Toward a Bad Job Economy: AI Adoption, Agency Costs, and Job Design

Matthias Fahn, Jin Li and Chang Sun

March 2026

Abstract

We study how AI affects compensation and job design when performance depends on non-contractible effort and humans remain needed to perform or verify tasks. In a principal–agent model with limited liability, AI reduces effort costs but disproportionately lowers the cost of achieving satisfactory performance. This raises the incentive cost of sustaining high effort and can induce firms to replace high-wage, high-effort good jobs with low-wage, low-effort bad jobs, even when good jobs create more total surplus. As a result, AI can lower wages, reduce worker welfare, and even depress profits. If workers can adopt AI unilaterally, adoption occurs even when the resulting equilibrium harms both parties; when adoption requires worker cooperation, resistance is strongest where AI erodes rents embodied in good jobs. In a search-and-matching extension, endogenous outside options amplify these forces and can generate multiple equilibria, reinforcing a bad-job economy.

1 Introduction

A pressing question in economics is how AI reshapes jobs—how work is organized and how workers who remain employed are compensated. Much of the literature (such as Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018a) takes a task-based view: jobs are bundles of tasks, and technology automates some tasks while leaving others to humans or creating new ones. In this view, workers' outcomes depend on which tasks remain human-performed given their skill set (a direct technological effect) and on how the resulting demand and supply for those skills adjust in general equilibrium (an indirect scarcity effect).

An aspect that receives less emphasis is that work is rarely a simple yes/no input that firms can "purchase" at a fixed quality. Performance—and thus the value created by labor—depends not only on workers' skills, but also on the effort and motivation they exert. When designing jobs and compensation schemes, firms must therefore take into account how they affect incentives. Since AI can change the costs of and returns to effort, its implications for jobs may extend well beyond the binary question of whether a task can be profitably automated.

Indeed, evidence shows that modern AI makes it easier to reach satisfactory performance across a wide range of tasks. Studies find sizable gains in mid-level writing, customer support, creative

content, and prediction, suggesting that AI sharply reduces the effort required to produce "good enough" output (Noy and Zhang, 2023; Dell'Acqua et al., 2024; Brynjolfsson et al., 2025; Kanazawa et al., 2025). These gains are concentrated in early stages—moving from nothing to a usable draft or plausible solution—while improvements are smaller in later stages that ensure correctness and full contextual fit. As a result, AI flattens the effort–output relationship and lowers the marginal return to additional discretionary effort at higher output levels (Chen et al., 2025). Put differently, when AI augments human effort, its productivity effects are often strongly nonlinear.

In this paper, we argue that this shift can undermine incentives because workers' temptation to go for the AI-enabled shortcut increases. We show how firms respond optimally by redesigning compensation schemes and, ultimately, jobs themselves, and we derive the consequences for profits and worker welfare. A central message is that these outcomes depend on the external environment, especially workers' outside options, which we treat first as exogenous and then endogenize in a general-equilibrium framework. We also suggest that policies may help sustain a "good job" equilibrium with high effort and high wages, rather than allowing a passive drift toward "bad jobs" with low effort and low pay.

To study these issues, we consider a setting in which a firm hires a worker to perform a task. This setup admits a broad interpretation. In one case, the task itself must remain human-performed, and AI primarily augments the worker's productivity. In another, the task could in principle be automated, but humans are still needed in practice to verify, refine, and supervise AI-generated output (see Lee et al., 2025; Fahn and Li, 2026; Catalini et al., 2026, for arguments and evidence that AI shifts human effort away from content generation and toward verification). Then, the key implication is that even ostensibly automatable tasks often continue to require human effort to ensure quality. Because this verification effort is costly as well and thus must be incentivized, also the success of AI-driven automation depends on the right design of incentives.

We formalize these ideas using a principal–agent model, where the agent's high effort guarantees a high output, while low effort yields a high output only with some probability. Because wages cannot be negative, inducing high effort requires rents whenever the agent's outside option is low. This creates agency costs: the principal faces a trade-off between maximizing total value and minimizing rent extraction.

A key implication is that the equilibrium can feature two qualitatively different employment relationships. When the agent's outside option is sufficiently high, the principal offers a *good job* with high wages and high effort. This occurs either because the outside option already exceeds the agency cost, so high effort can be implemented without rents, or because outside options are intermediate and the value of high effort still outweighs the agency cost, making it optimal to pay a rent. When the outside option is low, by contrast, inducing high effort becomes too costly, and the principal optimally offers *bad jobs* with low effort, a low wage, and no rent.

We model AI as reducing the worker's cost of both low and high effort, with a larger reduction for low effort. This "baseline-boosting" assumption is consistent with evidence that AI mainly accelerates the early stages of production and generates larger productivity gains for lower-baseline

workers in writing, customer support, and related prediction/assistance tasks (Chen et al., 2025; Noy and Zhang, 2023; Brynjolfsson et al., 2025; Kanazawa et al., 2025). This compression of the high–low effort cost gap has two immediate consequences, one for compensation design and one for job design. For **compensation design**, when the principal offers a good job, AI makes high effort cheaper for the worker but can raise the principal's incentive cost for sufficiently low outside options: the smaller gap between high- and low-effort costs increases the incentive payments required to sustain high effort, so the wage required to implement high effort rises in that range. For **job design**, the principal may therefore optimally switch from a good job to a bad job in order to avoid higher incentive costs, thereby expanding the range of outside options for which bad jobs are offered. In this sense, when "good enough" becomes easier to achieve, firms may find it optimal to redesign jobs toward merely satisfactory performance. The strength of this shift depends on the type of AI: it is more pronounced when AI primarily lowers low-effort costs, and weaker when it substantially lowers high-effort costs.

Such a shift from good to bad jobs harms workers because any rents previously associated with employment disappear. This contrasts with the standard labor-augmenting interpretation in the task-based literature, where technologies that raise workers' productivity in the tasks they continue to perform tend to increase wages, absent displacement or task-price effects (Acemoglu et al., 2025).

Moreover, firm profits need not rise with this shift either. This is because moving from a good to a bad job reduces not only the rent paid to the worker, but also the value generated by the relationship. When the worker's outside option, and thus his reservation wage, is relatively high, or when the good job did not involve a rent to begin with, the loss in relationship value dominates, so profits fall.

More generally, profits can decline even when AI does not induce a shift from good to bad jobs. For an intermediate range of outside options, organizational production costs—the sum of the worker's personal effort cost and the agency costs of eliciting effort—rise even though personal effort costs fall. In this region, the increase in agency costs absorbs much of the efficiency gain. By contrast, when outside options are relatively high (so good jobs remain optimal before and after AI) or relatively low (so bad jobs remain optimal throughout), the reduction in effort costs dominates and profits rise.

Next, we turn to AI adoption. A key distinction is whether workers can use AI even without the firm's approval or whether firms can control adoption. The first case is especially relevant for tools such as chatbots, which are difficult to restrict or monitor. When workers can adopt AI on their own, they will always do so, since for a given compensation scheme AI makes them weakly better off by lowering effort costs. Any decline in utility then arises only through the firm's equilibrium adjustment of compensation and job design. As a result, AI can leave both firms and workers worse off, despite raising individual productivity. When firms can in principle decide on adoption unilaterally, they will want to adopt whenever it raises profits. Even then, however, implementation may remain frictional. In the range where AI increases profits but harms workers, the resulting conflict of interest becomes important whenever adoption requires worker cooperation or formal

assent, as under works-council codetermination. Because the firm's profit gain is often too small to offset the worker's loss, this predicts organizational resistance precisely where AI induces a shift toward bad jobs.

Finally, we endogenize the agent's outside option and show that our main results carry over if it is determined in a standard Mortensen–Pissarides continuous-time random-matching framework (Pissarides, 2000). Firms post vacancies at a cost and vacancies match with unemployed workers through a random matching technology. Free entry (a zero-profit condition) pins down vacancy creation and thus the job-finding rate. The worker's outside option then consists of an exogenous component (e.g., unemployment benefits) plus the expected value of obtaining an alternative job, which depends on the job-finding rate and the utility associated with other matches.

Our main results survive in this environment: the exogenous component of the outside option continues to determine the relevant thresholds for when firms offer good versus bad jobs. But endogenizing the worker's outside option also yields additional implications. First, the impact of AI on *employment* depends on the prevailing "quality" of the jobs. When good jobs prevail, AI makes it harder to provide incentives, reduces firm profits and the number of good jobs through the labor market equilibrium. In contrast, when bad jobs prevail, AI reduces the low-effort costs, increases firm profits and the number of bad jobs.

Second, general-equilibrium feedback can amplify a shift toward bad jobs, because changes in profits affect vacancy creation, the job-finding rate, and thus workers' outside options. For example, suppose the direct effect of AI—holding vacancy posting fixed—initially keeps the economy in the good-job region, benefiting workers but reducing firms' profits. Lower profits then discourage vacancy posting, which depresses the job-finding rate and erodes outside options. If this decline is large enough, the economy may cross the threshold at which firms switch to bad jobs, leading to a sharp long-run drop in worker utility.

To conclude, our analysis suggests that the consequences of AI depend less on the technology itself than on the organizational response—specifically, on whether firms redesign incentives to sustain good jobs or instead shift toward bad jobs. This offers a different perspective on debates about "the right kind of AI" (Acemoglu, 2024). Policy discussions often emphasize steering innovation toward human-complementary technologies that create new tasks. Our analysis shows, instead, that the *same* AI technology can lead to very different outcomes depending on how firms organize work and incentives around it. A key implication is that sustaining good-job equilibria may require policy support, e.g., institutions that strengthen outside options or subsidies tied to good jobs, because otherwise the incentive costs of combining AI with high effort may make a shift toward bad jobs privately optimal for firms.

Related Literature

Our paper contributes to the following related literatures.

AI and productivity. A growing empirical literature documents substantial performance gains from modern AI tools in knowledge work and creative tasks. Field and experimental evidence shows that generative AI can raise productivity and quality in writing and knowledge-worker settings (Noy and Zhang, 2023; Dell'Acqua et al., 2024; Brynjolfsson et al., 2025), and can also improve outcomes in domains such as marketing content creation (Hartmann et al., 2025) and task performance in specific occupations (Kanazawa et al., 2025). These findings are consistent with the idea that AI can disproportionately lower the personal effort required to achieve satisfactory outcomes. Our analysis takes this "good enough becomes cheap" feature seriously and studies its implications for incentives and job design inside organizations.

Consistent with the notion that AI may weaken incentives, there is evidence consistent with the view that AI adoption is generally constrained by organizational frictions: while AI has generated notable productivity gains at the individual level, its effects on organizational efficiency remain uncertain. A report from the MIT Media Lab (Challapally et al., 2025) finds that 95% of organizations see no measurable return on their AI investments. Moreover, despite heavy investment, AI chatbots appear to have generated only modest productivity improvements in the aggregate (Humlum and Vestergaard, 2025). Consistent with this, a survey by the Centre for AI, Management and Organization at the University of Hong Kong (Fahn et al., 2025) shows that while many firms are piloting customer-facing and operational use cases, few have scaled them profitably, and nearly half report returns below expectations. Across these sources, managers point to organizational and execution barriers—rather than technical constraints—as the primary obstacles to realizing the potential of AI.

Task-based technological change. We complement the task-based literature (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019) by focusing on internal incentive costs, not task allocation alone. Recent work has started to incorporate an intensive-margin view by showing how GenAI can lower the expertise required for tasks that remain human-performed (Althoff and Reichardt, 2026; Hosseini Maasoum and Lichtinger, 2026). We interpret the same force as a reduction in the personal effort needed to reach a given performance threshold. Thus, beyond reallocating tasks between humans and machines, AI can reshape the effort–output relationship within human-performed tasks. In our framework, even when AI raises potential quality or lowers individual effort costs, the principal may optimally settle for lower effort because the cost of providing incentives increases. Moreover, while labor-augmenting AI raises wages (holding the task price constant) in task-based models, in our case higher agency costs may reduce wages in this case.

Motivation and "mediocrity traps." Our mechanism is closely related to concerns that better technology can worsen motivation when it makes acceptable performance easier to attain. Chen et al. (2025) study a "mediocrity trap" in which generative AI can reduce effort. We complement this perspective by providing a tractable contracting framework that links such motivational effects to

endogenous changes in job design: firms may respond to higher incentive costs by shifting from high-value good jobs to low-value, satisfactory-performance, bad jobs.

2 Model Setup

We now present the model and briefly discuss the key assumptions underlying the setup at the end of this section. One risk-neutral principal ("firm," "she") can hire one risk-neutral agent ("worker," "he"). The agent chooses his non-verifiable effort level $e \in \{L, H\}$, which is associated with effort costs c^e, with $c^H > c^L > 0$. While an effort level of $e = L$ reflects the "minimal effort" the agent would supply without additional incentives (e.g., due to intrinsic motivation or because this level is verifiable), high effort $e = H$ corresponds to working hard and exerting discretionary effort.

Output $Y \in \{0, y\}$ is either high ($Y = y > 0$) or low ($Y = 0$). The agent's effort determines the probability of achieving high quality. If the agent chooses $e = H$, high output $Y = y$ is obtained with probability 1. If the agent chooses $e = L$, high output is obtained with probability $q < 1$, and low output with probability $1 - q$.

We assume

$$(1 - q)y - \left(c^H - c^L\right) > 0,$$

thus high effort is efficient.

Compensation and Payoffs The agent receives a wage $w(Y)$, i.e., he is paid $w(y)$ if $Y = y$, and $w(0)$ if $Y = 0$. Thus, the principal's profit is

$$\begin{aligned} \pi^H =& y - w(y) \\ \pi^L =& q\left(y - w(y)\right) - (1 - q)w(0) \end{aligned}$$

The agent's utility from working for the principal is

$$\begin{aligned} u^H =& w(y) - c^H \\ u^L =& qw(y) + (1 - q)w(0) - c^L. \end{aligned}$$

If the agent does *not* work for the principal, he receives an outside option utility $\bar{u} \geq 0$, which is shaped by labor-market conditions such as the intensity of competition or the generosity of unemployment benefits. While we treat the outside option as exogenous in the first part of the paper, we endogenize it in a random matching model in Section 5. For now, we also assume the principal's outside option to be sufficiently low that hiring the agent is always profitable.

Finally, we assume that the agent is protected by *limited liability*, thus $w(Y) \geq 0$.

