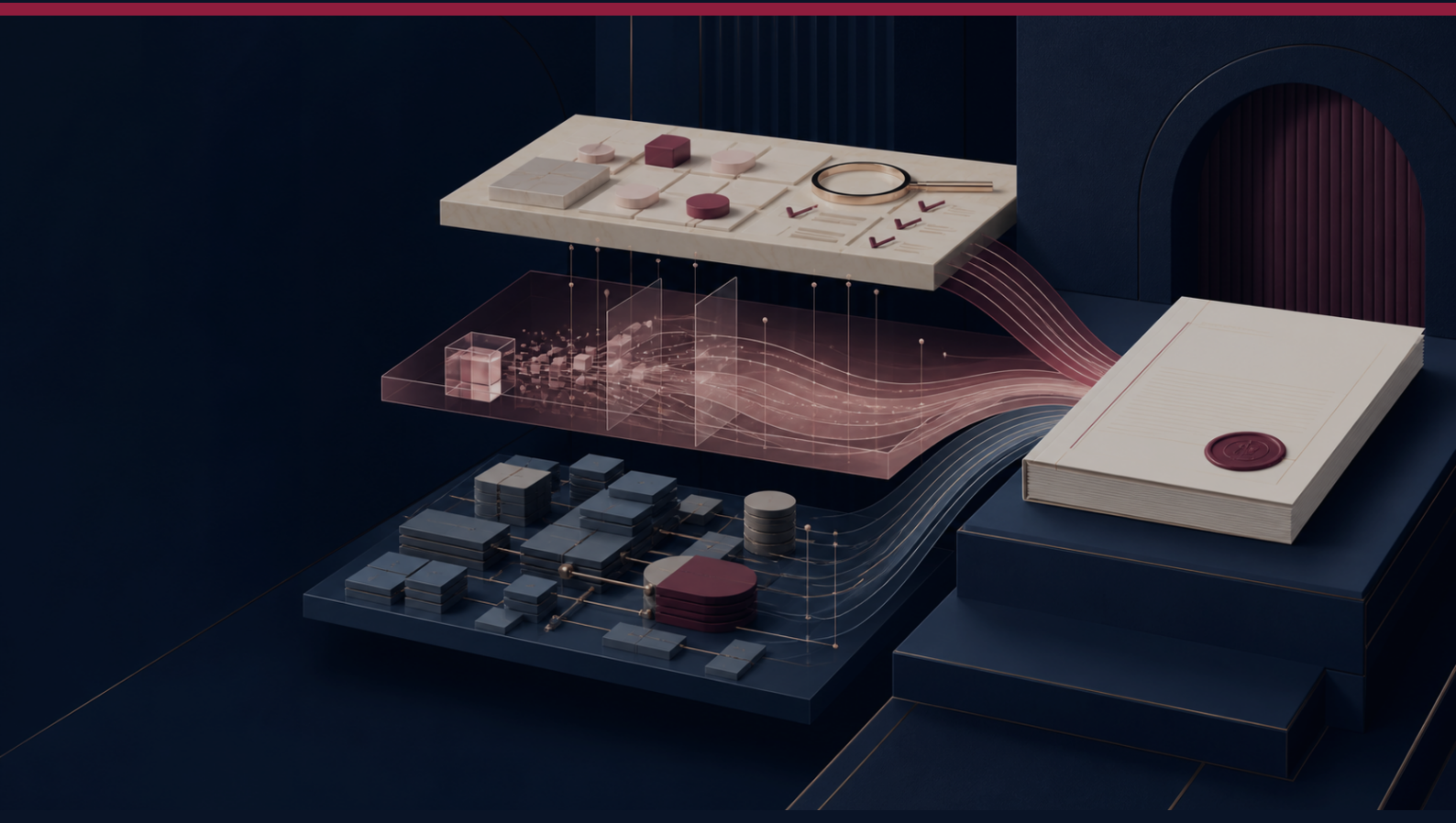
Managed Services

INSIGHT SERIES / WHITE PAPER

Beyond Automation: A New Model for Professional Services

Why the next step is not another tool, but a controlled combination of software, AI and human judgement that delivers finished professional outputs.



