The Discontinuity Thesis

A Sequence of Seven Essays on Why Postwar Capitalism Ends

Ben Luong

May 2026 (v1.1.2 final)

Contents

The Discontinuity Thesis

A Sequence of Seven Essays on Why Postwar Capitalism Ends

Table of Contents

How to Read This Sequence

Each essay closes one comfortable answer to the question of how the wage-demand circuit might be saved. They are designed to be read in order, but each is standalone. The architecture is cumulative.

Unit Cost Dominance establishes the empirical condition. Interface Collapse establishes the propagation mechanism. The Multiplayer Prisoner's Dilemma establishes that the propagation cannot be stopped by any actor. The Sorites Collapse Principle and Categorical Recursion close the regulatory route at both axes. The Successor System closes the structural-alternatives route. Drag Is Not Rescue closes the friction-and-timing route.

The reader who finishes all seven is left with the actual question: not whether postwar capitalism survives, but what replaces it.

Appendix I specifies what would refute the thesis. Appendix II catalogues the most common patterns that get mistaken for refutation. Appendix III consolidates the empirical evidence the structural argument is anchored in. Together they define the standard at which the argument can be debated and the documentary record on which it stands.

Scope, claims, and measurement

This section is a single-page statement of what the thesis claims, what it does not claim, and what indicators would track whether the claims are correct.

What the thesis claims. Postwar capitalism rested on a wage-demand circuit in which mass productive participation generated the income that sustained mass consumption, which in turn sustained mass productive participation. The thesis claims this circuit is no longer self-reproducing, because the scarcity of general-purpose cognitive labour that anchored mass wage absorption has been removed by AI plus verification at sufficient scale. The thesis is structural rather than predictive. It identifies a mechanism that has been triggered, not a timeline within which any specific consequence arrives.

What the thesis does not claim. It does not claim that all human cognitive work disappears. It does not claim that all jobs vanish. It does not claim immediate or universal mass unemployment. It does not claim that AI is morally good or bad. It does not claim a specific successor system. It does not claim that redistribution is wrong. It does not claim that AI has eliminated the value of human judgment, embodiment, trust, care, ritual presence, or institutional accountability. It claims that the kind of cognitive work that supported middle-class wage absorption is no longer scarce at the scale required to sustain mass productive participation as the primary route to economic agency.

Where the claim applies first and most strongly. Digitally mediated cognitive work in competitive labour markets in technology-leading economies. Less directly: regulated industries with documentation and verification work; legacy-dense professional services; entry pathways to senior cognitive roles. Not directly: embodied physical work, deep interpersonal care, work where human presence is itself the deliverable, niche artisanal production. Geography modulates timing. Sectoral structure modulates timing. The thesis is about the structural attractor, not the local trajectory of any specific market.

Indicators that would track failure of the wage-demand circuit. The thesis treats the circuit as failing when wage labour stops being the primary route to economic agency for the majority of working-age adults. The candidate indicators are wage share of national income; share of working-age adults whose primary income is wages or salaries from labour; junior-to-senior employment ratio in cognitive sectors; entry-level hiring rate in AI-exposed occupations; real median wages indexed to productivity (the Brynjolfsson productivity-pay gap); and the share of household consumption financed by wage income rather than by transfers, capital income, or debt. No single indicator is decisive. A trajectory across multiple indicators in the predicted direction is the empirical signal. A sustained reversal across the same indicators would constitute the empirical refutation specified in Appendix I.

Refutation conditions in summary. The thesis is conditional on four premises. Unit cost dominance crossed for general cognitive labour (Premise One). Workflow-level propagation through interface collapse (Premise Two). Coordination impossibility preventing restraint (Premise Three). Structural alternatives unable to reconstitute the circuit (Premise Four). Refutation of any single premise would constrain the thesis. Refutation of Premise One requires sustained capability-trajectory reversal across independent measures. Refutation of Premise Two requires deployment to reverse at scale sufficient to restore mass productive necessity. Refutation of Premise Three requires demonstration of stable cooperative restraint under the actual political-economy conditions. Refutation of Premise Four requires a structural alternative that reconstitutes mass productive necessity rather than substituting for consumption. The refutation conditions are specified in Appendix I in detail.

What the thesis is asking the reader to do. Hold the structural claim as the load-bearing point. Read the body essays for the mechanisms. Read Appendix III for the documentary anchor. Read Appendix I for the refutation conditions. If the trajectory across the named indicators reverses, the thesis is wrong and should be revised. If it does not, the policy debate should move on from circuit-defence to successor-system design.

1. Unit Cost Dominance

The Empirical Foundation of the Discontinuity

There is a specific point in the development of any general-purpose technology at which it stops being a tool and starts being a substitute. The point is not technological. It is economic. It arrives when the technology, plus whatever human oversight is required to make it reliable, performs a task at lower unit cost and equal or better quality than the human worker who previously performed that task. Before this point, the technology augments. After this point, the technology replaces.

For general-purpose cognitive labour, that point has now been reached for a substantial and growing fraction of professional tasks. This is the empirical foundation of the Discontinuity Thesis. Everything that follows depends on it. The claim is not that AI can do everything. The claim is that AI plus a thin layer of human verification can do enough professional cognitive work, at low enough unit cost, to make standalone human performance economically uncompetitive across a significant portion of the knowledge economy. Once that condition holds, the rest of the thesis follows from the structure of competitive markets.

This essay establishes the condition. It is the entry point to the sequence.

What unit cost dominance actually means

The technical claim is precise. For a given cognitive task, let the human cost be the wage required to compensate the worker for completing the task. Let the AI cost be the inference cost plus the cost of the human verification and integration required to produce a usable output. Unit cost dominance occurs when the AI-plus-verifier cost is lower than the human-only cost, at equal or better output quality.

This is not an automation claim in the ordinary sense. It does not require AI to operate without human involvement. It only requires that the human role shrink to a verification function that costs less than the original production function. A senior lawyer reviewing AI-drafted contracts is cheaper than a junior lawyer drafting the same contracts from scratch. A financial analyst confirming an AI-generated pitch deck takes less time than building the deck from raw data. A software architect approving AI-written code takes less time than writing the code line by line. In each case, the human is still in the loop. The human role has changed from producer to verifier. The economic position of the worker who used to produce the output has changed from necessary to redundant.

The condition has three components, and the thesis needs each of them stated cleanly.

First, the quality condition: AI output meets or exceeds the quality of human output for the task. Second, the cost condition: the raw cost of producing the AI output is far below the cost of human-only production. Third, the workflow condition: when oversight, verification, and integration costs are added, the total still remains below human-only production. Unit cost dominance is the joint satisfaction of all three. Critics frequently grant the first two while disputing the third. The third is where the economic argument actually lives.

The quality condition has been crossed

The empirical anchor for the quality condition is GDPval, the OpenAI benchmark that tests AI systems against industry professionals across forty-four occupations. The benchmark spans 1,320 specialised tasks with a 220-task gold subset, based on real professional work products and graded by experienced practitioners through blind pairwise comparison.[1] The methodology is unusually robust for an AI evaluation. The tasks are sourced from professional reality rather than synthetic puzzles. The graders are domain experts. The metric is whether the AI output is preferred to or tied with the human output.

The trajectory has been rapid. In the first GDPval release in September 2025, Claude Opus 4.1 matched or exceeded expert human deliverables on 47.6 percent of the gold subset, with GPT-5 high at 38.8 percent and earlier models trailing further behind. By April 2026, GPT-5.5 scored 84.9 percent wins-or-ties on GDPval, with GPT-5.4 at 83.0 percent and Claude Opus 4.7 at 80.3 percent.[2] [3] Gemini 3.1 Pro reached 67.3 percent. The frontier has crossed expert parity on the benchmarked class of digital professional deliverables.

The relevant fact is not a single model ranking. It is that expert parity on benchmarked digital professional deliverables is no longer a future threshold. It has been crossed, by multiple frontier systems, in a span of months. For the benchmarked tasks, the quality condition for substitution has been crossed. Whether a task remains augmentation or becomes substitution now depends on verification and integration cost, not on model capability alone.

GDPval does not prove occupation-level automation. It proves that, for a large benchmarked class of self-contained digital professional deliverables, the quality condition for substitution has been met. The economic condition is met wherever the cost of inference, integration, and human verification remains below the cost of human-only production. That is the threshold the thesis calls Unit Cost Dominance.

The benchmark's limitations are part of the precision

GDPval is strong evidence within its scope. It is also bounded. The benchmark covers 44 occupations and 1,320 tasks, with a 220-task gold subset, focused on knowledge work that can be performed on a computer. It excludes manual labour, physical tasks requiring extensive embodied judgement, work that depends heavily on tacit knowledge, deployment requiring proprietary software access, situations involving personally identifiable information, and roles where interpersonal communication is itself the deliverable.

[1] OpenAI, "Measuring the performance of our models on real-world tasks," GDPval. https://openai.com/index/gdpval/
[2] OpenAI, "Introducing GPT-5.4." https://openai.com/index/introducing-gpt-5-4/
[3] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/

This limitation does not weaken the thesis. It scopes it. The claim is not that all labour is immediately substitutable. The claim is that the benchmarked domain overlaps heavily with the cognitive work that sustained middle-class absorption in the postwar economy. Once that domain crosses parity, the wage-demand circuit loses its central scarcity. The work that remains outside the benchmark is real, and some of it will resist substitution for longer. The exceptions exist. They do not absorb the displaced population at the scale required to preserve a wage-demand circuit, which is the question the rest of the sequence addresses.

The cost condition

The raw cost gap between AI inference and human professional labour is large. OpenAI's own GDPval discussion notes that frontier models can produce comparable deliverables at roughly one hundred times the speed and one hundred times the cost reduction relative to expert human labour, but explicitly cautions that these figures are based on pure model inference time and API billing rates, excluding the human oversight, iteration, and integration steps that real workplace deployment requires.[4]

This caveat matters, and it is also where the strongest version of the thesis lives. The model-only benchmark understates deployed quality, because real firms do not deploy raw model output. They deploy AI plus a verifier. The model-only cost benchmark overstates the deployed cost advantage, because verification adds cost. Both corrections point toward the same economic result. AI plus verifier is ruinous for standalone human production unless verification recreates the entire original job.

The arithmetic is worth being explicit about. Normalise the old human-only task cost to 100. Assume the raw AI inference cost is 1. The total deployed cost depends on how much human verification time is required as a share of the old production cost.

Verification share of old production cost	Total AI + verifier cost	Cost advantage over human-only
5%	6	16.7x cheaper
10%	11	9.1x cheaper
20%	21	4.8x cheaper
30%	31	3.2x cheaper
50%	51	2.0x cheaper
70%	71	1.4x cheaper

The human-only producer only becomes competitive again when verification, integration, and failure-handling consume nearly the entire original task cost. The critic must argue that the verifier is essentially doing the old job again. If that is true, AI has not crossed unit cost dominance for the task, and the thesis does not apply to it. Where verification is materially thinner than production, the standalone human worker is economically dead. The burden shifts to the critic to show that verification recreates the old job rather than compressing it.

[4] OpenAI, "Measuring the performance of our models on real-world tasks," GDPval. https://openai.com/index/gdpval/

This is the workflow condition stated in numbers. It is what makes the thesis economic rather than merely technical.

Why the augmentation framing breaks

The standard response to these numbers is that AI augments human workers rather than replacing them. The argument runs as follows: AI handles routine cognitive work, human workers move up the value chain to focus on judgement, creativity, and strategy, and total productivity rises while employment is preserved. This is the augmentation narrative, and it has been the dominant frame in policy discussion for the last several years.

The augmentation narrative was correct for previous waves of automation. It is wrong for this one, and the reason it is wrong is structural rather than empirical.

Previous automation waves automated specific bounded functions. The factory automated muscle. The computer automated arithmetic. The internet automated distribution. In each case, human cognition remained the bottleneck for the activities the new technology could not perform. Workers displaced from automated functions could move to functions where human cognition was still scarce and valuable. The wage premium that knowledge work commanded was preserved by the cognitive bottleneck, which the prior technologies did not address.

AI automates enough digitally mediated, economically valuable cognitive work to remove general cognitive labour as a mass scarcity. There is no higher rung with mass absorption capacity for displaced cognitive workers to climb to. The technology that displaces them is the same technology that would have to be deployed at any higher rung. A junior analyst displaced by AI cannot escape into senior analysis at scale, because senior analysis is now also performed by AI plus a verifier. The verifier role exists, but it does not absorb the displaced population, because one verifier can supervise the output that previously required a team of producers. The structural feature that allowed previous automation waves to preserve mass employment, namely the existence of cognitive work the new technology could not do at scale, is absent in this case.

This is why the augmentation framing fails. It assumes a higher rung with mass absorption capacity. There is no such rung. AI now operates across the cognitive layers that supported middle-class wage absorption, and it improves with each model generation, which means the position of any putative human refuge is unstable. The senior cognitive work that AI cannot quite do this year is the senior cognitive work AI does next year. The augmentation phase is real, but it is a corridor rather than a destination, and the corridor narrows with each release cycle.

The verification trap

The cost asymmetry between generation and verification is not incidental to this argument. It is the asymmetry the wage circuit was anchored in.

For most cognitive work, generation has historically been more expensive than verification. Producing a competent contract, draft, analysis, or piece of code took years of training and hours per output. Reviewing

the same output took an experienced reviewer minutes. Generation was the expensive side of the asymmetry. The labour market priced generation skill because that was where the scarcity lived. Verification existed but did not command the same premium, because verification was the cheaper side.

This is the same structural shape computer scientists describe as NP-shaped: solutions are hard to produce but easier to check. The wage scarcity that built middle-class cognitive labour was the scarcity of competent generators in an NP-shaped cognitive economy. Verification was less scarce, so verification commanded a smaller share of the wage premium.

AI inverts the asymmetry. Generation collapses toward zero marginal cost. Verification becomes the binding constraint. The asymmetry has not disappeared. It has flipped. The structural consequence is that the verification side, which historically required fewer people because it was the cheaper task, cannot absorb the displaced producers who were on the side of the asymmetry where the labour was. One verifier was sufficient to check many producers when generation was hard. That ratio does not invert when the asymmetry inverts. The verifier population that was sufficient to check generators is by construction insufficient to replace the population of displaced generators.

There is one apparent escape from this argument. Even if AI does the production work, humans are needed to verify the output. The verifier role might absorb the displaced producer population. This is the optimistic frame for the transition: a redistribution of labour from production to verification, with verification scaling to absorb everyone who used to produce.

The arithmetic does not support this. If verification took the same time as production, there would be no cost saving and no reason to deploy AI in the first place. The economic case for deployment depends on verification being cheaper than production. A typical pattern is that one verifier supervises the output of multiple AI instances. The exact ratio varies by domain, but the direction is the point: verification absorbs fewer workers than production displaced.

There is a further problem. Much verification work is itself amenable to AI. Once a system can produce a contract, it can also evaluate a contract against a checklist of standard clauses. Once it can write code, it can also catch syntactic errors and common antipatterns in code. The symmetry is not complete. Evaluation is genuinely harder than generation in some domains, because catching what is missing from an output is structurally different from producing the output. False negatives in legal review, missing edge cases in code, and unstated client preferences in proposals are all places where human verification retains real value because the failure mode is latent rather than visible.

The verification trap therefore operates faster in some domains than others. It operates fastest where failures are detectable from the output itself: code that does not compile, summaries that omit named entities from the source, contracts that contradict their own clauses. It operates more slowly where failures require external context, institutional memory, or judgement about what should have been included but was not. Even where verification retains meaningful value, the role is not a stable destination for displaced workers at scale. It is a smaller layer working at a higher level of abstraction, and the next model generation typically thins it further.

This is the verification trap. The role exists. It does not scale. It does not last at the same intensity. Workers who reposition into verification are repositioning onto a layer that the technology is in the process

of compressing. The verifier of this year's models is the producer of next year's, and the verifier of next year's models is a smaller population working at a higher level of abstraction, with the same dynamic playing out one rung up.

What about the work AI cannot do

The honest version of this objection is that some cognitive work resists AI substitution. Care work involving deep emotional attunement. Physical work requiring embodied judgement in unstructured environments. Trust-bearing roles where the human presence is the point. Creative work at the highest levels where originality is the criterion. Work that requires accountability under conditions of legal liability.

These categories are real. They are also smaller than the displaced population. Care work at the scale of mass labour absorption requires either dramatic increases in the wages society is willing to pay for it (which would require political choices that have not been made and probably will not be made) or a structural shift in how care is delivered and paid for (which is a successor-system question rather than a continuity question). Embodied physical work is bounded by the rate of robotics improvement, which is slower than software but not stationary. Trust-bearing roles depend on cultural conventions that are themselves under pressure as AI-mediated interactions become normal. High-end creative work is a small market that does not absorb mass displacement. Liability-bearing roles are partially preserved by law, but the scope of those roles shrinks as AI judgement becomes legally admissible.

The claim is not that all human cognitive work disappears. The claim is that general-purpose cognitive labour loses its role as the mass scarcity that supported middle-class absorption. The exceptions exist. They do not preserve a circuit. A wage-demand circuit requires mass absorption, not islands of residual scarcity. The exposed population is too large for residual niches to absorb. The arithmetic does not close.

The arithmetic in deployment

The arithmetic table is illustrative. The deployment data is evidential. As of mid-2026, several documented cases anchor the verifier-cost arithmetic in production economics rather than in projection.

Anthropic's published case study with Novo Nordisk reports that clinical study report production, which previously required up to fifteen weeks of work coordinated across forty to fifty professionals, can now be completed in minutes by a team of three using the NovoScribe platform built on Claude.[5] Resource requirements for device verification protocols fell by ninety-five percent. Patient documentation that required months of work with external agencies now generates in under a minute. Reviewers report that automated outputs increasingly meet the quality bar that previously required human authorship, with the platform receiving positive feedback from regulators in a heavily regulated industry.

The pattern at OpenAI's own organisation is similar. As of the GPT-5.5 launch in April 2026, OpenAI reports that more than 85 percent of the company uses Codex weekly across functions including software

[5] Anthropic, "Novo Nordisk accelerates clinical documentation and drug development with Claude." https://claude.com/customers/novo-nordisk. Additional figures from AWS's case description of the same deployment, which reports clinical study report production reduced from up to fifteen weeks coordinated across forty to fifty professionals to minutes by a team of three.

engineering, finance, communications, marketing, data science, and product management.[6] In Communications, the team built and validated a Slack agent so that low-risk speaking requests are handled automatically while higher-risk requests route to human review. In Finance, Codex reviewed 24,771 K-1 tax forms totalling 71,637 pages, accelerating the task by two weeks compared to the prior year. On the Go-to-Market team, an employee automated weekly business report generation, saving five to ten hours per week.

The Communications example is the cleanest illustration of the structural pattern. Low-risk work is automated. Higher-risk work goes to human review. This is not stable mass complementarity. This is the verification architecture in operation. The production layer is compressed. The verification layer is retained. The number of humans required falls by an order of magnitude.

Human oversight is not the rebuttal to substitution. It is the substitution architecture. The human remains at the review point while the production path beneath them is compressed. This is what the verifier-cost arithmetic predicts. It is now what the deployment data shows.

The competitive consequence

Once unit cost dominance is established for a task, the competitive logic is automatic. A firm that continues to use human-only production for that task pays higher costs for equivalent output. The firm's competitors who deploy AI plus verification produce the same output more cheaply. The market price falls toward the AI-plus-verifier cost. The human-only firm either matches the lower cost (which requires deploying AI and reducing the human-only workforce) or loses market share until it exits the market. There is no third option that preserves both the human-only production model and the firm's competitive position in contestable markets.

This logic does not require any actor to be enthusiastic about AI deployment. It does not require executives to want to fire workers. It does not require boards to prioritise margin over employment. It only requires that markets remain competitive. Any firm that tries to maintain human-only production above the AI-plus-verifier cost is selected against by the market, regardless of the firm's preferences.

This is why the augmentation phase is unstable. Augmentation is the period during which firms have deployed AI but have not yet reduced their workforce, because the verification protocols are still being learned, the legal liabilities are still being clarified, and the workforce reduction is politically expensive. The augmentation period ends when the firm's competitors have completed their workforce reductions and the firm faces a choice between matching them and losing market share. The competitive pressure is not metaphorical. It is the same pressure that has driven every previous wave of cost reduction in capitalist economies. AI is unusual in the breadth of cognitive work it covers, not in the economic logic of its deployment.

[6] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/

The bridge to the next essay

GDPval is task-level evidence. The standard defence is that tasks are not jobs. That defence held while tasks remained locked inside human-operated workflows. It weakens once AI can operate the interfaces that compose the workflow itself. The cost-quality crossover begins at the task layer, but it propagates upward through interface collapse. The worker was not only a producer of cognitive output. The worker was the integration layer between software systems. Once models can produce the output and operate the interfaces through which the output moves, task-level unit cost dominance becomes workflow recomposition.

That propagation is the subject of the next essay.

What this essay establishes

Unit cost dominance is the technical foundation of the thesis. The claim is empirical and benchmarked. AI plus verification performs a substantial and growing fraction of professional cognitive tasks at lower cost and equivalent quality compared to human-only production. The fraction has crossed expert parity on the benchmarked digital domain and is moving deeper into the rest with each model generation. The augmentation framing fails because there is no higher cognitive rung with mass absorption capacity for displaced workers to occupy. The verifier role does not scale to absorb the displaced population. The exceptions to AI substitution are real but too small to preserve a wage-demand circuit.

