

Building Better Professional Services with AI

A discussion about knowledge, judgment and the design of firms whose product is expertise.

AI-first is not a claim that machines should lead. It is a commitment to decide, task by task, how institutional knowledge, machine capability and human judgment can produce the most rigorous work.

I / THE STARTING POINT

The hesitation is reasonable.

Professional firms are not built by producing words quickly. Their reputations are built by knowing what the evidence can support, when a client's stated ambition exceeds its operating capacity, and when a technically elegant recommendation will fail in practice.

For a firm of that kind, skepticism about artificial intelligence is not a failure of imagination. It is a professional instinct. A report, opinion, model or design may be the visible artifact, but the value of the work lies elsewhere: in the quality of the inquiry, the interpretation of imperfect evidence, the candour of the advice and the willingness of named people to stand behind the result.

Early generative systems gave experienced professionals good reasons to be wary. They produced confident prose without understanding why a claim mattered. They could invent a source, flatten a disagreement or turn an uncertain inference into a polished conclusion. The time saved in drafting could be lost again in checking. To people whose credibility depends on the distinction between what is known, what is inferred and what remains unresolved, fluency without responsibility was never an adequate bargain.

The deeper concern is therefore not whether AI can make a plausible professional work product. It plainly can. The concern is whether the use of such systems weakens the habits on which serious work depends: attention to context, respect for evidence, protection of confidential information, intellectual independence and accountability to the client. There is also a legitimate fear that convenience could erode craft—that junior professionals might stop learning how an argument is assembled, or that a firm's voice might be replaced by a synthetic competence that sounds acceptable and means little.

Any serious case for AI in professional services has to begin by conceding those risks. Starting with a catalogue of features or a promise of productivity would misunderstand the objection.

The question is not how to make professional firms more enthusiastic about technology. It is whether a technology that has changed materially can now strengthen the disciplines on which those firms depend. Each firm has to answer that question on the terms set by its work, its clients and its obligations.

II / THE PROFESSIONAL COMPACT

What professional judgment protects.

The knowledge in an advisory firm is not reducible to the documents on its server. Much of it resides in distinctions that experienced people have learned to make.

A credible professional knows that two sources of equal apparent authority may not deserve equal weight. A client's account can be accurate as testimony and misleading as evidence. A precedent can be technically relevant and practically inapplicable because the jurisdiction, institution, incentives or risk tolerance differ. A rigorous method does not remove judgment; it makes the use of judgment more visible.

This is why professional concern about AI is ultimately a concern about epistemology: how the firm knows what it claims to know. A system can retrieve a precedent without understanding why the precedent was persuasive. It can summarize a methodology without recognizing where that methodology ceased to be appropriate. It can reproduce the language of confidence while being unable to accept responsibility for the consequences.

Over years of practice, a firm accumulates more than finished reports. It accumulates methods, interview practices, source standards, client histories, failed approaches, proposal evidence and tacit knowledge about how decisions are actually made. If AI is introduced as a generic writing layer, detached from that body of knowledge, it will make the firm sound more fluent while making its work less particular. That would be a genuine loss.

The ethical boundary is equally important. Confidentiality, consent, intellectual property and client expectations cannot be delegated to a model. Nor can authorship. A person must remain able to explain why a consequential claim was included, what evidence supports it, what uncertainty remains and who approved its release. The presence of a human in the process is not sufficient; there must be meaningful human ownership of the judgment.

A firm protects its intellectual integrity by ensuring that every consequential claim has a source, a method and an accountable author—not by insisting that a person manually perform every step.

That distinction matters. It preserves what is essential in professional work without treating every inherited task as sacred. Research assembly, document comparison, retrieval, calculation and routine monitoring are means. Judgment and accountability are the professional obligations those means must serve.

III / THE OBLIGATION TO LOOK AGAIN

Intellectual honesty has to run in both directions.

Respect for professional judgment cannot become an assumption that a human being is the better instrument for every intellectual task simply because the human being is human.

People forget. We search unevenly, privilege recent experience and stop looking when a plausible answer appears. Our recall of earlier projects depends on who happens to be present. Under pressure, we reuse familiar sources, carry forward old language and make inconsistent choices about work that appears routine. Expertise reduces these failures; it does not abolish them.

