

When Solutions Are Abundant but Problems Scarce: AI in the Market for Knowledge

Felix Feng* Brett Green[†] Curtis Taylor[‡] Mark Westerfield[§]

September 23, 2026

Abstract

We study how artificial intelligence affects welfare in general equilibrium when production requires both finding and solving problems. Unsolved problems are unexploited opportunities or unmet needs that agents find in a congested sector. Problems are solved in layered hierarchies in which agents can ask for help. Firms may adopt a replicable AI agent in either activity. AI deployed in solving problems splits the human knowledge hierarchy: greater AI capability affects the layers above AI, whereas lower cost affects the layers below it. Either form of progress can lower welfare by reallocating solvers to the congested finding sector. A tax on finding corrects this distortion; when this is not possible, the optimal tax on AI can be positive or negative, depending on its effect on the congestion externality.

JEL Classification: O33, L23, D62, J24, D50.

Key Words: artificial intelligence, knowledge hierarchies, problem finding, training, congestion

*University of Georgia, ffeng@uga.edu.

[†]Washington University in St. Louis, b.green@wustl.edu.

[‡]Duke University, crtaylor@duke.edu.

[§]University of Washington, mwesterf@uw.edu.

1 Introduction

AI can make solutions abundant, yet turn unsolved problems into the scarce resource. We study artificial intelligence in a knowledge economy where firms must not only apply expertise but first find opportunities to apply it: client engagements, cases, projects, deals, and other valuable work. Our central departure from most theories of AI is simple. Problems must be found before they can be solved. If AI expands the capacity to solve problems faster than the supply of worthwhile problems grows, the bottleneck moves from solving problems to finding them.

This shift in scarcity changes the welfare consequences of automation. By making expertise cheap and replicable, AI expands solving capacity and raises the value of finding worthwhile applications. Displaced solvers can respond by seeking clients, starting businesses, originating projects, or working independently. Those responses are individually sensible, but several finders may pursue the same scarce opportunity, generating a congestion externality. Because their efforts partly duplicate one another, finders are paid their average product even though an additional finder contributes only the marginal product of search. AI can therefore push too many people into finding. A privately profitable and genuinely productive technology can expand solving capacity while reducing human welfare. This is our central mechanism.

Our notion of a *problem* is deliberately broad: any unmet need or opportunity to which knowledge can be productively applied. Finding valuable problems is a skilled activity, requiring knowledge and judgment to recognize a need, see the opportunity it presents, and bring it forward. Starting a business often begins with finding a problem – an entrepreneur recognizes a customer need but may require engineers, scientists, managers, or other experts to address it. A lawyer recognizes a potential claim, a security researcher discovers a vulnerability, an investor identifies an opportunity, and a scientist formulates a research question. Finding and solving problems may occur in the same sector or organization, but they remain economically distinct.

The process of finding a useful problem, transferring it to someone who can solve it, and then solving it, is most visible when it happens outside any particular firm, although our model can accommodate transfers inside a firm. A canonical example of the transfer step is referrals between professionals (see e.g., [Garicano and Santos \(2004\)](#)). Another is the separation of discovery and commercialization through the “market for ideas” (e.g. [Gans and Stern \(2003\)](#)). More recently, companies have formed to find and sell unsolved problems.

One example is Darrow, which finds legal risks (e.g. potential class actions) and sells them to law firms. From cofounder and CEO Ben Artzi: “Any legal risk is a future cash flow ... All you need to do is surface these opportunities, give them to the experts that can handle them ... and let them create the right trade.”¹

The distinction between finding a problem and solving it is the organizing principle of our general-equilibrium model, which jointly determines the supply of problems, workers’ occupations and training, and firms’ knowledge hierarchies. In the model, ex ante identical people choose between finding and solving problems. Knowledge is costly and makes people more effective in both activities. Solving occurs in knowledge hierarchies (Garicano, 2000; Garicano and Rossi-Hansberg, 2006). Less knowledgeable workers attempt problems first and pass difficult cases upward, allowing experts to concentrate where they are needed. Finding is decentralized and congested. Free entry and occupational choice connect firm organization to the economy-wide race for problems.

To this environment, we introduce AI – a replicable agent that can occupy any position in the knowledge hierarchy. Its capability determines what it can solve; its compute cost determines how cheaply it can be deployed.² We distinguish two forms of *AI progress*: becoming “smarter” through greater capability and becoming “cheaper” through lower compute cost.³ For humans, knowledge and its cost are linked by the training technology; for AI, the two dimensions can move independently.

We obtain three sets of results. First, AI splits the human knowledge hierarchy and makes AI cost improvements and AI capability improvements organizationally distinct. In an all-human hierarchy, problems are pushed up the hierarchy as agents ask for help and each layer connects those next to it, above and below. An agent’s knowledge governs how many problems reach the layer above, while the training cost of that knowledge governs the cost of helping the layer below. Therefore, a change at one layer propagates through the hierarchy in both directions, and the optimal skill for each layer reflects the layers on both sides. AI severs this recursion because its capability and compute cost are separate and

¹See <https://www.forbes.com/sites/jeffkaufman/2025/08/21/this-ai-fueled-startup-is-helping-attorneys-find-new-class-action-lawsuits/>. Similar to our AI as a finder extension (Section 7), Darrow uses a mix of AI and humans in finding legal risks to sell.

²By *compute cost* we mean the variable resources required to run AI on a problem, including inference hardware, energy, and related operating costs. We abstract from fixed development and training costs.

³Recent model launches illustrate the distinction. Anthropic described Claude Fable 5 as its most capable generally available model, with gains across capability benchmarks. DeepSeek-V3.2-Exp instead used sparse attention to preserve benchmark performance while reducing inference cost and cutting API prices by more than 50 percent.

exogenous to the hierarchy.

The break makes the two forms of AI progress organizationally distinct. The hierarchy separates into a set of human layers above AI, organized around its capability, and a set of human layers below AI, built on reducing compute cost. Smarter AI resolves more problems before they reach managers, so the layers above it shrink and acquire more knowledge. Cheaper AI reduces the incentive for workers below it to acquire knowledge merely to avoid referrals to the machine. Those layers de-skill and may disappear.

The same separation governs adoption. We characterize the adoption frontier – cost as a function of capability that makes a firm exactly indifferent from adopting AI or not – which is convex. The frontier touches the human cost curve when AI capability matches an existing layer’s capability and lies below the human cost curve between layers. When an AI exactly matches a human’s capability, it can replace a human layer without disturbing the rest of the ladder. If an AI is added between existing layers, then surrounding workers must retrain, so compute must be cheaper to make adoption worthwhile. We call this discount the *knowledge-mismatch gap*.

Because the supply of problems is endogenous, adoption of AI can be partial or full, and humans and AI can coexist within and across firms. An adopting firm expands solving capacity and bids up the price of problems until its profit disappears, while all-human firms may continue to break even. The two organizations then coexist over a range of technologies. Further AI progress can similarly produce coexisting configurations among all-AI firms. AI progress can thus generate gradual diffusion across firms but discrete changes in organization, training, and employment.

Second, human displacement, not AI adoption, determines welfare. At the adoption frontier, increasing AI capability or reducing its cost both lower human utility if and only if the adopting firm uses fewer humans per problem than the all-human firm it coexists with. Firms internalize the resources AI saves but not the congestion created when displaced solvers enter finding. Suppliers of scarce problems capture part of the technological gain, while duplicative search dissipates the rest. If adoption instead draws workers into solving, the welfare effect reverses. What matters is the direction of occupational reallocation, not whether firms profit from the technology.

If adoption is complete – AI is capable enough or cheap enough that all firms use the same AI configuration – then cheaper compute unambiguously raises utility. In contrast, greater capability can lower utility through displacement, shrinking managerial layers and releasing workers into finding. In either case, further improvements in cost or capability can

induce a transition in which firms using the old and new all-AI configurations coexist. During such transitions, replacing a human layer can reverse the welfare gains from lower compute costs. Thus, along a path of falling compute cost, the economy can alternate between smooth adjustment and transitions with opposing welfare effects so that utility traces a sawtooth – smooth gains interrupted by replacement losses and flat stretches – while employment adjusts in waves.

Third, first-best policy follows displaced workers rather than automatically targeting the technology. The primitive distortion is that finders receive their average rather than marginal social product, inducing excessive entry. A tax on finding income equal to the congestion externality, rebated lump-sum, decentralizes the first best. The first-best response to harmful AI adoption is therefore not an AI tax, but a tax on the activity into which displaced workers retreat. When AI sends workers crowding into an apparently safer occupation, efficiency may require taxing that refuge rather than protecting it.

When finding cannot be taxed directly, AI policy becomes a second-best instrument. A marginal tax on compute raises welfare if and only if it reduces aggregate finding. Depending on AI’s position and the stage of adoption, either a tax or a subsidy can do so. A tax on AI helps when further AI use releases solvers into finding; a subsidy helps when cheaper compute widens human solving layers and draws workers out of finding. Displacement therefore does not justify permanently taxing AI. The alternation of smooth adjustment and replacement transitions (the sawtooth utility pattern) implies that after a replacement episode, the tax can preserve the congestion it was meant to relieve.

Our results also have implications for what policymakers should measure. The practical test is not only who loses a job but where those workers go. Movement into sales, prospecting, business development, independent contracting, and other forms of problem finding may matter as much as headline employment. Capability benchmarks, task-level productivity, and job counts are therefore insufficient. Evaluating AI also requires tracking training, firm structure, problem finding, and occupational reallocation. Otherwise, we cannot tell whether AI’s private gains become social gains or instead ignite a costlier race for scarce work.

Finally, we consider a forward-looking extension in which AI finds as well as solves problems (as in our earlier example with Darrow). AI may increasingly recognize customer needs, originate business opportunities, identify legal claims or software vulnerabilities, and propose research directions – helping decide which questions are worth asking, not merely how to answer them. We model finding (discovery) as a non-hierarchical process, so AI competes with human finders on cost per unit of finding. While both coexist, any form of AI progress lowers

human utility through AI-driven congestion. Excessive entry persists, with compute joining human time in the race for problems. Once humans exit finding, further AI progress weakly raises utility because it cheapens problem supply without negatively impacting human efforts in finding.

When AI can be deployed in either sector, the sector of first deployment (the economy-wide adoption frontier) compares two protections for human labor. Human solvers are protected by the knowledge-mismatch gap, while human finders protected by their cost advantage in finding. Adoption in one sector lowers utility, and cheaper humans are less worth replacing, so adoption in one sector impedes adoption in the other.

Recent evidence maps naturally to these mechanisms. AI assistance has generated especially large productivity gains among inexperienced workers, even as early employment losses in exposed occupations appear concentrated among the young (Brynjolfsson et al., 2023, 2025). Our model shows how these patterns can coexist. Reorganizing around AI at a lower rung can raise productivity conditional on employment while reducing demand for lower-rung workers. AI also reduces escalation to supervisors in customer support (Brynjolfsson et al., 2023), performs unevenly across nearby consulting tasks (Dell’Acqua et al., 2026), and has diffused through narrow and uneven organizational deployments (Bonney et al., 2026; Humlum and Vestergaard, 2025). These findings support our focus on knowledge hierarchies, capability mismatch, and organizational coexistence.

2 Related Literature

This paper introduces problem finding into the theory of knowledge hierarchies (Garicano, 2000; Garicano and Rossi-Hansberg, 2006). Knowledge-hierarchy models explain how firms organize expertise to solve problems but generally take the supply of problems as given. In this tradition, heterogeneous agents sort across layers according to their cost of acquiring knowledge. We instead study ex ante identical agents whose training and organizational assignment are determined by free entry, and endogenize the supply of work through congested problem finding. Linking finding to solving makes the relative scarcity of problems and expertise endogenous. AI then changes not only how firms solve problems, but also how workers train, which opportunities firms pursue, and how intensely people compete to find worthwhile applications for the economy’s solving capacity.

On the organizational side, hierarchies allow less knowledgeable workers to handle routine problems and refer exceptions upward, economizing on expert time. Fuchs et al. (2015) study

the contracts that support such hierarchies when the difficulty of problems and the knowledge of consultants are hard to verify. We abstract from these informational frictions in the transfer of problems and instead . Closest to our analysis, [Ide and Talamàs \(2025\)](#) introduce AI into a knowledge hierarchy in which human knowledge is endowed, production opportunities are abundant, and a fixed stock of compute is embodied in AI agents with a common knowledge level. Their equilibrium is efficient, so AI reorganizes production and redistributes earnings without reducing aggregate welfare. We instead endogenize human training and the supply of problems, and allow AI capability and compute cost to vary independently. This independence breaks the recursive connection between adjacent human layers and generates our organizational separation, adoption frontier, and knowledge-mismatch gap. As we show (Section 5.5), if we freeze skills while allowing problems to be scarce and adding AI, then a market clearing argument generates price and wage changes but not AI adoption. Other recent work studies how AI affects corporate investment and the flow of information between managers and boards ([Chen and Han, 2026](#); [Ferreira and Li, 2026](#)).

Our displacement results also connect to task-based models of automation ([Acemoglu and Restrepo, 2018, 2019](#)), although the displaced margin here is a layer of an endogenous knowledge hierarchy rather than a task. Especially closely related is a growing literature on how automating entry-level work affects the production of expertise. In [Ide \(2026\)](#), novices acquire skill through apprenticeship, and automation reallocates them toward worse mentors, permanently lowering growth even without reducing novice employment. In [Chen \(2026\)](#), junior work produces both current output and future senior talent, so AI can thin the talent pipeline by eliminating the work through which expertise is created. In [Cai \(2026a\)](#), reliance on AI raises current productivity but slows learning-by-doing and degrades the human-generated data available to future systems, producing persistent skill polarization. Related work studies career viability under AI ([Garicano and Rayo, 2026](#)), the unbundling of production from learning ([Chaigneau and Jo, 2025](#)), and the effect of AI on human decision-making skills ([Chen et al., 2026](#)). Our analysis brings these questions into an endogenous multilayer hierarchy and distinguishes greater AI capability from lower compute cost. The two forms of progress act on opposite sides of the human hierarchy, while the connection to congested finding makes the destination of displaced workers central to welfare and policy.

A related strand studies how AI changes the stock, quality, and ownership of the knowledge on which future production depends. [Acemoglu et al. \(2026\)](#) study a learning externality in the accumulation of public knowledge. [Cai \(2026b\)](#) shows that AI adoption can degrade the data used to train future systems, while [Cullen et al. \(2026\)](#) study the ownership of the

human expertise on which AI is trained. These papers highlight dynamic feedback through human and machine knowledge. Our problem-finding margin introduces a contemporaneous externality even when those knowledge stocks are fixed.

Our finding sector shares the rent-dissipation logic of classic patent races, in which duplicative research effort can generate excessive entry (Loury, 1979; Reinganum, 1982). Hopenhayn and Squintani (2021) extend this logic across heterogeneous research problems, but those problems arrive exogenously and congestion arises as researchers compete to solve them. Here, agents compete to discover problems that knowledge hierarchies then solve. Linking the two activities makes problem prices, occupational allocation, training, and firm organization jointly determined.

3 The Model

3.1 Humans and Problems

The economy is populated by a unit measure of ex ante identical humans. Each human has one unit of time. Humans may use their time to find unsolved problems or to solve the problems that have been found. We refer to the former as *finders* and to the latter as *solvers*. Solved problems generate one unit of divisible output, which is the numeraire good. Problems have random difficulty $Z \sim H$.

Solving problems requires knowledge. A more knowledgeable solver can solve a strict superset of the problems that a less knowledgeable human can solve. The knowledge of a human can be characterized by $\tilde{z}$, denoting the most difficult problem she can solve. Let $q = H(\tilde{z}) \in [0, 1]$ denote the fraction of problems she can solve. Accumulating knowledge requires training, which is costly. The cost of acquiring knowledge q is $z(q)$. We assume throughout that $z'(q) > 0$, $z''(q) > 0$, $z(0) = 0$, and $z(1) = \infty$. For numerical examples, we adopt the log specification: $z(q) = -t \log(1 - q)$, with $t > 0$.

3.2 Firms and Hierarchies

A human engaged in solving can attempt one problem per unit of time, which she solves if its difficulty does not exceed her knowledge. In a firm, an unsolved problem can be passed up the hierarchy for help. A supervisor spends $h < 1$ units of time helping, and the problem is solved if it is within the supervisor's knowledge. Workers do not observe a problem's

difficulty, only whether they can solve it, so problems are passed up one layer at a time.⁴

For a given firm, let $L \geq 0$ denote the number of managerial layers; $L = 0$ corresponds to self-employment, and $L = 1$ to one layer of workers and one of managers. Label by q_l the knowledge of the humans in layer l and by n_l their mass. Layer $l + 1$ receives $n_0(1 - q_l)$ problems and solves the fraction $\frac{q_{l+1} - q_l}{1 - q_l}$ of them, so layer $l + 1$ is fully occupied helping layer l when

$$n_{l+1} = hn_0(1 - q_l). \quad (1)$$

When every layer is fully occupied, the firm attempts n_0 problems and solves $n_0 q_L$ of them. We write $s = \sum_{l=0}^L \frac{n_l}{n_0}$ for the number of people per problem attempted and refer to s as the *intensity* of solving.

Firms enter freely and arrange people within the firm to maximize profit. An optimizing firm leaves no manager with free time, so (1) describes the relative size of layers. Without loss, we assume that the firm reimburses workers for the cost of training.

We can interpret the model as one in which firms may produce problems internally by hiring additional workers as finders, or one in which firms buy problems externally in a decentralized market. Unsolved problems are bought and sold at price p , or p is the payment to finders within a firm. Our analysis is not sensitive to the interpretation, and we will use the decentralized market in our discussions.⁵

3.3 Finding Problems

Finding problems also requires knowledge, which is acquired via the same training technology as in the solving sector. A finder chooses knowledge q_F at cost $z(q_F)$. Each finder contributes finding effort equal to $f q_F$ for some $f > 0$. The aggregate finding effort is denoted by $Q = f n_F q_F$ where n_F is the measure of human finders. The quantity of problems found depends on aggregate finding effort, Q , and the aggregate needs in the economy, N :

$$\text{Problems found} = N^\gamma Q^{1-\gamma},$$

⁴Sequential referral is the efficient way to match problems with knowledge when problems cannot be cheaply labeled, as when knowledge is tacit (Garicano, 2000).

⁵In a knowledge-hierarchy framework, Fuchs et al. (2015) show that when knowledge is private information, spot contracting is inefficient and firm-like arrangements are needed.

where $\gamma \in (0, 1)$. The finding technology exhibits constant returns to scale in (N, Q) , but decreasing returns to scale in Q alone. Needs are proportional to the measure of humans, which we normalize to one, so that the measure of problems found is $G(Q) \equiv Q^{1-\gamma}$.

