
Nobody Gets Promoted for Deleting the Report

*Enterprise AI task inventories sort work into automate and keep.
The third bin, work that should stop, is the one with no artifact,
no owner, and no place in the operating model to record it.*

Gerald C. Kane, Ph.D.

C. Herman and Mary Virginia Terry Chair in Business Administration
Department Head, Management Information Systems
Terry College of Business, University of Georgia

© 2026 Gerald C. Kane. Licensed under CC BY-NC 4.0: share and adapt for noncommercial use with credit to the author; do not present as your own. Full terms on the last page.

Executive Summary

The best nationally representative measurement of what American firms actually do with artificial intelligence was published in April 2026 by economists at the U.S. Census Bureau. Among firms using AI, 44 percent reported that it augmented existing worker tasks, 10 percent that it substituted for tasks previously done by employees, and 11 percent that it created new tasks. And 52 percent reported none of those three effects at all. AI-related employment decreases were reported by 2 percent of AI-using firms, against 2.3 percent reporting AI-driven increases (Bonney et al., 2026).

Sit with the 52 percent. A slight majority of firms that have adopted AI reported none of those three task effects, and among the firms that do report an effect, augmentation alone accounts for about two thirds. The survey categories measure what firms say AI did to their tasks, nothing more. They do not measure speed, and a firm reporting no effect is not necessarily a firm where nothing changed. What the figures support is narrower and still striking: where AI is reported to touch tasks at all, the dominant reported effect is augmentation of existing tasks rather than substitution or new work.

The recommended fix is redesign: rewire the workflow, do not merely bolt the model onto it. That advice is right, and in the frameworks built specifically for allocating tasks to AI it is generally where the road ends. It also skips a question one step earlier. Before a firm decides whether a task should be performed by a machine or a person, it can decide whether the task should be performed at all.

The question is not new, and the field has not exactly forgotten it. Taiichi Ohno named overprocessing as one of the seven wastes of the Toyota Production System in 1978, placing work-that-should-not-exist alongside defects and inventory as a first-class category (Ohno, 1978/1988). Michael Hammer put it in a title in 1990, and the phrase he used for what firms were doing instead, paving the cow paths, describes enterprise AI programs now (Hammer, 1990). Process-redesign practice still carries the lesson: one current McKinsey workflow method makes elimination its first lever, ahead of automation (Maor et al., 2025), and BCG's reinvention guidance estimates that 20 to 30 percent of activities are legacy tasks and tells clients to remove them (Brunelli et al., 2025). Where the question goes missing is one layer over, in the frameworks built specifically for allocating tasks to AI. McKinsey's operating-model method decomposes workflows into activities that shift to AI and activities that remain under human control (Schmitz et al., 2026). The World Economic Forum's workforce projections sort tasks by whether humans, technology, or a combination will perform them (World Economic Forum, 2025). Both are two-bin sorts of existing work, and neither carries a bin for work that should stop. The same advisory firms that teach elimination in their process-redesign practice drop it when the frame becomes AI task allocation. That is the pattern Ohno corrected for production and Hammer corrected for information technology, resurfacing a third time at a third technology's allocation frame.

Three distinct mechanisms are usually run together here, and they have different fixes. **Capability:** can the work be stopped without breaking something downstream. **Measurement:** does the organization have any way to record that a recurring output stopped being produced. **Incentive:** does anyone's career improve if it does. This paper concerns the second and third, and the reason

is that of the three answers to "what should happen to this task," two produce something and one produces nothing. Automating yields a system, a vendor, a dashboard, and a capital request. Augmenting yields a tool, a training program, and an adoption metric. Eliminating yields an absence. That asymmetry is documented. What it explains is offered as this paper's leading explanation and testable hypothesis rather than as a demonstrated cause, because three different absences run through the evidence and they must be kept apart: an absence of published audited elimination cases, which is established by search; an absence of internal instrumentation for stopped outputs, which is documented in kind but not in rate; and a failure by firms to eliminate at all, which no instrument cited in this paper observes. The mechanism argument connects the three. The connection is the hypothesis.

That is sharper than it first appears, because elimination is instrumented perfectly well in some domains. Zero-based budgeting sees eliminated spend, application rationalization counts retired systems, SKU rationalization does it for product lines. Each has a dashboard, an owner, and a promotion path, and each watches a budget or balance-sheet object. Nothing in the mainstream AI task-allocation and benefit-tracking methods we could examine recognizes a stopped knowledge-work output as an owned, aggregated result, and that is where the third bin's contents mostly live. Section IV examines five named cases at that level, as an illustrative scan rather than a systematic review, and none carries an audited outcome. The boundary also runs through one of our own frameworks: the Automate, Informate, Transform taxonomy sorts what an organization does with a task (Kane et al., 2019), and elimination sits upstream of all three.

Agents change the picture, though not in the way the argument usually runs. The intuition is that automating unnecessary work creates supervision burden. The best measured evidence, from public GitHub rather than the enterprise, says something narrower: the share of contributions with any observable human participation was nearly identical for agent-authored and human-authored work, while the composition of the recorded review changed sharply. Among reviewed contributions, human-only review fell from 25.2 percent on human-authored work to 8.1 percent on agent-authored work, and 71.6 percent of review comments on agent work were written by other agents (Duma et al., 2026). Participation in that data is binary, a recorded touch or its absence, so it says nothing about how much attention humans paid. What it does show is that review records have become harder to read as evidence of oversight. A paper about a real thing that no instrument records should notice when a standard measure of oversight becomes harder to interpret, and should hold a GitHub analogy as an analogy.

We also take the strongest objection seriously, and it comes out of the lean literature, not out of AI research. Elimination creates value only if the released capacity has somewhere to go, and Section VI takes up the evidence that it often does not.

Why This Matters Now

Two things make this the moment to raise a question that has been available since 1978.

The first is that task-level decomposition has become a standard enterprise move. Eric Dozier, Executive Vice President and Chief People Officer at Eli Lilly, described the exercise on a July 2026 podcast in terms that will be familiar to anyone running one: a task-level understanding of every job across a workforce of roughly fifty thousand people, then a question about which parts of those jobs

could be, in his word, agentified (Green, 2026). Lilly is about halfway through the inventory and expects to reach the analysis stage in the fourth quarter of 2026. It is a serious, well-sequenced program at a serious company.

The second is a position rather than a finding, and it should be read as one. We were skeptical of agentic AI through late 2024, when vendor claims ran ahead of what the systems could do, and revised that view in spring 2025 when they caught up. Our judgment is that the tools now work well enough that the answer to this question has consequences, and that the organizations learning fastest in a compressed window pull ahead in ways that are awkward to reverse. That is a stated position, not a measured claim, and a reader is entitled to discount it.

