The Hidden Cost of AI Efficiency: Preventing Expertise Collapse in the Modern Workforce

Executive Summary

Artificial intelligence is producing real productivity gains, but the evidence is now clear that those gains are uneven, conditional, and sometimes reversible. In customer support, access to a generative AI assistant raised productivity by about 15% on average and helped less-experienced workers move down the experience curve faster; in three randomized field experiments with software developers, AI access increased completed tasks by about 26% on average, again with larger gains for junior workers. Yet a large field experiment with consultants found that AI improved speed and quality only for tasks inside the model's capability frontier, while harming performance on a task outside that frontier; a randomized study of experienced open-source developers found the opposite of the usual optimism narrative—developers believed AI made them faster, but completion time actually increased by 19%. [1]

The strategic risk is therefore not "AI lowers productivity." The deeper risk is that AI can raise output while quietly hollowing out the human capability base that creates resilience: judgment under uncertainty, tacit know-how, exception handling, peer teaching, and apprenticeship throughput. The human-factors literature reached this conclusion long before generative AI. A meta-analysis of 18 automation experiments found that higher automation improved routine performance and often lowered workload, but reduced situation awareness and worsened performance when the system failed; the classic "ironies of automation" argument similarly warned that people are often left responsible for rare abnormal conditions precisely when automation has deprived them of the repetitions needed to stay sharp. [2]

That risk is best described as **expertise collapse**: an organizational condition in which short-run efficiency rises while the stock of human expertise decays faster than it is replenished. The danger is greatest when firms use AI to eliminate intermediate-difficulty work, thin entry-level roles, reduce peer interaction, or convert skilled contributors into passive verifiers. Recent evidence from software development shows that AI can shift work toward more independent execution and away from collaborative/project-management activity; recent management research also warns that firms can become more automated yet less adaptive, less wise, and less able to sustain the human skills that make them competitive. [3]

The product implication is not to slow AI adoption, but to **redesign AI adoption as a human-capital system**. This report proposes **RUDY AI Human Expertise Preservation Engine**, an enterprise layer that sits on top of copilots and workflow systems to preserve the three inputs of durable expertise: **repetitions**, **reflections**, and **relationships**. RUDY uses risk-adaptive autonomy, challenge injection, rationale capture, spaced re-practice,

expert-to-junior exemplar routing, and fatigue-aware workflow design to convert AI from a substitution tool into an apprenticeship-preserving operating system. The result is a better balance between efficiency and capability formation. The evidence base for this design draws from official sources from MIT Sloan, HBR, OECD, and McKinsey, plus primary work in cognitive psychology, learning science, and human factors. [4]

Because no industry or company size was specified, the operating model below is cross-industry and the financial model is expressed **per 1,000 AI-exposed employees**. On illustrative assumptions that are deliberately more conservative than the strongest empirical productivity estimates, a three-year deployment can produce positive net present value while also improving learning velocity, exception-handling quality, and junior capability formation. The core recommendation is to pilot RUDY in a workflow that is measurable, repeatable, and exception-rich—such as customer operations, software engineering, claims review, finance operations, or IT service—using a cluster-randomized or staggered rollout design. [5]

Research Base and Problem Definition

At the macro level, the official message from labor-market institutions is not that AI mainly creates a need for more machine-learning engineers. The National Bureau of Economic Research[6]-adjacent field literature and the OECD evidence instead point to a dual shift: many jobs will become more AI-exposed, but most of those jobs will require **general AI literacy, critical evaluation, emotional/cognitive skills, and stronger teaching/training capabilities**, not specialist model-building. OECD reports that roughly one in three job vacancies has high AI exposure, while only about 1% require advanced AI-specialist skills; McKinsey projects that by 2030 up to 30% of current hours worked could be automated in the United States, with large-scale occupational transitions and rising demand for critical thinking, creativity, and teaching/training. [7]

That macro picture matters because organizational capability is not just the sum of current output. It is the sum of what employees can do **without** the tool, what they can diagnose when the tool is wrong, and what seniors can teach juniors when the tool is absent or uncertain. A helpful working definition is:

Expertise collapse = productivity gain – capability replenishment deficit

where the replenishment deficit is driven by fewer practice reps, weaker reflection, less peer teaching, and thinner apprenticeship pipelines. This formulation is consistent with the human-factors literature on automation trade-offs, recent field evidence on AI's heterogeneous effects, and recent management writing that frames AI adoption as a strategic choice between augmentation and mere automation. [8]