MANAGED SERVICES LIMITED

AUGUST 2026 | managed-group.co.uk

EXECUTIVE BRIEF

Automation is a component, not the outcome

Thesis

The familiar choice between doing work internally and commissioning a traditional consultancy is becoming less complete. A third model is emerging: a lean, output-led delivery service that orchestrates approved technology, repeatable methods and specialist human judgement around a defined result.

This is not a claim that AI can replace professional accountability. The evidence shows meaningful gains in some tasks and uneven or adverse results in others. The commercial and operating opportunity is therefore to design the combination—to route work according to risk and capability, then assure the output as a whole.

THREE TAKEAWAYS

1. Start with acceptance. Define the user, decision, evidence threshold and usable format before selecting technology. A fast draft that cannot pass review is unfinished inventory.

2. Engineer human judgement. Decide where a competent person frames, challenges, verifies, interprets and signs off. The phrase ‘human in the loop’ is too vague unless the loop has a named purpose and authority.

3. Contract for outputs and controls. Specify deliverables, service levels, source handling, quality tests, exceptions and knowledge transfer. This shifts attention from tool access or days supplied to work accepted.

PROBLEM FRAMING

The last mile is where value is won or lost

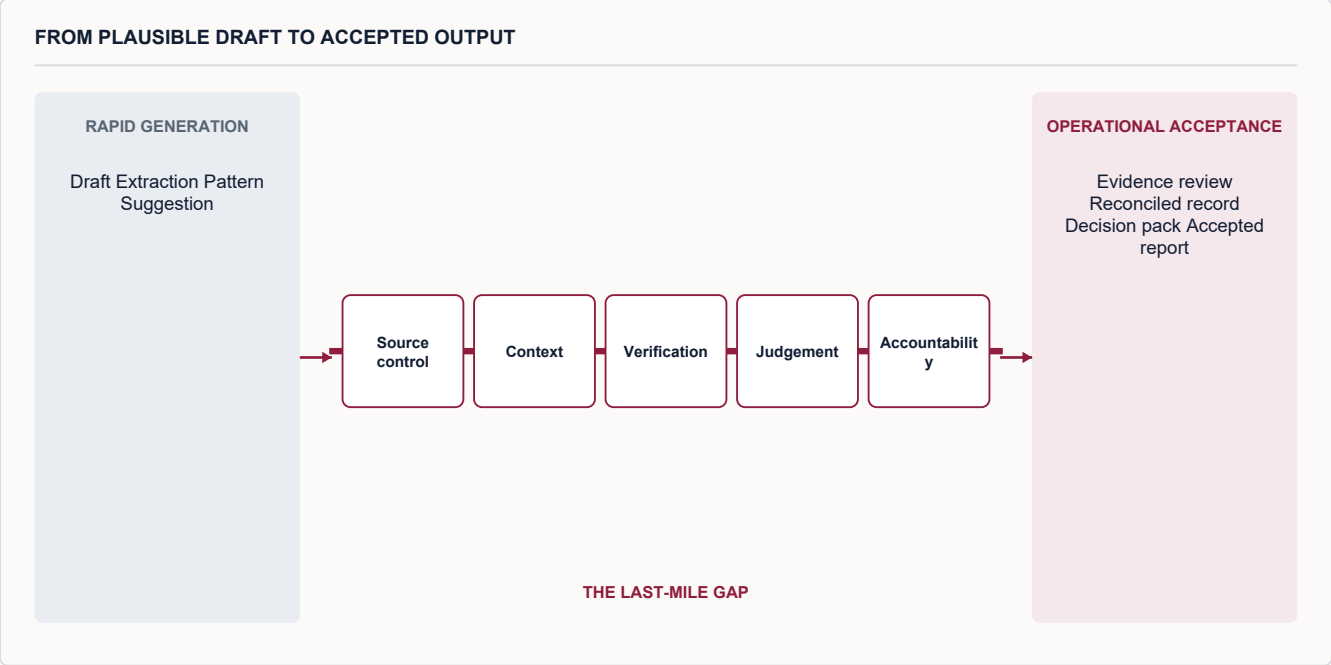


Figure 1. A gap separates rapid generation from operational acceptance. A bridge made of source control, context, verification, judgement and accountability crosses it.