A well-designed system may be able to search a firm's approved archive more broadly than any individual can remember it, compare more documents without fatigue and apply a specified check with greater consistency. It may notice that a current proposal resembles work completed twelve years ago, or identify a contradiction across evidence that no one person had reason to hold in mind. None of this gives the system wisdom. It does mean that human authorship is not, by itself, proof of superior care.

There is an ethical dimension to this counterargument. If scarce senior attention is consumed by retrieval, formatting and repeated assembly, less of it is available for interpretation, challenge and conversation with the client. If institutional memory remains dependent on individual recollection, the firm exposes clients to avoidable inconsistency. If recurring business-development or knowledge work disappears whenever delivery becomes busy, that too is a design choice with consequences.

The relevant comparison is therefore not between an ideal professional and a reckless machine. It is between actual ways of working, each with its own failure modes. AI can hallucinate, overstate and execute a poor instruction at speed. Humans can overlook, tire, rationalize and allow important work to stop. A serious operating model has to make both sets of weaknesses visible.

The role of professional judgment becomes more exacting: to specify the question, establish the evidentiary standard, interrogate the result, decide what matters and own the consequence.

For each part of the firm's work, what combination of institutional knowledge, machine capability and human judgment produces the most rigorous result?

IV / FROM PRINCIPLE TO INQUIRY

The technology moved. The standard did not.

Many experienced professionals tried generative AI in 2024 or 2025 and found it unreliable, cumbersome or intellectually mediocre. That judgment was often sound. It is no longer sufficient evidence about what is possible now.

The important change is not simply that models write better sentences. Current systems can be given a bounded objective, approved context, access to defined tools and a sequence of work to complete. They can search across files, compare evidence, calculate, edit documents, monitor a recurring process and return an artifact with sources and exceptions for review. Persistent context also makes it possible to connect the work to a firm's own methods and history rather than begin each exchange from a blank chat.

These capabilities do not make the systems reliable by default. They change the unit that must be evaluated. The old test was often whether a prompt produced an impressive answer. The relevant test now is whether a specified, controlled workflow produces better work after the cost of supervision, verification and correction is included.

That is why the familiar objections—confidentiality, hallucination, intellectual property, client expectations, quality control and over-reliance—belong inside the inquiry rather than outside it. Each should determine where information may go, what the system may do, what evidence it must return and where a person must intervene. The counter-risk is also real. Informal use may spread without common safeguards; competitors may combine equivalent judgment with broader research and faster execution; and the firm's accumulated knowledge may remain difficult to retrieve or reuse.

To be AI-first, in this context, does not mean asking a machine to do everything first. The phrase names a discipline of work design. Before scarce human time is assigned, the firm asks what a system could do independently, what it could prepare alongside a person and what must remain a wholly human act of interpretation, relationship or accountability.

THE PROPOSITION, CAREFULLY STATED

Begin with the work and the quality bar. Decide the right combination of institutional knowledge, machine capability and human judgment. Select tools only after those choices are clear.

This is not yet a recommendation to transform the firm. It is a case for a disciplined assessment: to identify where the newer capabilities are genuinely useful, where they remain unfit and what controls would be required before any client-facing use.

01

Professional AI use is already widespread; firm discipline has not caught up.

The strongest evidence is not that every competitor has solved AI. It is that informal use is becoming the default, leaving firms with the exposure but little of the operating benefit.

74%

of professionals in Thomson Reuters' 2026 survey use AI several times a week.

34%

use AI tools their organization has not sanctioned. That is a governance gap, not a non-use outcome.

THE CLIENT EXPECTATION GAP

Important		78%
Delivered		6%

Corporate clients rating AI-enabled quality as important or essential, versus those saying most or all providers deliver it.

Source: Thomson Reuters Institute, *Future of Professionals Report 2026*. Survey of 1,816 professionals across 62 countries in law, tax, audit, accounting, compliance, risk and global trade. It is a broad professional benchmark, not a universal proxy for every advisory or consulting practice.

WHAT THIS PROVES

AI is already inside reputation-sensitive professional work. Strategy and sanctioned tools affect how it is used.

