AIMONGER
WHITE PAPER

AI Discovery

Human Capacity Is the AI Operating Model

Leadership teams can buy AI tools faster than they can build the human capacity to use them well. McKinsey's high-performing AI organisations redesign workflows and engage senior leaders; most organisations report AI use without comparable enterprise impact. The missing layer is named owners, trained operators, reviewers who can verify, and workshops that change practice. Tools do not absorb themselves.

David Saliba . 2026-07-03 . aimonger.com/whitepapers/human-capacity-ai-operating-model/

The problem in one sentence

If nobody is trained to run it, you did not deploy AI. You deployed shelfware.

Budgets list models, vendors, and integrators. They under-list operator time to learn the workflow, reviewer time for human gates, owner time to change process, trainer time to keep skills current, and risk partner time for releases. What you see globally right now is tool access scaling faster than capability. Suites add AI buttons, vendors sell enablement as an add-on, and workers hear that everyone should be more productive by Monday. In a mid-market firm, the bottleneck is often simpler: who owns the workflow, who reviews exceptions, and who has been trained to use the system under real conditions?

Human capacity, plainly: the people systems, namely owners, operators, reviewers, and trainers, that let AI run in production without heroes. **In practice:** invoice matching works overnight only because three AP operators passed certification, a reviewer covers the exception queue, and the business owner can change the matching rules without opening a ticket.

McKinsey's State of AI 2025 survey implies that impact concentrates where organisations change how work is done. That is human work, not model shopping.

The forgotten line item

Treat human capacity as infrastructure, not culture fluff. When finance asks why AI spend rose and throughput did not, the honest answer is often that licences were capitalised while training and review hours were not.

Line item	What it buys	What happens when it is missing
Operator training	Correct daily use, overrides, feedback	Reversion to email and spreadsheets
Reviewer capacity	Sustainable human gates	Burnout or rubber-stamping
Owner time	Process change authority	Permanent pilot
Trainer capacity	Curriculum and refresh	Skills decay after go-live
Risk partner time	Release checkpoints	Shadow workarounds

Deloitte's 2026 enterprise survey reports insufficient worker skills as the biggest barrier to integrating AI into workflows. Among workers with sanctioned access, fewer than 60 percent use AI in daily workflow in the reported pattern. That is not a model problem. It is an enablement problem.

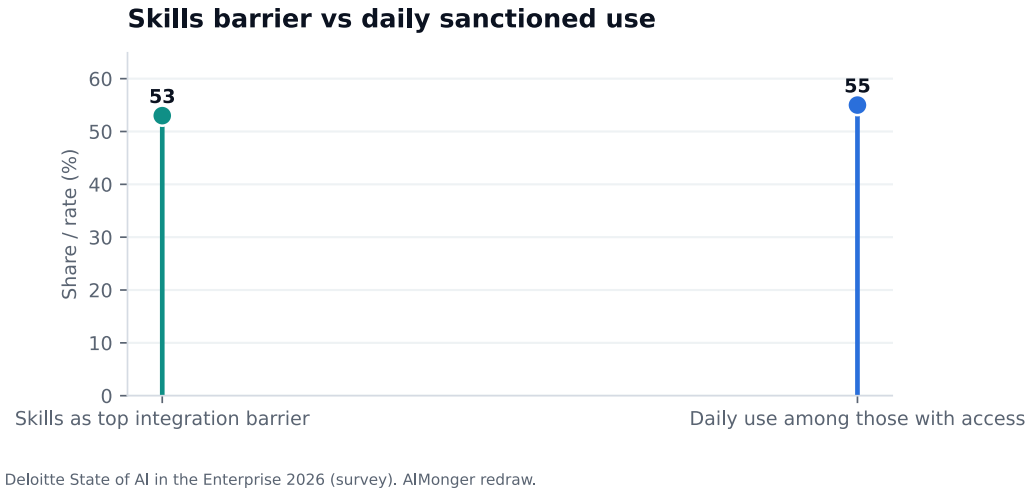

Figure 1. Reported skills barrier and incomplete daily use among workers with access. Plain read: access arrived; operating practice did not. **Worked case:** the board sees Copilot rollout complete; operations still runs on the old chase process. Source: Deloitte State of AI in the Enterprise 2026 (survey). AIMonger redraw.

