

# Artificial Intelligence in the Knowledge Economy\*

Enrique Ide & Eduard Talamàs

February 25, 2025

## Abstract

Artificial Intelligence (AI) can transform the knowledge economy by automating non-codifiable work. To analyze this transformation, we incorporate AI into an economy where humans form hierarchical organizations: Less knowledgeable individuals become “workers” doing routine work, while others become “solvers” handling exceptions. We model AI as a technology that converts computational resources into “AI agents” that operate autonomously (as co-workers and solvers/co-pilots) or non-autonomously (solely as co-pilots). Autonomous AI primarily benefits the most knowledgeable individuals; non-autonomous AI benefits the least knowledgeable. However, output is higher with autonomous AI. These findings reconcile contradictory empirical evidence and reveal tradeoffs when regulating AI autonomy.

---

\*Department of Economics, IESE Business School, Carrer d’Arnús i de Garí 3-7, 08034 Barcelona, Spain (eide@iese.edu and etalamas@iese.edu). We extend our deepest gratitude to Luis Garicano, whose invaluable suggestions have fundamentally shaped the presentation of this work. We have also benefited enormously from the comments of Ricardo Alonso, Nano Barahona, Markus Brunnermeier, Nicolás Figueroa, Andrea Galeotti, Ben Golub, Jason Hartline, Frank Levy, Niko Matouschek, Mike Ostrovsky, Cristóbal Otero, Sebastián Otero, Esteban Rossi-Hansberg (the Editor), four anonymous referees, and seminar participants at the various venues where we presented this work. We acknowledge the financial support of IESE through the High Impact Initiative-course 2024/2025. We declare we have no relevant or material financial interests that relate to the research described in this paper.

# 1 Introduction

Recent developments in Artificial Intelligence (AI) are driving a new wave of automation, enabling machines to perform sophisticated knowledge work such as coding, research, and problem-solving. While the potential of this technology to reshape the landscape of work is undeniable, its precise implications remain the subject of growing controversy (Brynjolfsson, 2022; Johnson and Acemoglu, 2023; Acemoglu, 2024; Autor, 2024; Korinek and Suh, 2024). This controversy arises from two factors. First, it is unclear whether lessons from previous waves of automation can inform our understanding of AI’s effects (Muro et al., 2019; Agrawal et al., 2019). Second, because AI is still in its infancy and individuals and firms are experimenting with its use (McElheran et al., 2023; Humlum and Vestergaard, 2024), current empirical evidence cannot yet fully capture its equilibrium effects.

In this paper, we develop a framework for studying the equilibrium effects of AI on the knowledge economy.<sup>1</sup> Our approach is grounded in two insights. First, the defining characteristic of knowledge work is that production know-how is predominantly tacit (or non-codifiable)—developed through repeated observations of practical successes and failures—and therefore inherently embodied in individuals (Polanyi, 1966; Garicano, 2000; Garicano and Wu, 2012). As a result, individuals’ time and knowledge are critical bottlenecks, making firms central to organizing production (Garicano and Rossi-Hansberg, 2015). Second, unlike traditional automation technologies, AI learns by example, uncovering patterns that cannot be codified into explicit rules (Autor, 2024; Brynjolfsson et al., 2025). This ability allows AI systems to acquire tacit knowledge and perform cognitive, non-codifiable tasks (see Table 1). By automating such work, AI has the potential to alleviate knowledge bottlenecks, opening the door to a fundamental reorganization of production.

More precisely, we incorporate AI into a canonical model of the knowledge economy: The knowledge hierarchies first introduced by Garicano (2000). This model emphasizes the central role of tacit knowledge in shaping organizations and labor market outcomes. Within this framework, we study the equilibrium effects of AI on occupational choice, organizational structure, and the distribution of labor income. Our analysis highlights that introducing AI agents is fundamentally different from

	<b>Codifiable</b>	<b>Non-Codifiable</b>
<b>Manual</b>	Robots	Physical AI
<b>Cognitive</b>	Enterprise Software	AI

Table 1: The Distinguishing Feature of AI

*Notes.* The defining feature of AI is its ability to perform cognitive work that relies on non-codifiable knowledge. This distinguishes it from enterprise software and robots, which automate codifiable cognitive and manual work, respectively. AI also differs from “Physical AI” (Nvidia, 2025), a concept that envisions robots powered by AI, which may, in the future, be capable of automating manual work that requires non-codifiable knowledge, such as caregiving or construction.

<sup>1</sup>According to data from the U.S. Bureau of Labor Statistics, as of December 2024, over 100 million people in the U.S. were employed in cognitive or knowledge-based roles (St. Louis Fed, 2024).

expanding the human labor force, and that AI’s impact crucially depends on its problem-solving capabilities and its degree of autonomy.

Our starting point—the pre-AI economy—is the baseline model of [Garicano and Rossi-Hansberg \(2004\)](#), [Antràs, Garicano, and Rossi-Hansberg \(2006\)](#), and [Fuchs, Garicano, and Rayo \(2015\)](#). Labor and knowledge are the sole inputs in production. Humans are endowed with one unit of time and are heterogeneous in terms of knowledge. Individuals use their time to pursue production opportunities that require cognitive work—such as handling customer support, drafting contracts, or designing products—but encounter problems of varying difficulty during the production process. Output is produced when an individual successfully solves the problem she confronts, which occurs when her knowledge exceeds the problem’s difficulty. If a human cannot solve a problem on her own, she may seek help from another human. Help, however, is costly in terms of time. Knowledge is tacit because matching problems to those who can solve them is difficult, and experts cannot outline a plan of action in advance. In this sense, individuals “know more than they can tell” ([Polanyi, 1966](#)).

In the competitive equilibrium of the pre-AI economy, humans either pursue production opportunities on their own or join hierarchical firms to optimize the use of their time and knowledge. These firms exhibit *management by exception*: Less knowledgeable individuals become “workers” who pursue production opportunities, while more knowledgeable individuals become “solvers” specializing in assisting workers with the exceptional problems they are unable to solve. This structure effectively shields the most knowledgeable individuals from routine work, enabling them to focus on their comparative advantage: solving complex problems.<sup>2</sup>

Our innovation is to incorporate AI into this otherwise canonical setting. We base our model on three key developments in modern AI, documented in Section 2. First, unlike earlier automation technologies, AI—like humans—can acquire tacit knowledge. However, unlike humans—who are constrained by their individual time—firms can leverage computational resources (“compute”) to use AI’s knowledge at scale. Second, the rise of flexible, general-purpose foundation models—adaptable to diverse tasks with minimal additional training—enables firms to rely on a shared AI model. Third, technology firms are developing and starting to deploy AI agents that can do cognitive work autonomously. These agents do not just passively answer questions but can pursue projects independently.

Guided by these developments, we model AI as a technology available to all firms that converts compute into AI agents, all endowed with the same exogenously fixed level of knowledge. As a benchmark, we first consider the case where these agents can do exactly the same as humans. These AI agents are “autonomous” because they can function both as “co-workers” (pursuing production opportunities) and as “solvers/co-pilots” (providing advice). A key contribution of our analysis is to contrast this benchmark with the case of “non-autonomous” AI, where—for technological or

---

<sup>2</sup>There is both anecdotal and systematic empirical evidence showing the emergence of such “knowledge hierarchies” (see, e.g., [Garicano and Hubbard, 2012](#); [Caliendo and Rossi-Hansberg, 2012](#); [Caliendo et al., 2015, 2020](#)). For instance, [Alfred Sloan \(1924\)](#), a former head of General Motors (GM), once wrote: “We do not do much routine work with details. They never get up to us. I work fairly hard, but on exceptions.”

regulatory reasons—AI agents are restricted to acting solely as co-pilots.

In the post-AI economy, firms decide their organizational structure and whether to integrate AI agents into their production processes. The total amount of compute available in the economy is exogenous, while its price is endogenously determined in equilibrium. We assume that compute is abundant relative to human time, making humans the binding constraint in human-AI interactions. This assumption reflects the exponential growth in computational capacity over the past two centuries (Nordhaus, 2007) and the fact that current AI models process information and generate outputs 10 to 100 times faster than humans (Amodei, 2024).

Regarding autonomous AI, we show that its impact critically depends on whether it is “basic” or “advanced.” The introduction of basic autonomous AI—which creates AI agents with the knowledge of pre-AI workers—induces the most knowledgeable pre-AI solvers to focus on supporting AI-driven production. This reallocation forces the least knowledgeable human workers to seek assistance from less knowledgeable individuals, pushing the marginal pre-AI workers to take specialized problem-solving roles. In other words, basic autonomous AI displaces humans from routine work towards complex problem-solving. In this scenario, the least knowledgeable individuals are harmed by the technology—they face competition from AI in production work and must rely on less knowledgeable individuals for support—whereas the most knowledgeable individuals benefit by using AI to leverage their expertise at a low cost.

In contrast, the introduction of an advanced autonomous AI—which creates AI agents with the knowledge of pre-AI solvers—induces the least knowledgeable humans to seek assistance from AI instead of humans. Consequently, the marginal pre-AI solvers are pushed into routine work, displacing humans from complex problem-solving. In this scenario, both the least and the most knowledgeable individuals benefit from the technology. The least knowledgeable gain access to relatively cheap AI agents that can help them solve complex problems, while the most knowledgeable continue to thrive by leveraging advanced AI agents for routine work.

Recent anecdotal evidence from professional services industries illustrates these reorganizations. For example, Beane and Anthony (2024) document that junior analysts in investment banking are increasingly distanced from senior partners as AI takes over tasks that once required junior involvement. Similarly, a 2022 survey of M&A lawyers found that AI, by automating routine knowledge work such as document review, has allowed junior lawyers to shift their focus to more complex responsibilities, including client engagement and legal analysis (Litera, 2022).

The impact of AI on labor outcomes also critically depends on its autonomy. We show that non-autonomous AI tends to benefit the least knowledgeable individuals the most, in contrast to autonomous AI, which mainly benefits the most knowledgeable. However, overall output is higher with autonomous AI than with non-autonomous AI.