Maximization Problem We derive a subgame perfect equilibrium that maximizes the principal's profits π: The principal's chooses (e^*, $w(0)$, $w(y)$) to maximize

$$
\begin{aligned}
\max_{w(\cdot),\, e^*} \quad & \pi = \mathbb{E}\big[Y - w(Y) \mid e^*\big] \\
\text{s.t.} \quad & e^* \in \arg\max_{e} \big\{\mathbb{E}\big[w(Y) \mid e\big] - c^e\big\} && \text{(IC)} \\
& \mathbb{E}\big[w(Y) \mid e^*\big] - c^{e^*} \geq \bar{u} && \text{(PC)} \\
& w(y) \geq 0, \quad w(0) \geq 0 && \text{(LL)}
\end{aligned}
$$

The Implications of AI We assume that AI enhances the agent's productivity, reducing c^L to c^L_{AI} and c^H to c^H_{AI}, but that it reduces low-effort cost by more than high-effort cost, i.e.,

$$0 \leq c^H - c^H_{AI} < c^L - c^L_{AI},$$

which implies

$$c^H_{AI} - c^L_{AI} \in \left(c^H - c^L,\, c^H\right].$$

This assumption is based on recent evidence that the adoption of AI is associated with a "baseline-boosting" pattern in which the tool disproportionately lowers the effort required to reach acceptable performance, while leaving the incremental effort needed for excellent/reliably correct performance comparatively high (Noy and Zhang, 2023; Brynjolfsson et al., 2025; Chen et al., 2025; Kanazawa et al., 2025). We allow AI to strictly lower the costs associated with high efforts or to leave the high-effort cost unchanged, i.e., $c^H_{AI} = c^H$.

Finally, we assume that

$$(1-q)y - \left(c^H_{AI} - c^L_{AI}\right) > 0,$$

ensuring that inducing high effort remains efficient.

2.1 Discussion of Assumptions

Before deriving equilibrium outcomes, we discuss key assumptions of our setup.

Augmentation, automation, and verification We focus on tasks for which some human effort remains necessary, but AI changes what that effort looks like. In one interpretation, the worker continues to produce the output directly and AI simply lowers the cost of exerting effort. In another, AI produces a first-pass output while the worker's role shifts toward verification, refinement, and exception handling Fahn and Li, 2026; Catalini et al., 2026. Importantly, verification itself requires effort and must be incentivized. The model therefore speaks not only to pure augmentation, but also to forms of automation in which human labor remains essential for quality control. Relative to task-based models, we abstract from the extensive employment margin and instead isolate how AI changes incentives, compensation, and job design conditional on a retained human role. The success of AI-driven automation therefore depends not only on technical capability, but also on incentive design.

Effort costs rather than output gains We model AI as lowering effort costs, with a larger reduction at low effort. This captures the paper's central mechanism: when AI makes satisfactory performance easier to reach, the cost gap between high and low effort narrows, so sustaining high effort becomes more expensive under limited liability. A related alternative would let AI raise the probability of a high outcome, especially under low effort. That would lead to similar qualitative results, because it would likewise weaken incentives for high effort, but it would combine the agency-cost channel with a direct positive effect on output. Modeling AI through effort costs keeps the mechanism transparent.

Binary effort Restricting effort to $e \in \{L, H\}$ is mainly an expositional device. The same logic extends to settings with continuous effort and convex effort costs, where effort raises the probability of a high outcome and AI reduces effort costs by less at higher effort levels. In such environments, AI still compresses the return to additional effort and can therefore raise the cost of implementing higher effort when limited liability binds.

3 Outcomes Without AI

3.1 Approach: Costs of Implementing Low vs. High Effort

The principal chooses between inducing high effort and accepting low effort. We follow the standard Grossman-Hart method for moral hazard problems with discrete effort: first derive the least-cost contract that implements low effort, then the least-cost contract that implements high effort, and finally compare the resulting profits to select the optimal contract. To induce *low effort*, the principal can offer an output-independent wage that merely satisfies the participation constraint:

$$w(y) = w(0) = \bar{u} + c^L.$$

With this contract, the worker's utility is equal to his outside option, $u^L = \bar{u}$.

To induce *high effort*, the principal must design a wage schedule that makes high effort incentive-compatible and at the same time satisfies the agent's participation constraint, i.e.,

$$w(y) - c^H \geq qw(y) + (1-q)w(0) - c^L \tag{IC}$$
$$w(y) - c^H \geq \bar{u}, \tag{PC}$$

must hold. Moreover, limited liability immediately gives $w(0) = 0$, therefore (IC) becomes

$$w(y) \geq \frac{c^H - c^L}{1-q}, \tag{IC}$$

and the optimal wage schedule is the minimal $w(y)$ that satisfies both constraints, as we describe in following lemma.

Lemma 1 *If the principal wants to implement high effort, the optimal contract satisfies the following:*

$$w(0) = 0, w(y) = max\left\{\frac{c^H - c^L}{1-q},\, \bar{u} + c^H\right\}.$$

The worker earns a rent if

$$\bar{u} < \frac{c^H - c^L}{1-q} - c^H \equiv \bar{u}^B.$$

For $\bar{u} < \bar{u}^B$, the rent received by the agent increases in q. Intuitively, as low effort becomes more likely to produce high quality, it becomes harder for the principal to distinguish and reward high effort. Moreover, in this range the wage does not depend on the agent's outside option. This rent reflects the *agency cost* of incentivizing high effort when the agent's incentive constraint pins down $w(y)$. Because the agent can obtain $w(y)$ with positive probability even under low effort, he must be compensated enough to resist this temptation. Put differently, the agent's private cost of choosing high rather than low effort is $\left(c^H - c^L\right)$, whereas the principal's cost of providing incentives for high effort is $\left(c^H - c^L\right)/(1-q)$. The difference between these two amounts,

$$\frac{q\left(c^H - c^L\right)}{(1-q)},$$

is the agency cost of implementing high effort. When $\bar{u}$ is small, limited liability prevents the principal from extracting the resulting rent through negative payments in the low-output state, so it remains with the agent; in this range, neither profits nor utility depend on $\bar{u}$. By contrast, when the agent's outside option is sufficiently large so that $\bar{u} + c^L$—the minimum total payment needed to secure participation—exceeds the agency cost $q(c^H - c^L)/(1-q)$, the principal can induce high effort while keeping the agent's utility equal to his outside option.

3.2 Optimal Contract

Next, we derive the profit-maximizing contract as a function of the agent's outside option $\bar{u}$. Using the least-cost implementations of low and high effort from above, profits under low effort are

$$\pi^L = qy - \bar{u} - c^L,$$

while profits under high effort are

$$\pi^H = y - \max\left\{\frac{c^H - c^L}{1-q},\, \bar{u} + c^H\right\}.$$

When $\bar{u} \leq \bar{u}^B$, we have $\pi^H = y - \frac{c^H - c^L}{1-q}$, in which case

$$\pi^H \geq \pi^L \;\Leftrightarrow\; \bar{u} \geq \frac{c^H - (2-q)c^L}{1-q} - (1-q)y \;\equiv\; \bar{u}^A.$$

Since $\bar{u}^B > \bar{u}^A$ follows from $(1-q)y > c^H - c^L$, we obtain the following proposition.

Proposition 1 *There exists $\bar{u}^A$ such that the following holds:*

- *When $\bar{u} \leq \bar{u}^A$, the principal implements low effort and sets wages $w(0) = w(y) = \bar{u} + c^L$.*
- *When $\bar{u} > \bar{u}^A$, the principal implements high effort and sets wages*

$$w(0) = 0, w(y) = max\left\{\frac{c^H - c^L}{1-q}, \bar{u} + c^H\right\}.$$

The worker earns a rent if $\bar{u} < \bar{u}^B = \frac{qc^H - c^L}{1-q}$, where $\bar{u}^B > \bar{u}^A$.

In general, the principal faces a trade-off between maximizing total surplus (which calls for high effort) and extracting as much rent as possible from the agent. As discussed above, this trade-off—and hence the optimal contract—depends not only on the value of high output and on the agency costs (captured by q), but also on the size of the agent's outside option.

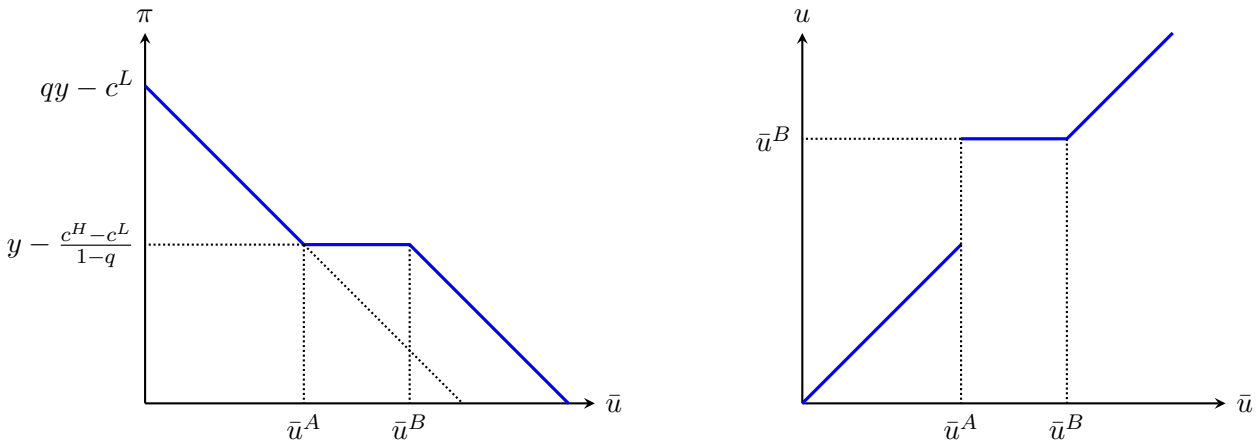
When the outside option is high enough that high effort can be implemented while keeping the agent's utility at his outside option, this is clearly optimal. If implementing high effort requires granting a rent, high effort remains optimal only when the value of high quality is large relative to the agency costs. By contrast, when the outside option is small and the value from high quality is low compared to these costs, inducing low effort is optimal and the agent earns no rent.

3.3 Interpretation: "Good Jobs" versus "Bad Jobs"

The cutoff $\bar{u}^A$ naturally lends itself to an interpretation in terms of job design. When $\bar{u} \leq \bar{u}^A$, the firm offers what we may call *bad jobs*: the contract induces low effort, pays a low wage $w \leq \bar{u}^A + c^L$, and leaves the worker with no rent. By contrast, when $\bar{u} > \bar{u}^A$, the firm offers *good jobs*: the contract induces high effort, pays a high wage $w \geq \bar{u}^B + c^H \left(> \bar{u}^A + c^L\right)$, and, for $\bar{u} < \bar{u}^B$, grants the worker a positive rent. The worker's utility therefore jumps discretely at $\bar{u}^A$, illustrating the transition from low-effort, low-pay "bad jobs" to high-effort, better-paid "good jobs" which may grant rents. It follows that, here, "good" and "bad" refer not to the intrinsic content of the task, but to how the employment relationship is organized. "Bad jobs" are roles structured around standardized, narrowly defined, "good-enough" performance: workers are expected to follow routines and deliver acceptable output with little discretion, initiative, or ownership, so pay remains low and typically contains no rent. "Good jobs," by contrast, are roles in which firms rely on workers to supply costly, noncontractible effort in the form of judgment, care, problem-solving, verification, or responsibility for quality. Because such jobs depend on worker initiative beyond the minimum, they carry higher pay and may include rents. The same underlying activity can often be organized in either way.

The following figure plots firm profits (a) and worker utility (b) as functions of $\bar{u}$, under the assumption that $\bar{u}^A > 0$.[1]

[1]Note that it may occur that $\bar{u}^A, \bar{u}^B \leq 0$, in which case one or both regions are empty under our assumption that $\bar{u} \geq 0$.



(a) Firms' profits

(b) Workers' utility

Figure 1: Optimal profits and worker utility across outside options (no AI). The solid blue line shows profits under the optimal contract. In Figure a), the dashed segment to the right of $\bar{u}^A$ represents the profits that would obtain if bad jobs were offered for $\bar{u} > \bar{u}^A$, while the dashed segment to the left of $\bar{u}^A$ continues the blue line and shows the profits that would obtain if good jobs were offered for $\bar{u} \leq \bar{u}^A$.

4 Outcomes With AI

Benchmark: Verifiable Effort We first derive a benchmark in which effort is contractible. Then, the principal can require $e = H$ and pay $w = c^H + \bar{u}$ conditional on the agent exerting high effort. The agent's utility then equals $\bar{u}$ and profits are $\pi = y - c^H - \bar{u}$, i.e., are equal to the total surplus. AI only reduces the cost term c^H,so $\Delta u = 0$, $\Delta w = c^H_{AI} - c^H < 0$ and $\Delta \pi = c^H - c^H_{AI} > 0$. Thus, with verifiable effort, profits are only affected by the reduced individual effort costs, and AI unambiguously raises profits.

4.1 Main Analysis

Recall that AI lowers the agent's effort costs, and does so more for low effort than for high effort. This forces the firm to adjust its contract because keeping the pre-AI scheme is generally bad for profits: If high effort was implemented before AI, existing incentives may no longer be strong enough. Even when high effort remains optimal for the worker without changing the contract, most of the efficiency gains would then be captured by the worker rather than the firm.

In the following, we therefore distinguish two types of responses. First, *compensation design*, where the principal adjusts the pay scheme while holding job design (good vs. bad jobs) fixed. Second, *job design*, where the principal decides whether the worker remains in a good or bad job, or is moved between them in response to AI.

Before doing that, we note that, with AI the structure of the optimal contract is identical to the one described in Proposition 1, only the cutoffs shift right. This is collected in the following Lemma:

Lemma 2 *With AI, there exist values $\bar{u}^B_{AI} > \bar{u}^A_{AI}$, where $\bar{u}^B_{AI} > \bar{u}^B$ and $\bar{u}^A_{AI} > \bar{u}^A$.*

4.2 Compensation Design

Bad jobs When the firm offers a bad job, it can cut compensation to $w_{AI} = \bar{u} + c^L_{AI}$, so profits rise to $\pi^L_{AI} = qy - \bar{u} - c^L_{AI}$. In other words, with bad jobs, AI lowers wages and raises profits, while the agent's utility remains at his outside option.

Good jobs With good jobs, the effect of AI on the optimal contract is twofold. First, AI raises the value created, which is good for the firm. Second, it makes this value harder to capture, because maintaining incentive compatibility now requires paying a higher wage to the worker:

$$w_{AI}(y) \geq \frac{c^H_{AI} - c^L_{AI}}{1-q} > \frac{c^H - c^L}{1-q},$$

so the agency cost of incentivizing high effort increases.

Put differently, AI makes it easier to do the right thing, but even easier to do the wrong thing. Under the pre-AI compensation scheme, the temptation to exert low effort becomes stronger, so the firm must raise wages to preserve incentives when the outside option is small. This may even reduce profits, as the following Lemma shows.

Lemma 3 *If the firm continues to offer good jobs with AI, then there exists a threshold $\bar{u}^{B*}$, such that the following holds:*

- *When $\bar{u} \leq \bar{u}^{B*}$, AI decreases profits.*
- *When $\bar{u} > \bar{u}^{B*}$, AI increases profits.*

Intuitively, if the participation constraint $u \geq \bar{u}$ determines the agent's compensation—so the incentive constraint (IC) is slack both before and after AI—then AI raises profits because the principal can lower the wage $w(y)$, i.e., AI reduces the cost of implementing high effort.

If instead (IC) binds both before and after AI, profits fall: higher agency costs force the principal to increase wages to preserve incentive compatibility. In the intermediate case, where (PC) determines wages before AI but (IC) binds afterward, the effect on profits is ambiguous and depends on which force dominates.