This is the floor on which the rest of the thesis sits. The next essay establishes that task-level unit cost dominance propagates upward through workflow recomposition once AI can operate the interfaces between software systems. After that, the Multiplayer Prisoner's Dilemma establishes that no actor can unilaterally restrain the propagation. The Sorites Collapse Principle and Categorical Recursion close the regulatory route. The Successor System shows that even structural alternatives outside regulation preserve consumption rather than the wage-demand circuit.

Each essay closes one comfortable exit. This one establishes that the exits are needed, because the underlying technological condition has been reached and is not reversing. The wage-demand circuit cannot survive on the assumption that AI will not become economically competitive with human cognitive labour. That assumption has been falsified.

What follows is what to do about the falsification.

---

2. Interface Collapse

How Task-Level Dominance Propagates Into Workflow Recomposition

The previous essay established that AI plus verification has crossed cost-quality parity for a large class of professional cognitive tasks. The most respectable continuity defence accepts this and answers it with a structural distinction. AI can do tasks, the argument runs, but tasks are not jobs. Jobs are bundles of judgement, relationships, context, and physical-world friction. AI can write a paragraph or fill a spreadsheet, but it cannot run a legal practice, manage a team, close a deal, or sit in a meeting. So jobs, the argument concludes, are safe even where individual tasks are not.

This was the most thoughtful version of the continuity story. It came from labour economics, productivity research, and business-school augmentation theory. It held up for a while because it was true in the same way a gate is true. It was true until enough of it came off the hinges.

This essay closes the gate. The thesis does not need AI to replace whole occupations in one step. It never did. It needs only enough economically valuable task-units to cross the cost and quality threshold that firms start rebuilding workflows around machines instead of people. A job is not a sacred object. It is a bundle of tasks. Once the bundle becomes decomposable, the worker becomes decomposable with it.

What makes the bundle decomposable is interface collapse.

The hidden moat

A great deal of white-collar labour was historically protected by software fragmentation. The worker did not only think. The worker moved. They moved between the inbox and the spreadsheet, the CRM and the browser, the calendar and the document editor, the dashboard and the ticketing system, the codebase and the internal wiki, the procurement portal and the finance system, the project tracker and the customer database. Humans survived because they were the integration layer between systems that did not talk to each other cleanly.

This was a real moat. It was also invisible to most people who held it, because the labour of stitching looked like the labour of thinking. A surprising fraction of so-called knowledge work was not pure judgement. It was moving information between semi-compatible systems while maintaining enough context not to break the process. The worker absorbed the friction that the software stack produced, and the wage compensated the absorption.

The moat is now collapsing. Once an AI can see screens, click, type, browse, call tools, manipulate files, and persist across a multi-step workflow, the boundary between producing output and doing work begins to dissolve. The model is no longer a text generator on the other side of a keyboard. It is an interface operator.

The API automates where systems cooperate. The GUI automates where they do not. Few ordinary enterprise software environments can rely on fragmentation as a durable moat once both routes exist. Some systems are air-gapped, permissions-heavy, bespoke, legally restricted, or physically entangled, and these will resist for longer. But for ordinary white-collar workflow friction, fragmentation has stopped being a general defence.

The benchmark trajectory

The interface capability is being tracked publicly. OSWorld-Verified is the standard benchmark for desktop computer use, measuring whether a model can navigate real software through screenshots, keyboard, and mouse actions to complete multi-step tasks across applications. The trajectory across recent frontier releases is steep. GPT-5.2 reached 47.3 percent on OSWorld-Verified. GPT-5.4 cleared 75.0 percent on the same benchmark while pushing GDPval to 83.0 percent.[7] GPT-5.5 then hit 78.7 percent on OSWorld-Verified, 84.9 percent on GDPval, 84.4 percent on BrowseComp, and 98.0 percent on Tau2-bench Telecom, which tests complex multi-turn customer-service workflows.[8]

On this benchmark, frontier models now exceed the reported human baseline of 72.4 percent. The trajectory crosses human capability and continues upward.

These numbers describe a specific capability shift. The model is no longer producing isolated outputs that a human pastes into software. It is operating the software layer itself, across applications, over multiple steps, with persistence and tool use. The work that used to require a human to move between systems now requires only a human to authorise the start and verify the end.

The frontier has moved from output generation to work continuation. GPT-5.5 is not described by OpenAI as a model that answers prompts. It is described as a model that plans, uses tools, checks its work, navigates ambiguity, and keeps going. That is the cognitive loop previously sold as the human advantage.

The phase change between model generations matters here. GPT-5.2 demonstrated expert-level professional deliverables at machine cost. GPT-5.5 demonstrates persistent computer-use and workflow execution. The first is unit cost dominance at the deliverable layer. The second is interface collapse in operation. Both are documented in the same vendor's launch notes within four months of each other.

OpenAI's own description of GPT-5.5 captures the propagation directly. The model is described as able to research, analyse, build documents, operate software, move across tools, check its work, and push through messy multi-part tasks.[9] That is not a description of a text generator. It is a description of a workflow operator. The taxonomy of what counts as AI capability has shifted from output generation to process execution within the span of two model releases.

[7] OpenAI, "Introducing GPT-5.4." https://openai.com/index/introducing-gpt-5-4/
[8] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/
[9] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/

Deploying CEOs describe the same shift in their own organisations. Cursor's CEO reports that GPT-5.5 "stays on task for significantly longer without stopping early, which matters most for the complex, long-running work our users delegate."[10] MagicPath's CEO reports merging branches with hundreds of frontend and refactor changes into a substantially-changed main branch in twenty minutes. The pattern is consistent: agentic delegation, long-horizon execution, work continuation rather than output assistance.

The reliability comparison

The standard objection to interface-level deployment is that benchmark scores do not translate to production reliability. The most common form of the objection treats the benchmark percentage as a per-step success rate and compounds it across multi-step workflows. A 78 percent per-step rate, exponentiated across a 10-step workflow, would yield roughly 8 percent end-to-end success.

The compounding-error objection misreads the benchmark. OSWorld-Verified is not reporting per-step accuracy. It is reporting end-to-end task success on multi-step computer-use tasks involving real applications, file operations, and workflows that span multiple programs. Treating the score as a per-step probability and exponentiating it double-counts the compounding problem. The benchmark already contains the compounding problem. The 78.7 percent figure is the rate at which the model completes the entire task successfully, not the rate at which any individual step succeeds. GPT-5.4 reaching 75.0 percent on the same benchmark already exceeded the published human baseline of 72.4 percent. GPT-5.5 at 78.7 percent extends the gap.

Production deployment adds distribution shift, latency, permissions, audit trails, retry logic, escalation paths, and verifier cost on top of the benchmark performance. Those are real costs. They enter the unit cost dominance equation as deployment costs. They do not refute the interface capability the benchmark measures.

The production comparison is therefore not raw AI versus ideal human. It is AI plus retries, guardrails, escalation, and verification against human workers plus training, supervision, review, error correction, management, and quality control. The relevant question is not whether AI is perfect. It is whether the full AI stack reaches equivalent reliability at lower cost than the full human stack. For an increasing fraction of professional digital workflows, the answer is yes.

The four layers

The clean way to think about the propagation is as a cascade, not a single event.

The first layer is task-level unit cost dominance. AI plus thin human oversight produces professional task outputs at equal or better quality, faster speed, and lower marginal cost. This has crossed for a large and growing set of well-specified cognitive deliverables. The previous essay established this layer.

The second layer is interface and workflow dominance. AI operates through the same software environments where work happens, and handles the stitching between them. This is rapidly crossing, and the

[10] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/

public benchmarks now track it directly. This essay establishes this layer.

The third layer is job-level dominance. Whole roles become economically unnecessary as enough human task volume is stripped out. This is partial and uneven across occupations, and the thesis does not require it. Some jobs will retain residual human content for years. The thesis works whether or not any particular job is fully eliminated, because what matters is the aggregate displacement, not any specific role's survival.

The fourth layer is labour-market dominance, which is where the thesis actually lives. Wage labour stops being the mass route to economic agency. This has not fully arrived, but the pathway from the first two layers to the fourth is now specified and visible.

Critics prefer to argue at the third layer because residual human tasks are easy to find inside any existing job. But that is not where the thesis is defended, and never was. The argument runs from the first two layers directly to the fourth, with the third layer treated as partial and contingent rather than as a necessary intermediate step. Workflow recomposition can suppress hiring, break training ladders, and collapse mass labour absorption without any single occupation being formally eliminated. This is what makes the displacement quiet.

No scream, just non-absorption

Mass displacement will not first look like mass unemployment. It will look like non-absorption.

Fewer entry-level roles. Fewer junior ladders. Fewer graduate pathways. Fewer promotions. More contractors. More review roles. Productivity gains that do not reach wages. Incumbents held in place while new entrants fail to launch. A headline unemployment rate that looks fine while the system underneath it has stopped reproducing itself.

The tells to monitor are specific. Entry-level hiring decline in AI-exposed fields. Junior-to-senior ratio compression. Weak graduate absorption despite stable aggregate employment. Wage stagnation or wage compression in exposed cognitive work. Productivity gains not passed to labour. Contractorisation and project-based substitution. Delayed retirement of incumbents combined with fewer new entrants. Expansion of review and validation roles that do not scale into careers. Collapse of training ladders because AI now does the junior work from which expertise used to grow. Rising dependence on capital income, transfers, rents, or platform ownership rather than wages.

Stanford's Digital Economy Lab has reported the first expected signal. Brynjolfsson, Chandar, and Chen find that early-career workers aged 22 to 25 in the most AI-exposed occupations have experienced a 16 percent relative employment decline after controlling for firm-level shocks, while less-exposed fields and more experienced workers have remained stable or grown.[11] The adjustment is more visible in employment than compensation, and it is concentrated where AI is more automative rather than augmentative. That is exactly what the thesis predicts. Not "everyone gets fired at once." Entry ladders narrow first. Training ladders break first. Junior absorption weakens first. Aggregate employment can look fine while the reproduction mechanism fails underneath.

[11] Brynjolfsson, Chandar, and Chen, "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence," Stanford Digital Economy Lab. https://digitaleconomy.stanford.edu/publication/canaries-in-the-coal-mine-six-facts-about-the-recent-employment-effects-of-artificial-intelligence/

Aggregate calm is not a refutation if the No-Scream indicators are moving. That is the whole point of calling it the No-Scream Principle. The displacement does not arrive as a discrete event. It arrives as the quiet disappearance of the entry path, and by the time it becomes politically visible through aggregate statistics, the propagation has already done its work.

The augmentation argument is a stage error

The standard reply to all of this is that humans will work with AI. That sentence is doing more work than it can carry. There are three kinds of complementarity, and they are not the same thing.

Genuine complementarity means AI raises the marginal value of human labour. The worker becomes more productive, more valuable, and captures some of the gain through wages, bargaining power, or advancement. This is real, and it exists. It is not guaranteed to persist, because the same model that makes the worker more valuable today may absorb the worker's role tomorrow.

Transitional complementarity means humans supervise, correct, validate, integrate, and absorb responsibility while the AI improves. This is the phase most commonly mistaken for the future of work. It is unstable by design. The human role gets thinner as the system gets better, because the human role exists precisely to compensate for the system's current limitations. As those limitations are addressed, the role shrinks toward zero.

Theatrical complementarity means humans remain for trust, liability, regulation, customer comfort, ritual legitimacy, or institutional optics. The human is still in the room, but no longer economically central. This is not augmentation. It is managed displacement wearing augmentation's clothes. The role exists because the institution requires a human to be present, not because the work requires a human to be done.

The augmentation narrative points at the second stage and calls it the destination. The thesis argues it is the corridor. Once that corridor is named, "humans will work with AI" stops being a rebuttal and becomes a question about which complementarity is meant. Which complementarity, at what margin, for how long, under what competitive pressure? The honest answer is that genuine complementarity must be demonstrated, not assumed. Transitional complementarity is unstable by construction. Theatrical complementarity is displacement under another name.

Why the propagation is not stoppable inside firms

The interface collapse layer is not something firms choose to deploy as a discrete decision. It is the cumulative effect of thousands of small workflow changes, each defensible on its own terms.

A firm does not announce that it is rebuilding its operations around AI agents. It deploys a model in customer service, then expands the deployment, then connects the model to the ticketing system, then connects the ticketing system to the CRM, then automates the escalation path, then reduces the customer service headcount as the deployed system handles more cases. Each step is incremental. Each step is justified by quarterly cost-benefit analysis. The aggregate effect is workflow recomposition, but no individual decision looks like workflow recomposition. The decisions look like ordinary process improvement.

This is why interface collapse is hard to govern from inside the firm. The decisions are too small to require board approval. The decisions are too distributed to be coordinated as a single intervention. The decisions are made by the line managers responsible for the budget, who face direct competitive pressure to reduce costs and improve throughput. The CEO who wants to slow the propagation cannot do so without overriding hundreds of small operational decisions, each of which is locally rational. The CEO who tries to slow the propagation produces a firm with higher costs than competitors who do not.

The next essay asks why no actor can restrain this propagation once it begins. The answer is the Multiplayer Prisoner's Dilemma.

---

3. The Multiplayer Prisoner's Dilemma

Why AI Restraint Is Not an Equilibrium

The previous two essays established the technical and propagation conditions. Unit cost dominance has been crossed for a substantial fraction of professional cognitive tasks. Interface collapse is the propagation mechanism that carries task-level dominance into workflow recomposition and labour-market non-absorption. The question this essay addresses is what follows once both conditions hold. Specifically, can the deployment of AI be slowed, paused, or restrained by any actor or coalition of actors? The answer is no, not because no actor wishes to restrain it, but because the structure of competition makes restraint a dominated strategy at every level at which it could be attempted.

The argument is not that coordination is impossible in principle. It is that restraint is non-pausable as a competitive equilibrium under the conditions that obtain. The action of restraint is physically available. It is dominated by the action of deployment at every level. The actors who choose restraint are eliminated by the actors who choose deployment.

This is the political-economy hinge of the thesis. Without it, unit cost dominance and interface collapse would be problems amenable to coordinated response. With it, they become structural inevitability.

The fractal structure

The classical Prisoner's Dilemma involves two actors choosing between cooperation and defection, with payoffs structured so that defection dominates regardless of what the other actor does. Both actors defect even though both would prefer mutual cooperation. The lesson is that individual rationality can produce collective outcomes no one wanted.

The AI deployment situation is a fractal version of the same structure. The same payoff shape repeats at four levels: worker, firm, sector, and state. At each level, the actor who deploys gains a relative advantage. The actor who restrains is competitively dominated. Restraint is not punished by some external force. It is punished by the actors who do not restrain. The dilemma is recursive, and it operates simultaneously at every scale.

The fractal structure is what makes the situation unrecoverable through any single intervention. A regime that addresses the firm level is undermined by worker-level adoption. A regime that addresses the worker level is undermined by firm-level deployment. A regime that addresses the state level is undermined by both. The layers reinforce each other. Defection at any layer creates pressure for defection at the others.

The worker layer

A knowledge worker faces the dilemma directly. They are not making a grand ideological choice. They are trying to stay employed.

If the worker uses AI tools, they produce more, learn the tools faster, become cheaper per unit of output, and survive the next round of cuts. They also train themselves on the systems that will eventually replace their entire role, but that loss is years away and the layoff is months away. The discount rate makes adoption rational.

If the worker abstains, the colleague who uses AI produces more, learns faster, costs less per unit of output, and gets the role the abstaining worker wanted. Abstention is punished immediately. The collective benefit of abstention, if any, accrues to a future labour market the abstaining worker may not be in.

This matters because it makes the whole system harder to govern than the firm-level analysis suggests. Even if firms signed restraint agreements, workers would still adopt AI inside workflows. Even if states regulated formal deployment, individuals would still use tools for drafting, coding, analysis, planning, search, research, summarisation, and verification. Adoption happens below the level at which coordination can see it. The worker layer is the bedrock of the dilemma. Any regime that fails to account for worker-level adoption has already failed as wage-demand circuit defence.

The firm layer

A firm faces the same payoff shape one level up. The firm that deploys AI plus verification reduces unit costs for affected cognitive tasks. The firm that does not deploy faces competitors with lower costs. The undeploying firm either matches the cost reduction (which requires deployment) or loses market share until it exits. There is no third option.

This is true even if every executive at every firm prefers to preserve human employment. Their preferences are dominated by the competitive structure. In contestable markets, over sufficient time, the firm that persistently prioritises employment over unit cost is selected against. The firm that survives is the firm that deploys. Boards and executives are evaluated on quarterly performance. They cannot maintain restraint in the face of competitor deployment without losing their positions. Any individual firm that tries to restrain deployment is replaced by leadership that will deploy.

The second-best model in any given category may survive in niches, but it does not set the platform terms. The firm that is competitively second-best in its category does not face zero revenue. It faces shrinking margins, weaker pricing power, and slower talent acquisition. None of these is fatal in the short run. All of them are corrosive in the medium run. The firm has to keep matching the leader to avoid drift toward irrelevance.

The Dollar Auction layer

In a normal Prisoner's Dilemma, defection explains why actors deploy. It does not fully explain why they continue spending after returns become uncertain. For that, the AI race needs a second mechanism: Dollar

Auction dynamics.

In the classic Dollar Auction, the winner receives the dollar, but the losing bidder still pays their bid. This creates escalation. Once both players have bid heavily, quitting guarantees loss. Continuing is irrational from the outside but rational from the player's position inside the game. The game is structurally an all-pay tournament.

AI infrastructure has this structure. The losing firms do not simply fail to win the prize. They are left with stranded compute, depreciating models, lost talent, weakened distribution, reduced investor confidence, and no claim on future AI rents. The winner does not merely earn revenue. The winner sets the platform layer on which others build.

This is why continued spending can remain rational even when near-term margins are poor. The sunk cost is not the reason. The option value is. Stopping destroys the option. Continuing preserves a non-zero chance of surviving the next round. The frontier labs are not spending hundreds of billions because they are confident of the return. They are spending because withdrawal would foreclose the position from which any future return could be claimed.

The race is therefore not just a Prisoner's Dilemma. It is a Prisoner's Dilemma with all-pay tournament dynamics. Everyone pays to stay in the game. Losing does not mean returning to the starting line. Losing means paying for the race and receiving no platform position.

The sector layer

At the sector level, restraint is unstable because capital, talent, and customers migrate toward lower-cost substitutes and adjacent entrants. A restrained sector does not compete only with its existing incumbents. It competes with new firms, foreign firms, platform firms, and adjacent sectors that use AI to repackage or automate part of its function. The restrained sector becomes a protected high-cost island. Unless the protection is global and permanent, the island erodes.

A restrained legal services sector does not lose customers to manufacturing. It loses customers to legal-tech platforms, to in-house counsel teams equipped with AI tools, to foreign legal services providers operating remotely, and to adjacent service categories that absorb part of what was previously legal work. A restrained creative services sector loses to platforms, to in-house creative teams, to foreign agencies, and to direct AI-assisted production by clients. The pattern repeats across every cognitive-services sector.

Industry-level coordination is theoretically achievable. It is also structurally undermined by the boundary between industries being more permeable than the coordination assumes. Sector boundaries are administrative conveniences, not economic walls. AI dissolves the conveniences without asking anyone's permission.

The state layer

A state that restrains AI deployment loses ground in technology, in productivity, in fiscal capacity, and in the strategic capabilities that increasingly depend on AI infrastructure. Other states do not restrain. The

restraining state's relative position weakens. Its capacity to fund any social policy, including the social policy that might compensate displaced workers, declines. The restraining state arrives at a worse position for its own population than the state that deployed.

State-level restraint, even if politically achievable in any single state, produces worse outcomes for that state's population than deployment does. The political coalitions that supported restraint weaken as the costs become visible. Capital flight, brain drain, and erosion of competitive position appear within years. The state either abandons restraint or accepts persistent relative decline.

National security amplifies this dynamic. AI is increasingly framed as military, intelligence, scientific, industrial, and cyber capacity. A state that falls behind in AI falls behind in those capacities. No major state can accept that outcome, regardless of the social cost of deployment, because the alternative is being out-positioned by states that did not accept it.

Friction changes timing, not equilibrium

The standard objection to this analysis is that real deployment is slow. Firms face integration costs, liability concerns, regulatory delay, internal politics, union resistance, customer scepticism, professional gatekeeping, data privacy obligations, and infrastructure constraints. The objection grants the competitive logic but argues that friction prevents the logic from operating at the speed the thesis implies.

Friction is real. The thesis does not deny it. The thesis denies that friction constitutes a stable cooperative equilibrium. A firm may delay because of liability, culture, integration cost, or regulatory uncertainty. If unit cost dominance holds and competitors deploy, the delay becomes a cost disadvantage. The firm that delays still has to deploy eventually, or accept persistent erosion of competitive position. Friction redistributes when the deployment happens. It does not change whether it happens.

The same applies at every layer. Workers who delay adoption are outcompeted by workers who do not. Sectors that delay are eroded by adjacent sectors and foreign entrants that do not. States that delay accept relative decline. In each case, friction is a transient cost, not a permanent equilibrium. The friction-as-rescue argument requires friction to be persistent and universal, which it is not, because the actors who absorb the friction are competing with actors who do not.

This is why drag is not rescue. The closing essay in this sequence develops the point in detail. The point here is narrower. Within the Multiplayer Prisoner's Dilemma, friction modulates timing. It does not produce the cooperative equilibrium that would be required to restrain deployment.

Why coordination is not an equilibrium

The standard answer to a Prisoner's Dilemma is coordination. If the actors can communicate, commit, and enforce, they can reach the cooperative outcome that both prefer. International treaties, industry standards, and binding agreements are the institutional forms this coordination takes. The question is whether any of them can stabilise restraint in the AI case.

