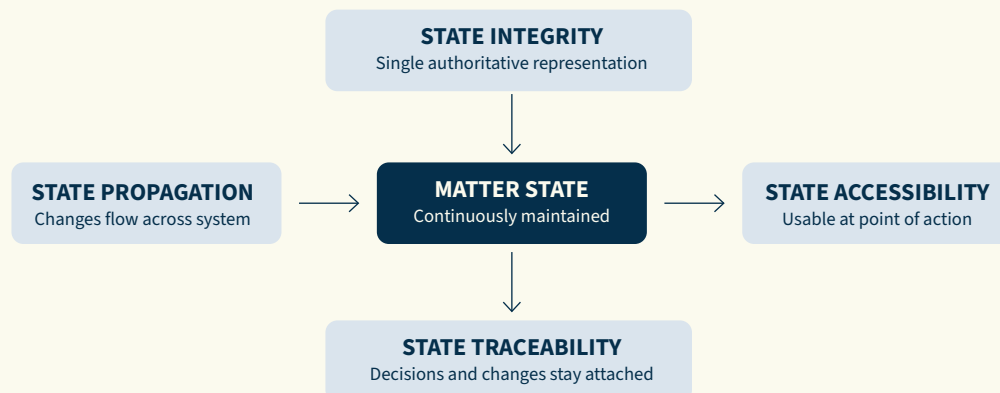
The gap between “produced” and “ready to act on” becomes more pronounced. Drafts arrive earlier but require extended review. Analyses are generated quickly but must be checked against underlying sources. Outputs need to be reconciled with parallel work occurring in other systems. At the same time, the volume of output increases: more drafts, more iterations, more intermediate artifacts, more points where alignment is required.

Decision-making does not become easier under these conditions. More available output introduces more potential interpretations. Establishing the correct context for a decision requires additional effort to confirm what is current, what has changed, and what remains valid. Teams spend more time reconciling differences between systems, clarifying assumptions, and confirming the current state of the matter before acting.

Recognizing the threshold

The shift in coherence becomes visible when certain conditions begin to appear consistently across matters:

- WORK IS PRODUCED FASTER THAN IT CAN BE REVIEWED AND ALIGNED.
- THE TIME BETWEEN DRAFT COMPLETION AND USABLE OUTPUT IS EXPANDING RATHER THAN CONTRACTING.
- DECISION-MAKING SLOWS AS MORE OUTPUT BECOMES AVAILABLE.
- DIFFERENT SYSTEMS REFLECT DIFFERENT VERSIONS OF THE SAME MATTER.
- PARTICIPANTS REGULARLY CONFIRM WHAT IS CURRENT BEFORE PROCEEDING.



None of these conditions, in isolation, indicate failure. Taken together, they point to an environment that is compensating for a lack of persistent state. The firm can generate work faster than it can maintain a single, trusted version of that work.

Why incremental improvement stops compounding

Under lower-speed conditions, firms compensate for structural gaps through experience and effort. Professionals carry the state of the matter mentally and absorb the cost of reconstruction as part of delivering the work. As speed and volume increase, that compensation becomes less effective. There is less time to reconcile differences, more outputs requiring validation, and more decisions being made against partially aligned context.

This is the point at which additional tools, more automation, and better access to information stop producing proportional returns. Each of these can increase output further, but none of them change the underlying requirement to align, validate, and integrate that work before it can move forward. As long as coherence depends on reconstruction, the cost of maintaining it scales with the volume of work produced.

Beyond a certain threshold, performance is determined less by how quickly work can be generated and more by whether the firm can maintain a continuous, shared understanding of that work as it evolves.

At a certain threshold, the conversation shifts from faster production to the structural conditions that determine whether faster production actually translates into better outcomes. The question becomes: *What does it take to maintain coherence as a property of the environment rather than a product of individual effort?*

4. Context as infrastructure

The response to a coherence constraint has to be architectural: a change in what systems maintain as work moves through them, not just a better tool layered on top.

Context, in this framing, is the maintained state of the matter. It is not limited to documents or background information. It includes the relationships between documents, the sequence of decisions, financial position, assumptions,

dependencies, and obligations: what has happened, what is happening, and what must happen next. In most firms, this state is not maintained as part of the environment. It is reconstructed as needed, by the people closest to the work.