This report therefore treats AI adoption as a **human-capital allocation problem** rather than only a software procurement problem. The central policy question is not "Where can AI save time?" but "Which tasks must remain developmental, which tasks can be accelerated, and what governance is needed so that efficiency does not cannibalize future

expertise?" That framing is increasingly aligned with recent MIT Sloan and HBR analysis emphasizing metacognition, accountability, learning design, and augmentation over pure substitution. [9]

Evidence on Expertise Collapse

Skill atrophy from automation

The most established result in the automation literature is that better routine performance can coexist with worse human backup performance. The 2014 meta-analysis of 18 experiments found a clear benefit of increasing automation for routine system performance, but a negative effect on failure performance and situation awareness, especially when automation crossed a boundary from supporting analysis to supporting action selection. This is the core technical precursor to expertise collapse: the organization becomes more efficient precisely by relocating humans away from the learning-rich parts of the task. [10]

The classic "ironies of automation" thesis sharpened the same point decades earlier. Automation often removes the frequent tasks through which operators build and maintain skill, then leaves them responsible for the rare, ambiguous, high-stakes situations when the system breaks or encounters novelty. In contemporary genAI terms, the operator becomes a fallback generalist without enough recent practice at diagnosis or recovery. [11]

The mitigation lesson is also well established: overreliance is not purely inevitable. Studies of automation bias and complacency show that training users on **rare automation failures** improves cross-checking behavior and reduces passive acceptance of automated advice. That matters for enterprise AI design because it implies that "AI safety" in the workforce is not only model governance; it is also **user conditioning**. Employees need structured exposure to model limits, not just policy documents about them. [12]

Decision fatigue and the shift from execution burden to verification burden

Decision fatigue is the tendency for decisions to become less effortful, more default-oriented, or lower quality after repeated decision-making demands. Recent reviews in clinical settings link decision fatigue to impaired judgment, reduced diagnostic quality, and higher error risk; field evidence from judicial and surgical decisions shows systematic within-day drift in outcomes as choice sequences lengthen or fatigue accumulates. Although those domains differ from general knowledge work, they provide the strongest current evidence that repeated micro-decisions can reliably distort expert performance. [13]

In AI-enabled knowledge work, the form of fatigue changes. Workers may make fewer first-draft decisions but more **verification, exception-handling, escalation, and accountability** decisions. That is why the promise that AI "reduces work" can be misleading. Recent HBR analysis argues that AI often intensifies rather than removes work,

because organizations layer new expectations of speed, responsiveness, and review on top of existing responsibilities. The hidden cost is that fatigue migrates from production into **monitoring and judgment**, where errors may be harder to observe but more consequential. [14]

For RUDY, the practical implication is that preserving expertise cannot mean bombarding users with more prompts, more nudges, and more approvals. The system must *reduce low-value cognitive friction while preserving high-value deliberate cognition*. This is why fatigue-aware routing, protected focus blocks, and selective challenge injection are part of the proposed architecture rather than ancillary change-management ideas. [15]

Junior employee capability erosion

The evidence on juniors is nuanced and strategically important. On the one hand, generative AI often benefits less-experienced workers the most. In the 2025 QJE customer-support study, newer and lower-skilled workers saw the largest gains, including a roughly 30% improvement in some productivity measures; the study also found that AI could help newer workers move down the experience curve faster. In large developer experiments, junior and short-tenured developers also saw larger gains than senior developers. These are real advantages and should not be dismissed. [16]

On the other hand, the same mechanism can become corrosive if firms interpret it as a reason to reduce junior hiring or compress developmental work. MIT Sloan's recent reporting on software development explicitly warns against replacing entry-level workers with models, noting that junior developers show the biggest benefits and that cutting these roles is a strategic error. In open-source settings, AI access has also been associated with shifts toward independent work and away from some collaborative or project-management activities, which threatens the very social pathways through which tacit knowledge is normally transferred. [17]

The strongest risk is therefore **junior output without junior formation**. A worker can look more productive because a model is carrying diagnostic structure, composition patterns, and tacit heuristics that the worker has not yet internalized. That is useful in the short run, but if the organization removes entry-level seats, fewer people accumulate intermediate competence; if collaboration declines, fewer people learn how experts reason; and if top workers increasingly defer to the model, the organization may weaken the very source material from which future models and future employees learn. The QJE study itself raises this longer-run concern by noting that top workers increased adherence to AI recommendations even when quality marginally declined, potentially reducing novel expert contributions over time. [18]

Learning theory and apprenticeship models

The learning-science literature is unusually consistent on one point: durable expertise is not created by frictionless fluency alone. Expert performance grows through **deliberate practice**, feedback, error correction, and repeated retrieval under conditions that are

effortful enough to strengthen storage rather than merely inflate short-term confidence. The deliberate-practice framework argues that expert performance is built through prolonged, effortful improvement activities; work on "desirable difficulties" similarly shows that spacing, interleaving, testing, and variation make learning feel harder in the moment but produce more durable and transferable knowledge. [19]