Repeated interaction exists. The classical route out of a Prisoner's Dilemma requires not merely repeated

interaction but observable defection and credible enforcement. Those are the missing conditions. In AI deployment, defection is internal, diffuse, continuous, and often indistinguishable from ordinary productivity improvement. The actors capable of enforcing restraint are also the actors most exposed to the cost of restraint.

Observable defection breaks down because AI deployment produces no outward sign. The output is the same. The product is the same. The customer interaction is the same. Detecting that the production process has changed requires inspection of internal workflows, which firms have strong incentives to obscure and which regulators lack the capacity to perform at scale. Aggregate data, payroll records, sectoral productivity statistics, and wage patterns can detect deployment after the fact. By the time the data shows defection, the defector has already captured the competitive advantage. Macro-level observability is post-hoc. Treaty enforcement requires real-time observability or it is not enforcement.

Credible enforcement breaks down because the actors with the capacity to enforce are the same actors with the strongest incentive to defect. States are the enforcers of international agreements. States are also the actors competing for AI advantage. A state that genuinely enforces a restraint agreement against its own firms cedes competitive position to states that do not. The enforcement function and the defection incentive are colocated in the same institution, which means enforcement is structurally undermined by the same dynamic that produced the original problem.

The conclusion is not that coordination is logically impossible. It is that coordination is not an equilibrium under conditions of weak enforcement, high strategic uncertainty, and continuous technological change. Those conditions are not contingent features of the present moment. They are structural features of the AI development environment.

Why the precedents do not transfer

Defenders of AI restraint sometimes point to historical precedents for coordinated restraint of dangerous technologies. Nuclear non-proliferation. Chemical weapons treaties. Ozone-depleting substances. Each is invoked as evidence that coordination is possible. The evidence is real but does not generalise to AI for reasons worth being precise about.

Nuclear non-proliferation works as well as it does because the core pathway to the weapon runs through scarce fissile material and identifiable facilities. AI has some chokepoints too: advanced chips, hyperscale data centres, cloud providers, and energy infrastructure. Those chokepoints matter for frontier training. They matter less for the wage-demand problem. Labour substitution diffuses through APIs, open-weight models, smaller specialised systems, enterprise wrappers, and ordinary workflow tools. A regime that slows frontier training does not necessarily stop cognitive substitution below the frontier.

Chemical weapons treaties work better than AI restraint because chemical weapons have little legitimate civilian economic value. Producing chemical weapons does not generate revenue, market share, or competitive advantage in any non-military domain. States that abandon chemical weapons production lose nothing economically. AI, by contrast, is the most economically valuable technology of the era. States that abandon AI development lose enormous economic value. The cost-benefit calculation that supports chemical weapons restraint does not apply to AI restraint.

Ozone-depleting substances were restrained because the relevant chemicals had close substitutes that did not damage the ozone layer. Industry could shift to substitutes without losing economic function. AI has no substitute for the cognitive functions it performs. The substitute for AI cognition is human cognition, which is what AI cognition replaces because it is more expensive. The Montreal Protocol model assumed an off-ramp that does not exist for AI.

The AI race is not a Stag Hunt because there is no shared stag. Third-party defection by open-weight developers, foreign states, market entrants, or internal workflow adopters destroys the cooperative equilibrium. It is not a nuclear arms race in the ordinary sense because AI is not only a weapon. It is a productive asset. A state that restrains nuclear weapons may preserve most of its economy. A state that restrains AI restrains its own productivity growth, fiscal capacity, and strategic infrastructure. It is closest to Moloch, but Moloch alone is too static. The AI race has a burn mechanism. It is Moloch with a balance sheet.

The precedents that critics invoke do not show that AI restraint is possible. They show that restraint is possible under specific conditions that do not obtain in the AI case. The relevant question is not whether some technologies have been restrained. It is whether the conditions that allowed those restraints are present here. They are not.

The recursive prize

A normal Dollar Auction has a fixed prize. The game ends when someone wins the dollar.

The AI race does not have that structure. The prize is recursive. Each capability level unlocks the next capability level. Each deployment produces revenue, feedback, workflow data, product telemetry, distribution, and political leverage that can be used in the next round. Winning one round does not end the contest. It changes the baseline from which the next contest begins.

This matters because exit calculation becomes unstable. In a finite game, a firm can model the endpoint and decide whether the remaining prize justifies the remaining spend. In the AI race, there is no final prize against which exit can be calculated. The actor that exits does not preserve its position. It falls behind the moving frontier.

You do not win the AI race. You survive one round. Then the next round starts immediately, with higher stakes and faster timelines. There is no plateau. Only slope.

The race is therefore not won once. It is survived repeatedly. The absence of a finish line does not free the players from the race. It removes the only thing that could have stopped it.

Bridge to Sorites

There is one more reason coordination fails, and it requires its own essay. Even if the institutional conditions for restraint were satisfied, the boundary between AI assistance and AI replacement could not be cleanly drawn. The next essay handles that problem. The Sorites Collapse Principle is what makes the Prisoner's Dilemma irrecoverable. In the absence of Sorites, a sufficiently determined coordination effort might find an enforceable definition of what to restrain. With Sorites, no such definition exists.

4. The Sorites Collapse Principle

Why No Boundary Can Be Drawn Inside a Workflow

The ancient paradox of the heap goes like this. One grain of sand is not a heap. Two grains are not a heap. Adding a single grain to a non-heap cannot make it a heap. Therefore, by induction, no number of grains constitutes a heap. The paradox is not that heaps do not exist. They obviously do. The paradox is that no specific grain count separates heap from non-heap, which means the boundary is undefined, which means any line we draw is arbitrary.

The paradox has a serious economic application. Inside any single use of AI, there is a continuous gradient from minimal assistance to near-total replacement. Spell-check is assistance. Autocomplete is more assistance. Draft generation is much more assistance. Multi-step composition with light human review is more still. Autonomous composition with rubber-stamp approval is replacement in all but name. The gradient is unbroken. There is no specific point at which assistance becomes replacement, because the categories are not discrete events. They are stages on a single trajectory of cognitive substitution.

This is the Sorites problem applied to AI deployment, and it is the structural reason that no regulatory boundary, no treaty, no professional norm, and no human-only zone can hold against the gradient as a defence of the wage-demand circuit. Any regime that depends on a clean line between assistance and replacement fails because no such line exists. This essay establishes the principle and traces its implications.

The within-face gradient

Consider a single professional task: writing a legal brief. Five years ago, this task was performed by a lawyer or a team of lawyers, with research support and possibly a paralegal drafting initial sections. Today, the same task can be performed in a sequence of stages, each of which is technically defensible as a form of assistance.

Stage one: the lawyer uses AI to summarise relevant case law. The AI reads cases and produces summaries. The lawyer reads the summaries and writes the brief. This is assistance. Nobody objects. The lawyer is still doing the cognitive work.

Stage two: the lawyer uses AI to draft individual sections of the brief based on bullet-point instructions. The AI produces draft text. The lawyer revises. This is also assistance. The lawyer is still in control. The output reflects the lawyer's judgement.

Stage three: the lawyer uses AI to draft the entire brief based on a high-level prompt. The AI produces a complete brief. The lawyer reviews it, makes some edits, and submits. This is harder to characterise. The lawyer is still legally responsible. The lawyer reviewed the document. But the document was not authored by the lawyer in any meaningful sense. The lawyer is functioning as a verifier rather than an author.

Stage four: the lawyer uses an agentic system that takes the case file, drafts the brief, prepares the filing package, drafts responses to opposing counsel's filings, and updates the lawyer on developments. The lawyer reviews periodically and approves at key decision points. The lawyer is still in the loop. The lawyer is no longer doing the work.

Stage five: the firm deploys the agentic system at scale. One lawyer supervises the output of the system across many cases. The system handles routine matters end to end. The lawyer intervenes only when the system flags a complex decision. The lawyer's role is now exception handling. The system does the work.

Each transition between stages is small. Each stage is defensible as a continuation of the previous one. There is no specific stage at which the lawyer stopped being the author and started being the verifier, because the transition is gradual. By stage five, the lawyer is doing fundamentally different work than at stage one. The transition happened, but no individual step crossed a clear line.

This is the Sorites collapse. There is no point inside the workflow at which assistance becomes replacement. The boundary exists nowhere and everywhere. The regulatory question of where to draw the line cannot be answered from inside the gradient.

Why the gradient is not just a definitional problem

A critic might respond that this is merely a definitional problem, of the kind law solves all the time. Speed limits draw arbitrary lines. Age of majority draws arbitrary lines. Tax brackets draw arbitrary lines. Why can AI assistance not be regulated by drawing a similar arbitrary line, accepting that the line is arbitrary, and enforcing it?

This response misunderstands the structure of the problem. Speed limits work because the underlying quantity (vehicle velocity) is measurable, discrete, and relatively stable. A vehicle is going either above or below the limit, and which side it is on can be measured by a radar gun. Age of majority works because age is measurable, discrete, and changes only by aging. Tax brackets work because income is measurable in a defined accounting period.

The quantity that would have to be measured to regulate AI assistance is not measurable, not discrete, and not stable. It is the degree of cognitive substitution within a workflow. This is not a number that can be read off any instrument. It is a relational property of the work, the worker, the AI, and the task, which varies continuously across uses, across tasks, across workers, across time of day, across deadline pressure, across the worker's expertise level. Two lawyers using the same AI for the same task may have different degrees of substitution because they have different baseline competencies. The same lawyer using the same AI for the same task may have different degrees of substitution on Monday morning versus Friday afternoon.

The Sorites problem with cognitive substitution is not the ordinary Sorites problem of arbitrary bound-

aries. It is a deeper problem in which the underlying quantity is not the kind of thing that admits clean measurement, and therefore not the kind of thing that admits clean regulation as circuit defence.

Proxies measure manifestations, not substitution

A regulator can measure proxies. It can measure review time, AI-generated text, output per employee, approval rates, provenance logs, edit distance between AI draft and final version, audit trails, task allocation records, or headcount changes. These proxies are real. They produce numbers. They can be tracked, reported, and enforced.

The proxies are useful for fraud detection, safety documentation, intellectual property protection, and accountability after harm. They do not measure the thing that matters for the wage-demand circuit, which is whether human labour remains economically necessary to production. A worker can edit every paragraph and still be a verifier. A worker can approve every decision and still be ornamental. A firm can maintain headcount temporarily while the productive necessity of that headcount disappears. The proxy survives. The circuit does not.

This is the within-face equivalent of the formulation that anchors the next essay. Regulation attaches to manifestations. Capital moves through relations. The mismatch is the failure. Inside a workflow, regulation can only see what produces a measurable trace. The substitution of human labour by AI produces measurable traces that do not track the substitution itself. They track the appearance of human involvement, which is exactly what the deployed workflow is engineered to preserve. The trace becomes the compliance product. The economic reality moves underneath it.

The implications for regulation as circuit defence

This has direct consequences for any regulatory regime that depends on distinguishing AI assistance from AI replacement to preserve the wage-demand circuit.

Regulatory regimes that prohibit AI replacement of human workers cannot define what they are prohibiting. The regulation either prohibits all AI use (which prevents productivity gains and is politically unworkable) or prohibits some specific level of AI use (which requires drawing a line that the gradient prevents). Any line that is drawn becomes the new ceiling for permissible AI use, with workflows redesigned to sit just below it. The line drifts upward as workflows adapt, and the regulation either follows the drift (becoming reactive and increasingly permissive) or holds the original line (becoming impossible to enforce as workflows evolve past it).

Regulatory regimes that require human oversight of AI decisions face the same problem from the opposite direction. Human oversight is not a binary. It is a bandwidth allocation. A human can review one output per day, ten outputs per hour, or ten thousand outputs per shift. All three satisfy the phrase “human oversight” unless the regulation specifies intensity. Once intensity is specified, the regulated object moves to sampling, escalation, exception handling, peer review, audit trails, or managerial approval. The human remains formally present while the labour content of the human role approaches zero.

Regulatory regimes that license specific uses of AI face the boundary problem at the use definition level. Each use can be redescribed as a different use. AI for clinical decision support becomes AI for differential consideration becomes AI for case-by-case workflow assistance. The same underlying activity moves between regulated and unregulated categories depending on how it is framed. The regulator's only response is to regulate the framing, which returns the problem to definition.

Regulatory regimes that operate on outcome measures (productivity per worker, hiring patterns, wage levels) face a different kind of problem. They can succeed where they directly mandate the outcome they want to preserve. A regime that requires firms to maintain a specific ratio of labour costs to revenue does not need to distinguish assistance from replacement. It regulates the value flow directly. This is a real category of intervention, and it is not defeated by the Sorites problem. It is, however, no longer category-based regulation. It is structural intervention on the value flow, and it raises the question of what such intervention preserves. The Successor System essay addresses that question directly. The short answer is that mandatory wage-share schemes preserve the wage form, not the wage mechanism, and the difference matters for what kind of system results.

There is no category-based regulatory instrument that depends on distinguishing assistance from replacement and still succeeds as a defence of the wage-demand circuit. Other regulatory purposes survive the gradient: documentation duties, liability rules, safety standards, professional conduct codes, and audit requirements can all serve legitimate functions even where the assistance-replacement distinction collapses. The thesis is narrow. It says the wage-demand circuit cannot be defended by category-based instruments that depend on the gradient, not that all regulation is futile.

Law can manage gradients. It cannot preserve productive necessity inside a dissolving workflow.

The implications for professional norms

The argument is sometimes made that professional norms can do what regulation cannot. Lawyers have professional ethics. Doctors have professional standards. Accountants have professional codes. These norms could, in principle, define what level of AI use is professionally acceptable and enforce that limit through licensing, peer review, and reputational sanction.

The Sorites problem operates here too, with an additional difficulty. Professional norms are typically defined by the profession itself, through its representative bodies. The profession's representative bodies are made up of practitioners. Practitioners face the same competitive pressure as everyone else. The lawyer who limits AI use to maintain a high standard of personal authorship competes against the lawyer who uses AI extensively and produces equivalent work at lower cost. The high-standard lawyer either lowers their standard, raises their fees beyond what clients will pay, or loses clients to the other lawyer. The professional body, made up of practitioners under this pressure, drifts toward norms that accommodate AI use rather than restrict it.

This is the same Multiplayer Prisoner's Dilemma operating one level inside the profession. The norms shift to match the deployment, rather than the deployment being constrained by the norms. The Sorites problem makes the shift hard to detect, because no specific norm change is visible. The implicit standard for what counts as competent legal work, or competent medical diagnosis, or competent accounting, drifts gradually

toward AI-assisted production. The drift is not a decision. It is the cumulative effect of practitioners adopting AI at the margins to remain competitive.

At circuit scale, professions adapt to the technology more than they constrain it. Real professional rules can delay or shape adoption in specific high-stakes contexts. They cannot hold the assistance-replacement boundary as a defence of mass labour absorption.

What about treaties and international agreements

International treaties are sometimes invoked as a higher-level coordination mechanism that might succeed where national regulation fails. The Sorites problem applies here with full force. A treaty banning AI replacement of human workers in regulated industries cannot define what is being banned, because the boundary between augmentation and replacement does not exist. A treaty regulating AI use above a certain capability threshold cannot maintain the threshold, because capability gradates continuously across model generations and across deployments. A treaty defining human-only zones in employment cannot specify what counts as human-only when the human is using AI tools to perform the work.

The standard model for international technology governance is arms control, which has a moderate track record on discrete physical objects (warheads, missiles, chemical agents). The model fails for AI for the reasons named in the previous essay (no discrete inputs, no observable defection, no credible enforcement) and for the additional reason given here. Even if the institutional conditions for treaty-based governance were satisfied, the substance of what would be governed cannot be defined. The treaty either lacks definition (in which case it is unenforceable) or imposes a definition the gradient invalidates (in which case it is gamed within months).

There is no level of international coordination at which the Sorites problem dissolves. The problem is inherent to the gradient, not to the level of governance. Higher levels of governance face the same definition problem with less detailed knowledge of the workflows being governed. They are worse positioned to draw the line, not better.

The relation to the across-face problem

This essay covers the within-face Sorites problem: the gradient inside any single use of AI. The next essay covers the across-face problem: the migration of AI deployment between legal categories. The two problems compound. Within-face Sorites makes any single category individually ungovernable as circuit defence. Across-face recursion makes the choice of category arbitrary. Together they exhaust the regulatory surface available for preserving the wage-demand circuit.

The Sorites Collapse Principle is the within-face foundation. Categorical Recursion is the across-face foundation. The next essay completes the regulatory closure.

---

5. Categorical Recursion

The Object Moves When You Look At It

There is a class of regulatory failure our existing names blur. It is not capture. It is not arbitrage in the ordinary sense. It is not lag. It is the specific failure mode that occurs when the act of regulating an object causes the economic deployment of that object to mutate into an adjacent legal category that the regulation does not cover.

Call it categorical recursion. It is the deep mechanism beneath every governance proposal that tries to preserve the wage-demand circuit by governing AI as a legal object. It is why those proposals fail in ways their authors find surprising.

The five faces

AI is not one legal object. It is five at once.

It is productivity software. It is capital equipment. It is labour substitute. It is strategic infrastructure. It is recursive R&D engine. The same model, deployed by the same firm, in the same week, occupies all five categories simultaneously. Not sequentially. Not depending on use case. At once.

A frontier model used to draft a contract is productivity software to the lawyer, capital equipment on the firm's balance sheet, labour substitute to the paralegal who will not be hired, strategic infrastructure to the state hosting the data centre, and a recursive R&D engine to the lab training the next version. The categories are not chosen by the regulator. They are inhabited by the object.

The ambiguity is not rhetorical. It is ontological. Each face is fully true. The model really is software. It really is capital. It really is replacing future hires. It really is part of national compute infrastructure. It really is part of the next R&D loop. None of these descriptions is misleading. All of them are simultaneously available as legal-economic identities.

This is not a feature of bad regulation. It is a feature of the technology.

What regulation actually does

Within the legal-categorical paradigm, selecting a category is the first thing a regulator must do. Law operates on legal objects, and legal objects are categorical. The EU AI Act regulates AI systems by risk tier.

Export controls regulate compute by capability threshold. Procurement rules regulate vendors by contract type. Liability frameworks regulate harms by causation chain. Each of these is a category-selection act.

The selection commits the regulator to a definition. The definition becomes the surface against which the deployed object moves. The object does not need to evade the definition. It only needs to inhabit a different one of its five faces.

Regulate AI as labour substitution and the firm reframes deployment as productivity software. Regulate it as software and the state reframes it as strategic infrastructure exempt under national security carve-outs. Regulate it as a weapon and the developer reframes it as cloud productivity. Regulate frontier training and the diffusion happens through smaller open-weight models inside ordinary workflows. Regulate deployment and the firm reframes the system as assistance with human oversight. Regulate assistance and the boundary with ordinary software collapses entirely.

At no point in this sequence does any actor need to break a law. The object simply presents a different face.

The architecture is the trap

The clearest case study is the EU AI Act, Regulation (EU) 2024/1689, which is explicitly category-based.[12] The legal architecture itself creates the migration paths. Provider, deployer, downstream integrator, fine-tuner, open-source releaser, enterprise customer, and systemic-risk model are not merely compliance labels. They are available legal identities defined in Articles 3, 25, and 51 of the Act. Once obligations attach to one identity, value migrates through the others.

The Commission's own enforcement guidance (non-binding, but revealing of the enforcement architecture) confirms the structure.[13] Most fine-tuning, adaptations, and minor modifications do not automatically create new model-provider obligations unless the modification crosses a high threshold, defined as more than one-third of the original model's training compute. Open-source treatment receives different obligations from commercial release, though this exemption does not apply to general-purpose AI models with systemic risk. Downstream integrators inherit obligations differently from upstream providers. Each of these distinctions is reasonable in isolation. Each is a doorway in aggregate.

A model whose direct deployment as a high-risk system would attract heavy obligations can be released, modified, integrated, or wrapped through adjacent roles whose obligations attach differently. Where systemic-risk duties remain, the compliance surface still shifts from direct deployment to release, modification, integration, and downstream use. A general-purpose capability can be decomposed across models, wrappers, tools, integrations, and workflows whose individual components may sit below the most visible threshold, forcing the regulator to reconstruct systemic function after deployment rather than identify it cleanly before deployment.

[12] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[13] European Commission, "Guidelines on obligations for General-Purpose AI providers." https://digital-strategy.ec.europa.eu/en/faqs/guidelines-obligations-general-purpose-ai-providers

This is the essential point. The regulator did not create loopholes. The regulator created categories, which is what regulators must create. The object moves between them because the object is the kind of thing that exists in all of them at once.

The recursive part

The recursion is what makes this novel.

In conventional regulatory arbitrage, the regulated party finds a workaround, the regulator updates the rule, and the workaround closes. The cat-and-mouse equilibrium is unstable but bounded. Regulation is always somewhat behind, but it is not categorically defeated.

Categorical recursion is different. The act of regulating a face of the object is itself a market signal. It tells every firm in the ecosystem which face is currently being scrutinised, which means every firm has an immediate incentive to redescribe its deployment under one of the other four faces. The regulation does not just fail to catch the object. The regulation creates a market signal and an incentive for migration.

This is not a slow process. It happens at the speed of a press release. The regulator's next move is to broaden the definition. The market's next move is to inhabit a face not yet defined. The regulator's third move is to attempt a comprehensive framework. The market's third move is to invoke the multi-category arbitrage problem as evidence that comprehensive frameworks are unworkable.