A finder's output is proportional to her effort: a finder with knowledge q_F produces $f q_F Q^{-\gamma}$ problems. Decreasing returns to scale in Q generates a congestion externality. When one finder increases her effort, every other finder's output falls.

Remark 1 *To further motivate the congestion externality – the decreasing-returns technology for finding – we can micro-found it as part of a constant-returns, undirected search technology.⁶ Let there be a pool of N needs and let Q be the aggregate effort spent finding them. Each need has a visibility $\nu > 0$ describing how easily effort leads to discovery, with Pareto density $g(\nu) = (1 - \gamma)\nu_0^{1-\gamma}\nu^{-2+\gamma}$ on $[\nu_0, \infty)$. Effort is spread evenly over the pool and discovery events are independent Poisson, so a need of visibility ν is discovered with probability $1 - e^{-\nu Q/N}$, and a found need counts once no matter how many times it is found. The total measure discovered is $N \int_{\nu_0}^{\infty} (1 - e^{-\nu Q/N}) g(\nu) d\nu$, which for N large relative to Q is proportional to $N^\gamma Q^{1-\gamma}$. Doubling effort does not double finding output because the same need cannot be found twice.*

3.4 AI Technology

We first study a benchmark economy without AI. Beginning in Section 5, we introduce a replicable AI agent with capability q^A , the fraction of problems it can solve, measured on the same scale as human knowledge. A firm can deploy any number of AI agents. Each uses one unit of compute per problem, at cost κ , and has the same processing capacity as one human worker. A firm may deploy AI at any solving position in its hierarchy or in finding. The pool of needs, however, continues to be generated by people. Deploying AI to find or solve problems does not create additional needs in the economy. Throughout, phrases such as “AI at position a ” are shorthand for a firm deploying AI there. Firms make all AI deployment decisions.

⁶This construction combines the search technology of [Kortum \(1997\)](#) with duplicative search over a pool of problems, as in [Butters \(1977\)](#) or [Hopenhayn and Squintani \(2021\)](#). The resulting congestion externality is related to [Stokey \(1995\)](#) and reflects the “stepping on toes” effect emphasized by [Jones and Williams \(2000\)](#).

4 The Benchmark Economy without AI

People are ex ante identical and free to enter either role at any knowledge level, so in equilibrium every human obtains the same utility, which we denote by U and sometimes refer to as *welfare*. Thus, the total cost to a firm of employing a human with knowledge q is

$$w(q) = U + z(q).$$

We refer to $w(q)$ as the wage paid to a human with knowledge q , which includes her cost of training.

4.1 Firms and Solving

Firms require two inputs: problems, which they buy at price p , and solvers, which they hire at an equilibrium wage. Firms are price takers for both inputs. A *configuration* is a depth L . Given its configuration, a firm chooses a *hierarchy* of knowledge levels $0 \leq q_0 < \dots < q_L \leq 1$, one for each layer of the hierarchy. Because knowledge is ordered, the sets of problems the layers can solve are nested. Each problem travels up until it is solved or reaches the top, and the firm therefore ultimately solves the fraction q_L of the n_0 problems it buys.

Adding up the layers, a firm's profit is

$$\begin{aligned} \pi &= n_0 q_L - n_0 p - \sum_{l=0}^L n_l [U + z(q_l)] \\ &= n_0 \left(q_L - p - h \sum_{l=1}^L (1 - q_{l-1}) [U + z(q_l)] - [U + z(q_0)] \right), \end{aligned} \quad (2)$$

where the second equation uses (1). Profit is linear in the scale n_0 of the firm, so the relevant object is profit per problem attempted: given (p, U) , the maximized profit of a depth- L firm is

$$\Pi_L(p, U) \equiv \max_{0 \leq q_0 < \dots < q_L \leq 1} q_L - p - h \sum_{l=1}^L (1 - q_{l-1}) [U + z(q_l)] - [U + z(q_0)]. \quad (3)$$

Notice that the objective in (3) is supermodular in $(q_0, \dots, q_L, U)$. Therefore, holding L fixed, an increase in U raises the knowledge of every layer. More expensive human solvers make the manager's time worth saving.

Maximizing with respect to knowledge at an interior layer l gives

$$FOC(q_l) : -h(1 - q_{l-1})z'(q_l) + h[U + z(q_{l+1})] = 0 \quad (4)$$

or $z'(q_l) = hw(q_{l+1})/n_l$. The interior first-order conditions have a common structure that will be central when we introduce AI. The marginal value of knowledge at layer l trades off the mass of that layer, which is set by the knowledge of the layer *below*, against the cost of passing problems to the layer *above*.

Free entry drives each firm's maximum profit at the optimal configuration to zero,

$$\max_{L \geq 0} \Pi_L(p, U) = 0. \quad (5)$$

Observe that Π_L is continuous and strictly decreasing in the utility the firm must deliver and the price p it must pay per problem. Thus, for each price p , there is a unique utility level satisfying (5), which we denote as $U_S(p)$ and refer to it as the “firm break-even curve for solving”. The curve captures the highest utility to the solvers a firm can pay while breaking even at a given problem price p . Let $p(U)$ denote the inverse, the highest price that any firm can pay for a problem while still breaking even given that humans command utility U .

We are particularly interested in how many agents a firm uses per problem attempted, which we define as s :

$$s \equiv \sum_{l=0}^L \frac{n_l}{n_0} \quad (6)$$

By the envelope theorem, $\partial \Pi_L / \partial U = -s$. Similarly, by applying the implicit function and envelope theorems to (5) yields

$$\frac{\partial}{\partial p} U_S(p) = -\frac{1}{s} \in [-1, 0) \quad (7)$$

wherever the maximizing depth L is unique. When problems become more expensive, the loss is spread over the s humans who work on each problem.

Lemma 1 *$U_S(p)$ is convex and strictly decreasing, and s is weakly increasing in p . Within a fixed configuration, an increase in p weakly reduces the knowledge of every layer.*

When problems are more expensive, the utility firms can pay solvers decreases, and solvers are relatively cheap. Then, knowledge, which allows agents lower in the hierarchy to

economize on the manager's time, is less valuable. By the supermodularity of (3), the set of optimal hierarchy layers shifts downward, so when the optimizer is unique, every layer de-skills.⁷ The other content of the lemma is the shape of the envelope. As the price of problems rises, firms first *retrain*, shifting optimal knowledge downward within the configuration, and then *transition* to a different configuration. The transitions correspond to kinks of $U_S(p)$ at which s jumps.

4.2 The Finders

Recall that p denotes the price of a problem. It will be convenient to measure the congestion-adjusted price $\rho \equiv pQ^{-\gamma}$. A finder with knowledge q finds $f q Q^{-\gamma}$ problems and earns utility

$$U_F = f q \rho - z(q), \quad (8)$$

Each atomistic finder takes ρ as given. She does not internalize her effect on Q . Finder's utility is maximized at q_F satisfying

$$z'(q_F) = \rho f. \quad (9)$$

Substituting (9) into (8) and applying the envelope theorem,

$$U_F = z'(q_F) q_F - z(q_F), \quad (10)$$

$$\frac{\partial}{\partial \rho} U_F(\rho) = f q_F. \quad (11)$$

The finder's problem is strictly concave, so its solution is unique, and the right side of (10) is strictly increasing in q_F (since $q z''(q) > 0$). Finder utility, finder knowledge, and the congestion-adjusted price move together.

Notice that each finder is paid her average product per unit of effort $f\rho$, while that effort adds only the marginal product $(1-\gamma)f\rho$. Thus, finding effort is inefficiently high. The social value of finding, $n_F U_F = p(f n_F q_F)^{1-\gamma} - n_F z(q_F)$, is maximized where $z'(q_F) = (1-\gamma)\rho f$. This *finding-congestion* wedge is the single distortion in the benchmark economy.

⁷ If $1 \leq z'(0)$, then $q_0 = 0$ in every firm, and no human is self-employed. When the first unit of knowledge costs more than the output of a problem, production workers are left untrained. Their role is to attempt problems and pass the failures up. This is still valuable because unsolved problems require only h units of manager's time. The same condition rules out self-employment, since a self-employed human would optimally choose zero knowledge and earn negative utility.

4.3 Market Clearing

Because profit is linear in scale, free entry permits several firm configurations (hierarchy depths L) to break even at the same (p, U) . We index configurations by c , write s^c for the intensity of configuration c , and let $\theta_c \geq 0$ denote the share of problems attempted by firms of configuration c , with $\sum_c \theta_c = 1$. The average intensity of solving is then

$$\bar{s} \equiv \sum_c \theta_c s^c. \quad (12)$$

All finders choose the same knowledge, so aggregate finding effort is $Q = fn_F q_F$ and $Q^{1-\gamma}$ problems are found. Solving them occupies $\bar{s}$ people per problem, and finding occupies the rest of the population, so market clearing in problems and people requires

$$\bar{s} Q^{1-\gamma} + n_F = 1. \quad (13)$$

When a single configuration is active, $\bar{s} = s$, and we drop the index.

4.4 Equilibrium

An *equilibrium candidate* is a problem price p , a mass n_F of finders with knowledge q_F , and a finite set of active configurations with problem shares $\theta_c > 0$ summing to one. The remaining objects constructed from the candidate: aggregate finding effort is $Q = fn_F q_F$, the congestion-adjusted price is $\rho = pQ^{-\gamma}$, and utility is the highest available in either role, $U = \max\{U_F(\rho), U_S(p)\}$.

Definition 1 *An equilibrium candidate is a competitive equilibrium if: (i) q_F maximizes (8) given ρ ; (ii) every active configuration maximizes profit and earns zero profit; (iii) every occupied role attains U : $U_F(\rho) = U$ if $n_F > 0$, and $U_S(p) = U$ if humans are employed in solving; (iv) the markets for problems and people clear: (13) holds.*

In the benchmark economy, both roles must be populated in an equilibrium so condition (iii) reduces to the indifference $U_F(\rho) = U_S(p)$. We state condition (iii) in its weaker form because it survives the introduction of AI. Thus, the equilibrium can be organized around a single utility level U . In the finding sector, U determines q_F and ρ , both increasing, by (9) and (10). In the solving sector, U determines the set of optimal hierarchies. Then, the zero-profit condition determines the break-even price $p(U)$, which falls as U rises by (7).

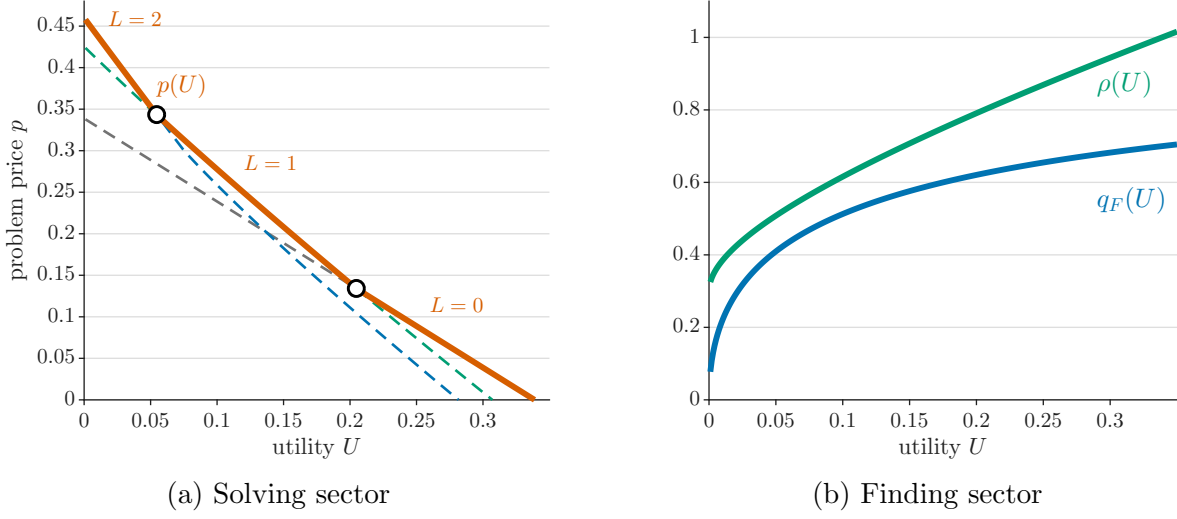


Figure 1: The solving sector of the equilibrium construction. Panel (a): the break-even price $p(U)$ is the upper envelope over $p_L(U)$ (dashed). Circles mark the kinks in $p(U)$ where the optimal depth switches. Panel (b): the finder's knowledge $q_F(U)$ and the congestion-adjusted price $\rho(U)$, both strictly increasing. Log specification, $h = 0.8$, $t = 0.3$, $f = 1$.

Figure 1a illustrates how U determines the price of problems and firm depth. Figure 1b illustrates how U determines finder knowledge and the congestion-adjusted price.

Given p and ρ , the congestion identity $\rho = pQ^{-\gamma}$ then implies the aggregate finding effort Q . Market clearing for problems implies the finder mass $n_F = Q/(fq_F)$. The equilibrium utility is then determined by market clearing for people, the level of U at which the resulting demand for people, $\bar{s}G(Q) + n_F$, equals the population. The configuration shares θ_c enter only through the average solving intensity $\bar{s}$. When a single configuration is active, $\bar{s}$ is pinned down by (6). When several configurations are active, market clearing determines the required mix. Figure 2 illustrates these two possibilities.

Assumption 1 (Viability) $h z'(0) < 1$.

Assumption 1 is necessary and sufficient for production to be viable and we maintain it throughout. If the assumption fails, the training bill of every organizational configuration weakly exceeds its output even when workers receive zero utility and problems are free. No firm can then cover the cost of its workforce.⁸

⁸To see that Assumption 1 is sufficient for production, note that $q - hz(q)$ is the profit of an $L = 1$ firm with a worker of zero skill at layer zero when workers receive zero utility and problems are free. It equals zero at $q = 0$ and is strictly increasing there, since $1 - hz'(0) > 0$. Hence $p(0) > 0$. Conversely, if Assumption 1 fails, then convexity and $z(0) = 0$ imply $z(q) \geq z'(0)q \geq q/h$. Hence, $h(1 - q_{l-1})z(q_l) \geq (1 - q_{l-1})q_l \geq q_l - q_{l-1}$.

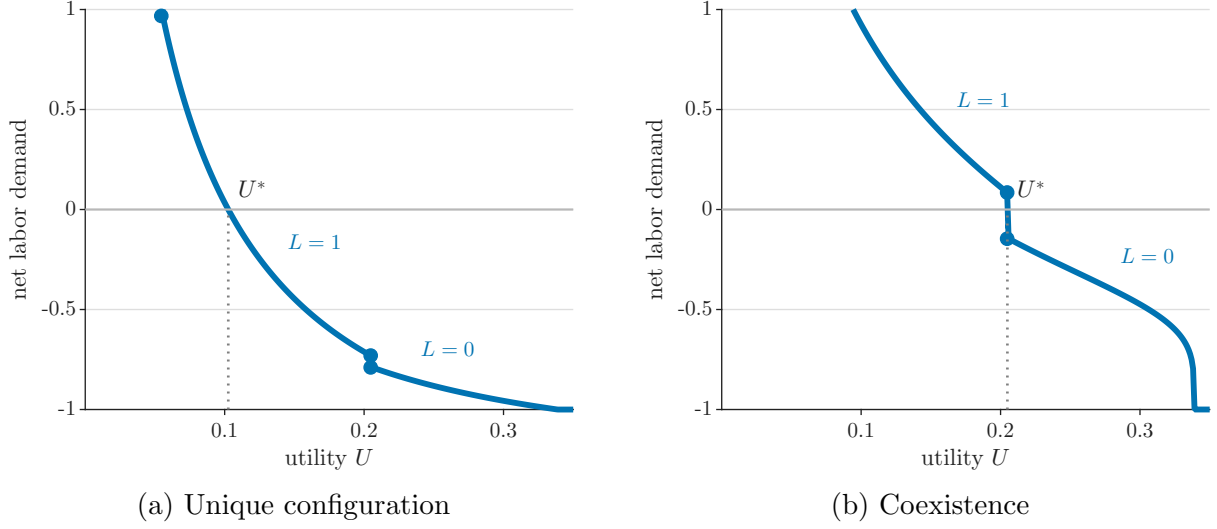


Figure 2: The net demand for people, $\bar{s}G(Q) + n_F - 1$, is strictly decreasing in U , with downward jumps where the optimal configuration switches. Panel (a) illustrates an equilibrium where a single $L = 1$ configuration is active. Panel (b): net demand crosses zero inside a jump and two configurations coexist. Log specification, $h = 0.8$, $t = 0.3$, $f = 1$, $\gamma = 0.5$ in (a), $\gamma = 0.82$ in (b).

Theorem 1 *A competitive equilibrium exists. Its utility, price, finder knowledge, finding effort, finder mass, and average intensity $(U^*, p^*, q_F^*, Q^*, n_F^*, \bar{s}^*)$ are unique and positive, and the depths of the active configurations differ by at most one layer. Generically, the configuration shares are also unique and at most two configurations are active.*⁹

Existence follows from the boundary behavior and upper hemicontinuity of labor demand. At $U = 0$, finders acquire no knowledge and no problems are found, while labor is cheap and firms would pay a high price for problems meaning the problem market cannot clear. At high U , firms break even only at a low problem price while finders require a high reward. A low p and a high ρ mean a small Q , so few problems are found, and the labor market cannot clear. In between, demand falls strictly as U rises. When a single configuration is active, demand for solvers falls continuously with U . At a kink of the break-even price it

Summing across layers and adding $z(q_0) \geq q_0$ shows that the wage bill of any configuration is at least q_L even when $U = 0$.

⁹Generically means for all parameters outside a knife-edge set, here the set on which three or more configurations break even at U^* . Lemma A.3 bounds the depths of those configurations, but not their number. Any two break even jointly only at a single utility, so a third requires a second such coincidence at U^* , which requires one further equation on the primitives (h, t, f, γ) . In this non-generic case, the active shares are not uniquely determined, only $\bar{s}^*$ is pinned down.

jumps down, because the shallower configuration uses fewer people per problem. Mixtures of the two configurations attain every level of demand in between. Demand is therefore strictly decreasing and upper hemicontinuous with convex values, so it crosses the population exactly once.

5 AI as a Solver

We now allow firms to deploy AI agents within the solving hierarchy. A configuration consists of the depth L and, when the technology is adopted, a position $a \in \{0, \dots, L\}$ at which AI is deployed. The hierarchy specifies knowledge levels for the remaining human layers. A change of configuration is a *transition*, whereas a change in the hierarchy at a fixed configuration *retrains* the firm's workers. To keep every layer fully occupied, a firm deploying AI at position a uses $h(1 - q_{a-1})$ units of compute per problem attempted, with one unit at $a = 0$. For a configuration that uses AI, let s_h denote the number of humans per problem attempted (analogous to s from (6)). Only humans enter labor-market clearing, so $\bar{s}$ in (13) averages s_h across active configurations.