WHAT IT DOES NOT PROVE

The survey does not establish a return for any particular firm, show that automated advice is superior, or justify moving client work into a system without contract, privacy and source controls.

The status quo also makes a choice

Professional firms lose value when opportunity monitoring, knowledge capture, client follow-up and communications are treated as campaigns. Delivery becomes busy, the campaign stops, and the same discussion restarts months later.

AI is relevant because some of that work can become a maintained system. It can keep searching, sorting and preparing while people preserve their attention for client judgment and relationships.

THE COMPETITIVE CONSEQUENCE

A competitor does not need better judgment to gain an advantage. Equivalent judgment combined with broader research, faster proposal assembly and more reliable reuse of past work may be enough.

02

Intellectual honesty has to run in both directions.

A fluent model is not an expert. A human professional is not automatically the better source simply because the professional is human.



For each part of the work, the question is which combination of firm knowledge, machine capability and human judgment produces the most defensible result.

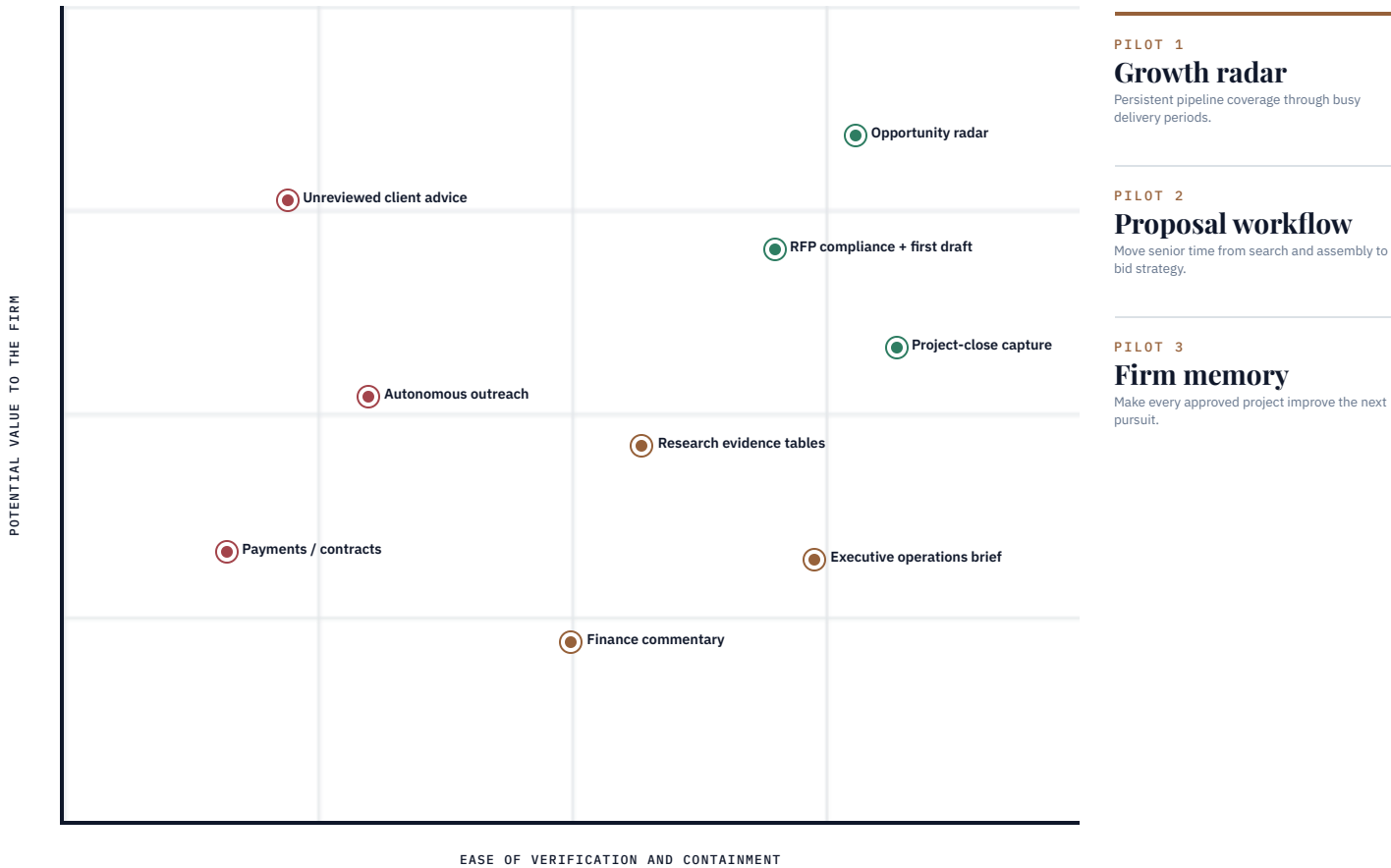
MACHINE FAILURE MODES	HUMAN FAILURE MODES
Hallucination, weak causal reasoning, overconfidence, hidden assumptions, prompt injection, permission errors and rapid execution of a bad instruction.	Forgetfulness, uneven search coverage, fatigue, recency bias, inconsistent process and the tendency for recurring work to stop when projects become busy.