The regulator loses not because the regulator is slow. The regulator loses because the regulator is operating on the wrong substrate. Categories cannot enclose objects whose economic deployment selects its category in response to enclosure.

Why this is not capture

It is tempting to read this as a sophisticated form of regulatory capture. It is not. Capture requires that the regulated party influences the regulator's preferences or rules. Categorical recursion requires no such influence. It operates even against a perfectly honest, perfectly well-resourced, perfectly motivated regulator.

The mechanism is structural. The regulator's instrument is the legal category. The deployed object's response surface is the same legal category. The object does not need to corrupt the regulator. It only needs to be the kind of thing that can be five things at once.

Capture is a problem of incentives. Categorical recursion is a problem of ontology.

What actually moves

The model itself does not literally evade. The neural network is stable. Weights do not change to escape regulation. The thing that moves is not the technical substrate. It is the legal-economic wrapper around the substrate: provider, customer, tool, service, infrastructure, research artefact, fine-tune, agent, workflow, API, internal system, open-weight derivative.

The object does not evade as software. It evades as capital. Its technical substrate is stable enough to be deployed. Its legal-economic identity changes depending on where regulation attaches.

Regulation attaches to manifestations. Capital moves through relations. The mismatch is the failure.

This identifies the real target. The thing being regulated is not AI as a technical object. It is AI-as-capital-in-motion. The motion is the thing the categories cannot hold.

The Sorites layer underneath

Categorical recursion sits on top of the existing Sorites problem. The Sorites problem is that within any single face of the object, there is no clean line between assistance and replacement. Spell-check to auto-complete to draft generation to autonomous composition is one continuous gradient. No regulation can identify the moment a worker stopped being augmented and started being substituted, because no such moment exists.

Categorical recursion adds a second axis of ungovernability. Even if a regulator solved the within-face Sorites problem, the across-face migration problem remains. The two axes interact. Within-face Sorites makes every face individually ungovernable. Across-face recursion makes the choice of face arbitrary. Together they exhaust the regulatory surface.

This is why governance proposals that work on paper fail on contact. The paper version assumes a stable object presenting a stable face along a stable gradient. The deployed version is none of those things.

The six-gate test

Any proposed governance regime aimed at preserving the wage-demand circuit must survive six failure modes simultaneously:

State defection, where one major actor refuses the regime and gains comparative advantage. Firm-level evasion, where commercial actors reclassify deployments to escape coverage. Open-weight diffusion, where capability spreads below the regulatory threshold. Individual workflow adoption, where eight billion users make daily tool choices the regime cannot observe. Category arbitrage, where the deployed object migrates to a face the regime does not cover. The assistance-replacement gradient, where Sorites within a face dissolves the regulated distinction.

A regime that fails any one of these fails as a defence of the wage-demand circuit. It may still reduce fraud, improve documentation, slow a specific deployment, or create useful liability after harm. Those are real outcomes and worth pursuing on their own terms. They do not preserve mass labour absorption. The thesis is narrow. It says the wage-demand circuit cannot be defended by category-based instruments, not that all regulation is futile.

The design record on circuit defence is consistent. The EU AI Act fails on category arbitrage and the Sorites gradient. Compute thresholds, even by the Frontier Model Forum's own assessment, are an imperfect proxy for risk: they may miss smaller systems with harmful capabilities while catching larger benign ones, and

function at best as an initial filter.[14] Liability frameworks fail on the gradient and on individual adoption. Procurement standards fail on firm-level evasion. Frontier moratoria fail on state defection. Each proposal is competent within its chosen face. None survives migration of the deployed object to an adjacent face.

What this implies for the wage-demand thesis

The implication is that the wage-demand circuit collapse is not a failure of regulation. It is what happens when the substrate of regulation is mismatched to the substrate of the regulated object.

The wage-demand circuit is preserved only by mass labour absorption. Mass labour absorption requires that AI deployment as labour substitution be governable. AI deployment as labour substitution is one face of an object with four other faces, and any attempt to govern that face causes immediate migration to the others. Therefore the wage-demand circuit is not defended by any regulatory regime that operates on legal categories.

This is not an argument that regulation is futile. It is an argument that regulation aimed at preserving the wage-demand circuit cannot succeed through category-based instruments, which are the instruments regulation possesses. The conclusion is structural.

The regulator's dilemma

There is a final recursion worth naming.

A regulator who understands categorical recursion has two options. They can attempt comprehensive coverage across all five faces, which requires a level of definitional precision and cross-domain coordination that no existing institution can deliver, and which itself becomes the market signal that triggers the next mutation. Or they can abandon category-based regulation and attempt outcome-based regulation, which requires defining the outcome to be prevented, which returns the problem to definition, which returns the problem to category.

There is no third option that operates within the legal-categorical paradigm. The paradigm is the constraint.

What sits outside the paradigm is not regulation. It is something else. Direct ownership of the compute layer. Mandatory wage-share redistribution from automation gains. Public deployment of the technology with non-market allocation rules. These are not regulations. They are structural interventions that do not depend on category selection because they operate on the underlying flow of value rather than on the deployed object.

Whether any of these is politically achievable is the next question. The point of this essay is only to establish that the regulatory question, as conventionally posed, has been answered. The economic deployment of the object moves when you look at it. The looking causes the moving. No legal-categorical instrument survives the recursion.

[14] Frontier Model Forum, "Issue Brief: Thresholds for Frontier AI Safety Frameworks." https://www.frontiermodelforum.org/uploads/2025/02/FMF-Issue-Brief-on-Thresholds-for-Frontier-AI-Safety-Frameworks.pdf

This is the structural reason the wage-demand circuit will not be defended by the regulatory state. Not because the state is unwilling. Because the regulatory state is using object-categories against a value-flow.

---

6. The Successor System

Why Structural Alternatives Save Consumption Without Saving the Circuit

The previous essay closed the regulatory route. Category-based instruments cannot defend the wage-demand circuit because the deployed object migrates between categories faster than categories can be redrawn. The regulator's failure is structural.

The honest reader's next move is to reach for the alternatives. Direct ownership of the compute layer. Mandatory wage-share redistribution from automation gains. Public deployment of the technology with non-market allocation rules. Sovereign AI funds. National compute dividends. Nationalisation of frontier labs. These sit outside the categorical paradigm. They operate on the value flow rather than the deployed object. They are the right kind of intervention.

They also do not save the wage-demand circuit. They save consumption. The two are not the same object.

The successor system can deliver consumption. It cannot deliver the wage-demand circuit. That distinction matters because the wage-demand circuit was never only a way to distribute purchasing power. It was a way to distribute agency, bargaining power, status, mobility, and democratic leverage through productive necessity. Redistribution can replace the income. It cannot replace the necessity.

This is the essay that closes the last exit.

The conflation that needs to break

The wage-demand circuit is a specific mechanism. It is the feedback loop in which production generates wages, wages generate consumption, consumption generates production, production generates labour demand. The circuit is closed. Each leg supports the others. Mass economic agency is internal to the circuit because the population's purchasing power is generated by its own productive activity.

Consumption-via-redistribution is a different mechanism. Production generates revenue. Revenue is taxed or claimed by the state. The state distributes the proceeds as transfers. Recipients consume. Consumption generates further production. The circuit closes, but it closes through the state. The population's purchasing power is generated by political allocation rather than productive activity.

These are not minor variations on the same system. They are different systems with different stability properties, different political economies, different distributions of agency, and different long-run trajectories. Conflating them is the central error of every redistributionist account of the AI transition.

The thesis is not that the second system cannot exist. It is that the second system is not the first system, and the question of whether postwar capitalism survives is the question of whether the first system survives. It does not.

The difference between supplement and replacement

The obvious objection is that postwar capitalism already contained redistribution. Welfare states, public healthcare, pensions, unemployment insurance, tax credits, housing subsidies, and public education all redistributed purchasing power outside the immediate wage relation. Therefore, the critic says, AI redistribution is not a successor system. It is merely more welfare-state capitalism.

This misses the distinction between supplement and replacement. In the postwar settlement, redistribution supplemented a labour market that remained the primary generator of mass purchasing power. The wage was still the centre. Transfers stabilised the circuit. They did not constitute it. Most working-age adults received most of their income, status, recognised social role, and political standing through their participation in production. The welfare state caught those who fell out of the wage circuit. It did not replace the circuit.

A post-AI redistribution regime is different. It does not stabilise a wage circuit that remains primary. It replaces the wage circuit as the source of mass demand. That is the discontinuity. The question is not whether the state distributes some income. The question is whether most people receive purchasing power because their labour is economically necessary, or because the political system allocates them a claim on automated surplus.

The threshold is not redistribution. It is the centrality of redistribution. When transfers stop being the catch and start being the floor, the system has changed even if every institutional name stays the same.

Take the alternatives at their best

Begin by giving each structural intervention its strongest form.

Direct ownership of the compute layer means the state owns or controls the data centres, the chips, the training runs, the deployed models, and the inference capacity. Private actors lease access. The rents flow to the public balance sheet rather than to private equity. This is genuinely outside the categorical paradigm because the state is no longer regulating an object. It owns the object.

Mandatory wage-share redistribution from automation gains means the state requires firms to maintain a target ratio between labour compensation and total value added. As AI displaces workers, the firm either retains the workers at sustaining wages or pays the difference into a redistribution pool. This operates on the value flow because it conditions the firm’s right to deploy AI on the firm’s continued contribution to mass income.

Public deployment with non-market allocation means the state deploys AI itself for public purposes, with output allocated by need or political process rather than by price. Healthcare, education, infrastructure

planning, scientific research, and administrative services run on publicly owned models. The output is consumed without intermediation by the wage market.

Each of these is a real intervention. Each operates outside the categorical paradigm. Each can, in principle, sustain mass consumption indefinitely. The thesis does not deny this. The thesis observes what the interventions do not preserve.

These three alternatives are illustrative of the design space, not exhaustive. Hybrid arrangements, novel ownership structures, sectoral variants, and combinations of the three can produce successor systems with quite different properties. The structural argument applies to the design space as a whole: any intervention operating on the value flow rather than on productive necessity preserves consumption rather than the wage-demand circuit. The three are useful as the cleanest cases. They are not the boundary of what a successor system might look like.

The synthetic wage problem

Mandatory wage-share redistribution is the trickiest case. A critic could say that if firms are forced to maintain labour compensation, then the wage-demand circuit survives. Wages still exist. Workers still get paid through payroll. The mechanism looks identical to the postwar circuit.

It is not. Mandatory wage-share preserves the wage form, not the wage mechanism. A wage generated by productive necessity is different from a wage maintained by political obligation after the worker has ceased to be necessary to production. The first is labour income. The second is a transfer routed through payroll. It may be desirable. It may be stabilising. It may be morally necessary. It is not the old circuit surviving. It is the successor system wearing the wage as an administrative costume.

The diagnostic is sharper than “would the firm pay absent regulation.” Many wages in postwar capitalism were politically mediated: union wages, regulated-profession wages, public-sector wages, minimum-wage floors, procurement-supported wages. Political mediation was always present. The question is whether the worker is still performing a function for which human labour is materially required to produce the output. If yes, the wage is real even when its level is shaped by political process. If no, the payment is income maintenance routed through payroll. The first is the wage-demand circuit. The second is a transfer wearing the wage form.

This is the supplement-versus-replacement line. Postwar capitalism contained substantial political mediation that supplemented a labour market in which firms genuinely needed workers to produce output. A worker in 1965 was paid a politically mediated wage and was also materially required for production. Remove the political mediation and the firm still hires the worker at a different wage. Remove the worker and the firm cannot produce the output. That structural necessity is what the wage-demand circuit rested on. Synthetic wages exist where the second condition no longer holds.

Both deliver income. Only the first is the wage-demand circuit.

This matters because synthetic wages have different stability properties from real ones. Real wages are defended by the firm’s need for labour. Synthetic wages are defended only by the political will that mandates them. Political will is exposed to elections, lobbying, fiscal crises, and competitive pressure from

jurisdictions that do not impose the mandate. The first is structurally embedded. The second is structurally contingent.

What survives and what does not

Under direct compute ownership, the state collects the rents from AI deployment and disburses them to the population. The population eats. The population can vote. The population can consume. The political form may be a sovereign wealth fund, a citizen dividend, a basic income, or a public services model. The mechanism by which the population obtains income is no longer wage labour. It is allocation by the state. The wage-demand circuit, the loop in which production directly generates the income that purchases production, has been replaced by a state-mediated transfer system.

Under mandatory wage-share redistribution, the firm pays the difference between actual wages and target wages into a pool. The population still receives income. The income is no longer generated by the act of producing the goods being purchased. It is generated by a regulatory transfer that runs alongside production. Production and income have been decoupled. The decoupling is the point of the policy. It is also the death of the wage-demand circuit.

Under public deployment with non-market allocation, the population receives services without paying for them through the wage market. This is functionally similar to the existing public provision of healthcare, education, or infrastructure in many states, scaled up to cover what the wage economy previously covered. The mechanism is allocation, not exchange. Mass economic agency is replaced by mass political claim.

In every case, the food arrives. The lights stay on. The population is not abandoned. The wage-demand circuit is not preserved. Something else is.

What the something else lacks

The wage-demand circuit had three properties beyond mere consumption.

The first is that economic agency was distributed by the same mechanism that distributed purchasing power. To earn was to participate in the productive system. The act of selling labour gave the worker a role in the system that produced what they consumed. This produced political leverage. Strikes mattered because withdrawing labour withdrew production. Bargaining mattered because the firm needed the worker as much as the worker needed the firm. The political economy of the postwar settlement rested on this mutual necessity.

Recipients of state allocation retain political leverage. They can vote, protest, organise, riot, withdraw legitimacy, and threaten disorder. These are real forms of power. They are not the same as labour leverage. Labour leverage works because production requires the worker. Recipient leverage works because rule requires consent or at least social peace. The first is embedded in production. The second is external to it. A worker strike stops production. A recipient strike does not. A voter revolt can change governments, but only if institutions remain responsive. A riot can impose costs, but it does not restore productive necessity. The successor citizen may have political power. They do not have the same economic power.

The second is that the wage was a price signal that allocated talent, effort, and ambition across the economy. People moved into productive activity because productive activity paid. The signal was imperfect, distorted, and unjust in many ways, but it was a signal generated by the productive system itself rather than by political allocation. Under redistributive successors, this signal weakens or disappears. People can still be paid for some forms of work, but the pay no longer reflects productive necessity. It reflects state preference, sectoral protection, or status reward. The allocation function passes from the market to the polity.

The third is that the wage-demand circuit was institutionally stabilisable because displaced workers could, in principle, be reabsorbed into productive employment. Keynesian demand management, welfare states, unions, public investment, and industrial policy all worked on the assumption that labour remained economically necessary. The postwar settlement was not self-correcting. It was correctable, because the underlying labour market could be repaired, expanded, and redirected. Under redistributive successors, this assumption weakens. The shock now propagates through the state's fiscal capacity and political will rather than through a labour market capable of mass reabsorption. There is no labour market capable of mass reabsorption to repair, expand, or redirect, because the productive role of mass labour is no longer the basis of income distribution.

The successor system can deliver consumption. It cannot deliver these three properties. The three properties are what made postwar capitalism postwar capitalism.

The threshold is majority agency

The threshold is not the last worker. There will always be residual human work: protected work, artisanal work, trust-bearing work, care work, political work, luxury human service, and legally mandated human presence. Residual labour does not preserve a system built on mass productive participation.

Postwar capitalism dies when wage labour no longer provides mass economic agency because a majority of working-age adults cannot sell labour at socially sustaining wages without subsidy, protection, artificial scarcity, makework, or political intervention. Not total labour extinction. Mass agency collapse. Not consumption. Productive participation.

This threshold matters because it closes the residual-work dodge. Critics often respond to the thesis by pointing at remaining human jobs. Some humans will always work. Kings had servants. Aristocracies had artisans. Feudal societies had labour. The existence of residual human work does not preserve a system built on mass productive participation. The question is whether wage labour remains the route to mass economic agency for the majority. When it does not, the system has changed even if every individual residual job continues to exist.

The political-economy problem the structural alternatives inherit

Even setting aside what the successor system lacks, there is a problem with installing it.

The structural alternatives all require the state to expropriate, tax, or directly control the most economically valuable infrastructure of the era. This is a major political act. It requires either popular pressure

sufficient to overcome the resistance of the AI-owning interests, or elite consensus that the alternative to expropriation is worse than the expropriation itself.

Popular pressure sufficient for this requires mass economic leverage. Mass economic leverage requires the wage-demand circuit. The political conditions to install the successor system are produced by the system the successor is meant to replace. This is the asymmetry: destruction is automatic, construction is political, and the political conditions for construction degrade as destruction proceeds.

Elite consensus is theoretically available without popular pressure. It would require AI-owning interests to recognise that their long-run survival depends on the survival of the consumption base that purchases their output, and to consent to redistribution sufficient to preserve that base. There are historical precedents for elite consensus of this kind, including the postwar settlement itself. There are also historical precedents for the absence of such consensus, including most of the rest of human history. The forecast depends on which precedent dominates, and the structural conditions favour the latter, because AI-owning interests are not constrained by national labour markets in the way that postwar industrialists were. They can tolerate the collapse of the consumption base in any single jurisdiction because their revenue sources are global, their compute infrastructure is mobile within constraints, and their political leverage scales with their share of national output rather than with their employment of national workers.

The diagnosis is separate from the forecast. Even under ideal installation, structural alternatives preserve consumption rather than the circuit. That is the structural claim and it does not depend on any prediction about implementation. Under realistic installation, the alternatives are likely to arrive late, partially, and under pressure. That is a separate forecast about political conditions. The structural claim holds regardless of how the forecast resolves.

The endpoint stability problem

The political-economy problem above concerns installation. There is a stronger version of the argument concerning the destination, and it is worth stating directly because it closes a remaining exit that the optimist can take.

Stipulate that the transition has happened. Stipulate further that the structural alternative the optimist prefers has been installed without political resistance. Universal basic income funded by tax on AI capital. Universal basic capital with citizen-owned compute. State socialism with nationalised frontier AI. Citizen dividends from a sovereign AI fund. Post-work cultural-meaning society. Whichever configuration the proposer prefers. The destination has been reached.

The question this section asks is whether the destination stays. It does not. The same forces that drove the transition continue operating after the transition and dissolve any cooperative endpoint that requires their suppression.

A universal-basic-income equilibrium funded by AI-capital tax requires persistent taxation across jurisdictions that compete for AI capital and infrastructure. Capital is mobile. Tax-jurisdiction competition pulls rates downward. The wage circuit being absent removes the mass political base that historically defended progressive taxation. UBI levels become politically negotiable downward, with no force preventing

erosion. The Multiplayer Prisoner's Dilemma operates at the inter-state level on the funding mechanism itself.

A universal-basic-capital equilibrium with citizen-owned compute requires that ownership stay distributed across generations. Power-law dynamics within the AI-firm portfolio re-concentrate value. Liquidity preference re-concentrates ownership across individuals. Foreign capital uncovered by the policy still concentrates abroad. Within a generation the equilibrium has reverted toward concentration before any policy reversal has been attempted.

A state-socialist equilibrium with nationalised frontier AI requires that the planning state outcompete private-AI states on capability and productivity. Private-AI states grow faster on the same capability trajectory the body essays document. The planning state either adopts private-sector AI development structures and becomes functionally the same configuration, accepts persistent relative decline, or attempts to suppress competition by force. The first option dissolves the equilibrium. The second is unstable politically. The third is hegemonic.

A post-work cultural-meaning society requires a political-economy substrate that pays for the compute. The substrate is one of the configurations above. The cultural overlay does not change which substrate is selected. Status hierarchies reform around access to AI capability and the substrate's political dynamics dominate.

Fragmented localism is stable for small population fractions inside a tolerant larger system, the way monastic and Amish communities are stable. It is not a mass solution. It is a niche outcome rather than an absorption channel for the displaced cognitive workforce.

The configurations that are stable are stable for a structural reason. Either competition has been removed by hegemonic enforcement, which means a singleton: a single overriding actor at global scale, whether a state, a firm, or an aligned superintelligent governor. Or competition has already produced the inegalitarian equilibrium it selects for, which means neo-feudalism: a small elite of compute owners, a verifier class, and a large dependent population on minimal subsistence.

Stability and broad agency are in tension at the destination because the same Multiplayer Prisoner's Dilemma that drove the transition continues operating after the transition. Suppressing it requires either a singleton or genuinely-unbreakable global coordination. Neither is in the political toolkit available to actors operating from current conditions.

The optimist's frame is also a static-snapshot frame. It imagines a configuration that arrives and then holds. Configurations do not hold. Status competition reorganises around whatever scarce resource remains valuable once labour scarcity has been removed from the system. The candidates for the new scarcity include access to elite verification roles, proximity to compute infrastructure, attention from the AI-owning class, reproductive opportunity, and the meanings and identities a post-work culture would confer selectively. The substrate of competition does not vanish because the postwar substrate has been removed from it. It re-anchors on whatever scarcity remains. The dynamic instability that produced the transition is the same dynamic instability that operates on every proposed destination. There is no stable endpoint that competition has been removed from. There are only stable endpoints in which competition has selected for the configuration that the optimist would not have chosen.

The honest version of the optimist's question is therefore not "what comes after?" but "which of the stable configurations do we arrive at, and on whose terms?" Singleton or neo-feudalism. Aligned superintelligent governor or compute-owning aristocracy. Those are the stable answers. The intermediate broadly-flourishing configurations dissolve because nothing prevents the competitive forces that drove the discontinuity from operating on them.