The third thing worth saying at the outset is where this argument does not go. It is not a claim that organizations are full of useless jobs. The representative evidence says they are not, and Section VI engages it directly. The claim concerns specific recurring outputs, which is why Section IV spends its length on why those outputs are invisible to the systems that would otherwise catch them.

A note on what the evidence does and does not establish. The two-bin characterization in this paper's subtitle, that AI task inventories sort work into automate and keep, rests on a documented sample of prominent task-allocation methods plus one executive account, not on a survey of how enterprises actually design their inventories. On the method side the sample is small but named: McKinsey's operating-model approach decomposes workflows into activities that shift to AI and activities that remain under human control (Schmitz et al., 2026), and the World Economic Forum's Future of Jobs framework sorts tasks by whether humans, technology, or a combination performs them, with no elimination category (World Economic Forum, 2025). On the practice side there is Dozier's account of his own program, in which the elimination question arrives as a by-product rather than as a designed output. Against these stand the same advisory firms' process-redesign methods, which retain elimination explicitly (Maor et al., 2025; Brunelli et al., 2025). So the defensible claim is narrower than field-wide amnesia: the AI-specific allocation frameworks in this sample omit the elimination bin, the broader process-redesign practice has kept it, and whether the omission carries through to how individual enterprises run their inventories is unmeasured. The Census data cannot settle it either way, since it measures what firms report AI *did* to their tasks, not how inventory exercises are designed. And if the two-bin reading of enterprise practice turns out wrong, the reader should be clear about what falls and what stands: the claim that AI frameworks are where the question dies falls; the artifact asymmetry, the work-versus-people distinction, the capacity objection of Section VI, the managerial-legibility conclusion, and the provisional Stop Inventory stand on their own evidence.

I. What the Numbers Show

Begin with what is measured, not with what is recommended.

The Census Bureau's Business Trends and Outlook Survey is the only nationally representative instrument tracking AI use across American firms, and in April 2026 a team including John Haltiwanger and Lucia Foster published a working paper analyzing its 2026 AI supplement, fielded between November 2025 and January 2026. Eighteen percent of firms used AI in a business function, rising to 32 percent on an employment-weighted basis. Among adopting firms, 57 percent

confined AI to three or fewer functions and 65 percent limited worker-task use to three or fewer task types (Bonney et al., 2026).

The task-effect figures given above need their denominators stated, because the headline is easy to misread. The widely quoted "two thirds only augment" is true of firms already reporting some task effect, roughly half the adopters, and not of AI-using firms generally, where augmentation-only works out closer to a third. This is a working paper rather than a peer-reviewed article, and the analysis is the authors' rather than official Census output, but the underlying data is a government probability sample and there is nothing else of comparable quality in the field.

Consultancy surveys, sampling very differently and skewing toward large enterprises, report a compatible shape. Deloitte's 2026 State of AI in the Enterprise, fielded across 3,235 leaders in 24 countries in August and September 2025, found 37 percent using AI at surface level with little or no change to existing processes (Deloitte, 2026). Deloitte describes its sample as organizations on the leading edge of AI, so this is not an economy-wide figure, and leaders grade their own transformation generously. Both push the same way, which makes 37 percent a floor.

The standard corrective is redesign, and it has real support. McKinsey's March 2025 survey tested 25 organizational attributes against self-reported EBIT impact and reported that "the redesign of workflows has the biggest effect on an organization's ability to see EBIT impact from its use of gen AI" (McKinsey & Company, 2025). The finding is correlational, from a Johnson's relative weights analysis with an R-squared of 0.20, on a self-selected panel, fielded in July 2024. It is the most-quoted number in enterprise AI advisory work and considerably softer than its reputation. On its own terms it points the right way: among respondents whose organizations use gen AI, only 21 percent said any workflows had been fundamentally redesigned.

BCG's 10-20-70 formulation, holding that roughly 70 percent of AI value comes from people and process, deserves a note because it is so often cited as evidence. Across BCG's own publications it is framed three different ways: as an observed split among top performers (Boston Consulting Group, 2025), as a claim about where value originates (Boston Consulting Group, 2026a), and as recommended effort allocation (Boston Consulting Group, 2026b). Those are three different claims, and we could locate no study, sample, or method behind any of them. Sharper still, the three do not agree on what the numbers are attached to: two assign 10 percent to algorithms and 20 percent to data and technology, while the third assigns 10 percent to technology and 20 percent to algorithms and data. It is a serviceable framing and not a measurement.

II. Ohno Named It in 1978, Hammer Titled an Article With It in 1990

There is a version of this argument that claims the elimination question is a discovery. That version is wrong, and the true history is more useful.

Ohno's account of the Toyota Production System organizes waste into seven categories: overproduction, waiting, transportation, processing itself, inventory, movement, and defects (Ohno, 1978/1988). Five of those describe inefficiency inside work that needs doing. Two do not. Overprocessing, the fourth item, names work performed to a standard nobody requires. Overproduction, the first, names output nobody asked for, and Ohno treats it as the most damaging of the seven, because it manufactures the appearance of productivity while consuming real

resources. The distinction between waste inside a necessary task and a task that is itself the waste is not a subtlety we are importing. It sits in the middle of the most widely taught taxonomy in operations management, and has since 1978.

Michael Hammer made the same cut for white-collar work twelve years later and made it the title. "Reengineering Work: Don't Automate, Obliterate" appeared in the July-August 1990 *Harvard Business Review*. His complaint was that firms were using information technology to mechanize old ways of doing business while leaving the underlying processes intact. Two consecutive sentences carry it: "It is time to stop paving the cow paths." And: "Instead of embedding outdated processes in silicon and software, we should obliterate them and start over" (Hammer, 1990).

Substitute artificial intelligence for information technology and the 1990 diagnosis describes the 2026 Census data without adjustment. Hammer's central example was Ford's accounts payable function, which employed more than five hundred people in North America. Benchmarked against Mazda's equivalent operation of five, Ford abandoned an initial 20 percent efficiency target, redesigned so that invoice matching relied on three data items instead of fourteen, and reported a 75 percent headcount reduction. Two caveats belong on those numbers. The 75 percent is a reduction within Ford's North American accounts payable function, not company-wide, and the Mazda comparison was Ford's own benchmarking against a far smaller company, not a size-normalized one. And all of it is Hammer relaying a client engagement from the client's internal analysis; none of it was independently audited. The mechanism is nonetheless unambiguous: Ford did not reach 75 percent by automating invoice matching. It got there by removing most of the matching.

