Operating Standard.

FOUNDING THESIS / 001

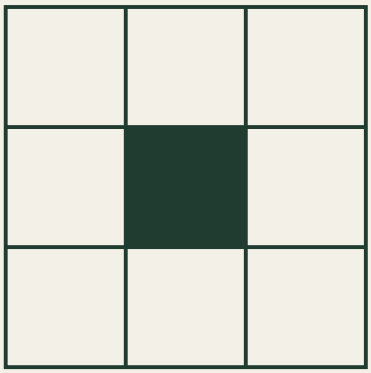
The Deployment of Intelligence.

Operating knowledge, reliable execution and the last mile of AI.

OPERATING KNOWLEDGE

CONTEXT + AUTHORITY

EVIDENCE + EXECUTION



VERSION 1.0 / 28 NOVEMBER 2026

OPSTD.CO

Working towards a better operating standard.

Our conviction

Artificial intelligence can perform extraordinary intellectual work. Yet an ordinary business process can still stall because an invoice is missing, two departments use different definitions, or nobody knows who is authorised to resolve an exception.

The contrast is revealing. In 2025, an advanced Gemini Deep Think system reached the gold-medal standard at the International Mathematical Olympiad, solving five of six problems. That is a specific achievement in mathematical reasoning; it does not establish dependable performance across business operations. [S1]

The image of an “Einstein in a box” captures the surprise of powerful reasoning becoming accessible. But intelligence still needs a place in the organisation: a purpose, relevant information, permission to act and a standard against which its work can be checked.

We believe the next opportunity is to help businesses build those conditions.

THESIS

Intelligence needs an operating standard

A capable model becomes useful when people can connect it to their work, define what good looks like and trust the process that carries an objective through to completion.

Operating Standard is being built around this conviction. We help the people who run businesses put AI to work, combining operating knowledge with the engineering needed for reliable execution.

1. The gap is between capability and dependable use

The case for deployment does not require a claim that AI has created no economic value. There is already evidence of gains in particular settings. A published study of 5,172 customer-support agents found an average 15% increase in issues resolved per hour, with substantial differences across workers. That result belongs to the studied setting; it is not a promise about every workflow. [S2]

The harder question is how improvements become repeatable across organisations, survive real exceptions and exceed their full delivery costs. A task can become faster while the wider process remains slow. More documents can be prepared while a reviewer becomes overloaded. Hours can be released without becoming lower expenditure or additional revenue.

Research on the productivity J-curve provides a useful foundation: new technologies require complementary investments in processes, skills and organisational capital, and measured productivity can lag those investments. It does not tell us when a particular AI deployment will pay back. [S3]

Our interest is in that practical work of conversion. We want to understand which processes can improve, what must change around the model and how to establish that the improvement is worth sustaining.

EVIDENCE

Keep the levels separate

Benchmark performance shows a capability. A workplace study shows a result in a setting. A deployment must demonstrate reliable completion. A business must establish the effect on quality, capacity, cost and cash. Evidence at one level does not automatically establish the next.

2. Operating knowledge gives intelligence a job

In a real organisation, the difficult part of a task is often knowing which interpretation applies.

A finance operator may know which source should prevail when records disagree. A procurement lead may recognise that a delivery is partial rather than late. An account manager may understand that a customer's familiar exception is acceptable only with a particular approval. These judgments shape what the system must do, when it must wait and when it must ask for help.

Some knowledge can be documented. Some needs examples. Some reveals a policy the business has never agreed. More context cannot settle a decision that still needs an owner.

This is why we place operating knowledge at the centre of deployment. The person closest to the work should help specify its inputs, exceptions, consequences and acceptance conditions. Engineering turns that specification into a maintained system with appropriate access, validation and recovery.

Recent research provides a related hypothesis. A 2026 working paper associates AI investment and productivity gains with the accumulation of firm-specific organisational capital. Its observational evidence supports investigating this connection; it does not establish the performance of our proposed delivery model. [S8]

Limestone's discussion of knowledge maintenance makes a related point: resolving conflicting business definitions remains an organisational responsibility, even when agents help maintain the documentation. [S4]

PRINCIPLE

Put the operator in the design loop

The operator helps define the outcome and its exceptions. The accountable owner approves policy and acceptance criteria. Engineers establish technical reliability. AI supports the work within those boundaries.

This is a division of responsibility, not a claim that domain expertise eliminates the need for engineering. We intend to make operating knowledge easier to express, test and improve, while making technical expertise more effective through closer contact with the business.

3. Give business people the means to shape AI

We see an opportunity to build from the knowledge already inside the organisation.

An operator should be able to explain a recurring problem, show a few accepted and rejected examples, identify the evidence required and describe the point at which someone must make a decision. Those inputs should become a working specification that both people and software can use.

The aim is to give business people practical control over the parts they understand. They should be able to review a proposed rule, correct an example, see what a system has done and decide whether the result meets the agreed standard. They should also be able to recognise the boundary of their own authority.