Intuitively, non-autonomous AI primarily benefits the least knowledgeable because it does not compete with them for production work and provides a means of solving complex problems without human assistance. Its lower opportunity cost relative to autonomous AI—stemming from its inabil-

ity to perform production work independently—reinforces this advantage, allowing the least knowledgeable to capture a larger share of AI-generated benefits. For the most knowledgeable individuals, the effect is reversed. Non-autonomous AI is less advantageous than autonomous AI because it lacks the capacity to handle routine work and competes with them for advisory roles.

Our findings reveal a trade-off between output and inequality when regulating AI autonomy. They also reconcile seemingly contradictory empirical findings on AI’s impact on work. On the one hand, some evidence suggests that AI disproportionately benefits the least knowledgeable individuals and reduces performance inequality (e.g., Dell’Acqua et al., 2023; Noy and Zhang, 2023; Peng et al., 2023; Wiles et al., 2023; Caplin et al., 2024; Brynjolfsson et al., 2025). On the other hand, there is also evidence that AI complements high-skilled knowledge workers while substituting for their low-skilled counterparts (Berger et al., 2024; Alekseeva et al., 2024; Toner-Rodgers, 2024). Although these findings may seem contradictory, they can be rationalized by our framework as we explain in Section 7: They align with the distinct impacts of non-autonomous AI and basic autonomous AI, respectively.

## Related Literature

This paper builds on the literature on knowledge hierarchies—which investigates how tacit (or non-codifiable) knowledge shapes organizational structures and labor market outcomes—to deepen our understanding of the effects of automation. Specifically, we explore the implications of AI, a new technology that can automate work traditionally reliant on tacit knowledge.

The literature on knowledge hierarchies originates with Garicano (2000), who introduces the model and shows that knowledge hierarchies are optimal when production requires the application of tacit knowledge. Garicano and Rossi-Hansberg (2004, 2006) extend this model to include heterogeneous agents, exploring how the endogenous organization of knowledge work affects labor income inequality. Fuchs et al. (2015) further develop this literature by characterizing the equilibrium contractual arrangements when there is asymmetric information about knowledge.<sup>3</sup> Building on this foundation, we develop a framework to examine how modern AI can transform knowledge work.

In the context of knowledge hierarchies, the most closely related paper is Antràs et al. (2006). They study the effects of offshoring by comparing the equilibrium of a closed economy with one in which firms can form international teams. Our paper differs from theirs in two key respects. First, while offshoring gives firms access to a human population with varying knowledge levels, AI allows firms to automate knowledge work at scale. As discussed in Section 7, this leads to qualitatively different outcomes. Second, we introduce AI-specific dimensions—such as AI’s knowledge, autonomy, and computational resources—to explore how AI’s impact is shaped by these dimensions. These considerations, absent from Antràs et al. (2006), are crucial for understanding AI’s unique effects.<sup>4</sup>

---

<sup>3</sup>Other important contributions to this literature include Garicano and Hubbard (2007, 2009), Garicano and Rossi-Hansberg (2012), Bloom et al. (2014), Garicano and Hubbard (2018), Caicedo et al. (2019), Gumpert et al. (2022), Adhvaryu et al. (2023), and Carmona and Laohakunakorn (2024).

<sup>4</sup>Another related contribution is Pans’s (2018, ch. 5) textbook example, which shows how superintelligent ma-

With regard to the existing literature on automation, the first important recent contribution is Zeira (1998), who shows how automation can lead to a decline in the labor share as the economy develops. Acemoglu and Restrepo (2018) contend that, by depressing wages, automation also encourages the creation of new tasks in which labor has a comparative advantage. Further developing these ideas, Acemoglu and Restrepo (2022) and Acemoglu et al. (2024) study how different types of technological advances—such as automation and labor-augmenting technologies—affect wages, inequality, and productivity. Complementary research by Autor et al. (2003) and Acemoglu and Autor (2011) emphasizes how the automation of codifiable tasks—driven by the adoption of computers and robots—explains employment and wage polarization, as these tasks are predominantly performed by middle-skilled workers.<sup>5</sup>

Our work contributes to this literature by focusing on AI, a fundamentally new form of automation. Unlike earlier technologies that primarily automated codifiable tasks, AI can acquire and apply tacit knowledge, allowing it to perform cognitive, non-codifiable work. As discussed above, by automating such work, AI has the potential to ease knowledge bottlenecks and enable a fundamental reorganization of production.

Traditional models of automation, while useful for understanding earlier automation waves, have a hard time capturing this transformation. These models often overlook firms' central role in structuring production—an essential factor for assessing AI's impact on the knowledge economy. In contrast, our framework explicitly accounts for the organization of knowledge work, showing how AI reshapes production and how shifts in firm composition create new pathways for AI to influence labor outcomes.<sup>6</sup> Moreover, we introduce AI-specific dimensions absent from existing models, leading to new insights. For example, we find that non-autonomous AI disproportionately benefits the least knowledgeable individuals, while autonomous AI primarily benefits the most knowledgeable. As far as we know, these results have no parallel in the existing literature on automation or AI.

Finally, we also contribute to the broader literature on the economics of knowledge and ideas (e.g., Romer, 1990, 1992; Jones, 2005), the knowledge-based view of the firm (e.g., Nonaka, 1991; Grant, 1996; Garicano, 2000; Garicano and Wu, 2012), and the economics of data (e.g. Jones and Tonetti, 2020; Farboodi and Veldkamp, 2023, 2024). Specifically, we highlight how modern AI systems blur the conventional boundary between codifiable and non-codifiable knowledge.

Both types of knowledge are, in principle, non-rival—one person's use does not reduce availability for others—but their practical application depends on rival resources such as time, effort, or compu-

---

chines—matching the abilities of the most capable humans—eliminate labor income inequality. In contrast, we demonstrate that AI's effects on labor income distribution are nuanced, depending critically on AI's proficiency and autonomy.

<sup>5</sup>Other important contributions include Aghion et al. (2017), Moll et al. (2022), Korinek and Suh (2024), Acemoglu and Loebbing (2024), Acemoglu (2024), and Jones (2024).

<sup>6</sup>For instance, we show that in the case of a basic autonomous AI, the most knowledgeable human solvers switch toward using AI for routine production work. As a result, a lower-quality pool of solvers is left to assist those who continue as workers in the post-AI world, reducing their productivity. As discussed in Section 5.1, this outcome is consistent with anecdotal evidence from sectors such as investment banking and legal services.



tational power. The key distinction between codifiable and non-codifiable knowledge lies in their transferability. Codifiable knowledge, such as information stored in databases or manuals, is highly transferable because it can be expressed as explicit rules and procedures. In contrast, non-codifiable knowledge has limited transferability, spreading primarily through direct experience. Consequently, individuals who possess tacit knowledge often become critical bottlenecks in its application.

Modern AI systems do not fit neatly into this classification. These systems learn from examples, identifying patterns that cannot always be reduced to explicit, formal descriptions. Once trained, however, they can apply these insights at scale using compute. In this sense, AI effectively encodes tacit-like knowledge into model parameters—making it codifiable to machines but not humans. This creates a hybrid form of “embodied” knowledge that is neither purely human nor strictly codified, challenging conventional wisdom on how knowledge is transferred, scaled, and applied.

## Roadmap

The rest of the paper is organized as follows. Section 2 documents three key developments in modern AI that shape the assumptions of our baseline model, introduced in Section 3, which focuses on autonomous AI. Section 4 characterizes the equilibrium with autonomous AI, while Section 5 examines the impact of this technology relative to the pre-AI equilibrium. Section 6 explores the implications of non-autonomous AI. Finally, Section 7 concludes by discussing additional implications of our analysis and linking our findings to emerging empirical evidence about AI’s effects on the labor market. We relegate all formal mathematical proofs to the Online Appendix.

## 2 Three Key Developments in Modern AI

In this section, we document three key developments in modern AI that are essential for understanding the excitement and concerns surrounding its impact on labor and organizational outcomes: (i) AI’s ability to scale the use of tacit knowledge, (ii) the advent of foundation models, and (iii) the rise of AI agents capable of performing cognitive work autonomously. These three developments guide the assumptions of our baseline model in Section 3.

1. *AI Can Use Tacit Knowledge at Scale.*— AI represents a fundamentally new form of automation. Unlike earlier technologies, AI—like humans—can acquire tacit knowledge, enabling it to perform cognitive, non-codifiable work. As Autor (2024, p. 7) observes:

Pre-AI, computing’s core capability was its faultless and nearly costless execution of... procedural tasks. Its Achilles’ heel was its inability to master... tasks requiring tacit knowledge. Artificial Intelligence’s capabilities are precisely the inverse [...] Rather than relying on hard-coded procedures, AI learns by example, gains mastery without explicit instruction and acquires capabilities that it was not explicitly engineered to possess [...] Like a

human expert, AI can weave formal knowledge (rules) with acquired experience to make — or support — one-off, high-stakes decisions.

However, unlike humans—who can only apply their tacit knowledge within the limits of their own time—firms can leverage AI across all available compute in the economy. This reflects the nonrival nature of digital information, which has a negligible marginal cost of reproduction (Brynjolfsson and McAfee, 2016; Goldfarb and Tucker, 2019). As Waisberg and Hudek (2021, p. 177) explain:

You can ask a question at a seminar or webinar, but you can only get the data as provided by the host or speaker, and a given speaker can only be in one seminar at a time... [In contrast, AI] scales; no longer are you limited by the time of a single domain expert. If their knowledge and behavior are encoded into an AI, you can have as many copies of that AI working at the same time as you want.

2. *The Advent of Foundation Models.*— Foundation models represent a transformative shift in AI. These are models characterized by their ability to handle a wide variety of tasks through training on large-scale datasets using self-supervised learning techniques (Bommasani et al., 2022; Council of Economic Advisers, 2024). Models like GPT-4, Gemini, and Claude Sonnet exemplify this approach, serving as flexible, general-purpose platforms that can be adapted to diverse downstream tasks with minimal additional training. This adaptability enables firms to leverage shared AI models, removing the need for costly development of proprietary, domain-specific systems.<sup>7</sup> As a result, a small number of powerful AI models can function across multiple domains, driving significant economic and industry-wide impact.<sup>8</sup>

Foundation models have demonstrated their versatility across various industries, from customer service automation to legal research and financial analysis. For instance, Klarna, a fintech company, automated over 700 customer service roles using a version of OpenAI’s models. In the legal field, Harvey leverages OpenAI’s technology to automate research and document analysis. Similarly, Claude Sonnet powers applications at Lonely Planet for itinerary creation and at Bridgewater Associates for tasks like research, coding, and data visualization.