THE KNOWLEDGE ADVANTAGE	A controlled system may retrieve and contextualize decades of work more consistently than a person relying on memory. That is not a reason to trust it blindly. It is a reason to stop making individual recollection the only index to the firm.
-------------------------	---

03

Start with high-value work whose output can be checked against a known source.

The best first workflow is rarely the most dramatic. It repeats often enough to learn, contains a visible quality signal and fails safely before anything leaves the firm.



Illustrative portfolio for expertise-led professional firms. Each firm must confirm actual frequency, systems, permissions, review cost and quality sensitivity before selecting pilots.

04 Pipeline coverage should continue when delivery becomes busy.

The system prepares the field. Leadership still decides what to pursue and makes the relationship move.

01 Scan approved sources MDBs, DFIs, donors, governments, foundations, partners and selected procurement portals.	02 Normalize and deduplicate Resolve repeats, deadlines, eligibility rules, geography, funding source and buyer.	03 Match firm evidence Capability, assignments, staff credentials, client history, partners and likely gaps.	04 Rank and brief Produce a short list with fit rationale, missing evidence, deadline and recommended action.	05 Human gate Leadership selects pursuits, sets the relationship path and authorizes outreach.
--	---	---	--	---

THE USEFUL OUTPUT A five-to-ten-item weekly brief Short enough to support a decision. Detailed enough to avoid reopening the entire search.	Opportunity Buyer, instrument, scope, geography and deadline.		Fit Why the firm is relevant, with source-linked evidence.	Eligibility Mandatory conditions and any unresolved question.
	Evidence gap Missing project, credential, country presence or partner.		Relationship path Known client, institution, consultant or partner route.	Next action Pursue, monitor, partner, decline or seek clarification.

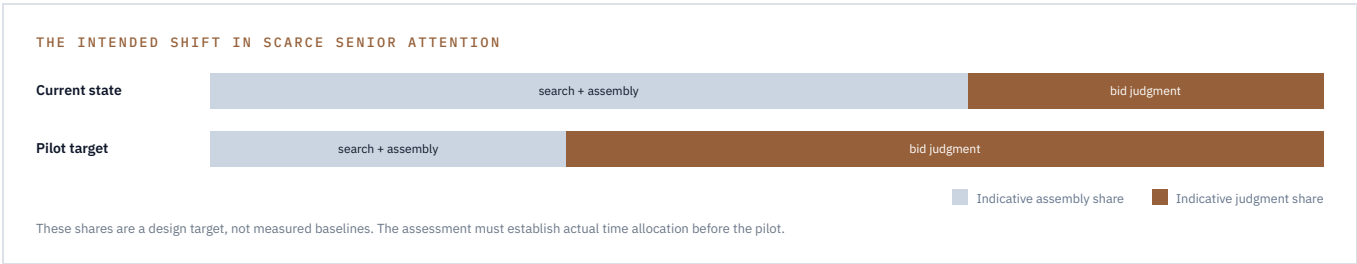
VALUE MEASURE Useful opportunities surfaced per week and staff time per selected pursuit.	QUALITY MEASURE False positives, missed eligibility conditions, stale deadlines and missing duplicates.	STOP CONDITION The brief creates more checking than it removes or does not change pursuit decisions.
---	---	--

05

Senior time should move from proposal assembly to the argument that wins the work.