Roles that must exist

Operating model, in short: who does what when AI touches a real workflow, not the org chart on the website. **On Monday:** every consequential AI action has a named owner for the metric, a reviewer for the gate, and a trainer who certifies access.

Role	Responsibility	Failure mode if vacant
Business owner	Metric, kill criterion, process change	Pilot never graduates
Technical owner	Release, eval, isolation	Fragile production
Operators	Daily use, overrides, feedback	Shelfware
Reviewers	Gates on consequential actions	Incidents or rubber stamps
Trainers	Curriculum and certification	Untrained production access
Risk partner	Security/legal checkpoints	Blocked releases or bypass

If two titles are the same person by accident, say so, and watch for burnout and bottlenecks. Heroics do not scale and create key-person risk.

Workshop design that changes metrics

Certification means: proof an operator can run, refuse, and escalate on the production system, not attendance at a keynote. **Operating case:** production access opens only after passing provenance reading, override protocol, and data-class rules on the live environment.

Good workshops:

- **Real systems and documents.** Operators practise on the firm's actual production environment and permitted document sets, so muscle memory transfers directly to Monday's queue. A generic sandbox hides the permission and provenance quirks that cause real incidents, which means staff trained only on toy data freeze the first time they meet a messy supplier format. We have seen cohorts pass a sandbox exercise and still file a severity-one ticket on day one because the live corpus contained scanned PDFs the trainer never showed them.

- **Failure modes on the curriculum.** Workshops deliberately teach temporal errors, citation theatre, and cost runaway so certified staff recognise when the model is confidently wrong. A model that invents a clause number and cites a nonexistent section is not a bug to fix later; it is a failure class operators must flag before it reaches a customer or ledger. Teaching these failure modes by name gives reviewers a shared vocabulary for exception triage under peak volume.
- **Override and provenance drills.** Every cohort must demonstrate how to refuse an answer, read source citations, and escalate exceptions using the same controls they will face under peak volume. Refusal is a skill: an operator who cannot articulate why a draft is unsafe will either send it or silently delete it, and both choices hide the defect from engineering. Provenance drills close that gap by making the operator trace each consequential claim to a document version before signing it off.
- **Certification tied to access.** Production credentials release only after passing exercises on the live system, which prevents the common pattern where attendance certificates precede unsafe daily use. Tying access to a demonstrated competency, rather than to a calendar date, means a rollout cannot outrun operator readiness. A mid-market insurer we advised held the line on this rule and shipped a workflow two weeks late rather than release uncertified staff onto claims triage.

Bad workshops:

- **Generic prompt entertainment.** Keynote-style "10 tricks" sessions produce enthusiastic chat users who still cannot run the firm's workflow, override safely, or file failures into evaluation. The cohort leaves with a positive feeling and zero transferable skill, which is the worst possible outcome because leadership reads the feedback scores as success. Six weeks later nobody on that cohort has touched the sanctioned tool and the steering deck still cites the session as enablement complete.
- **Vendor keynotes without exercises.** Passive listening leaves operators reverting to email and spreadsheets because nobody practised the exception queue on real supplier formats. A vendor demo looks polished precisely because it sidesteps the messy documents, ambiguous fields, and edge cases that define daily operations. When the only hands-on time is a curated sandbox, the gap between training and production widens instead of closes.
- **No link to a workflow metric.** When training is disconnected from a named KPI, the board sees completion rates while cycle time, error rate, and reviewer load stay flat. Completion is a process metric, not an outcome metric, and conflating the two is how programmes accumulate cost without surplus. A training line item without a paired workflow baseline is a leading indicator of shelfware.

Soft launches without certification recreate shadow-AI behaviour inside the official tool. Operators paste sensitive content into consumer tools because the sanctioned path feels harder than the training they never received.