Ford is worth holding onto, because the same company supplies the counter-case in the present decade. In 2026 Ford disclosed that it had hired, promoted, or rehired about 350 veteran technical specialists over three years, many of them retired employees or engineers from key suppliers, after automated quality-control systems failed to catch defects that experienced inspectors had routinely caught. Ford's vice president of vehicle hardware engineering, Charles Poon, said the tools had been trained on the company's design requirements rather than on the judgment of its most experienced engineers, many of whom had left before that knowledge could be used to improve them: "Artificial intelligence is a fantastic tool, but it's only as good as the information you use to train it" (Bloomberg, 2026). Set the two episodes side by side. In 1990 Ford removed work that a supplier relationship had made unnecessary, and the gain held. In 2026 it thinned a population of experts on the premise that a system had made them redundant, and had to rebuild. The distinction between eliminating an activity and eliminating the person performing it runs through the whole of Section IV, and Ford has now been on both sides of it.

The pattern is older than either. Electric power reached American manufacturing in the 1880s and produced no measurable productivity gain for roughly four decades. The gain arrived in the 1920s, when factories stopped retrofitting motors onto existing plants and started building differently. Paul David's explanation is architectural. "Factory structures could be radically redesigned," he writes, "once the need for bracing (to support the heavy shafting and belt-housings for the transmission apparatus that typically was mounted overhead) had been dispensed with" (David, 1990). The gain came from what could be torn out. Line shafting was a structure whose reason for existing had lapsed the moment power stopped needing to be distributed mechanically, and removing it is what let the building be laid out around the work instead of around the power train. By David's own estimate, approximately half the five-percentage-point acceleration in U.S. manufacturing total

factor productivity growth between the 1909-19 and 1919-29 decades is accounted for statistically by the growth in secondary electric motor capacity during that decade.

Brynjolfsson and Hitt reached the same conclusion for the computer era with firm-level data: IT investment pays only in combination with organizational change, the ratio of intangible organizational assets to IT assets may run as high as ten to one, and changing incrementally can create significant productivity losses, because the benefits of computerization are more than outweighed by negative interactions with existing organizational practices (Brynjolfsson & Hitt, 2000). Half-measures are not a slower path to the same place. They can be worse than nothing.

So the position was fully articulated, in three separate research traditions, before most of the people running today's task inventories entered the workforce, and parts of the advisory field still teach it: elimination survives as the first lever in current workflow-redesign methods (Maor et al., 2025) and as an explicit step in reinvention roadmaps (Brunelli et al., 2025). What needs explaining is more specific than amnesia. The question keeps its place in general process redesign and loses it in the frameworks built for allocating work to AI, the third time a new technology's allocation frame has arrived without it. The interesting question is why the elimination bin fails to survive the translation.

III. The Statistic That Buried the Question

Reengineering collapsed within about five years of Hammer's article, hard enough that the vocabulary became unusable. Thomas Davenport, who had helped create the field, published its obituary in the first issue of *Fast Company* in November 1995. His opening line is the whole story: "Reengineering didn't start out as a code word for mindless bloodshed" (Davenport, 1995). By 1995, he wrote, it stood for restructuring, layoffs, and failed change programs.

Two facts about that collapse have been almost entirely lost, and both matter for anyone deciding whether to ask the elimination question in 2026.

The first concerns the number everyone remembers. The claim that 50 to 70 percent of reengineering efforts failed is among the most-cited statistics in the change-management literature. It originates in Hammer and Champy's 1993 book, where the authors describe it as "our unscientific estimate" and where the quantified claim is that this share of organizations "do not achieve the dramatic results they intended" (Hammer & Champy, 1993). In fairness to the folklore, the preceding sentence does say that many companies beginning reengineering "don't succeed at it," so the authors are not wholly innocent of the reading. But the number itself is attached to falling short of dramatic, which is a different thing, and it is self-labelled as not a study.

The distortion has been traced. Grant's 2002 article in *Communications of the ACM* states that approximately 70 percent of reengineering projects fail, citing a 1994 practitioner-journal article, not an outcome study (Grant, 2002). A self-declared unscientific estimate about not achieving dramatic results became, through one restatement, a flat failure rate in a peer-reviewed venue, where it acquired the authority of a finding. Mark Hughes reviewed five separate published instances and concluded that in each case there was an absence of valid and reliable empirical evidence, ending with the judgment that "there is no valid and reliable empirical evidence to

support such a narrative" (Hughes, 2011). We have treated that figure as debunked folklore in prior work and do not repeat it here except as the object of study.

The second lost fact is what the real outcome data said. CSC Index, the consultancy that commercially owned reengineering, surveyed the field in 1994, and Davenport reported the results: of 99 completed initiatives, 67 percent were judged to have produced mediocre, marginal, or failed results, and 73 percent of participating companies said they were using reengineering to eliminate an average of 21 percent of jobs (Davenport, 1995). The second figure is the one Davenport builds his diagnosis on, and the diagnosis is his rather than the survey's: a method for redesigning processes had been turned into cover for headcount reduction, so that "eliminate the work" was in practice read as "eliminate the people." The number shows the substitution happening. It does not by itself prove the substitution is what killed the movement, though Davenport, writing from inside it, argues exactly that.

Which leaves an obvious question: did reengineering, done as designed instead of as practiced, actually work? The econometrics arrived late and say yes. Ozcelik analyzed a data set of large U.S. firms and found that labor productivity, return on assets, and return on equity all improved after reengineering projects were completed, while remaining unaffected during execution. The second finding is the one worth pinning to the wall: functionally focused projects on average contributed more to performance than those with broader cross-functional scope, which the author reads as evidence that failure risk rises past a certain level of scope (Ozcelik, 2010).

Narrow beat broad, associationally but consistently, and it cuts against Hammer's own prescription, which held that reengineering could not be accomplished in small and cautious steps and was an all-or-nothing proposition. The question was right. The method was wrong, and wrong in a specific and correctable way: too big, consultant-led, executed as a transformation program, and eventually a euphemism.

That reading fits a position we have held for a long time and that has nothing to do with AI. Do not make big bets. Start many small experiments, invest in what works, and kill what does not (Kane et al., 2021). It also explains the durability of the amnesia. The wreckage was public enough that a generation of executives learned to treat the underlying question as radioactive, and nobody wants to convene the meeting where a consultant asks which work should not exist. The question was buried with the method's name, on the strength of a number its own authors called unscientific, and where it survives today it survives quietly, as a lever inside workflow-redesign methods rather than as anything a career gets built on.

IV. Elimination Leaves No Artifact

Here is this paper's leading explanation for why the question fails to survive the translation into AI task allocation, and it should be read as a hypothesis with evidence behind its premise rather than as a demonstrated cause. What this section establishes is an asymmetry: elimination leaves no artifact where automation and augmentation leave several, and no published audited case of knowledge-work elimination could be found. What no evidence here establishes is the causal chain, that the missing artifact is why enterprises omit the question; firms may be omitting it for other reasons, or quietly eliminating work in ways neither their instruments nor ours can see. It is worth being precise, too, that this is a claim about measurement and incentives, not about capability.