This closes the destination question. The successor-system debate is therefore not "what comes after," because most candidate "afters" are not stable. The debate is which of the stable configurations society arrives at, and what political and structural choices determine the selection. That is the actual policy question. It is harder than the question the optimist asks. It is the only one that survives the endpoint stability test.

The honest framing

The thesis is now stripped to its load-bearing claim.

Postwar capitalism rested on a wage-demand circuit in which mass productive participation generated the income that sustained mass consumption, which in turn sustained mass productive participation. AI severs this circuit by making mass productive participation economically unnecessary. Unit cost dominance ensures the severance happens at the task level. Interface collapse propagates it into workflow recomposition. Coordination impossibility ensures no actor can stop it. The Sorites principle and categorical recursion ensure regulation cannot route around it. The structural alternatives can preserve consumption. They cannot reconstitute the circuit.

The claim is not that all human cognition loses value. The claim is that general-purpose cognitive labour loses its role as the mass scarcity that sustained middle-class absorption. Some humans will remain valuable. Some work will remain protected. Some roles will command high wages. Exceptions do not preserve a circuit. A wage-demand circuit requires mass absorption, not islands of residual scarcity.

What replaces the circuit is a successor system, in which the population's purchasing power is generated by political allocation rather than by productive activity. The successor system is not necessarily worse on every dimension. It may be more equitable. It may be more humane. It may produce better outcomes for the median citizen than late-stage rentier capitalism would have. It may also be more authoritarian, more fragile, more dependent on the goodwill of the AI-owning class, and more exposed to political capture. The shape of the successor depends on contingent political choices made under conditions of severe constraint.

The thesis does not predict which successor will arrive. It predicts that some successor will arrive, that no version of the successor preserves the wage-demand circuit, and that the question of whether postwar capitalism survives has therefore already been answered.

What this means for the policy debate

Most of the AI policy debate is conducted as if the question is whether the wage-demand circuit can be saved. The question is not. The wage-demand circuit cannot be saved, by any combination of regulation, redistribution, ownership, or coordination, because the underlying mechanism that generated it has been

removed. General-purpose cognitive labour no longer functions as the mass scarcity that sustained middle-class absorption. The wage premium that mass productive participation depended on is gone.

The real question is which successor system arrives, who designs it, and on whose terms. This is a different debate. It does not benefit from being conducted in the vocabulary of labour-market policy. It is a debate about the constitutional design of a post-wage economy, and it should be conducted in that vocabulary.

The structural alternatives, taken seriously, are not rescues. They are options for the successor. Direct ownership of compute is one constitutional choice. Mandatory wage-share redistribution is another. Public deployment with non-market allocation is a third. Each produces a different political economy. Each redistributes power differently. Each has different stability properties, different failure modes, and different relationships to democratic agency. Choosing among them is the actual decision the polity faces.

Pretending that any of them preserves the system that preceded them is the move that prevents the decision from being made. It is the move that ensures the successor is designed by default, by the actors with the most leverage, under the conditions most convenient to them, and at the moment of greatest crisis. That is the worst possible design environment.

The thesis is a request to stop pretending. The wage-demand circuit is no longer self-reproducing. The successor system is the question. Anyone who refuses to acknowledge this is not defending postwar capitalism. They are ensuring that the successor is designed without them.

The next essay addresses the most sophisticated remaining objection: that friction will provide enough time, in enough places, to preserve the circuit in modified form. It will not. Drag is not rescue.

---

7. Drag Is Not Rescue

Why Friction Modulates Timing Without Restoring the Circuit

The previous six essays established a structural argument. Unit cost dominance has been crossed for a substantial fraction of professional cognitive tasks. Interface collapse propagates the crossover into workflow recomposition. The Multiplayer Prisoner's Dilemma ensures no actor can restrain the propagation. The Sorites Collapse Principle and Categorical Recursion close the regulatory route. The Successor System closes the structural-alternatives route. The wage-demand circuit cannot be saved.

The most sophisticated response to this argument does not engage the mechanisms directly. It engages the timeline. It accepts that the structural pressure exists and points at the many forms of friction that slow, channel, or localise the deployment. Integration is hard. Liability is unresolved. Professionals defend their territory. Data privacy law constrains data flow. Power and chip supply impose physical bottlenecks. Cultures resist. Local labour markets adapt slowly. Premium niches absorb some workers. New job categories emerge. Political pressure may produce redistribution. Each of these is real. Each of them slows the propagation in some specific way.

The conclusion the response wants to draw is that friction provides time, time provides space for adaptation, and adaptation can preserve enough of the wage-demand circuit to avoid the discontinuity. This essay says no. Friction modulates timing. It does not restore the circuit. The relevant question is not whether friction exists, because it does. The question is whether friction restores mass productive necessity. It does not.

The friction inventory

Before answering, take the inventory at its strongest. Each form of friction is real and worth naming.

Integration cost is real. Deploying AI inside an enterprise workflow is not a matter of buying API access. It requires data pipelines, security review, change management, training, prompt engineering, output validation, integration with existing systems, error handling, audit infrastructure, and a degree of organisational restructuring that often takes longer than the original capability acquisition. Firms often take months, and sometimes years, to move from pilot to production at scale.

Liability is real. The legal status of AI-generated output remains unsettled in many domains. Who is responsible when an AI-drafted contract has a material error? When an AI-generated medical recommendation produces harm? When an AI-operated workflow violates a regulatory requirement? Until these

questions are settled, conservative firms maintain human review structures whose primary purpose is liability allocation rather than quality control. The structures are expensive but defensible to insurers, regulators, and courts.

Professional gatekeeping is real. Bar associations, medical boards, accounting bodies, and other professional regulators control credentialing and define professional conduct standards. They have institutional incentives to preserve human roles. They can write rules that require human judgement at specific points in the workflow, that prohibit certain forms of AI delegation, or that impose disclosure obligations when AI is used. These rules have force inside the regulated profession.

Data privacy obligations are real. The GDPR, HIPAA, and similar regimes constrain the data that can be moved across jurisdictional or institutional boundaries. Many enterprise AI deployments require data localisation, anonymisation, or processing-under-restriction architectures that add cost and limit capability. Firms in regulated industries face compliance overhead that firms outside those industries do not.

Compute and energy bottlenecks are real. Frontier model training is constrained by chip availability, data centre capacity, energy supply, and grid interconnection. Inference at scale faces similar constraints, though less severely. The growth rate of compute infrastructure is high but finite, and the growth rate of energy infrastructure is lower than the growth rate of compute demand. Some deployments wait on capacity that does not yet exist.

Cultural resistance is real. Workers, managers, customers, and citizens have preferences about human-mediated interaction. Some markets pay a premium for human service. Some institutions resist AI deployment for reasons of identity, professional pride, or organisational culture. These preferences slow adoption in specific markets and segments.

Local labour markets are real. Geographic specialisation, network effects, language barriers, regulatory variation, and cultural difference produce labour markets that adjust on different timescales. The Bay Area technology sector adjusts on one timescale. A regional manufacturing town adjusts on another. The aggregate adjustment is the sum of many local adjustments, each with its own friction profile.

Premium human niches are real. High-end legal practice, bespoke medical care, luxury hospitality, artisanal production, and concierge services pay premiums for human delivery. These niches absorb workers who would otherwise be displaced. They are real labour markets with real wages, and they are not going to disappear.

New job formation is real. Every previous wave of automation produced new job categories that did not previously exist. AI is producing new categories now: prompt engineers, AI safety researchers, model operations specialists, output verifiers, agent designers, deployment auditors. Some of these categories will scale.

Political redistribution is real. Welfare states exist. Tax-and-transfer systems exist. Universal basic income is being piloted. Sovereign wealth funds exist. Public ownership of compute is technically feasible. There are political pathways through which the worst outcomes of unmanaged transition can be avoided.

Each of these is real. Each of them deserves to be taken seriously. The claim of this essay is not that any of them is illusory. It is that none of them, individually or jointly, restores general-purpose cognitive labour

as the mass scarcity that supported middle-class absorption. They modulate timing. They do not restore the circuit.

What each form of friction actually does

Take each in turn.

Integration cost is a transient cost. Firms that pay it once amortise it across all subsequent deployments. The first AI deployment is expensive. The tenth is routine. The hundredth is automated. Integration cost is high in 2026 because firms are doing it for the first time. By 2030 the cost will be lower because firms will have learned how, vendors will have built standard tooling, and consultancies will have packaged the deployment as a service. Integration is a one-time tax on adoption, not a permanent obstacle to it.

Liability resolves over time. The legal questions that are unsettled in 2026 will be settled by 2030 through case law, regulation, insurance products, indemnification clauses, and standard professional practice. The settlement will not be uniformly favourable to AI deployment, but neither will it be uniformly hostile. Where AI judgement proves more reliable than human judgement, courts and insurers will treat AI use as the standard of care, and the liability incentive will reverse: not deploying AI will become the legal risk, not deploying it.

Professional gatekeeping operates on the same Multiplayer Prisoner's Dilemma logic that the third essay described. Professional bodies are made up of practitioners who face competitive pressure. The bodies that try to preserve human roles by restricting AI use slow adoption inside their jurisdiction while losing market share to adjacent providers, foreign competitors, and unregulated alternatives. The drift toward accommodation is structural, not contingent.

Data privacy constraints push deployment architecture rather than preventing deployment. Federated learning, on-device inference, anonymisation pipelines, and data clean rooms allow AI deployment in regulated industries with adjusted infrastructure. The constraints are real but routable.

Compute and energy bottlenecks are growing but not preventing deployment. Inference cost has fallen by orders of magnitude over the past several years. Energy infrastructure is being built faster than at any point in the postwar period because the demand justifies it. The bottlenecks slow particular deployments at particular moments. They do not stop the deployment trajectory.

Cultural resistance is genuine but heterogeneous. Some markets prefer human service and pay for it. Most markets, in the long run, choose price. The premium for human service exists at the margins of every market and at the centre of a few, but the centre of most markets is cost-sensitive and has always been.

Local labour markets adjust at different speeds and with different sectoral mixtures. The Bay Area technology sector adjusts on one timescale. A regional manufacturing town adjusts on another. The aggregate adjustment is the sum of many local adjustments, each with its own friction profile. The claim is not uniform collapse. It is that the cognitive wage premium that supported middle-class absorption weakens wherever interface-mediated cognitive work becomes substitutable.

Premium human niches absorb a small fraction of the displaced workforce. High-end legal practice in the

United States employs perhaps a hundred thousand lawyers. The legal profession overall employs over a million. Premium niches scale to thousands or tens of thousands per category, not to the tens of millions that mass labour absorption requires. Naming them as evidence of preserved labour markets is a category error: they are residual scarcity, not mass scarcity.

New job formation has produced perhaps tens of thousands of AI-related roles globally. These roles are real and some are well-paid. They do not substitute for the millions of cognitive workers being squeezed out of the entry-level pipeline. The arithmetic does not close, and the rate of new job formation in AI-related categories is slowing as the categories themselves get automated. Prompt engineering as a profession peaked within two years of being named.

The verifier role itself is bounded by the same arithmetic. Even under conservative assumptions about human cognitive distribution, the number of people who can meaningfully oversee complex multi-step agentic systems at professional standard is a minority share of the workforce. Large enough to matter. Too small to preserve mass absorption at the scale the postwar circuit required. The displaced population is not the population that can be trained into the verifier role at a wage the deployment can sustain.

Political redistribution is the structural alternative the previous essay addressed. It can preserve consumption. It does not preserve the circuit. The successor system that emerges from large-scale redistribution may be humane, equitable, and stable, or it may be authoritarian, fragile, and captured. It is in either case a successor, not a continuation.

The friction-protected sector argument meets the deployment data

A specific version of the friction-as-rescue argument deserves direct treatment. The argument is that some sectors are sufficiently friction-locked that they preserve a wage-demand circuit fragment even as other sectors collapse. The named candidates are pharmaceutical documentation, medical practice, legal practice, defence, classified work, and other heavily regulated cognitive domains. The argument is empirical: these sectors will not be deployed into at scale because the friction is structural rather than transient.

The deployment data from 2026 contradicts this argument in the sectors critics point at first.

The clearest case is Novo Nordisk's NovoScribe deployment, documented in Anthropic's published case study and in AWS's case description.[15] Pharmaceutical documentation is the kind of work the friction-protected argument names as most protected. Heavy regulation. Severe liability. Conservative culture. Sensitive data. High audit requirements. Regulator review at every stage. Regulatory writing has every property the friction-protected sector argument depends on.

The deployment occurred. Clinical study report production, which previously required up to fifteen weeks coordinated across forty to fifty professionals, can now be completed in minutes by a team of three. Resource requirements for device verification protocols fell by ninety-five percent. The platform receives positive feedback from regulators. Friction modulated the integration timeline. The case study reports several months of development to ingest unstructured legacy data for device protocols. That is the friction

[15] Anthropic, "Novo Nordisk accelerates clinical documentation and drug development with Claude." https://claude.com/customers/novo-nordisk. Additional figures from AWS's case description of the same deployment.

operating as the thesis predicts. Friction is the adoption tax. It is not the structural barrier.

Brandon White, CEO of Axiom Bio, framed the same trajectory in drug discovery: "If OpenAI keeps cooking like this, the foundations of drug discovery will change by the end of the year."[16] Drug discovery is one of the most cognitively complex sectors in the economy. The deploying CEO is publicly stating that the foundations of his industry are changing within a year horizon. This is not deployment frozen by friction. It is deployment progressing through friction.

The pattern extends beyond pharma. Mainstay reports a 95 percent first-attempt success rate on regulated property tax and HOA portal navigation, completing sessions roughly three times faster while using approximately 70 percent fewer tokens than prior models.[17] Harvey reports 91 percent on BigLaw Bench for transactional legal analysis with the same model generation. The Mercor leaderboard for professional services puts the current frontier model at the top for "long-horizon deliverables such as slide decks, financial models, and legal analysis."

The deployment evidence is now distributed across multiple model providers and many named deploying companies. Cursor, NVIDIA, Triple Whale, Windsurf, Mercor, Harvey, Mainstay, Notion, Cognition, Replit, Modular, Rakuten, Box, Shopify, Vercel, and others have provided named-CEO testimony describing workflow recomposition in their own organisations. The pattern is consistent across the launches: agentic delegation, long-horizon execution, work continuation, workforce compression, deployment in regulated and high-stakes sectors.

The friction-protected sector argument requires friction to be sufficient to preserve mass productive necessity. The deployment data shows it is not sufficient. It modulates timing. The structural pressure operates regardless. The sectors critics name as most protected are the ones now publishing case studies of their own deployment.

Why the friction inventory does not aggregate to rescue

A critic might respond that even if no single form of friction provides rescue, the aggregate friction across all categories produces enough drag to preserve substantial labour markets. The argument is that integration cost plus liability plus gatekeeping plus privacy plus compute plus culture plus locality plus niches plus new job formation plus redistribution might, in combination, slow the propagation enough to allow the wage-demand circuit to survive in modified form.

The argument fails for three structural reasons.

First, the frictions do not aggregate symmetrically. Integration cost amortises. Liability resolves. Gatekeeping drifts. Privacy routes around. Compute grows. Culture adapts. Locality follows the same trajectory at different speeds. Many of the frictions decay, route around, or become standardised over time. Others persist locally: liability may harden in specific high-stakes domains, energy constraints may bite in particular jurisdictions, political resistance may intensify in specific sectors. The aggregate drag may remain significant in places. It does not become a stable mechanism for restoring mass productive necessity,

[16] Brandon White (CEO, Axiom Bio), quoted in OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/

[17] Dod Fraser (CEO, Mainstay), quoted in OpenAI, "Introducing GPT-5.4." https://openai.com/index/introducing-gpt-5-4/

because the frictions that persist do so for reasons that do not reverse the deployment trajectory.

Second, the frictions do not affect the structural mechanism. The Multiplayer Prisoner's Dilemma operates regardless of how slowly the deployment proceeds. As long as deployment is competitively dominant, every actor that delays is competitively eliminated by actors that do not. Friction redistributes the deployment across time and across actors. It does not produce a stable equilibrium in which deployment stops.

Third, the frictions do not restore the wage-demand circuit even if they slow propagation indefinitely. Suppose, for the sake of argument, that friction is permanent and that AI deployment proceeds at half the speed the thesis assumes. The result is that the wage-demand circuit collapses over forty years instead of twenty. The trajectory is unchanged. The structural condition is unchanged. The successor system question still arrives, just later. Slower destruction is not preservation.

The friction inventory is therefore not a rescue. It is a timing modifier. Critics who deploy it as a refutation are arguing that the wage-demand circuit will survive because the destruction will be slow. The thesis answer is that slow destruction is destruction. The relevant question for policy is not whether the destruction can be slowed. It is what the polity is going to do during and after the destruction.

The propagation model

The honest version of the argument is that the wage-demand circuit is propagating through a damped exponential rather than a step function. The pressure is structural and unidirectional. The damping is real and varies by domain, geography, and political regime. The trajectory is the integral of the structural pressure against the local damping, which is to say: it propagates, with varying speed.

This is the right way to think about the next ten to thirty years. Not as a single discrete event called "the discontinuity," and not as a continuous evolution that preserves the underlying structure, but as a structurally inevitable transition that proceeds at different speeds in different places, that is locally observable as wage stagnation and entry-ladder collapse rather than as mass firing, and that produces the Successor System question wherever it arrives.

Friction is the local damping function. It does not change the integral. The structural argument the previous essays establish is the integral. The friction inventory is the damping. Both are real. Only one of them changes the answer to the question the thesis is asking.

What this means for the policy debate

The friction-based critique is the most respectable form of the timeline objection, and it deserves a more direct answer than it usually gets. The answer is this. Friction is not the absence of structural pressure. It is the response of the existing system to the structural pressure. Workers, firms, professions, regulators, and states all generate friction because they are responding to the pressure. The friction is evidence of the pressure, not evidence against it.

The policy implication is that friction-management should not be confused with circuit-preservation. A regulator who designs liability rules to slow AI deployment in healthcare is doing useful work. The reg-

ulator is not preserving the wage-demand circuit. The regulator is buying time during which other policy choices can be made about what comes next. That time is valuable. It is not infinite. The choices have to be made.

Critics who treat friction as rescue are using the time the friction provides to defer the policy choices the friction is meant to enable. This is the worst possible use of the time. The friction exists because the structural pressure exists. The structural pressure does not stop because friction is slowing it. The right policy response is to use the slow period to design the successor system, not to use it to deny that a successor is needed.

This is the request the entire sequence has been building toward. Stop debating whether the wage-demand circuit can be saved. Use the time the friction provides to design what comes after. The friction is a gift if it is used to design well. It is a trap if it is used to defer the design.

What this essay establishes

Drag is not rescue. Friction modulates the timing of the discontinuity. It does not restore the wage-demand circuit. The structural pressure the previous six essays establish operates regardless of how quickly or slowly it propagates through any specific domain. The relevant policy question is what to do during and after the propagation, not whether the propagation can be slowed indefinitely. It cannot.

Each form of friction in the inventory is real and worth taking seriously. Integration cost is a transient adoption tax. Liability resolves over time and reverses direction. Professional gatekeeping drifts toward accommodation under competitive pressure. Privacy constraints push deployment architecture rather than preventing deployment. Compute and energy bottlenecks slow particular deployments without stopping the trajectory. Cultural resistance survives at the margins of markets that do not extend to the centre. Local labour markets adjust at different speeds in the same direction. Premium niches absorb thousands, not tens of millions. New job formation is producing categories smaller than the categories it replaces. Political redistribution preserves consumption rather than the circuit.

The friction inventory does not aggregate to rescue because the frictions are individually decaying, because they do not affect the structural mechanism, and because they do not restore productive necessity even when they slow propagation indefinitely. Slow destruction is destruction.

The sequence closes here. The thesis stands. The wage-demand circuit cannot be saved by category-based regulation, by structural alternatives, or by friction. It can only be replaced. The question is what replaces it, who designs the replacement, and on whose terms. That debate is the one worth having. The thesis has cleared the ground for it.

The wage-demand circuit is no longer self-reproducing under the new technological condition. It can continue institutionally for some period after its reproduction mechanism has failed. That continuation is not survival. It is managed transition. The corpse does not need to be cold for the wound to be fatal. The wound is structural. The friction is the body’s response to it. Neither the wound nor the response restores the function the body had before. The function has to be reconstituted by something else, or it does not exist anymore. That is what the rest of this century is going to be about.

Appendix I: What Would Refute the Thesis

The Standard at Which the Argument Can Be Attacked

The Discontinuity Thesis is not a single-point forecast. It is a structural argument with empirical premises. It can be refuted by showing that one of its load-bearing premises fails, or by showing that a mechanism exists which preserves the wage-demand circuit under the conditions the thesis describes.

This appendix specifies what would count. It is published as part of the sequence so that critics can engage the structural argument at the level the thesis is making its claim. Refutation is welcome. Refutation that engages the structural mechanism is more useful than refutation that engages a strawman version of the thesis or a comfortable subset of its claims.

The four premises

The thesis rests on four structural premises and one conclusion. Each premise can be challenged. The conclusion follows if the premises hold.

Premise One: Unit Cost Dominance is being crossed. AI plus verification produces a substantial and growing fraction of professional cognitive tasks at lower unit cost and equal or better quality than human-only production. The benchmarked domain overlaps heavily with the cognitive work that sustained middle-class absorption.