4.3 Job Design

If the principal offers a good job before AI, there are two margins along which she can adjust. First, she can modify compensation when the agent's outside option is low in order to preserve the good job, as described in the previous section. Second, she may find it optimal to switch from a good to a bad job; this latter adjustment is characterized in the following Proposition.

Proposition 2 *AI favors bad jobs.*

- *If the firm offers a bad job before AI, it continues to offer a bad job after AI.*
- *If the firm offers a good job before AI, it switches to offering a bad job if* $\bar{u} < \bar{u}^A_{AI}$.

The proposition follows directly from the upward shift in the cutoff, $\bar{u}^A \to \bar{u}^A_{AI}$ (provided $\bar{u}^A_{AI} > 0$). This shift implies that, for $\bar{u} \in \left[\bar{u}^A, \bar{u}^A_{AI}\right]$, the principal offers a good job before AI but a bad job after AI. This is because, with higher agency costs and a weak outside option, the trade-off between maximizing firm value and limiting rents to the agent becomes more severe. Conversely, if $\bar{u} > \bar{u}^A_{AI}$, a good job is offered both before and after the introduction of AI, while if $\bar{u} \leq \bar{u}^A$, a bad job is offered throughout.

This result contributes to the discussion on the potential downsides of AI, where much of the existing analysis emphasizes task substitution: if AI automates tasks previously performed by humans, employment in affected occupations may fall, and even when workers remain employed, wages can decline if the relative demand for human-performed tasks falls compared to the supply of workers with the relevant skills (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019). By contrast, in the standard task-based view, productivity gains in tasks that remain human-performed tend to raise wages, absent displacement or task-price effects (Acemoglu et al., 2025). Our results highlight an additional within-task channel. Even when the task is not technologically eliminated, because a human remains necessary either to perform it directly or to verify and refine AI-generated output, AI can still worsen job outcomes by changing the incentive structure of the job. In particular, holding labor-market demand constant (we consider general-equilibrium feedbacks in Section 5), the firm may optimally shift from a good job to a bad job if sustaining high effort becomes too costly.

We next study how the "quality" of AI affects a firm's propensity to switch from good to bad jobs, distinguishing improvements that reduce c^H_{AI} from those that reduce c^L_{AI}. The extent of this shift is captured by the size of the range $\bar{u} \in (\bar{u}^A, \bar{u}^A_{AI})$, which the next lemma is based on.

Lemma 4 *The length of the range in which AI causes a shift from good to bad jobs,* $\bar{u} \in [\bar{u}^A, \bar{u}^A_{AI}]$, *shrinks when* c^H_{AI} *falls and expands when* c^L_{AI} *falls.*

To summarize, our results predict a shift toward bad jobs, especially under two conditions. First, the shift is more likely when workers' outside options are low, for example because alternative jobs are scarce or unemployment benefits are limited. Second, it is strongest when AI primarily boosts low-effort productivity. These forces also suggest potential countervailing responses: governments can raise outside options, highlighting a complementarity between the social safety net and the viability (and even improvement) of good jobs under AI, and firms can mitigate the drift by investing in AI tools that complement high effort so that sustaining it remains profitable.

4.4 Consequences of AI-Induced Job Design

Now we examine the consequences of the AI-induced shift from good to bad jobs for worker utility and firm profits, and then discuss AI adoption.

Proposition 3 *The worker's utility decreases with the shift from a good job to a bad job.*

A move from a good job to a bad job typically eliminates the worker's rent: under a bad job the wage only needs to satisfy the agent's participation constraint, whereas under a good job (for sufficiently low $\bar{u}$) incentive compatibility forces the firm to leave rents to the worker. Hence, the shift implies a reduction in utility whenever the good job previously paid a rent. The only exception is the case in which the economy moves from a good job *without* rent to a bad job (which can occur only if $\bar{u}^B \leq \bar{u}^A_{AI}$); then the worker's utility equals $\bar{u}$ both before and after, so utility does not change.

We next turn to firm profits. Interestingly, the shift from good to bad jobs can also *reduce* profits.

Proposition 4 *Suppose* $(1-q)y > c^H - c^L_{AI}$. *Then there exists* $\bar{u}^{A*} \in \left(\bar{u}^A, \bar{u}^A_{AI}\right)$ *such that:*

- *When* $\bar{u} \leq \bar{u}^{A*}$, *a shift from good to bad job caused by AI increases the firm's profit.*
- *When* $\bar{u}^{A*} < \bar{u}$, *a shift from good to bad job caused by AI decreases the firm's profit.*

If $(1-q)y \leq c^H - c^L_{AI}$, *a shift from good to bad job caused by AI increases the firm's profit.*

Switching from a good to a bad job has two countervailing effects on profits.[2] First, the value generated by the relationship shrinks. Second, if the good job paid a rent, moving to a bad job eliminates that rent, raising the firm's share of the surplus. If the good job paid no rent (which is possible only if $\bar{u}^B \leq \bar{u}^A_{AI}$), then the second channel is absent: the wage is pinned down by the agent's participation constraint both before and after, so the shift only reduces value—and profits must fall.

If the good job did pay a rent, the rent-saving channel works in the opposite direction. For relatively low $\bar{u}$, the agent's reservation wage in the bad job is low, so eliminating rents can more than offset the loss in value, and profits rise. As $\bar{u}$ increases within the shift region, however, the wage required to fill a bad job rises with $\bar{u}$, shrinking the rent-saving gain; beyond the cutoff $\bar{u}^{A*}$, the value loss dominates and profits fall.

Finally, when $(1-q)y \leq c^H - c^L_{AI}$, low-effort production costs are low relative to the incremental value of high effort. Then, the bad-job option is sufficiently attractive that shifting to a bad job raises profits throughout (indeed, in this case AI raises profits for all $\bar{u}$).

AI Adoption The previous analysis established that when $(1-q)y > c^H - c^L_{AI}$, there exists a cutoff $\bar{u}^{A*}$ such that, within the good-job-to-bad-job region, profits—and (weakly) worker utility—decline for $\bar{u} > \bar{u}^{A*}$. Put differently, AI can be *lose-lose*: the firm optimally switches from a good job to a bad job, yet both parties end up worse off. This raises the question of why a firm would adopt AI in the first place. Indeed, adoption would not be optimal in this range if it were fully under managerial

[2] In the Appendix, we characterize the effects of AI on profits and utility for all values of $\bar{u}$, i.e., not only for the range in which AI induces a shift from good to bad jobs.

control. In practice, however, many tools—especially ubiquitous chatbots—are difficult to restrict or monitor, so firms must take worker AI use as given and respond by redesigning contracts and jobs. This is particularly true in settings where AI can be used "quietly" as a drafting or idea creation tool and its use is hard to observe, such as content-generation and marketing roles, customer-support, HR and recruiting, and software development or data analysis. Consistent with this, a recent survey (Challapally et al., 2025) reports that only about 40% of firms purchase official LLM subscriptions, yet employees in over 90% of those firms already use personal AI tools for work. Importantly, in our model the worker is always better off with AI *holding the compensation scheme fixed*, simply because effort becomes less costly; any decline in equilibrium utility reflects the firm's optimal adjustment of compensation or job design.

For $\bar{u} \leq \bar{u}^{A*}$ within the good-job-to-bad-job region, AI raises the firm's profits while (weakly) reducing the worker's utility, i.e., it generates a conflict of interest. When adoption is under managerial control (rather than driven by unavoidable worker use, as discussed above)—as is often the case for systems beyond simple chatbots—the firm adopts AI *whenever it can decide unilaterally.* Regulation, labor-representation rules, and the need for tacit employee cooperation can, however, create effective worker veto power and thereby alter this trade-off. In many European systems, employee representation provides such veto points over workplace technologies; for example, German works councils have codetermination and consultation rights under the Works Constitution Act (Federal Republic of Germany, 2001), while in France collective economic redundancies must be accompanied by a "Plan de sauvegarde de l'emploi" (République française, 2017).

Importantly, when $(1-q)y > c^H - c^L_{AI}$, the profit gain in this conflict-of-interest range is smaller than the worker's utility loss, so compensating the worker to secure cooperation is not profitable. By contrast, when $(1-q)y \leq c^H - c^L_{AI}$, the profit gain exceeds the utility loss, making compensation feasible in principle.

Moreover, the size of this range depends on how strongly AI boosts low-effort productivity. If $(1-q)y > c^H - c^L_{AI}$, this range is $\bar{u} \in \left(\bar{u}^A, \bar{u}^{A*}\right)$ and has length

$$\bar{u}^{A*} - \bar{u}^A = c^L - c^L_{AI},$$

which expands as c^L_{AI} becomes smaller.

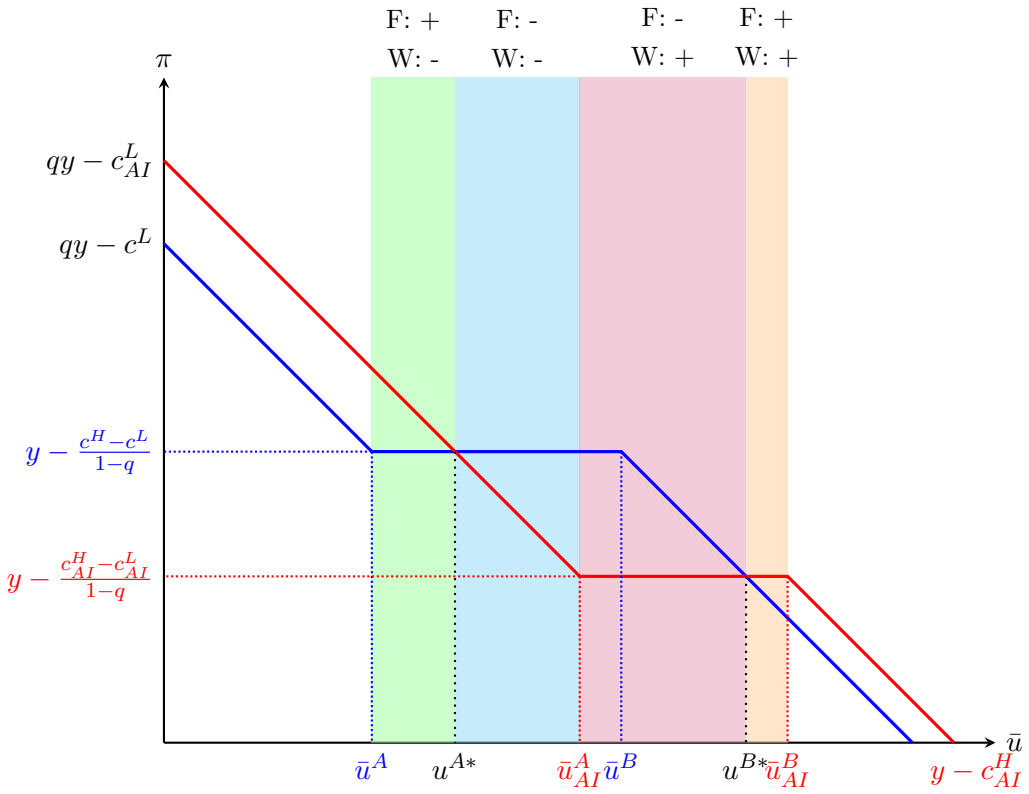


Figure 2: Firm Profit Function with and without AI. Box colors: light green, cyan, purple, orange.

5 General Equilibrium: Search, Matching and AI

In this section, we embed the earlier contracting problem into a general equilibrium model with search and matching. In this setup, we show that endogenizing workers' outside options $\bar{u}$—by interpreting them as the value of unemployment in a frictional labor market—leaves our main qualitative results unchanged, while also yields additional insights on the impact of AI on job design and employment.

5.1 The Model and Equilibrium

We consider a continuous-time economy populated by a fixed mass of workers L and an endogenous mass of firms N. The labor market is characterized by search frictions, where the formation of productive matches is governed by a matching function $\mu(U, V)$. Here, U represents the number of unemployed workers and V denotes the number of open vacancies posted by firms. The matching function is assumed to be homogeneous of degree one. We define $\theta = V/U$ as labor market tightness, which determines the job-finding rate $f(\theta)$ and the vacancy-filling rate $s(\theta) = f(\theta)/\theta$. Matches separate exogenously at rate λ.

Firms enter the market by paying a flow vacancy-posting cost κ. Once a match is formed, production occurs according to the time-invariant contract $w(Y)$ and the effort choice $e \in \{L, H\}$.

The firm's flow profit π is the residual of output minus the wage paid. We assume a discount rate r for all agents. In equilibrium, the number of firms N is determined by a free-entry condition, which dictates that firms will continue to post vacancies until the expected value of a vacancy is driven to zero.

A worker's welfare is determined by their state. While matched, a worker receives a wage and incurs an effort cost $c(e)$. While unemployed, the worker receives a flow utility b, which represents unemployment benefits, the value of leisure, or home production. The critical departure in this section is the endogenization of the worker's outside option $\bar{u}$. In our baseline, $\bar{u}$ was a fixed reservation utility; here, it is the flow value of unemployment (rV^U). This value is endogenous because it depends on the equilibrium contract (the expected utility from future employers) and the job-finding rate, $f(\theta)$. In this setup, the contracts offered by firms determine the value of being in the labor force, while the aggregate state of the labor market, i.e., the outside option, determines which contracts are optimal.

In equilibrium, the market settles into different regimes based on the unemployment benefit, b, which plays a similar role as the exogenous outside option, $\bar{u}$, in the baseline model. When b is high, firms are willing to offer "Good Jobs" – contracts that implement high effort and may provide workers with information rents. When b is low, firms find it more profitable to offer "Bad Jobs" – low-effort contracts in which workers receive no rents and are indifferent between working and unemployment.

Different from the baseline equilibrium, multiple equilibria exist when b is in an intermediate range in this general equilibrium setup. The economy might settle into either a "Good Jobs" or "Bad Jobs" state. For example, a "Good Job" equilibrium can be self-fulfilling: if all firms offer high-effort contracts with rents, the outside option stays high, which in turn makes it easier for each individual firm to provide incentives. Similarly, when all firms offer low-effort contracts, the outside option is low, making it optimal for an individual firm to offer low-effort contracts.

The existence of these distinct regimes implies that job quality is often a matter of coordination and equilibrium selection, where industry norms or "corporate culture" determine which outcome prevails. Moreover, the stability of these outcomes remains fragile, and as we will see, the introduction of AI provides a structural shock that can collapse the "Good Job" equilibrium entirely.

5.2 AI, Job Design and Employment

We model the introduction of AI as reductions in the effort costs as in the baseline model. We assume that AI reduces low-effort costs more than high-effort costs, i.e., $c^L - c^L_{AI} > c^H - c^H_{AI} \geq 0$. We maintain the assumption that inducing effort is efficient under AI, i.e., $(1-q)H > c^H_{AI} - c^L_{AI}$.

The impact of AI on job design is similar to that in the baseline model. Intuitively, AI increases information rents and makes it harder for firms to implement high efforts. The introduction of the labor market equilibrium, however, allows us to discuss the impact of AI on employment, i.e., the number of matches, M. We have the following proposition:

Proposition 5 *1. Suppose the economy is in a good job equilibrium in which the workers earn rents before and after AI adoption. The equilibrium employment level decreases after AI adoption.*

2. Suppose the economy is in a bad job equilibrium before and after AI adoption. The equilibrium employment level increases after AI adoption.