Whether a given task *can* be stopped without breaking something downstream is a separate question with a separate answer, and Section VII treats it separately.

Consider the three possible answers to "what should happen to this task," and ask what each leaves behind.

Automate it, and the organization acquires a system: a vendor, a contract, an implementation plan, a go-live date, an adoption dashboard, and a capital request with a stated return. Augment it, and the organization acquires a tool and a training program, with a license count, a seat-utilization metric, and a set of before-and-after time studies. Less capital, same essential property. Both decisions produce objects that can be counted and presented, several of them visible to the board.

Eliminate it, and the organization acquires nothing. No vendor, no contract, no adoption curve, because there is nothing to adopt. No dashboard, because the metric is the absence of an activity that formerly had one. No implementation team, and therefore no team lead whose promotion case rests on the outcome. The financial effect is real and shows up, if at all, as a small reduction in a cost line nobody was watching, six quarters later, indistinguishable from ordinary drift.

Managerial courage has little to do with it. The problem is informational, and it maps onto a specific dimension of Digital Congruence, which holds that value stalls when strategy, structure, talent, leadership, culture, and operating model fail to move in the same direction (Kane et al., 2019). The dimension at issue is the **operating model**: the workflow architecture, evaluation infrastructure, and human-in-the-loop boundaries through which a firm actually runs its work. An operating model with no way to record that a recurring output stopped being produced cannot support a decision to stop producing it, however well the strategy is set. A leader who eliminates thirty hours of monthly reporting has created more value than one who automates it, and has less to show for it than one who piloted a tool and produced a slide.

Two objections arrive here, and both sharpen the claim instead of defeating it.

The first is that organizations already know how to instrument elimination. They do. Zero-based budgeting exists precisely to make eliminated spend visible, budgeted, and career-relevant, and it has been run at scale with published targets and named owners. Application rationalization is an established discipline whose central metric is the count of systems retired. SKU rationalization does the same for product lines, sunset review for regulations. Each has a dashboard, an owner, and a promotion path. So the claim was never that elimination cannot be measured. Every one of these instruments watches a budget or balance-sheet object: a cost line, an application, a product, a statute. What is seldom instrumented is a recurring knowledge-work output: the weekly reconciliation memo that goes to nine people, two of whom open it. That is the gap, and it explains a pattern that would otherwise look contradictory. Where elimination attaches to a countable object, the evidence exists, as Section VII shows for process automation. Where it attaches to knowledge work, the published evidence stops. Whether the internal evidence also stops, and whether the eliminating itself stops, are the two further absences this paper infers but cannot observe.

The second is a counterexample to the title, and it demonstrates the mechanism instead of breaking it. Executives absolutely do get promoted for eliminating things. They get promoted for eliminating people. Headcount reduction is instantly visible, arrives in a named line item, and fits on one slide. That is exactly the substitution the 1994 CSC Index data captured. When an organization can see

one kind of elimination and not the other, it will produce the one it can see. The work stays, the people go, and the remaining people absorb the work. Any firm asking the elimination question in 2026 should expect this failure mode to appear.

An illustrative scan of five cases

If the artifact problem is real, we should observe a specific signature: organizations announcing that work should not exist more readily than they demonstrate having removed it and been better off. The five cases below are an illustrative scan, not a systematic review. They were selected as the most prominent and best-documented examples reachable in the research pass, without a pre-registered search protocol or formal inclusion criteria, and a different search would surface a different five. What they can establish is that the pattern exists in prominent cases and that a reader should not assume the evidence base is stronger than it is. What they cannot establish is a rate.

Eli Lilly is at the announcement stage. Dozier described a point past the inventory where the company could ask "what tasks are being done today that don't even need to be done anymore," on the grounds that the business model has moved past them. He went further, naming the sequencing error: "I think too often, we're starting with a people-first process and laying AI over top of it," as opposed to designing an AI-enabled process and placing people into it (Green, 2026). That is a clear-eyed statement from the chief people officer of a company of roughly 50,000 people (Eli Lilly and Company, 2026), and on his own account it is scheduled for the fourth quarter and has not yet happened. (Dozier says "the fourth quarter" without naming a year; in a July 2026 interview the reading is obvious, but the year is ours.) Lilly demonstrates that the question is being asked out loud at the top of a large organization, which is genuinely new. It does not demonstrate that asking it produces anything. Note too that Dozier presents no tidy three-bucket taxonomy: the elimination question arrives as a by-product, framed as something available "even beyond just AI," which is how a question with no artifact behaves.

Bayer offers the most aggressive contemporary version. Chief Executive Bill Anderson, writing under his own byline in *Fortune* in March 2024, described internal rules running to 1,362 pages and twelve levels of hierarchy, and argued that cutting a thousand rules to nine hundred changes nothing, so lasting progress requires eliminating all of them and starting fresh (Anderson, 2024). The Dynamic Shared Ownership program reports a 50 percent cut in hierarchy levels, a two-thirds cut in management positions, and a target of two billion euros in savings between 2023 and 2026 (Bayer AG, 2026a). The outcome data is the problem. In the full-year release of March 2026, in the third year of a program ending in 2026, Bayer states that it expects to realize the targeted savings (Bayer AG, 2026b). Expects. There is no achieved-savings figure and no bridge from positions removed to euros saved, and the program ran under acute distress from glyphosate litigation, which confounds elimination with cost-cutting.

General Electric's Work-Out, begun in late 1987, is the historical precedent most often invoked, and its design was sound: cross-functional groups of 45 to 100 people, each session led by a senior executive who had to participate rather than merely bless it, and a requirement that the leader answer yes or no on the spot to every idea presented, with unnecessary reports, approvals, and meetings as the targets (Kleiner, 2003). The results were not measured. Kleiner reports that nobody knows how many people went through Work-Out, because Welch discouraged record-keeping, and the one figure available is that all but 9 percent of *approved* ideas were followed through, which

measures compliance with decisions already taken and says nothing about the ideas rejected on the spot, let alone about value created. A flagship elimination program that declined to instrument itself is a close to perfect illustration of the artifact problem, and thirty-eight years on there is still no way to check.

Amazon is the boundary case, instructive because it does not qualify. Bezos wrote in the 2017 shareholder letter that the company uses narratively structured six-page memos instead of slide presentations (Amazon.com, 2018). This is frequently cited as work elimination and is not: Amazon substituted a more expensive artifact for a cheaper one on a quality argument, the letter concedes a good memo takes a week or more, and no outcome claim is made anywhere in it. Substitution miscounted as elimination is the most common error in this space.