Curriculum modules (minimum)

1. **Permitted actions.** Operators learn exactly what the system may propose, execute, or draft autonomously so they never assume authority the workflow has not granted. Without this boundary, a confident model can talk a user into approving an action that bypasses a control the firm spent years building. The curriculum teaches the difference between a draft a human sends and an action the system takes, because conflating the two is where most early incidents begin.
2. **Provenance reading.** Staff must trace every consequential answer to document version, section, and retrieval step before trusting it for a client, regulator, or ledger decision. A model that cites a policy section by number is not evidence; the operator must open the document and confirm the text matches the cited version. This drill exists because citation theatre, where a plausible reference points at the wrong or outdated passage, is the most common failure mode in retrieval-augmented systems.
3. **Refusal and override protocol.** Certification requires demonstrating when to stop the model, how to override safely, and who to notify when the system exceeds its tier. An operator who cannot refuse an unsafe answer will either send it or silently delete it, and both choices hide the

defect from engineering. The protocol turns refusal into a logged, reviewable event so that escalation feeds the evaluation fixture instead of vanishing into a private chat.

4. **Evaluation failure filing.** Operators learn how to log wrong answers, near-misses, and edge cases into the evaluation fixture so engineering can promote fixes instead of repeating incidents. Without a structured filing path, failures stay as individual war stories and the same defect surfaces across cohorts for months. The skill is not debugging the model; it is writing down what happened in a form another engineer can reproduce and test against.

5. **Data classes and shadow AI rules.** Every cohort receives plain-language rules on what may enter which tool, because permission without data-class discipline manufactures shadow AI inside the sanctioned product. An operator who pastes a class 4 client document into a consumer tier because the sanctioned path felt slower has created an incident, not a shortcut. Teaching data classes by name, with examples staff recognise from their own queue, is what makes the rule stick at 11 p.m. before a deadline.

6. **Incident basics.** Staff know the severity classes, first responder, and notification path so a harmful answer does not wait for a security review that starts after reputational damage. The first hour of a wrong-document leak determines whether it is a contained event or a board-level crisis. Operators do not need to run incident response, but they must know who to call and what to preserve before they close the tab.

7. **Cost awareness.** Operators understand when not to run heavy jobs, such as batch re-runs, wide corpus searches, and agent loops, so FinOps does not discover runaway spend from uncertified experimentation. A single agent stuck in a retry loop can consume a month of token allocation overnight if no budget cap stops it. Cost awareness is operator hygiene, not a finance problem to retrofit after the invoice arrives.

8. **Domain proposition workshop.** Business owners and operators co-design the workflow promise in plain language so the metric, kill criterion, and human gates stay aligned after go-live. When the proposition stays implicit, operators optimise for throughput while the owner expects quality, and the dashboard reports a number nobody agreed on. Co-designing the promise before certification is how the team avoids a metric that moves in the wrong direction.

Certification unlocks production access. Refresh on every material release. Training debt is a leading indicator of incident risk.