This section restricts AI to solving. Section 7 studies AI in finding.

5.1 AI and Human Knowledge

The AI deployed at position a decomposes the firm's organizational structure into two components. The human layers above form an *upper-hierarchy* built on capability q^A : at market utility U , the value of the best upper-hierarchy above knowledge x is

$$V(x) = \max_{m \geq 0, x < q_1 < \dots < q_m < 1} q_m - h \sum_{l=1}^m (1 - q_{l-1}) [U + z(q_l)] \quad (q_0 \equiv x), \quad (14)$$

where an empty upper-hierarchy ($m = 0$) delivers value x itself. The human layers below AI, together with the compute bill, form a *lower-hierarchy*: the cheapest lower-hierarchy of depth $a \geq 1$ costs

$$B_a(\kappa) = \min_{0 \leq q_0 < \dots < q_{a-1} < q^A} [U + z(q_0)] + h \sum_{l=1}^{a-1} (1 - q_{l-1}) [U + z(q_l)] + h(1 - q_{a-1}) \kappa, \quad (15)$$

and $B_0(\kappa) = \kappa$ when the firm deploys AI at the bottom layer. Profit per problem attempted is then

$$\pi = V(q^A) - B_a(\kappa) - p. \quad (16)$$

Critically, the upper- and lower-hierarchies share no variables. The upper-hierarchy depends on q^A , the AI capability on which it is built, but not on κ ; the lower-hierarchy depends on κ , through the compute bill, but not on q^A . AI therefore severs the human hierarchy, and the two halves are linked only through zero profit and the common market utility. The layer first-order conditions have the same interpretation at the margin. Spelling out the interior condition,

$$FOC(q_l) : \quad - \underbrace{h(1 - q_{l-1})}_{n_l/n_0, \text{ set by } q_{l-1}} \times \underbrace{z'(q_l)}_{\text{marginal cost}} + \underbrace{h[U_S + z(q_{l+1})]}_{\text{cost of the layer above}} = 0 : \quad (17)$$

Human knowledge and human cost are jointly determined in equilibrium, so a human layer transmits changes from the bottom of the ladder to the top and back. The two independent AI parameters interrupt this transmission.¹⁰ If a firm deploys AI in layer $l + 1$, the relevant cost above layer l is κ , regardless of q^A ; if the AI is in layer $l - 1$, the mass of human layer l is $h(1 - q^A)$, regardless of κ .

Proposition 1 (AI severs the knowledge hierarchy) *Fix a configuration with AI deployed at position a , and suppose that $q_{a-1} < q^A < q_{a+1}$ at the optimum. An increase in q^A weakly raises the knowledge of every human layer above position a and does not change the layers below. A decrease in κ weakly lowers the knowledge of every human layer below position a and does not change the layers above.*

As AI becomes smarter or cheaper, the human layers on either side of it change to move away from its capability in a form of *skill repulsion*. A smarter AI leaves fewer unsolved problems for the manager directly above, which shrinks that layer and makes knowledge cheap relative to mass. Managers above it therefore up-skill. Cheaper AI reduces the value to the worker directly below of asking the AI for help, so workers below it de-skill.

Pushed far enough, this de-skilling eliminates affected layers in the lower hierarchy. As κ falls, optimality pushes q_{a-1} down until it is below q_{a-2} and a layer must exit. Once a

¹⁰The distinction reflects the independence of κ and q^A , whereas a human worker's knowledge and training cost are linked by $z(q)$. The ladder split does not obtain in [Ide and Talamàs \(2025\)](#).

firm deploys AI directly above the bottom layer, those workers' only reason to train is to economize on compute for their AI managers, and their first-order condition is $z'(q_0) = h\kappa$. The optimum is at the corner $q_0 = 0$ whenever the first unit of knowledge costs more than the compute it saves, $z'(0) \geq h\kappa$. Under the condition of footnote 7, $z'(0) \geq 1$, the bottom layer is untrained for every $\kappa \leq 1/h$. As $\kappa \rightarrow 0$, layers continue to exit until every human layer below AI is eliminated. Because $U > 0$ in equilibrium, even an untrained human costs more than a very inexpensive AI agent.

5.2 The AI Adoption Frontier

Fix the no-AI equilibrium, with utility U^* and break-even price p^* , and suppose the profit-maximizing configuration and its ladder $q_0^* < \dots < q_L^*$ are unique. The *adoption frontier* $\bar{\kappa}(q^A)$ is the highest compute cost at which some AI configuration can break even at p^* . If $\kappa > \bar{\kappa}(q^A)$, every configuration using AI earns negative profit at p^* , so no firm adopts it. If $\kappa < \bar{\kappa}(q^A)$, some configuration using the AI earns positive profit at the no-AI equilibrium price, so some or all firms adopt it.¹¹

Proposition 2 (The adoption frontier) *The adoption frontier, $\bar{\kappa}$, is convex, and $\bar{\kappa}(q^A) \leq U^* + z(q^A)$, with equality at the layers of the human ladder, $\bar{\kappa}(q_l^*) = U^* + z(q_l^*)$, and tangency to the human cost curve at the interior layers, $\bar{\kappa}'(q_l^*) = z'(q_l^*)$.*

We illustrate the results of Proposition 2 in Figure 3. The adoption frontier lies below the human cost curve and touches it where AI replicates an existing human layer, tangentially at every interior layer. When $q^A = q_l^*$ and compute cost equals the corresponding human's wage, $U^* + z(q_l^*)$, a firm can substitute AI for that layer without otherwise changing the rest of the hierarchy. This makes the firm indifferent. Away from the benchmark layers, fitting AI into the hierarchy requires a different surrounding human hierarchy, which creates a mismatch cost. Compute must be strictly cheaper than the cost of a human with knowledge q^A . We refer to the gap $U^* + z(q^A) - \bar{\kappa}(q^A)$ as the *knowledge-mismatch gap*. It is the cost of the mismatch between AI capability and the surrounding human ladder, and it will reappear in Section 7 as the solvers' protection against replacement.

¹¹For sufficiently low q^A , no AI configuration can break even even when compute is free. In that region, we define $\bar{\kappa}(q^A) < 0$ to be the virtual compute cost at which the best AI configuration would just break even. This convention defines the frontier on all of $[0, 1)$ and makes $\bar{\kappa}(q^A) \geq 0$ a necessary condition for adoption at a nonnegative compute cost.

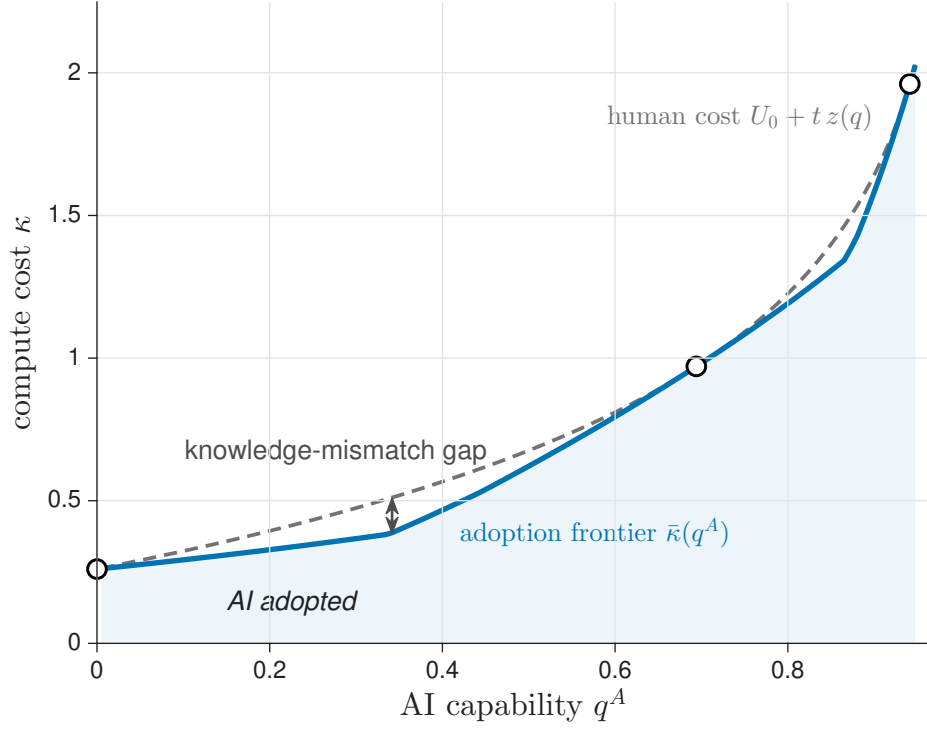


Figure 3: The solver adoption frontier $\bar{\kappa}(q^A)$ (solid) and the human cost curve $U^* + z(q)$ (dashed) at the no-AI equilibrium. Circles mark the rungs of the human ladder, where the frontier touches the cost curve, tangentially at every interior rung.

5.3 Private versus Social Gains from AI

An improvement in AI corresponds to a movement in the (q^A, κ) plane: an increase in AI capability, reduction in compute cost, or both. We evaluate such improvements from two perspectives. First, we measure their value to an individual firm, holding the rest of the economy fixed. Second, we measure their value to the economy as a whole, allowing equilibrium prices to adjust.

For a firm deploying AI at position a , define

$$\delta^+ = \begin{cases} h[U + z(q_{a+1})] & a < L, \\ 1 & a = L, \end{cases} \quad \delta^- = \begin{cases} h(1 - q_{a-1}) & a \geq 1, \\ 1 & a = 0 : \end{cases} \quad (18)$$

at $a = L$ the solutions produced by the AI are the firm's output, and at $a = 0$ the firm uses one unit of compute per problem. δ^+ is the marginal value of AI capability to the firm, and δ^- the marginal value of cheaper compute. Equivalently, δ^+ is the slope of V at q^A and δ^-

the slope of B_a in κ , so the two cases in (18) are each one envelope derivative of the split (16). The value of an increase in capability is the wage bill that each additional solved problem saves, as determined by the human layer above the slot. The value of a reduction in κ is proportional to the amount of compute used, as determined by the layer below. Therefore, the two gains separate in the hierarchy – the wage bill above grows with the AI’s position, while compute use is largest at the bottom – so greater AI capability is more valuable near the top of the hierarchy, whereas cheaper compute is more valuable near the bottom.

Now, we differentiate profit per problem attempted along any improvement path in the (q^A, κ) plane. By the envelope theorem, the induced adjustment of human knowledge levels within the fixed configuration does not enter the profit differential directly. Only the market prices and the terms associated with AI remain:

$$d\pi = -s_h dU - dp + \delta^+ dq^A - \delta^- d\kappa. \quad (19)$$

Two rates of substitution between capability and compute cost follow from (19), depending on what is held fixed.

For the private substitution rate, a single firm takes the market as given, so $dU = dp = 0$. Equation (19) then collapses to the iso-profit curve of a firm operating the given configuration,

$$\left. \frac{d\kappa}{dq^A} \right|_{dU=dp=0} = \frac{\delta^+}{\delta^-}. \quad (20)$$

The iso-profit curve is the firm’s private substitution rate between AI capability and compute cost that keep profits constant. Thus, an improvement in AI that crosses iso-profit curves is one the firm will pay for. The curve for zero profit at the no-AI equilibrium is the adoption frontier of Section 5.2.

For the general-equilibrium iso-utility curve, we hold utility fixed but allow the problem price to adjust. Free entry keeps the active AI configuration at zero profit, so with $dU = 0$ and $d\pi = 0$, (19) becomes

$$\left. \frac{d\kappa}{dq^A} \right|_{dU=0} = \frac{\delta^+}{\delta^-} - \frac{1}{\delta^-} \frac{dp}{dq^A}. \quad (21)$$

The social substitution rate is the private one plus a correction for the movement in the price of problems.

To continue, we will assume that the economy only has one configuration (one firm type). This allows us to infer the impact of that firm on the economy, and we loosen this assumption in Section 5.4. Fixing U freezes the finding sector (q_F and ρ cannot change if U cannot change), so the price can move only through congestion. Differentiating the congestion identity $\rho = pQ^{-\gamma}$ and market clearing at fixed (ρ, q_F) yields

$$dp = -C ds_h, \quad C \equiv \frac{\gamma p Q^{1-\gamma}}{1 - \gamma s_h Q^{1-\gamma}} \geq 0, \quad (22)$$

where the congestion factor C describes the impact of reallocating people on the price. In the solving sector, the price does not enter any first-order condition, so within a configuration the humans per problem depend on (U, q^A, κ) alone: $ds_h = c_q dq^A + c_\kappa d\kappa$, where $c_q \equiv \partial s_h / \partial q^A \leq 0$ and $c_\kappa \equiv \partial s_h / \partial \kappa \leq 0$ by Proposition 1. Substituting $dp/dq^A = -C ds_h/dq^A$ into (21) and solving for the slope yields the social substitution rate:

Proposition 3 (The iso-utility and iso-profit wedge) *In an economy with a single active configuration, iso-utility curves are flatter than iso-profit curves,*

$$\left. \frac{d\kappa}{dq^A} \right|_{dU=0} = \frac{\delta^+ + C c_q}{\delta^- - C c_\kappa} \leq \frac{\delta^+}{\delta^-} = \left. \frac{d\kappa}{dq^A} \right|_{d\pi=0}, \quad (23)$$

with equality if and only if the improvement reallocates no humans between sectors ($c_q = c_\kappa = 0$).

In words, a firm will accept a larger increase in compute cost for a unit of AI capability than society would. At $\gamma = 0$ there is no finding congestion – the price term in (22) vanishes and the two substitution rates coincide. With $\gamma > 0$, the two substitution rates differ by the congestion terms Cc_q and $-Cc_\kappa$, and both arise from the reallocation of people, which the planner internalizes but the firm does not. An increase in AI capability displaces humans into finding, which exacerbates congestion, while a cost reduction pulls people out of finding and relieves congestion.

5.4 The Welfare Effects of AI Improvements

When does an improvement in AI make people better off, and when worse? Consider a path of static equilibria as AI improves, beginning with (q^A, κ) above the adoption frontier. Figure 4 traces one such path, along which compute becomes cheaper at fixed capability.

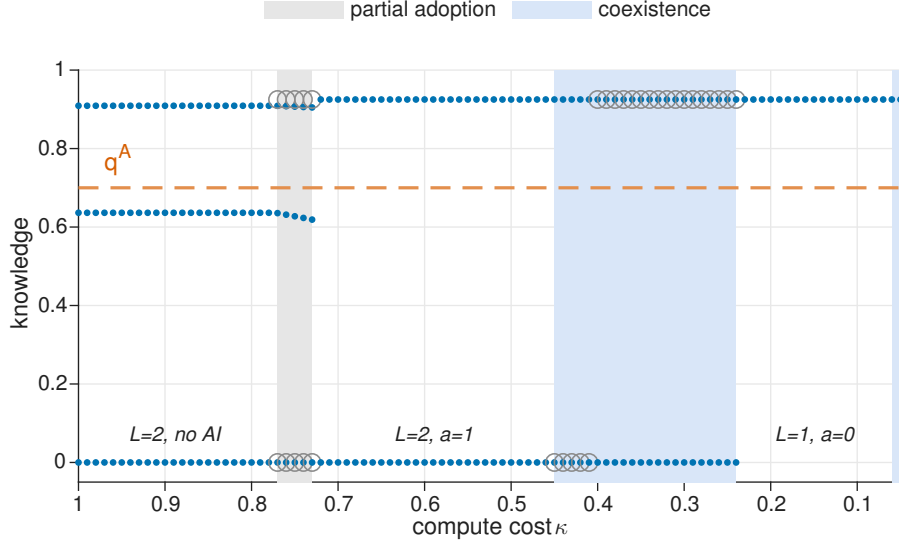


Figure 4: Equilibrium configurations and human knowledge as compute becomes cheaper. Dots: the rungs of the active configuration. Circles: the rungs of the second configuration where two coexist. Dashed line: the AI's capability. Log specification, $q^A = 0.7$, $h = 0.5$, $t = 0.5$, $f = 1$, $\gamma = 0.75$.

When the frontier is crossed, some firms adopt AI while others do not, and the economy enters a *partial adoption region* in which all-human and AI-adopting firms coexist. As AI improves further, the all-human configuration ceases to be competitive, every firm adopts AI, and a single configuration is active. Further cost improvements make a different AI configuration competitive by eliminating the unskilled human layer, and two configurations coexist again, both using AI, until only the shallower configuration remains. The last two kinds of regions can alternate as AI continues to improve, and each switch of configuration is a transition in the sense of Section 4. In what follows, we describe the welfare effects of AI improvements in each of the three regions (partial adoption, single AI firm, coexistence).

5.4.1 The Partial Adoption Region

With partial adoption, all-human and AI-adopting firms coexist. Thus, both firms must earn zero profits before and after an AI improvement. For the all-human firm, this means $dp = -s dU$, an application of the envelope theorem (7). For the AI-adopting firm, this means $dp = -s_h dU + \delta^+ dq^A - \delta^- d\kappa$, which adds the direct value of the improvement to the single layer in which AI is used. Subtracting the first from the second eliminates the price and delivers the welfare effect:

Proposition 4 (Welfare at adoption) *Suppose all-human and AI-adopting firms coexist. Then $dU = -(\delta^+ dq^A - \delta^- d\kappa) / (s - s_h)$: every improvement in AI strictly lowers equilibrium utility and raises the price of problems if and only if adoption displaces human workers, $s_h < s$.*

When adoption displaces humans, AI progress makes problems scarce relative to solving capacity. Competition shifts the rents to the problems, and people, who are paid out of what is left, lose. For the all-human firm $dp = -s dU$, so a higher price of problems *is* a lower wage. With problems more valuable and utility lower, finding expands while solving sheds people – Q rises and q_F falls, so the finder mass grows – and the displaced solvers are absorbed at the congested margin, where the entrant is paid her average product but adds only her marginal product. Falling utility and displacement into finding are two sides of the same event. When $s_h < s$, the improvement’s gains are absorbed by the higher price of problems and dissipated by congestion. If adopting firms instead use more humans per problem, $s_h > s$, both the utility and price responses are reversed.

The following lemma provides multiple sufficient conditions for displacement at the adoption frontier:

Lemma 2 (Displacement at the frontier) *Let the configuration that uses AI deploy it at position a . Then,*

- (i) *If $a \leq L$ and $q^A \geq q_a^*$ then $s_h \leq s$.*
- (ii) *If $a \in \{0, 1\}$, then $s_h \leq s$ for every q^A , with strict inequality unless $q^A = q_0^*$ or $L = 0$.*
- (iii) *For each $a \leq L$, there is a neighborhood $\mathcal{N}(q_a^*)$ such that if $q^A \in \mathcal{N}(q_a^*)$ then $s_h < s$.*

Part (i) is the general principle. When AI is *qualified* for the slot it occupies (i.e., $q^A > q_a^*$), the surrounding hierarchy can use fewer humans per problem. The upper-hierarchy is built on more knowledge and thins out. The lower-hierarchy faces compute costs that exceed what the displaced human earned and therefore uses more knowledge to economize on compute. Part (ii) applies the same logic at the bottom of the hierarchy, where replacing the production layer dominates the possible addition of one span of managers. Part (iii) covers AI capabilities near the benchmark layers, where the AI capability closely resembles the knowledge of a human layer.