Zappos is the failure case. In a March 2015 internal memo, Tony Hsieh announced that in order to eliminate the legacy management hierarchy there would effectively be no more people managers as of April 30, offering severance of at least three months to anyone not committed to holacracy (Hsieh, 2015). Zappos told the press that 210 employees took the holacracy-specific offer, about 14 percent of a roughly 1,500-person workforce, and by January 2016 *Fortune* calculated cumulative departures across both buyout offers at about 18 percent. The percentages are reporters' arithmetic on company-supplied headcounts, not audited disclosures, and they do not share a denominator, so they should not be added or compared closely. The memo itself survives only as reproduced by *Quartz*; Zappos never published it. Zappos ran the purest available experiment in eliminating an entire category of work, and what went wrong is specific: the elimination was executed as a structural decree with a deadline, not as the conclusion of an examination of what the work was actually doing. That is the first question asked badly rather than the third asked well. Failure cases are where the mechanism is legible, because practitioner literature and access both skew toward winners, and this one is legible.

Across the scan, one organization has scheduled the question, one has announced a target it has not yet evidenced, one deliberately declined to measure, one performed a substitution regularly miscounted as elimination, and one executed by decree and lost a fifth of its workforce. None published an auditable demonstration that identifying and removing unnecessary knowledge work produced a measured improvement. Read that as a signal about how thin the demonstrated evidence is in prominent cases, not as a measured base rate. Comparable evidence does exist one level down, at the scale of a single process with a countable output, and Section VII presents it. What is missing is the enterprise-level version, run across a task inventory of the kind Lilly is building, with an outcome anyone outside the company can check.

V. What Agents Change

The argument so far would have held in 2015. Agents change it, and the mechanism is not the one usually proposed.

The intuitive version runs like this: automate unnecessary work and you have not eliminated it, you have industrialized it, generating output someone must now review, which generates supervision work, which justifies more automation. It is a tidy story, and the best available measured evidence does not support it.

Researchers have assembled a data set of 932,791 pull requests authored by AI coding agents across 116,211 GitHub repositories and 72,189 developers, spanning five agents: Codex, Devin, GitHub Copilot, Cursor, and Claude Code (Li et al., 2026). An analysis of that corpus examined 33,596 agent-authored pull requests in repositories with more than a hundred stars and found that 61.4 percent carried no recorded review activity (Duma et al., 2026). That figure is widely quoted and needs three qualifications before it can be used at all.

First, "no recorded review activity" means no review comment in the GitHub record. The authors are explicit that this does not imply the code went unreviewed, since a silent approval is indistinguishable from no review in the data.

Second, and more important, the 61.4 percent comes from the full population of repositories containing agent work. The paper's matched comparison, restricted to repositories containing both agent- and human-authored contributions, runs the other way: agent-authored pull requests went unreviewed *less* often than human-authored ones, 28.9 percent against 34.5 percent, and the authors' own summary is that observable human participation is "nearly identical," at 30.1 against 30.8 percent. Quoting the 61.4 percent as evidence that humans have stopped looking, without that baseline, inverts what the study found.

Third, the study controls for neither pull request size, complexity, nor repository workflow, and makes no claim about whether the output was needed, wanted, or read before merging.

What the study does establish is a change in the *composition* of oversight, and it is substantial. Among reviewed contributions in the matched set, human-only review fell from 25.2 percent on human-authored work to 8.1 percent on agent-authored work. Of all review comments on agent-authored work, 71.6 percent were written by other agents. And the character of human involvement shifted: comments that steer the agent rather than evaluate its output rose from 1.6 percent of human review comments on human work to 25.9 percent on agent work. The authors' framing is that review activity no longer corresponds reliably to direct human evaluation, and that conventional review metrics should be interpreted with care as indicators of oversight.

That is a narrower claim than abdication and a more useful one for this argument. What held steady is binary participation, whether any human touch appears in the record at all; the data does not measure how much attention a touch represents, so nothing here shows attention holding constant, and nothing shows it degrading. What changed, measurably, is the composition of the recorded review and therefore what the record can be read to mean: review metrics that once indicated direct human evaluation no longer reliably do. A paper about work that is real but invisible to measurement should notice when a standard measure of oversight becomes harder to interpret, and should resist the temptation to overstate it in the process.

The honest limitation is that all of this is public GitHub, and the right way to hold this section is as a narrow analogy about review composition and metric interpretability in the one setting where agent-authored work is measured at scale. It offers no evidence that enterprise measurement of oversight has degraded. Whether the pattern generalizes to an enterprise finance function or a regulatory affairs group is an assumption, and we state it as one.

A companion study found humans performing 83.2 percent of follow-up commits on agent-generated files, against 93.0 percent on human-generated ones, across 100 repositories over six months (Sawada et al., 2026). Humans are doing most of the downstream maintenance on code

they did not write. Read alongside the composition shift, human effort appears to be moving from the gate toward the cleanup, which is the less useful position, because a reviewer at the gate can still ask whether the change was needed and a maintainer six weeks later cannot.

Three adjacent measurements suggest output volume is rising without anyone being accountable for its consumption, and each carries a caveat that limits what it can support. GitClear reports duplicated code blocks per million changed lines rising 81 percent between 2023 and 2026, from 40.3 to 73.0 across a corpus of 623 million changes, on an uncontrolled time series across a period that also included layoffs and hiring-mix shifts (GitClear, 2026). Microsoft's telemetry found mass emails to twenty or more recipients up 7 percent year over year while one-to-one threads fell 5 percent, which establishes load and not waste, and which Microsoft does not attribute to AI (Microsoft, 2025). And a survey of 1,150 U.S. full-time employees by BetterUp Labs with Stanford's Social Media Lab named the receiving end of it workslop, "AI generated work content that masquerades as good work, but lacks the substance to meaningfully advance a given task," with 40 percent reporting having received it in the prior month (Niederhoffer et al., 2025). That study is vendor-affiliated and self-reported; its dollar extrapolations should not be used, and its defensible finding is attitudinal. Recipients judged senders less creative and less trustworthy afterward, which is the first market signal against unnecessary output this paper has found anywhere.

One theoretical result should temper confidence in task-level decomposition generally. Demirer and colleagues model production as sequences of steps automated within contiguous runs they call chains, and find that comparative-advantage logic can fail under chaining, because the value of automating a step depends on whether its neighbors are automated; occupations whose AI-suitable steps are dispersed among human steps show weak automation despite high exposure (Demirer et al., 2026). The model establishes chain dependence, adjacency effects, and the weakness of atomized task-by-task analysis. The further step, that removing intervening steps is therefore the move, is our inference and not a result the authors derive; consolidation or resequencing would satisfy the same logic.

VI. The Objection That Has to Be Answered

The strongest argument against everything above comes out of the lean literature, from people who spent two decades transplanting exactly this thinking into service organizations and measuring what happened.