The required shift forces law firms to treat context as infrastructure: something the operating environment preserves and carries forward, instead of something that is assembled at the point of use.

What a context-enabled environment looks like

A maintained state exhibits four key properties:

1. **State integrity** means a single, authoritative representation of the matter exists and is continuously updated as work progresses.
2. **State propagation** means changes to the matter are reflected across systems without requiring manual reconciliation.
3. **State traceability** means decisions, assumptions, and changes remain attached to the work, so the path from input to output can be understood without reconstruction.
4. **State accessibility** means the current state of the matter is directly usable at the point of action, without participants needing to assemble it before proceeding.

These are properties of system behavior, not interface design. They describe how the environment carries meaning forward, as opposed to how it presents information on screen.

The complexity of legal data

Legal work cannot be represented as a single dataset. It is a composition of distinct data types that exist across different systems and evolve at different rates: unstructured work product such as contracts, pleadings, and emails; semi-structured matter data including status, phases, staffing, and timelines; financial and commercial data spanning time entries, budgets, realization, and billing arrangements; communication and decision data from emails, calls, internal discussions, and client direction; and knowledge and precedent from templates, prior matters, and playbooks.

Each of these captures a dimension of the matter. None of them, independently, represent the state of the matter. The state emerges from how these elements relate to one another over time. When those relationships are not maintained, context

must be reconstructed. That is why even firms with strong individual systems can still struggle with coherence at the matter level.

What changes when context is maintained

In a context-enabled environment, the matter exists as a continuously updated state. Documents, financials, communications, discovery, and decisions contribute to and operate against that state. Changes are reflected as they occur, and relationships between elements are preserved.

Participants operate within that state rather than assembling it. AI operates with the same structure: inputs reflect the current state of the matter, outputs align with active assumptions and constraints, decisions are grounded in consistent context, and traceability is preserved without reconstruction. Work advances through continuity rather than repeated reassembly.

This is also where the data architecture requirement becomes clear. Bringing data into a single location does not establish context if relationships are not preserved and state is not maintained. What is required is a shift from storing artifacts to maintaining stateful, relational models of the matter: models that preserve relationships between entities, maintain temporal state, distinguish between drafts and authoritative outputs, and link decisions to the artifacts and assumptions they affect.

*Context is not an enhancement layered onto AI.
It is the condition that determines whether AI
can operate reliably within legal workflows.*

5. From fragmented systems to a coherent operating model

Large law firms are not going to consolidate into a single system. They operate across platforms that carry governance, permissions, and audit requirements that cannot be displaced without introducing risk. Document management systems, financial systems, matter systems, communication platforms, and knowledge repositories will continue to coexist.

The question is whether the overall environment can behave coherently across touchpoints, and whether the state of the matter can persist as work moves rather than resetting at each boundary.

Stored data versus maintained state

Most firms already have access to the information required to understand a matter. What many lack is a mechanism to maintain that understanding as the matter changes. Data remains distributed, but the relationships that give it meaning are not preserved in a durable way.

This is the key distinction between stored data and maintained state. Stored data can be retrieved. Maintained state can be acted on. A maintained state reflects the current position of the matter, incorporates changes as they occur, and preserves the relationships between legal, financial, and operational elements. It allows participants to operate within a shared understanding rather than having to reconstruct one.

Constructing context without displacing system authority

In a multisystem environment, context must be constructed as a governed representation of the matter that draws from existing systems while respecting their authority. This involves normalizing entities across systems, establishing relationships between them, and maintaining those relationships over time as work progresses.

In practice, this means documents, clauses, parties, financial elements, tasks, and decisions are represented in a consistent structure even though they originate from different systems. Changes are tracked as part of the state itself, and rules of authority determine how conflicts between sources are resolved. The result is not a duplicate of underlying data, but a model of the matter that preserves meaning across system boundaries.

How does the environment shift affect AI

In a fragmented environment, AI operates on artifacts. It retrieves documents, summarizes communications, and analyzes inputs independently. Each interaction depends on how well context has been assembled at that moment. Outputs reflect the completeness and quality of that assembly.

In a maintained environment, AI operates on a structured representation of the matter. It can evaluate work in relation to the broader state, identify changes

as they occur, and generate outputs that align with current assumptions, dependencies, and constraints. Less effort is spent assembling context. More effort is applied to judgment, decision-making, and execution.