A useful system does not produce generic prose. It builds the compliance architecture, retrieves approved evidence and makes every gap visible.

01 Parse the full RFP Mandatory conditions, scoring logic, deliverables, format rules, dates, work packages and questions.	02 Build compliance matrix Every requirement linked to a proposed section, evidence source, owner and open issue.	03 Retrieve approved evidence Projects, methodologies, bios and boilerplate from a controlled library.	04 Draft with lineage Complete architecture and first-pass sections with source links and explicit gaps.	05 Human gate Senior lead owns positioning, partners, interpretation, pricing and final narrative.
--	---	--	--	--

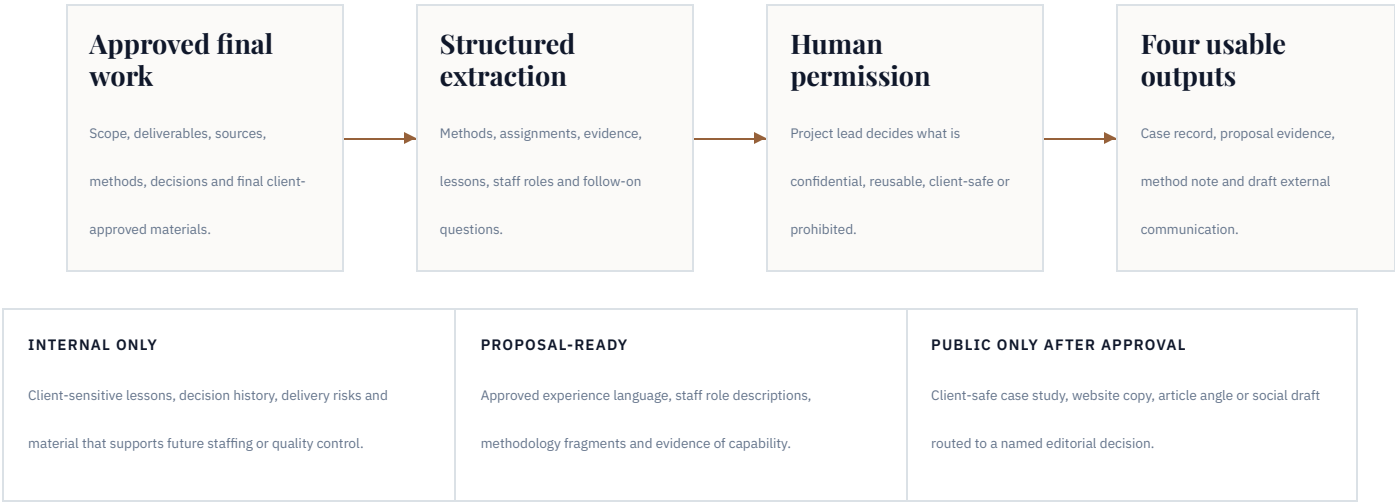


VALUE MEASURE Elapsed time to a compliant first draft and senior hours spent locating evidence.	QUALITY MEASURE Missed requirements, unsupported reuse, source failures and material senior rewrites.	STOP CONDITION The system creates fluent copy that cannot be traced to approved firm evidence.
---	---	--

06

Every approved project should improve the next proposal and the firm's public evidence.

The archive becomes valuable when project-close decisions determine what may be retained, reused and published.



The value is cumulative

The first project record may save little time. The operating advantage appears when the next pursuit can retrieve relevant assignments, approved evidence and delivery lessons without asking who remembers them.

The risk is architectural

Firm memory should be permissioned, source-linked and exportable. A firm should not bury its intellectual property so deeply inside one vendor that it cannot reconstruct or move the corpus.

VALUE MEASURE Project-close capture rate, time to retrieve experience and later proposal reuse.	QUALITY MEASURE Incorrect permissions, missing provenance, duplicate records and stale approved language.	STOP CONDITION The workflow cannot separate client-confidential material from reusable firm knowledge.
---	---	--

07

Clear operating lanes are more useful than a long AI policy.