Why tool-led adoption stalls

A tool can make an individual step faster while leaving the surrounding service unchanged. Someone must still decide which sources are authoritative, resolve contradictions, check completeness, adapt the result to the organisation's context and take responsibility for use. If those activities are not redesigned, the saved production time can reappear as review, rework or uncertainty. Appetite is not the constraint: in Microsoft's 2023 global survey, 76% of respondents said they would be comfortable using AI for administrative tasks, 79% for analytical work and 73% for creative work. These are attitudes, not measured benefits.[5]

This is the last-mile problem of knowledge work. A generated summary is not yet an evidence synthesis. Extracted fields are not yet a reconciled register. A dashboard is not yet a management judgement. An action list is not yet programme control. Professional value lies in crossing the distance between plausible material and an output that meets an explicit purpose.

Three traps

The speed trap treats cycle time as the only result. The substitution trap assumes an entire role can be automated because some constituent tasks can be. The confidence trap mistakes fluent presentation for reliable evidence. Each is avoided by managing the work at task level while assuring it at output level.

There is also a sourcing trap: buying access to named people without defining the product of their work. Time-based support remains useful when demand cannot be bounded. But where an output can be specified, an output-led service creates a clearer relationship between fee, responsibility and acceptance.

A more useful question

Instead of asking, ‘Can AI do this job?’, ask: ‘What sequence of tasks produces the output; which tasks are within the technology's demonstrated capability; what can go wrong; and where must a person exercise judgement?’ That question turns adoption from a software experiment into delivery design.