Moreover, evidence suggests that general-purpose models often outperform specialized AI systems. For instance, ChatGPT—a fine-tuned version of GPT-4—has demonstrated superior performance to BloombergGPT—a Large Language Model (LLM) trained specifically on financial data—in various finance-related tasks (Li et al., 2023). Similarly, Nori et al. (2023) demonstrate that GPT-4 surpasses specialist models across medical benchmarks and argue that this advantage extends to other domains, including accounting, law, nursing, and clinical psychology. This shows that even

---

<sup>7</sup>For instance, Andreessen Horowitz, a venture capital firm, surveyed around 100 Fortune 500 companies across industries such as technology, telecom, banking, healthcare, and energy. The survey found that 94% of these firms were using existing foundation models, while only 6% were creating their own specialized models (Wang and Xu, 2024).

<sup>8</sup>Using the language of computer scientists, foundation models exhibit *emergence*—the ability to develop unexpected capabilities that enable them to perform tasks beyond their explicit training. Combined with scalability, emergence then leads to *homogenization*, defined as the reliance on a few powerful models across multiple domains (Bommasani et al., 2022).



in fields that traditionally relied on highly specialized models, general-purpose foundation models offer better performance, increased flexibility, and lower costs.

3. *The Rise of AI Agents*.—The adaptability, efficiency, and scalability of foundation models are driving a new wave of automation: the rise of AI agents capable of performing cognitive work autonomously. Technology firms such as DeepMind and OpenAI are actively developing and starting to deploy these AI agents. As Dario Amodei, co-founder and CEO of Anthropic, explains (Amodei, 2024):

I am not talking about AI as merely a tool to analyze data [...] It does not just passively answer questions; instead, it can be given tasks that take hours, days, or weeks to complete, and then goes off and does those tasks autonomously, in the way a smart employee would, asking for clarification as necessary.

The emergence of autonomous AI agents raises the crucial question of whether current AI models are sophisticated enough for such responsibilities. Sarah Tavel, a general partner at the venture capital firm Benchmark, is optimistic (Stebbing, 2024b):

There are certainly use cases that can be automated right now, and we see, we see a lot of them... beyond that, there is no question that as the foundation models improve, we are going to see the ability to take on more and more complex tasks.

This optimism is grounded in the fact that much of today's economy is driven by knowledge work, an area where AI excels compared to industries reliant on manual labor. For instance, Eloundou et al. (2024) estimate that 46% of jobs could have more than half of their tasks affected by AI foundation models in the near future. Similarly, Goldman Sachs (2023) projects that two-thirds of U.S. occupations are at least partially exposed to AI-driven automation, with approximately a quarter of all work potentially automatable when weighted by employment share.

These developments call for new research into AI's economic implications. As Hadfield (2025) remarked during a panel at the ASSA 2025 meetings in San Francisco:

All of this is about autonomy [...] Think about what that means [...] [AI agents would be] engaging in transactions, entering into contracts, hiring people, designing products, setting prices for products, advertising, doing all of those things. So I think it's really important for us to be shifting our thinking a little bit. [...] Not just thinking about it as a technology or a tool but rather as a novel economic agent.

## 3 Autonomous AI

### 3.1 The Model

*The Pre-AI Economy*.— There is a unit mass of humans, each endowed with one unit of time and exogenous knowledge  $z \in [0, 1]$ . The distribution of knowledge in the population is given by a cu-

mulative distribution function  $G$  with density  $g$ . The only assumption we impose on this distribution is that  $g$  is continuous and strictly positive for all  $z$ . The knowledge of each individual is perfectly observable.

There is a large measure of identical competitive firms. Production occurs inside firms, which are the residual claimants of all output. Labor and knowledge are the sole inputs in production. Firms have no fixed costs and, for simplicity, two layers at most.<sup>9</sup>

Single-layer firms rely on a single individual to carry out production. This “independent producer” dedicates her entire time to pursuing a single production opportunity. These opportunities can be thought of as work requiring cognitive ability, such as handling customer support, drafting contracts, hiring employees, and designing products. Each production opportunity is linked to a problem whose difficulty  $x$  is ex-ante unknown and distributed uniformly on  $[0, 1]$ , independently across problems. If the knowledge of the human engaging in production exceeds the problem’s difficulty, she solves the problem and produces one unit of output. Otherwise, no output is produced.<sup>10</sup>

Two-layer firms hire one “solver” and multiple “workers,” where all workers have the same knowledge. The restriction that all workers share the same knowledge is without loss because—as we show below—the equilibrium matching arrangement between (human) workers and (human) solvers exhibits strict positive assortative matching.

As in single-layer firms, each worker in a two-layer firm devotes her time to a single production opportunity. The difference is that if the worker cannot solve the problem independently, she can ask the solver for help. If the solver’s knowledge exceeds the problem’s difficulty, she communicates the solution to the corresponding worker, who then produces a unit of output. Otherwise, the worker fails to produce. However, communication is costly in that each request for help consumes  $h \in (0, 1)$  units of the solver’s time, regardless of whether the solver knows the solution. Hence, a two-layer organization optimally hires exactly  $n(z)$  workers of knowledge  $z < 1$  to fully exploit its solver’s time, where  $n(z)$  satisfies  $h \times n(z) \times (1 - z) = 1$ .

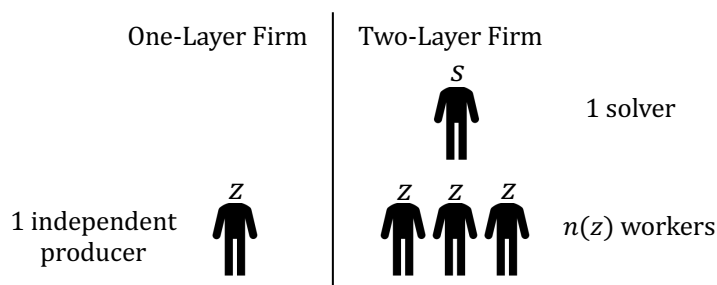


Figure 1: The Two Possible Firm Configurations in the Pre-AI World

<sup>9</sup>In the Online Appendix, we show that many insights we uncover—such as the conditions under which the least and most knowledgeable individuals benefit from AI—naturally generalize to settings with an arbitrary number of layers.

<sup>10</sup>The assumption that  $x \sim U[0, 1]$  is without loss, as the distribution of knowledge in the population is arbitrary. Under this assumption, an individual’s knowledge  $z$  is interpreted as the fraction of problems she can solve on her own.

Figure 1 depicts the two possible firm configurations of the pre-AI world, where the letter attached to each human corresponds to her knowledge. In this context, knowledge is tacit because solvers cannot prescribe a plan of action to workers in advance, and a worker must interact with her solver to determine whether the solver knows the solution to a problem.

*Autonomous AI.*—We model AI as a technology that converts compute into “AI agents,” all with the same exogenously fixed level of knowledge,  $z_{AI} \in [0, 1]$ .<sup>11</sup> Each AI agent requires one unit of compute and serves as a perfect substitute for a human with equivalent knowledge. These AI agents are “autonomous” because they can function both as “co-workers” (pursuing production opportunities) and as “co-pilots” (providing advice). In Section 6, we explore the case where the AI agents can only be co-pilots, rendering them non-autonomous.

All firms have access to AI. Post-AI, firms decide their organizational structure and whether to integrate AI agents into their production processes. The total amount of compute in the economy, denoted by  $\mu$ , is exogenous. We further assume that production opportunities exceed the combined capacity of humans and AI agents to pursue them, meaning opportunities are abundant relative to available resources. As we discuss below, this implies that there is no unemployment in equilibrium.

Figure 2 illustrates the five possible post-AI firm configurations. In addition to single- and two-layer firms that hire only humans (the only possible pre-AI firms), there are three additional possibilities: Single-layer automated firms (which use AI agents as independent producers), bottom-automated firms (which use AI agents as workers), and top-automated firms (which use AI agents as solvers). Note that a firm will never use AI agents in both layers of the organization because an AI solver knows the solution to the same set of problems as an AI worker.

Our modeling approach is motivated by the three AI developments described in Section 2. First, unlike humans—who can only apply their knowledge within the limits of their own time—AI is scalable, meaning that once knowledge  $z_{AI}$  is encoded into the technology, such knowledge can be used simultaneously across all available compute. Second, foundation models—flexible and general-purpose—allow firms to access AI on equal terms through a shared AI model. Third, firms can

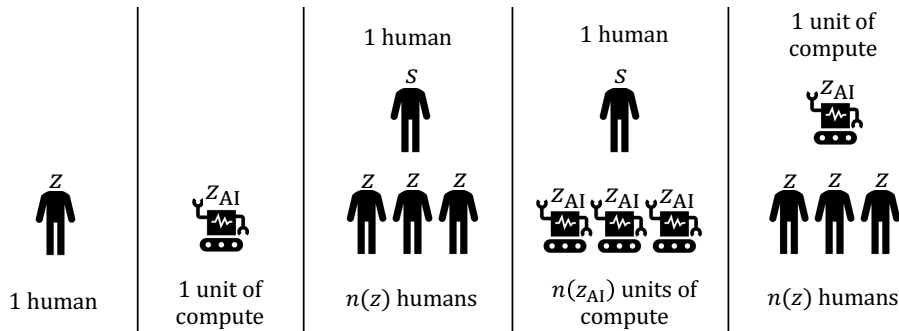


Figure 2: The Five Possible Firm Configurations in the Post-AI World

<sup>11</sup>We assume  $z_{AI} < 1$  because the equilibrium has a discontinuity at  $z_{AI} = 1$ . See the Online Appendix for the case  $z_{AI} = 1$ .

combine AI with compute to deploy autonomous AI agents capable of performing cognitive work independently.

*Wages, Prices, and Profits.*—Let  $w(z)$  be the wage of a human with knowledge  $z$  and denote by  $r$  the rental rate of one unit of compute (i.e., the rental rate of an AI agent). All humans in this economy are risk neutral and maximize their income. In particular, the owners of compute rent it to the highest bidder. We normalize the value of each unit of output to one.