The impact of AI on employment reveals a fundamental divergence based on the prevailing "quality" of the labor market. In a Good Job equilibrium, AI adoption acts as a barrier to incentive provision: because AI lowers the costs of low-effort production more, the information rents required to motivate high effort must increase, squeezing firm profits. Consequently, firms reduce vacancy postings, leading to a tighter labor market and fewer equilibrium matches. In contrast, in a Bad Job equilibrium, AI serves only as a productivity-enhancing tool for low-effort tasks. By reducing production costs without the burden of increasing worker rents, AI raises the flow profit of each match, inducing firms to post more vacancies and increasing total employment—albeit in the form of low-utility, rent-free positions. Under certain unemployment benefit levels, the only path to employment growth after AI adoption is through the proliferation of "bad jobs," while "good jobs" become increasingly scarce and expensive to maintain.

Regarding the impact of AI on job design, the general equilibrium amplifies the difficulty of sustaining the good jobs. In Appendix Lemma **??**, we show that the cutoff unemployment benefit, below which no Good Job equilibrium can be sustained, increases. The increase can be decomposed into two parts: a direct effect holding the job-finding rate fixed at its initial level, and an indirect effect operating through the general-equilibrium adjustment of the job-finding rate. Therefore, for certain unemployment benefit levels, a Good Job equilibrium can be maintained after AI adoption if we ignore general equilibrium feedback effects, but not if we account for the change in the job-finding rate.

6 Conclusion

This paper studies how modern AI affects incentives and job design within firms when performance is verifiable only through coarse output measures and workers are protected by limited liability. The central mechanism is that AI can reduce the personal cost of producing "good enough" outcomes more than it reduces the cost of sustained high effort. By compressing the effort-cost gap, AI raises the incentive payments required to implement high effort and increases the agency cost of maintaining high-performance relationships.

This mechanism generates a shift in optimal job design. For sufficiently low outside options, firms optimally offer "bad jobs" that implement low effort and pay low wages without rents. When outside options are higher, firms offer "good jobs" with high effort, high wages, and potentially rents. AI shifts the boundary between these regimes, expanding the set of environments in which bad jobs are optimal. As a result, firm profits need not rise with AI adoption: for intermediate outside options, higher agency costs can absorb efficiency gains and can even reduce profits despite

higher individual productivity. The model also predicts when AI adoption is contested inside organizations: resistance is most likely precisely when AI induces a transition from rent-bearing good jobs to low-effort bad jobs.

Embedding the contracting problem in a frictional labor market shows that these effects can be amplified in general equilibrium. When AI reduces profits in the high-effort regime, vacancy creation falls, job-finding rates decline, and outside options weaken—potentially pushing the economy across the threshold where firms switch to bad jobs. The model thus highlights that the labor-market consequences of AI depend not only on the technology but also on equilibrium organizational responses, and expectations about other firms' behavior. Policies that strengthen outside options or subsidize high-effort implementations can help sustain a good-job equilibrium and avoid a passive drift toward bad jobs.

References

Acemoglu, Daron, "The World Needs a Pro-Human AI Agenda," Project Syndicate November 2024. Accessed 2025-12-21.

__ **and David Autor**, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1043–1171.

__ **and Pascual Restrepo**, "Artificial Intelligence, Automation and Work," SSRN Scholarly Paper ID 3098384, Social Science Research Network, Rochester, NY January 2018.

__ **and** __ , "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 2018, *108* (6), 1488–1542.

__ **and** __ , "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, 2019, *33* (2), 3–30.

__ **, Fredric Kong, and Pascual Restrepo**, "Tasks at Work: Comparative Advantage, Technology and Labor Demand," in "Handbook of Labor Economics," Vol. 6, Elsevier, 2025, pp. 1–114.

Althoff, Lukas and Hugo Reichardt, "Task-Specific Technical Change and Comparative Advantage," Technical Report January 2026.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond, "Generative AI at Work," *Quarterly Journal of Economics*, May 2025, *140* (2), 889–942.

Catalini, Christian, Xiang Hui, and Jane Wu, "Some Simple Economics of AGI," February 2026. SSRN Working Paper / MIT Sloan Research Paper, February 24, 2026.

Challapally, Aditya, Chris Pease, Ramesh Raskar, and Pradyumna Chari, "State of AI in Business 2025 Report," Technical Report 2025. Industry research report.

Chen, Yvonne Jie, Jie Gong, Jin Li, and Zibo Zhao, "Better Technology, Worse Motivation: GenAI's Mediocrity Trap," *Working Paper*, 2025.

Dell'Acqua, Fabrizio, Edward McFowland, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Working Paper 24-013 2024.

Fahn, Matthias and Jin Li, "Verification Tax: Why the AI Era Needs to Shift from "Production" to "Verification"," 2026. FT (Financial Times Chinese), Chinese-language article.

__ **, Jie Gong, and Jin Li**, "AI Adoption in Enterprise: Deloitte-HKU C-Suite Survey," Technical Report 2025.

Federal Republic of Germany, "Works Constitution Act (Betriebsverfassungsgesetz – BetrVG)," Official English translation provided by the Federal Ministry of Labour and Social Affairs (BMAS), published via Gesetze im Internet 2001. Full citation on source page: Works Constitution Act of 25 September 2001 (Federal Law Gazette I, p. 2518), as amended by Article 1 of the Act of 19 July 2024 (Federal Law Gazette 2024 I p. 248). Accessed 2025-12-23.

Hartmann, Jochen, Yannick Exner, and Samuel Domdey, "The power of generative marketing: Can generative AI create superhuman visual marketing content?," *International Journal of Research in Marketing*, March 2025, *42* (1), 13–31.

Humlum, Anders and Emilie Vestergaard, "Large Language Models, Small Labor Market Effects," NBER Working Paper 33777 May 2025.

Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe, "AI, Skill, and Productivity: The Case of Taxi Drivers," *Management Science*, 2025. Forthcoming. Earlier version: NBER Working Paper No. 30612.

Lee, Hao-Ping (Hank), Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson, "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers," in "Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems" CHI '25 Association for Computing Machinery New York, NY, USA April 2025, pp. 1–22.

Maasoum, Seyed Mahdi Hosseini and Guy Lichtinger, "Generative AI, Expertise, and Effective Labor Supply," Technical Report January 2026.

Noy, Shakked and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Science*, July 2023, *381* (6654), 187–192.

Pissarides, Christopher A., *Equilibrium Unemployment Theory*, 2 ed., Cambridge, MA: MIT Press, 2000.

République française, "Code du travail: Article L1233-61 (Plan de sauvegarde de l'emploi)," Légifrance (official portal for French law) 2017. Version en vigueur depuis le 22 décembre 2017; modifié par Ordonnance n°2017-1718 du 20 décembre 2017, art. 1. Accessed 2025-12-23.

A Appendix – General Results

First, we describe the consequences of AI on profits for all values of the outside option.

Proposition A-1 *Suppose* $(1-q)y > c^H - c^L_{AI}$. *Then there exist* $\bar{u}^{A*}$, $\bar{u}^{B*}$ *such that:*

- *When* $\bar{u} \leq \bar{u}^{A*}$, *AI increases the firm's profit.*
- *When* $\bar{u}^{A*} < \bar{u} < \bar{u}^{B*}$, *AI decreases the firm's profit.*
- *When* $\bar{u} \geq \bar{u}^{B*}$, *AI increases the firm's profit.*

If $(1-q)y \leq c^H - c^L_{AI}$, *AI increases profits for all* $\bar{u}$.

Profits can decline with AI when outside options fall within an intermediate range. In this range, *organizational production costs increase even though personal effort costs have gone down.* To identify the conditions under which profits rise or fall, we examine all possible moves within or between the regions defined in Propositions 1 and 2:

- Region 1: Bad job (low effort, no rent)
- Region 2: Good job (high effort, rent)
- Region 3: Good job (high effort, no rent).

If outside options are either sufficiently high or sufficiently low so that the system remains in Region 3 or Region 1, then profits rise. In these cases, no rent is paid to the agent either before or after AI, and optimal effort remains unchanged. The only effect is the agent's reduced personal cost of effort, which increases profits. (Note that the latter scenario may not apply if the relevant cutoff values are negative.)

If the system remains in Region 2 (rent is paid both before and after AI), profits fall. The higher agency cost of sustaining high effort dominates, and organizational production costs increase.

Profits also fall when the system transitions from Region 3 to Region 1 (which is possible if $\bar{u}^B \leq \bar{u}^A_{AI}$). In this case, no rent is paid before or after AI, but effort falls because high agency costs make it unprofitable to sustain high effort. The principal still captures the entire surplus, but the overall relationship value declines with quality.

When AI induces a transition from Region 3 to Region 2 or from Region 2 to Region 1, the effect on profits can go in either direction. A move from Region 3 to Region 2 introduces rents for the agent. This reduces profits, but lower effort costs raise them. Profits fall when the outside option exceeds the cutoff $\bar{u}^{B*}$, and rise when it does not. A move from Region 2 to Region 1 eliminates agent rents, which raises profits, but lowers implemented quality, which reduces them. Profits increase if the outside option is below the cutoff $\bar{u}^{A*}$, and decrease otherwise.

Note that the relative position of the cutoff values depends on whether $\bar{u}^A_{AI} < \bar{u}^B$ or $\bar{u}^A_{AI} > \bar{u}^B$. In the former case, $\bar{u}^{B*}$ lies between $\bar{u}^B$ and $\bar{u}^B_{AI}$, while $\bar{u}^{A*}$ lies between $\bar{u}^A$ and $\bar{u}^A_{AI}$. In the latter case, $\bar{u}^{B*}$ lies between $\bar{u}^A_{AI}$ and $\bar{u}^B_{AI}$, while $\bar{u}^{A*}$ lies between $\bar{u}^A$ and $\bar{u}^B$.

If $(1-q)y \leq c^H - c^L_{AI}$, profits increase for all values of $\bar{u}$. Since we also assume $(1-q)y \geq c^H_{AI} - c^L_{AI}$, this case corresponds to situations with relatively small y and a highly effective AI.
The following figure displays profits with and without AI, for the case in which a profit reduction happens for intermediate values of the outside option.

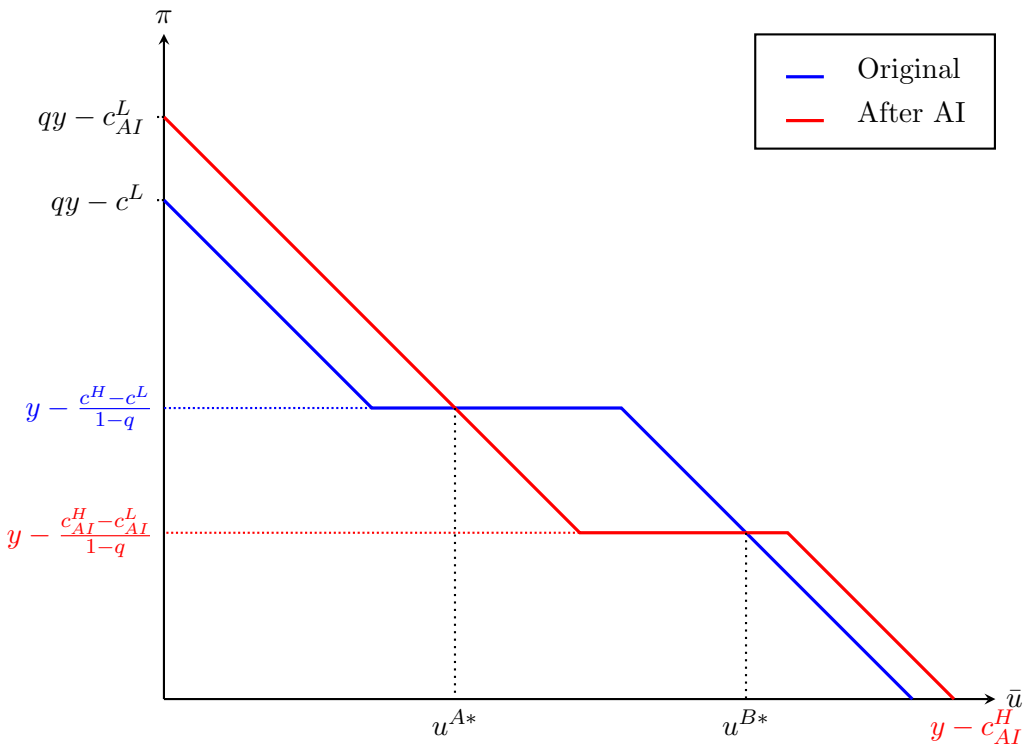


Figure A-1: Firm Profit Function with and without AI

Utility Next, we show how AI affects the agent's utility for all values of his outside option.

Proposition A-2 *The following holds:*

1. *When $\bar{u} \leq \bar{u}^A_{AI}$, the worker's utility weakly decreases with AI*
2. *When $\bar{u} > \bar{u}^A_{AI}$, the worker's utility weakly increases with AI.*

Generally, the agent's utility either increases or remains unaffected by AI, except in the case where the agent received a rent prior to AI but the introduction of AI makes high effort unprofitable. In that case, the principal switches to low effort, and the agent's payoff falls to the outside option.

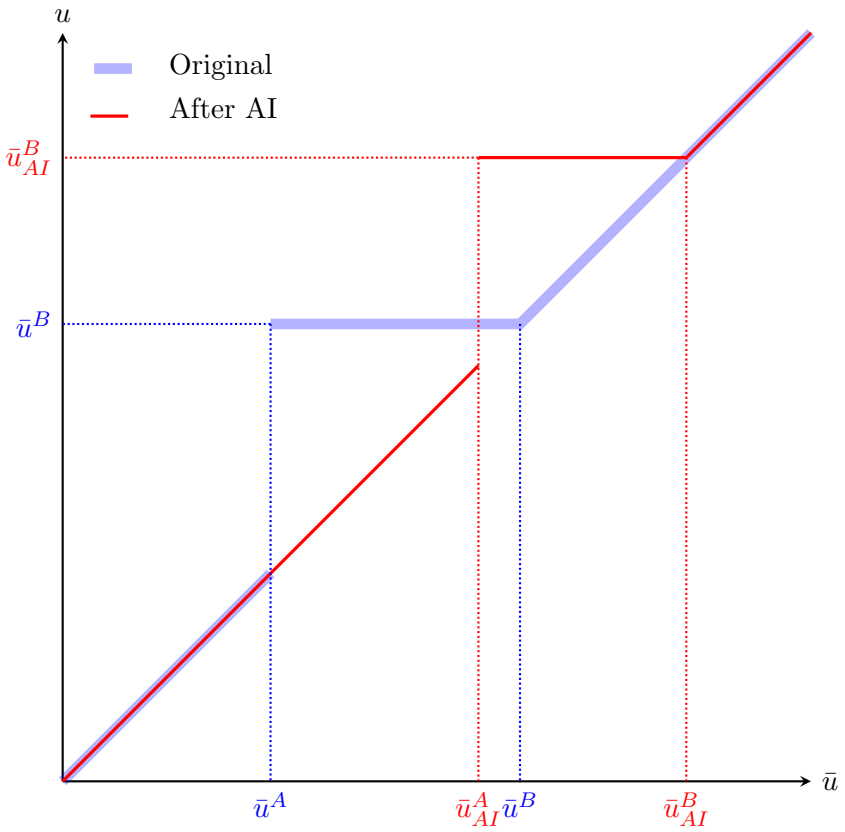


Figure A-2: Worker Utility Function with and without AI

A.1 Attitudes Towards AI

Here, we show when interests between firm and worker are aligned for all $\bar{u}$

Corollary A-1 *The following holds:*

1. ***Win-Win:*** *When $\bar{u} \in \left(\bar{u}^{B*}, \bar{u}^B_{AI}\right)$, AI strictly increases the firm's profit and the agent's utility.*

2. ***Lose-Lose:***

   (a) *Suppose $(1-q)y > c^H - c^L_{AI}$ and $\bar{u}^B > \bar{u}^A_{AI}$: When $\bar{u} \in \left(\bar{u}^{A*}, \bar{u}^A_{AI}\right)$, AI strictly reduces the firm's profit and the agent's utility.*

   (b) *Suppose $(1-q)y > c^H - c^L_{AI}$ and $\bar{u}^B \leq \bar{u}^A_{AI}$: When $\bar{u} \in \left(\bar{u}^{A*}, \bar{u}^B\right)$, AI strictly reduces the firm's profit and the agent's utility.*

For relatively high outside options, interests are fully aligned: the relationship moves from a good job without rent to a good job with rent, and the productivity gain from the reduction in c^H_{AI} dominates the loss from reduced rent extraction. Note that even for $\bar{u} > \bar{u}^B_{AI}$, i.e., to the right of the win–win region, the agent is not worse off, since his utility remains at $\bar{u}$. Thus, whenever AI adoption raises the principal's profits, the agent is at least weakly better off as well.