Staff need to know what information may be used, where it may be used, how far a system may act and who approves the result.

OPEN	Typical material Public information, formatting, brainstorming, generic market scans and low-sensitivity internal work.	Environment Approved firm workspace. No client-confidential or sensitive personal data.	Operating rule AI may run with normal professional review.
CONTROLLED	Typical material Client work where contract permits, proposals, research, internal finance and structured firm knowledge.	Environment Enterprise workspace, explicit access, minimum necessary data, source lineage and logs.	Operating rule Human review before the output drives a decision or leaves the firm.
APPROVAL-GATED	Typical material Final advice, public release, payments, contracts, hiring, sensitive personal data and irreversible actions.	Environment Tool use only where specifically approved, with least privilege and a checkpoint before execution.	Operating rule AI may prepare or recommend. A named human authorizes the action.

One workspace

Named owner and approved enterprise terms.

Data rule

Public, internal, client-confidential and highly sensitive.

Least privilege

Each workflow sees only what it needs.

Source lineage

Claims and calculations can be reconstructed.

Known tests

Important workflows run against expected results.

Portability

Firm memory can be exported and re-routed.

RULE OF THUMB

Autonomy rises with reversibility and verifiability.

The ability to detect and undo a bad result should determine how much the system may do. Excitement about a model is not an operating control.

The lane structure translates the bounded-autonomy and accountability principles in Singapore's 2026 Model AI Governance Framework for Agentic AI into a practical firm operating rule.

08

Diagnose the work, not the firm.

A maturity score would create false precision. The useful output is a practical answer to where AI adds value, what controls it requires and what the firm should test first.

01 Leadership and risk Strategy, growth bottlenecks, client promises, risk tolerance, non-negotiables and the work leadership most wants to stop carrying.	02 Workflow census Real artifacts, recurring tasks, triggers, hand-offs, wait time, rework, decisions and what disappears when delivery gets busy.	03 Systems and knowledge Email, files, CRM, finance, project tools, archives, public sources, client restrictions and the reliability of available context.	04 Opportunity portfolio Value, frequency, feasibility, data sensitivity, reversibility, review cost and the human judgment required for each workflow.	05 Pilot specifications Inputs, permissions, tools, steps, expected output, checkpoint, exception path, acceptance test, owner and measures.
--	--	--	--	---

THE DELIVERABLE A short operating blueprint	■ Working principles and red lines ■ Three-to-five pilot specifications ■ 90-day implementation plan	■ Prioritized workflow portfolio ■ Initial tool, data and permission architecture ■ Measurement and ownership cadence
--	--	---

TIMING

Two to three weeks

Long enough to inspect real work and short enough to avoid turning an operating question into an IT transformation programme.

DESIGN CONSTRAINT

Usable without the assessor

The firm should own the specifications, evidence, permissions and measures. The blueprint should not require a consultant to remain in the loop.

09

Build enough structure to learn quickly.

Governance comes before access to sensitive data or external action. A firm does not need a perfect policy before testing low-risk work.

DAYS 1-30 Foundation Choose one primary enterprise workspace and a controlled secondary model. Set client-use and data rules. Prepare the initial knowledge corpus. Baseline the selected workflows. Train staff on specification, verification and escalation. Exit: clear permission to use AI well and a measurable baseline.	DAYS 31-60 Pilots Run three to five workflows with weekly review. Track failures, review time and exceptions. Revise source rules, prompts, steps and checkpoints from observed evidence. Do not add automation to a workflow whose output is still unstable. Exit: evidence about where AI adds net value in the firm's work.	DAYS 61-90 Integrate Standardize the workflows that show value. Add schedules and connectors only after quality is stable. Assign owners and keep a change log. Retire low-value experiments. Set the next-quarter expansion decision. Exit: a small operating system, not a collection of demonstrations.
---	--	---

Net capacity Time saved minus review, correction and rework.	Cycle time Elapsed time through the full workflow.	Quality Errors, missed requirements and material rewrites.	Coverage Opportunities and knowledge that no longer disappear.	Risk Unsanctioned use, exceptions, near misses and corrections.
---	---	---	---	--

By quarter end, the test is whether staff trust a few governed workflows because they have survived real client and firm work.