The profit of a single-layer organization is:

$$\Pi_1 = \begin{cases} z - w(z) & \text{if the firm hires a human with knowledge } z \\ z_{\text{AI}} - r & \text{if the firm uses an AI agent} \end{cases}$$

In other words, it is equal to the expected output of its independent producer minus the hiring cost.

The profit of a two-layer organization, in turn, depends on whether it uses AI as a solver (i.e., automates the top layer, a “*tA*” firm), uses AI as a worker (i.e., automates the bottom layer, a “*bA*” firm), or it does not use AI (an “*nA*” firm):

$$\begin{aligned} \Pi_2^{\text{tA}}(z) &= n(z)[z_{\text{AI}} - w(z)] - r && \text{(where } z \leq z_{\text{AI}}) \\ \Pi_2^{\text{bA}}(s) &= n(z_{\text{AI}})[s - r] - w(s) && \text{(where } z_{\text{AI}} \leq s) \\ \Pi_2^{\text{nA}}(s, z) &= n(z)[s - w(z)] - w(s) && \text{(where } z \leq s) \end{aligned}$$

where  $z$  and  $s$  denote the knowledge of a human worker and a human solver, respectively, and we use the fact that no two-layer firm hires a solver who is less knowledgeable than its workers. In all three cases, the profit of a firm is its expected output minus the cost of the resources it uses. For instance, in the case of a *tA* firm that hires workers with knowledge  $z$ , its total expected output is  $n(z)z_{\text{AI}}$ , while the cost of resources is  $n(z)w(z) + r$ .

*Competitive Equilibrium.*— A competitive equilibrium comprises the following objects: (1) an allocation of compute across its various potential uses; (2) a partition of the human population into distinct occupations; (3) a description of the matching between different types of agents; and (4) a wage schedule,  $w(z)$ , and a rental rate of compute,  $r$ .

Let  $\mu_i$ ,  $\mu_w$ , and  $\mu_s$  be the amount of compute used to create AI independent producers, workers, and solvers, respectively. We denote by  $I$  the set of humans hired as independent producers, and by  $W_p$  and  $W_a$  the set of human workers assisted by humans and AI agents, respectively.<sup>12</sup> Similarly, let  $S_p$  be the set of humans who assist other humans, and  $S_a$  the set of humans assisting AI agents in production. We assume that the sets  $I$ ,  $W_p$ ,  $W_a$ ,  $S_p$ , and  $S_a$  are measurable with respect to  $G$ . Since a *tA* firm that hires human workers with knowledge  $z \in W_a$  rents one AI agent,  $\mu_s = \int_{W_a} h(1-z)dG(z)$ . Similarly, since a *bA* firm that hires a human solver with knowledge  $s \in S_a$  rents  $n(z_{\text{AI}})$  AI agents,  $\mu_w = n(z_{\text{AI}}) \int_{S_a} dG(z)$ .

Finally, let  $m : W_p \rightarrow S_p$  represent the function describing the pointwise matching arrangement generated by the hiring decisions of *nA* firms. Specifically,  $m(z)$  denotes the knowledge of the human

<sup>12</sup>We use the subscript “*p*” (for people) instead of “*h*” (for humans), to avoid any confusion with the helping cost  $h$ .

solver assisting human workers with knowledge  $z$ .<sup>13</sup> As mentioned earlier, the equilibrium matching arrangement between human workers and human solvers exhibits strict positive assortative matching, so  $m$  is strictly increasing. Additionally, it must satisfy the following resource constraint:

$$(1) \quad \int_Y h(1-u)dG(u) = \int_{m(Y)} dG(u), \text{ for any measurable set } Y \subseteq W_p$$

where  $m(Y)$  denotes the set of solvers assigned to the set of workers  $Y \subseteq W_p$ . This condition ensures that the total time required to consult on problems left unsolved by the workers in  $Y$  equals the total time available from the solvers matched to these workers,  $m(Y)$ . In essence, it reflects that  $nA$  firms optimally hire  $n(z)$  workers with knowledge  $z$  to fully use the time of their solvers.

**Definition (Competitive Equilibrium).** An equilibrium consists of nonnegative amounts  $(\mu_i, \mu_w, \mu_s)$ , measurable sets  $(I, W_p, W_a, S_p, S_a)$ , a function  $m : W_p \rightarrow S_p$  satisfying (1), a wage schedule  $w : [0, 1] \rightarrow \mathbb{R}_{\geq 0}$  and a rental rate of compute  $r \in \mathbb{R}_{\geq 0}$ , such that:

1. Firms optimally choose their structure (while earning zero profits).
2. Markets clear: (i)  $\mu_i + \mu_w + \mu_s = \mu$ , and (ii) the union of the sets  $(I, W_p, W_a, S_p, S_a)$  is  $[0, 1]$  and the intersection of any two of these sets has measure zero.

Note that the human workers in  $W_p$  are endogenously matched with the human solvers in  $S_p$  according to the pointwise matching function  $m$ . In contrast, all the humans in  $W_a$  are matched with an AI solver (which has knowledge  $z_{AI}$ ), while all the humans in  $S_a$  are matched with AI workers (whose knowledge is  $z_{AI}$ ).

Total output in this economy is  $\int_I z dG(z) + z_{AI} \mu_i + \int_{W_p} m(z) dG(z) + \int_{S_a} n(z_{AI}) z dG(z) + \int_{W_a} z_{AI} dG(z)$ , where each term represents the total output of each of the five firm configurations depicted in Figure 2 (in the same order). Total labor income, in turn, is equal to  $\int_0^1 w(z) dG(z)$ , whereas total capital income is  $\mu r$ . Since firms earn zero profits in equilibrium, total output equals the sum of total labor income and total capital income.

*Compute is “Abundant” Relative to Human Time.*— We focus on the case in which compute is sufficiently abundant so that the binding constraint in human-AI interactions is human time, not compute. In other words, there are more AI agents than can be matched with humans, implying that some AI agents must engage in independent production.<sup>14</sup>

We consider this the most relevant case due to the exponential growth in computational capacity over the past two centuries (Nordhaus, 2007) and the fact that current AI models process information and generate outputs 10 to 100 times faster than humans (Amodei, 2024). However, to explore the effects of this assumption, we also examine the case of limited compute in the Online Appendix.

<sup>13</sup>As discussed in Fuchs et al. (2015),  $m(z)$  is an approximation of a firm that hires a small interval of human workers  $(z - \varepsilon_1, z + \varepsilon_1)$  and a small interval of human solvers  $(m(z) - \varepsilon_2, m(z) + \varepsilon_2)$  with the requirement that the mass of solvers in the firm  $\int_{m(z)-\varepsilon_2}^{m(z)+\varepsilon_2} dG(u)$  is equal to  $h$  times the mass of problems left unsolved by the workers in the firm  $\int_{z-\varepsilon_1}^{z+\varepsilon_1} (1-u)dG(u)$ . The latter requirement captures that  $nA$  firms optimally hire  $n(z)$  workers of knowledge  $z$ .

<sup>14</sup>A sufficient condition for this to hold is  $\int_0^{z_{AI}} n(z)^{-1} dG(z) + n(z_{AI})(1 - G(z_{AI})) < \mu$ . For any distribution  $G$  and helping cost  $h \in (0, 1)$ , there exists a finite  $\mu$  that satisfies this condition for all  $z_{AI} \in [0, 1]$ .

*Some Notation.*— For future reference, we define  $W \equiv W_a \cup W_p$  and  $S \equiv S_a \cup S_p$  as the overall set of human workers and solvers of the economy, respectively. We also denote by  $e : S_p \rightarrow W_p$  the inverse of the matching function  $m$ . That is,  $e$  is the “employee matching function” denoting the knowledge of the human worker matched with a human solver with knowledge  $s \in S_p$ . This function always exists given that, as argued above, the matching function  $m$  is strictly increasing in equilibrium.

Finally, for any arbitrary set  $B$ , we denote by  $\text{int}B$  and by  $\text{cl}B$  the interior and closure of  $B$ , respectively. We also use  $B \preceq B'$  to indicate that the set  $B \subseteq [0, 1]$  “lies below” the set  $B' \subseteq [0, 1]$ . Formally,  $B \preceq B'$  if  $\sup B \leq \inf B'$ . For example,  $W_a \preceq W_p$  means that the most knowledgeable human worker assisted by AI is weakly less knowledgeable than the least knowledgeable human worker assisted by another human.

### 3.2 Discussion of the Model

Before proceeding with the analysis, we map our model to real-world applications. We also clarify how to interpret compute in our framework, justify our focus on a single AI technology, and explain why we do not consider the possibility of unemployment. The Online Appendix discusses some limitations of our model and suggests avenues for future research.

*Interpretation.*— In our benchmark model, AI agents are autonomous in that they can operate both as co-workers—capable of pursuing production opportunities—and as co-pilots—providing assistance.<sup>15</sup> A real-world example of an AI co-worker is Klarna, which automated over 700 customer service agents using OpenAI’s models (see Section 2). Another is Agentforce by Salesforce, an AI agent that not only identifies business leads using company data but also contacts prospects and schedules meetings on behalf of salespeople (Rosenbush, 2024). Similarly, All Day TA, a startup founded by University of Toronto Professors Kevin Bryan and Joshua Gans, leverages foundation models to fully automate the work of university teaching assistants.

In contrast, Brynjolfsson et al. (2025) provide an example of an AI co-pilot in the customer support industry: their study highlights the productivity effects of equipping human agents with a real-time, AI-based conversational assistant to manage customer inquiries. Another example is GitHub Copilot, which aids developers by suggesting code snippets, autocompleting functions, and recommending improvements to code quality. Similarly, BrainTrust AIR assists HR professionals by reviewing candidates, grading applications, and generating detailed scorecards to streamline recruitment decisions.

Regarding AI autonomy, for simplicity, we focus on the two extreme cases. First, we consider a scenario in which AI agents are fully autonomous and capable of performing all tasks that humans possessing knowledge  $z_{\text{AI}}$  can accomplish. Then, in Section 6, we examine the opposite case, where AI agents are entirely non-autonomous and can only provide advice. Most AI applications, however,

---

<sup>15</sup>The model is silent about where production opportunities come from. The role of organizations is to decide who to assign to routine work and who to assign to specialized problem solving. Hence, workers and independent producers—human or AI—should not be thought of as being responsible for generating production opportunities; they are only responsible for pursuing them.



operate between these extremes, with autonomy levels varying by sector. For example, tools like Harvey automate many routine legal tasks, such as document review and contract analysis. Yet, they remain unable to engage directly with clients or argue cases in court.