Why integration alone is not sufficient

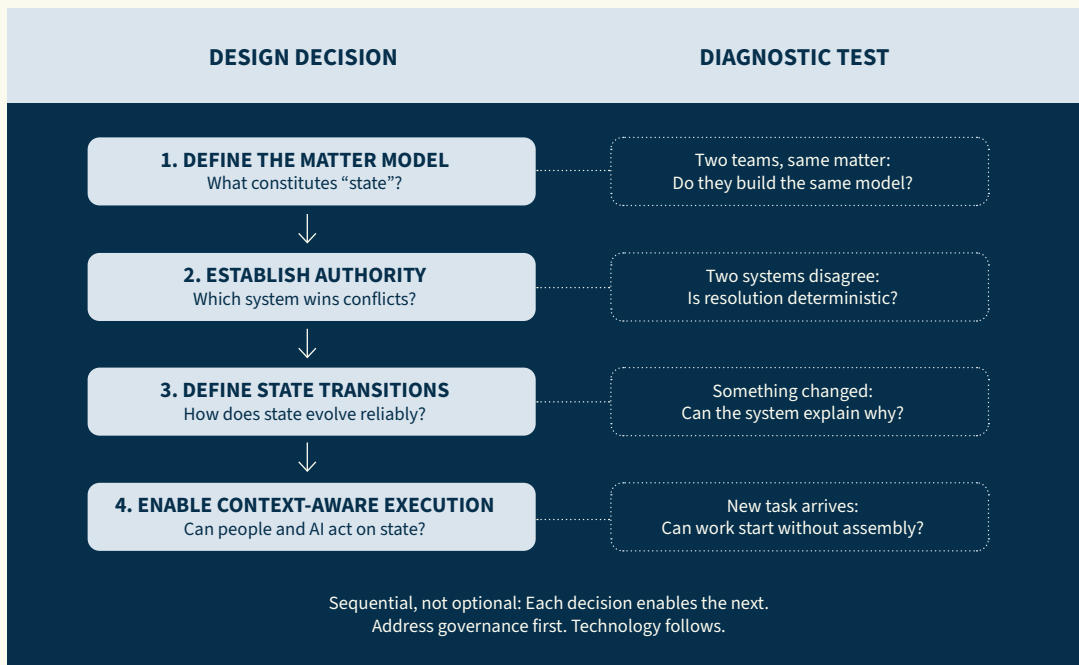
Many firms have already invested in integration layers, unified search, and data aggregation. These efforts improve access and reduce obvious inconsistencies, but they do not establish coherence. Integration enables data exchange. It does not preserve interpretation. Without a maintained representation of the matter, each interaction still requires context to be re-established, and alignment continues to depend on human effort. A coherent operating model requires that interpretation itself is carried forward as part of the system, not recreated at each step.

The harder challenge: Governance

The technical components of this shift are increasingly well understood. The more difficult challenge is governance. Maintaining a shared representation of the matter requires explicit decisions about authority, precedence, and accountability: which systems are authoritative under different conditions, how conflicts are resolved, and how changes are validated before they become part of the durable state. Without this clarity, attempts to construct context will either introduce inconsistency or revert to reconstruction at the point of use.

Building the operating model: Four design decisions

For firms evaluating how to move from fragmented context to preserved, continuous state, the challenge is not identifying systems. It is defining how the matter is represented across all of them. This requires four design decisions, each of which carries both technical and governance implications.



1. DEFINE THE MATTER MODEL

The firm must establish what constitutes the “state” of a matter, not as a system schema, but as a working model of how legal work is structured. This includes the core entities (documents, clauses, parties, obligations, tasks, financial elements), how those entities relate to one another, and which elements are critical to understanding the current position. Without a defined matter model, context cannot be constructed consistently. Each use case rebuilds its own version.

DIAGNOSTIC TEST: IF TWO TEAMS RECONSTRUCT THE SAME MATTER INDEPENDENTLY, DO THEY PRODUCE THE SAME STRUCTURE?

2. ESTABLISH AUTHORITY AND PRECEDENCE

In a multisystem environment, the same concept can exist in multiple places. A deal term may appear in a draft, an email, and a financial model. Matter status may differ between a system record and actual execution. The firm must define which system is authoritative for which elements, how conflicts are resolved, and when derived context can override source data. This is as much a governance decision as a technical one.