10

The assessment should produce a decision, not another AI strategy document.

Continue only if the pilots show net value under the firm's actual conditions after review, correction and risk controls.

CONTINUE AND EXPAND

Evidence supports the operating thesis

- Staff spend less net time without material quality loss.
- Outputs are traceable to approved sources and reusable firm evidence.
- The workflow persists through busy project periods.
- Human gates catch exceptions without becoming the entire process.
- Users trust the result because the acceptance test is visible.

CHANGE OR STOP

The workflow does not earn its place

- Review and repair consume the time supposedly saved.
- Source lineage is unreliable or permissions remain ambiguous.
- The system produces generic work that weakens the proposal or analysis.
- Quality depends on one enthusiastic user rather than a repeatable method.
- The activity was not valuable enough to automate in the first place.

- 1 Where does scarce senior attention currently limit growth, delivery quality or continuity?
- 2 Which recurring activities matter but repeatedly stop when client work becomes intense?
- 3 What client, confidentiality or intellectual-property boundaries are non-negotiable?
- 4 Which parts of the firm's archive are reliable, reusable and permitted to enter a shared knowledge layer?
- 5 What evidence would make a 90-day pilot worth continuing, and what failure would end it?

NOTES

Sources and evidence base

External evidence is used to establish the operating context, not to claim a predetermined return for any firm.

[1] **Thomson Reuters Institute.** *Future of Professionals Report 2026*, June 2026. Survey of 1,816 professionals across 62 countries.
thomsonreuters.com/en/institute/future-of-professionals-2026/report

[3] **Microsoft.** *2026 Work Trend Index: Agents, human agency, and the opportunity for every organization*, 5 May 2026. Survey of 20,000 AI-using knowledge workers and Microsoft 365 telemetry.
microsoft.com/en-us/worklab/work-trend-index/...

[5] **OpenAI.** *Codex for every role, tool, and workflow*, 2 June 2026. Provider evidence for connected tools, artifact creation and non-developer use.
openai.com/index/codex-for-every-role-tool-workflow/

[7] **OpenAI.** *How HSP GRUPPE builds AI capabilities for tax advisory*, 7 August 2026. Vendor customer story used only as a directional small-firm deployment example.
openai.com/index/hsp-gruppe/

[2] **Thomson Reuters Institute.** *2026 AI in Professional Services Report*, February 2026. More than 1,500 professionals in adjacent reputation-sensitive fields.
thomsonreuters.com/.../2026-ai-in-professional-services-report.pdf

[4] **IMDA Singapore.** *Model AI Governance Framework for Agentic AI*, version 1.0, 22 January 2026. Used for bounded autonomy, accountability and lifecycle controls.
imda.gov.sg/.../mgf-for-agentic-ai.pdf

[6] **OpenAI.** *Enterprise Signals: What frontier firms are doing differently*, updated 12 August 2026. Aggregated provider usage data; token volume is an imperfect proxy for business value.
openai.com/signals/enterprise-data/

SCOPE AND METHOD

The argument and workflow examples are intended for professional firms whose value depends on specialist knowledge, evidence, discretion and accountable judgment. The examples are hypotheses, not a universal operating model. Each firm must validate them against its work, contracts, data, client expectations and risk tolerance through staff interviews, artifact review and workflow observation.

EVIDENCE HIERARCHY

Firm work, contracts and approved methods first; current public and institutional sources second; adjacent professional benchmarks third; product-provider telemetry and case studies labeled as such.

DESIGN AUTHORITY

Doyen Collective's current website system and *The Capture Gap*: Playfair Display, Spectral, IBM Plex, Oxford blue, muted copper, square geometry and restrained information graphics.

ABOUT DOYEN COLLECTIVE

Prepared for people who make consequential economic decisions.

Doyen Collective produces evidence-led analysis, decision tools and advisory work for the institutions and professionals who shape economic growth. The work is direct, sourced and built to survive a reader who has done the job for 20 years.