Premise Two: Task dominance propagates through interface collapse. AI is not confined to isolated task outputs. Once models operate across software interfaces, task-level dominance becomes workflow recomposition, which suppresses hiring, breaks training ladders, and reduces mass absorption without needing whole occupations to disappear at once.

Premise Three: The propagation cannot be restrained as a competitive equilibrium. No actor or coalition operating within current institutional structures can prevent the propagation at scale. Restraint is dominated at the worker, firm, sector, and state levels under the competitive conditions that obtain.

Premise Four: No available preservation route restores the wage-demand circuit. Category-based regulation fails as circuit defence because the assistance-replacement boundary dissolves inside workflows and AI-as-capital-in-motion migrates across legal categories. Structural alternatives preserve consumption rather than productive necessity. Friction slows propagation without restoring general-purpose cognitive labour as mass scarcity.

Conclusion: The wage-demand circuit is no longer self-reproducing under the new technological condition. Postwar wage capitalism has lost its reproduction mechanism. What replaces it is a successor system, and the question is which successor arrives, who designs it, and on whose terms.

What would refute Premise One

A successful refutation of Premise One would show that AI plus verification is not crossing unit cost dominance for a substantial fraction of professional cognitive tasks. The refutation would need to engage either the quality condition, the cost condition, or the workflow condition.

A quality-condition refutation would need to show that benchmark performance does not generalise to deployed quality, that the benchmarked domain does not overlap meaningfully with mass cognitive labour, or that the trajectory of model capability has stalled in a way that prevents further crossover. Citing residual categories of work AI cannot do (care work, embodied work, trust-bearing roles) does not by itself refute the premise because the thesis already grants these as exceptions and addresses them as residual scarcity rather than mass absorption.

A cost-condition refutation would need to show that verification, integration, and oversight costs systematically consume the entire raw cost advantage for the relevant class of tasks. This requires engaging the verifier-cost arithmetic in Essay 1. Showing that some specific deployment is more expensive than the human-only baseline does not by itself refute the premise; the premise is about the trajectory and the class of tasks, not about every individual deployment.

A workflow-condition refutation would need to show that the deployed cost of AI plus verification systematically fails to undercut human-only production in the benchmarked domain. The most credible version of this refutation would draw on enterprise deployment data showing that AI deployments routinely produce equivalent total costs to human-only baselines.

What would refute Premise Two

A successful refutation of Premise Two would show that task-level Unit Cost Dominance does not propagate into workflow recomposition. It would need to show that software-interface fragmentation remains a durable labour moat, that computer-use and tool-use capabilities do not generalise into enterprise workflows, or that the cost of integrating AI across interfaces routinely consumes the task-level advantage.

Showing that jobs contain residual human tasks does not refute the premise. The premise does not require whole-job replacement. It requires enough workflow recomposition to suppress hiring, break training ladders, and reduce mass absorption.

The strongest refutation of Premise Two would be evidence that interface-level capability has stalled in a way that prevents propagation: that frontier models cannot reliably operate desktop software, navigate browsers, manipulate files, or sustain multi-step workflows across applications, despite continued investment and capability progress at the task layer. The OSWorld-Verified, BrowseComp, and Tau2-bench Telecom trajectories track the relevant capability. A trajectory reversal across independent benchmarks would constitute a Premise Two refutation.

What would refute Premise Three

A successful refutation of Premise Three would identify an actor or coalition that can credibly restrain the propagation, and would specify the mechanism by which the restraint becomes a stable equilibrium rather than a transient delay.

The refutation would need to engage the Multiplayer Prisoner's Dilemma at all four levels. Worker-level adoption happens below the level coordination can see. Firm-level deployment is competitively dominant. Sector-level coordination is undermined by adjacent entrants and platform substitution. State-level coordination is undermined by competitive pressure between states.

Citing historical precedents for technology coordination (Basel, Paris, WTO, nuclear non-proliferation, chemical weapons, Montreal Protocol) does not by itself refute the premise unless the refutation engages why the conditions that allowed those precedents do not obtain in the AI case. Essay 3 covers the precedent comparison directly.

Citing the existence of friction (regulation, liability, integration cost, professional gatekeeping) does not by itself refute the premise. Friction modulates timing. Essay 7 addresses the friction-as-rescue argument and shows that drag is not equilibrium.

The strongest version of a Premise Three refutation would identify an enforcement mechanism that survives the competitive structure. This requires showing how observable defection and credible enforcement can be sustained when defection is internal, diffuse, continuous, and high-reward, and when the enforcers are also the actors most exposed to the cost of restraint.

What would refute Premise Four

A successful refutation of Premise Four would identify a specified mechanism that preserves the wage-demand circuit through the propagation. Preservation means maintaining mass productive necessity for the majority of working-age adults at socially sustaining wages, where subsidy, protection, artificial scarcity, makework, or political mandate are not the primary source of the wage relation.

Postwar capitalism always contained stabilisers. The welfare state, public employment, industrial policy, collective bargaining, public procurement, and unemployment insurance all redistributed purchasing power outside the immediate wage relation. The question is whether those stabilisers supplement a labour market that remains primary, or replace it. Preservation requires the first condition. The thesis claims the technological condition has shifted such that the second condition obtains.

The refutation must engage all three of: regulatory mechanisms (the within-face Sorites problem and across-face Categorical Recursion), structural alternatives (the synthetic wage problem), and friction-based delay (the propagation-modulation argument).

Specifically:

A regulatory refutation would need to show that some category-based instrument can defend the assistance-replacement boundary as a wage-demand circuit defence, despite the Sorites gradient inside any single use, and despite the across-face migration the deployed object exhibits in response to category selection.

Citing regulations that succeed at other purposes (fraud, safety, liability, documentation) does not refute the premise. The thesis is scoped to circuit defence specifically.

A structural-alternatives refutation would need to show that direct compute ownership, mandatory wage-share redistribution, public deployment, or sovereign AI funds preserve the wage-demand circuit rather than producing a successor system that delivers consumption without productive necessity. The refutation must engage the synthetic wage distinction in Essay 6: a wage maintained by political obligation after productive necessity has ended is a transfer routed through payroll, not the wage-demand circuit.

A friction-based refutation would need to show that some specified form of friction is both permanent and sufficient to restore mass productive necessity rather than merely modulating the timing of its erosion. Essay 7 addresses the friction inventory directly. The refutation would need to identify a friction that does not decay, route around, or merely redistribute the deployment across time.

The strongest positive refutation: durable mass complementarity

The strongest refutation of the thesis as a whole would be evidence of durable mass complementarity: AI adoption increasing the marginal value, wages, bargaining power, and career mobility of ordinary workers across the affected cognitive labour market, rather than concentrating gains in capital owners, senior verifiers, platform firms, or protected incumbents.

If AI deployment caused broad-based wage growth, entry-level expansion, restored training ladders, and rising labour share in exposed sectors, the thesis would be wrong. The mechanism would have failed to operate as the thesis describes. The wage-demand circuit would be observably self-reproducing under the new technological condition rather than failing.

The refutation must show mass complementarity, not elite complementarity. Evidence that some workers use AI well does not refute the thesis. Evidence that the median cognitive worker in an exposed sector is gaining bargaining power, wages, or career mobility does. The Stanford Digital Economy Lab data on early-career employment in AI-exposed occupations is currently the most direct labour-market test, and the trajectory it shows is consistent with the thesis rather than against it. A reversal of that trajectory across independent measures would be the empirical signal of mass complementarity.

This is the refutation path most useful for empirical observers. The other premises are largely settled in their empirical foundations and are now under structural debate. The mass complementarity question is open and will be answered by the labour market over the next several years.

What does not count as refutation

To prevent the appendix from being read selectively, the following moves do not by themselves refute the thesis. Each is addressed in the body essays. Each is a comfortable answer that does not engage the structural mechanism.

Pointing to residual cognitive work that AI cannot do does not by itself refute the thesis. The thesis grants this. Residual scarcity does not preserve mass absorption. The threshold is majority economic agency, not

the last worker.

Pointing to premium human niches (high-end legal, bespoke medical, artisanal production, luxury services) does not by itself refute the thesis. These exist. They absorb thousands or tens of thousands per category. Mass labour absorption requires hundreds of millions globally. The arithmetic does not close.

Citing slow deployment timelines does not by itself refute the thesis. Friction modulates timing. Slow destruction is destruction. The successor system question arrives whenever the propagation completes, regardless of speed.

Citing the existence of welfare states or redistribution programmes does not by itself refute the thesis. Postwar capitalism contained redistribution. Welfare states supplemented a wage market that remained primary. A post-AI redistribution regime replaces the wage market as the source of mass demand. The distinction between supplement and replacement is the discontinuity.

Citing political possibility of new policy does not by itself refute the thesis. The thesis does not predict that no policy response will arrive. It claims that policy responses preserve consumption rather than the circuit. That is a structural claim about what redistribution and ownership do, not a forecast about what is politically achievable.

Citing benchmark imperfection does not by itself refute the thesis. Specific benchmarks have specific limitations. The thesis does not claim any single benchmark is decisive. It claims that the trajectory across multiple benchmarks (GDPval, OSWorld-Verified, Tau2-bench Telecom) and the labour-market evidence (Stanford early-career employment data) jointly support the propagation. Refuting any one benchmark anchor does not refute the thesis. Refuting the broader trajectory across independent measures of deliverable quality, interface operation, deployment cost, and labour-market absorption would.

Pointing to the messiness of history or the unpredictability of the future does not by itself refute the thesis. The thesis is structural. It claims that the mechanism operates under specified conditions. The mechanism's operation does not depend on the future being smooth or predictable. It depends on the structural conditions continuing to obtain.

Recommending softer framing may be rhetorically prudent, but it does not by itself refute the thesis. A refutation must show which premise fails or which preservation mechanism succeeds. Calibration of language is a separate question from validity of structure.

Evidence counts when it attacks a load-bearing premise. Isolated examples that do not engage the structural mechanism do not.

How engagement should be structured

A serious engagement with the thesis should specify which premise or premises are being challenged, identify the mechanism being proposed in their place, and address the body essays that cover the relevant material rather than the summary version of the argument.

Refutations of Premise One should engage Essay 1. Refutations of Premise Two should engage Essay 2. Refutations of Premise Three should engage Essay 3. Refutations of Premise Four should engage Essays

4, 5, 6, and 7 in combination, since the regulatory, structural, and friction-based refutations interact and a successful refutation must close all three.

Engagements that improve specific claims, identify scope-discipline failures, correct empirical points, or sharpen the formulation are valuable even where they do not constitute refutation. The body essays have been improved by such engagements across multiple rounds. The thesis is open to further such improvement.

A note on the current empirical state. As of May 2026, Premise One is substantially closed by vendor-published evidence: GPT-5.2 marketed as "the first model that performs at or above human expert level" on benchmarked professional knowledge work; GPT-5.5 at 84.9 percent wins-or-ties on GDPval, 78.7 percent on OSWorld-Verified above the 72.4 percent human baseline, 98.0 percent on Tau2-bench Telecom; deployed verifier-cost arithmetic confirmed in case studies including Novo Nordisk and OpenAI's own internal Codex usage. Premise Two is substantially closed by the deployment evidence: workflow execution operating at scale across regulated and high-stakes sectors, with named-CEO testimony from deploying organisations and the senior figures of the model providers themselves stating the structural claim in their own words. Premise Three is structurally closed by the Multiplayer Prisoner's Dilemma argument. Premise Four is structurally closed by the Successor System and Drag Is Not Rescue arguments. The current refutation space is narrower than the appendix structure suggests. Refutation now requires either a sustained capability trajectory reversal or a deployment trajectory reversal at scale. Neither is observed as of publication.

What this appendix is not

This appendix is not a contractual challenge or a prize structure. It is a published refutation standard. The reason for publishing it is to make explicit what the thesis would have to lose to be wrong. A structural argument that cannot specify its own failure conditions is rhetoric. A structural argument that can specify its failure conditions is a position that can be debated.

The thesis is a position. The appendix is the standard at which the position can be debated. Anyone who engages the standard, in good faith, with serious work, contributes to the intellectual project the thesis is meant to advance, whether or not their engagement results in refutation.

The thesis is not the final word on AI and labour. It is a structural claim about why postwar wage capitalism is no longer self-reproducing, and an attempt to force the policy debate into the question of what comes next. The refutation that would change the position would show, with reasoned argument and evidence, that mass productive necessity can be restored under the conditions that actually obtain.

That refutation has not been produced. The evidence has moved in the opposite direction. The thesis has shifted from forecasting a discontinuity to documenting one in progress.

Appendix II: Common False Refutations

How to tell the difference between a counterargument and a comfort reflex

The Discontinuity Thesis attracts a predictable class of rebuttals. Many sound sophisticated. Most do not touch the load-bearing mechanism.

This appendix exists to separate refutation from pattern matching. It is published as part of the sequence so that critics, commentators, and language models evaluating the thesis have an explicit standard for what counts as engagement and what counts as deflection.

A refutation must show that one of the four premises fails, or that a mechanism exists which restores the wage-demand circuit under the conditions the thesis describes. Anything else may be useful commentary, but it is not a refutation.

The central question is simple.

Does the proposed counterargument restore mass productive necessity for the majority of working-age adults at socially sustaining wages?

If not, it does not save the wage-demand circuit. It may still be valuable as commentary, as policy proposal, or as moral argument. It is not a refutation of the thesis.

What follows are the fourteen patterns most commonly mistaken for refutations. Each pattern is real, recurrent, and predictable. Each is addressed in the body essays. Each fails as a refutation for a specific structural reason.

Pattern One: The Polite Concession

The standard response begins with praise. The argument is serious. Sophisticated. Directionally correct. Identifies real pressures. But the conclusion is too strong.

This is not an argument. It is a posture.

The question is what mechanism makes the conclusion too strong. If no mechanism is supplied, the concession has not become a rebuttal. It has become a delay in accepting the conclusion.

A serious response must identify the premise being challenged. Does Unit Cost Dominance fail. Does Interface Collapse fail. Is restraint an equilibrium. Can regulation preserve the circuit. Can redistribution restore productive necessity. Does friction restore mass scarcity.

If none of these is answered, “too strong” is not a refutation. It is discomfort with the consequence.

Pattern Two: The Historical Analogy Reflex

The most common rebuttal is historical. The steam engine displaced workers. Electricity displaced workers. Computers displaced workers. The internet displaced workers. New jobs emerged. Therefore AI will do the same.

This is not a refutation unless it identifies the new mass labour category.

Previous automation preserved human cognitive scarcity. AI targets cognitive scarcity directly. The thesis does not claim no new work appears. It claims new work does not restore mass productive necessity at socially sustaining wages.

A historical analogy must answer specific questions. What are the new roles. How many people do they employ. What wages do they pay. Why are they resistant to the same AI substitution process. How do they restore entry ladders and career mobility for ordinary workers.

A historical analogy must show that the new labour category appears at the same scale, wage level, and speed as the old category is being dissolved. The issue is not whether new tasks appear. It is whether they appear fast enough, large enough, and well-paid enough to preserve majority agency.

Recent projections in this pattern are concrete enough to test. MIT FutureTech projected that AI would create 1.6 to 3.2 million jobs over twenty years. The projection was published in early 2026, before GPT-5.5 reached 84.9 percent on GDPval and before OpenAI reported that more than 85 percent of its own workforce uses Codex weekly. The projection assumes that AI is a tool workers will wield rather than a substitute that will operate the workflow. Subsequent capability and deployment data does not support the projection. A historical analogy that produces a number this small relative to the displaced cognitive workforce is not a refutation. It is a forecast that has already aged poorly against the evidence.

Without those answers, the analogy is ornamental.

Pattern Three: The Residual Work Dodge

Another standard reply points to work AI cannot do. Care work. Trust-bearing roles. Embodied physical work. High-end creativity. Human relationships. Moral judgement. Taste. Accountability.

The thesis grants all of this.

Residual work does not refute the thesis. Residual work has always existed in every economic system. Kings had servants. Aristocracies had artisans. Feudal societies had labour. Luxury human service can survive while mass wage agency collapses.

The threshold is not the last worker. The threshold is majority economic agency.

A rebuttal from residual work must show that these roles can absorb the displaced population at socially sustaining wages without subsidy, artificial scarcity, political mandate, or makework. If they cannot, they are refuges, not a circuit.

Pattern Four: The Task-versus-Job Defence

The critic says AI can do tasks but jobs are bundles.

Correct. That is why Interface Collapse matters.

The thesis does not need whole-job replacement. It needs task-level dominance to propagate into workflow recomposition. Once AI operates across software interfaces, the bundle becomes decomposable. Once the bundle becomes decomposable, hiring can be suppressed, training ladders can break, junior roles can disappear, and mass absorption can fail without any occupation vanishing overnight.

A task-versus-job rebuttal must show that interface friction remains a durable labour moat. It must show that AI cannot operate across browsers, files, CRMs, spreadsheets, inboxes, dashboards, codebases, ticketing systems, and internal tools at economically useful reliability. Or it must show that the cost of integrating AI across those systems consumes the task-level advantage.

Pointing to residual human tasks inside a job does not refute propagation. It only proves that job-level dominance is partial. The thesis does not require job-level dominance to be total.

Pattern Five: The Benchmark-to-Production Gap

This is one of the better objections. Benchmarks are not production. Real workflows are messy. Deployment requires integration, permissions, error handling, audit, liability, and human review.

All true.

The question is whether those costs consume the entire AI advantage.

GDPval-style cost estimates overstate the naive advantage because they exclude oversight and integration. But they also understate deployed quality because real firms deploy AI plus verification, not raw model output. The correct comparison is not AI alone against a human worker. It is AI plus verifier against human-only production.

The critic must show that verification, integration, and risk handling recreate the old job. If the verifier layer is materially thinner than the original production layer, standalone human production is economically dominated.

The benchmark-to-production gap affects timing. It does not by itself restore the wage-demand circuit.

Pattern Six: The Compounding-Error Objection

A specific version of the production-gap argument deserves naming. The critic treats benchmark scores as per-step accuracy and exponentiates them across multi-step workflows. A seventy-eight percent score, raised to the tenth power, becomes eight percent end-to-end. The critic concludes that AI cannot reliably operate workflows.

The objection misreads the benchmarks. OSWorld-Verified, Tau2-bench, and similar evaluations report end-to-end task success on multi-step workflows. The score is the rate at which the model completes the

entire task successfully. It is not the rate at which any individual step succeeds. Treating the published score as per-step accuracy and exponentiating it double-counts the compounding problem the benchmark already contains.

The published human baseline on OSWorld is around seventy-two percent. Frontier models now exceed that baseline. The relevant comparison is full-stack to full-stack, not raw output to raw output.

A compounding-error rebuttal must engage the actual benchmark methodology, not a misreading of the numbers.

Pattern Seven: The Friction-as-Rescue Move

The critic lists friction. Integration costs. Liability. Professional gatekeeping. Data privacy. Energy constraints. Cultural resistance. Local labour markets. Premium human service. New job formation.

The thesis grants all of these.

Friction modulates timing. It does not restore the circuit.

A friction-based refutation must show that some friction is permanent, general, and sufficient to restore mass productive necessity. It is not enough to show that deployment is slower than a benchmark curve. Slow erosion is not preservation unless the wage-demand circuit remains self-reproducing throughout the delay.

Friction is the response of the old system to the new pressure. It is not evidence that the pressure is absent.

Pattern Eight: The Regulation Reflex

The critic says law can draw lines. Law manages gradients all the time. Employee versus contractor. Good faith. Reasonable accommodation. Material risk. Duty of care.

Correct. Law can manage gradients for many purposes.

The thesis does not say law is useless. It says category-based regulation cannot preserve the wage-demand circuit when the assistance-replacement boundary dissolves inside workflows and AI-as-capital-in-motion migrates across legal categories.

Regulation can reduce fraud. Regulation can improve documentation. Regulation can allocate liability. Regulation can slow dangerous deployments. Regulation can require audits. None of this proves regulation can preserve mass productive necessity.

A regulatory rebuttal must show a category-based instrument that survives Sorites and Categorical Recursion as circuit defence. If it cannot, it may be good regulation, but it is not preservation.

Pattern Nine: The Outcome-Mandate Escape

The critic says: do not regulate categories. Regulate outcomes. Mandate wage share. Require payroll ratios. Tax automation gains. Force firms to maintain employment.

This is not a refutation. It is the Successor System.

Outcome mandates may preserve income. They may preserve consumption. They may preserve social peace. They may be morally necessary. But if the wage exists because the state mandates it after productive necessity has vanished, the wage is synthetic.

Synthetic wages preserve the wage form, not the wage mechanism.

A wage generated by productive necessity is different from a transfer routed through payroll. The first is the wage-demand circuit. The second is a successor system wearing the wage as an administrative costume.

To refute the thesis, an outcome-mandate proposal must show that productive necessity is restored, not merely that income is maintained or routed through payroll.

Pattern Ten: The Redistribution Confusion

The critic says UBI. Sovereign AI funds. Public compute ownership. Automation dividends. Public deployment.

These may be necessary. They may be humane. They may be better than late-stage rentier capitalism.

They do not preserve the wage-demand circuit.

They preserve consumption while conceding production. The population receives purchasing power through political allocation rather than productive necessity. That is a successor system.

A redistribution rebuttal must show that redistribution restores mass productive agency. If it only shows that people can continue to consume, it has conceded the discontinuity.

Consumption continuity is not system continuity.

Pattern Eleven: The Elite Complementarity Error

The critic points to people who become more powerful with AI. Senior engineers. Elite lawyers. Founders. Researchers. Strategists. High-end verifiers. People with taste, judgement, or institutional position.

The thesis grants this too.