It is possible to construct examples in which adoption at the frontier does not involve displacement. For example, a configuration with AI deployed above the entire human hierarchy can have weakly more humans per problem than the all-human benchmark. It crosses the adoption frontier only when compute is sufficiently cheap, $\bar{\kappa} < 1/h$, and at unchanged

depth the lower human layers de-skill and widen: $s_h \geq s$.¹² Thus the sign of the welfare effects in the partial adoption region depend on the human intensity of the AI configuration, not on adoption alone.

5.4.2 One Representative Adopting Firm

Consider next a full adoption region with a single active AI configuration. We can re-use the analysis of Section 5.3, differentiating the whole equilibrium – zero profit, the finder block, congestion, and market clearing – to obtain the welfare gradient of the AI technology parameters:

Lemma 3 (The welfare gradient) *With a representative firm that has adopted AI, there is a positive factor Φ such that*

$$dU = \Phi \left[(\delta^+ + C c_q) dq^A - (\delta^- - C c_\kappa) d\kappa \right]. \quad (24)$$

In particular, $\partial U / \partial \kappa < 0$, so cost reductions always raise welfare. $\partial U / \partial q^A$ shares the sign of $\delta^+ + C c_q$.

Both forms of AI progress carry a positive direct effect. Cheaper compute saves δ^- per unit and greater capability is worth δ^+ per problem. The indirect effect is the movement of people across the congested finding margin, and there the two differ in sign. Cheaper AI leads the workers below to de-skill and their layers to widen, so solving absorbs more people per problem and draws them out of finding ($-C c_\kappa \geq 0$). Smarter AI thins the layers above and displaces people into finding ($C c_q \leq 0$). Cost reductions therefore raise welfare unambiguously, whereas capability improvements are ambiguous. When congestion is severe the indirect effect dominates, the numerator of (23) turns negative, and the iso-utility curve slopes downward. An increase in capability then reduces welfare.

5.4.3 Transitions After Full Adoption

Finally, consider a coexistence region between two AI configurations. The two zero-profit conditions pin down p and U by themselves, so the differencing argument from Section 5.3

¹²Writing $B(\kappa) \equiv \min_a B_a(\kappa)$, the human configuration solves the same lower-hierarchy program at compute cost $1/h$ – its top layer’s first-order condition – so $p^* = 1 - B(1/h)$, while a top-entry firm breaks even at $q^A - B(\bar{\kappa})$. At $\bar{\kappa}$, $B(\bar{\kappa}) = B(1/h) - (1 - q^A) < B(1/h)$, hence $\bar{\kappa} < 1/h$. Cheaper compute at the lower-hierarchy’s top then weakly lowers every layer below that one (Fact A.4).

applies again, across the two configurations:

$$dU = \frac{\Delta\delta^+ dq^A - \Delta\delta^- d\kappa}{\Delta s_h}. \quad (25)$$

The numerator is the relative marginal value of the improvement in the two configurations. The denominator is their difference in humans per problem. Inside a coexistence region, utility must adjust to keep the two configurations equally profitable. The intuition is familiar from the partial adoption region (Proposition 4). If an AI improvement favors the configuration with a lower human intensity, then utility must fall to offset the differential gain, and conversely.

To continue, we assess what transitions look like after full AI adoption. By (16) a firm optimizes its upper- and lower-hierarchies separately, so a coexisting pair must tie in one optimization problem or the other.

Lemma 4 (Coexistence is one-sided) *Outside a set of (q^A, κ) of Lebesgue measure zero, two coexisting configurations differ on only one side of the AI: either their upper-hierarchies differ or their lower-hierarchies differ, but not both.*

Intuitively, the upper-hierarchy determines the value created once problems reach it, while the lower-hierarchy determines the cost of delivering them. For two AI configurations to coexist, their total profits must be equal. Such equality cannot come from simultaneous changes on both parts of the hierarchy, except in the knife-edge case. Thus coexisting configurations differ on one side of the AI only.

If two configurations share upper-hierarchies, then they value AI capability equally, $\Delta\delta^+ = 0$, so greater AI capability therefore has no effect on utility. Only compute cost matters, and its sign depends on whether humans are displaced into or out of solving (Δs_h). If the two configurations share lower-hierarchies, then they use the same amount of compute, $\Delta\delta^- = 0$, so lower compute cost therefore has no effect on utility. Only AI capability matters, the configuration benefiting more from a capability increase is the one whose upper-hierarchy uses fewer humans. Using these restrictions in (25) gives the following result:

Proposition 5 (Welfare at transitions) *Inside a coexistence region with the same upper-hierarchy for both firms, utility is independent of q^A and increasing in κ . Inside a coexistence region with the same lower-hierarchy for both firms, utility is independent of κ and decreasing in q^A .*

To illustrate the first result, consider an economy in which two AI-adopting firms share upper-hierarchies but differ in how problems are initially attempted. One firm puts AI in position $a = 0$, so that the AI processes every problem and requires one unit of compute per problem. The other firm deploys AI at $a = 1$, so that an untrained human worker attempts each problem first and refers it to the AI's layer. Although the worker solves no problems alone, the hierarchy uses the AI as a managerial input and reduces compute use from one unit to h units per problem.¹³ In this example, direct calculation shows that the firm with the AI at position 1 has $B_1(\kappa) = U + h\kappa$, while the firm with AI at position 0 has $B_0(\kappa) = \kappa$. Coexistence requires the two lower-hierarchies to have the same cost, so $U + h\kappa = \kappa$, or equivalently $U = (1 - h)\kappa$.

The worker is therefore paid exactly the value of the compute she saves. As compute becomes cheaper, that saving becomes less valuable, and equilibrium utility falls. By contrast, an increase in q^A affects the shared hierarchy above the slot in the same way under both configurations and therefore does not disturb the coexistence condition. Hence $\partial U / \partial \kappa = 1 - h > 0$ and $\partial U / \partial q^A = 0$.

Assembling the pieces, we see that the welfare effects of AI progress along a path of static equilibria may be non-monotone. The same form of progress can raise utility at one point and lower it at another. The organizing principle is displacement. An improvement can lower utility only when it displaces solvers into finding, and raises utility when it does not. Figure 5 provides an illustration of the welfare effects from a decrease in compute costs (left) and an increase in capability (right), and Figure 4 shows the configurations and human knowledge along the first of these paths.

5.5 The Role of Adjustments

The welfare effects discussed above rely on two types of adjustments: humans can retrain and switch occupations. We first ask what the introduction of AI does when neither type of adjustment is possible. We then discuss how these adjustments, when they do occur, are informative about the direction of welfare.

¹³This configuration pair is an equilibrium for a nonempty range of κ whenever training is sufficiently costly. More specifically, the following three conditions are sufficient: (i) $z'(0) \geq h\kappa$ ensures the human solver below AI has $q_0 = 0$, (ii) $h(1 - q^A)z'(q^A) \geq 1$ ensures any solver above AI costs more than her value added, and (iii) $p((1 - h)\kappa) \leq q^A - \kappa$ ensures that an all-human configuration cannot profitably enter. This configuration pair is an equilibrium for the parameters used in Figure 5 when $\kappa \in (0.24, 0.46)$.

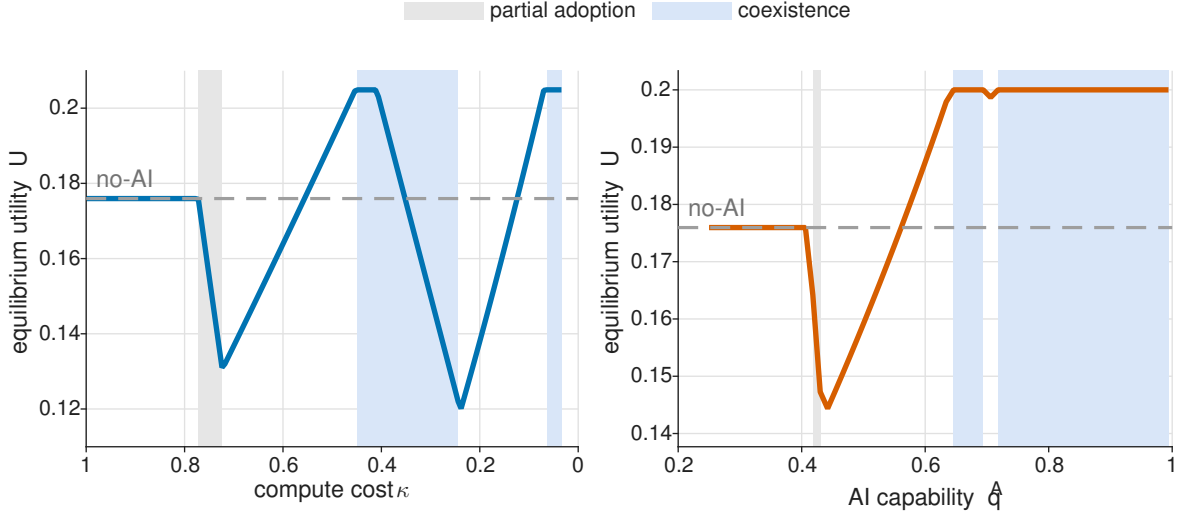


Figure 5: Equilibrium utility against the AI technology parameters with AI as a solver. Left: $U(\kappa)$ at $q^A = 0.7$. Right: $U(q^A)$ at $\kappa = 0.4$. Dashed line: no-AI utility. Shading marks partial adoption (grey) and coexistence of two AI configurations (blue). Log specification, $h = 0.5$, $t = 0.5$, $f = 1$, $\gamma = 0.75$.

AI without adjustment. Suppose adjustments are not possible, because they are costly or because the horizon is short. Fix the benchmark equilibrium and let AI arrive with human knowledge and occupations fixed, and training costs sunk. Each layer is then a type in fixed supply with its own wage w_l in place of $U^* + z(q_l^*)$, the number of problems is fixed, and firms may form any hierarchy from the existing types with or without AI. This frozen economy is the setting of [Ide and Talamàs \(2025\)](#) with one difference: problems are scarce in our economy whereas they are abundant in theirs. Appendix A.4 formalizes the economy and proves the claims that follow.

If $q^A < q_L^*$, AI is not adopted for any $\kappa > 0$. Firm hierarchies and the allocation of problems to solvers are the same as in the no-AI benchmark. Wages nevertheless adjust. A worker whom AI could replace cannot be paid more than the compute that would replace her, so every type with $q_l^* \leq q^A$ earns at most κ . The threat of adoption transfers income from the types AI could replace to the owners of problems, whom it cannot replace. In [Ide and Talamàs \(2025\)](#), by contrast, wages adjust only when AI is adopted, in which case output rises because a worker replaced by AI is free to produce. Here AI is not adopted, output is unchanged, and what the replaced types lose is a transfer to the finding sector.

When workers can retrain or switch occupations, a wage cannot fall below what the worker would earn after retraining or in finding. The threat of replacement lowers wages

only to this floor, and once compute is cheaper than a worker paid the floor wage, adoption begins. Adjustment therefore facilitates adoption. Knowledge and occupational adjustments also determine where the effects of AI are observed. In the frozen economy they are confined to wages and the price of problems, and firms are organized as in the benchmark. With adjustment, wages are $U + z(q_t)$, and the effects of AI appear in how firms organize and how workers train.

Organization and training as a welfare signal. Suppose a planner observes firms' organizational choices and the price of compute, but not AI capability. The pattern of firm organization and worker training is nevertheless informative about welfare. If adopting firms coexist with all-human firms and use fewer people per problem, every improvement lowers utility, and no further diagnosis is needed. If firms are replacing a layer of workers with AI, the economy is inside a lower-hierarchy coexistence region, in which the coexisting configurations share the upper hierarchy and differ below AI. Utility does not respond to capability and falls as compute becomes cheaper for as long as the replacement continues. If instead every firm uses the same configuration, cheaper compute raises welfare, and the ambiguous margin is greater AI capability. The workers below AI become the diagnostic. Their training responds only to utility and κ , so if they de-skill while compute has not become cheaper, welfare is decreasing.

6 Policy

Our economy features one primitive distortion. Too many humans enter finding because finders are paid their average product rather than their social marginal product. AI adoption can reduce welfare by reallocating labor through this distorted margin. What corrective policies improve welfare? We show that a tax on finding income corrects the finding margin and implements the first best. We then ask what a tax or subsidy on AI can accomplish when taxing finder income is infeasible.

We consider policies whose revenue is rebated lump-sum, T per capita. A tax τ on finding income reduces the finder's income to $(1 - \tau)f q \rho$, and a tax t_A on compute raises the cost firms face to $\kappa_e = \kappa + t_A$. In Proposition 6, $t_A = 0$, so $\kappa_e = \kappa$.

Proposition 6 (A finding tax decentralizes the first best) *With $\tau = \gamma$ and the revenue rebated lump-sum, the competitive equilibrium allocation maximizes net output over all feasible allocations, and U attains the planner's value.*

The intuition is straightforward if somewhat surprising. The efficient response to AI encroachment on the solving sector is to levy a tax on humans who seek refuge by becoming finders. The reasoning is familiar from other tragedy-of-the-commons settings. Here, a finder keeps the average product of her effort while contributing only the marginal product. The tax makes her retain the marginal product, which aligns the private and social returns to finding. With the entry margin into finding corrected, no other distortion remains.¹⁴ With the corrective tax in place, welfare rises along either dimension of AI progress and eliminates the gap between the private and social returns to improvements identified in Proposition 3.

Suppose instead that taxing finders is infeasible (perhaps for political reasons). A tax or subsidy on AI affects firms' adoption decisions directly and the finding entry margin indirectly. Under what conditions does it improve welfare?

Proposition 7 (A tax on AI) *Let the compute tax t_A be rebated lump-sum. Then at any equilibrium that varies differentiably in the tax,*¹⁵

$$\left. \frac{dU}{dt_A} \right|_{t_A=0} = \frac{dU}{d\kappa} + N_A = -\gamma\rho \frac{dQ}{d\kappa}, \quad (26)$$

where N_A is the aggregate number of AI agents and $dQ/d\kappa$ is the equilibrium finding response.

Starting from $t_A = 0$, a marginal increase in the tax on AI affects welfare only through its effect on the finding sector. The marginal tax therefore raises welfare if and only if it reduces Q .¹⁶ Note that $dU/d\kappa < 0$ does not imply that the tax is harmful. By (26), the tax reduces welfare only if $dU/d\kappa < -N_A$, that is, only if a cost reduction raises welfare by more than the resources it saves.

Whether a tax or subsidy on AI improves welfare depends on the equilibrium configuration. If a single representative firm exists, the tax moves Q through two opposing channels: it lowers the value of problems, which contracts finding effort and relieves congestion, but it also leads the firm to substitute trained workers for compute, which displaces humans into finding and exacerbates congestion. The sign is ambiguous in general. However, when firms

¹⁴The tax can equivalently be placed on the buyer side of the problem market: firms pay $p(1 + \tau_b)$ while finders receive p . At $\tau_b = \gamma/(1 - \gamma)$, the revenue, and the allocation are identical to those under the finding tax.

¹⁵Differentiability holds in the interior of the single firm and coexistence regions and fails only at the boundaries between them, where (26) holds for one-sided derivatives.

¹⁶When $t_A \neq 0$, both the response from the finding sector and how firms adopt AI contribute to the welfare effects of a tax on AI.

deploy AI at or near the bottom of the ladder, $a \in \{0, 1\}$, only the first channel operates and taxing AI raises welfare.

In a coexistence region there is no ambiguity. A marginal tax on AI raises welfare by (26) and Proposition 5. In a lower-hierarchy coexistence region, welfare is increasing in κ so taxing AI would be beneficial even if the tax revenue was burned. In the upper-hierarchy coexistence region, the tax raises welfare even though $dU/d\kappa = 0$. Cheaper compute neither helps nor harms humans, yet taxing it raises welfare. Absent a tax, the direct savings from lower compute cost are exactly offset by the loss from increased congestion. The tax on compute relieves congestion and rebates the proceeds to humans.

7 AI as a Finder

We now allow firms to deploy AI agents in finding, beginning with finding alone and then comparing adoption across the two roles. Problems found with AI sell on the same market as those found by humans. Finding differs from solving because there is no hierarchy. A finder's output is one-dimensional, so AI competes with human finders on the cost per unit of finding alone. The welfare effects are therefore starker than in solving. When humans and AI coexist in finding, every improvement to AI hurts, but once humans exit, every improvement helps.

7.1 AI Finders Alone

We build up AI-based finding in parallel to the construction of the human finding sector in Section 4.2, remembering that $\rho = pQ^{-\gamma}$ is the congestion-adjusted price of problems. Profit from one AI agent deployed in finding is $f q^A \rho - \kappa$, and free replication implies:

Lemma 5 (Adoption of AI in finding) *Firms adopt AI in finding if the no-AI economy has $\rho > \frac{\kappa}{f q^A}$. With positive AI deployment, we have $\rho = \frac{\kappa}{f q^A}$.*

This is simply a zero-profit condition. AI agents enter finding if they would make positive profit, and replicability expands their number until profits are driven to zero. This mechanism gives AI finders a conceptually simple role. They fix the congestion-adjusted price ρ at their own cost per problem discovered, $\kappa/(f q^A)$. As long as this price still attracts human finders, human and AI finders coexist.¹⁷

¹⁷The proof of Proposition 8 shows that AI and human finders coexist on an open set of (q^A, κ) .

Within this coexistence region, driving down the congestion-adjusted price of problems reduces welfare, so improvements in AI reduce utility:

Proposition 8 (Coexistence in finding lowers welfare) *If human and AI finders coexist, then any technological improvement (a decrease in κ or an increase in q^A) reduces utility, lowers the congestion-adjusted price ρ , and raises the problem price p and the quantity of problems Q . Welfare is also lower than in the corresponding equilibrium without AI.*

These results stem from AI-driven congestion: adding AI increases the supply of problems in a welfare-decreasing way.¹⁸ First, human-finder utility increases in the congestion-adjusted price of her output (Section 4.2). AI entry reduces that price (Lemma 5), and an improvement in AI reduces it further, since the improved AI finder is profitable at the old price and AI agents enter until it is not. Since human finders are active, they are indifferent between finding and solving, so utility is lower for the economy as a whole.