As laid out in the main part, both parties are worse off when the relationship switches from a good job with rents to a bad job, and the loss in value from lower effort outweighs both the full rent extraction in the bad job and the reduction in low-effort cost c^L_{AI}.[3]

Next, we describe when conflicts of interests emerge.

Corollary A-2 *The following holds:*

1. ***Pro-Worker:*** *Suppose* $(1-q)y > c^H - c^L_{AI}$. *When* $\bar{u} \in \left(\bar{u}^A_{AI}, \bar{u}^{B*}\right)$, *AI reduces the firm's profit but increases the agent's utility.*

2. ***Pro-Firm:***

   (a) *Suppose* $(1-q)y > c^H - c^L_{AI}$: *When* $\bar{u} \in \left(\bar{u}^A, \bar{u}^{A*}\right)$, *AI strictly increases the firm's profit and reduces the agent's utility. Within this range the profit gain is smaller than the utility loss.*

   (b) *Suppose* $(1-q)y \leq c^H - c^L_{AI}$: *When* $\bar{u} \in \left(\bar{u}^A, \bar{u}^B\right)$, *AI strictly increases the firm's profit and reduces the agent's utility. Within this range the profit gain exceeds the utility loss.*

The following figure displays all potential cases stated in the Corollaries.

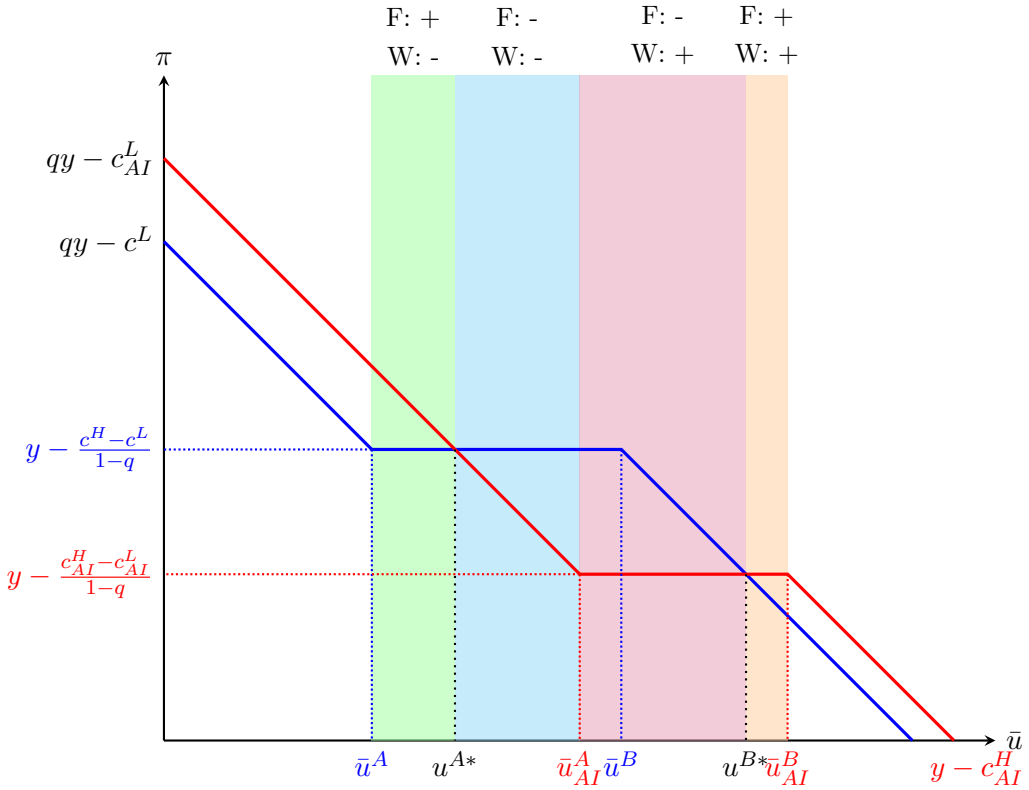


Figure A-3: Firm Profit Function with and without AI. Box colors: light green, cyan, purple, orange.

[3]Note that when $(1-q)y \leq c^H - c^L_{AI}$, this region is empty, because in that case AI increases profits for all values of the outside option.

A.2 Effectiveness of AI: Comparative Statics Results

We now provide a systematic analysis of how the effectiveness of AI—i.e., the differential effects of reductions in c_{AI}^H and c_{AI}^L—shapes equilibrium outcomes.

Our first lemma investigates how reductions in c_{AI}^H versus c_{AI}^L change the sign and magnitude of AI's impact on equilibrium outcomes. Using profits as an example, we analyze how the relative sizes of c_{AI}^H and c_{AI}^L affect $\Delta\pi \equiv \pi^{AI} - \pi^{noAI}$, and delineate the regions in which making AI "better" renders $\Delta\pi$ positive, negative, or unchanged. The same logic extends to the utility and wages.

Lemma A-1 *For $\bar{u} > \bar{u}_{AI}^A$, a smaller c_{AI}^H enhances the impact of AI on profits but worsens its effects on utilities (for utilities no effect for $\bar{u} > \bar{u}_{AI}^B$). For $\bar{u} \leq \bar{u}_{AI}^A$, c_{AI}^H does not affect AI's impact on payoffs.*

For $\bar{u} > \bar{u}_{AI}^B$, a smaller c_{AI}^L has no effect on payoffs. For $\bar{u} \in \left(\bar{u}_{AI}^A, \bar{u}_{AI}^B\right)$, a smaller c_{AI}^L worsens the effects of AI on profits, but enhances its effects on utilities. For $\bar{u} \leq \bar{u}_{AI}^A$, a smaller c_{AI}^L enhances the effects of AI on profits, while there is no effect on utilities.

In general, improvements in high-effort AI are beneficial for profits—but only when the agent's outside option is sufficiently high to keep high effort optimal after AI adoption. By contrast, improvements in low-effort AI (a lower c_{AI}^L) leave profits unchanged if high effort remains optimal post-AI and the agent does not earn a rent. If high effort remains optimal but the agent is granted a rent, profits fall because the agent captures the efficiency gains. If low effort is optimal post-AI, however, profits rise, as production costs fall without the need to pay rents.

Next, we are interested in how changes in c_{AI}^H and c_{AI}^L affect the size of the ranges in which effort, profits, and utilities move up or down. These comparative-statics results clarify not only the direction of AI's impact but also the breadth of situations in which different adoption outcomes arise.

All of these findings are summarized in the following Lemma.

Lemma A-2

- ***Effort.*** *AI reduces effort for outside options $\bar{u} \in \left(\bar{u}^A, \bar{u}_{AI}^A\right)$. The length of this range shrinks when c_{AI}^H falls and expands when c_{AI}^L falls.*
- ***Profits.*** *If $(1-q)y > c^H - c_{AI}^L$, AI reduces profits for $\bar{u} \in \left(\bar{u}^{A*}, \bar{u}^{B*}\right)$. The length of this range shrinks when c_{AI}^H falls and expands when c_{AI}^L falls.*
- ***Utility.*** *AI increases the agent's utility for $\bar{u} \in \left(\bar{u}_{AI}^A, \bar{u}_{AI}^B\right)$. The length of this range expands when c_{AI}^H falls and shrinks when c_{AI}^L falls.*
  - *For $\bar{u}^B > \bar{u}_{AI}^A$, AI decreases the utility for $\bar{u} \in \left(\bar{u}^A, \bar{u}_{AI}^A\right)$. The length of this range shrinks when c_{AI}^H falls and expands when c_{AI}^L falls.*
  - *For $\bar{u}^B \leq \bar{u}_{AI}^A$, AI decreases the utility for $\bar{u} \in \left(\bar{u}^A, \bar{u}^B\right)$. The length of this range is unaffected by c_{AI}^H or c_{AI}^L.*

B Appendix – Omitted Proofs

B.1 Proof to Lemma 2

The value of the cutoffs,

$$\begin{aligned}\bar{u}_{AI}^{B} =& \frac{q c_{AI}^{H} - c_{AI}^{L}}{1-q} \\ \bar{u}_{AI}^{A} =& \frac{c_{AI}^{H} - c_{AI}^{L}(2-q)}{1-q} - (1-q)\, y,\end{aligned}$$

follows immediately, as well as that

$$\begin{aligned}\bar{u}_{AI}^{B} - \bar{u}^{B} &= \frac{q\left(c_{AI}^{H} - c^{H}\right) + c^{L} - c_{AI}^{L}}{1-q} > \frac{\left(c_{AI}^{H} - c^{H}\right) + c^{L} - c_{AI}^{L}}{1-q} > 0 \\ \bar{u}_{AI}^{A} - \bar{u}^{A} &= \frac{c_{AI}^{H} - c_{AI}^{L} - c^{H} + c^{L} + \left(c^{L} - c_{AI}^{L}\right)(1-q)}{1-q} > 0.\end{aligned}$$

B.2 Proof to Lemma 3

Profits with high effort are $\pi^{H} = y - \max\left\{\frac{\left(c^{H}-c^{L}\right)}{1-q},\, \bar{u} + c^{H}\right\}$ before AI, and $\pi_{AI}^{H} = y - \max\left\{\frac{\left(c_{AI}^{H}-c_{AI}^{L}\right)}{1-q},\, \bar{u} + c_{AI}^{H}\right\}$ after AI, and the cutoff $\bar{u}^{B} = \frac{\left(c^{H}-c^{L}\right)}{1-q} - c^{H}$ increases with AI. For the profit difference $\Delta\pi^{H} \equiv \pi_{AI}^{H} - \pi^{H}$, we thus have the following cases:

- $\bar{u} \geq \bar{u}_{AI}^{B}$: $\Delta\pi^{H} = c^{H} - c_{AI}^{H} > 0$
- $\bar{u} \in \left(\bar{u}^{B},\, \bar{u}_{AI}^{B}\right)$: $\Delta\pi^{H} = \bar{u} + c^{H} - \frac{\left(c_{AI}^{H}-c_{AI}^{L}\right)}{1-q}$
  - For $\bar{u} \to \bar{u}_{AI}^{B}$, $\Delta\pi^{H} = c^{H} - c_{AI}^{H} > 0$
  - For $\bar{u} \to \bar{u}^{B}$, $\Delta\pi^{H} = \frac{\left(c^{H}-c^{L}\right)-\left(c_{AI}^{H}-c_{AI}^{L}\right)}{1-q} < 0$
- $\bar{u} \leq \bar{u}^{B}$: $\Delta\pi^{H} = \frac{\left(c^{H}-c^{L}\right)-\left(c_{AI}^{H}-c_{AI}^{L}\right)}{1-q} < 0.$

It follows that profits increase for $\bar{u} > \bar{u}^{B*} \equiv \frac{\left(c_{AI}^{H}-c_{AI}^{L}\right)}{1-q} - c^{H} \in \left(\bar{u}^{B},\, \bar{u}_{AI}^{B}\right)$, and decrease for $\bar{u} \leq \bar{u}^{B*}$.

B.3 Proof to Propositions A-1 and A-2 and Lemma ?? (Complete Characterization of Outcomes With and Without AI)

Here, we compare the situations with and without AI, and derive the consequences for AI on profits, as well as wages and the agent's utility. Note that the total value is maximized if and only if high effort is implemented. This follows from our assumption that high effort is efficient even with AI, an assumption we might eventually relax.

Before doing that, let us first collect the outcomes for all ranges, separately for the cases without and with AI.