Because our framework models the broader knowledge economy, we interpret our results as approximations of AI’s impact based on its average level of autonomy across the different relevant sectors. Accordingly, the practical implications of our analysis will likely fall between the predictions of these two benchmark cases. To determine which extreme better reflects current and future realities, the critical question is: How many tasks traditionally performed by humans can AI agents like Harvey handle autonomously, and how will this capability evolve over time?

*On Compute.*— Our main goal is to analyze how AI affects human labor outcomes. For this reason, we take AI technology as given. Thus, in our framework, “compute” refers to the resources required during the inference phase—the stage in which an AI system generates outputs based on new input data after it has already been trained.

To facilitate comparison between AI output and human labor, we express compute in terms of human time. In practice, however, compute is often measured in FLOPs or, for LLMs, in tokens. In an earlier version of this paper (Ide and Talamàs, 2025), we explain how compute in our model translates into these real-world units.

*Why Only One AI?*— A key property of AI is that it relies on compute, which is general-purpose hardware that can be rented on demand. This implies that competition among AI providers drives compute toward the highest-performing AI systems—a phenomenon that computer scientists and venture capitalists call the commoditization of foundation models (Navani, 2023; Stebbings, 2024a).<sup>16</sup> As Aravind Srinivas, founder of PerplexityAI, explains (Stebbing, 2024a):

[It] is a losing game... every time you end up finishing a large training run, you burn a lot of money, you have a great model, and then you watch it be destroyed by the next update... So where are you recovering all that money back? (...) Nobody wants to use the API [i.e., the model] if somebody else is offering a better model at a cheaper price.

This competition for compute underpins our assumption of focusing on a single AI technology. Indeed, in the Online Appendix, we show that if multiple AI technologies are available and all of them require similar compute to create AI agents, only the most advanced AI will be used in equilibrium. However, if less advanced AI technologies require substantially less compute, they may coexist alongside more advanced models. This result is consistent with the development of smaller foundation models, such as GPT-4o Mini and Claude 3 Haiku.

Furthermore, this extension highlights a fundamental distinction between human and AI hetero-

---

<sup>16</sup>Compute can be rented on demand through cloud computing providers and easily redeployed across various AI foundation models. According to a recent survey by Andreessen Horowitz, “Most enterprises are designing their applications so that switching between models requires little more than an API change. Some companies are even pre-testing prompts so the change happens literally at the flick of a switch” (see Wang and Xu, 2024).

geneity. While humans with varying levels of knowledge can coexist in equilibrium despite requiring the same resources—time—to perform knowledge work, AI heterogeneity emerges only when models differ significantly in both knowledge and resource requirements. Hence, introducing a population of AI agents is fundamentally different from expanding the human labor force.

*No Unemployment.*— In defining our competitive equilibrium, we excluded the possibility of unemployment. This is without loss of generality, as we assume that total available resources—human time and compute—are scarce relative to production opportunities (even though compute is abundant relative to time). This scarcity ensures that, in equilibrium, it is always worthwhile for every human to be employed in some capacity, even if AI is more knowledgeable than a fraction of the human population.

In the Online Appendix, we explore an extension where compute is so abundant that it outstrips production opportunities. In that case, we show that AI leads to technological unemployment: All humans who are less knowledgeable than AI become unemployed. Organizations, however, still display a hierarchical structure in that the most knowledgeable humans specialize in tackling the problems that AI cannot solve.

### 3.3 Benchmark: The Pre-AI Equilibrium

We begin by presenting a partial characterization of the equilibrium without AI. This is also the equilibrium when the amount of compute  $\mu$  is zero, and was originally described by [Fuchs et al. \(2015\)](#).<sup>17</sup> Note that in this case,  $W_a = S_a = \emptyset$ , so  $W = W_p$  and  $S = S_p$ .

**Proposition 1.** *In the absence of AI, there is a unique equilibrium. The equilibrium is efficient (i.e., it maximizes total output) and has the following features:*

- *Occupational stratification:*  $W \preceq I \preceq S$ .
- *Positive assortative matching:* The function  $m : W \rightarrow S$  is strictly increasing.
- $W \neq \emptyset$  and  $S \neq \emptyset$ . However,  $I \neq \emptyset$  if and only if  $h > h_0 \in (0, 1)$ .

Moreover, the wage function  $w$  is continuous, strictly increasing, and convex (strictly so when  $z \in W \cup S$ ), and is given by:

- $w(z) = m(z) - w(m(z))/n(z)$  for all  $z \in W$ .
- $w(z) = z$  for all  $z \in I$ .
- $w(z) = C + \int_{\inf S}^z n(e(u))du > z$  for all  $z \in S$ , where  $C > \inf S$  when  $h < h_0$  (and  $C = \inf S$  otherwise).

In particular,  $w(z) > z$  for all  $z \notin \text{cl}I$  (so  $w(z) > z$  for all  $z \in [0, 1]$  when  $h < h_0$ ).

The equilibrium without AI—which we illustrate in Figure 3—has several salient features. First, it is efficient, as the First Welfare Theorem holds in this economy (with perfect competition, complete

<sup>17</sup>An economy with  $\mu = 0$  is different from one with  $\mu > 0$  and  $z_{\text{AI}} = 0$ . This is because, even if AI cannot solve any problems, it can still pursue production opportunities, thereby expanding the economy’s production possibility frontier.

information, and only pecuniary externalities). Moreover, the wage of an individual with knowledge  $z$  corresponds to her marginal product, defined as the increase in output from introducing one additional human with this knowledge into the economy.

Second, the equilibrium exhibits occupational stratification: Workers are less knowledgeable than independent producers, who are, in turn, less knowledgeable than solvers. Intuitively, more knowledgeable individuals have a comparative advantage in specialized problem solving, as this allows them to leverage their knowledge by applying it to more than one problem. Hence,  $(W \cup I) \preceq S$ . Similarly, less knowledgeable individuals have a comparative advantage in assisted rather than independent production work, as they are less likely to succeed on their own. Consequently,  $W \preceq I$ .

Third, there is strict positive assortative matching: More knowledgeable workers in  $W$  match with more knowledgeable solvers in  $S$ . The reason is that worker and solver knowledge are complements: For a given team of workers, a more knowledgeable solver increases expected output, while for any given solver, more knowledgeable workers increase team size.

Fourth, the set of workers and the set of solvers are always nonempty, but the set of independent producers is empty when the communication cost  $h$  is below a threshold  $h_0 \in (0, 1)$ . Intuitively, lower values of  $h$  make two-layer organizations more attractive than single-layer ones, leading all humans to be employed by two-layer organizations.

Fifth, workers and solvers earn strictly more than their expected output as independent producers

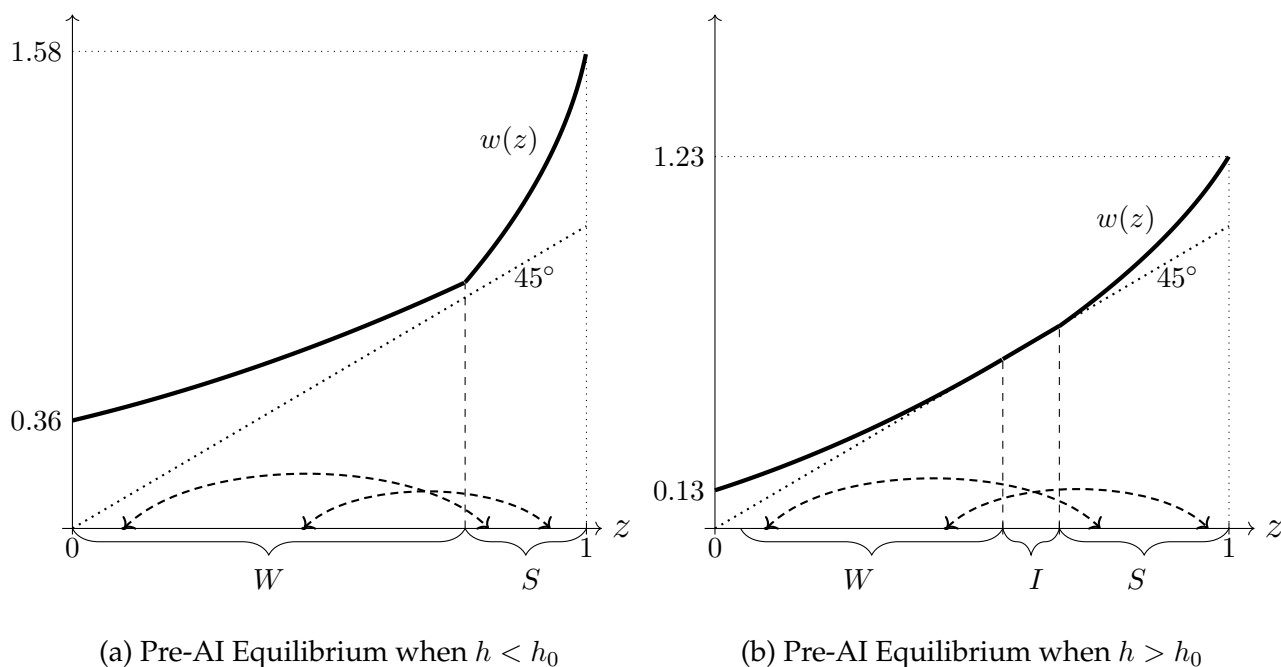


Figure 3: Illustration of the Pre-AI Equilibrium.

Notes. Distribution of knowledge:  $G(z) = z$ . Parameter values: For panel (a),  $h = 1/2 (< h_0 = 3/4)$ , while for panel (b),  $h = 0.8125 (> h_0 = 3/4)$ . The thick line depicts the equilibrium wage function. The dashed arrows illustrate the matching between workers and solvers.

(except in the case of the most knowledgeable worker and the least knowledgeable solver when  $h \geq h_0$ ). The reason for this is that the marginal value of knowledge is strictly lower than 1 for workers (as their knowledge is used to free up solver time),<sup>18</sup> exactly equal to 1 for independent producers (as their expected output equals their knowledge), and strictly greater than 1 for solvers (as they can use their knowledge in more than one problem). Hence, if solvers earned their expected output as independent producers, then all two-layer firms would try to hire the most knowledgeable individuals as solvers. Similarly, if workers earned their expected output as independent producers, then all two-layer firms would try to hire the least knowledgeable individuals as workers.