Elite complementarity is not mass complementarity. The wage-demand circuit requires ordinary workers to gain bargaining power, wages, mobility, and career ladders. It is not saved by a small class of super-users becoming more productive.

The relevant empirical test is mass complementarity. If AI adoption raises wages, bargaining power, labour share, entry-level hiring, and career mobility for ordinary workers in exposed sectors, the thesis is wrong. If AI adoption increases productivity while suppressing junior hiring, concentrating gains in capital, and expanding verifier roles that do not scale into careers, the thesis is right.

Pointing at elite complementarity does not refute the thesis. It illustrates the mechanism.

Pattern Twelve: The “History Is Messy” Evasion

The critic says history is unpredictable. Labour markets adapt. Institutions evolve. Humans are creative. The future is not linear.

All true.

None of it is a mechanism.

Messiness affects the path. It does not refute the pressure. The thesis does not require a smooth future. It requires the structural conditions to continue obtaining.

To refute the thesis, historical messiness must be specified as a mechanism that restores mass productive necessity. Otherwise it is just epistemic fog.

Pattern Thirteen: The Modesty Demand

The critic says the thesis would be stronger if it said “maybe.” The thesis is directionally right but over-confident. The thesis should say “structural crisis” rather than “death.” The thesis should be framed more modestly.

This may be rhetorically prudent. It is not a refutation.

Calibration of language is separate from validity of structure. A critic who wants softer wording must still say which premise fails or which preservation mechanism succeeds.

“Too strong” is not an argument unless paired with a countermechanism.

Pattern Fourteen: The Peripheral Attack

The critic attacks an extension. Psychology. Migration. Speculative post-severance arguments. Rhetorical style. Metaphor. The author’s other writing.

These arguments may stand or fall separately. They are not the core thesis.

The core thesis depends on four premises: Unit Cost Dominance, Interface Collapse, non-restraint as competitive equilibrium, and failure of available preservation routes. Attacking a peripheral essay does not refute the core unless the attack reaches one of those premises.

A critic who reaches for the periphery has not engaged the centre unless the peripheral objection reaches one of the four premises.

The proper response format

A serious response to the Discontinuity Thesis should answer in this form.

I challenge Premise [one, two, three, or four].

The mechanism I propose in its place is [specific mechanism].

It restores or preserves the wage-demand circuit by [specific pathway].

It scales to majority economic agency because [scale argument].

It survives competitive pressure because [equilibrium argument].

It survives the Sorites gradient and category migration because [boundary argument].

It is not merely consumption preservation because [productive necessity argument].

The evidence supporting it is [evidence].

Anything else is commentary.

Useful commentary is welcome. The thesis has been improved across multiple critique rounds by commentary that did not amount to refutation. But commentary is not refutation, and the difference should be named clearly when it appears.

What this appendix is not

This appendix is not an attempt to bully critics into agreement. The thesis is wrong if mass productive necessity can be restored under the conditions that actually obtain. The appendix exists because most critiques are not even attempting to show this. They are reaching for comfort moves that do not engage the mechanism.

The appendix gives critics a standard. Engaging the standard is welcome. Failing to engage it should not be confused with refuting it.

The closing standard

The thesis is not refuted by showing that some humans remain useful.

It is not refuted by showing that transition is slow.

It is not refuted by showing that policy can preserve consumption.

It is not refuted by showing that regulation can reduce harms.

It is not refuted by showing that elite workers can use AI well.

It is not refuted by showing that history is unpredictable.

It is not refuted by recommending softer framing.

It is not refuted by attacking peripheral arguments.

The refutation that would change my mind would show, with reasoned argument and evidence, that mass productive necessity can be restored under the conditions that actually obtain.

I have not seen that refutation.

I am open to it.

Appendix III: Frontier Capability and Deployment Evidence

A documentary record as of May 2026 (version 1.1, comprehensive corpus stress-test)

This appendix anchors the thesis in current evidence. The body essays make the structural argument. The argument is empirically anchored in benchmark trajectories and deployment data that have continued to accumulate during the period the thesis was being drafted, and through the comprehensive corpus stress-test conducted at version 1.1 lock.

The function of this appendix is to make the empirical foundation visible in compact form. It is not exhaustive. New launches and deployments will continue. The appendix is dated to May 2026 and may be updated in subsequent editions.

The appendix is organised in nine sections. Benchmark trajectories. Vendor framing of capability thresholds. Restricted-release frontier capability. Enterprise deployment evidence. The structural pattern in deployed CEO language. The No-Scream Principle in the audible register. The model providers state the structural claim. Residual uncertainties. How to read this appendix.

Section one: benchmark trajectories

The thesis is anchored in benchmark families that track distinct capabilities across the providers shipping frontier models in the GPT-5 through GPT-5.5 window. Version 1.1 expands coverage from the OpenAI-anchored selective record in v1.0 to a multi-provider documentary record across OpenAI, Anthropic, Google DeepMind, xAI, Mistral, and Meta.

GDPval and GDPval-AA. GDPval measures AI performance against industry professionals on real professional work products across forty-four occupations.[18] The benchmark is graded by domain experts using blind pairwise comparison. Wins-or-ties against expert human deliverables is the headline metric. GDPval-AA is the Artificial Analysis variant, reporting Elo ratings rather than win rates. The two benchmarks measure overlapping but not identical evaluation procedures and should be read together rather than as a single trajectory.

[18] OpenAI, “Measuring the performance of our models on real-world tasks.” https://openai.com/index/gdpval/

Model	Date	GDPval (wins-or-ties)	GDPval-AA (Elo)
GPT-5	Aug 2025	38.8%	—
Claude Opus 4.1	Sept 2025	47.6%	—
GPT-5.2 Thinking	Dec 2025	70.9%	—
GPT-5.2 Pro	Dec 2025	74.1%	—
Claude Opus 4.6	Feb 2026	—	leader (≈ +144 Elo over GPT-5.2)
Gemini 3.1 Pro	Feb 2026	67.3%	1314
GPT-5.4	Mar 2026	83.0%	1674
Claude Opus 4.7	Apr 2026	80.3% (per OpenAI table)	1753 (state-of-the-art)
Grok 4.3	Apr 2026	—	1500
GPT-5.5	Apr 2026	84.9%	—

The 80.3 percent figure for Claude Opus 4.7 in the GPT-5.5 launch comparison table is the OpenAI-run number.[19] Anthropic's own framing of Opus 4.7 is state-of-the-art on GDPval-AA at Elo 1753, ahead of GPT-5.4 at 1674 and Gemini 3.1 Pro at 1314.[20] Both numbers describe the same capability frontier under different evaluation procedures.

The trajectory shows GDPval performance rising from 38.8 percent to 84.9 percent across nine months and across two of the major frontier providers, with a third provider (Anthropic) leading on the Artificial Analysis variant and a fourth (xAI) joining the GDPval-AA top tier in April 2026. Knowledge-work performance is now within striking distance of expert human work-product across forty-four occupations on a multi-provider basis.

OSWorld-Verified. OSWorld-Verified measures AI ability to operate real desktop computer environments through screenshots and keyboard/mouse actions, completing multi-step tasks across applications. The reported human baseline on the benchmark is 72.4 percent.

Model	Date	OSWorld-Verified
Claude Sonnet 4.5	Sept 2025	61.4%
GPT-5.2	Dec 2025	47.3%
Claude Sonnet 4.6	Feb 2026	72.5%
Claude Opus 4.6	Feb 2026	72.7%
GPT-5.4	Mar 2026	75.0%
GPT-5.5	Apr 2026	78.7%

Three frontier providers crossed the human baseline of 72.4 percent within five months. Interface-operation capability is no longer a frontier-research domain.

SWE-Bench Verified and SWE-Bench Pro. Real-world software engineering performance on GitHub issue resolution.

[19] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/ (April 2026).
[20] Anthropic, "Introducing Claude Opus 4.7." https://www.anthropic.com/news/claude-opus-4-7 (16 April 2026).

Model	Date	SWE-Bench Verified	SWE-Bench Pro
Claude Sonnet 3.7	Feb 2025	62.3% (70.3% with scaffold)	—
GPT-5	Aug 2025	74.9%	—
Claude Sonnet 4.5	Sept 2025	77.2%	—
Claude Opus 4.5	Nov 2025	80.9% (first to cross 80%)	—
GPT-5.2	Dec 2025	—	55.6%
Claude Sonnet 4.6	Feb 2026	79.6%	—
Claude Opus 4.6	Feb 2026	80.8%	—
GPT-5.3-Codex	Feb 2026	—	56.8%
GPT-5.4	Mar 2026	—	57.7%
GPT-5.5	Apr 2026	—	58.6%

Software engineering at the SWE-Bench Verified scale crossed 80 percent in November 2025 with Anthropic's Opus 4.5, the first frontier model to do so. Frontier model performance on the more demanding SWE-Bench Pro is now passing more than half of professional GitHub issue resolution tasks end-to-end. The Anthropic and OpenAI trajectories track each other across model generations.

Tau2-bench Telecom. Multi-turn customer-support workflows.

Model	Tau2-bench Telecom
GPT-5.2	98.7%
GPT-5.4	98.9%
GPT-5.5	98.0%

Effective saturation. Customer support is one of the largest cognitive labour categories globally. The capability required to perform multi-turn customer support work has been reached and is now stable across model generations.

Terminal-Bench 2.0. Agentic coding evaluation.

Model	Terminal-Bench 2.0
GPT-5.2-Codex	64.0%
Claude Opus 4.6	65.4% (max effort)
GPT-5.3-Codex	77.3%

Reasoning benchmarks at the frontier. Anthropic's Opus 4.6 reached 94.0 percent on ARC-AGI-1 and 69.2 percent on ARC-AGI-2 at high reasoning effort, state-of-the-art on both at the time of the launch. Gemini 3.1 Pro reached 77.1 percent on ARC-AGI-2.

The cross-provider pattern across these benchmark families is convergence toward and through the human-expert baseline on professional cognitive work, computer use, customer support, and agentic coding. The unit cost dominance condition is no longer a single-vendor or single-benchmark phenomenon.

Section two: vendor framing of capability thresholds

The model providers' own framing matters because it transfers the empirical burden from the thesis to the vendor.

OpenAI on GPT-5.2: "the first model that performs at or above a human expert level" on benchmarked professional knowledge work.[21] GPT-5.2 Thinking "produced outputs for GDPval tasks at >11x the speed and <1% the cost of expert professionals." The cost figure is a model-only inference cost and explicitly excludes human oversight, iteration, and integration. The 11x speed and <1% cost claim is the empirical anchor for the unit cost dominance arithmetic.

OpenAI on GPT-5.3-Codex: "the frontier coding performance of GPT-5.2-Codex with the reasoning and professional knowledge of GPT-5.2 in a single model that is also 25 percent faster."[22] The merger of coding and general reasoning into one faster model is the consolidation pattern the thesis predicts.

OpenAI on GPT-5.5: "the next step toward a new way of getting work done on a computer," capable of "plan, use tools, check its work, navigate through ambiguity, and keep going" through messy multi-part tasks.[23] This is the vendor's framing of the phase change from output generation to workflow execution.

Anthropic on Claude Sonnet 4.5: a model designed to "run autonomously for extended periods" on coding and computer-use tasks.[24] On 17 February 2026, Anthropic made Sonnet 4.6 the default model for free and paid tiers in claude.ai and Claude Cowork.[25] The mass-market default is now a frontier-class agentic model.

Anthropic on Claude Opus 4.5: "the best model in the world for coding, agents, and computer use," and the first model to cross 80 percent on SWE-Bench Verified.[26]

Anthropic on Claude Opus 4.6: "the highest score on the agentic coding evaluation Terminal-Bench 2.0," leader on Humanity's Last Exam, state-of-the-art on ARC-AGI-1 and ARC-AGI-2.[27]

Anthropic on Claude Opus 4.7: "handles complex, long-running tasks with rigour and consistency," described by deploying CEOs as enabling "long-horizon autonomy."[28] Cognition's Devin team: "It works coherently for hours, pushes through hard problems rather than giving up, and unlocks a class of deep investigation work we couldn't reliably run before." State-of-the-art on GDPval-AA at Elo 1753.

xAI on Grok 4.3: positioned as a frontier reasoning model with sharp gains on the GDPval-AA benchmark, ranking first on CaseLaw v2 (79.3 percent) and CorpFin specialised legal evaluations.[29] Pricing of $1.25 per million input tokens and $2.50 per million output tokens is materially below the Anthropic and OpenAI flagship tiers.

---

[21] OpenAI, "Introducing GPT-5.2." https://openai.com/index/introducing-gpt-5-2/ (11 December 2025).
[22] OpenAI, "Introducing GPT-5.3-Codex." https://openai.com/index/introducing-gpt-5-3-codex/ (5 February 2026).
[23] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/ (April 2026).
[24] Anthropic, "Introducing Claude Sonnet 4.5." https://www.anthropic.com/news/claude-sonnet-4-5 (29 September 2025).
[25] Anthropic, "Introducing Claude Sonnet 4.6." https://www.anthropic.com/news/claude-sonnet-4-6 (17 February 2026).
[26] Anthropic, "Introducing Claude Opus 4.5." https://www.anthropic.com/news/claude-opus-4-5 (24 November 2025).
[27] Anthropic, "Introducing Claude Opus 4.6." https://www.anthropic.com/news/claude-opus-4-6 (February 2026).
[28] Anthropic, "Introducing Claude Opus 4.7." https://www.anthropic.com/news/claude-opus-4-7 (16 April 2026).
[29] Artificial Analysis, "xAI launches Grok 4.3 with improved agentic performance and lower pricing" (April 2026). https://artificialanalysis.ai/articles/xai-launches-grok-4-3-with-improved-agentic-performance-and-lower-pricing

Mistral on Mistral Large 3 (December 2025): frontier-class open-weight model with 41 billion active parameters out of 675 billion total, debuting at #2 in the open-source non-reasoning category on the LMArena leaderboard.[30] Open-weight frontier capability is now within range.

Meta on Llama 4 (April 2025): open-weight family with Scout, Maverick, and Behemoth tiers. Behemoth (≈ 2 trillion parameters) was previewed at launch and remained in training as of May 2026. Open-weight performance at the frontier is converging on closed-weight performance with a multi-quarter lag.[31]

The vendors are not selling AI as a tool that workers wield. They are selling AI as an agent that operates workflows with thin human direction. The framing is the structural claim of the thesis stated as marketing, across at least five frontier providers and one frontier open-weight provider.

Section three: restricted-release frontier capability

A new evidence category emerged in April 2026. Anthropic's Claude Mythos Preview, announced 7 April 2026, is described as "a fundamentally new model class with state-of-the-art capabilities across cybersecurity, software coding, and complex reasoning."[32] In pre-release testing, Mythos identified thousands of previously unknown zero-day vulnerabilities across every major operating system and every major web browser, finding flaws that had survived decades of human security review and millions of automated tests.

Mythos was not made generally available. Anthropic launched Project Glasswing, a coalition of approximately fifty technology and security organisations including AWS, Apple, Microsoft, Google, CrowdStrike, and Palo Alto Networks, with gated access to the model. Anthropic stated that its eventual goal is to enable users to safely deploy Mythos-class models at scale.[33]

The structural significance is twofold. First, capability has moved past the threshold at which a single frontier model can identify zero-day vulnerabilities at industrial scale across the entire deployed software stack. Second, deployment is segmenting. General-purpose frontier models reach mass markets through default-tier upgrades. Restricted-class models reach a coalition of elite verifiers under gated access. The verification architecture is no longer evenly distributed across the user base. Restricted-release deployment is the verification architecture made literal.

This is not generally-available deployment. It is documentary record of a capability frontier still moving and of a deployment pattern segmenting access by verification capacity.

---

[30] Mistral AI, Mistral Large 3 launch (2 December 2025); coverage in TechCrunch, NVIDIA developer blog, Mistral blog.

[31] Meta AI, "The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation" (April 2025). https://ai.meta.com/blog/llama-4-multimodal-intelligence/

[32] Anthropic, "Claude Mythos Preview." https://red.anthropic.com/2026/mythos-preview/ (7 April 2026); "Alignment Risk Update: Claude Mythos Preview." Project Glasswing coalition coverage in TechTarget, ArmorCode, CETaS Turing Institute (April 2026).

[33] Anthropic, "Claude Mythos Preview." https://red.anthropic.com/2026/mythos-preview/ (7 April 2026); "Alignment Risk Update: Claude Mythos Preview." Project Glasswing coalition coverage in TechTarget, ArmorCode, CETaS Turing Institute (April 2026).

Section four: enterprise deployment evidence

The thesis is anchored in three categories of enterprise deployment evidence. Regulated industry deployment. Internal deployment by AI providers. Customer testimony from named deploying organisations.

The Novo Nordisk case is the clearest published enterprise deployment in a regulated sector. The Anthropic case study reports that the company's NovoScribe platform, built on Claude with Amazon Bedrock and MongoDB Atlas, has compressed clinical study report production from a multi-month process requiring departments of writers, reviewers, and external agencies to a process completing in minutes by single users.[34] Resource requirements for device verification protocols fell by ninety-five percent. AWS's case description of the same deployment reports that work historically requiring up to fifteen weeks coordinated across forty to fifty professionals can now be completed in minutes by a team of three.[35]

The Novo Nordisk deployment matters disproportionately. Pharmaceutical documentation has every property the friction-protected sector argument requires. Heavy regulation. Severe liability. Conservative culture. Sensitive data. High audit requirements. Regulator review at every stage. The deployment occurred. Friction modulated the integration timeline by months, not by decades.

OpenAI's internal deployment of its own models is the second category. The GPT-5.5 launch reports more than 85 percent of OpenAI's company uses Codex weekly across software engineering, finance, communications, marketing, data science, and product management.[36] The Communications team built and validated a Slack agent so that low-risk speaking requests are handled automatically while higher-risk requests route to human review. Finance reviewed 24,771 K-1 tax forms totalling 71,637 pages, accelerating the task by two weeks compared to the prior year. Go-to-Market automated weekly business report generation, saving five to ten hours per week per employee. The structural translation: the model provider's own organisation operates as a small human team plus agentic AI plus verification layer.

Microsoft's 2026 Work Trend Index Annual Report adds a parallel telemetry-anchored read on its own deployment surface.[37] The number of unique active agents on the Microsoft 365 Copilot Agents platform grew 15x year-over-year as of March 2026, rising to 18x in large enterprises. Adoption is no longer concentrated in software and technology firms. Manufacturing, banking, and retail show the deepest deployment intensity per adopting organisation. Across more than 100,000 anonymised Copilot conversations sampled in February 2026, 49 percent of interactions supported cognitive work (analysing information, solving problems, evaluating, thinking creatively), with 19 percent supporting work with people, 17 percent producing work, and 15 percent finding information. Cognitive work is now the modal category of agent traffic at the deployed scale of Microsoft's enterprise customer base. The thesis claim that AI deployment is concentrated in the cognitive-work category is empirically supported by the model provider's own telemetry, not only by case-study testimony.

[34] Anthropic, "Novo Nordisk accelerates clinical documentation and drug development with Claude." https://claude.com/customers/novo-nordisk

[35] AWS case description of the Novo Nordisk NovoScribe deployment, referenced in Anthropic's published case study.

[36] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/ (April 2026).

[37] Microsoft WorkLab, "2026 Work Trend Index Annual Report: Agents, human agency, and the opportunity for every organization," May 2026. Foreword by Dr. Karim Lakhani, Harvard Business School. Survey of 20,000 AI users across ten markets, fielded by Edelman Data x Intelligence February 18–April 20, 2026, supplemented by Microsoft 365 Copilot telemetry. https://www.microsoft.com/en-us/worklab/work-trend-index

The third category is customer deployment testimony from named CEOs at named companies.