A useful toolset therefore needs more than a chat window. It needs a way to define a workflow, maintain approved knowledge, examine evidence, review exceptions, test changes and observe performance.

> DESIGN COMMITMENT
>
> ### Make expertise usable
>
> Help an operator turn “this is how we know the work is right” into examples, checks, decisions and a repeatable process. Make the resulting system understandable enough for the business to question and improve it.

Our ambition is an increase in organisational capability. The client should understand its process better after a deployment and retain access to the definitions, evidence and procedures on which that process depends.

4. Define a better operating standard

We use operating standard to mean a maintained agreement about how a particular piece of work should be completed.

It sets out the objective, required information, responsible people, permitted actions and evidence of completion. It explains what should happen when information is incomplete, sources disagree or the case exceeds the system’s authority. It also identifies who can change the standard and when a revision takes effect.

Part of the standard	Question it answers
Outcome	What should be completed, and for whom?
Context	Which records, definitions and policies apply?
Authority	What may the system do, and who decides the rest?
Evidence	What supports the proposed result?
Exceptions	What requires clarification, waiting or escalation?
Acceptance	Who confirms completion and on what basis?
Maintenance	Who owns changes and how are they checked?

The standard should be proportionate to the work. A simple reminder does not need the same controls as a financial adjustment. A mature deployment should be able to distinguish them.

A standard also needs room for legitimate variation. An exceptional decision should remain an exception until the responsible owner decides that it should become a rule. This protects the organisation from turning an isolated correction into a policy nobody intended.

5. The last mile is the whole path to completion

The phrase last-mile deployment can make the work sound small. In practice, it can include connecting systems, clarifying responsibilities, obtaining missing evidence, managing approvals, recovering from failure and supporting the process after launch.

We use the phrase to describe the distance between an available capability and an accepted result in a particular business.

Consider a missing supplier invoice. A model may be able to read an invoice accurately, but the invoice is not yet available. The useful system must identify the relevant receipt, find the responsible person, request the right document, assess the response and keep the case visible until it is resolved or escalated.

> ILLUSTRATIVE WORKFLOW
>
> ### An invoice received is not a case resolved
>
> Goods have arrived, but the supporting invoice is missing. The system links the purchase order and receipt, assigns an owner and prepares a request. When a document arrives, it checks identity, quantity, amount and reference. A mismatch stays open for review. A finance owner accepts the evidence or decides the next step. The record preserves the result and its basis.
>
> This is a proposed workflow, not a claim about a deployed client system. Accounting treatment and any funds movement remain separately authorised decisions.

Work can continue across several days and several people. Its state must survive a failed agent run, an unavailable application or a change of personnel. A reply is an event; completion is a judgment against the agreed standard.

That distinction is central to the systems we intend to build.

6. Combine reasoning with reliable execution

Different parts of a process need different tools. Models can help interpret messy documents, investigate ambiguity and draft contextual requests. Deterministic software can calculate amounts, compare identifiers and enforce explicit conditions. Workflow infrastructure can preserve state, handle waiting and record transitions. People retain decisions that require their authority and judgment.

Anthropic’s engineering guidance distinguishes predictable workflows from more open-ended agents and recommends using the simplest design that meets the task. That is a useful implementation discipline, rather than an argument for making every process autonomous. [S5]

We begin with existing systems and capabilities. New infrastructure must earn its place through a requirement that simpler alternatives cannot meet adequately.

ENGINEERING STANDARD

The workflow must outlast the agent

Preserve the objective, evidence, ownership and decision history independently of a particular model run. Detect incomplete work. Recover without duplicating actions. Make unresolved conditions visible to someone who can act.

A reliable process also needs an explicit way to stop. When the evidence is insufficient or the action is outside the permitted scope, the appropriate outcome may be a question, a hold or an escalation. Those outcomes must be tested alongside successful completion.

7. Learn from work without quietly changing the rules

The most useful learning often appears when an experienced person corrects a proposed result. But a correction can mean several things: a genuine error, missing information, a new circumstance, a preference or a policy change.

A system should retain enough evidence to distinguish those possibilities. Proposed improvements can then be tested before they change production behaviour.

OpenAI and Thrive Holdings describe a tax-preparation system developed with practitioners at Crete, using expert corrections, production traces and evaluations to guide engineering changes. It is a relevant company-reported implementation example, rather than evidence of OPSTD’s capabilities or of unattended accounting. [S6]

Our proposed learning loop is straightforward: record the outcome, understand the correction, propose a change, evaluate it on representative cases, obtain the appropriate approval and monitor the result. Client-specific rules remain client-specific unless there is a justified, authorised basis for generalisation.

KNOWLEDGE STANDARD

Experience should become a candidate improvement

A correction is evidence to investigate. It becomes a shared rule only after its meaning, scope and authority are established. Test the proposed change against cases it was not developed to pass.