Applied review evidence on retrieval practice reinforces the same point. A 2021 systematic review of real classroom studies found robust benefits from retrieval practice across settings, levels, and test formats. Taken together with broader cognitive reviews, the implication for AI-enabled work is straightforward: if AI removes retrieval, comparison, and self-explanation from the workflow, it may raise immediate throughput while reducing retention and transfer. [20]

Scaffolding matters, but it must be **faded**. The expertise-reversal literature shows that heavily guided instruction helps novices but can become redundant or even counterproductive as expertise rises. Conversely, less structured tasks may require examples and support for longer than highly structured tasks. That directly supports a dynamic AI policy: junior employees should receive more worked examples, prompts, and coaching early on, but support should fade into challenge mode as their competence grows. [21]

Apprenticeship theory fills the social half of the model. Cognitive apprenticeship argues that learning improves when thinking is made visible through modeling, coaching, scaffolding, articulation, reflection, and exploration. Modern apprenticeship research also shows that apprenticeship is a durable workforce-development model, while noting that the literature is still thin on the specific curricular elements that drive retention, completion, advancement, and learning. In enterprise AI, the lesson is not to romanticize old mentorship systems; it is to **instrument them**. RUDY should make expert reasoning visible, route juniors into graduated responsibility, and preserve teaching as a measurable organizational output rather than an untracked side activity. [22]

Intervention comparison

Intervention	Mechanism	Evidence base	Best use case	Failure mode if absent	RUDY implementation
Graduated automation	Match autonomy level to task criticality and user expertise	Higher automation improves routine performance but worsens failure performance and situation awareness at higher levels. [10]	High-volume workflows with meaningful exceptions	Passive overreliance; brittle fallback performance	Risk engine selects Assist, Coach, or Challenge mode
Rare-failure	Train	Exposure to rare	Safety-,	Blind trust	Scheduled

Intervention	Mechanism	Evidence base	Best use case	Failure mode if absent	RUDY implementation
exposure	vigilance and calibration by exposing users to model limits	automation failures improves cross-checking and reduces complacency. [23]	compliance-, and quality-sensitive work	in AI suggestions	“failure drills” and counterexample libraries
Self-explanation and rationale capture	Force articulation of reasoning, not just answer acceptance	Self-explanation improves learning from examples; MIT Sloan finds AI boosts creativity mainly for workers with strong metacognitive strategies. [24]	Knowledge work requiring judgment	Shallow adoption with weak transfer	“Explain before accept” and post-task reasoning summaries
Retrieval practice and spaced re-practice	Convert exposure into durable memory and transfer	Retrieval practice yields robust learning gains; spacing and testing are desirable difficulties. [25]	Onboarding, new process adoption, policy-heavy roles	Fast forgetting and tool dependence	Auto-generated retrieval cards, spaced follow-ups, mini-sims
Faded scaffolding	Increase challenge as competence increases	Novices benefit from more guidance; support should fade as expertise develops. [21]	Junior ramp-up and cross-skilling	Eternal novice mode; support becomes crutch	Competence-based progression from examples to challenge tasks
Instrumented apprenticeship	Preserve tacit knowledge through modeled thinking and coaching	Cognitive apprenticeship emphasizes modeling, coaching, articulation, reflection, exploration; apprenticeship	Teams with specialist expertise and scarce seniors	Lost tacit knowledge; thin leadership bench	Exemplar graph, shadow reviews, mentor routing

Intervention	Mechanism	Evidence base	Best use case	Failure mode if absent	RUDY implementation
		literature shows enduring workforce value. [26]			
Collaboration protection	Defend learning-rich peer interaction	AI can shift work toward independent activity and away from collaborative/project-management activity. [27]	Engineering, consulting, operations	Individual productivity rises while collective learning falls	Mandatory peer review thresholds and collaborative retros

RUDY AI Human Expertise Preservation Engine

RUDY AI should be positioned as a **control layer for capability-preserving augmentation**. It does not replace enterprise copilots. It orchestrates how, when, and under what learning conditions those copilots are used. The design goal is to optimize for two outputs simultaneously: **productive work now** and **stronger human judgment later**. That is the practical translation of recent MIT Sloan and HBR guidance that AI adoption works best when organizations emphasize learning, accountability, role redesign, and augmentation over pure labor substitution. [28]