Finally, the equilibrium wage function  $w$  is continuous, strictly increasing, and convex (strictly so when  $z \in W \cup S$ ). The wages for the different occupations are obtained as follows. For independent producers,  $w(z) = z$  is an immediate implication of the zero-profit condition of single-layer firms. For workers and solvers, consider the problem of a two-layer organization that recruited  $n(z)$  workers with knowledge  $z \in W$  and is deciding which solver  $s \in S$  to hire:

$$\max_{s \in S} \left\{ n(z)[s - w(z)] - w(s) \right\}$$

The corresponding first-order condition evaluated at  $s = m(z)$  implies that  $w'(m(z)) = n(z)$ , or, equivalently,  $w'(z) = n(e(z))$  for any  $z \in S$ . Thus  $w(z) = C + \int_{\inf S}^z n(e(u))du$  for any  $z \in S$ , where the constant  $C$  is chosen so that the wage function is continuous, as the latter condition is necessary for market clearing. The wages of workers are then determined by the zero-profit condition of two-layer organizations:  $w(z) = m(z) - w(m(z))/n(z)$  for any  $z \in W$ .

For simplicity, we restrict attention to  $h < h_0$ . This implies that there are no independent producers in the pre-AI equilibrium. In a previous version of this paper (Ide and Talamàs, 2024), we show that virtually all of our results—except one, discussed in Section 5.2—extend to the case where  $h \geq h_0$ .

## 4 The Equilibrium with Autonomous AI

Our objective is to understand the effects of autonomous AI by comparing the pre- and post-AI equilibrium. In this section, we provide a partial characterization of the post-AI equilibrium, focusing on the essential information needed for the comparison (which is carried out in Section 5).

For future reference, we index the post-AI equilibrium using the superscript “\*” (note that the pre-AI equilibrium has no superscript). Furthermore, recall that  $W^* \equiv W_a^* \cup W_p^*$  is the overall set of human workers and that  $S^* \equiv S_a^* \cup S_p^*$  is the overall set of human solvers.

**Proposition 2.** *In the presence of an autonomous AI, there is a unique equilibrium. The equilibrium is efficient and has the following features:*

- *Occupational stratification:  $W^* \preceq I^* \preceq S^*$ .*

<sup>18</sup>A marginal increase in the knowledge of a worker with knowledge  $z$  frees  $h < 1$  units of her solver’s time. This allows her firm to hire  $1/(1 - z)$  extra workers, with an expected net output gain of  $(m(z) - w(z))/(1 - z) < 1$  (as the wage  $w(z)$  of that worker is strictly greater than  $z$  in the interior of  $W$ ).

- No worker is more knowledgeable than AI, and no solver is less knowledgeable than AI:  $W^* \preceq \{z_{AI}\} \preceq S^*$ .
- Positive assortative matching:  $m^* : W_p^* \rightarrow S_p^*$  is strictly increasing and  $W_a^* \preceq W_p^*$  and  $S_p^* \preceq S_a^*$ .

Furthermore, AI is always used for independent production, and whether it is also used as a worker or as a solver depends on its knowledge level **relative to the pre-AI equilibrium**:

- If  $z_{AI} \in W$ , then AI is necessarily used as a worker (and possibly also as a solver).
- If  $z_{AI} \in S$ , then AI is necessarily used as a solver (and possibly also as a worker).

Finally, the rental rate of compute  $r^*$  is equal to  $z_{AI}$ , and the wage function  $w^*$  is continuous, strictly increasing, and convex (strictly so when  $z \in W_p^* \cup S_p^*$ ), and is given by:

- $w^*(z) = z_{AI}(1 - 1/n(z)) > z$  for all  $z \in W_a^*$ .
- $w^*(z) = m^*(z) - w^*(m^*(z))/n(z)$  for all  $z \in W_p^*$ .
- $w^*(z) = z$  for all  $z \in I^*$ .
- $w^*(z) = C^* + \int_{\inf S_p^*}^z n(e^*(u))du$  for all  $z \in S_p^*$ , where  $C^* = \inf S_p^*$ .
- $w^*(z) = n(z_{AI})(z - z_{AI}) > z$  for all  $z \in S_a^*$ .

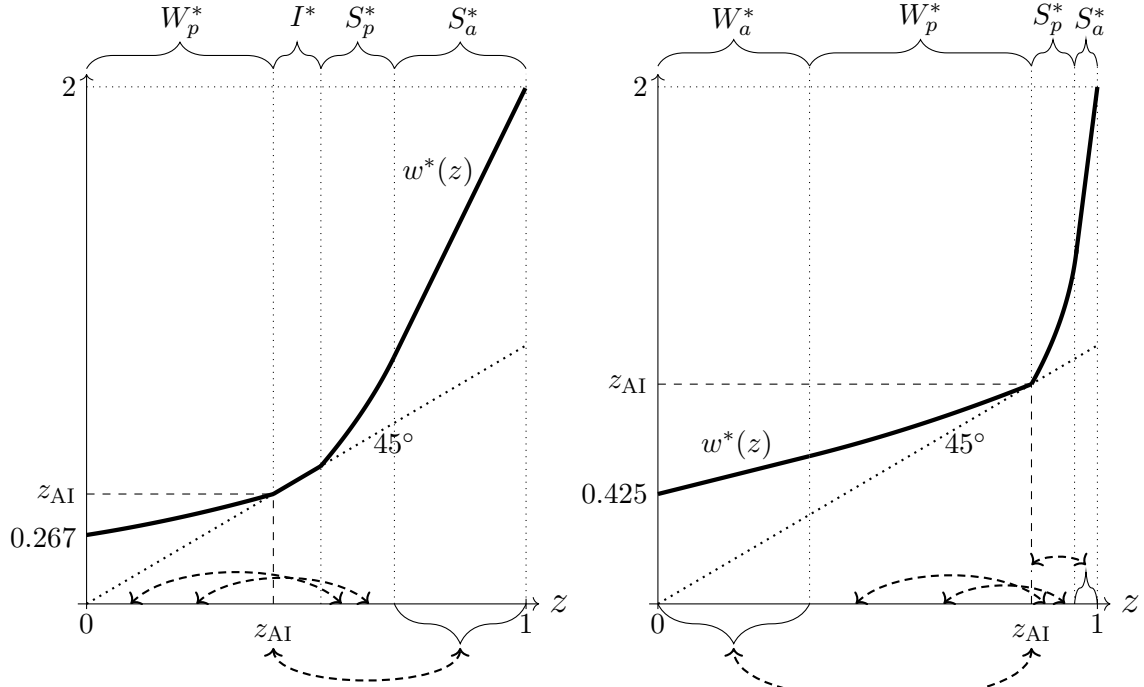
In particular,  $w^*(\sup W^*) = \sup W_p^*$ ,  $w^*(z_{AI}) = z_{AI}$ , and  $w^*(\inf S^*) = \inf S_p^*$ .

Figure 4 illustrates the post-AI equilibrium in two scenarios. In panel (a), AI agents serve only as workers and independent producers but not as solvers. In panel (b), AI agents perform all three roles. In both cases, AI agents are the most knowledgeable workers, so they are assisted by the most knowledgeable human solvers. In panel (b), AI agents are also the least knowledgeable solvers, so they assist the production work of the least knowledgeable human workers. Regardless of the scenario, humans with knowledge  $z_{AI}$  earn their output as independent producers because they are perfect substitutes for AI agents (each priced at  $r^* = z_{AI}$ ).

Proposition 2 has three parts. The first part states the basic properties of the equilibrium. The second part describes how firms use AI agents as a function of their knowledge. The third part characterizes the equilibrium prices.

Let us consider part one. As in the pre-AI world, the post-AI equilibrium is unique and efficient, exhibits occupational stratification, and features positive assortative matching. The remaining properties follow from occupational stratification, positive assortative matching, and AI's scalability. Indeed, because compute is abundant relative to time, some AI agents must engage in independent production. Hence, occupational stratification implies no worker is more knowledgeable than AI, and no solver is less knowledgeable than AI (i.e.,  $W^* \preceq \{z_{AI}\} \preceq S^*$ ). Moreover, by positive assortative matching, if AI is used as a worker, then it is supervised by the most knowledgeable solvers (i.e.,  $S_p^* \preceq S_a^*$ ), and if AI is used as a solver, then it assists the least knowledgeable workers (i.e.,  $W_a^* \preceq W_p^*$ ).

For part two, the result that AI agents are necessarily used as workers when  $z_{AI} \in W$  is intuitive. Indeed, AI agents must be employed either as workers or solvers; otherwise, they would not interact with humans, so humans would match exactly as in the pre-AI equilibrium, preserving pre-AI wages. This would imply  $w^*(z_{AI}) = w(z_{AI}) > z_{AI}$ , a contradiction.



(a) AI is used as a worker but not as a solver      (b) AI is used as a worker and a solver

Figure 4: Illustration of the Post-AI Equilibrium for Two Different Values of  $z_{AI}$

*Notes.* Distribution of knowledge:  $G(z) = z$ . Parameter values: Both panels have  $h = 1/2$ . For panel (a),  $z_{AI} = 0.425$ , while for panel (b),  $z_{AI} = 0.85$ . The thick line depicts the equilibrium wage function. The dashed arrows illustrate the matching between workers and solvers, both humans and AI. Human workers in  $W_p^*$  are endogenously matched with the human solvers in  $S_p^*$  according to  $m^*$ . All the humans in  $W_a^*$  are matched with an AI solver, while all humans in  $S_a^*$  are matched with AI workers.

Hence, if AI agents are not workers, they must be solvers. However, this would imply that, post-AI, the economy has more solvers and weakly fewer workers than before AI (as  $W^* \preceq \{z_{AI}\}$ ), violating market clearing. A similar logic explains why AI agents are necessarily used as solvers in equilibrium when  $z_{AI} \in S$ . Determining whether AI agents are simultaneously used in all three roles when  $z_{AI} \in W$  or  $z_{AI} \in S$  is not fundamental for our main results. For this reason, we relegate such details to the Online Appendix.