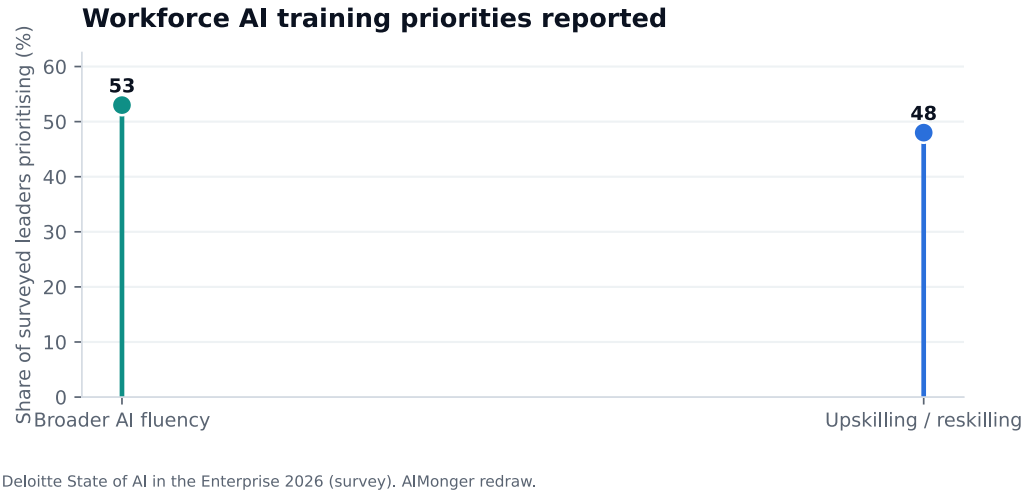

Figure 2. Shares of surveyed leaders prioritising fluency and upskilling. In short: the market knows training matters; budgets still underfund it. **On the P&L:** compare training line items to licence spend; if the ratio is inverted, expect shelfware. Source: Deloitte State of AI in the Enterprise 2026 (survey). AIMonger redraw.

Operating model blueprint

Treat AI enablement as a production system with four queues:

1. **Workflow queue.** The steering group maintains a ranked list of which processes are being absorbed this half so ownership, training, and reviewer staffing stay tied to named metrics rather than vendor demos. Ranking the queue forces a trade-off conversation, because a firm that tries to absorb twelve workflows at once absorbs none of them well. The queue is also the place where a permanent pilot becomes visible: an entry that never graduates is a cost centre the board can finally name.

2. **Training queue.** Every role that will touch production access appears here with certification dates, refresh rules, and blockers so go-live cannot proceed with uncertified operators holding credentials. The queue makes training debt visible before it becomes incident risk, because a slipped certification date is a cheaper signal than a wrong-document leak. When the training queue and the workflow queue fall out of sync, the programme ships a tool it cannot safely operate.

3. **Reviewer queue.** Human-gate staffing is modelled against expected case volume each week so propose-only mode triggers before reviewers rubber-stamp or burn out under backlog. Modelling reviewer hours against volume is the difference between a sustainable gate and a gate that quietly collapses into auto-approval. A reviewer queue that exceeds available hours is the earliest leading indicator that autonomy was promised before capacity was built.

4. **Exception queue.** Failures, overrides, and near-misses filed into evaluation fixtures feed engineering priorities instead of disappearing into individual inboxes after each incident. Without a structured exception queue, the same defect surfaces across cohorts for months because nobody wrote it down in a form engineering can reproduce. This queue is how the system learns from its own mistakes, which is the only honest mechanism for improving retrieval and routing over time.

If only the workflow queue is funded, the other three become silent failure modes. PwC's 2025 AI Jobs Barometer reports a 56 percent wage premium for AI-skilled workers and 66 percent faster skill change in AI-exposed occupations. That is a tight external labour market. Build and certify internally or pay the premium forever.

Worked example: accounts payable exception queue

Before: Three AP staff sample 20 percent of invoices because full review is unaffordable. Exceptions surface late. The metric is days payable outstanding, not AI usage.

After absorption with human capacity: Every invoice is extracted overnight. Operators certified on the system handle the dashboard. Reviewers see only PO mismatches above a threshold. The business owner owns exception rate and mean time to clear.

Human capacity budgeted:

- **Operator certification (16 hours, three people).** Three AP operators each receive sixteen hours of hands-on certification so overnight extraction runs without a hero on call when supplier formats or exception rules change. Sourcing three operators, rather than one, removes the key-person risk that turns a single holiday into a production outage. The sixteen-hour figure reflects time on the live system, not a slide deck, because muscle memory is what survives a Monday-morning queue.

- **Reviewer capacity (2 hours/day at peak).** Peak exception volume receives two dedicated reviewer hours per day so PO mismatches above threshold clear within SLA instead of forcing propose-only mode or silent rubber-stamping. Funding reviewer hours as a line item, rather than assuming they will be absorbed into existing roles, is what keeps the gate honest under load. When reviewer time is implicit, the queue backlogs and the dashboard reports a system that looks autonomous but is actually rubber-stamping.

- **Trainer refresh (4 hours/month).** The trainer spends four hours monthly updating curriculum on new supplier formats so certified operators do not revert to manual chase when the corpus shifts. Supplier formats change more often than the model does, which means curriculum that is not refreshed decays within a quarter. A trainer with a published refresh cadence is the difference between a workforce that adapts and one that quietly reverts to the old workflow.
- **Owner review (2 hours/month).** The business owner allocates two hours monthly to false-positive rate review so matching rules evolve with the metric instead of living permanently in the exception queue. Without scheduled owner review, matching rules drift out of alignment with the business reality they were built to reflect. Two hours a month is a small investment to prevent the rule set from becoming the next exception backlog.