Outcomes

Without AI

1. $\bar{u} \geq \bar{u}^B$

   $e = H$

   $w = \bar{u} + c^H$

   $\pi = y - w = y - \bar{u} - c^H$

   $u = w - c^H = \bar{u}$

2. $\bar{u} \in \left[\bar{u}^A, \bar{u}^B\right)$

   $e = H$

   $w = \frac{\left(c^H - c^L\right)}{1-q}$

   $\pi = y - w = y - \frac{\left(c^H - c^L\right)}{1-q} > 0$

   $u = w - c^H = \frac{qc^H - c^L}{1-q} > \bar{u}$

3. $\bar{u} < \bar{u}^A$

   $e = L$

   $w = \bar{u} + c^L$

   $\pi = qy - w = qy - \bar{u} - c^L > y - \frac{c^H - c^L}{1-q} > 0$

   $u = w - c^L = \bar{u}$

With AI

1. $\bar{u} \geq \bar{u}^B_{AI}$

   $e = H$

   $w = \bar{u} + c^H_{AI}$

   $\pi = y - w = y - \bar{u} - c^H_{AI}$

   $u = w - c^H_{AI} = \bar{u}$

2. $\bar{u} \in \left[\bar{u}^A_{AI}, \bar{u}^B_{AI}\right)$

   $e = H$

   $w = \frac{\left(c^H_{AI} - c^L_{AI}\right)}{1-q}$

   $\pi = y - w = y - \frac{\left(c^H_{AI} - c^L_{AI}\right)}{1-q} > 0$

   $u = w - c^H_{AI} = \frac{qc^H_{AI} - c^L_{AI}}{1-q} > \bar{u}$

3. $\bar{u} < \bar{u}_{AI}^{A}$

   $e = L$

   $w = \bar{u} + c_{AI}^{L}$

   $\pi = qy - w = qy - \bar{u} - c_{AI}^{L} > y - \frac{c_{AI}^{H} - c_{AI}^{L}}{1-q} > 0$

   $u = w - c_{AI}^{L} = \bar{u}$

Consequences of AI

We now describes the consequences of AI, as a function of the agent's outside option. Recall the values of the cutoffs,

$$\begin{aligned}
\bar{u}_{AI}^{B} =& \frac{q c_{AI}^{H} - c_{AI}^{L}}{1-q} \\
\bar{u}_{AI}^{A} =& \frac{c_{AI}^{H} - c_{AI}^{L}(2-q)}{1-q} - (1-q)\, y \\
\bar{u}^{B} =& \frac{q c^{H} - c^{L}}{1-q} \\
\bar{u}^{A} =& \frac{c^{H} - c^{L}(2-q)}{1-q} - (1-q)\, y
\end{aligned}$$

It follows that $\bar{u}_{AI}^{B} > \bar{u}^{B}$, $\bar{u}_{AI}^{A} > \bar{u}^{A}$, $\bar{u}_{AI}^{B} > \bar{u}_{AI}^{A}$, and $\bar{u}^{B} > \bar{u}^{A}$. The only ambiguity holds for $\bar{u}^{B}$ vs. $\bar{u}_{AI}^{A}$, and we have to distinguish between the cases $\bar{u}^{B} > \bar{u}_{AI}^{A}$ and $\bar{u}^{B} \leq \bar{u}_{AI}^{A}$. There, note that

$$\begin{aligned}
& \bar{u}^{B} > \bar{u}_{AI}^{A} \\
\Leftrightarrow & (1-q)^{2} y + q c^{H} - c^{L} > c_{AI}^{H} - c_{AI}^{L} - c_{AI}^{L}(1-q). \\
\Leftrightarrow & \left[(1-q)\, y - c^{H} + c_{AI}^{L}\right](1-q) > c_{AI}^{H} - c_{AI}^{L} - \left(c^{H} - c^{L}\right).
\end{aligned}$$

We have assumed that $c_{AI}^{H} - c_{AI}^{L} \in \left(c^{H} - c^{L},\, c^{H}\right]$. Thus, if this condition is satisfied for $c_{AI}^{H} - c_{AI}^{L} = c^{H}$ (i.e., $c_{AI}^{H} = c^{H}$ and $c_{AI}^{L} = 0$), it holds for all AI levels. If this condition is violated for $c_{AI}^{H} - c_{AI}^{L} \to c^{H} - c^{L}$, it holds for no AI

This gives the following three cases:

A)

$$c^{H} \leq (1-q)\, y + c_{AI}^{L} - \frac{c^{L}}{(1-q)},$$

then $\bar{u}^{B} \geq \bar{u}_{AI}^{A}$ for all c^{AI}.

B)

$$(1-q)\, y + c_{AI}^{L} - \frac{c^{L}}{(1-q)} < c^{H} < (1-q)\, y + c_{AI}^{L},$$

then there exists some Δc^{AI}, such that

B.1) $c_{AI}^H - c_{AI}^L - \left(c^H - c^L\right) \leq \Delta c^{AI}$: $\bar{u}^B \geq \bar{u}_{AI}^A$

B.2) $c_{AI}^H - c_{AI}^L - \left(c^H - c^L\right) > \Delta c^{AI}$: $\bar{u}^B < \bar{u}_{AI}^A$

C)

$$c^H \geq (1-q)\, y + c_{AI}^L,$$

then $\bar{u}^B < \bar{u}_{AI}^A$ for all c^{AI}. However, note that this constrains c_{AI}^H, since $c_{AI}^H - c_{AI}^L < (1-q)y$ holds by assumption.

In the following, we analyze these cases separately, defining situations a)-e) to describe the regions before and after the introduction of AI. However, since $\bar{u}_{AI}^B$ is the highest cutoff in each case, we have

a) Situation $\bar{u} > \bar{u}_{AI}^B$: Remain in region 3 (good job, no rent)

- High effort remains optimal
- $\Delta w = c_{AI}^H - c^H < 0$
- $\Delta \pi = c^H - c_{AI}^{AI} > 0$
- $\Delta u = 0$

CASES A) and B.1): $\bar{u}^B > \bar{u}_{AI}^A$**:** $(1-q)\, y > \frac{c_{AI}^H - q c^H + c^L - c_{AI}^L (2-q)}{(1-q)}$ We also have $c^H - c_{AI}^L < (1-q)\, y$

b) Situation $\bar{u} \in \left(\bar{u}^B,\, \bar{u}_{AI}^B\right]$: Move from region 3 to 2 (good job, no rent $\rightarrow$ good job, rent)

- High effort remains optimal
- $\Delta w = \frac{\left(c_{AI}^H - c_{AI}^L\right)}{1-q} - \left(\bar{u} + c^H\right) \gtrless 0$
- $\Delta \pi = c^H - \frac{\left(c_{AI}^H - c_{AI}^L\right)}{1-q} + \bar{u} \gtrless 0$
  $\bar{u} = \bar{u}_{AI}^B$: $\Delta \pi = c^H - c_{AI}^H > 0$
  $\bar{u} = \bar{u}^B$: $\Delta \pi = \frac{c^H - c^L - \left(c_{AI}^H - c_{AI}^L\right)}{1-q} < 0$
- $\Delta u = \frac{q c_{AI}^H - c_{AI}^L}{1-q} - \bar{u} > \frac{q c_{AI}^H - c_{AI}^L}{1-q} - \bar{u}_{AI}^B = 0$

c) Situation $\bar{u} \in \left(\bar{u}_{AI}^A,\, \bar{u}^B\right]$: Remain in region 2 (good job, rent)

- High effort remains optimal
- $\Delta w = \frac{\left(c_{AI}^H - c_{AI}^L\right) - \left(c^H - c^L\right)}{1-q} > 0$
- $\Delta \pi = \frac{\left(c^H - c^L\right)}{1-q} - \frac{\left(c_{AI}^H - c_{AI}^L\right)}{1-q} < 0$
- $\Delta u = \frac{q\left(c_{AI}^H - c^H\right) + c^L - c_{AI}^L}{1-q} \geq \frac{c_{AI}^H - c^H + c^L - c_{AI}^L}{1-q} > 0$

d) Situation $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^A_{AI}\right]$: Shift from region 2 to 1 (good job, rent $\rightarrow$ bad job)

- Effort goes down
- $\Delta w = \bar{u} + c^L_{AI} - \frac{\left(c^H - c^L\right)}{1-q} < 0$, since $(1-q)\, y > \frac{c^H_{AI} - q c^H + c^L - c^L_{AI}(2-q)}{(1-q)}$

$$
\begin{aligned}
& \bar{u} + c^L_{AI} - \frac{\left(c^H - c^L\right)}{1-q} \\
<& \bar{u}^A_{AI} + c^L_{AI} - \frac{\left(c^H - c^L\right)}{1-q} \\
=& \frac{c^H_{AI} - c^L_{AI} - \left(c^H - c^L\right)}{1-q} - (1-q)\, y \\
<& \frac{c^H_{AI} - c^L_{AI} - \left(c^H - c^L\right)}{1-q} - \frac{c^H_{AI} - q c^H + c^L - c^L_{AI}\,(2-q)}{(1-q)} \\
=& - c^H + c^L_{AI} < 0.
\end{aligned}
$$

- $\Delta \pi = \frac{\left(c^H - c^L\right)}{1-q} - c^L_{AI} - (1-q)\, y - \bar{u} \gtrless 0$

  $\bar{u}^A_{AI}$: $\Delta \pi = \frac{c^H - c^L - \left(c^H_{AI} - c^L_{AI}\right)}{1-q} < 0$

  $\bar{u}^A$: $\Delta \pi = c^L - c^L_{AI} > 0$
- $\Delta u = \bar{u} - \frac{q c^H - c^L}{1-q} = \bar{u} - \bar{u}^B < 0$
- $\Delta \pi + \Delta u = c^H - c^L_{AI} - (1-q)\, y < 0$, thus it is not profitable to compensate the agent

e) Situation $\bar{u} \leq \bar{u}^A$: Remain in region 1 (bad job)

- Low effort remains optimal
- $\Delta w = c^L_{AI} - c^L < 0$
- $\Delta \pi = c^L - c^L_{AI} > 0$
- $\Delta u = 0$

Conclusion:

- There are 2 cutoffs, $\bar{u}^{A*} \in \left(\bar{u}^A,\, \bar{u}^A_{AI}\right)$ and $\bar{u}^{B*} \in \left(\bar{u}^B,\, \bar{u}^B_{AI}\right)$, with $\bar{u}^{B*} > \bar{u}^{A*}$, such that:
  - $\bar{u} > \bar{u}^{B*}$: $\Delta \pi > 0$, $\Delta w < 0$
  - $\bar{u}^{A*} < \bar{u} < \bar{u}^{B*}$: $\Delta \pi < 0$
    - $\bar{u}^A_{AI} < \bar{u} < \bar{u}^{B*}$: $\Delta w > 0$
    - $\bar{u}^{A*} < \bar{u} \leq \bar{u}^A_{AI}$: $\Delta w < 0$
  - $\bar{u} < \bar{u}^{A*}$: $\Delta \pi > 0$, $\Delta w < 0$
- Utility
  - remains unaffected for $\bar{u} > \bar{u}^B_{AI}$, even though wage goes down.

- increases for $\bar{u} \in \left(\bar{u}^A_{AI},\, \bar{u}^B_{AI}\right]$, even though w goes down for $\bar{u} \in \left(\bar{u}^{B*},\, \bar{u}^B_{AI}\right]$, it goes up for $\bar{u} \in \left(\bar{u}^A_{AI},\, \bar{u}^{B*}\right)$
- decreases for $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^A_{AI}\right]$, wage goes down
- remains unaffected for $\bar{u} \leq \bar{u}^A$, even though wage goes down.

CASE B2): $\bar{u}^B \leq \bar{u}^A_{AI}$**:** $\bar{u}^A_{AI} \geq \frac{qc^H - c^L}{1-q}$**, and** $(1-q)\,y > c^H - c^L_{AI}$

b) Situation $\bar{u} \in \left(\bar{u}^A_{AI},\, \bar{u}^B_{AI}\right]$: Move from region 3 to 2 (good job, no rent → good job, rent)

- High effort remains optimal
- $\Delta w = \frac{\left(c^H_{AI} - c^L_{AI}\right)}{1-q} - \left(\bar{u} + c^H\right) \gtrless 0$
- $\Delta \pi = \bar{u} + c^H - \frac{\left(c^H_{AI} - c^L_{AI}\right)}{1-q} \gtrless 0$

  $\bar{u} = \bar{u}^B_{AI}$: $\Delta \pi = c^H - c^H_{AI} > 0$

  $\bar{u} = \bar{u}^A_{AI}$: $\Delta \pi = c^H - c^L_{AI} - (1-q)\,y < 0$
- $\Delta u = \frac{qc^H_{AI} - c^L_{AI}}{1-q} - \bar{u} > \frac{qc^H_{AI} - c^L_{AI}}{1-q} - \bar{u}^B_{AI} = 0$

c) Situation $\bar{u} \in \left(\bar{u}^B,\, \bar{u}^A_{AI}\right]$: Move from region 3 to 1 (good job, no rent → bad job)

- Effort goes down
- $\Delta w = c^L_{AI} - c^H < 0$
- $\Delta \pi = c^H - c^L_{AI} - (1-q)\,y < 0$
- $\Delta u = 0$

d) Situation $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^B\right]$: Move from region 2 to 1 (good job, rent → bad job)

- Effort goes down
- $\Delta w = \bar{u} + c^L_{AI} - \frac{\left(c^H - c^L\right)}{1-q} \leq \bar{u}^B + c^L_{AI} - \frac{\left(c^H - c^L\right)}{1-q} = c^L_{AI} - c^H < 0$
- $\Delta \pi = \frac{\left(c^H - c^L\right)}{1-q} - c^L_{AI} - (1-q)\,y - \bar{u} \gtrless 0$

  $\bar{u}^B$: $\Delta \pi = c^H - c^L_{AI} - (1-q)\,y < 0$

  $\bar{u}^A$: $\Delta \pi = c^L - c^L_{AI} > 0$
- $\Delta u = \bar{u} - \frac{qc^H - c^L}{1-q} = \bar{u} - \bar{u}^B < 0$
- $\Delta \pi + \Delta u = c^H - c^L_{AI} - (1-q)\,y < 0$, thus it is not profitable to compensate the agent

e) Situation $\bar{u} \leq \bar{u}^A$: Remain in region 1 (bad job)

- Low effort remains optimal
- $\Delta w = c^L_{AI} - c^L < 0$
- $\Delta \pi = c^L - c^L_{AI} > 0$
- $\Delta u = 0$

Conclusion:

- There are 2 cutoffs, $\bar{u}^{A*} \in \left(\bar{u}^{A},\, \bar{u}^{B}\right)$ and $\bar{u}^{B*} \in \left(\bar{u}^{A}_{AI},\, \bar{u}^{B}_{AI}\right)$, with $\bar{u}^{B*} > \bar{u}^{A*}$, such that:
  - $\bar{u} > \bar{u}^{B*}$: $\Delta\pi > 0$, $\Delta w < 0$
  - $\bar{u}^{A*} < \bar{u} < \bar{u}^{B*}$: $\Delta\pi < 0$
    - $\bar{u}^{A}_{AI} < \bar{u} < \bar{u}^{B*}$: $\Delta w > 0$
    - $\bar{u}^{A*} < \bar{u} \leq \bar{u}^{A}_{AI}$: $\Delta w < 0$
  - $\bar{u} < \bar{u}^{A*}$: $\Delta\pi > 0$, $\Delta w < 0$
- Utility
  - remains unaffected for $\bar{u} > \bar{u}^{B}_{AI}$, even though wage goes down.
  - increases for $\bar{u} \in \left(\bar{u}^{A}_{AI},\, \bar{u}^{B}_{AI}\right]$, even though w goes down for $\bar{u} \in \left(\bar{u}^{B*},\, \bar{u}^{B}_{AI}\right]$, it goes up for $\bar{u} \in \left(\bar{u}^{A}_{AI},\, \bar{u}^{B*}\right)$
  - remains unaffected for $\bar{u} \in \left(\bar{u}^{B},\, \bar{u}^{A}_{AI}\right]$, even though wage goes down.
  - decreases for $\bar{u} \in \left(\bar{u}^{A},\, \bar{u}^{B}\right]$, wage goes down
  - remains unaffected for $\bar{u} \leq \bar{u}^{A}$, even though wage goes down.