WHAT THE EVIDENCE SAYS

The opportunity is real — and uneven

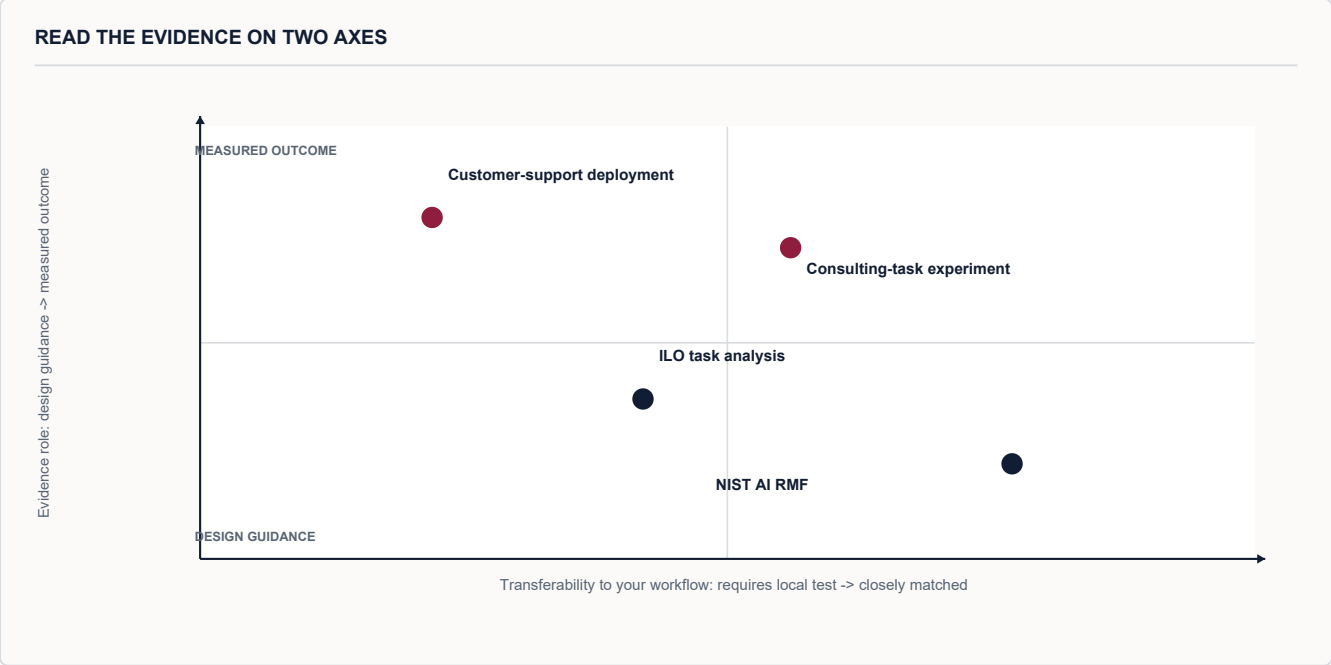


Figure 2. A matrix plots strength of measured gain against transferability to a new setting. Customer support and experimental consulting show measured gains but require local validation; ILO and NIST guide system design rather than predicting savings.

Measured gains in bounded settings

An NBER study examined the staggered introduction of a generative AI assistant among 5,179 customer-support agents. Access increased issues resolved per hour by 14% on average, with larger gains for novice and lower-skilled workers and minimal effects for the most experienced and highest-skilled workers. The result is strong evidence in this setting, not a transferable forecast for research or consulting work.[1]

A field experiment involving 758 consultants tested work resembling aspects of consulting practice. For tasks assessed as inside the model's capability frontier, access to GPT-4 increased task completion by 12.2%, speed by 25.1% and human-rated quality by more than 40%. The same research also demonstrated the ‘jagged frontier’: capability varies between apparently similar tasks, and using AI for a task outside that frontier reduced the likelihood of a correct solution. Configuration and task selection matter.[2]

Augmentation is not automatic

The International Labour Organization's global task analysis concluded that generative AI was more likely to augment jobs than fully automate them, because most occupations were only partly exposed. It also warned that outcomes depend on how adoption is managed, including effects on job quality, autonomy and work intensity.[3]

NIST's voluntary AI Risk Management Framework provides a useful governance discipline: Govern, Map, Measure and Manage. It treats risk management as continuous and socio-technical, not as a test applied only at launch. For professional services, that means controls must cover the brief, data, model, people, process and use of the result.[4]

CENTREPIECE FRAMEWORK

The Managed Intelligence Loop

Figure 3. Six stages form a loop: Define, Ground, Orchestrate, Challenge, Assure and Deliver. Human decision gates sit at scope, exceptions and release. Feedback connects acceptance back to definition.

Six controls around one outcome

The model begins and ends with the intended use. Define names the output, user, decision, constraints and acceptance criteria. Ground assembles authorised evidence, records provenance and limits the material available to tools and people. Orchestrate decomposes the work and routes each task to software, AI or a specialist according to capability and risk.

Challenge is the deliberate search for error: contradictory sources, omissions, weak reasoning, bias, implausible values and unintended disclosure. Assure applies proportionate checks and records unresolved exceptions. Deliver converts the working material into the agreed format, secures acceptance and captures feedback for the next cycle.

Human involvement changes across the loop. At Define, people establish intent. At Ground, they approve sources and access. At Orchestrate, they set task boundaries and intervene on exceptions. At Challenge, a suitably competent reviewer tests rather than merely reads. At Assure, an accountable owner decides whether the evidence is sufficient. At Deliver, the client owns the decision made with the output.

Technology also has more than one role. Deterministic automation moves and transforms data where rules are stable. Search and retrieval narrow the evidence set. Generative AI supports classification, synthesis and drafting where variability is useful. Monitoring records performance and drift. None of these roles, by itself, owns the conclusion.

OPERATING MODEL

From software access to output accountability

THREE MODELS, THREE UNITS OF MANAGEMENT

UNIT OF MANAGEMENT	Software access	Time-based support	Managed output
WHAT IS BOUGHT	Capability	Expert capacity	Defined result + controls
WHO INTEGRATES	Client user	Usually client lead	Delivery provider, within client governance
SUCCESS MEASURE	Use and task-level benefit	Effort, milestones and deliverables	Acceptance, quality, flow and value

Figure 4. Three columns compare software, time-based professional services and managed professional outputs by what is bought, who integrates the work and how success is measured.