Architecture

```
flowchart TB
    A[Enterprise workflow systems<br/>email, docs, chat, CRM, ITSM, code,
tickets] --> B[Event and context ingestion]
    B --> C[Task classification and risk engine]
    C --> D{Policy engine}
    D -->|low risk / high proficiency| E[Assist mode]
    D -->|medium risk / developing proficiency| F[Coach mode]
    D -->|high risk / edge case / novice critical| G[Challenge mode]
    D -->|scheduled capability maintenance| H[Drill mode]

    E --> I[Foundation models + enterprise knowledge]
    F --> I
    G --> I
    H --> I

    I --> J[User workspace]
    J --> K[Verification capture and outcome logging]
    K --> L[Skill ledger and expertise graph]
    L --> M[Spaced retrieval and simulation engine]
```

```
L --> N[Mentor routing and exemplar library]
K --> O[KPI, experiment, and ROI analytics]
O --> P[Governance, audit, privacy, policy controls]
```

The architectural logic is evidence-linked. **Assist mode** captures straightforward efficiency opportunities. **Coach mode** inserts prompts that require users to explain decisions, compare alternatives, or identify where the model might fail. **Challenge mode** is the anti-collapse mechanism: the worker must produce an initial judgment, exception diagnosis, or plan before seeing—or before fully accepting—the AI answer. **Drill mode** deliberately creates protected reps in which the tool is restricted or delayed so that humans continue to exercise the underlying skill. This mirrors the learning-science distinction between assistance that speeds task completion and practice that builds durable capability. [29]

Core features

Feature	What it does	Why it matters
Risk-adaptive autonomy	Uses task criticality, user tenure, prior accuracy, and novelty to determine AI mode	Prevents a one-size-fits-all policy from over-automating developmental work
Rationale capture	Requires concise user explanation for high-impact AI-assisted decisions	Preserves articulation and auditability; builds metacognitive skill
Challenge injection	Withholds or delays AI in selected cases; asks for first-pass human judgment	Protects diagnostic and exception-handling capability
Exemplar distillation	Converts expert reviews and high-quality decisions into reusable worked examples	Scales expert teaching without flattening nuance
Apprenticeship graph	Matches juniors to reviewers, mentors, or shadow opportunities based on skill gaps	Rebuilds learning relationships that AI can otherwise thin
Spaced re-practice	Creates follow-up quizzes, simulations, and edge-case drills from real work	Converts work episodes into durable memory
Fatigue-aware orchestration	Detects end-of-shift drift, queue overload, and excessive review burden; changes flow accordingly	Prevents verification burden from becoming hidden decision fatigue
Skill ledger	Tracks capability growth by skill, not just aggregate productivity	Gives leadership a human-capital dashboard rather than only an automation dashboard

Workflow design

A typical RUDY workflow has five stages.

1. **Classify the work episode.** The system tags the task by risk, novelty, reversibility, and developmental value. High-frequency low-risk work can be automated more aggressively; medium-risk work is coached; high-risk and edge-case work is challenge-routed.
2. **Select the autonomy mode.** A senior employee handling a routine case may see AI immediately. A junior employee handling a complex exception may first be asked for a diagnosis, then shown the model's proposal, then asked to compare the two.
3. **Capture comparison and reasoning.** On selected cases, the user must state why the final choice was accepted, edited, or rejected. Over time this becomes a searchable evidence base of human-AI disagreement.
4. **Route for learning.** Real cases generate worked examples, retrieval questions, micro-simulations, and matched mentor reviews. This is where routine work becomes an apprenticeship substrate.
5. **Measure capability outcomes.** The same event stream feeds dashboards on manual fallback performance, judgment quality, junior independence, expert teaching yield, collaboration density, and fatigue drift.

Integration points

No single vendor stack is required. A pilot should start with the integrations that make capability formation observable.

System layer	Example integration point	RUDY data / action
Productivity suites	Microsoft[30] 365, Google Workspace	Draft/accept/edit patterns, rationale prompts, retrieval follow-ups
Collaboration tools	Slack, Teams	Mentor routing, postmortems, collaboration-density measurement
Engineering tools	GitHub[31], Jira, CI/CD	Code review coverage, edge-case drills, junior independence tracking
Customer/operations systems	Salesforce, Zendesk, ServiceNow	Resolution quality, escalation patterns, time-to-proficiency
Learning systems	LMS, LXP, simulation tools	Spaced practice, certification, knowledge checks
People systems	HRIS, identity, skills taxonomies	Tenure-aware policies, role mapping, experimental stratification
Data/BI	Warehouse, dashboard tools	KPI computation, ROI, experimentation, governance audit