DIAGNOSTIC TEST: WHEN TWO SYSTEMS DISAGREE, IS RESOLUTION DETERMINISTIC OR DEPENDENT ON INDIVIDUAL INTERPRETATION?

3. DEFINE STATE TRANSITION LOGIC

A maintained state is not static. It evolves as work progresses. The firm must define what constitutes a meaningful change to the matter, how changes are validated before being incorporated, and how historical state is preserved alongside the current state. This is what allows context to be persistent rather than reassembled, and it is where many early attempts at integration fall short, as they focus on synchronizing data rather than governing how state changes.

DIAGNOSTIC TEST: CAN THE SYSTEM EXPLAIN WHAT CHANGED IN THE MATTER, AND WHY, WITHOUT SOMEONE RECONSTRUCTING IT MANUALLY?

4. ENABLE CONTEXT-AWARE EXECUTION

Once the state is defined and maintained, it must be usable. This requires exposing the state in a way that humans can operate within directly, AI can act against without reconstructing context, and workflows can reference as a source of truth. This is where the value becomes operational, and where the investment in maintaining state translates into a different experience of work, for both people and systems.

These four decisions are the foundation of the operating model. They are not implementation details to be resolved later. They are the choices that determine whether the environment will carry context forward or continue to depend on reconstruction. Firms that address them explicitly tend to find that the technical path becomes clearer once the structural and governance questions have been answered.

DIAGNOSTIC TEST: CAN WORK BE ADVANCED WITHOUT FIRST ASSEMBLING CONTEXT FROM MULTIPLE SYSTEMS?

6. What changes: Capacity, clarity, and controlled scale

When coherence is maintained by the environment rather than reconstructed through effort, the effects are not incremental. They change how the firm converts effort into progress, how decisions are made, and how the system behaves as volume increases.

Capacity

In many firms, a significant proportion of effort is not applied to advancing the matter. It is applied to maintaining a usable understanding of it: locating information across systems, confirming whether it is current, reconciling differences between sources, and re-establishing context before acting. This effort stabilizes the work but does not move it forward and does not scale.

When the state of the matter is maintained, that stabilization effort is reduced. Participants operate within a shared representation of the work. The current state is available, prior decisions remain attached, and dependencies are visible without reconstruction. The effect is a reallocation of effort: more time applied to substantive legal work, less time spent holding the environment together. Junior lawyers and AI operate within a coherent structure rather than assembling one, which increases the usability of their output. Senior lawyers spend less time validating context and more time exercising judgment.

Capacity expands when the system stops consuming effort to hold itself together.

Decision clarity

Legal decision-making is rarely constrained by a lack of information. More often it's about the effort required to establish what is actually current and reliable. Where context must be reconstructed, decisions are preceded by interpretation. Participants assemble a view of the matter, often from partially aligned sources, before they can act. This introduces delay and creates variability in how different people understand the same situation.

When the state of the matter is maintained, decisions are made against a current, shared representation of the work. Legal, financial, and operational signals exist within the same frame. Dependencies and prior decisions are visible as part of the state, not inferred from separate systems. Effort shifts from determining what is true to determining what should be done. Decisions occur earlier, with fewer dependencies on clarification, and with greater confidence in their downstream impact.

Governance and trust

Legal work depends on accuracy, traceability, and defensibility. In many environments, these are enforced retrospectively: Work is produced, then reviewed, validated, and corrected. Establishing trust requires reconstructing how an output was generated and whether it can be relied upon. As output increases, that approach becomes progressively more expensive.

When governance is embedded in how work is maintained, trust is established as part of execution rather than applied after the fact. Relationships between inputs, decisions, and outputs are preserved within the state. Changes are recorded with their rationale. Dependencies remain visible over time. Verification still occurs but no longer requires rebuilding the conditions for trust from scratch.

This is also what allows AI to operate within a governed environment. Outputs exist as part of a traceable system of work, not as isolated artifacts that must be independently validated.

Controlled scale

Fragmented environments scale through coordination. As volume increases, coordination effort expands alongside it. More work produces more versions, more dependencies, and more divergence. It requires alignment to be re-established more frequently. Inconsistencies persist longer. Risk accumulates across systems. Scale, in this model, introduces instability.

When coherence is maintained, scale behaves differently. Work remains connected as it moves. Context persists across participants, phases, and systems. As volume increases, coordination effort does not expand at the same rate. The system absorbs additional work without proportional degradation in clarity or control. Scale becomes a function of system design rather than coordination capacity.