The tool-led model

In a tool-led model, licences are distributed and individuals decide how to use them. Benefits are often described through adoption, prompts or anecdotes. Quality depends on local practice; evidence trails are inconsistent; difficult cases quietly revert to manual work. This can be an effective space for learning, but it is not yet a dependable service.

Traditional professional services solve a different part of the problem by supplying expertise and governance. They can, however, be too broad or expensive for bounded operational outputs, and time-based teams can leave the client managing utilisation, hand-offs and the last mile.

The output-led model

An output-led managed service sits between these approaches. The client provides purpose, approved access, decision authority and acceptance. The provider designs and runs the workflow, assembles proportionate capability, operates controls and supplies the evidence trail. Technology is selected within that system rather than sold as the system.

The statement of work should define units and tolerances: what counts as an item or output; required fields and formats; source hierarchy; turnaround bands; first-time acceptance target; sampling or full-review rules; material exception categories; security and retention requirements; and the route for change. Where uncertainty prevents fixed output pricing, use a discovery or pilot phase before agreeing the service baseline.

Good measures cover four dimensions. Flow: lead time, queue age and on-time delivery. Quality: first-time acceptance, defect severity and traceability. Control: exceptions, access breaches and completion of required reviews. Value: user adoption, decision timeliness and the stated use of released capacity. Volume alone is not evidence of value.

PRACTICAL ACTION

A 90-day route to a controlled service

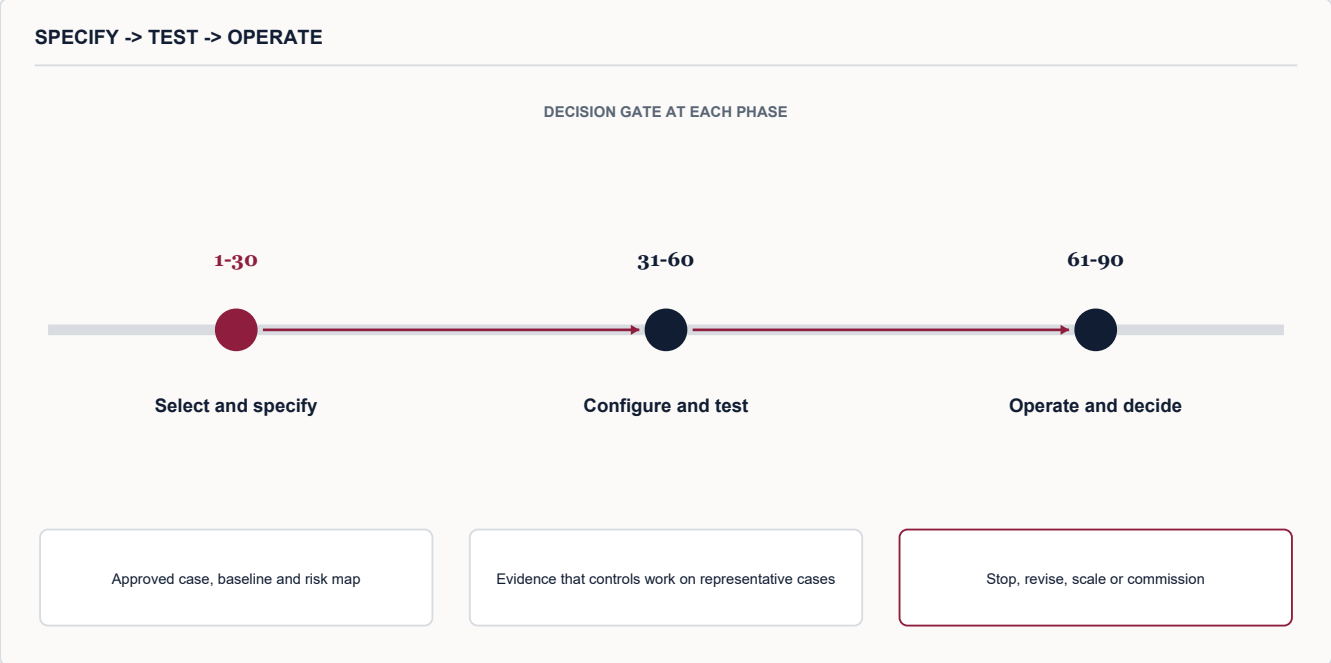


Figure 5. A 90-day roadmap moves through selection and specification, controlled testing, and limited operation, with a decision gate at each phase.