Second, lower utility for humans implies that firms can and do pay more for problems (see (7)). The congestion-adjusted price is nevertheless lower, because the quantity of problems is higher. Humans are displaced from finding, but not enough to offset the entry of AI agents, so congestion worsens. The increase in congestion allows both the price and quantity of found problems to rise, and the added congestion is what dissipates welfare.

Once humans have fully exited finding, lower problem-supply costs benefit human solvers:

Proposition 9 (Welfare gains when all finding uses AI) *(i) Humans exit finding once $\frac{\kappa}{q^A}$ falls below a threshold. (ii) Further improvements to AI weakly increase welfare: utility rises strictly except where two firm configurations coexist, where it is constant. (iii) AI improvements remain privately beneficial.*

As AI improves, the congestion-adjusted price of problems falls far enough that humans are driven out of finding. At that point the congestion externality is entirely AI-on-AI, so AI improvements do not increase duplication of human efforts. The over-entry continues, but it dissipates compute rather than human time, and human solvers benefit from cheaper problems. Figure 6 traces the resulting path of equilibrium utility.

¹⁸A bug-bounty program offers a partial illustration of the contribution of AI to finding problems, the coexistence regime, and the congestion externality. Curl paid posted rewards for vulnerabilities, which makes it a market for found problems. After AI entered the market, submissions dramatically increased and duplicate submissions, detailed in a way that manual search does not produce, became common. The program had to shut down and re-start under different terms. Curl’s Daniel Stenberg reports a similar pattern at more than two dozen other projects. From <https://daniel.haxx.se/blog/2026/01/26/the-end-of-the-curl-bug-bounty/> and <https://daniel.haxx.se/blog/2026/04/22/high-quality-chaos/>.

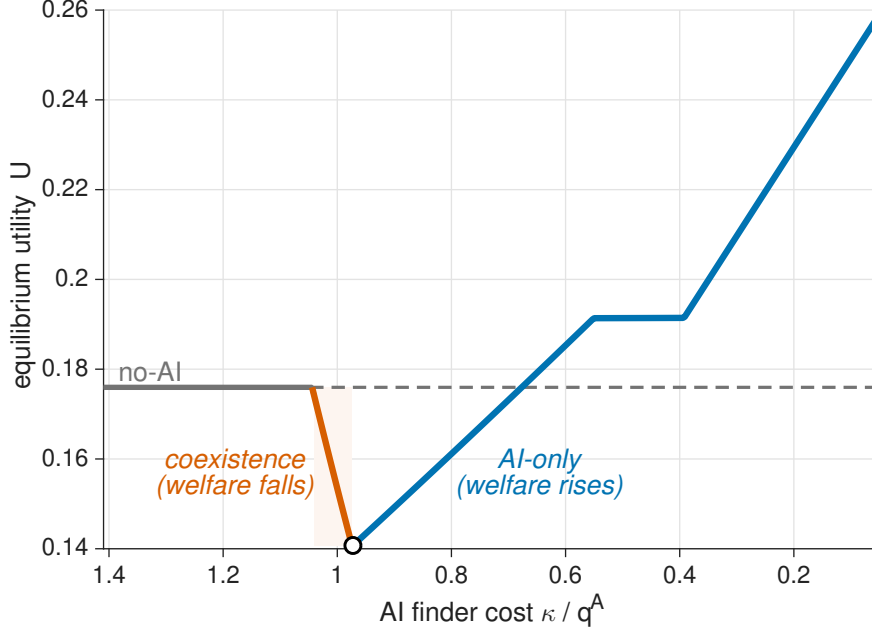


Figure 6: Equilibrium utility against the AI finder cost per unit of capability, κ/q^A (a rightward movement is an improvement), with AI finding only. Utility falls while human finders coexist with AI (shaded region) and rises once finding is AI-only. Along the second flat segment two AI firm configurations coexist. Log specification, $h = 0.5$, $t = 0.5$, $f = 1$, $\gamma = 0.75$.

The contrast with solving is sharp. In solving, AI improvements can have ambiguous effects after full adoption. Smarter AI can lower welfare through the reorganization and displacement it induces. In finding there is no ladder to reorganize, so once humans have exited finding, an improvement only makes problems cheaper.

7.2 AI Adoption in Finding and Solving

We consider whether AI is adopted in the finding or solving sectors. This is a problem of the break-even quality and cost of AI, and the maximum cost that induces adoption for the economy as a whole is simply the maximum of the cost in each sector. Thus, the adoption frontier (κ as a function of q^A) is the maximum of the solver frontier from Section 5.2 and the finder frontier from Section 7.1.

Proposition 10 (Finder efficiency and joint AI adoption) *Without AI, the human finder's knowledge minimizes the cost per unit of finding input: $q_F = \arg \min \frac{U_F + z(q)}{q}$.*

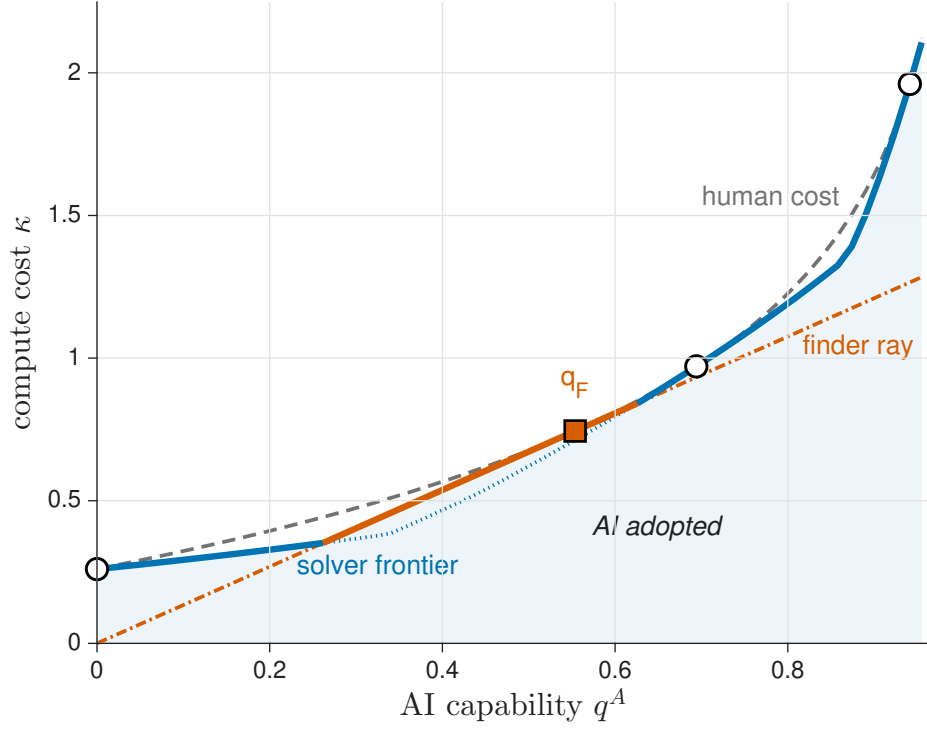


Figure 7: The joint adoption frontier (solid): the upper envelope of the solver frontier $\bar{\kappa}(q^A)$ and the finder ray, drawn dotted and dash-dotted respectively where each is not the maximum. The ray is tangent to the human cost curve (dashed) at the finder knowledge q_F (square). Circles mark the rungs of the solver ladder. Log specification, $h = 0.32$, $t = 0.6$, $f = 1$, $\gamma = 0.8$.

The AI adoption frontier in finding is $\frac{\kappa}{q^A} = \frac{U_F + z(q_F)}{q_F} = z'(q_F)$, and the joint AI adoption frontier in finding or solving is $\max\{\bar{\kappa}(q^A), q^A z'(q_F)\}$, with $\bar{\kappa}(q^A)$ defined in Proposition 2.

The first part explains why the finder frontier is a ray. AI finding scales with the number of AI agents deployed, so what matters for finding is the cost per unit of finding input, κ/q^A . Human finders choose q_F to minimize the same cost per unit. Their total cost is $U_F + z(q)$ and their finding input is q .

The human finder is protected by efficiency selection. The finder has already chosen the knowledge level that minimizes cost per unit of finding input, so an AI with different capability must be sufficiently cheap per unit of finding input to be adopted as a finder. Solving is protected by a different force: the knowledge-mismatch gap. The solver frontier $\bar{\kappa}(q^A)$ lies below the human cost curve except at the rungs of the benchmark ladder, because a capability level that does not match a rung requires a different surrounding hierarchy. Joint adoption is therefore a comparison of two protection mechanisms: knowledge mismatch

in solving and cost-per-input efficiency in finding. Figure 7 illustrates the joint adoption frontier.

This comparison has two implications. First, exposure is local in knowledge. Capability levels close to a benchmark solver rung face little mismatch in solving, while capability levels close to the human finder’s knowledge face the strongest comparison in finding. Second, the finder can be protected even when an equally knowledgeable solver is exposed. If the human finder and a solver rung have the same knowledge level, the finder frontier is the tangent ray to the human cost curve at that point, while the solver frontier is convex and touches the same curve there. The two frontiers agree at the shared knowledge level, but away from it the solver frontier is crossed first. For finder adoption to occur first, the AI’s capability must be sufficiently close to the finder’s knowledge and sufficiently efficient per unit of finding input.

The frontier comparison above is evaluated at the no-AI benchmark. Once AI is adopted in one role, utility changes and the other frontier moves.

Proposition 11 (Adoption in one role lowers the other frontier) *(i) Suppose all-human and adopting firms coexist as solvers and $s_h < s$. Then $U < U^*$, and the finder frontier lies below its no-AI level. (ii) Suppose instead human and AI finders coexist. Then $\rho < \rho^*$ and $U < U^*$, and if the configuration attaining the solver frontier has $s_h < s$ at each utility in $[U, U^*]$, then $\bar{\kappa}(q^A)$ lies below its no-AI level.*

In both roles, the AI replaces a human input whose cost includes $U + z(q)$ with one whose cost is κ , so lower utility leaves fewer rents to capture. Adoption in one role thus makes adoption in the other harder whenever the adopting configuration uses fewer humans per problem than the all-human one. Lemma 2 gives sufficient conditions for this at the frontier.

8 Conclusion

We study AI in a knowledge economy organized around two activities: finding problems and solving them. Solving takes place within firms whose knowledge hierarchies are staffed by endogenously trained workers, while finding is subject to a classical congestion externality. This externality creates a wedge between the productivity gains generated inside firms and the broader effects of AI on training, organizational structure, occupational reallocation, and welfare.

Three extensions seem most promising. First, the model is static, a collection of snapshots along a path of AI improvement. Workers may invest in knowledge before future AI capability is known, firms may delay reorganization because restructuring is costly, and displacement today may alter the future supply of managers. Adjustment costs and durable human capital would permit a study of anticipatory training, and of whether gradual adoption reflects technology, organizational inertia, or strategic delay. Second, the direction of AI innovation and the creation of problems could be endogenized. Developers may choose whether to improve capability in finding, solving, or coordinating human knowledge, and problems that differ in value and appropriability would allow a richer account of congestion, in which additional finding duplicates discoveries or exhausts a finite stock. Finally, our results suggest an empirical agenda centered on organizational adjustment rather than employment counts: the number and spacing of managerial layers, training above and below the position at which a firm deploys AI, and the movement of workers between solving and finding. These margins distinguish displacement from reorganization, and identify when privately valuable adoption raises productivity at the expense of welfare.

References

- Acemoglu, Daron and Pascual Restrepo**, “The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment,” *American Economic Review*, 2018, 108 (6), 1488–1542.
- **and** —, “Artificial Intelligence, Automation, and Work,” in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019, pp. 197–236.
- , **Dingwen Kong, and Asuman Ozdaglar**, “AI, Human Cognition and Knowledge Collapse,” Working Paper 34910, National Bureau of Economic Research 2026.
- Bonney, Kathryn, Cory Breaux, Emin Dinlersoz, Lucia Foster, John Haltiwanger, and Aditya Pande**, “The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks,” CES Working Paper 26-25, U.S. Census Bureau, Center for Economic Studies April 2026.
- Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen**, “Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence,” Working Paper, Stanford Digital Economy Lab November 2025.

- , **Danielle Li**, and **Lindsey R. Raymond**, “Generative AI at Work,” NBER Working Paper 31161, National Bureau of Economic Research April 2023.
- Butters, Gerard R.**, “Equilibrium Distributions of Sales and Advertising Prices,” *Review of Economic Studies*, 1977, 44 (3), 465–491.
- Cai, Zhifeng**, “The AI Skill Trap,” 2026. Working paper, Rutgers University.
- , “Generative AI and Data Quality: Implications for Productivity, Labor Displacement, and Policy,” *Review of Financial Studies*, 2026. Advance article.
- Chaigneau, Pierre and Evan Jo**, “Artificial Intelligence Investments and Expertise Erosion,” 2025. Working paper, December 2025.
- Chen, Alvin**, “The Organizational Cost of Artificial Intelligence,” 2026. Working paper, Stockholm School of Economics.
- Chen, Joanne and Brandon Yueyang Han**, “When Corporate AI Adoption Backfires,” 2026. Working paper.
- , – , and **Hongda Zhong**, “Human Skills in the Age of AI,” 2026. Working paper.
- Cullen, Zoë, Danielle Li, and Shengwu Li**, “Labor as Capital: AI and the Ownership of Expertise,” 2026. Working paper, March 2026.
- Dell’Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz, Katherine C. Kellogg, Saran Rajendran, Lisa Kraye, François Candelon, and Karim R. Lakhani**, “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality,” *Organization Science*, 2026, 37 (2), 403–423.
- Ferreira, Daniel and Jin Li**, “Artificial intelligence in the boardroom,” 2026. Working paper.
- Fuchs, William, Luis Garicano, and Luis Rayo**, “Optimal Contracting and the Organization of Knowledge,” *Review of Economic Studies*, 2015, 82 (2), 632–658.
- Gans, Joshua S. and Scott Stern**, “The product market and the market for “ideas”: commercialization strategies for technology entrepreneurs,” *Research Policy*, 2003, 32 (2), 333–350.
- Garicano, Luis**, “Hierarchies and the Organization of Knowledge in Production,” *Journal of Political Economy*, 2000, 108 (5), 874–904.
- and **Esteban Rossi-Hansberg**, “Organization and Inequality in a Knowledge Economy,” *Quarterly Journal of Economics*, 2006, 121 (4), 1383–1435.

- **and Luis Rayo**, “Training in the Age of AI: A Theory of Career Viability,” 2026. Working paper, March 2026.
- **and Tano Santos**, “Referrals,” *American Economic Review*, June 2004, *94* (3), 499–525.
- Hopenhayn, Hugo and Francesco Squintani**, “On the Direction of Innovation,” *Journal of Political Economy*, 2021, *129* (7), 1991–2022.
- Humlum, Anders and Emilie Vestergaard**, “Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI,” NBER Working Paper 33777, National Bureau of Economic Research May 2025. Revised March 2026.
- Ide, Enrique**, “Automation, AI, and the Intergenerational Transmission of Knowledge,” 2026. Working paper, IESE Business School.
- **and Eduard Talamàs**, “Artificial Intelligence in the Knowledge Economy,” *Journal of Political Economy*, 2025, *133* (12). DOI: 10.1086/737233.
- Jones, Charles I and John C Williams**, “Too much of a good thing? The economics of investment in R&D,” *Journal of Economic Growth*, 2000, *5* (1), 65–85.
- Kortum, Samuel S.**, “Research, Patenting, and Technological Change,” *Econometrica*, 1997, *65* (6), 1389–1419.
- Loury, Glenn C.**, “Market Structure and Innovation,” *Quarterly Journal of Economics*, 1979, *93* (3), 395–410.
- Reinganum, Jennifer F.**, “A Dynamic Game of R&D: Patent Protection and Competitive Behavior,” *Econometrica*, 1982, *50* (3), 671–688.
- Stokey, Nancy L**, “R&D and economic growth,” *Review of economic studies*, 1995, *62* (3), 469–489.

A Appendix

A.1 Preliminaries

The appendix keeps shorter names for the objects of Section 5. A *chain* is an upper-hierarchy and a *block* is a lower-hierarchy. A *rung* is the knowledge of a layer, and a *ladder* is the ordered rungs $q_0 < \dots < q_L$ of a configuration.

Fact A.1 By (10), *finder utility is $\phi(q_F)$ with $\phi(q) \equiv z'(q)q - z(q)$, strictly increasing, $\phi(0) = 0$, and $\phi(q) \rightarrow \infty$ as $q \rightarrow 1$. Write $q_F(U) \equiv \phi^{-1}(U)$ and $\rho(U) \equiv z'(q_F(U))/f$, continuous and strictly increasing, with $q_F(0) = 0$ and $\rho(0) = z'(0)/f$.*

Proof. Differentiating, $\phi'(q) = q z''(q) > 0$, and for $q > \frac{1}{2}$, $\phi(q) - \phi(\frac{1}{2}) = \int_{1/2}^q s z''(s) ds \geq \frac{1}{2} [z'(q) - z'(\frac{1}{2})] \rightarrow \infty$ as $q \rightarrow 1$. ■

For a configuration and ladder $c \equiv (L, q_0, \dots, q_L)$, the price at which c breaks even at utility U is

$$P_c(U) = q_L - [U + z(q_0)] - h \sum_{l=1}^L (1 - q_{l-1}) [U + z(q_l)],$$

which is affine in U with slope $-s^c \leq -1$. The break-even price of the economy is $p(U) \equiv \sup_c P_c(U)$, and profit per problem attempted at price p is $P_c(U) - p$. A configuration is optimal at U if $P_c(U) = p(U)$. The chain value V and the block cost B_a are those of (14)–(16). Among optimal chains above x , let $S(x)$ denote the smallest human mass, let $m^*(x)$ be its depth, and let $q_1(x)$ be its first rung. Then $S(x)$ satisfies

$$S(x) = \begin{cases} h(1 - x) + S(q_1(x)) & m^*(x) \geq 1, \\ 0 & m^*(x) = 0. \end{cases} \quad (27)$$

The recursion holds by Lemma A.1(ii). The best configuration with the AI at position a breaks even at $p_a(\kappa) = V(q^A) - B_a(\kappa)$. Write $B_a(\tilde{\kappa}; \bar{q})$ for the block cost with $\tilde{\kappa}$ in place of κ and rungs below $\bar{q}$ in place of q^A , so that $B_a(\kappa) = B_a(\kappa; q^A)$. A human configuration is a bottom worker of knowledge q , of mass one and cost $U + z(q)$, under a chain, so

$$p(U) = \max_{q \in [0,1]} [V(q) - (U + z(q))]. \quad (28)$$

In an optimal configuration at U , the rungs above any layer form an optimal chain above that layer, and $V(q_L) = q_L$ at the top rung. At the no-AI equilibrium, q_0^* solves (28).