Finally, the third part of the proposition addresses equilibrium prices. The result that  $r^* = z_{AI}$  follows because some compute must be used for independent production, and single-layer automated firms must break even. The wages of the humans employed by  $tA$  and  $bA$  firms—those in  $W_a^*$  and  $S_a^*$ —are determined by the zero-profit conditions of these firms, along with the fact that  $r^* = z_{AI}$ :

$$\begin{aligned} n(z)[z_{AI} - w^*(z)] - r^* &= 0 \implies w^*(z) = z_{AI}(1 - 1/n(z)), \text{ for any } z \in W_a^* \\ n(z_{AI})[z - r^*] - w^*(z) &= 0 \implies w^*(z) = n(z_{AI})(z - z_{AI}), \text{ for any } z \in S_a^* \end{aligned}$$

The wages of the human workers and solvers hired by  $nA$  firms—those in  $W_p^*$  and  $S_p^*$ —are derived following a similar logic as in the pre-AI equilibrium. The only difference is that the most



knowledgeable workers and least knowledgeable solvers earn their expected output as independent producers (while they earn strictly more than that in the pre-AI equilibrium). The reason for this is that the equilibrium wage function must be continuous, and there is always some independent production in the post-AI equilibrium (done by AI agents and possibly also humans).

## 5 The Impact of Autonomous AI

We now analyze the impact of an autonomous AI by comparing the pre- and post-AI equilibrium. Figure 5 illustrates this comparison, showing substantial changes in three areas: (i) occupational choice, (ii) the matching of workers and solvers, and (iii) the distribution of labor income. In the Online Appendix, we further examine how these changes influence the distribution of firm size, productivity, and decentralization.

For brevity, we focus on two cases: (i) "basic" AI, which creates AI agents with the knowledge of pre-AI workers ( $z_{AI} \in \text{int}W$ ), and (ii) "advanced" AI, which creates AI agents with the knowledge of pre-AI solvers ( $z_{AI} \in \text{int}S$ ). The results for the knife-edge case where AI has the knowledge of both ( $z_{AI} \in W \cap S$ ) are in the Online Appendix.

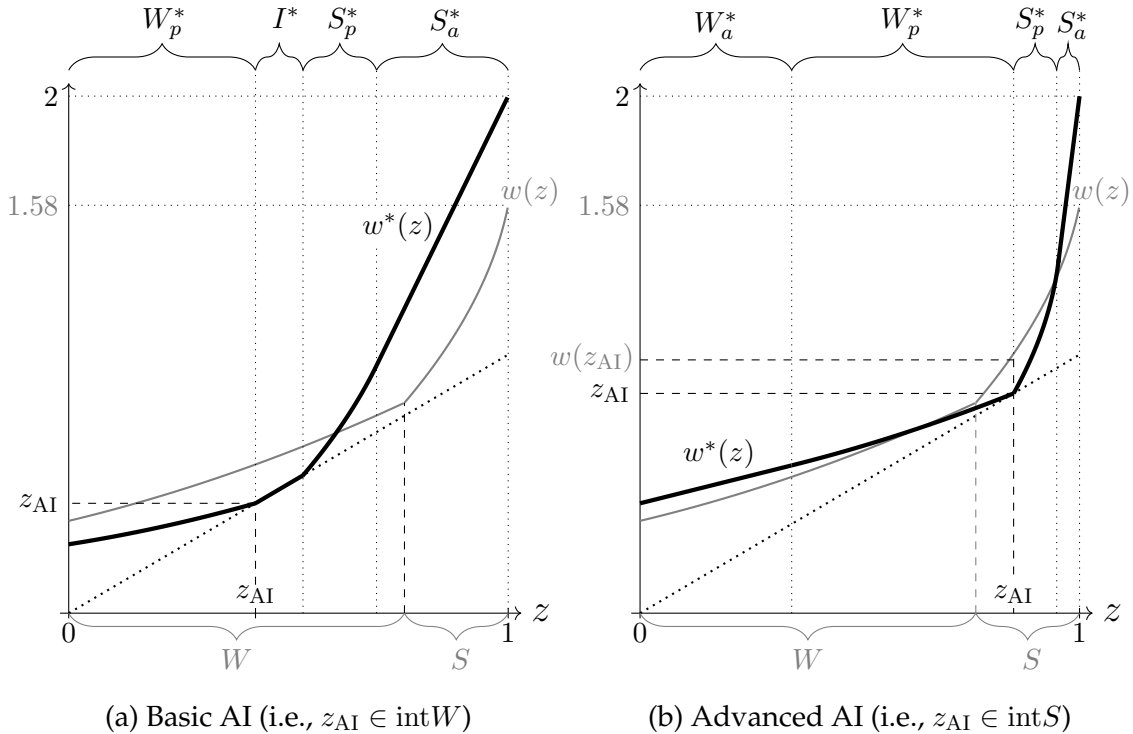


Figure 5: Comparison of the Pre- and Post-AI Equilibrium

Notes. Distribution of knowledge:  $G(z) = z$ . Parameter values: For both panels,  $h = 1/2$  ( $< h_0 = 3/4$ ). For panel (a),  $z_{AI} = 0.425$ , while for panel (b),  $z_{AI} = 0.85$ . In the post-AI equilibrium depicted in panel (a), AI is used as a worker and independent producer. In panel (b), AI is used in all three possible roles.

## 5.1 Occupational Choice and Worker-Solver Matching

We begin by analyzing the effects of AI on occupational choice.

**Proposition 3.** *If  $z_{AI} \in \text{int}W$ , AI displaces humans from routine knowledge work to specialized problem solving, i.e.,  $W^* \subset W$  and  $S^* \supset S$ . In contrast, if  $z_{AI} \in \text{int}S$ , AI displaces humans in the opposite direction, i.e.,  $W^* \supset W$  and  $S^* \subset S$ .*

Intuitively, when AI is basic ( $z_{AI} \in \text{int}W$ ), it serves as a relatively inexpensive technology for routine work, lowering workers' wages. This increases demand for solvers to match with the less expensive, more abundant workers (both humans and AI), drawing marginal pre-AI workers into specialized problem solving. Conversely, when AI is advanced ( $z_{AI} \in \text{int}S$ ), it serves as a relatively inexpensive technology for specialized problem solving. This increases the demand for workers to match with the less expensive, more abundant solvers (both humans and AI), inducing the marginal pre-AI solvers to shift to routine knowledge work.

Next, we examine how AI affects the productivity of pre-AI workers who continue as workers and the span of control of pre-AI solvers who continue as solvers. We define a worker's (average) productivity as her expected output per unit of time. Thus, it is equal to the knowledge of the solver with whom she is matched. Similarly, a solver's span of control is equal to the number of workers (human or AI) under her supervision. Hence, a solver's span of control is increasing in the knowledge of the workers with whom she is matched. For the following result, recall that  $e(z)$  is the pre-AI equilibrium employee function:

**Proposition 4.**

- *If  $z_{AI} \in \text{int}W$ , then the productivity of  $z \in W^* \subset W$  is strictly lower post-AI than pre-AI. Moreover, the span of control of  $z \in S \subset S^*$  is strictly larger post-AI than pre-AI if  $e(z) < z_{AI}$ , and strictly smaller post-AI than pre-AI if  $e(z) > z_{AI}$ .*
- *If  $z_{AI} \in \text{int}S$ , then the productivity of  $z \in W \subset W^*$  is strictly higher post-AI than pre-AI if  $z < e(z_{AI})$ , and strictly lower post-AI than pre-AI if  $z > e(z_{AI})$ . Moreover, the span of control of  $z \in S^* \subset S$  is strictly larger post-AI than pre-AI.*

Proposition 4 is illustrated in Figure 6 when  $z_{AI} \in \text{int}W$  and AI is used as a worker and as an independent producer (but not as a solver). As the figure shows, AI worsens the matches of all pre-AI workers who remain workers (i.e., those humans in  $W_p^*$ ), reducing their productivity. Moreover, AI improves the match of the least knowledgeable solvers who are solvers both pre- and post-AI—increasing their span of control—but worsens the match of the most knowledgeable solvers.

For intuition, consider the case  $z_{AI} \in \text{int}W$  (the argument for  $z_{AI} \in \text{int}S$  is similar). In this case, (i) AI agents are the best workers in the post-AI economy and are therefore assisted by the most knowledgeable solvers (Proposition 2), and (ii) the knowledge of the marginal solvers decreases with AI's introduction (Proposition 3). Both effects worsen the pool of solvers available for those pre-AI workers who remain workers post-AI, thus reducing their productivity.

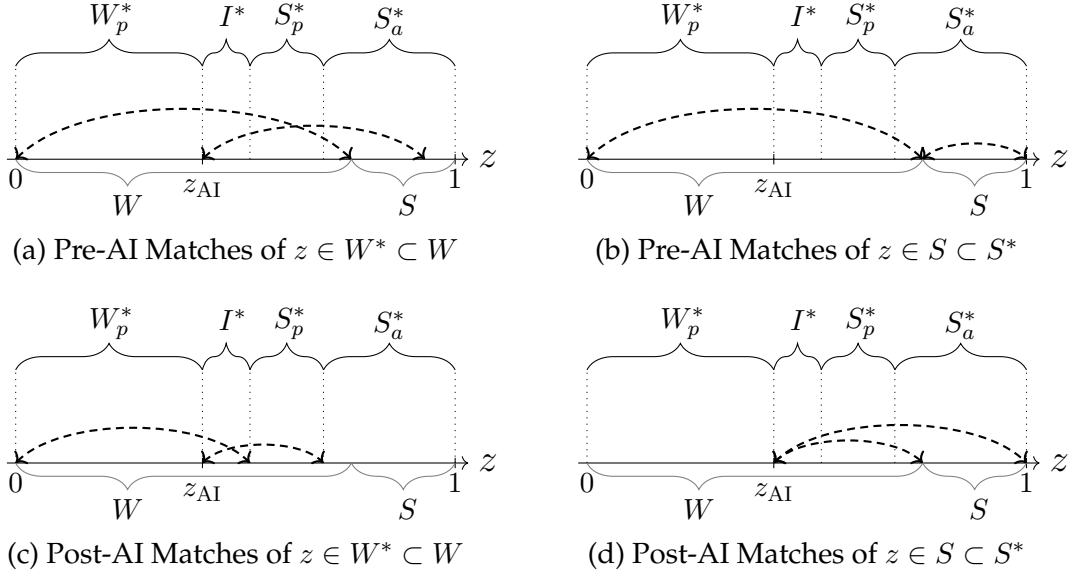


Figure 6: An Illustration of Proposition 4 -  $z_{AI} \in \text{int}W$

Notes. Distribution of knowledge:  $G(z) = z$ . Parameter values:  $h = 1/2$  and  $z_{AI} = 0.425$ . In the post-AI equilibrium shown in the figure, AI is used as a worker and independent producer. The dashed arrows illustrate the matching between workers and solvers.