KPI Models and Experimental Framework

The central measurement error in enterprise AI today is that organizations often track adoption and output but not **capability preservation**. RUDY requires a dual-scorecard: one layer for business performance and another for expertise formation. This is consistent with MIT Sloan's emphasis on learning and accountability, HBR's warning about organizational skill erosion, and OECD's emphasis on AI literacy plus cognitive and social skills rather than narrow tool familiarity. [32]

A practical composite metric is the **Expertise Preservation Index**:

$$\begin{aligned}\text{EPI} = {} & 0.25(\text{manual exception accuracy}) + 0.20(\text{rationale quality}) \\ & + 0.20(\text{junior independence}) + 0.15(\text{verification depth}) \\ & + 0.10(\text{expert teaching yield}) + 0.10(\text{fallback readiness})\end{aligned}$$

Each component should be standardized to a 0–100 scale within role family. The point is not false precision; the point is to make expertise a managed asset.

KPI model

KPI	Operational definition	Baseline	Recommended pilot target	Why it matters
Expertise Preservation Index	Composite score above	8–12 week pre-period average	+10 to +15 points by month 6	Makes capability visible
Time to proficiency	Days until worker performs target task bundle at agreed quality without close supervision	Current median by role	-15% to -25%	Captures whether AI accelerates learning, not just output
Manual exception accuracy	Accuracy on sampled "AI-restricted" or edge-case tasks	Audit-based baseline	+8% to +15%	Tests whether workers can still reason without the tool
Junior independence ratio	Share of tasks completed by juniors without escalation after training threshold	Historical cohort baseline	+10 to +20 percentage points	Signals real capability accretion
Verification depth rate	Share of AI-assisted tasks with substantive review action, source check, or reasoned edit	Current AI workflow rate	60%+ low-risk, 85%+ high-risk	Distinguishes real judgment from rubber-stamping

KPI	Operational definition	Baseline	Recommended pilot target	Why it matters
Expert teaching yield	Validated exemplars, reviews, or learning artifacts contributed per expert per month	Historical average	+25% to +50%	Prevents experts from becoming only bottleneck reviewers
Collaboration density	Peer reviews, comments, or consults per 100 complex tasks	Pre-pilot average	No decline in complex work; ideally +5% to +10%	Guards against thinning apprenticeship relationships
Decision-fatigue drift	End-of-shift decline in quality, escalation appropriateness, or defaulting behavior; paired with workload survey	Pre-pilot hourly/shift slope	20%+ reduction in adverse drift	Detects hidden review and monitoring burden
Retention in junior cohort	6- and 12-month retention for early-career employees	Historical cohort retention	+0.5 to +1.5 percentage points	Connects learning design to talent economics

These target bands are intentionally moderate. They sit below the strongest short-run productivity estimates in the field literature and are more appropriate for a controlled enterprise pilot than for a marketing forecast. [33]

Experimental design

The recommended primary design is a **cluster-randomized controlled trial** at the team or manager level to reduce contamination.

Element	Recommended design
Units	Teams or pods with comparable workflow, manager span, and task mix
Arms	Control: standard enterprise AI; Treatment A: standard AI + RUDY Coach; Treatment B: standard AI + RUDY Coach + Apprenticeship
Duration	8–12 week baseline + 24 week intervention
Primary outcomes	Time to proficiency, manual exception accuracy, EPI
Secondary outcomes	Throughput, quality, rework, collaboration density, fatigue drift, retention intent
Data sources	System logs, blinded quality audits, simulations, micro-assessments, surveys, HRIS

Element	Recommended design
Randomization strata	Role family, tenure band, baseline performance, geography if relevant
Sample size	A practical minimum for a two-arm cluster RCT targeting a moderate effect of about **0.30 SD**, with **ICC ≈ 0.05** and **~12 employees per team**, is roughly **24 teams per arm** (about **576 employees total**). For three arms, a prudent target is **~720–900 employees**, depending on expected attrition and outcome variance.
Estimand	Intention-to-treat primary; treatment-on-treated secondary

If randomization is infeasible, the next-best design is a **staggered rollout quasi-experiment** using difference-in-differences with unit and time fixed effects, plus an event-study plot to test pre-trends. The preferred setting is one with rich logs and auditable output, following the logic of recent field studies in customer support, software development, and consulting. [34]

Experimental flow

```
flowchart LR
    A[Select comparable teams] --> B[Collect 8-12 week baseline]
    B --> C[Stratify by role, tenure, baseline performance]
    C --> D[Randomize or stagger rollout]
    D --> E1[Control<br/>standard AI]
    D --> E2[Treatment A<br/>RUDY Coach]
    D --> E3[Treatment B<br/>RUDY Coach + Apprenticeship]
    E1 --> F[24 week pilot]
    E2 --> F
    E3 --> F
    F --> G[Collect logs, audits, simulations, surveys, HR data]
    G --> H[Primary and secondary analyses]
    H --> I[Scale, refine, or stop]
```