AI increases output. Context determines whether that output moves. The distinction between firms that scale coordination and firms that scale coherence is where separation is now emerging.

7. Assessing context maturity

The distinction between fragmented and coherent environments is not theoretical. It is visible in how work behaves as it moves through the firm, and it can be observed without reviewing architecture diagrams or cataloguing systems.

Start with a live matter and follow it across participants, systems, and stages of work. The question is not what tools are in place, but whether the state of the matter persists as work advances, or whether it must be reconstructed each time someone needs to act.

Where to start

1. ORIENTATION

When a lawyer joins or returns to a matter, how quickly can they establish a reliable understanding of its current position? In many firms, this still requires pulling together documents, communications, and prior decisions to form a working view. Where that synthesis is necessary, the state of the matter is not directly available.

2. TRANSITIONS

When work moves between individuals, teams, or phases, context either carries forward or resets. If the receiving party must re-establish what has already been decided, why it was decided, and what has changed since, then continuity is not being maintained by the environment. Rework often concentrates at these moments, because the underlying state had to be rebuilt before progress could continue.

3. POST-PRODUCTION

The constraint in many firms is no longer how quickly drafts or analyses can be generated. It is how much additional effort is required before they can be used. Where outputs consistently require clarification, validation, or alignment before they can move forward, the system is compensating for a lack of continuity.

4. ALIGNMENT UNDER PRESSURE

As volume increases, environments that depend on reconstruction become more dependent on it. Coordination expands, visibility becomes harder to maintain, and small inconsistencies take longer to resolve. Environments in which the state of the matter persists tend to remain more stable as activity increases, because work can continue without proportional increases in alignment effort.

Four stages of context maturity

Taken together, the signals of context maturity tend to cluster around four observable conditions. The stages below are not a certification model or a scoring rubric. They describe how the environment behaves, and where the most consequential shifts occur.



Where most firms operate today

THE CRITICAL TRANSITION
Connected to contextual

Before this point: AI gains remain local, and effort redistributes.

After this point: Returns compound, and work carries forward.

The divide is not who adopted AI first, but who built the environment for it.

STAGE	HOW THE ENVIRONMENT BEHAVES	WHAT MOVING FORWARD REQUIRES
Fragmented	The state of the matter is assembled manually across systems, conversations, and individual knowledge. Progress depends heavily on reconstruction at each step.	Identify the highest-friction handoffs and reconstruction points. Map which matter elements are reconstructed most frequently and where the most time is lost.
Connected	Access to information improves and some inconsistencies are reduced through point-to-point integrations. The state of the matter is still re-established at critical points in the workflow.	Define what “matter state” means for the firm’s primary work types. Establish authority rules for conflicting data. Begin building the matter model described earlier.
Contextual	A shared representation of the matter begins to persist across systems, handoffs, and stages of work. Participants can operate from a more consistent view without rebuilding it each time.	Extend state maintenance to cover transitions, governance checkpoints, and AI-assisted workflows. Define state transition logic so that context evolves reliably as work progresses.
Operationalized	The state of the matter is maintained as part of the environment itself. Work, decisions, and execution carry forward without depending on repeated reconstruction.	Extend the model to new practice areas and matter types. Context compounds into institutional learning, enabling the firm to improve how it operates, not just what it produces.

Most large firms operate somewhere between fragmented and connected conditions, with pockets of more advanced behavior in specific workflows or teams. That is why many AI efforts produce visible local gains without changing how work moves across the broader matter lifecycle. As a result, access improves, output increases, and the system still resets at key points.