Radnor, Holweg and Waring studied lean implementation across four multi-level case studies in the English National Health Service and found that it generally amounted to the application of specific tools producing small-scale, localized productivity gains. Their explanation identifies two breaches of lean's underlying assumptions in health care. The first concerns customer value, which is hard to define where the customer and the commissioner are not the same party. The second bears directly here: health care is "predominantly designed to be capacity-led, and hence there is limited ability to influence demand or make full use of freed-up resources" (Radnor et al., 2012). Their conclusion is that these constraints are organizational and managerial in origin, not professional, which is a diagnosis we would make in the same words.

Apply it directly. Eliminating a task frees capacity, and freed capacity creates value only if it has a destination. In a capacity-led setting, where demand is determined by something other than the

organization's throughput, released hours do not convert into output; they are absorbed by whatever else is queued, or by nothing. A firm that eliminates thirty hours of monthly reporting without deciding in advance what those hours are for will find, two quarters later, that they have been consumed by the ambient expansion of everything else and that no measurable improvement occurred. It will then conclude, wrongly, that the elimination did not work.

We have written elsewhere about the downstream half of this problem, arguing that an efficiency gain and its allocation are two separate managerial events, and we do not re-derive that argument here (Kane, 2026). The two failure modes compound. Eliminate without deciding where the capacity goes and the gain evaporates; decide where capacity goes without asking what should be eliminated and no capacity is released in the first place. The evidence for evaporation points two ways: Suh and Oh measure saved time correlating with output change at effectively zero across 5,512 Korean workers, while Ranganathan and Ye observe capacity absorbed by intensification rather than released at all (Suh & Oh, 2026; Ranganathan & Ye, 2026). They disagree about the destination and agree on what matters here, which is that freed hours do not become value on their own.

One further caution, on a claim this paper deliberately does not make. It is tempting to argue that organizations are full of useless jobs, and there is a well-known book making that case. The best representative evidence contradicts it. Soffia, Wood and Burchell tested Graeber's hypotheses against European Working Conditions Survey data covering 21,536 to 29,784 respondents across three waves and found that the share of EU workers who rarely or never feel they do useful work was 4.8 percent in 2015, 5.6 percent in Anglo-Saxon economies, and declining rather than rising, against Graeber's claim of 20 to 50 percent (Soffia et al., 2022). The defensible claim is much narrower and much better evidenced than the popular one: specific work outputs go unconsumed, and the systems producing them have no mechanism for noticing.

VII. What to Do Instead

The historical record supplies unusually specific guidance, because the last attempt failed in ways diagnosed at the time and confirmed by econometrics fifteen years later. Four rules follow, and one instrument.

Run it narrow. Ozcelik's finding that functionally focused reengineering projects were associated more positively with firm performance than broader cross-functional ones is the most actionable result in the historical literature (Ozcelik, 2010). Three disclosures, since it carries weight here. The study panel covers the 1998 Fortune 1000, 832 firms over 1984–2004, of which only **93** implemented a reengineering project: **37 functionally focused against 56 cross-functional**, so the scope comparison rests on a small subsample. The method is fixed-effects panel regression with industry and firm controls, which makes the result associational rather than causal; there is no instrument and no matched control group. And 1990s reengineering scope and 2026 task-elimination scope are different interventions, so the transfer is an inference. With all three stated, it still contradicts the transformation-program instinct directly. Ask the question of one function, one process, or one recurring artifact, and resist running it enterprise-wide as a program with a name.

Sequence elimination before automation. Telefónica O2's robotic process automation program, one of the best-documented in the academic literature, deployed more than 160 robots handling

400,000 to 500,000 transactions monthly with a twelve-month payback, and the researchers stressed the sequencing: the company eliminated processes that added nothing and simplified what remained before automating anything (Lacity et al., 2015). A systematic review of 83 peer-reviewed studies of robotic process automation reaches the same conclusion at the level of the field, finding pre-automation process quality to be the primary moderator of benefit and poor process selection a leading cause of failure (Khantong & Sriboonlue, 2026).

A caution before this carries more weight than it can bear. The O2 numbers are combined-program results: the payback and the transaction volumes belong to the whole initiative, elimination and simplification and automation together, and no separately audited effect of the elimination step exists in the study. What the case documents is the practice of eliminating first, and the researchers' judgment that the sequencing mattered, which the systematic review echoes at the level of the field. It still speaks to the objection Section IV raised against itself, though only partially: process automation is a domain where outcomes get audited at all, because a process has a transaction count, a cycle time, and a payback period. Whether the lesson transfers to the reconciliation memo and the quarterly deck is the open question, and the honest answer is that nobody has built the instrument that would tell us, for the elimination step even here.

Decide the destination before you eliminate. This is Radnor's capacity-led breach turned into a rule. Before removing a task, name what the freed hours are for, in advance and in writing, at the level of the specific team. Learning, customer contact, cycle-time reduction, and headcount are all legitimate answers. No answer is not a legitimate answer, and it is the one that produces evaporation.

Separate the work decision from the people decision, explicitly and in public. The 1994 CSC Index data is the warning, read through Davenport's diagnosis: 73 percent of firms were using a process-redesign method to cut an average of 21 percent of jobs, and Davenport's argument, made from inside the movement, is that the substitution killed it by making the method untrustworthy to the people whose knowledge it required. The survey shows the substitution happening; the causal reading is his. Anyone running a task inventory in 2026 is asking employees to describe their own work accurately to a process that might conclude the work should stop. They will do that honestly exactly once, and only if the first round demonstrably did not become a layoff. Treat that as a data-quality prerequisite rather than a communications problem.

The Stop Inventory

The artifact problem needs an artifact. If elimination produces nothing the organization can see, the intervention is to build the instrument that makes it visible, and to build it before the first elimination rather than after.

A Stop Inventory is the counterpart to the task inventory, and a firm running one without the other has built half an instrument. It carries one row per eliminated activity, opened at the moment of the decision and closed no earlier than two quarters later. Each row records six things: what was being produced, who consumed it and how that was determined, the recurring hours it consumed and how they were estimated, what those hours were reassigned to, what would have to become true for the activity to be reinstated, and what happened at the two-quarter check. The last two columns are what distinguish it from a cost-savings tracker. The reinstatement condition is a falsifiable

claim, not a justification, and firms should expect some rows to trigger it, because a program that never reinstates anything is not testing its own judgment.

Two design notes, and a reconciliation. Keep it small: one function, roughly ten rows, one owner. Name a consumption test, not a value judgment. "Who read this, and how do we know" is answerable from access logs, distribution lists, and a single question to three named recipients; "is this valuable" is not answerable and will produce a meeting. And this is deliberately an inventory, not another audit, test, or diagnostic; it is distinct from the *Saved-Hours Review*, which examines what happened to capacity after a workflow was made faster, where the Stop Inventory examines whether the activity should have been running at all.