Fact A.2 *Any configuration that cannot be improved by deleting one of its layers satisfies the following three conditions:*

(i) If its top layer is human,

$$h[U + z(q_L)] \leq \frac{q_L - q_{L-1}}{1 - q_{L-1}} \leq 1, \quad (29)$$

and every rung lies weakly below $\bar{q}_U < 1$, defined by $z(\bar{q}_U) = 1/h - U$.

(ii) If a human layer sits directly below AI, $U + z(q_{a-1}) \leq \kappa$.

(iii) If all rungs lie below some $\hat{q} < 1$ and the wage bill is at most $\bar{W}$ then the depth is at most $\bar{W}/[h(1 - \hat{q})U]$.

Proof. For (i), deleting the top layer lowers output by $q_L - q_{L-1}$ and saves the wage bill $h(1 - q_{L-1})[U + z(q_L)]$. The saving is at most the lost output, which is the first inequality in (29); the second holds because $q_L \leq 1$. For (ii), at $a \geq 2$, deleting the human layer directly below the AI saves its wage $h(1 - q_{a-2})[U + z(q_{a-1})]$ and raises the compute bill by $h(q_{a-1} - q_{a-2})\kappa$. The saving is at most the extra compute, so $U + z(q_{a-1}) \leq \kappa(q_{a-1} - q_{a-2})/(1 - q_{a-2}) \leq \kappa$. At $a = 1$ the deleted layer is the bottom worker and the same comparison gives $U + z(q_0) \leq \kappa[1 - h(1 - q_0)] \leq \kappa$. For (iii), each managerial layer's wage bill is at least $h(1 - \hat{q})U$, so $L \cdot h(1 - \hat{q})U \leq \bar{W}$. ■

A chain of value at least x has wage bill at most one, so by Fact A.2(i) and (iii) its rungs lie below $\bar{q}_U$ and its depth is bounded. The maximum in (14) is therefore over a compact set, and an optimal chain exists at every x .

Fact A.3 Let $\hat{x}$ maximize $\alpha(\cdot) - \mu\beta(\cdot)$ and $\hat{x}'$ maximize $\alpha(\cdot) - \mu'\beta(\cdot)$, with $\mu < \mu'$ and each maximizer feasible in the other problem. Then $\beta(\hat{x}') \leq \beta(\hat{x})$.

Proof. Adding the optimality inequalities $\alpha(\hat{x}) - \mu\beta(\hat{x}) \geq \alpha(\hat{x}') - \mu\beta(\hat{x}')$ and $\alpha(\hat{x}') - \mu'\beta(\hat{x}') \geq \alpha(\hat{x}) - \mu'\beta(\hat{x})$ cancels the α terms and leaves $(\mu' - \mu)[\beta(\hat{x}) - \beta(\hat{x}')] \geq 0$. ■

Fact A.4 At fixed prices, Π_L , V , and $-B_a$ have increasing differences in $(q_0, \dots, q_L; U)$, $(q_{a+1}, \dots, q_L; q^A, U)$, and $(q_0, \dots, q_{a-1}; \kappa, U)$, respectively. By Topkis's theorem, every rung weakly rises with U , the chain's rungs weakly rise with q^A , and the block's rungs weakly rise with κ .

Proof. Each objective is a sum of terms $-h(1 - q_{l-1})[U + z(q_l)]$, plus an output and a wage. Each cross-partial in the listed pairs is nonnegative. ■

Fact A.5 At utility U , a configuration with the AI at position a , using s_A units of compute per problem at cost κ , breaks even at a price of at most $p(U) + s_A[(U + z(q^A)) - \kappa]$.

Proof. Replacing the AI with humans of knowledge q^A preserves every ordering constraint and every mass. It changes the cost from κ to $U + z(q^A)$ per unit, and yields a human configuration, whose break-even price is at most $p(U)$. ■

Fact A.6 Fix $U > 0$ with human finders, and let $\tilde{p}(s)$ be the price of problems at which the market clears when solving uses s people per problem. Then $\tilde{p}$ is decreasing in s , with $d\tilde{p}/ds = -C$ for the congestion factor C of (22), and strictly increasing in U .

Proof. At fixed U , $q_F(U)$ and $\rho(U)$ are fixed, so clearing, $sQ^{1-\gamma} + Q/(fq_F) = 1$, determines Q uniquely, its left side being strictly increasing in Q , and the congestion identity gives $\tilde{p} = \rho Q^\gamma$. For the derivative in s , take logarithmic differentials of the clearing condition:

$$Q^{1-\gamma} ds + (1 - \gamma s Q^{1-\gamma}) d \ln Q = 0,$$

where the coefficient of $d \ln Q$ is $(1 - \gamma)sQ^{1-\gamma} + Q/(fq_F)$ after substituting $Q/(fq_F) = 1 - sQ^{1-\gamma}$, and it is positive because $sQ^{1-\gamma} \leq 1$. Since $d \ln \tilde{p} = \gamma d \ln Q$,

$$\frac{d\tilde{p}}{ds} = -\frac{\gamma \tilde{p} Q^{1-\gamma}}{1 - \gamma s Q^{1-\gamma}} = -C.$$

For monotonicity in U , a rise in U raises $q_F(U)$, so the left side of the clearing condition falls and Q rises. Since $\rho(U)$ also rises, so does $\tilde{p} = \rho Q^\gamma$. ■

Lemma A.1 (i) V is convex and nondecreasing. (ii) At any rung of an optimal chain, V is differentiable, and the optimal chain above the rung is unique. (iii) If $V(y) = y$, then $V(y') = y'$ for every $y' \geq y$, and any rung $y' \geq y$ of an optimal chain is the top rung of that chain. (iv) If $x < x'$, the first rung of any nonempty optimal chain above x is weakly below the first rung of any nonempty optimal chain above x' .

Proof. Given $q = (q_1, \dots, q_j)$ with $q_i < q_{i+1}$, let $A(x; q_1, \dots, q_j) = q_j - h(1 - x)w(q_1) - \sum_{i=2}^j h(1 - q_{i-1})w(q_i)$, which is affine in x with slope $h w(q_1) > 0$, and $A(x; \emptyset) = x$. By definition, $V(x)$ is the maximum of $A(x; q)$ over q with $q_1 > x$. A layer $q_1 \leq x$ adds no output, costs its wage, and raises the mass of the layer above from $h(1 - x)$ to $h(1 - q_1)$, so deleting it raises $A(x; q)$ and the constraint $q_1 > x$ is slack. Hence V is the maximum of $A(x; q)$ over all q , $A(\cdot; \emptyset)$ included. Each is affine with positive slope, so V is convex and nondecreasing, which is (i).

Two facts will be useful for establishing (ii) and (iii). First, the slope of any optimal chain q above x , namely $h w(q_1)$, or 1 if q is empty, is a subgradient of V at x .¹⁹ Second, the top rung of an optimal chain q satisfies $h w(q_j) < 1$ by Fact A.2(i), and w is increasing, so $h w(q_1) < 1$.

For (ii), let $\hat{q}$ be a rung of an optimal chain and $\hat{x}$ the knowledge beneath it. The rungs above $\hat{q}$ form an optimal chain above $\hat{q}$, so the chain's value is its value up to $\hat{x}$, less $h(1 - \hat{x})w(\hat{q})$, plus $V(\hat{q})$. Optimality of the chain therefore makes $\hat{q}$ a maximizer of $V(y) - h(1 - \hat{x})w(y)$ over $y \in (\hat{x}, 1)$. Let $\lambda \in \partial V(\hat{q})$, where the subdifferential $\partial V(x)$ is the set of subgradients of V at x , a closed interval that is a single point exactly when V is

¹⁹A subgradient of V at x is any slope λ with $V(y) \geq V(x) + \lambda(y - x)$ for all y .

differentiable at x . Optimality of $\hat{q}$ gives $V(y) - V(\hat{q}) \leq h(1 - \hat{x})[w(y) - w(\hat{q})]$ for y near $\hat{q}$, and the subgradient inequality bounds the same difference below by $\lambda(y - \hat{q})$. As $y \downarrow \hat{q}$ and $y \uparrow \hat{q}$, the two bounds give $\lambda = h(1 - \hat{x})w'(\hat{q})$. The subdifferential is therefore a singleton, and V is differentiable at $\hat{q}$. Every optimal chain above $\hat{q}$ has slope $V'(\hat{q})$. Distinct first rungs have distinct slopes, and the empty chain's slope 1 exceeds $hw(q_1)$ for every nonempty optimal chain q , so the optimal chain above $\hat{q}$ has a unique first rung. Applying this at each rung in turn (depth is finite), the optimal chain above $\hat{q}$ is unique.

For (iii), $A(\cdot; \emptyset)$ equals V at y , so 1 is a subgradient of V at y . Every subgradient at $y' > y$ is then at least 1. An optimal chain above y' with first rung q_1 would give the subgradient $hw(q_1) < 1$, so $V(y') = y'$. At a rung $y' \geq y$ of an optimal chain, the optimal chain above y' is unique by (ii), and $V(y') = y'$, so the chain has no rung above y' .

For (iv), the first rung of an optimal chain above x maximizes $V(y) - h(1 - x)w(y)$ over $y > x$, and the first rung of an optimal chain above x' maximizes $V(y) - h(1 - x')w(y)$ over $y > x'$, with $h(1 - x') < h(1 - x)$. Either the first is at most x' , in which case it is below the second, or both are feasible in both problems and Fact A.3 gives the order. ■

Lemma A.2 (i) If $x \leq x'$ and the chain above x' is nonempty, then the chain above x is nonempty and $q_1(x) \leq q_1(x')$. (ii) If $\tilde{q}$ solves (28) and $x < \tilde{q}$, then every nonempty optimal chain above x has first rung at least $\tilde{q}$. (iii) S is nonincreasing.

Proof. For (i), if $V(x') > x'$, then $V(x) > x$ for $x < x'$ by Lemma A.1(iii). An empty chain has zero mass, so $m^*(x') \geq 1$ implies $V(x') > x'$; hence $m^*(x) \geq 1$. If both chains are nonempty and $x < x'$, Lemma A.1(iv) gives $q_1(x) \leq q_1(x')$. For (ii), let q_1 be the first rung of an optimal chain above x . Then $\tilde{q}$ maximizes $V(y) - w(y)$, while q_1 maximizes $V(y) - h(1 - x)w(y)$ with $h(1 - x) < 1$. Since $\tilde{q} > x$ is feasible above x and q_1 is feasible in (28), Fact A.3 gives $q_1 \geq \tilde{q}$.

For (iii), we show by induction on n that $S(x') \leq S(x)$ whenever $x \leq x'$ and both optimal chains have depth at most n . If $m^*(x') = 0$, then $S(x') = 0 \leq S(x)$. Otherwise $m^*(x) \geq 1$ and $q_1(x) \leq q_1(x')$ by (i), and (27) gives $S(x) = h(1 - x) + S(q_1(x))$ and $S(x') = h(1 - x') + S(q_1(x'))$. The chains above $q_1(x)$ and $q_1(x')$ are optimal chains of depth at most $n - 1$, so $S(q_1(x')) \leq S(q_1(x))$ by the inductive hypothesis, and $h(1 - x') \leq h(1 - x)$. ■

Lemma A.3 Fix $U > 0$. Optimal configurations have at most two depths, and they are consecutive. Any two optimal configurations have distinct intensities, and when their depths differ the deeper configuration has a larger intensity. Two optimal configurations with the same positive top rung coincide.

Proof. For the first statement, suppose configurations $q = (q_0, \dots, q_L)$ and $q' = (q'_0, \dots, q'_{L'})$ are optimal at U with $L' \geq L + 2$. The rung q'_{L+1} has a rung above it, so $q'_{L+1} < q_L$ by Lemma A.1(iii), since $V(q_L) = q_L$. Let k be the smallest index with $q'_{k+1} \leq q_k$. For $k \geq 1$,

minimality gives $q_{k-1} < q'_k$, so Lemma A.1(iv) gives $q_k \leq q'_{k+1}$. For $k = 0$, $q'_0 < q'_1 \leq q_0$, and q_0 solves (28), so Lemma A.2(ii) gives $q'_1 \geq q_0$. In either case $q'_{k+1} = q_k$, and $k < L$ because $q'_{L+1} < q_L$. The common rung belongs to both optimal chains, so the optimal chain above it is unique (Lemma A.1(ii)). The rungs of q above q_k and the rungs of q' above q'_{k+1} are both optimal chains above it, with depths $L - k \geq 1$ and $L' - k - 1 \geq L - k + 1$, a contradiction.

For the second statement, let $q = (q_0, \dots, q_L)$ and $q' = (q'_0, \dots, q'_{L'})$, $L \leq L'$, be distinct optimal configurations at U , and let j be the first index at which they differ. Then $j \leq 1$, since at $q_0 = q'_0$ and $q_1 = q'_1$ the optimal chain above the common rung q_1 is unique (Lemma A.1(ii)), and the configurations coincide. We use one further fact twice. Two points $x < x'$ with $V(x) = x$ and $V(x') = x'$ cannot both maximize $V(\cdot) - \mu w(\cdot)$ for a constant $\mu > 0$, since $V(y) \geq y$ for all y , so both would maximize the strictly concave $y - \mu w(y)$.

Suppose $L < L'$, $L \geq j$, and $q_j < q'_j$. Then $q_l \leq q'_l$ for $j \leq l \leq L$ by Lemma A.1(iv). At $q_l = q'_l$ the optimal chains above q_l coincide (Lemma A.1(ii)), so $L = L'$. Hence $q_L < q'_L$. Since $V(q_L) = q_L$ and $q'_L > q_L$, Lemma A.1(iii) gives $V(q'_L) = q'_L$. For $L \geq 1$, q'_L is a rung of the chain above q'_0 , so it is that chain's top rung and $L' = L$, contradicting $L < L'$. For $L = 0$, $q_0 < q'_0$ both maximize (28) with $V(q_0) = q_0$ and $V(q'_0) = q'_0$, which the fact above excludes. Hence $q'_j < q_j$ when $L < L'$ and $L \geq j$. At equal depths, we label the configurations so that $q'_j < q_j$. Then $q'_l \leq q_l$ for $l \leq L$ (Lemma A.1(ii),(iv)), and

$$s^{q'} - s^q = h \sum_{l=j+1}^L (q_{l-1} - q'_{l-1}) + h \sum_{l=L+1}^{L'} (1 - q'_{l-1}).$$

The first sum contains $q_j - q'_j > 0$ when $L > j$, and the second is positive when $L < L'$, so $s^{q'} > s^q$ unless $L = L' = j$. In the remaining case $L = L' = j$, the top rungs q_j and q'_j satisfy $V(q_j) = q_j$ and $V(q'_j) = q'_j$ and both maximize $V(\cdot) - \mu w(\cdot)$ with $\mu = 1$ at $j = 0$ and $\mu = h(1 - q_0)$ at $j = 1$, which the fact above excludes.

For the third statement, at a positive rung r of an optimal configuration the first-order condition is $\nu w'(r) = h w(r^+)$, where r^+ is the rung above r , with $h w(r^+)$ read as 1 when r is the top rung, and where $\nu = 1$ if $r = q_0$ and $\nu = h(1 - r^-)$ if r^- is the rung beneath it. Since $h(1 - r^-) \leq h < 1$, the condition determines from r and r^+ whether r is the bottom rung and, if not, which rung is beneath it. Two optimal configurations with the same positive top rung therefore have the same rung beneath it, and so on down to the bottom. ■

Lemma A.4 *Suppose two distinct optimal chains above the same knowledge x have depths $m < m'$. Then (i) the deeper chain's first rung is strictly lower, so δ^+ is strictly smaller for the deeper chain; (ii) the deeper chain has strictly larger human mass.*

Proof. Write S and S' for the human masses of the two chains. If $m = 0$, then $S = 0 < S'$, and $\delta^+ = 1$ for the empty chain exceeds $h w(q'_1) \leq h w(q'_{m'}) < 1$ by Fact A.2(i). Let $m \geq 1$. The first rungs differ, since at $q_1 = q'_1$ the optimal chain above this common rung is unique

(Lemma A.1(ii)) and the depths would agree. Whichever chain has the lower first rung is lower at every common index k : from $q_{k-1} < q'_{k-1}$, Lemma A.1(iv) gives $q_k \leq q'_k$, and at $q_k = q'_k$ the chains would coincide from index k on (Lemma A.1(ii)), so their depths would agree; likewise with q and q' exchanged.

For (i), suppose instead $q'_1 > q_1$. Then $q_k < q'_k$ for $k \leq m$, and $V(q_m) = q_m$, so the rung $q'_m > q_m$ is the top rung of the deep chain by Lemma A.1(iii), contradicting $m < m'$. Hence $q'_1 < q_1$, and $hw(q'_1) < hw(q_1)$ since w is increasing.

For (ii), from the rung comparison above and (i), we have $q'_k < q_k$ for $k \leq m$. The masses share the term $h(1-x)$, so $S' - S = h \sum_{k=1}^{m-1} (q_k - q'_k) + h \sum_{k=m}^{m'-1} (1 - q'_k)$. The first sum is nonnegative, and the second has $m' - m$ terms, each at least $h(1 - \bar{q}_U) > 0$, since every rung lies below $\bar{q}_U$ (Fact A.2(i)). ■

Lemma A.5 *Suppose two distinct blocks of depths $a < a'$ both cost $\min_{i \geq 0} B_i(\kappa)$. Then (i) the deeper block's top rung is strictly higher, so δ^- is strictly smaller; (ii) the deeper block has strictly larger human mass.*

Proof. If $a = 0$, the shallow block is empty, with $\delta^- = 1$ and no human mass, while the deep block has $\delta^- = h(1 - q'_{a'-1}) < 1$ and positive mass. Let $a \geq 1$.

Dividing (15) by $h\kappa$ gives $\tilde{w}(q_0) + h \sum_{l=1}^{a-1} (1 - q_{l-1})\tilde{w}(q_l) + (1 - q_{a-1})$ with $\tilde{w} = w/(h\kappa)$, which is one minus the break-even price of the configuration $(q_0, \dots, q_{a-1})$ when wages are $\tilde{w}$ and output is q_{a-1} . An optimal block is therefore an optimal configuration for the wage function $\tilde{w}$, with rungs below q^A . Lemmas A.1–A.3 use only that the wage function is positive, increasing, strictly convex, and differentiable, so they hold for blocks. In particular, the rungs above any layer of an optimal block form an optimal chain above it, and part (ii) is the second statement of Lemma A.3.