Where the consequential shift happens

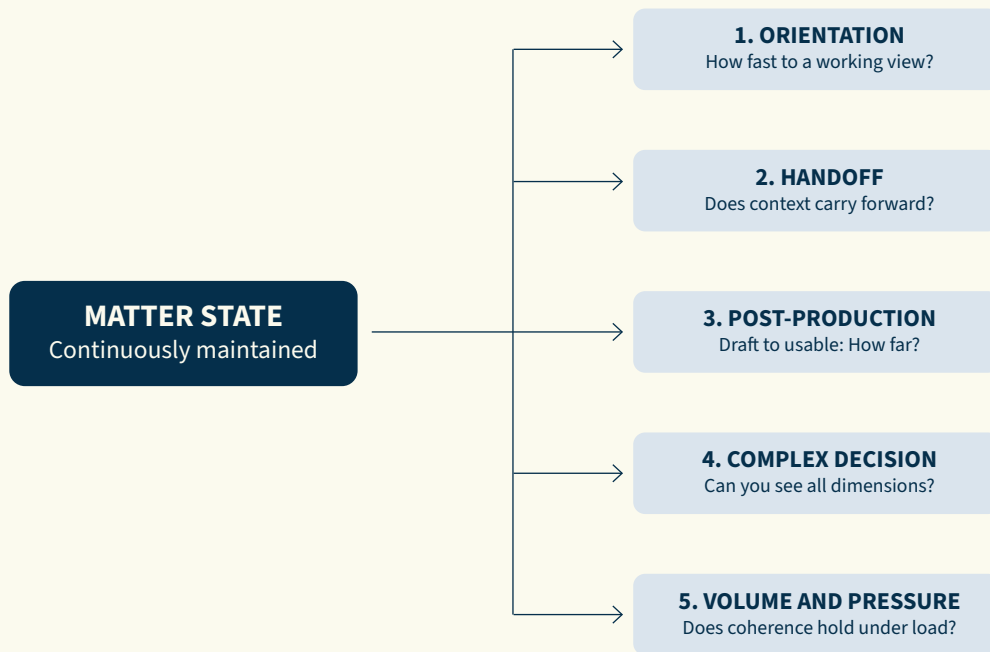
The important transition is not just from fragmented to connected. Many firms are already making progress there, and integration investments are delivering real value. The more consequential shift is from connected to contextual: the point at which the state of the matter begins to persist rather than being reassembled. That is where the environment starts to behave differently, and where the returns on AI investment begin to compound rather than plateau.

Applying context maturity to your environment

Maturity frameworks are useful for orientation, but they rarely change how a firm thinks about its own environment. What does change thinking is observation: watching how a specific matter actually moves through the firm and noticing where context persists and where it has to be rebuilt.

The exercise below is designed to be conducted against a single live matter. Choose one that is active, involves multiple participants, and has moved through at least one transition (a handoff between teams, a shift in phase, or a change in staffing). Follow the matter through five moments, each chosen because it reveals something specific about how context behaves in the environment.

At each moment, the question is the same: Did the participant operate from a state that was carried forward, or did they have to reconstruct one before they could act?



AT EACH MOMENT, ONE QUESTION:
Maintained, or reconstruction?

The pattern reveals where context persists and where it breaks down.
Repeat across matter types for a fuller picture.

1. ORIENTATION

A lawyer joins or returns to the matter after time away: a new team member staffed midstream, or a partner returning after several weeks focused elsewhere. Before they can contribute, they need to understand the current position. What has happened, what has changed, what decisions are active, and what is expected next.

ASK: HOW LONG DOES IT TAKE THEM TO REACH A RELIABLE WORKING UNDERSTANDING? DO THEY GET IT FROM THE ENVIRONMENT, OR DO THEY ASSEMBLE IT FROM CONVERSATIONS, DOCUMENTS, AND EMAILS?

2. HANDOFF

Work moves from one person or team to another: a draft passes from associate to partner for review, a matter transitions from origination into active execution, or a workstream shifts between offices. This is the point where context either carries forward or resets.

ASK: DOES THE RECEIVING PARTY PICK UP FROM WHERE THE PREVIOUS PARTICIPANT LEFT OFF, OR DO THEY SPEND MEANINGFUL TIME RE-ESTABLISHING WHAT HAS ALREADY BEEN DECIDED AND WHY?

3. POST-PRODUCTION

An AI-generated or associate-produced draft, analysis, or summary has been completed. It exists as an artifact. Before it can be relied upon, it needs to be evaluated against the current state of the matter: checked for alignment with negotiated positions, financial assumptions, strategic direction, and prior decisions.

ASK: HOW MUCH EFFORT IS REQUIRED BETWEEN “DRAFT COMPLETE” AND “READY TO USE”? IS THAT EFFORT SPENT ON SUBSTANTIVE JUDGMENT, OR ON LOCATING AND CONFIRMING THE CONTEXT NEEDED TO EXERCISE JUDGMENT?