If reviewer load exceeds available hours, the system stays in propose-only mode. Model the human minutes before promising autonomy.

Time budgets (make them explicit)

If a reviewer needs three minutes per case and volume is 200 cases/day, that is 10 hours/day of human capacity, a hiring and training fact, not a footnote. Boards should see reviewer utilisation beside token spend.

Metric	Meaning	Cadence
Certified operators	Who can run without heroes	Per release
Reviewer utilisation	Gate sustainability	Weekly
Time-to-override competence	Training quality	Per cohort
Owner coverage	Workflows with named owners	Monthly
Trainer backlog	Curriculum debt	Monthly

Capacity and culture

Public AI narratives can frighten staff into paralysis or reckless shadow use. Leadership's job is a precise story and a training path. Mid-market firms, including Malta-origin teams serving EU clients, win on trust and competence, not on headcount of tools.

Eurostat's 20.0 percent EU enterprise AI adoption figure (2025) shows many firms are still at the beginning of the capability curve. The opportunity is not to buy the largest catalogue. It is to train operators before local peers turn access into operating practice.

OECD work on AI and skills emphasises that only a small fraction of roles require advanced model-building skills, while digital, data, judgment, and collaboration skills matter across much of the workforce. Fund operator certification and job redesign beside licences, or expect shelfware and shadow AI.

Organising delivery

Large catalogues of half-trained tools recreate the productivity paradox. Deliberately small, invitation-quality delivery, meaning fewer concurrent programmes, higher ownership, and real training, beats spray enablement. McKinsey's high performers more often redesign workflows and engage senior leadership. Human ownership sits inside the value mechanism.

Evidence base: skills and redesign are the binding constraints

Source	Finding	Operating-model implication
Deloitte 2026	Insufficient worker skills reported as biggest barrier; 53% prioritise broader AI fluency; 48% upskilling/reskilling	Training is a production dependency
Deloitte 2026	Among workers with access, fewer than 60% use sanctioned AI in daily workflow	Access without enablement is shelfware
PwC 2025	AI-skilled workers commanded 56% wage premium; skills in AI-exposed occupations change 66% faster	Capability markets are tight; build internally
McKinsey 2025	High performers more often redesign workflows and engage senior leadership	Human ownership sits inside value
Gartner 2025	Over 40% of agentic AI projects predicted cancelled by end 2027, citing unclear value and weak controls	Underbuilt human capacity drives cancellation

Plain read: the data does not say training alone guarantees ROI. It says tool-first programmes without operators, reviewers, and owners predictably stall.

Methodology note

Wage premiums and skill-change rates from job-ad analyses describe employer demand signals. They do not prove every firm should hire the same specialty. Survey percentages differ by sample and question wording. Treat triangulation as directional for management priorities.

Counter-position: experts will figure it out

Another board might argue that a few power users can carry the programme while everyone else catches up. A few experts can carry a demo. Production systems need certified operators, staffed reviewers for gates, and owners who can change process. Heroics do not scale and create key-person risk. The counter-position is valid for sandbox learning; it fails for operating capacity.

CIO decision criteria

Clear investment committee only when the programme identifies:

1. **Named owners per workflow.** Every priority workflow lists a business owner for the metric and a technical owner for release so accountability survives steering-deck reshuffles and vendor churn. Without named owners, the workflow becomes an orphan that nobody can graduate, pause, or kill when the metric stalls. The naming also forces the conversation about whether one person is accidentally holding two seats, which is a burnout risk disguised as efficiency.

2. **Certification before access.** Production credentials issue only after documented certification requirements pass, which prevents the pattern where licences arrive months before anyone can run the system safely. Tying access to demonstrated competency, rather than to a procurement date, is the single most effective control against shelfware. A CIO who cannot answer “who is certified on this workflow today” is governing a licence estate, not a production system.