Statistical analysis plan

The analysis should be pre-registered. For continuous outcomes such as EPI or time-to-proficiency scores, use mixed-effects ANCOVA or linear mixed models with baseline adjustment, time fixed effects, and team-level random intercepts. For binary outcomes such as correct exception handling, use mixed-effects logistic models. For time-to-proficiency, use discrete-time hazard models or Cox models where appropriate. The primary analysis should be **intention-to-treat**, with treatment-on-treated estimates as secondary analyses using assignment as the primary source of exogenous variation in use. [35]

Heterogeneity analyses should be pre-specified for tenure, role complexity, baseline performance, and task risk. Mediation analyses should test whether gains in proficiency are transmitted through rationale quality, verification depth, or mentor exposure. Multiple testing should be controlled using a hierarchical testing plan or false-discovery-rate

procedure. Missing survey or audit data should be handled with multiple imputation or inverse-probability weighting, while system-log outcomes should be analyzed as close to census data as possible. These choices matter because the central aim is to distinguish **novelty and throughput effects** from **true learning and resilience effects**. [36]

Enterprise ROI Models

Because industry and company size are unspecified, the financial model below is **illustrative and scalable**. It is expressed per **1,000 AI-exposed employees** over a **three-year horizon** at a **10% discount rate**. The assumptions are intentionally conservative relative to the strongest field estimates reported in customer support, developer studies, and writing tasks. [37]

Core formulas

$$\begin{aligned}&\text{Productivity benefit}\\&= N \times \text{fully loaded cost} \times \text{AI-exposed work share} \times \text{time saved}\\&\times \text{capture rate}\end{aligned}$$

$$\begin{aligned}\text{Ramp benefit} = \text{junior cohort} \times \text{cost per workday} \times \text{days to proficiency saved}\\\times \text{realization rate}\end{aligned}$$

$$\text{NPV} = -\text{setup cost} + \sum_{t=1}^{3} \frac{\text{gross benefits}_t - \text{recurring cost}_t}{(1+r)^t}$$

Assumptions for scenario modeling

The model below uses the following baseline assumptions:

- 1,000 AI-exposed workers
- Fully loaded annual cost per worker: **$120,000**
- Share of work materially AI-exposed: **35%**
- Junior cohort: **15%** of workforce
- One-time setup cost: **$0.8M**
- Annual recurring cost: **$0.9M**

These assumptions are not claims about a specific industry. They are a cross-industry template that can be rescaled.

ROI scenarios

Scenario	Productivity recapture	Faster junior ramp	Error / rework reduction	Attrition reduction	Gross annual benefit	Net annual benefit after recurring cost	3-year NPV	Approx. IRR
Conservative	$0.67M	$0.30M	$0.21M	$0.20M	**$1.38M**	**$0.48M**	**$0.39M**	~23%
Base	$1.62M	$0.72M	$0.42M	$0.40M	**$3.16M**	**$2.26M**	**$4.82M**	~180%
Aggressive	$2.73M	$1.30M	$0.63M	$0.60M	**$5.26M**	**$4.36M**	**$10.04M**	>300%

The logic behind these scenarios is that the firm does **not** need to capture the full headline productivity effect from the research literature to justify RUDY. It only needs to convert a modest proportion of time savings into usable capacity while also shortening junior ramp time, lowering rework, and preserving retention. Put differently: RUDY's business case does not depend on believing maximalist AI claims. It depends on preventing a hidden human-capital liability while capturing a realistic portion of measurable gains. [38]

Sensitivity analysis

Variable around base case	Low case	High case	Approx. NPV impact
AI-exposed work share	25%	45%	~$3.0M to ~$6.3M
Productivity capture rate	40% effective capture	70% effective capture	~$3.2M to ~$6.4M
Active adoption in treatment group	55%	85%	Largest non-cost driver
Annual recurring cost	$1.2M	$0.6M	~$4.1M to ~$5.6M

The model is most sensitive to **capture rate**, **share of work genuinely AI-exposed**, and **active adoption of the right workflows**, not to modest license-cost changes. That makes the operating model more important than the model vendor. A poorly governed rollout can deliver impressive adoption numbers and still destroy ROI by weakening judgment, increasing verification burden, and thinning apprenticeship pipelines. [39]