4. DECISION UNDER COMPLEXITY

A decision needs to be made that draws on more than one dimension of the matter: legal, financial, commercial, or procedural. A pricing decision that depends on current exposure. A strategic call that requires understanding both the legal position and the client relationship. A resource allocation that depends on where multiple matters stand simultaneously.

ASK: CAN THE DECISION-MAKER SEE THE RELEVANT DIMENSIONS WITHIN A SHARED FRAME, OR DO THEY NEED TO PULL INFORMATION FROM SEPARATE SYSTEMS AND RECONCILE IT BEFORE THEY CAN ACT?

5. VOLUME AND PRESSURE

Activity on the matter increases: a compressed timeline, parallel workstreams, or a surge in output from multiple contributors. This is where the environment’s ability to maintain coherence is tested. Under normal conditions, reconstruction may be manageable. Under pressure, the cost of it becomes visible.

ASK: AS ACTIVITY INCREASES, DOES COORDINATION EFFORT SCALE WITH IT (MORE CHECK-INS, MORE RECONCILIATION, MORE TIME SPENT CONFIRMING WHAT IS CURRENT) OR DOES THE ENVIRONMENT ABSORB THE ADDITIONAL LOAD WITHOUT PROPORTIONAL DEGRADATION?

Reading the pattern

While the context maturity exercise does not produce a score, it produces a pattern. Across the five moments, the question is how consistently the environment carries context forward versus how frequently participants have to rebuild it.

Where reconstruction dominates (where most moments require the participant to assemble context before they can act) the environment is operating in the fragmented-to-connected range. Gains from AI and other tools will tend to remain localized, because each step still depends on human effort to re-establish the conditions for work to move.

Where maintained state begins to appear (where participants can operate from a shared, current understanding without rebuilding it) the environment is approaching contextual maturity. This is where returns on capability investment begin to compound because work carries forward rather than resetting.

DOMINANT PATTERN	WHAT IT INDICATES	WHERE TO FOCUS NEXT
Reconstruction at most moments	Context is not maintained by the environment. Coherence depends on individual effort and coordination. AI and automation gains are likely being redistributed rather than compounding.	Start with the matter model. Define what constitutes the state of the matter for your highest-volume work types, and identify where the most frequent reconstruction occurs.
Mixed: Some moments carry forward, others reset	Pockets of maintained state exist, often within specific teams or workflows, but context still resets at key boundaries. Integration has improved access without establishing coherence.	Focus on authority and state transitions. Define which systems are authoritative, how conflicts are resolved, and how changes are validated. This is typically the governance gap that prevents connected environments from becoming contextual ones.
Maintained state at most moments	The environment is beginning to carry context forward. Participants operate from a shared understanding more often than they rebuild one. The conditions for compounding returns are forming.	Extend the model. Apply it to additional practice areas, matter types, and AI-assisted workflows. Define how context-aware execution is exposed to both people and systems. This is where institutional advantage begins to accumulate.

This exercise is most useful when conducted by someone close enough to the work to observe these moments directly, and senior enough to act on what they find. It is also more revealing when repeated across different matter types and practice areas. Spotting variation here often matters as much as spotting the overall pattern.

Where external support matters

Firms that approach the problem as a sequence of tools, pilots, or point solutions often improve parts of the workflow without changing how work moves overall. What is required instead is support at the level of structure: how the matter is represented, how changes are carried forward, how authority is defined across systems, and how workflows are shaped around persistent context.

That distinction, between vendors who deliver functionality and partners who help the firm define and build its operating architecture, is increasingly what separates firms making incremental gains from firms making structural ones. Clio's enterprise practice is built around this kind of engagement: working with firms to define use cases against real infrastructure, navigate the data and governance implications, and support the transition from isolated improvements to durable operating change.

The compounding divide

Over the next decade, every serious firm will deploy AI in some form. Access to capable tools will not be the dividing line. The divide will emerge between firms whose individual parts move faster and firms that become structurally better at converting capability into outcomes.

As capability increases, the difference between these two positions compounds. Where work resets at each step, increased capability produces more activity that must be aligned, validated, and reassembled. Where work carries forward, the same capability produces progress that builds on itself.

The firms that recognize this now, and invest in the operating architecture that makes it possible, will be better positioned than those still compensating for fragmentation through effort alone. They will have built the environment in which AI, and the people using it, can operate at their best.

 **Clio** *for Enterprise*