This is a paper-specific construct offered provisionally. It has not been field-tested, it is not part of the established framework set, and it should not be treated as one until it has been run somewhere and reported on. It is a way of generating the evidence the field does not currently have, which is what a firm asking this question in 2026 will have to do for itself.

VIII. What Would Falsify This

Three claims, stated so they can be scored wrong.

By mid-2028, at least one large enterprise will publish a knowledge-work elimination result with an auditable methodology attached, and it will come from a regulated industry where documentation of process change is already mandatory. Regulated firms already own the instrumentation; everyone else has to build it.

The share of AI-using firms reporting no task effect will fall from the 52 percent Census measured for late 2025, and the composition of that fall is what matters. If it drops while the share reporting AI-related employment decreases rises faster than the share reporting broader functional integration, the visible form of elimination is crowding out the invisible one and the 1994 pattern is repeating. The same BTOS supplement carries both items, so the comparison is available. If the two move together, we are wrong.

Within twenty-four months a major consultancy will market a task-elimination offering sold as a transformation program at enterprise scope. If Ozcelik's associational scope finding from 1990s reengineering transfers to this setting, an inference rather than a result, that is the configuration most at risk of failing, and its failure will be read as evidence against the question instead of against the packaging. That is the mistake of 1995 in a new decade.

IX. Making an Absence Legible

There is a version of transformation that shows up as something new: a platform, a capability, a headcount plan, a slide. That version is easy to manage because it is easy to see. The version this paper has been circling shows up as nothing at all: a meeting that stopped, a report nobody writes, a review cycle that quietly ended. It is a real gain of the same standing, the kind Ford's accounts payable and David's factory floors were built on, but an organization cannot manage what it has no way to record.

So the practical work is instrumentation, not exhortation. Nobody needs to be told that unnecessary work is bad. Someone has to build the operating-model capacity to notice an absence, assign it an owner, and let it accumulate into something a person can point at in a review. Until that exists, the elimination question will keep being asked in good faith and answered with nothing.

No organization has demonstrated this at enterprise scale yet, and the honest state of the evidence is that nobody has published the receipts. But the contours are becoming clearer, and the first firm to make a subtraction legible on its own books will have built something more durable than another dashboard.

Methodology Note

This analysis is built from published research, government statistics, first-party corporate disclosure, and executive statements on the record. The empirical spine is a U.S. Census Bureau working paper analyzing a nationally representative probability sample, peer-reviewed econometrics on reengineering outcomes, analyses of a large corpus of agent-authored code contributions, and peer-reviewed evaluation of lean implementation in health care. Consultancy surveys are used where sample and fielding window are disclosed and labeled as self-reported; vendor-affiliated research is identified as such and vendor causal framings are not adopted. Forecasts are excluded from the evidentiary base entirely, on the grounds that a projection is not evidence of a mechanism.

Five limitations bear on the conclusions. First, the two-bin characterization of AI task inventories is supported by a small documented sample of prominent task-allocation frameworks and one executive account, not by direct measurement of how enterprise inventories are designed; the same advisory firms retain elimination in their process-redesign methods, so the claim concerns the AI allocation frame, not the advisory field as a whole. Second, the measurement-and-incentive mechanism is this paper's leading explanation for the missing bin and is offered as a testable hypothesis: the artifact asymmetry and the absence of published audited cases are established, the absence of internal instrumentation is inferred from how benefit-tracking systems are built, and no instrument cited here observes whether firms in fact fail to eliminate. Third, the agent-behavior evidence comes entirely from public open-source repositories, measures binary participation rather than quantity of attention, and its generalization to enterprise settings is an assumption. Fourth, the five cases in Section IV are an illustrative scan without a documented search protocol, so they show that a pattern exists in prominent cases and cannot establish its rate. Fifth, the historical material relies in two places on figures the original authors reported from their own client engagements, and those are labeled in the text.

Where a widely circulated statistic could not be traced to a primary source, we have said so instead of repeating it, and in one case the untraceability is itself part of the argument.

References

Amazon.com, Inc. (2018). *2017 letter to shareholders*. <https://www.aboutamazon.com/news/company-news/2017-letter-to-shareholders>

- Anderson, B. (2024, March 21). Bayer CEO: Corporate bureaucracy is a relic of 19th-century leadership. *Fortune*. <https://fortune.com/2024/03/21/bayer-ceo-bill-anderson-corporate-bureaucracy-19th-century-leadership/>
- Bayer AG. (2026a). *Dynamic Shared Ownership*. <https://www.bayer.com/en/strategy/dynamic-shared-ownership>
- Bayer AG. (2026b, March 4). *Bayer delivers upgraded 2025 ambition; projects solid 2026 focused on strategic priorities*. <https://www.bayer.com/media/en-us/bayer-delivers-upgraded-2025-ambition-projects-solid-2026-focused-on-strategic-priorities/>
- Bonney, K., Breaux, C., Dinlersoz, E., Foster, L., Haltiwanger, J., & Pande, A. (2026). *The microstructure of AI diffusion: Evidence from firms, business functions, and worker tasks* (Center for Economic Studies Working Paper CES-WP-26-25). U.S. Census Bureau. <https://www2.census.gov/library/working-papers/2026/adrm/ces/CES-WP-26-25.pdf>
- Boston Consulting Group. (2025). *From potential to profit: Closing the AI impact gap*. <https://www.bcg.com/publications/2025/closing-the-ai-impact-gap>
- Boston Consulting Group. (2026a, January 26). *What is driving AI value*. <https://www.bcg.com/news/26january2026-what-is-driving-ai-value>
- Boston Consulting Group. (2026b). *How AI leaders create competitive advantage*. <https://www.bcg.com/publications/2026/how-ai-leaders-create-competitive-advantage>
- Brunelli, J., Sissimatos, E., Meissner, N., Cali, S., & Bengtsson, A. (2025, July 11). *End-to-end reinvention unleashes a technology's full potential*. Boston Consulting Group. <https://www.bcg.com/publications/2025/the-roadmap-for-end-to-end-reinvention>
- Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14(4), 23–48. <https://doi.org/10.1257/jep.14.4.23>
- Bloomberg. (2026, June 25). *Ford has been rehiring quality inspectors after AI fell short*. <https://www.bloomberg.com/news/articles/2026-06-25/>
- Davenport, T. H. (1995, November). The fad that forgot people. *Fast Company*, 1. <https://www.fastcompany.com/26310/fad-forgot-people>
- David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *American Economic Review*, 80(2), 355–361.
- Deloitte. (2026). *The state of AI in the enterprise, 2026*. <https://www.deloitte.com/global/en/issues/generative-ai/state-of-ai-in-enterprise.html>
- Demirer, M., Horton, J. J., Immorlica, N., Lucier, B., & Shahidi, P. (2026). *Chaining tasks, redefining work: A theory of AI automation* (NBER Working Paper 34859). National Bureau of Economic Research. <https://www.nber.org/papers/w34859>
- Duma, K., Wróblewski, P., Bobińska, J., Winiarska, J., & Przymus, P. (2026). *These aren't the reviews you're looking for: How humans review AI-generated pull requests* (arXiv:2605.02273). <https://arxiv.org/abs/2605.02273>
- Eli Lilly and Company. (2026). *Key facts*. <https://www.lilly.com/about/key-facts>
- GitClear. (2026). *The maintainability gap: 2026 AI code quality research*. https://www.gitclear.com/the_ai_code_quality_maintainability_gap
- Grant, D. (2002). A wider view of business process reengineering. *Communications of the ACM*, 45(2), 85–90. <https://doi.org/10.1145/503124.503128>
- Green, D. (Host). (2026, July 28). How HR can lead AI and workforce transformation across the enterprise (No. 280) [Audio podcast episode]. In *Digital HR Leaders*. myHRfuture. <https://www.myhrfuture.com/digital-hr-leaders-podcast/how-hr-can-lead-ai-and-workforce-transformation-across-the-enterprise>