Ethical Implementation and the Future of Human Capital

The first ethical principle is that RUDY should optimize for **human development, not worker surveillance**. Employees should know which behaviors are being logged, why those logs are being used, how long they are retained, and which metrics are used for

developmental coaching versus performance management. Capability-preserving systems will fail if workers experience them as covert productivity policing. This is especially important because the OECD frames AI literacy partly in terms of critical evaluation and ethics, not just tool use. [40]

The second principle is **augmentation before substitution**. HBR's recent argument that firms choosing augmentation may outperform those choosing pure automation is not just a cultural recommendation; it is a human-capital strategy. When junior roles disappear, so do future experts. When experts cease to practice or teach, the organization becomes technically advanced but strategically fragile. AI should therefore be deployed with explicit role-design safeguards: protected developmental tasks, required mentor throughput, and minimum collaboration thresholds in knowledge-rich work. [41]

The third principle is **calibrated trust**. Workers should not be pushed toward either blind reliance or blanket rejection. The goal is informed, well-calibrated use. That requires transparent uncertainty cues where possible, routine exposure to failure cases, and workflow prompts that ask users to identify where the system might be wrong. In critical work, "accepting the answer" should not count as a competent action unless it is paired with evidence of verification. [42]

The fourth principle is **social preservation**. Work is not only an information-processing pipeline. It is also a network in which trust, tacit knowledge, and professional identity are formed. Recent HBR writing notes that employees are beginning to rely on AI for social support and that AI can weaken human connection at work; MIT Sloan reporting on software development similarly raises concerns about reduced teamwork as AI use grows. Human-capital governance for the AI era therefore must monitor not only productivity and quality, but also whether the organization is losing the interactions through which judgment is transmitted. [43]

The longer-run implication is that labor advantage will increasingly belong to firms that can do four things at once: deploy AI rapidly, preserve human fallback competence, accelerate junior formation, and codify tacit expertise without flattening it into generic templates. In this future, the scarce asset is not merely model access. It is **institutional learning velocity**. McKinsey's labor-transition work and OECD's skill analyses both point in that direction: the premium will rise on organizations that can systematically reskill, redeploy, teach, and preserve judgment rather than merely automate current tasks. [44]

Pilot Design and Recommended Next Steps

The recommended pilot for RUDY AI is a **measurable, exception-rich, cross-functional workflow** with visible outputs and meaningful learning stakes. Good candidates are technical customer support, software engineering, IT service, compliance review, claims operations, and finance operations. These environments are preferable because productivity, quality, escalation, and proficiency are easier to measure, following the logic of the strongest field evidence. [45]

A strong pilot sequence has three phases.

First, run a **diagnostic design sprint** of four to six weeks. Build the task taxonomy, identify developmental versus automatable tasks, define which edge cases should be challenge-routed, and lock the KPI definitions. This phase should also identify where collaboration matters most and where AI may be thinning it.

Second, execute a **24-week controlled rollout**. Start with one role family and one adjacent comparison group. Instrument the workflow, train managers on rationale capture and coaching norms, and pre-register the primary analysis. Do not begin with a companywide deployment; the point is to learn the learning system.

Third, move to **selective scale-up**. Expand only after the pilot demonstrates that efficiency gains are not being purchased by declines in manual exception handling, collaboration density, or expert teaching yield. Scaling criteria should include EPI improvement, not just adoption or throughput.

A practical implementation roadmap is shown below.

Phase	Duration	Deliverables
Diagnostic	4–6 weeks	Skill taxonomy, task-risk matrix, KPI definitions, baseline data pipeline
Controlled pilot	24 weeks	RCT or staggered rollout, manager enablement, ongoing quality audits, monthly capability review
Scale-up gate	2–4 weeks	Go / refine / stop decision based on EPI, throughput, quality, fatigue drift, and ROI
Expansion	2–3 quarters	Additional workflows, integrated skill ledger, enterprise mentor graph, finance tracking

Open questions and limitations

The empirical literature is stronger on **short-run productivity** than on **long-run expertise dynamics**. The customer-support, developer, consulting, and writing studies give credible evidence that AI can change performance quickly, but the causal evidence on what happens to organizations after several years of AI-mediated work is still developing. Some of the most relevant recent evidence remains in working-paper or preprint form, though the direction of findings is already informative. [46]

The evidence on decision fatigue is also more mature in medicine and law than in ordinary enterprise knowledge work. That means generalization should be careful: the strongest claim is not that generative AI has already been proven to create classic decision fatigue in all office settings, but that it predictably relocates effort toward review, exception handling, and accountability—conditions under which fatigue-related distortions are plausible and should be measured directly. [47]