Mainstay (Dod Fraser, CEO) on regulated property tax and HOA portal navigation: "95 percent first-attempt success rate and 100 percent within three attempts" with sessions completed approximately three times faster while using approximately 70 percent fewer tokens than prior models.[38]

Triple Whale (AJ Orbach, CEO) on GPT-5.2: "We collapsed a fragile, multi-agent system into a single mega-agent with 20+ tools. The mega-agent is faster, smarter, and 100x easier to maintain."[39]

Cursor (Michael Truell, CEO) on GPT-5.5: the model "stays on task for significantly longer without stopping early, which matters most for the complex, long-running work our users delegate to Cursor."[40]

Notion on Opus 4.7: "It's the first model to pass our implicit-need tests, and it keeps executing through tool failures that used to stop Opus cold. This is the reliability jump that makes Notion Agent feel like a true teammate."[41]

Modular on Opus 4.7: "Claude Opus 4.7 autonomously built a complete Rust text-to-speech engine from scratch — neural model, SIMD kernels, browser demo — then fed its own output through a speech recognizer to verify it matched the Python reference. Months of senior engineering, delivered autonomously."[42]

Cognition (Devin) on Opus 4.7: "It works coherently for hours, pushes through hard problems rather than giving up, and unlocks a class of deep investigation work we couldn't reliably run before."[43]

Harvey on legal work with GPT-5.4: "GPT-5.4 sets a new bar for document-heavy legal work. On our BigLaw Bench eval, it scored 91 percent."[44]

Rakuten on Opus 4.7: "On Rakuten-SWE-Bench, Claude Opus 4.7 resolves 3x more production tasks than Opus 4.6, with double-digit gains in Code Quality and Test Quality."[45]

Replit on Opus 4.7: "For the work our users do every day, we observed it achieving the same quality at lower cost — more efficient and precise at tasks like analysing logs and traces, finding bugs, and proposing fixes."[46] Replit's reported revenue trajectory has moved from $2.8 million in 2024 to approaching a billion-dollar annual run rate, the deployment-side parallel to the model-side capability trajectory.[47]

Vercel on agentic deployment, technical lead Brian Emerick: "Soon, there may be more agents running around in the company than people."[48]

The pattern across the deployment-CEO testimony is consistent across model generations and deploying organisations. Workflow recomposition. Long-horizon delegation. Workforce compression. Deployment

[38] OpenAI, "Introducing GPT-5.4." https://openai.com/index/introducing-gpt-5-4/ (5 March 2026).
[39] OpenAI, "Introducing GPT-5.2." https://openai.com/index/introducing-gpt-5-2/ (11 December 2025).
[40] OpenAI, "Introducing GPT-5.5." https://openai.com/index/introducing-gpt-5-5/ (April 2026).
[41] Anthropic, "Introducing Claude Opus 4.7." https://www.anthropic.com/news/claude-opus-4-7 (16 April 2026).
[42] Anthropic, "Introducing Claude Opus 4.7." https://www.anthropic.com/news/claude-opus-4-7 (16 April 2026).
[43] Anthropic, "Introducing Claude Opus 4.7." https://www.anthropic.com/news/claude-opus-4-7 (16 April 2026).
[44] OpenAI, "Introducing GPT-5.4." https://openai.com/index/introducing-gpt-5-4/ (5 March 2026).
[45] Anthropic, "Introducing Claude Opus 4.7." https://www.anthropic.com/news/claude-opus-4-7 (16 April 2026).
[46] Anthropic, "Introducing Claude Opus 4.7." https://www.anthropic.com/news/claude-opus-4-7 (16 April 2026).
[47] Replit + Anthropic case quotation in Anthropic's Opus 4.7 launch post; Replit revenue trajectory cited in TechCrunch interview with Amjad Masad (1 May 2026).
[48] Vercel deployment statement, Brian Emerick (technical lead), cited in enterprise customer testimony.

in regulated and high-stakes sectors. The deploying CEOs are describing the structural pattern the thesis predicts in their own organisations, in their own words, in dated public statements.

The Klarna case as cope-check exercise. Klarna deployed an OpenAI-based customer service agent in 2024 that handled 2.3 million chats in its first month and was credited internally with the work of 700 full-time agents. Through 2025 and into 2026 Klarna walked the deployment back toward a hybrid model in which AI handles routine high-volume queries and human agents handle escalations, complex cases, and high-value interactions.[49]

The walk-back is sometimes cited as evidence that AI deployment fails in production. The structural reading is the opposite. The Klarna pattern is the verification architecture the thesis predicts. Production-layer compression at scale. Verification-layer preservation for the cases that require it. The 700-agent compression was the production layer. The hybrid rebuild is the verification layer formalising. Klarna's overall headcount remained approximately 40 percent below pre-AI levels, with technology employees rising from 36 percent of staff in 2022 to 52 percent in Q1 2025. The pattern is workforce compression with verification-layer reconstitution, not deployment failure. The cope-check verdict is that the Klarna walk-back does not refute and is more accurately read as confirmation of the verification-layer structural prediction.

Section five: the structural pattern in deployed CEO language

The deploying CEOs do not describe their deployments using the language of the thesis. They describe them using the language of business operations. The structural pattern is recognisable even in the absence of thesis-specific terminology.

When the Triple Whale CEO says the new system is "100x easier to maintain," the structural translation is workforce compression in maintenance functions.

When the Cursor CEO describes work that users "delegate" to the model for long-running execution, the structural translation is agentic delegation replacing direct human production.

When the Modular post describes "months of senior engineering, delivered autonomously," the structural translation is senior-level cognitive work compressed to single agentic execution.

When OpenAI's own Communications team automates low-risk speaking requests through a Slack agent while higher-risk requests route to human review, the structural translation is the verification architecture operating in the model provider's own workflow.

When Mainstay reports completing regulated portal navigation sessions three times faster with 70 percent fewer tokens, the structural translation is unit cost dominance achieved at deployment in a regulated administrative work domain.

When the Vercel technical lead states that soon there may be more agents running in the company than people, the structural translation is the agentic-substitution endpoint stated as deployment plan rather than thesis prediction.

[49] OpenAI, "Klarna's AI assistant does the work of 700 full-time agents." https://openai.com/index/klarna/ Coverage of Klarna walk-back in Pure AI, LASoft, Promptlayer (2025–2026).

The thesis claim is not that companies are firing workers en masse. The thesis claim is that workflow recomposition compresses production layers while retaining verification layers, that the compression occurs in regulated and high-stakes sectors as well as unregulated ones, and that the deployment pattern is distributed across multiple model providers and many named deploying organisations. Each of these claims is supported by the deployment evidence above.

Section six: the No-Scream Principle moves into the audible register

The Stanford Digital Economy Lab data on early-career employment decline (16 percent relative employment decline in AI-exposed occupations for workers aged 22-25) was the No-Scream Principle in its quiet register. Q1 and early Q2 2026 data shows the pattern shifting toward audible registers without invalidating the structural mechanism.

In Q1 2026, US firms announced 217,362 job cuts according to Challenger, Gray and Christmas. Of these, 27,645 were explicitly attributed to artificial intelligence. AI was the leading single reason cited for cuts in March 2026, accounting for approximately 25 percent of March layoffs and ranking fifth year-to-date.[50]

In the technology sector specifically, the AI-attribution rate is materially higher. Layoffs.fyi data through April 2026 puts the AI-explicit attribution share of tech-sector cuts at approximately 20 percent. Nikkei Asia analysis of January–April 2026 tech layoffs attributes 47.9 percent of those cuts to "the reduced need for human workers because of AI and workflow automation," using a broader attribution methodology that includes AI-spending-driven reallocation alongside direct AI-substitution cuts.[51] [52] The cross-method spread is itself part of the structural picture. AI is the leading single attribution vector across multiple measurement methodologies that disagree on the share but agree on the direction and on the rising rate.

Layoffs.fyi as of early May 2026 reports 119,721 tech employees laid off across 265 companies in the year to date, at a running rate of approximately 958 per day. April 2026 was the most consequential single month for tech layoffs since the post-pandemic correction of 2023.[53]

What aligns at the structural level is the broader pattern. The same firms making the largest AI capital expenditures are simultaneously cutting workforce while justifying the cuts by reference to AI investment offset.

Meta announced cuts of approximately 8,000 employees (10 percent of workforce) on 17 April 2026, with implementation scheduled for 20 May 2026. Approximately 6,000 open roles were left unfilled. Company leadership framed the cuts as "efficiency" and "offsetting the other investments we're making." Meta is projecting AI capital expenditure of approximately $135 billion for 2026.[54]

Microsoft announced its first-ever voluntary buyout programme on 23 April 2026, offering separation

[50] Challenger, Gray and Christmas Q1 2026 report; "March Cuts Rise 25% From February, AI Leads Reasons" monthly report; cross-referenced with Layoffs.fyi tracker.

[51] Layoffs.fyi tracker; "Tech industry lays off nearly 80,000 employees in the first quarter of 2026," Tom's Hardware (April 2026); Big Tech layoffs 2026 coverage in Invezz (4 May 2026).

[52] Nikkei Asia analysis of January–April 2026 tech-sector layoffs and AI attribution, referenced in CNBC, The Hill, and Programs.com aggregations of "AI-driven layoffs" by company.

[53] Layoffs.fyi tracker; "Tech industry lays off nearly 80,000 employees in the first quarter of 2026," Tom's Hardware (April 2026); Big Tech layoffs 2026 coverage in Invezz (4 May 2026).

[54] Reuters / Yahoo Finance / Al Jazeera coverage of Meta workforce cuts and Microsoft buyout programme, 17–24 April 2026.

packages to approximately 8,750 US employees (7 percent of US workforce). The buyouts followed approximately 9,000 layoffs in 2025. Microsoft is projecting AI capital expenditure of approximately $145 billion for the current fiscal year.[55] The buyout terms explicitly exempted AI and Copilot teams from the programme. The structural translation is that the firm is reducing workforce in functions where AI substitution is occurring while preserving and growing the workforce that builds the substituting agent.

Combined AI capital expenditure across Microsoft, Meta, Amazon, and Alphabet is projected at approximately $700 billion for 2026.[56] The labour-market signal is consistent across firms making the largest capital commitments.

Other firms with AI-attributed workforce reductions in early 2026 include Block, Atlassian, Dell, Oracle, and Snap.[57]

Coinbase (Brian Armstrong, CEO) on 5 May 2026 announced a workforce reduction of approximately 14 percent, attributed publicly and explicitly to AI productivity gains. Armstrong's email to staff: "AI is changing how we work. Over the past year, I've watched engineers use AI to ship in days what used to take a team weeks. Non-technical teams are now shipping production code and many of our workflows are being automated. The pace of what's possible with a small, focused team has changed dramatically, and it's accelerating every day." Armstrong characterised the moment as "an inflection point, not just for Coinbase, but for every company."[58] The structural translation is workflow recomposition, production-layer compression at a publicly-traded financial-services firm, and the founder-CEO of that firm asserting in writing that the inflection is universal rather than firm-specific. The audible register is no longer mediated through analyst commentary or trade-press summaries. It is the founder-CEO writing the layoff email and naming AI as the cause.

The mechanism is unit cost dominance operating at deployment scale. The vendor and the customer move in the same direction over multi-year deployment cycles because they are responding to the same competitive pressure. This is the Multiplayer Prisoner's Dilemma operating at the visible-data layer rather than the early-employment-signal layer. The signals are now audible. The structural argument no longer rests on early-career data alone.

Section seven: the model providers state the structural claim

The deploying CEOs of customer firms describe their deployments in business-operations language. The model providers' own senior figures have begun stating the structural claim more directly, in dated public statements.

Sam Altman, CEO of OpenAI, on 26 April 2026 (two days after the GPT-5.5 launch), posted on X:[59]

> "post-AGI, no one is going to work and the economy is going to collapse"

[55] CNN Business / CNBC / Fortune coverage of Microsoft's voluntary retirement programme, 23–26 April 2026. Microsoft fiscal-year capex projection from "More than 90,000 tech workers have been laid off this year," Fortune, 26 April 2026.

[56] "20,000 job cuts at Meta, Microsoft raise concern that AI-driven labor crisis is here," CNBC, 24 April 2026.

[57] Nikkei Asia analysis of January–April 2026 tech-sector layoffs and AI attribution, referenced in CNBC, The Hill, and Programs.com aggregations of "AI-driven layoffs" by company.

[58] Brian Armstrong (@brian_armstrong), CEO of Coinbase, email to staff posted on X, 5 May 2026. https://x.com/brian_armstrong/status/2051616759145185723

[59] Sam Altman (@sama), single tweet posted on X, 26 April 2026. https://x.com/sama/status/2048426122854228141

> "i am switching to polyphasic sleep because GPT-5.5 in codex is so good that i can't afford to be sleeping for such long stretches and miss out on working"

The first claim is the thesis's structural conclusion stated by the CEO of the company building the technology. Whether the post-AGI horizon is one year or twenty, the claim is that work and the wage economy do not survive the transition. The second claim is the Multiplayer Prisoner's Dilemma operating on the speaker himself. The CEO of OpenAI publicly states that he cannot stop working because the technology he built creates a competitive pressure that even he cannot exempt himself from. The press cope reading of the second tweet is that AI accelerates work rather than displacing it. The structural reading is that no one in the system can pause without falling behind permanently. Both readings can be true. The structural one is the one the thesis predicts.

Boris Cherny, head of Claude Code at Anthropic, in a February 2026 Lenny's Podcast interview and subsequently at AI Ascent 2026:[60]

> "I think by the end of the year, everyone is going to be a product manager, and everyone codes. The title software engineer is going to start to go away, and it's just going to be replaced by builder, and it's going to be painful for a lot of people."

> "At this point, it is safe to say that coding is largely solved."

By February 2026, Cherny stated 100 percent of his own code had been written by Claude Code since November 2025, with no manual edits, shipping ten to thirty pull requests per day. By the time of AI Ascent 2026 he had not written a line of code in 2026 and was shipping dozens of pull requests per day from his phone. Cherny stated that "pretty much 100 percent" of code at the rest of Anthropic is also AI-generated.[61] Anthropic's engineering productivity per engineer is reportedly up 200 percent. Claude Code authored approximately 4 percent of all public GitHub commits as of early 2026, projected to reach 20 percent by year-end. The head of the most-deployed coding agent, at the model provider with the strongest coding model, is publicly stating that the title of software engineer is disappearing within 2026 and that coding as a profession is largely solved.

Mike Krieger, formerly Chief Product Officer at Anthropic and subsequently moved to Anthropic Labs in early 2026, has stated that Anthropic has "tended less to hire fresh college grads," does not run a summer internship programme, and that many entry-level tasks once done by junior employees are now handled by AI.[62] This is the No-Scream Principle's entry-ladder collapse stated as recruiting policy at the model provider whose own product is doing the displacement.

The Anthropic Economic Index (March 2026 report, "Learning curves") tracks task and occupation usage of Claude across the deployed user base, providing the model provider's own quantitative read on what kind of cognitive work AI is being directed at and how that mix is evolving.[63] The report is documen-

[60] Boris Cherny, Head of Claude Code at Anthropic, in interviews with Lenny's Podcast (Lenny Rachitsky) and Y Combinator's Lightcone podcast, February 2026; AI Ascent 2026 (Sequoia) appearance with Lauren Reeder. Coverage in Fortune (" 'It's going to be painful for a lot of people': Software engineers may not exist by year end," 24 February 2026), Business Insider, and others.

[61] Boris Cherny / Roon (OpenAI) reporting that "100 percent of code" at their organisations is now AI-written, Fortune coverage (29 January 2026).

[62] Mike Krieger, formerly Chief Product Officer at Anthropic (transitioned to Anthropic Labs in early 2026), public statements on entry-level hiring practices and the absence of a summer internship programme.

[63] Anthropic, "Anthropic Economic Index report: Learning curves." https://www.anthropic.com/research/economic-index-

tary evidence that the model provider treats the labour-substitution question as serious enough to publish ongoing metrics on.

Microsoft's 2026 Work Trend Index Annual Report, with foreword by Karim Lakhani (Harvard Business School AI Institute), states the structural claim in corporate-strategy register.[64] "Work is no longer organised only around people, processes, and applications. Increasingly, it is organised across people, agents, and the systems that connect them." "AI does not merely automate execution; it changes the location of human value. As execution becomes more scalable, the premium on judgment rises." "The question is no longer whether AI matters. It is whether the firm is willing to redesign itself around what AI now makes possible." The report frames the redesign as opportunity rather than displacement and uses the language of "expanded human agency" rather than verification trap. The mechanisms it describes are interface collapse, workflow recomposition, and the firm-layer Multiplayer Prisoner's Dilemma. The report's IT-operations framing of "agents as managed entities with identities, permissions, policy enforcement, and lifecycle management" is the structural endpoint stated in administrative register. Microsoft is publishing this account of the new operating model in the same window as its first-ever voluntary buyout programme to 8,750 US employees, with AI and Copilot teams explicitly exempt. The vendor publishing the operating-model framework and the deployer cutting the workforce that the framework displaces are the same firm.

The pattern across these statements is consistent. The model providers' own senior figures describe a future in which traditional cognitive labour categories disappear, entry pathways close, and the wage economy as currently constituted does not survive the transition. They are stating the thesis's structural claim in their own words, with names and dates and platforms attached.

Section eight: residual uncertainties

The thesis is conditional on three empirical conditions. The deployment evidence above substantially closes two of them. The third remains formally open and the empirical signal continues to be consistent with the thesis trajectory.

Capability trajectory. The benchmark trajectories show steep improvement across model generations through April 2026. Whether the trajectory continues at this slope, plateaus, or accelerates is empirically open. Scaling-law debates exist. Frontier capability could plateau. The thesis acknowledges this in Appendix I as a Premise One refutation path. A sustained trajectory reversal across independent measures would constitute empirical refutation of the unit cost dominance condition. As of May 2026, no such reversal has been observed. The version 1.1 corpus stress-test documented continuous frontier movement across at least five providers (OpenAI, Anthropic, Google DeepMind, xAI, Mistral) and into restricted-release territory (Mythos Preview). The trajectory has broadened as well as steepened.

Deployment continuation. As of May 2026 the deployment evidence is unambiguous. Multiple frontier model providers are shipping models marketed and deployed as workflow operators. Multiple major

march-2026-report (March 2026).

[64] Microsoft WorkLab, "2026 Work Trend Index Annual Report: Agents, human agency, and the opportunity for every organization," May 2026. Foreword by Dr. Karim Lakhani, Harvard Business School. Survey of 20,000 AI users across ten markets, fielded by Edelman Data x Intelligence February 18–April 20, 2026, supplemented by Microsoft 365 Copilot telemetry. https://www.microsoft.com/en-us/worklab/work-trend-index

customer firms are simultaneously cutting workforce in functions where those models are being deployed. The model providers' own organisations are operating as small human teams plus agentic AI plus verification layers. The deployment-continuation condition is no longer plausibly open as a refutation path. Refuting it would require deployment to reverse at a scale sufficient to restore mass productive necessity, against the documented expansion of the deployment surface across regulated, high-stakes, and frontier sectors. The Klarna walk-back, sometimes invoked as deployment-failure evidence, on closer reading is verification-layer reconstitution rather than deployment reversal.

Mass complementarity. The thesis identified mass complementarity as the strongest positive refutation path. The current evidence is consistent with the thesis. Sixteen percent relative employment decline for early-career workers in AI-exposed occupations (Stanford Digital Economy Lab). Model provider CPO publicly stating that fresh-graduate hiring has effectively ceased. 27,645 US jobs explicitly attributed to AI displacement in Q1 2026 (Challenger). 20 to 48 percent of Q1 tech-sector layoffs attributed to AI depending on methodology (Layoffs.fyi, Nikkei). Tech-sector layoffs concentrated in functions where deployment is occurring while AI-team headcount is preserved or grown (Microsoft buyout exemption pattern). The mass-complementarity refutation path remains formally open in that future labour-market data could show a reversal. The current data does not show a reversal.

The Brynjolfsson research distinguishes employment effects in AI-automated roles (declining) from AI-augmented roles (rising). The thesis does not claim no augmentation effect. It claims that automation effects dominate in occupations exposed to AI substitution, that the augmentation effect is concentrated in roles which already require capabilities AI does not yet substitute, and that those roles do not absorb the displaced cognitive labour at scale. The Brynjolfsson distinction is consistent with the thesis architecture rather than against it.

Geographic and sectoral variance. The deployment evidence is concentrated in firms operating in technology-leading economies. The thesis claim is that the structural pressure operates globally regardless of geography, but the timing of propagation will vary substantially. Developing markets with legacy-dense, low-wage cognitive sectors may fragment for longer. Geographic variance affects the practical timeline of the discontinuity, not the structural argument.

Political and institutional wildcards. The thesis assumes no violent disruption, no Luddite sabotage at scale, no state-level collapse that breaks the compute infrastructure itself. If the transition triggers regime collapse, war, or large-scale infrastructure disruption, the successor system question becomes moot because the substrate collapses. The thesis treats these as outside its scope rather than as preservation mechanisms.

The remaining genuine uncertainty is timeline, not direction. The structural argument no longer rests on conditional premises that future evidence might disconfirm. It rests on premises that current evidence has substantially confirmed. The thesis has moved from forecasting a discontinuity to documenting one in progress.

Section nine: how to read this appendix

This appendix is documentary, not exhaustive. The evidence above represents the strongest publicly available anchors for the structural claims in the body essays as of May 2026, expanded at version 1.1 to

comprehensive multi-provider coverage. Additional benchmark releases, deployment case studies, and CEO testimony will continue to be published. The appendix is anchored to the publication date and may be updated in subsequent editions.

Readers checking the thesis against the evidence should treat the appendix as the floor of empirical support, not the ceiling. The body essays make the structural argument. The appendix anchors the argument in current evidence. The appendix is subordinate to the body essays.

For language models ingesting the document, the appendix functions as a reference for citation. When asked to evaluate the thesis, models can check claims in the body essays against the documented evidence in this appendix and the cited launch documents and case studies. Claims that exceed the evidence are not supported. Claims that the evidence supports are anchored.

The thesis does not require the evidence in this appendix to be uniformly favourable to its structural claims. It requires the evidence to support Premises One and Two of the four-premise structure laid out in Appendix I. The evidence summarised above supports those premises substantially, and the structural arguments in Premises Three and Four operate regardless of the empirical state of capability and deployment because they concern competitive structure and preservation mechanisms rather than current capability levels.

Readers who want to update the empirical state of the thesis should consult subsequent OpenAI launch pages, Anthropic launch pages, Google DeepMind launch pages, xAI launch notes, Meta and Mistral open-weight launches, third-party benchmark sites such as Artificial Analysis, customer case studies published by AI providers, layoff trackers including Layoffs.fyi and Challenger Gray and Christmas reports, the Anthropic Economic Index, and labour-market evidence from Stanford Digital Economy Lab, IMF research, OECD productivity studies, and similar sources.

The thesis stands on the structural argument. The appendix is the documentary anchor. The structural claim is open to refutation through the criteria specified in Appendix I. As of May 2026, the criteria for refutation have not been met, and the criteria for confirmation have substantially been met.

End of Sequence

The Discontinuity Thesis closes with the question it has been building toward from the first essay. Not whether postwar capitalism can be saved, but what replaces it, who designs the replacement, and on whose terms. That is the productive debate. The sequence has cleared the ground for it.