Days 1–30 | Select and specify

Choose a bounded, repeatable output with a real backlog, delay or capacity constraint. Avoid the highest-impact decision for the first pilot. Map tasks and failure modes; identify personal, confidential or sensitive data; define authorised sources, acceptance criteria and the accountable owner. Establish a baseline for lead time, effort, quality and exceptions.

Classify each task: deterministic, assistive or judgement-led. Record why. This prevents enthusiasm for a particular tool from determining the workflow.

Days 31–60 | Configure and test

Build the minimum controlled workflow. Use representative test cases, including incomplete, contradictory and out-of-scope inputs. Compare technology-supported outputs against an agreed reference or expert review. Test access, retention, traceability, bias, security, accessibility and failure recovery as relevant. Define confidence or tolerance thresholds and the mandatory path for exceptions.

Run a shadow pilot before relying on the output operationally. Measure reviewer effort as well as generation time; otherwise a faster first step can hide a slower end-to-end process.

Days 61–90 | Operate and decide

Release a limited batch under active monitoring. Review first-time acceptance, material defects, turnaround, exception volume, reviewer load and user usefulness. Interview affected staff about work intensity and autonomy, not only speed. Update controls from observed failures.

At day 90, decide explicitly: stop because the case is weak; revise the configuration; scale with defined limits; or commission a managed service. Record the decision, residual risks, benefit owner and next review date.

CONCLUSION

The future of professional services is assembled around outcomes

A practical synthesis

Automation changes the economics of individual tasks. It does not remove the need to define the problem, understand institutional context, judge conflicting evidence or own the consequences of a recommendation. Those responsibilities become more important when plausible material can be produced cheaply and at volume.

The strongest professional-services model therefore combines three forms of capability. Software provides consistency and reach. AI provides flexible acceleration across selected cognitive tasks. People provide intent, contextual understanding, challenge, empathy and accountability. Workflow and governance turn those ingredients into a dependable service.

This model should be judged neither by the sophistication of its technology nor by the size of its team. It should be judged by whether it delivers the required output, within agreed controls, with an evidence trail the client can understand and use. The organisations that learn to commission at that level can gain leverage without surrendering judgement.

Questions for commissioners

What finished output do we need? Which evidence must support it? What is the consequence of error? Which tasks can be accelerated, and which demand judgement? Who can challenge and stop release? How will we know the result is useful—not simply fast?

WHAT TO REMEMBER

Buy the outcome. Inspect the controls. Keep judgement accountable.

SOURCES AND FURTHER READING

1. National Bureau of Economic Research (April 2023; revised November 2023). Generative AI at Work. Brynjolfsson, Li and Raymond studied a staggered deployment among 5,179 customer-support agents. Productivity was issues resolved per hour; effects varied substantially by worker experience and skill. Source link

2. Harvard Business School (September 2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Working Paper 24-013 by Dell'Acqua and colleagues. Field experiment with 758 consultants; results depend on whether tasks fell inside or outside the assessed AI capability frontier. Source link

3. International Labour Organization (August 2023). Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality. Task-exposure analysis indicating that augmentation is more likely than full job automation across most occupations. Exposure is not the same as realised adoption or job loss. Source link

4. US National Institute of Standards and Technology (January 2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). Voluntary, rights-preserving and non-sector-specific framework organised around Govern, Map, Measure and Manage. It supports risk management; it does not certify an AI system or prescribe one control set for every use. Source link

5. Microsoft (9 May 2023). Will AI Fix Work? Work Trend Index Annual Report. Survey of 31,000 workers in 31 markets on attitudes to AI-supported work and skills. Source link

This paper is a general discussion of delivery design. Evidence is presented with its source and scope; illustrative frameworks are not universal benchmarks or client performance claims.