Finally, the literature does **not** support a blanket anti-AI conclusion. Some studies show durable worker learning, faster junior ramp-up, improved customer interactions, and higher creativity under the right metacognitive conditions. The correct conclusion is therefore more demanding: AI should be adopted, but it should be adopted as a **capability-preserving socio-technical system**. RUDY AI Human Expertise Preservation Engine is the product translation of that conclusion. [48]

---

[1] [5] [6] [16] [18] [30] [33] [34] [36] [37] [38] [45] [46] [48] https://academic.oup.com/qje/article/140/2/889/7990658

https://academic.oup.com/qje/article/140/2/889/7990658

[2] [8] [10] https://journals.sagepub.com/doi/abs/10.1177/0018720813501549

https://journals.sagepub.com/doi/abs/10.1177/0018720813501549

[3] [27] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5007084

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5007084

[4] [9] https://mitsloan.mit.edu/press/does-generative-ai-actually-enhance-creativity-workplace

https://mitsloan.mit.edu/press/does-generative-ai-actually-enhance-creativity-workplace

[7] [40] https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/04/bridging-the-ai-skills-gap_b43c7c4a/66d0702e-en.pdf

https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/04/bridging-the-ai-skills-gap_b43c7c4a/66d0702e-en.pdf

[11] https://www.researchgate.net/publication/372281355_Ironies_of_Public_Service_Automation_-_Bainbridge_Revisited

https://www.researchgate.net/publication/372281355_Ironies_of_Public_Service_Automation_-_Bainbridge_Revisited

[12] [23] https://www.sciencedirect.com/science/article/abs/pii/S1071581908000724

https://www.sciencedirect.com/science/article/abs/pii/S1071581908000724

[13] [31] [47] https://pmc.ncbi.nlm.nih.gov/articles/PMC11808891/

https://pmc.ncbi.nlm.nih.gov/articles/PMC11808891/

[14] [15] https://hbr.org/2026/02/ai-doesnt-reduce-work-it-intensifies-it

https://hbr.org/2026/02/ai-doesnt-reduce-work-it-intensifies-it

[17] https://mitsloan.mit.edu/ideas-made-to-matter/generative-ai-changes-how-employees-spend-their-time

https://mitsloan.mit.edu/ideas-made-to-matter/generative-ai-changes-how-employees-spend-their-time

[19] https://gwern.net/doc/psychology/writing/1993-ericsson.pdf

https://gwern.net/doc/psychology/writing/1993-ericsson.pdf

[20] [25] [29] https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf

https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf

[21] https://www.uky.edu/~gmswan3/EDC608/Kalyuga2007_Article_ExpertiseReversalEffectAndItsI.pdf

https://www.uky.edu/~gmswan3/EDC608/Kalyuga2007_Article_ExpertiseReversalEffectAndItsI.pdf

[22] [26] https://www.aft.org/ae/winter1991/collins_brown_holum

https://www.aft.org/ae/winter1991/collins_brown_holum

[24] https://www.sciencedirect.com/science/article/abs/pii/0364021389900025

https://www.sciencedirect.com/science/article/abs/pii/0364021389900025

[28] [32] https://mitsloan.mit.edu/ideas-made-to-matter/how-generative-ai-can-boost-highly-skilled-workers-productivity

https://mitsloan.mit.edu/ideas-made-to-matter/how-generative-ai-can-boost-highly-skilled-workers-productivity

[35] https://pubsonline.informs.org/doi/10.1287/mnsc.2025.00535

https://pubsonline.informs.org/doi/10.1287/mnsc.2025.00535

[39] https://hbr.org/2026/04/dont-let-ai-destroy-the-skills-that-make-your-company-competitive

https://hbr.org/2026/04/dont-let-ai-destroy-the-skills-that-make-your-company-competitive

[41] https://hbr.org/2026/04/why-companies-that-choose-ai-augmentation-over-automation-may-win-in-the-long-run

https://hbr.org/2026/04/why-companies-that-choose-ai-augmentation-over-automation-may-win-in-the-long-run

[42] https://journals.sagepub.com/doi/abs/10.1177/0018720815581940

https://journals.sagepub.com/doi/abs/10.1177/0018720815581940

[43] https://hbr.org/2026/05/employees-are-relying-on-ai-for-personal-support-thats-risky

https://hbr.org/2026/05/employees-are-relying-on-ai-for-personal-support-thats-risky

[44] https://www.mckinsey.com/mgi/our-research/a-new-future-of-work-the-race-to-deploy-ai-and-raise-skills-in-europe-and-beyond

https://www.mckinsey.com/mgi/our-research/a-new-future-of-work-the-race-to-deploy-ai-and-raise-skills-in-europe-and-beyond