Reusable expertise may include process patterns, software components and evaluation methods. Access to client information does not itself grant permission to reuse that information for another customer or product.

8. Measure the value that survives delivery

We intend to evaluate deployments against accepted work and its complete cost.

That includes preparation, review, rework, escalation, model usage, infrastructure, maintenance and the time people spend keeping the system useful. A higher automation rate is insufficient if it creates a larger review queue or more difficult failures.

The benefit also needs an identified mechanism. Released capacity may serve additional customers. More reliable evidence may reduce rework. Faster collections may improve cash timing. A shorter close may make information available for an earlier decision. These are distinct effects and should be measured separately.

Improvement	What must be established
Time released	Net effort falls after review and rework
Capacity created	The business can use that capacity productively
Cost reduced	An actual cost is avoided or removed
Cash accelerated	Timing improves without relabelling cash as new revenue
Quality improved	A defined error or service measure improves
Better decisions	The information changes an identified decision

> VALUE STANDARD
>
> ### Count what the business can use
>
> Measure quality, total human effort, elapsed time and full operating cost. Explain how the improvement affects the business. Keep capacity, cash and profit distinct.

Our responsibility is to make the evidence clear enough for a customer to decide whether to continue, expand, simplify or stop.

9. Complex implementation should leave lasting capability

Some work is necessarily specific to a business. Its systems, relationships and policies deserve attention. The commercial challenge is to preserve that depth while finding what can be repeated responsibly.

We intend to begin with focused engagements, use them to learn which problems recur and build shared capabilities only where repeated delivery justifies the investment. A maintained standard, useful evaluation cases and an understandable handover can retain value even as individual models change.

Long Lake’s published approach combines engineers and operators working within its companies with long-term ownership. We see operating proximity and continuity as useful principles; ownership brings separate capital and organisational requirements. [S7]

For Operating Standard, long-term operating value starts with a simpler question: does the improvement continue to work, can another person maintain it and does the customer still consider it worth paying for?

> CLIENT COMMITMENT
>
> ### Leave the business more capable
>
> Provide understandable definitions, documented responsibilities, exportable client evidence and a usable handover. An ongoing service should earn renewal through useful work and reliability.

10. A roadmap earned through evidence

Our intended development has five stages. They describe capabilities to establish, not milestones already achieved or a timetable we can promise.

Stage	Intended capability	Evidence required to progress
Deploy	Improve a bounded business workflow	Accepted work, dependable quality and paid demand
Standardise	Reuse the parts that recur	Similar work delivered by another operator with bounded variation
Operate	Maintain useful execution over time	Renewal, manageable support and accountable ownership
Compound	Connect related capabilities	Better delivery economics and customer value after full costs
Own, selectively	Consider deeper participation in operating value	A distinct investment case, capable leadership and adequate capital

Ownership is a possible later choice. A durable deployment and operating business is a worthwhile outcome in its own right. Any decision to invest in or acquire a business needs evidence beyond a successful software implementation.

Working towards a better operating standard

We believe AI deployment will be strongest when it gives the people who understand the work a greater role in shaping the system.

The ambition is practical: better definitions, more useful tools, clearer responsibilities, reliable execution and improvements that last. We expect to learn where this approach creates value and where simpler tools or different methods are more appropriate.

Operating Standard begins with that work. We intend to combine business understanding with engineering discipline, improve one meaningful process at a time and carry what we learn into the next deployment.

Working towards a better operating standard.

References

[S1] Google DeepMind. Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad. 21 July 2025. https://deepmind.google/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/

[S2] Brynjolfsson, E., Li, D. and Raymond, L. Generative AI at Work. Quarterly Journal of Economics, 140(2), 889-942, May 2025. https://doi.org/10.1093/qje/qjae044

[S3] Brynjolfsson, E., Rock, D. and Syverson, C. The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. American Economic Journal: Macroeconomics, 13(1), 333-372, 2021. Working-paper access: https://www.nber.org/papers/w25148

[S4] Ajzenstadt, M. The knowledge factory. Limestone Digital, 24 November 2026. https://thefoundation.limestonedigital.com/p/the-knowledge-factory

[S5] Anthropic. Building effective agents. 19 December 2024. https://www.anthropic.com/research/building-effective-agents

[S6] OpenAI and Thrive Holdings. Building self-improving tax agents with Codex. 27 May 2026. https://openai.com/index/building-self-improving-tax-agents-with-codex/

[S7] Long Lake. Company approach, accessed 28 November 2026. https://llmh.com/

[S8] Babina, T., He, A. and Jiang, R. Canaries in the Gold Mine: Early Productivity Gains from Artificial Intelligence Creating Organization Capital. CEPR Discussion Paper 21894, 31 August 2026. Publisher abstract: https://cepr.org/publications/dp21894

This is a founding position and intended approach. Proposed workflows and roadmap stages do not represent a record of completed OPSTD deployments.