For (i), suppose $q'_{a'-1} \leq q_{a-1}$. Compare the blocks level by level from the top, the rung q_{a-1-i} of the shallow block with the rung $q'_{a'-1-i}$ of the deep block. If at some level $i \leq a-1$ the deep block's rung is weakly above the shallow block's, so is the rung one level up: by Lemma A.1(iv) if it is strictly above, and by Lemma A.1(ii) if the two rungs are equal and positive, while if they are equal and zero both are bottom rungs and $a = a'$. Repeating this up to the top gives $q'_{a'-1} \geq q_{a-1}$, so the top rungs are equal. If they are zero, each block is the single layer at zero and $a = a'$. If they are positive, the blocks coincide by the third statement of Lemma A.3, and again $a = a'$. Hence the deep block's rung is strictly below the shallow block's at every level $i \leq a-1$.

At level $a-1$ the shallow block's rung is its bottom worker q_0 , and the deep block's rung $q'_{a'-a}$ lies strictly below it. Since $a' > a$, the rung $q'_{a'-a}$ has a rung beneath it, $q'_{a'-a-1} < q_0$, and the rungs of the deep block from $q'_{a'-a}$ upward form a nonempty optimal chain above $q'_{a'-a-1}$. Since q_0 solves (28), Lemma A.2(ii) gives $q'_{a'-a} \geq q_0$, a contradiction. Hence $q'_{a'-1} > q_{a-1}$. ■

A.2 Proofs for Section 4

Proof of Lemma 1. The break-even price $p(U)$ is a supremum of functions affine in U , hence convex and strictly decreasing. U_S is its inverse, and the inverse of a strictly decreasing

convex function is strictly decreasing and convex, so $U_S(p)$ is convex; by convexity its slope $-1/s$ is nondecreasing in p , so s is weakly increasing in p . For the last claim, by Fact A.4 every q_l is weakly increasing in U ; since the price enters no first-order condition and U_S is decreasing, every q_l is weakly decreasing in p . ■

Proof of Theorem 1.

Step 1. By (9) and Fact A.1, condition (i) of Definition 1 with finder utility U holds if and only if $q_F = q_F(U)$ and $\rho = \rho(U)$.

Step 2. The break-even price p is the inverse of U_S , so by Lemma 1 and its proof, it is convex, continuous, and strictly decreasing. Under Assumption 1, $p(0) \geq \max_q [q - hz(q)] > 0$; every configuration has output at most one and wage bill at least U , so $P_c(U) \leq 1 - U$ and $p(U) \leq 1 - U$. Let $\bar{U} \in (0, 1]$ denote the unique zero of p . Let $\mathcal{C}(U)$ denote the set of optimal configurations at U . We establish three facts about $\mathcal{C}(U)$ on $(0, \bar{U}]$.

- (i) $\mathcal{C}(U)$ is nonempty and compact. A configuration with wage bill $q_L - P_c(U) > 2$ has $P_c(U) < -1 < p(U)$, so restricting to configurations with wage bill at most 2 that cannot be improved by deleting a layer leaves $p(U)$ unchanged. If such a configuration has $L = 0$, then $z(q_0) \leq 2$. If such a configuration has $L > 0$, its rungs lie below $\bar{q}_U$ by Fact A.2(i), and its depth is at most $\bar{L}(U) \equiv 2/[h(1 - \bar{q}_U)U]$ by Fact A.2(iii), with $\bar{L}$ decreasing. The set is therefore compact, and $P_c(U)$ is continuous on it, so the maximum exists.
- (ii) Intensities fall in U . Configurations in $\mathcal{C}(U)$ maximize $P_c(0) - s^c U$, so by Fact A.3, $U < U'$ implies $s^{c'} \leq s^c$ for every $c \in \mathcal{C}(U)$ and $c' \in \mathcal{C}(U')$.
- (iii) $\mathcal{C}$ has closed graph. If $U_k \rightarrow U \in (0, \bar{U}]$ and $c_k \in \mathcal{C}(U_k)$, some subsequence has a common depth and converging rungs. The limit attains $p(U)$ by continuity, so it lies in $\mathcal{C}(U)$, and $s^{c_k} \rightarrow s^c$.

Step 3. For $U \in (0, \bar{U})$, the congestion identity $\rho = pQ^{-\gamma}$ is equivalent to $Q(U) \equiv [p(U)/\rho(U)]^{1/\gamma}$, and $Q = fn_F q_F$ to $n_F(U) \equiv Q(U)/[fq_F(U)]$. Both are continuous and strictly decreasing on $(0, \bar{U})$, since p falls while ρ and q_F rise, with limits $Q(0^+) = [p(0)/\rho(0)]^{1/\gamma} \in (0, \infty)$, $n_F(0^+) = \infty$ (the numerator stays bounded away from zero while $q_F(0^+) = 0$), and $Q(\bar{U}^-) = n_F(\bar{U}^-) = 0$.

Step 4. For $U \in (0, \bar{U})$ and $s \in \mathcal{S}(U) \equiv \{s^c : c \in \mathcal{C}(U)\}$, let

$$R(U, s) = sQ(U)^{1-\gamma} + n_F(U) - 1,$$

and let $\bar{R}(U)$ and $\underline{R}(U)$ denote its value at the largest and smallest elements of the compact set $\mathcal{S}(U)$. For $U < U'$ we have $\underline{R}(U) > \bar{R}(U')$. Every element of $\mathcal{S}(U)$ weakly exceeds every element of $\mathcal{S}(U')$ by Step 2(ii), while Q and n_F strictly fall. Moreover $\underline{R}(U) \geq n_F(U) - 1 \rightarrow \infty$ as $U \downarrow 0$, and, since $s^c \leq 1 + h\bar{L}(\bar{U}/2)$ for $U \geq \bar{U}/2$, $\bar{R}(U) \rightarrow -1$ as $U \uparrow \bar{U}$. The set $B = \{U : \bar{R}(U) \geq 0\}$ is therefore nonempty and bounded away from $\bar{U}$; let $U^* = \sup B$. By the closed graph of $\mathcal{C}$ (Step 2(iii)), $\bar{R}$ is upper semicontinuous and $\underline{R}$ is lower semicontinuous,

so $\bar{R}(U^*) \geq 0$ and $\underline{R}(U^*) \leq 0$. If $\underline{R} \leq 0 \leq \bar{R}$ held at U and at $U' > U$, then $0 \geq \underline{R}(U) > \bar{R}(U') \geq 0$. Thus U^* is the unique utility level at which $\underline{R}(U) \leq 0 \leq \bar{R}(U)$.

Step 5. $R(U^*, \cdot)$ is affine and strictly increasing in s , so there are $\theta \in [0, 1]$ and configurations $c^-, c^+ \in \mathcal{C}(U^*)$ with the largest and smallest intensities such that $\bar{s} = \theta s^{c^-} + (1 - \theta)s^{c^+}$ solves $R(U^*, \bar{s}) = 0$. The price $p^* = p(U^*) > 0$, the finder objects of Step 1, the aggregates of Step 3, and the configurations c^-, c^+ with shares $(\theta, 1 - \theta)$ satisfy (i)–(iv). An equilibrium with at most two active configurations exists, and $U^* > 0$.

Conversely, consider any equilibrium, and let U denote its common utility. First, $Q > 0$. If $Q = 0$, then conditions (i)–(iv) give $q_F = 0$, $U = 0$, and $p = p(0) > 0$, so the congestion-adjusted price $\rho = pQ^{-\gamma}$ diverges and the finder's problem has no solution, contradicting (i). Hence $n_F > 0$ and $q_F > 0$, so $U = \phi(q_F) > 0$. The finder's first-order condition gives $\rho = \rho(U) > 0$, so $p = \rho Q^\gamma > 0$; (ii) gives $p = p(U)$, so $U < \bar{U}$. Steps 1–3 then determine q_F, ρ, Q, n_F at their U -values, and (iv) reads $R(U, \bar{s}) = 0$ with $\bar{s}$ the share-weighted intensity of the active configurations, which lie in $\mathcal{C}(U)$; thus $\underline{R}(U) \leq 0 \leq \bar{R}(U)$ and $U = U^*$. All aggregates coincide with those constructed above, and market clearing determines the average intensity $\bar{s}^* = (1 - n_F^*)/(Q^*)^{1-\gamma}$, so the active shares must be a distribution over $\mathcal{C}(U^*)$ averaging to $\bar{s}^*$. The configurations in $\mathcal{C}(U^*)$ have depths differing by at most one and distinct intensities (Lemma A.3), so when at most two of them attain $p(U^*)$, that distribution is unique. ■

A.3 Proofs for Section 5

Proof of Proposition 1. The human layers above solve the chain problem above q^A , in which κ does not appear, and the layers below solve the block problem, in which q^A does not appear. By Fact A.4, the chain's rungs weakly rise with q^A and the block's rungs weakly rise with κ . ■

Proof of Proposition 2. Fix a configuration together with its ladder; the break-even price is then affine in (q^A, κ) : κ enters through the compute bill $-h(1 - q_{a-1})\kappa$ (or $-\kappa$ at $a = 0$), and q^A enters through the cost of the layer above, $-h(1 - q^A)[U^* + z(q_{a+1})]$, or through the output q^A when AI is deployed at the top. The best AI break-even price $p_{AI}(q^A, \kappa)$ is the supremum of these affine functions, hence convex, and it is strictly decreasing in κ . By Fact A.2(ii), a maximizing configuration has $z(q_{a-1}) \leq \kappa - U^*$, so its compute use per problem, $h(1 - q_{a-1})$ (one unit at $a = 0$), is bounded away from zero on bounded sets of κ . The no-adoption region $\{(q^A, \kappa) : p_{AI} < p^*\}$ is therefore the strict epigraph $\{\kappa > \bar{\kappa}(q^A)\}$, and $\{p_{AI} \leq p^*\}$ is convex as a sublevel set of a convex function, so $\bar{\kappa}$ is convex.

For the bound, Fact A.5 gives $p_{AI}(q^A, \kappa) < p^*$ whenever $\kappa > U^* + z(q^A)$, so $\bar{\kappa}(q^A) \leq U^* + z(q^A)$. At a rung, the firm using AI with $q^A = q_l^*$ and $\kappa = U^* + z(q_l^*)$ is the human firm with rung l relabeled, so it breaks even at p^* and $\bar{\kappa}(q_l^*) = U^* + z(q_l^*)$.

For tangency at an interior rung $q_l^* > 0$, the bound and the equality at the rung make any subgradient of $\bar{\kappa}$ at q_l^* a subgradient of $U^* + z(\cdot)$ there. This function is differentiable, so the subgradient equals $z'(q_l^*)$, and $\bar{\kappa}$ is differentiable at the rung with slope $z'(q_l^*)$. ■

Proof of Lemma 4. Fix (U, q^A, κ) . By (16), a configuration is profit-maximizing if and only if its chain is optimal above q^A and its block costs $\min_a B_a(\kappa)$, so two coexisting configurations that differ on both sides give two distinct optimal chains and two distinct blocks of minimum cost. The value of a chain with fixed rungs is affine in U with slope equal to minus its human mass, so two chains that differ strictly in mass (Lemma A.4) attain $V(q^A)$ jointly at most at one utility $U^{ch}(q^A)$. At that utility the two values move with q^A at the distinct rates δ^+ (Lemma A.4), so U^{ch} is strictly monotone in q^A . Likewise, two distinct blocks attain $\min_a B_a(\kappa)$ only at a unique utility $U^{bl}(\kappa)$ (Lemma A.5). Thus, coexisting configurations that differ on both sides require $U^{ch}(q^A) = U^{bl}(\kappa)$, which for each κ and pair of depths admits at most one q^A . For each pair of chain depths and pair of block depths, this set is the graph of a function of κ , so it has Lebesgue measure zero in the (q^A, κ) plane. ■

Proof of Proposition 3. At fixed U the finder block is fixed, $q_F = q_F(U)$ and $\rho = \rho(U)$, so along an iso-utility curve, equilibrium requires the break-even price to equal the clearing price of Fact A.6: $P(U, q^A, \kappa) = \tilde{p}(s_h)$, where P has envelope derivatives $P_{q^A} = \delta^+$ and $P_\kappa = -\delta^-$. The price enters no first-order condition, so at fixed U , $ds_h = c_q dq^A + c_\kappa d\kappa$. The signs follow from Proposition 1. A higher q^A weakly raises every rung above position a and none below, so every human mass above $(h(1 - q^A))$ at layer $a + 1$, $h(1 - q_{l-1})$ higher up) weakly falls, $c_q \leq 0$; a lower κ weakly lowers the rungs below and none above, so the masses below weakly rise, $c_\kappa \leq 0$.

Differentiate $P = \tilde{p}(s_h)$ at fixed U ; by Fact A.6, $d\tilde{p} = -C ds_h$. This gives $\delta^+ dq^A - \delta^- d\kappa = -C(c_q dq^A + c_\kappa d\kappa)$, i.e. $(\delta^+ + Cc_q) dq^A = (\delta^- - Cc_\kappa) d\kappa$, with $\delta^- - Cc_\kappa \geq \delta^- > 0$, which is the equality in (23). The inequality is immediate from $Cc_q \leq 0$ and $-Cc_\kappa \geq 0$, and strict if c_q or c_κ is nonzero, since $C > 0$ in equilibrium ($p, Q > 0$ and $\gamma > 0$). ■

Proof of Proposition 4 and of (25). Each active configuration has zero profit throughout the coexistence region, so along any path of (q^A, κ) within it, (19) gives $0 = -s_h^{c_i} dU - dp + \delta_{c_i}^+ dq^A - \delta_{c_i}^- d\kappa$ for $i = 1, 2$, with dU and dp common. Subtracting eliminates dp and yields $\Delta s_h dU = \Delta \delta^+ dq^A - \Delta \delta^- d\kappa$, which is (25) when $\Delta s_h \neq 0$. With c_2 the all-human firm, $\delta_{c_2}^\pm = 0$ and $s_h^{c_2} = s$, so $dU = -(\delta^+ dq^A - \delta^- d\kappa) / (s - s_h)$. For any improvement ($dq^A \geq 0 \geq d\kappa$, not both zero) the numerator is strictly positive, since $\delta^+, \delta^- > 0$; hence $dU < 0$ if and only if $s_h < s$, and then $dp = -s dU > 0$. When $s_h < s$ and $\gamma > 0$, solvers are displaced into finding. With p rising and U falling, $Q = (p/\rho(U))^{1/\gamma}$ rises while $q_F(U)$ falls, so $n_F = Q/(fq_F)$ rises and the mass of human solvers $1 - n_F$ falls. ■

Fix $U = U^*$ and write $p^H = p^*$. By the uniqueness assumed in Section 5, $s = 1 + S(q_0^*)$ and $q_1^* = q_1(q_0^*)$. The equilibrium ladder splits at every slot $a \leq L$. Rungs $a+1, \dots, L$ are an optimal chain above q_a^* , rungs $0, \dots, a-1$ solve the block program at compute cost $w(q_a^*)$ and bound q_a^* , and

$$p^H = V(q_a^*) - B_a(w(q_a^*); q_a^*), \quad (30)$$

since a better chain or a cheaper block, substituted into the configuration, would give a

break-even price above p^H . The masses are $s = m_a^* + S(q_a^*)$ with $m_a^* \equiv 1 + h \sum_{l=1}^a (1 - q_{l-1}^*)$, while a configuration with the AI at slot a employs $s_h = \tilde{b}_a + S(q^A)$, $\tilde{b}_a$ its block's human mass; hence

$$s - s_h = [m_a^* - \tilde{b}_a] + [S(q_a^*) - S(q^A)]. \quad (31)$$

At $a = 0$ the best configuration breaks even at $p_0(\kappa) = V(q^A) - \kappa$ and employs $s_h = S(q^A)$; at $a = 1$ it breaks even at $p_1(\kappa) = V(q^A) - B_1(\kappa)$ and employs $s_h = 1 + S(q^A)$. When a solution exists, let κ_a solve $p_a(\kappa_a) = p^H$. The configuration on the frontier has $\bar{\kappa} = \max_a \kappa_a$. Writing $D \equiv V(q^A) - p^H$, we have $\kappa_0 = D$, and Fact A.5 gives $D \leq w(q^A)$.

Proof of Lemma 2. For (i), at $\kappa = w(q_a^*)$,

$$p_a(w(q_a^*)) = V(q^A) - B_a(w(q_a^*); q^A) \geq V(q_a^*) - B_a(w(q_a^*); q_a^*) = p^H,$$

since V is nondecreasing (Lemma A.1(i)) and raising the bound from q_a^* to q^A enlarges the block's feasible set. Since p_a is strictly decreasing in κ , $\kappa_a \geq w(q_a^*)$ and $\bar{\kappa} \geq \kappa_a$. The block of the configuration on the frontier has a weakly higher compute cost and a weakly looser bound than the human rungs $0, \dots, a-1$, which solve the same program at $(w(q_a^*), q_a^*)$ by (30). By Fact A.4 the block's rungs weakly exceed the human ones (selecting the highest rungs where several ladders are optimal), so $\tilde{b}_a \leq m_a^* - h(1 - q_{a-1}^*)$, and at $a = 0$, $\tilde{b}_0 = 0 = m_0^* - 1$. Since $q^A \geq q_a^*$, $S(q^A) \leq S(q_a^*)$ by Lemma A.2(iii). Both terms of (31) are then nonnegative, so $s_h \leq s$.

For (ii) at $a = 0$, $s - s_h = 1 + S(q_0^*) - S(q^A)$. If $q^A \geq q_0^*$, then $S(q^A) \leq S(q_0^*)$ by Lemma A.2(iii), so $s - s_h \geq 1$. If $q^A < q_0^*$ and the chain above q^A is empty, $s - s_h = 1 + S(q_0^*) > 0$. If it is nonempty, Lemma A.2(ii) gives $q_1(q^A) \geq q_0^*$, so by (27) and (iii), $S(q^A) = h(1 - q^A) + S(q_1(q^A)) \leq h(1 - q^A) + S(q_0^*)$, and $s - s_h \geq 1 - h(1 - q^A) \geq 1 - h > 0$.

For (ii) at $a = 1$, first let $q^A < q_0^*$ and suppose κ_1 exists. Any bottom worker feasible at position 1 has $q_0 < q^A < q_0^*$. The configuration $(q_0, q_0^*, \dots, q_L^*)$ is then a feasible human configuration, so its break-even price is at most p^H , which reads

$$w(q_0) + h(1 - q_0)w(q_0^*) \geq w(q_0^*). \quad (32)$$

Since $D \leq w(q^A) < w(q_0^*)$, (32) bounds the block cost at $\kappa = D$ away from D :

$$w(q_0) + h(1 - q_0)D \geq w(q_0^*) - h(1 - q_0)[w(q_0^*) - D] \geq (1 - h)w(q_0^*) + hD > D,$$

so $B_1(D) > D$. Since B_1 is nondecreasing and $B_1(\kappa_1) = D$, $\kappa_1 < D = \kappa_0$. The configuration on the frontier therefore does not deploy AI at position 1.