- Hammer, M. (1990). Reengineering work: Don't automate, obliterate. *Harvard Business Review*, 68(4), 104–112. <https://hbr.org/1990/07/reengineering-work-dont-automate-obliterate>
- Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*. HarperCollins.
- Hsieh, T. (2015, March 24). *Internal memo: Zappos is offering severance to employees who aren't all in with holacracy* [Memorandum]. Reproduced in *Quartz*. <https://qz.com/370616/internal-memo-zappos-is-offering-severance-to-employees-who-arent-all-in-with-holacracy>
- Hughes, M. (2011). Do 70 per cent of all organizational change initiatives really fail? *Journal of Change Management*, 11(4), 451–464. <https://doi.org/10.1080/14697017.2011.630506>
- Kane, G. C. (2026). *Where the saved hours go* (Working paper, July 2026).
- Kane, G. C., Nguyen Phillips, A., Copulsky, J. R., & Andrus, G. R. (2019). *The technology fallacy: How people are the real key to digital transformation*. MIT Press.
- Kane, G. C., Nanda, R., Nguyen Phillips, A., & Copulsky, J. R. (2021). *The transformation myth: Leading your organization through uncertain times*. MIT Press.
- Khantong, S., & Sriboonlue, P. (2026). Robotic process automation in business process management: A systematic literature review and an integrated framework. *Technologies*, 14(4), 225. <https://doi.org/10.3390/technologies14040225>
- Kleiner, A. (2003). GE's next workout. *strategy+business*, 33. <https://www.strategy-business.com/article/03403>
- Lacity, M., Willcocks, L. P., & Craig, A. (2015). *Robotic process automation at Telefónica O2* (The Outsourcing Unit Working Research Paper Series, Paper 15/02). London School of Economics and Political Science. https://eprints.lse.ac.uk/64516/1/OUWRPS_15_02_published.pdf
- Li, H., Zhang, H., & Hassan, A. E. (2026). *AIDev: Studying AI coding agents on GitHub* (arXiv:2602.09185). <https://arxiv.org/abs/2602.09185>
- Maor, D., Guggenberger, P., & Holzer, A. (2025, August 27). *Want to break the productivity ceiling? Rethink the way work gets done*. McKinsey & Company. <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/want-to-break-the-productivity-ceiling-rethink-the-way-work-gets-done>
- McKinsey & Company. (2025, March 12). *The state of AI: How organizations are rewiring to capture value*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-how-organizations-are-rewiring-to-capture-value>
- Microsoft. (2025, June 17). *Breaking down the infinite workday* (Work Trend Index special report). <https://www.microsoft.com/en-us/worklab/work-trend-index/breaking-down-infinite-workday>
- Niederhoffer, K., Kellerman, G. R., Lee, A., Liebscher, A., Rapuano, K., & Hancock, J. T. (2025, September 22). AI-generated "workslop" is destroying productivity. *Harvard Business Review*. <https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity>
- Ohno, T. (1988). *Toyota production system: Beyond large-scale production*. Productivity Press. (Original work published 1978)
- Ozcelik, Y. (2010). Do business process reengineering projects payoff? Evidence from the United States. *International Journal of Project Management*, 28(1), 7–13. <https://doi.org/10.1016/j.ijproman.2009.03.004>
- Radnor, Z. J., Holweg, M., & Waring, J. (2012). Lean in healthcare: The unfilled promise? *Social Science & Medicine*, 74(3), 364–371. <https://doi.org/10.1016/j.socscimed.2011.02.011>
- Ranganathan, A., & Ye, X. M. (2026, February 9). AI doesn't reduce work—it intensifies it. *Harvard Business Review*. <https://hbr.org/2026/02/ai-doesnt-reduce-work-it-intensifies-it>

- Sawada, S., Shirai, T., Kashiwa, Y., Yamaguchi, K., Iwata, H., & Iida, H. (2026). *To what extent does agent-generated code require maintenance? An empirical study* (arXiv:2605.06464). <https://arxiv.org/abs/2605.06464>
- Schmitz, C., Diedrich, D., Cilento, G., & Fuchs, S. (2026, July 7). *The operating model advantage: Why AI winners are rewiring their organizations*. McKinsey & Company. <https://www.mckinsey.com/industries/industrials/our-insights/the-operating-model-advantage-why-ai-winners-are-rewiring-their-organizations>
- Soffia, M., Wood, A. J., & Burchell, B. (2022). Alienation is not "bullshit": An empirical critique of Graeber's theory of BS jobs. *Work, Employment and Society*, 36(5), 816–840. <https://doi.org/10.1177/09500170211015067>
- Suh, D., & Oh, S. (2026). *Generative AI and the reallocation of time: Productivity, leisure, and fulfilling work* (arXiv:2602.12695). <https://arxiv.org/abs/2602.12695>
- World Economic Forum. (2025, January 7). *The future of jobs report 2025*. <https://www.weforum.org/publications/the-future-of-jobs-report-2025/>

Copyright, reuse, and citation

© 2026 Gerald C. Kane. This paper is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), <https://creativecommons.org/licenses/by-nc/4.0/>.

You may share and adapt it for noncommercial purposes, provided you credit Gerald C. Kane as the author, link to the license, and indicate any changes you made. You may not present this material, or adaptations of it, as your own work, and you may not sell it or use it in paid training or consulting without written permission.

Cite as: Kane, G. C. (2026). *Nobody Gets Promoted for Deleting the Report*. <https://www.profkane.com/papers/deleting-the-report.pdf>

Permissions and commercial use: jerry@profkane.com