3. **Modelled reviewer capacity.** Reviewer hours are calculated against expected case volume and reported weekly so the CIO can trigger propose-only mode before gates collapse under load. The model converts an abstract promise of autonomy into a staffing fact that finance can cost. When reviewer capacity is modelled rather than assumed, the CIO gains a leading indicator that precedes every gate collapse by weeks.

4. **Trainer with refresh cadence.** A named trainer owns curriculum updates on a published schedule so skills decay after go-live does not recreate shadow workarounds within a quarter. Supplier formats and exception rules shift faster than the model does, which means a curriculum without a refresh owner is already going stale. The cadence also gives the steering group a checkpoint to retire training that no longer matches the live workflow.
5. **Training budget beside licences.** Finance sees a training line item adjacent to licence and integration spend so human capacity is capitalised like infrastructure rather than treated as optional culture spend. Placing the two lines beside each other makes the inverted ratio visible to the investment committee without a special report. When training sits in a separate culture budget, it is the first item cut when the programme needs to show savings.
6. **Operating metrics on the dashboard.** Certified operators, reviewer utilisation, and owner coverage appear beside token spend so the board governs capacity, not only consumption. Token spend measures how much the firm is buying; capacity metrics measure whether the firm can actually use what it bought. A dashboard that shows only consumption is a vendor scorecard, not a board instrument.
7. **Go-live kill criterion.** If training completion falls below the published threshold at go-live, promotion pauses automatically instead of shipping uncertified operators into consequential workflows. The kill criterion removes the social pressure to ship on a date regardless of readiness, which is how most early incidents begin. Publishing the threshold before go-live is what makes it enforceable rather than aspirational.

Appendix: enablement plan template

Use this table for each priority workflow before licences go live. Leave the Entry column blank until the steering group fills it.

Field	Entry
Workflow	
Audience roles	
Workshop dates	
Certification rule	
Refresh cadence	
Owner	

Appendix: ninety-day human-capacity charter

Complete one charter per workflow before production access. Go-live stays blocked until both conditions below are true: certification is complete, and reviewer capacity is staffed.

Field	Entry
Workflow name	
Operators required (count and roles)	
Certification date target	
Reviewer FTE or hours per day	
Trainer owner	
Business owner	
Technical owner	
Go-live blocked until	Certification complete, and reviewer capacity staffed

Plain read: empty cells are intentional. A filled charter is the go-live gate, not a decorative appendix.

Closing position

Human capacity is not a soft afterthought. It is the operating model.

Fund it like infrastructure. Measure it like delivery. Without it, Gartner's cancellation drivers, namely unclear value and weak controls, arrive on schedule.

References

1. Deloitte AI Institute, "The State of AI in the Enterprise: The Untapped Edge" (2026 edition); survey of 3,235 leaders, Aug-Sep 2025. https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html
2. PwC, "The Fearless Future: 2025 Global AI Jobs Barometer," 3 June 2025. https://www.pwc.com/gx/en/news-room/press-releases/2025/ai-linked-to-a-fourfold-increase-in-productivity-growth.html
3. McKinsey & Company / QuantumBlack, "The State of AI: Global Survey 2025." https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
4. Eurostat, "20% of EU enterprises use AI technologies," 11 December 2025 (20.0% in 2025; 13.5% in 2024). https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/ddn-20251211-2
5. Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," press release, 25 June 2025. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027
6. Gartner, "Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026," press release, 11 October 2023. https://www.gartner.com/en/newsroom/press-releases/2023-10-11-gartner-says-more-than-80-percent-of-enterprises-will-have-used-generative-ai-apis-or-deployed-generative-ai-enabled-applications-by-2026
7. Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU AI Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj
8. OECD, AI and skills publications hub. https://oecd.ai/en/ai-principles
9. NIST, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," NIST AI 600-1, July 2024. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.600-1.pdf
10. Stanford Institute for Human-Centered Artificial Intelligence, "The 2026 AI Index Report," Economy chapter. https://hai.stanford.edu/ai-index/2026-ai-index-report/economy

Published by AIMonger . https://aimonger.com . Malta . Europe