CASE C): $\bar{u}^{B} \leq \bar{u}^{A}_{AI}$**:** $\bar{u}^{A}_{AI} \geq \frac{qc^{H}-c^{L}}{1-q}$**, and** $(1-q)\,y \leq c^{H} - c^{L}_{AI}$

b) Situation $\bar{u} \in \left(\bar{u}^{A}_{AI},\, \bar{u}^{B}_{AI}\right]$: Move from region 3 to 2 (good job, no rent → good job, rent)

- High effort remains optimal
- $\Delta w = \frac{\left(c^{H}_{AI}-c^{L}_{AI}\right)}{1-q} - \left(\bar{u}+c^{H}\right) < 0$
- $\Delta\pi = \bar{u} + c^{H} - \frac{\left(c^{H}_{AI}-c^{L}_{AI}\right)}{1-q} > 0$, since

$$\begin{aligned} &\bar{u} + c^{H} - \frac{\left(c^{H}_{AI}-c^{L}_{AI}\right)}{1-q} \\ >&\bar{u}^{A}_{AI} + c^{H} - \frac{\left(c^{H}_{AI}-c^{L}_{AI}\right)}{1-q} + c^{H} - \frac{c^{AI}}{1-q} \\ =&c^{H} - c^{L}_{AI} - (1-q)\,y \geq 0 \end{aligned}$$

- $\Delta u = \frac{qc^{AI}}{1-q} - \bar{u} > \frac{qc^{AI}}{1-q} - \bar{u}^{B}_{AI} = 0$

c) Situation $\bar{u} \in \left(\bar{u}^{B},\, \bar{u}^{A}_{AI}\right]$: Move from region 3 to 1 (good job, no rent → bad job)

- Effort goes down
- $\Delta w = c^{L}_{AI} - c^{H} < 0$
- $\Delta\pi = c^{H} - c^{L}_{AI} - (1-q)\,y > 0$
- $\Delta u = 0$

d) Situation $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^B\right]$: Move from region 2 to 1 (good job, rent $\rightarrow$ bad job)

- Effort goes down
- $\Delta w = \bar{u} + c_{AI}^L - \frac{\left(c^H - c^L\right)}{1-q} \leq \bar{u}^B + c_{AI}^L - \frac{\left(c^H - c^L\right)}{1-q} = c_{AI}^L - c^H < 0$
- $\Delta \pi = \frac{\left(c^H - c^L\right)}{1-q} - c_{AI}^L - (1-q)\, y - \bar{u} > 0$, since

$$\begin{aligned} & \frac{\left(c^H - c^L\right)}{1-q} - c_{AI}^L - (1-q)\, y - \bar{u} \\ > & \frac{\left(c^H - c^L\right)}{1-q} - c_{AI}^L - (1-q)\, y - \bar{u}^B \\ = & c^H - c_{AI}^L - (1-q)\, y > 0 \end{aligned}$$

- $\Delta u = \bar{u} - \frac{q c^H - c^L}{1-q} = \bar{u} - \bar{u}^B < 0$
- $\Delta \pi + \Delta u = c^H - c_{AI}^L - (1-q)\, y > 0$, thus possible to compensate the agent.

e) Situation $\bar{u} \leq \bar{u}^A$: Remain in region 1 (bad job)

- Low effort remains optimal
- $\Delta w = c_{AI}^L - c^L < 0$
- $\Delta \pi = c^L - c_{AI}^L > 0$
- $\Delta u = 0$

Conclusion:

- Profits go up, wage goes down for all $\bar{u}$
- Utility
  - remains unaffected for $\bar{u} > \bar{u}_{AI}^B$, even though wage goes down.
  - increases for $\bar{u} \in \left(\bar{u}_{AI}^A,\, \bar{u}_{AI}^B\right]$, even though w goes down
  - remains unaffected for $\bar{u} \in \left(\bar{u}^B,\, \bar{u}_{AI}^A\right]$, even though wage goes down.
  - decreases for $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^B\right]$, wage goes down
  - remains unaffected for $\bar{u} \leq \bar{u}^A$, even though wage goes down.

Q.E.D.

B.4 Proof of Corollary A-2.

Suppose $(1-q)y > c^H - c_{AI}^L$. In Proposition A-1, we have shown that AI increases profits unless the agent's outside option is in the "middle range" $\bar{u} \in \left(\bar{u}^{A*},\, \bar{u}^{B*}\right)$, with $\bar{u}^{A*} = \frac{\left(c^H - c^L\right)}{1-q} - c_{AI}^L - (1-q)\, y$ and $\bar{u}^{B*} = \frac{\left(c_{AI}^H - c_{AI}^L\right)}{1-q} - c^H$. In Lemma A-2, we have demonstrated that AI decreases the agent's

utility for $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^A_{AI}\right]$ (if $\bar{u}^B > \bar{u}^A_{AI}$) or for $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^B\right]$ (if $\bar{u}^B \leq \bar{u}^A_{AI}$). Since $\bar{u}^{B*} > \bar{u}^A_{AI}$ and $\bar{u}^{B*} > \bar{u}^B$, the effect of AI on the agent's utility is non-negative at and beyond the upper bound of the "middle range." However, since $\bar{u}^A < \bar{u}^{A*}$, the effect of AI on the agent's utility is negative at and in the left neighborhood of the lower bound of the "middle range." Thus, in the range $\left(\bar{u}^A, \bar{u}^{A*}\right)$, AI increases profits but decreases utility, with $\bar{u}^{A*} - \bar{u}^A = c^L - c^L_{AI}$. In this range, $\Delta\pi + \Delta u = c^H - c^L_{AI} - (1-q)\, y < 0$.

The results for $c^H - c^L_{AI} - (1-q)\, y \geq 0$ follow immediately from Proposition A-1 and Lemma A-2

Q.E.D.

B.5 Proof of Lemma A-1.

Here, we compute the partial derivatives of Δw, $\Delta\pi$, Δu with respect to c^H_{AI} and c^L_{AI} for all potential cases, as derived in the proof to Proposition A-1 and Lemma A-2.

a) Situation $\bar{u} > \bar{u}^B_{AI}$: Remain in region 3

- $\frac{\partial \Delta\pi}{\partial c^H_{AI}} = -1 < 0$
- $\frac{\partial \Delta\pi}{\partial c^L_{AI}} = 0$
- $\frac{\partial \Delta u}{\partial c^H_{AI}} = \frac{\partial \Delta u}{\partial c^L_{AI}} = 0$

CASES A) and B.1): $\bar{u}^B > \bar{u}^A_{AI}$

b) Situation $\bar{u} \in \left(\bar{u}^B,\, \bar{u}^B_{AI}\right]$: Move from region 3 to 2

- $\frac{\partial \Delta\pi}{\partial c^H_{AI}} = -\frac{1}{1-q} < 0$
- $\frac{\partial \Delta\pi}{\partial c^L_{AI}} = \frac{1}{1-q} > 0$
- $\frac{\partial \Delta u}{\partial c^H_{AI}} = \frac{q}{1-q} > 0$
- $\frac{\partial \Delta u}{\partial c^L_{AI}} = \frac{-1}{1-q} < 0$

c) Situation $\bar{u} \in \left(\bar{u}^A_{AI},\, \bar{u}^B\right]$: Remain in region 2

- $\frac{\partial \Delta\pi}{\partial c^H_{AI}} = -\frac{1}{1-q} < 0$
- $\frac{\partial \Delta\pi}{\partial c^L_{AI}} = \frac{1}{1-q} > 0$
- $\frac{\partial \Delta u}{\partial c^H_{AI}} = \frac{q}{1-q} > 0$
- $\frac{\partial \Delta u}{\partial c^L_{AI}} = \frac{-1}{1-q} < 0$

d) Situation $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^A_{AI}\right]$: Move from region 2 to 1

- $\frac{\partial \Delta\pi}{\partial c^H_{AI}} = 0$
- $\frac{\partial \Delta\pi}{\partial c^L_{AI}} = -1 < 0$

- $\frac{\partial \Delta u}{\partial c_{AI}^H} = \frac{\partial \Delta u}{\partial c_{AI}^L} = 0$

e) Situation $\bar{u} \leq \bar{u}^A$: Remain in region 1

- $\frac{\partial \Delta \pi}{\partial c_{AI}^H} = 0$
- $\frac{\partial \Delta \pi}{\partial c_{AI}^L} = -1 < 0$
- $\frac{\partial \Delta u}{\partial c_{AI}^H} = \frac{\partial \Delta u}{\partial c_{AI}^L} = 0$

CASE B2) and Case C: $\bar{u}^B \leq \bar{u}_{AI}^A$

b) Situation $\bar{u} \in \left(\bar{u}_{AI}^A,\, \bar{u}_{AI}^B\right]$: Move from region 3 to 2

- $\frac{\partial \Delta \pi}{\partial c_{AI}^H} = -\frac{1}{1-q} < 0$
- $\frac{\partial \Delta \pi}{\partial c_{AI}^L} = \frac{1}{1-q} > 0$
- $\frac{\partial \Delta u}{\partial c_{AI}^H} = \frac{q}{1-q} > 0$
- $\frac{\partial \Delta u}{\partial c_{AI}^L} = \frac{-1}{1-q} < 0$

c) Situation $\bar{u} \in \left(\bar{u}^B,\, \bar{u}_{AI}^A\right]$: Move from region 3 to 1

- $\frac{\partial \Delta \pi}{\partial c_{AI}^H} = 0$
- $\frac{\partial \Delta \pi}{\partial c_{AI}^L} = -1 < 0$
- $\frac{\partial \Delta u}{\partial c_{AI}^H} = \frac{\partial \Delta u}{\partial c_{AI}^L} = 0$

d) Situation $\bar{u} \in \left(\bar{u}^A,\, \bar{u}^B\right]$: Move from region 2 to 1

- $\frac{\partial \Delta \pi}{\partial c_{AI}^H} = 0$
- $\frac{\partial \Delta \pi}{\partial c_{AI}^L} = -1 < 0$
- $\frac{\partial \Delta u}{\partial c_{AI}^H} = \frac{\partial \Delta u}{\partial c_{AI}^L} = 0$

e) Situation $\bar{u} \leq \bar{u}^A$: Remain in region 1

- $\frac{\partial \Delta \pi}{\partial c_{AI}^H} = 0$
- $\frac{\partial \Delta \pi}{\partial c_{AI}^L} = -1 < 0$
- $\frac{\partial \Delta u}{\partial c_{AI}^H} = \frac{\partial \Delta u}{\partial c_{AI}^L} = 0$

To conclude, these results deliver the following links

- Smaller c_{AI}^L and profits: no effect if $\bar{u} > \bar{u}_{AI}^B$, negative effect if $\bar{u} \in \left(\bar{u}_{AI}^A, \bar{u}_{AI}^B\right)$, positive effect if $\bar{u} \leq \bar{u}_{AI}^A$
- Smaller c_{AI}^L and utility: no effect if $\bar{u} > \bar{u}_{AI}^B$, positive effect if $\bar{u} \in \left(\bar{u}_{AI}^A, \bar{u}_{AI}^B\right)$, no effect if $\bar{u} \leq \bar{u}_{AI}^A$

- Smaller c_{AI}^H and profits: positive if $\bar{u} > \bar{u}_{AI}^A$, no effect if $\bar{u} \leq \bar{u}_{AI}^A$
- Smaller c_{AI}^H on utility: No effect if $\bar{u} > \bar{u}_{AI}^B$, negative effect if $\bar{u} \in \left(\bar{u}_{AI}^A, \bar{u}_{AI}^B\right)$, no effect if $\bar{u} \leq \bar{u}_{AI}^A$.

Q.E.D.

B.6 Proof of Lemma A-2.

For effort, the statement follows form

$$\begin{aligned}\frac{\partial \bar{u}_{AI}^A}{\partial c_{AI}^H} =& \frac{1}{1-q} > 0\\ \frac{\partial \bar{u}_{AI}^A}{\partial c_{AI}^L} =& \frac{-(2-q)}{1-q} < 0.\end{aligned}$$

For profits, note that

$$\bar{u}^{B*} - \bar{u}^{A*} = \frac{c_{AI}^H - q c_{AI}^L}{1-q} - c^H - \frac{\left(c^H - c^L\right)}{1-q} + (1-q)\, y,$$

thus

$$\begin{aligned}\frac{\partial\left(\bar{u}^{B*} - \bar{u}^{A*}\right)}{\partial c_{AI}^H} =& \frac{1}{1-q} > 0\\ \frac{\partial\left(\bar{u}^{B*} - \bar{u}^{A*}\right)}{\partial c_{AI}^L} =& \frac{-q}{1-q} < 0.\end{aligned}$$

For the agent's utility, note that

$$\bar{u}_{AI}^B - \bar{u}_{AI}^A = -c_{AI}^H + c_{AI}^L + (1-q)\, y,$$

thus

$$\begin{aligned}\frac{\partial\left(\bar{u}_{AI}^B - \bar{u}_{AI}^A\right)}{\partial c_{AI}^H} =& -1 < 0\\ \frac{\partial\left(\bar{u}_{AI}^B - \bar{u}_{AI}^A\right)}{\partial c_{AI}^L} =& 1 > 0.\end{aligned}$$

Moreover,

$$\frac{\partial\left(\bar{u}_{AI}^{A}-\bar{u}^{A}\right)}{\partial c_{AI}^{H}}=\frac{1}{1-q}>0$$
$$\frac{\partial\left(\bar{u}_{AI}^{A}-\bar{u}^{A}\right)}{\partial c_{AI}^{L}}=\frac{-\left(2-q\right)}{1-q}<0.$$

Q.E.D.

C Appendix – Detailed Derivations of the General Equilibrium

C.1 Value Functions and Equilibrium

Given the matching function, $\mu(U,V)$, the job finding rate for workers as $\frac{\mu(U,V)}{U}=\mu(1,V/U)\equiv f(\theta)$, where $\theta\equiv\frac{V}{U}$ is the labor market tightness. Similarly, we denote the matching rate for a vacancy as $\frac{\mu(U,V)}{V}=f(\theta)/\theta\equiv s(\theta)$. A higher θ means more vacancies per unemployed worker, which leads to a higher job finding rate $f(\theta)$ and a lower vacancy filling rate $s(\theta)$. We consider a stationary equilibrium where all endogenous variables are constant over time. Let M denote the number of matches between workers and firms. Therefore, the number of unemployed workers is $U=L-M$, and the number of vacancies is $V=N-M$. In the stationary equilibrium, the number of matches satisfies

$$\lambda M=\mu(U,V). \tag{C-1}$$

Each match produces a flow profit of π for the firm, which will be determined by the equilibrium contract and output. Denoting the discount rate as r, values of a vacancy and a filled job as Π^V and Π^J, we can write the value functions (Hamilton-Jacobi-Bellman equations) of firms as follows:

$$r\Pi^J=\pi+\lambda(\Pi^V-\Pi^J),$$
$$r\Pi^V=-\kappa+s(\theta)(\Pi^J-\Pi^V).$$

An unemployed worker receives a flow utility of b. For example, b can represent unemployment benefits or the value of leisure. We do not exclude the possibility that b is negative, representing the disutility of being unemployed. However, similar to the case with an exogenous outside option, we assume that b is not too low such that firms can make positive profits. When matched with a firm, the worker receives a take-it-or-leave-it offer, which takes the form of $w(Y)$, where Y is the output produced by the match. We assume that the firm commits to the same contract until the match is separated and do not allow for renegotiation or dynamic contracts. The flow utility of an employed worker exerting effort e is therefore $w(Y|e)-c(e)$. Denoting the optimal effort choice as

e^*, the values of being unemployed (V^U) and employed V^E can be written as follows:

$$rV^U = b + f(\theta)(V^E - V^U),$$
$$rV^E = \mathbb{E}[w(Y|e^*)] - c(e^*) + \lambda(V^U - V^E).$$

The worker's participation constraint requires that $V^E \geq V^U$. Combining with the second equation above, it is equivalent to $\mathbb{E}[w(Y|e^*)] - c(e^*) \geq rV^U$. Compared to the model with an exogenous outside option, the worker's outside option $\bar{u}$ is now endogenized as the flow value of unemployment rV^U. This is the only difference from the baseline contracting problem described in Section 2. The incentive compatiblity constraint remains unchanged.