Now let $q^A \geq q_0^*$. Then $s - s_h = S(q_0^*) - S(q^A) \geq 0$ by Lemma A.2(iii). At $q^A = q_0^*$ the two chain problems coincide and $s_h = s$. Suppose $q^A > q_0^*$. If $L = 0$, the chain above q_0^* is empty, hence so is the chain above q^A , and $s_h = s = 1$. If $L \geq 1$ and the chain above q^A is

empty, $s - s_h = S(q_0^*) > 0$; if not, Lemma A.2(i) gives $q_1(q^A) \geq q_1^*$, so by (27) and (iii),

$$S(q_0^*) - S(q^A) = h(q^A - q_0^*) + [S(q_1^*) - S(q_1(q^A))] \geq h(q^A - q_0^*) > 0.$$

For (iii), at $q^A = q_a^*$, the AI replicates rung a , and the optimal configuration is the equilibrium ladder with the layer- a humans replaced by compute. It is unique. At $\kappa = w(q_a^*)$, replacing the AI with humans of knowledge q_a^* preserves the break-even price (Fact A.5), so a second optimal configuration would yield a second optimal human configuration, contradicting the assumed uniqueness. For this configuration, $s - s_h = h(1 - q_{a-1}^*) > 0$, and $s - s_h = 1$ at $a = 0$. There are finitely many positions, and each $\kappa_{a'}$ is continuous in q^A ; at q_a^* , κ_a is strictly the largest, so the optimal configuration deploys the AI at position a for every q^A near q_a^* . The optimal ladder is unique and varies continuously, so s_h is continuous in q^A . Since $s - s_h > 0$ at q_a^* , $s - s_h > 0$ on a neighborhood $\mathcal{N}(q_a^*)$. ■

Proof of Lemma 3. Equilibrium at parameters $\eta \in \{q^A, \kappa\}$ solves $\mathcal{E}(U, \eta) \equiv P(U, q^A, \kappa) - \tilde{p}(s_h(U, q^A, \kappa); U) = 0$, with $\tilde{p}$ the clearing price of Fact A.6. $\mathcal{E}$ is strictly decreasing in U . By the envelope theorem, $P_U = -s_h < 0$. At fixed s , $\tilde{p}$ strictly rises in U (Fact A.6). By Fact A.4, every rung rises with U , so every human mass falls and s_h weakly falls in U . With $d\tilde{p}/ds = -C \leq 0$, $\tilde{p}(s_h(U); U)$ rises in U . At fixed U , $\partial\mathcal{E}/\partial\eta = P_\eta + C c_\eta$, with $P_{q^A} = \delta^+$ and $P_\kappa = -\delta^-$. The implicit function theorem gives $dU/d\eta = (P_\eta + C c_\eta)/(-\partial\mathcal{E}/\partial U)$, which is (24) with $\Phi = 1/(-\partial\mathcal{E}/\partial U) > 0$. The sign of $\partial U/\partial\kappa$ is that of $-(\delta^- - C c_\kappa) \leq -\delta^- < 0$. ■

Proof of Proposition 5. In a coexistence region whose configurations share the chain, $\Delta\delta^+ = 0$, and their blocks tie at the common (q^A, κ) at different depths (at equal depths the two blocks would be two optimal ladders of one configuration, which uniqueness excludes). By Lemma A.5 the deeper block has strictly more mass and strictly smaller δ^- , so (25) gives $dU/d\kappa = -\Delta\delta^-/\Delta s_h > 0$ and $dU/dq^A = 0$. In a coexistence region whose configurations share the block, Lemma A.4 gives $dU/dq^A = \Delta\delta^+/\Delta s_h < 0$ and $dU/d\kappa = 0$. ■

A.4 AI Without Adjustment

Consider the frozen economy of Section 5.5, with the benchmark ladder $q_0^* < \dots < q_L^*$ assumed unique, and let $w_l^* = U^* + z(q_l^*)$. Type l has knowledge q_l^* , wage w_l , and supply m_l per problem, with $m_0 = 1$ and $m_l = h(1 - q_{l-1}^*)$, and the number of problems is $G(Q^*)$. For a hierarchy c built from the existing types, with or without AI, write q^c for the knowledge of its top layer, m_l^c for the mass of type l it uses per problem attempted, and k^c for its compute per problem attempted, which is $h(1 - q)$ for the AI above a layer of knowledge q , one at the bottom, and zero without the AI. At wages w , hierarchy c breaks even at $P_c(w) = q^c - \sum_l m_l^c w_l - \kappa k^c$. The benchmark hierarchy c^* has $m_l^{c^*} = m_l$ and $k^{c^*} = 0$. An equilibrium is a wage vector w , a price p , and shares $\theta_c \geq 0$ of problems attempted by each hierarchy such that every type is employed, $\sum_c \theta_c m_l^c = m_l$ for every l , every active hierarchy

breaks even, $P_c(w) = p$, and no hierarchy earns positive profit, $P_c(w) \leq p$. Let $\hat{\kappa}(q^A)$ be the highest compute cost at which some hierarchy using AI breaks even at (w^*, p^*) .

Proposition A.1 *Suppose $q^A < q_L^*$. (i) An equilibrium exists, and in every equilibrium the allocation is the benchmark allocation and no firm uses AI. (ii) Every type with $q_l^* \leq q^A$ earns at most κ . (iii) If $\kappa < \hat{\kappa}(q^A)$ and no wage rises, then some wage falls, and the price of problems and finders' income rise.*²⁰

Proof of Proposition A.1. For (i), let $w = \lambda w^*$ and $p(\lambda) = P_{c^*}(\lambda w^*)$. The profit $P_c(\lambda w^*) - p(\lambda)$ of any c is affine in λ . It is nonpositive at $\lambda = 0$, since $q^c \leq q_L^* = p(0)$. For c without AI it is also nonpositive at $\lambda = 1$, since c^* is optimal at benchmark wages, and therefore on all of $[0, 1]$. The largest $\lambda \in [0, 1]$ at which no AI hierarchy earns positive profit exists by continuity. At it, c^* breaks even and no hierarchy earns positive profit. For the allocation, a hierarchy uses one unit of type 0 per problem if type 0 is its bottom layer and none otherwise. Clearing for type 0 therefore gives $\sum_c \theta_c = 1$, with type 0 at the bottom of every active hierarchy. Suppose types $0, \dots, l-1$ sit in order at the bottom of every active hierarchy. Type l 's usage per problem is m_l if it sits directly above type $l-1$, $h(1-q^A) < m_l$ if the AI sits between them, which requires $q^A > q_{l-1}^*$, and zero if type l is absent. Clearing for type l therefore puts type l directly above type $l-1$ in every active hierarchy. After type L , the only remaining position for the AI is above type L , which requires $q^A > q_L^*$. Hence every active hierarchy is c^* .

For (ii), let $q_l^* \leq q^A$, let l' be the highest type with $q_{l'}^* \leq q^A$, so $l' < L$, and let c replace types $l, \dots, l'$ by AI and keep every other layer. Then $k^c = m_l$, $q^c = q_L^*$, type $l' + 1$ has mass $h(1-q^A) \leq m_{l'+1}$, and every other mass is unchanged. The condition $P_c(w) \leq P_{c^*}(w)$ reads $\sum_{j=l}^{l'} m_j w_j + [m_{l'+1} - h(1-q^A)] w_{l'+1} \leq \kappa m_l$. Every term on the left is nonnegative, so $w_l \leq \kappa$.

For (iii), the AI hierarchy that breaks even at (w^*, p^*) when $\kappa = \hat{\kappa}(q^A)$ earns positive profit there when $\kappa < \hat{\kappa}(q^A)$, so $w \neq w^*$. With $w \leq w^*$, some wage falls and $\sum_l m_l w_l < \sum_l m_l w_l^*$. Hence $p = P_{c^*}(w) > p^*$. Finding effort is Q^* , so $\rho = p(Q^*)^{-\gamma}$ and finders' income $f q_F^* \rho$ rise. ■

A.5 Proofs for Section 6

We begin with the planner's problem. An allocation consists of a mass n_F of finders with knowledge q_F (so that $Q = f n_F q_F$) and weights θ_c that split the found problems across configurations; it is feasible if it clears the market for people, $G(Q)\bar{s} + n_F = 1$, where bars denote θ -averages. A configuration c uses s_h^c humans and s_A^c units of compute per problem

²⁰Equilibrium wages are not unique. They form the polytope $\{w : P_c(w) \leq P_{c^*}(w) \text{ for all } c\}$, which contains w^* if and only if $\kappa \geq \hat{\kappa}(q^A)$.

and generates the per-problem surplus

$$v_c = q_L^c - Z^c - \kappa s_A^c, \quad Z^c = \sum_l (n_l/n_0) z(q_l), \quad (33)$$

where Z^c is the configuration's training bill. Net output, which the planner distributes equally, is therefore

$$W = G(Q) \sum_c \theta_c v_c - n_F z(q_F). \quad (34)$$

Because W is linear in the weights, the planner operates only surplus-maximizing configurations. We denote the planner's value by W^* .

Proof of Proposition 6. Let (p, ρ, Q) be the taxed equilibrium. Since total finder revenue is $pG(Q)$ and $\tau = \gamma$, the lump-sum rebate is $T = \gamma pG(Q)$; the active finder's after-tax surplus is $U - T$. Let $\widetilde{W}$ denote (34) at an arbitrary feasible allocation (marked with tildes). Three optimality conditions of the taxed equilibrium hold at every feasible configuration and finder knowledge level:

- (a) Free entry: $\tilde{v}_c \leq p + (U - T) \tilde{s}_h^c$.
- (b) Finder optimality at $\tau = \gamma$: $(1 - \gamma)fq\rho - z(q) \leq U - T$, for every q ,
- (c) Concavity of G : $pG(\tilde{Q}) \leq pG(Q) + (1 - \gamma)\rho(\tilde{Q} - Q)$.

By (a) and clearing, $G(\tilde{Q}) \sum_c \tilde{\theta}_c \tilde{v}_c \leq pG(\tilde{Q}) + (U - T)(1 - \tilde{n}_F)$; (b) at $q = \tilde{q}_F$, multiplied by $\tilde{n}_F$, gives $(1 - \gamma)\rho\tilde{Q} \leq \tilde{n}_F[(U - T) + z(\tilde{q}_F)]$. Adding the two inequalities and applying (c),

$$\widetilde{W} \leq pG(\tilde{Q}) + (U - T) - (1 - \gamma)\rho\tilde{Q} \leq pG(Q) - (1 - \gamma)\rho Q + (U - T) = U, \quad (35)$$

where the final equality uses $\rho Q = pG(Q)$ and $T = \gamma pG(Q)$. Since the population has measure one, all humans obtain utility U , and transfers cancel in aggregate, the equilibrium's net output equals U . Therefore $\widetilde{W} \leq U$ for every feasible allocation; the equilibrium allocation is feasible and attains $W = U$, so $U = W^*$. ■

Lemma A.6 *Let η be any parameter of the firm's profit-maximization problem, with direct effect $\partial_\eta \bar{v} = \sum_c \theta_c \partial_\eta v_c$, and let any associated revenue be rebated lump-sum, dT per capita. Then*

$$dU = G(Q) \partial_\eta \bar{v} d\eta + dT - \gamma \rho dQ. \quad (36)$$

Proof. Zero profit, $0 = v_c - p - s_h^c(U - T)$ for each active c , averaged with the weights θ_c , gives by the envelope theorem $dp = \partial_\eta \bar{v} d\eta - \bar{s}(dU - dT)$. Finder entry gives $f q_F d\rho = dU - dT$.

The congestion identity $\rho = pQ^{-\gamma}$ gives $d\rho = Q^{-\gamma}dp - \gamma\rho dQ/Q$. Substituting the first and third into the second, with $f q_F = Q/n_F$,

$$(dU - dT)(n_F + \bar{s} G(Q)) = G(Q) \partial_\eta \bar{v} d\eta - \gamma\rho dQ, \quad (37)$$

and by market clearing $n_F + \bar{s} G(Q) = 1$, which gives (36). ■

Proof of Proposition 7. Let $\bar{s}_A = \sum_c \theta_c s_A^c$ be the share-weighted compute per problem, so that $N_A = \bar{s}_A G(Q)$. Apply Lemma A.6 twice with $\partial_\eta \bar{v} = -\bar{s}_A$. For $\eta = \kappa$ there is no rebate, so $dU/d\kappa = -N_A - \gamma\rho dQ/d\kappa$. For $\eta = t_A$ the rebate $dT = N_A dt_A$ cancels the direct term, and the allocation depends only on κ_e , so $dQ = (dQ/d\kappa) dt_A$; hence $dU/dt_A|_0 = -\gamma\rho dQ/d\kappa = dU/d\kappa + N_A$. ■

A.6 Proofs for Section 7

Throughout this subsection, write $r \equiv \kappa/(f q^A)$, so an improvement in AI is a fall in r . An equilibrium with AI finders adds to Definition 1 a mass $n_A \geq 0$ of AI agents in finding, each contributing effort $f q^A$, with the entry condition $f q^A \rho - \kappa \leq 0$, with equality if $n_A > 0$. Aggregate finding effort is $Q = f(n_F q_F + n_A q^A)$; only humans enter the population constraint.

Proof of Lemma 5. If $\rho^* > r$, one AI agent earns $f q^A \rho^* - \kappa > 0$ at the no-AI equilibrium, so that allocation violates the entry condition. Any equilibrium therefore has $n_A > 0$, and $\rho = r$. If $\rho^* \leq r$, the no-AI equilibrium together with $n_A = 0$ satisfies it. ■

Proof of Proposition 8. Suppose $n_F > 0$ and $n_A > 0$. By Lemma 5, $\rho = r$, and human-finder indifference gives $U = U_F(r)$, with U_F strictly increasing (Fact A.1). Any improvement lowers r , hence lowers U . Zero profit gives $p = p(U)$, so p rises. The congestion identity gives $Q = (p(U)/r)^{1/\gamma}$, so Q rises. For the second claim, p and Q are as in the benchmark at utility U . Clearing gives $n_F = 1 - \bar{s}(U)G(Q)$, and

$$n_A f q^A = Q - f q_F n_F \geq 0 \quad \Longleftrightarrow \quad \bar{s}(U) G(Q) + \frac{Q}{f q_F} \geq 1,$$

which is the residual condition $R(U) \geq 0$ from the proof of Theorem 1 (Step 4). Since R is strictly decreasing with $R(U^*) = 0$, coexistence implies $U \leq U^*$, and at $U = U^*$ the residual is zero, so $n_A = 0$. Hence $U < U^*$ and $r < \rho^*$.

A remark used in the text: coexistence occurs exactly for r in the interval (r_F, ρ^*) , where r_F is the exit threshold of Proposition 9; the interval is nonempty since $n_F \rightarrow n_F^* > 0$ as $r \rightarrow \rho^*$, and it is open, so coexistence holds on an open set of (q^A, κ) . ■

Proof of Proposition 9. For (i), in the coexistence region, $n_F = 1 - \bar{s}(U)G(Q)$ with $U = U_F(r)$. As r falls, U falls, so $\bar{s}$ weakly rises (Fact A.3), $p(U)$ rises, and Q rises; hence $\bar{s}G(Q)$ rises and n_F falls, strictly except at the utilities where two configurations attain p , where it jumps down. Thus $r_F \equiv \inf\{r : n_F > 0\}$ is well defined and positive (as $r \downarrow 0$, $\bar{s}G \rightarrow \infty$), and below it $n_F = 0$.

For (ii), with $n_F = 0$, equilibrium is the root of

$$R_A(U; r) \equiv \bar{s}(U) \left(\frac{p(U)}{r} \right)^{\frac{1-\gamma}{\gamma}} - 1 = 0.$$

R_A is strictly decreasing in U with downward jumps where two configurations attain p , so existence and uniqueness, with mixing between tied configurations, follow as in Theorem 1, Steps 4–5. A root exists for every $r \leq r_F$: $R_A(U_F(r_F); r_F) = 0$ at the threshold, R_A at that utility strictly rises as r falls, and $R_A \rightarrow -1$ as $U \uparrow \bar{U}$. A fall in r raises R_A pointwise, so the root weakly rises. At a root with a single active configuration, the implicit function theorem gives $dU/dr < 0$ strictly. At a root inside a coexistence region, U satisfies an equality between two configurations' break-even conditions, in which r does not appear, so $dU/dr = 0$ there. The human-finder participation constraint $U \geq U_F(r)$ binds at r_F and is slack below it, since the left side weakly rises and the right side strictly falls as r falls.

For (iii), at the pre-improvement equilibrium an AI agent deployed in finding earns zero; after any improvement, at the old prices, it earns $f q^A \rho - \kappa' > f q^A \rho - \kappa = 0$. ■

Proof of Proposition 10. The minimand $(U_F + z(q))/q$ has first-order condition $z'(q)q - z(q) = U_F$, i.e. $\phi(q) = U_F$; since ϕ is strictly increasing and $U_F = \phi(q_F)$ (Fact A.1), the unique minimizer is q_F , with minimized value $(U_F + z(q_F))/q_F = z'(q_F)$. By Lemma 5 and the finder first-order condition $\rho^* = z'(q_F)/f$, AI is adopted in finding if and only if $\kappa/q^A < z'(q_F)$. The finder frontier is the ray $\kappa = q^A z'(q_F)$. Some AI use is profitable if and only if κ is below the solver frontier or the finder ray, so the joint frontier is $\max\{\bar{\kappa}(q^A), q^A z'(q_F)\}$. ■

Proof of Proposition 11. For (i), at a partial adoption equilibrium the all-human configuration is active, so $p = p(U)$, and clearing holds at the average intensity $\bar{s} \in [s_h, s]$, strictly below s when $s_h < s$. Then $R(U, \bar{s}) = 0$ implies $\bar{R}(U) \geq R(U, s) > 0$. Under the uniqueness assumed in Section 5, $\bar{R}(U^*) = R(U^*, s) = 0$, so $U \neq U^*$. The monotonicity of $\underline{R}$ and $\bar{R}$ (Theorem 1, Step 4) rules out $U > U^*$. The finder frontier at utility U is $\kappa/q^A = f\rho(U)$, and $\rho(U)$ is strictly increasing (Fact A.1), so at $U < U^*$ the finder frontier lies below its benchmark level.

For (ii), the levels $\rho = r < \rho^*$ and $U = U_F(r) < U^*$ are those of Proposition 8. The solver frontier $\bar{\kappa}(q^A; U)$ is defined by $p_{AI}(q^A, \bar{\kappa}; U) = p(U)$, a static program whose maximizing configuration is defined at every utility, whether or not any firm is active. The envelope derivatives $\partial p_{AI}/\partial U = -s_h$ and $p'(U) = -s$ give $\delta^- \partial \bar{\kappa}/\partial U = s - s_h$, evaluated at the maximizing configuration; under the hypothesis the right side is positive at each utility in $[U, U^*]$, so $\bar{\kappa}(q^A; U) < \bar{\kappa}(q^A; U^*)$. ■