A general equilibrium of the model is defined as follows:

Definition 1 (General Equilibrium) *A general equilibrium consists of the number of matches M, vacancies V, unemployed workers U, firms N, contracts $w(Y)$, effort level e^*, and values Π^J, Π^V, V^E, V^U such that*

1. *The contract $w(Y)$ and effort level e^* solve the contracting problem described in section 2 with the endogenous outside option rV^U.*
2. *The values Π^J, Π^V, V^E, V^U satisfy the HJB equations above.*
3. *The number of matches satisfies $M = L - U = N - V$.*
4. *The labor market clearing condition* (C-1) *holds.*
5. *The free entry condition holds: $\Pi^V = 0$.*

C.2 Equilibrium without AI

We are now ready to characterize the equilibrium without AI in the following proposition.

Proposition C-3 (Equilibrium under random matching) *There exist values $b^A > \underline{b}$ such that:*

1. *When $b \leq \underline{b}$, firms implements low effort $e^* = L$. Workers receive no rent and are indifferent between employment and unemployment.*
2. *When $b \geq b^A$, firms implement high effort $e^* = H$.*
3. *When $b \in (\underline{b}, b^A)$, there are multiple equilibria:*
   (a) ***bad jobs only**: no firm implements effort, and workers earn no rent.*
   (b) ***good jobs only**: all firms implement high effort, and workers earn rents.*

(c) ***mixed**: a fraction of firms implement high effort, and the rest impelment low effort. Firms are indifferent between implementing high and low efforts. Workers earn rents from the high-effort firms and are indifferent between employment and unemployment when matched to low-effort firms.*

The equilibria can be ranked by firm profits/values and total employment M as follows:

$$\textit{Bad jobs only} > \textit{Good jobs only} = \textit{Mixed}.$$

They can be ranked by the value of unemployment and employment as follows:

$$\textit{Good jobs only} > \textit{Mixed} > \textit{Bad jobs only}.$$

Proof of Proposition C-3.

To implement high effort $e = H$, the firm sets $w(0) = 0$ and chooses

$$w(y) = \max\left\{\frac{c^H - c^L}{1-q},\, rV^U + c^H\right\}.$$

To implement low effort $e = L$, the firm sets $w(Y) = rV^U + c^L$, regardless of output Y.

We consider two cases, depending on which term is larger in the expression of $w(y)$ when implementing high effort.

1. $\frac{c^H - c^L}{1-q} \leq rV^U + c^H$. In this case, the worker's flow utility of employment equals $w(y) - c^H = rV^U$. Subsituting into the HJBs of V^U, V^E, we obtain $V^U = V^E = b/r$. We define the cutoff

$$b^B \equiv \frac{qc^H - c^L}{1-q}. \tag{C-1}$$

   The initial inequality condition is equivalent to $b \geq b^B$.

   Firm's flow profits under high and low efforts become

$$\pi(e = H) = y - w(y) = y - rV^U - c^H,$$
$$\pi(e = L) = qy - w(0) = qy - rV^U - c^L$$

   For firms to adopt incentive contracts, we need $\pi(e = H) \geq \pi(e = L)$, i.e.,

$$(1-q)y \geq c^H - c^L,$$

   which always holds in our setup.

2. $\frac{c^H - c^L}{1-q} > rV^U + c^H$. This further implies that, when implementing efforts, workers' flow utility is

$$\mathbb{E}[w(Y|e = H)] - c(e = H) = \frac{c^H - c^L}{1-q} - c^H = b^B.$$

The initial inequality thus implies that $b^B > rV^U$. Combined with HJBs of V^U, V^E, we have $V^E > V^U$ and $b < b^B$.

Firm profits become

$$\pi(e = H) = y - w(y) = y - \frac{c^H - c^L}{1 - q},$$
$$\pi(e = L) = qy - w(0) = qy - rV^U - c^L.$$

We have three possible cases:

(a) All firms adopt incentive contracts. In this case,

$$V^U = \frac{(r + \lambda)b + f(\theta)b^B}{r(r + \lambda + f(\theta))}.$$

The condition that no single firm wants to switch to no incentive contracts is

$$y - \frac{c^H - c^L}{1 - q} > qy - rV^U - c^L.$$

We define the threshold

$$b^A \equiv \frac{c^H - (2 - q)c^L}{1 - q} - (1 - q)y, \tag{C-2}$$

and the above condition is equivalent to $rV^U > b^A$. Substituting the solution of V^U, we have a new threshold for b:

$$b > \underline{b} \equiv \frac{r + \lambda + f(\theta)}{r + \lambda} b^A - \frac{f(\theta)}{r + \lambda} b^B. \tag{C-3}$$

Since $b^A < b^B$, it is straightforward that $\underline{b} < b^A$.

(b) No firm implements incentive contracts. In this case, $w(e = L) - c^L = rV^U$, which further implies that $V^E = V^U = b/r$. For firms not to implement high efforts, we need $\pi(e = H) < \pi(e = L)$, i.e.,

$$y - \frac{c^H - c^L}{1 - q} < qy - b - c^L \Rightarrow b < b^A.$$

(c) Firms are indifferent between implementing efforts and not. We solve for the fraction, γ, of firms that implement efforts. Due to random matching, workers matched to high-effort firms with probability γ and earn flow utility of b^B. With probability $1 - \gamma$, they are matched to low-effort firms and earn flow utility of rV^U. The expected flow utility of employment is therefore

$$\gamma b^B + (1 - \gamma) rV^U.$$

For firms to be indifferent between implementing efforts and not, we need $\pi(e = H) = \pi(e = L)$:

$$y - \frac{c^H - c^L}{1 - q} = qy - rV^U - c^L \Rightarrow b^A = rV^U.$$

Substituting back into the equation for the flow utility of employment and the solution of V^U, we obtain

$$\gamma = \frac{(r + \lambda)(b^A - b)}{f(\theta)\,(b^B - b^A)}. \tag{C-4}$$

One can verify that $\gamma \in (0, 1)$ when $b \in (\underline{b}, b^A)$.

We now compare the three equilibria when $b \in (\underline{b}, b^A)$. The firms' flow profits are

$$\pi_{\text{good-job}} = \pi_{\text{mixed}} = y - \frac{c^H - c^L}{1 - q}, \quad \pi_{\text{bad-job}} = qy - b - c^L.$$

When $b < b^A$, we have $\pi_{\text{bad-job}} > \pi_{\text{good-job}} = \pi_{\text{mixed}}$.

The equilibrium market tightness is determined by the zero-profit condition for posting a vacancy, $\Pi^V = 0 \Leftrightarrow s(\theta)\pi = (r + \lambda)\kappa$. Since $s(\theta)$ is decreasing in θ, we must have

$$\theta_{\text{bad-job}} > \theta_{\text{good-job}} = \theta_{\text{mixed}}.$$

Using the steady-state condition, and the formula for the job finding rate, we have

$$\lambda M = \mu(U, V) = f(\theta)U = f(\theta)(L - M).$$

Since $f(\theta)$ is increasing in θ, we must have

$$M_{\text{bad-job}} > M_{\text{good-job}} = M_{\text{mixed}}.$$

Next, we consider workers' utilities. For unemployed workers, we have

$$\begin{aligned} V^U_{\text{good-job}} &= \frac{(r + \lambda)b + f(\theta^*)b^B}{r(r + \lambda + f)}, \\ V^U_{\text{bad-job}} &= \frac{b}{r}, \\ V^U_{\text{mixed}} &= \frac{b^A}{r}. \end{aligned}$$

For employed workers, we have

$$V^E_{\text{good-job}} = \frac{\lambda b + (r + f(\theta^*))b^B}{r(r + \lambda + f(\theta^*))},$$
$$V^E_{\text{bad-job}} = \frac{b}{r},$$
$$V^E_{\text{mixed}} = \frac{(r + f(\theta^*))b^A - rb}{rf(\theta^*)}.$$

In the expressions of $V^U_{\text{good-job}}$, $V^E_{\text{good-job}}$ and V^E_{mixed}, the market tightness θ^* refers to that corresponding to the profit level $\pi = H - \frac{c^H - c^L}{1-q}$. One can verify that

$$V^U_{\text{good-job}} > \frac{(r + \lambda)b^A_{u0} + f(\theta^*)b^B}{r(r + \lambda + f(\theta^*))} = \frac{b^A}{r} = V^U_{\text{mixed}}$$
$$V^U_{\text{mixed}} = \frac{b^A}{r} > \frac{b}{r} = V^U_{\text{bad-job}},$$
$$V^E_{\text{good-job}} > \frac{\lambda b^A_{u0} + (r + f(\theta^*))b^B}{r(r + \lambda + f(\theta^*))} = \frac{(r + f(\theta^*))b^A - rb^A_{u0}}{rf(\theta^*)} > V^E_{\text{mixed}},$$
$$V^E_{\text{mixed}} = \frac{(r + f(\theta^*))b^A - rb}{rf(\theta^*)} > \frac{b}{r} = V^E_{\text{bad-job}},$$

where we have used the expression of b^A_{u0} in equation (C-3) to derive the first equality in the first and third lines. ■

The equilibrium structure is similar to the one with an exogenous outside option. In general, firms are more likely to implement high efforts when the unemployment benefit, b, thus the endogenous outside option V^U, is high. A new feature is the existence of multiple equilibria when b is in an intermediate range. The outside option, V^U, depends on the prevailing contracts offered by firms in the market. When more firms offer incentive contracts and implement high efforts, workers receive higher wages and earn rents, which raises V^U. This in turn makes it easier for firms to implement high efforts. Therefore, there exists a self-fulfilling good-jobs-only equilibrium where all firms implement high efforts, and workers earn rents. Conversely, when no firm implements effort, workers receive low wages and earn no rent, leading to a self-fulfilling bad-jobs-only equilibrium. Finally, there also exists a mixed equilibrium where a fraction of firms implement high efforts, and the rest implement low efforts. Firms are indifferent between implementing high and low efforts. Workers earn rents from the high-effort firms and are indifferent between employment and unemployment when matched to low-effort firms.

C.3 The Impact of AI

When firms adopt AI, low-effort costs are reduced more than high-effort costs, i.e., $c^L - c^L_{AI} > c^H - c^H_{AI} \geq 0$. We can prove the impact of AI on employment as follows:

Proof of Proposition 5. In a good job equilibrium with worker rents, all firms induce high efforts and obtain profits $\pi = y - \frac{c^H - c^L}{1-q}$ before AI adoption and $\pi = y - \frac{c^H_{AI} - c^L_{AI}}{1-q}$ after AI adoption.

Since $c^L - c^L_{AI} > c^H - c^H_{AI}$, firm profits decline after AI adoption. In the steady-state equilibrium, the value of the vacancy is zero, and we have

$$\Pi^V = 0 \Rightarrow s(\theta)\pi = (r+\lambda)\kappa.$$

Since $s(\theta)$ decreases in θ, a decline in π implies a drop in θ. Moreover, we have

$$\lambda M = \mu(U,V) = f(\theta)(L-M).$$

A drop in θ means a smaller $f(\theta)$, thus lower M. The number of matches, or equivalently the employment level (good jobs), drops after AI adoption.

In a bad job equilibrium, firm profits are

$$\pi = qy - b - c^L.$$

Since AI reduces c^L, firm profits increase. Therefore, the impacts of AI on labor market tightness and equilibrium employment are exactly the opposite of those in the good job equilibrium with worker rents. ■

The follow lemma describes how the unemployment benefit cutoffs for different equilibrium regimes change after AI adoption.

Lemma C-3 *The unemployment benefit cutoffs for different equilibrium regimes, $\underline{b}, b^A, b^B$, all increase after AI adoption. The cutoff $\underline{b}$, under which no good job equilibrium can be sustained, increases more when we allow the labor market tightness, θ^*, to adjust in equilibrium.*

Proof of Lemma C-3. Applying the formula for b^A, we have

$$b^A_{AI} - b^A = \frac{c^H_{AI} - c^L_{AI} - c^H + c^L + \left(c^L - c^L_{AI}\right)(1-q)}{1-q} > 0.$$

To analyze the change in $\underline{b}$, we first consider the change in the cutoff b^B, above which high effort is optimal and workers receive no rent in equilibrium. Applying the formula for b^B in equation C-1, we have

$$b^B_{AI} - b^B = \frac{q\left(c^H_{AI} - c^H\right) - \left(c^L_{AI} - c^L\right)}{1-q} > \frac{\left(c^H_{AI} - c^H\right) - \left(c^L_{AI} - c^L\right)}{1-q} > 0.$$

One can further show that

$$b^A_{AI} - b^A - (b^B_{AI} - b^B) = c^H_{AI} - c^L_{AI} - c^H + c^L > 0. \tag{C-1}$$

Finally, applying the formula for $\underline{b}$ in equation C-3, we have

$$\underline{b}_{AI} - \underline{b} = b^A_{AI} - b^A + \frac{f(\theta^*_{AI})}{r+\lambda}(b^A_{AI} - b^B_{AI}) - \frac{f(\theta^*)}{r+\lambda}(b^A - b^B),$$

where θ^* and θ^*_{AI} are the equilibrium labor market tightness correponding to the profit levels under incentive contracts with positive worker rents, i.e., $\pi = y - \frac{c^H - c^L}{1-q}$ and $\pi_{AI} = y - \frac{c^H_{AI} - c^L_{AI}}{1-q}$. Since $\pi_{AI} < \pi$, we have fewer vacancies created and $\theta^*_{AI} < \theta^*$, and thus $f(\theta^*_{AI}) < f(\theta^*)$. We know from equation (C-1) that $b^A - b^B < b^A_{AI} - b^B_{AI} < 0$. Therefore, we must have $\underline{b}_{AI} - \underline{b} > 0$.

The above equation also implies a decomposition of the change in $\underline{b}$ into two parts: the change when the job finding rate is fixed at the original $f(\theta^*)$, and the change due to the general equilibrium feedback through the job finding rate. To see this, we rewrite the change as

$$\underline{b}_{AI} - \underline{b} = \underbrace{b^A_{AI} - b^A + \frac{f(\theta^*)}{r+\lambda}\left[b^A_{AI} - b^A - (b^B_{AI} - b^B)\right]}_{\text{change when } \theta^* \text{ is fixed}} + \underbrace{\frac{f(\theta^*_{AI}) - f(\theta^*)}{r+\lambda}\left(b^A_{AI} - b^B_{AI}\right)}_{\text{GE feedback}}.$$

The first component keeps the job finding rate fixed at the original level $f(\theta^*)$, and it is straightforward to see that it is positive. The second component captures the general equilibrium feedback through the job finding rate. Since $f(\theta^*_{AI}) < f(\theta^*)$ and $b^A_{AI} - b^B_{AI} < 0$, the general equilibrium feedback further increases $\underline{b}$. ■