Regarding the solvers who are not occupationally displaced, the match of the most knowledgeable solvers worsens with AI's introduction because post-AI, they assist production by AI, while pre-AI, they assist humans who are more knowledgeable than AI. However, the match of the least knowledgeable solvers improves, as post-AI, the least knowledgeable workers are assisted by the newly appointed solvers (humans and possibly AI), who are less knowledgeable than the least knowledgeable pre-AI solvers.

There is anecdotal evidence of these reorganizations in professional services. For example, a 2022 survey of M&A lawyers found that AI, by automating routine knowledge work such as document review, has allowed junior lawyers to focus on more complex tasks, including client engagement and legal analysis (Litera, 2022). This aligns with Proposition 3, as the introduction of basic AI pushes marginal human workers into problem-solving roles.

Similarly, in investment banking, Beane and Anthony (2024) document that junior analysts are becoming increasingly distanced from senior partners as AI takes over tasks that once required junior involvement. This is consistent with Proposition 4, which indicates that basic AI forces those workers who remain in their positions to rely on less knowledgeable individuals for support, as the most knowledgeable pre-AI solvers shift their focus to assisting AI-driven production.

## 5.2 Labor Income

We now analyze the effects of AI on total labor income and its distribution. To begin, note that total labor income increases with AI's introduction: if this were not the case, then the set of firms that

existed pre-AI could collectively earn strictly positive aggregate profits post-AI—a contradiction. This also immediately implies that total output increases with the introduction of AI, as it is the sum of labor income and capital income, and pre-AI, capital income is zero in this knowledge economy.

Next, we examine AI’s impact on the distribution of labor income. Analyzing these distributional effects is challenging due to the presence of two potentially opposing forces. On one hand, AI alters firm composition and, therefore, the quality of matches (as shown by Proposition 4). On the other hand, by automating humans with knowledge  $z_{AI}$ , AI shifts the relative scarcities of different knowledge levels, influencing how each firm’s output is divided between workers and solvers.

To begin, note that the individual with knowledge  $z_{AI}$  always loses from AI, i.e.,  $w^*(z_{AI}) < w(z_{AI})$ . Define then  $B$  and  $T$  as the sets of individuals below and above  $z_{AI}$  who benefit from AI, respectively, i.e.,  $B \equiv \{z \in [0, z_{AI}] : w^*(z) > w(z)\}$  and  $T \equiv \{z \in [z_{AI}, 1] : w^*(z) > w(z)\}$ . As we show in the Online Appendix,  $B$  takes the form  $[0, z_b)$ , while  $T$  takes the form  $(z_t, 1]$  for some  $z_b \geq 0$  and  $z_t \leq 1$ . In other words, the individuals who gain from AI are necessarily located at the extremes of the knowledge distribution. The next proposition characterizes when  $B$  and  $T$  are non-empty:

**Proposition 5.** *There are winners at the bottom if AI is good enough, i.e.,  $B \neq \emptyset$  if and only if  $z_{AI} > \bar{z}_{AI}$ , where  $\bar{z}_{AI} \in \text{int}W$ . In contrast, there are always winners at the top, i.e.,  $T \neq \emptyset$  for all  $z_{AI} \in [0, 1]$ .*

Figure 5, at the beginning of this section, provides an illustration of Proposition 5. As shown in panel (a) of the figure, only the most knowledgeable humans benefit when  $z_{AI}$  is relatively low. In contrast, as shown in panel (b), both the least and the most knowledgeable humans benefit when  $z_{AI}$  is relatively high.

We now provide intuition for this proposition. We do this using Figure 7, which builds on Figure 5(a) and illustrates the pre- and post-AI scenarios when  $z_{AI} \in \text{int}W$ . Since  $B$  takes the form  $[0, z_b)$  and  $T$  takes the form  $(z_t, 1]$ , to verify the existence of winners below or above  $z_{AI}$ , it suffices to analyze AI’s impact on the wages of the least and most knowledgeable individuals. Accordingly, Figure 7(a) shows how AI affects the match of individuals with  $z = 0$  and the wages of their match before and after AI, while Figure 7(b) provides the same analysis for individuals with  $z = 1$ .

Consider Figure 7(a) first. AI’s introduction has two opposing effects on the wages of the least knowledgeable individuals. On the one hand, AI reduces their wages by lowering their productivity, as they are now matched with a less knowledgeable solver (a negative “match” effect). On the other hand, AI increases their wages by allowing them to appropriate a larger share of the firm’s output (a positive “share” effect). This occurs because the least knowledgeable solvers (with whom they are matched) now earn their expected output as independent producers—their wages fall on the 45° line—whereas pre-AI, they earned strictly more.

Thus, when  $z_{AI} \in \text{int}W$ , AI benefits the least knowledgeable individuals if the positive share effect outweighs the negative match effect. This happens when  $z_{AI}$  is sufficiently high (i.e., when  $z_{AI} > \bar{z}_{AI}$ , where  $\bar{z}_{AI} \in \text{int}W$ ). Conversely, when  $z_{AI} \in \text{int}S$ , the least knowledgeable individuals always benefit from AI’s introduction because both the match and share effects are positive: AI always improves the match of the least knowledgeable individuals in this case (see Proposition 4).

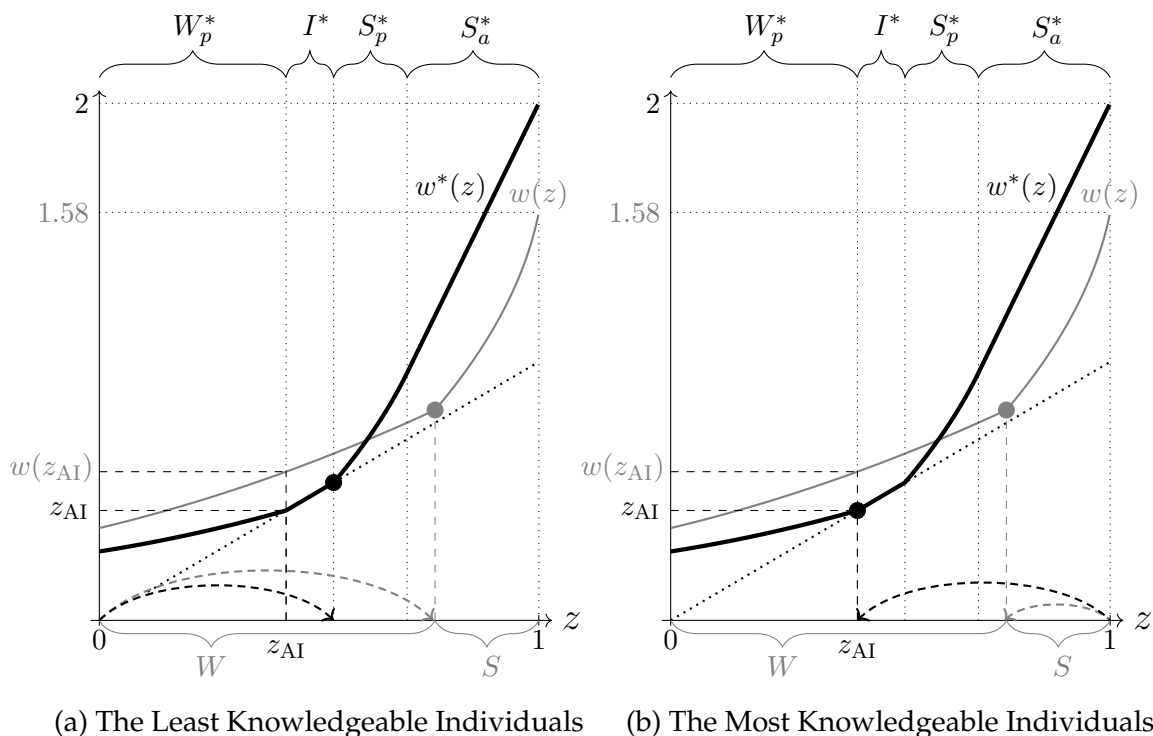


Figure 7: Decomposing the Effects of AI on Wages

Notes. Distribution of knowledge:  $G(z) = z$ . Parameter values:  $h = 1/2 (< h_0 = 3/4)$  and  $z_{AI} = 0.425$ , so AI is used as a worker and independent producer. Panel (a) shows how AI affects the match of  $z = 0$  and the wages of their match before and after AI, while panel (b) provides the same analysis for  $z = 1$ .

Now consider Figure 7(b). As in the case of the least knowledgeable individuals, AI generates both a negative match effect and a positive share effect for the most knowledgeable individuals. However, in this case, the positive share effect always dominates, explaining why there is always a set of winners above  $z_{AI}$ . The key driver behind this result is that  $h < h_0$ , meaning the size of the team each solver assists is large. Consequently, even a small decrease in the wages of the most knowledgeable workers (with whom the most knowledgeable individuals are matched) is amplified by team size, helping the positive share effect outweigh any negative match effect. In a previous version of this paper (Ide and Talamàs, 2025), we show that AI can sometimes harm the most knowledgeable individuals when  $h \geq h_0$ .<sup>19</sup>

## 6 Non-Autonomous AI

In our baseline model, AI is a technology that transforms units of compute into AI agents that can do exactly the same as humans with knowledge  $z_{AI}$ . This type of AI is “autonomous” because it

<sup>19</sup>The finding that the most knowledgeable individuals always benefit from AI’s introduction depends not only on  $h < h_0$  but also on the assumption that  $z_{AI} \in [0, 1)$ , meaning AI is not “superintelligent.” As we show in the Online Appendix, when  $z_{AI} = 1$ , the most knowledgeable humans necessarily lose from AI’s introduction.

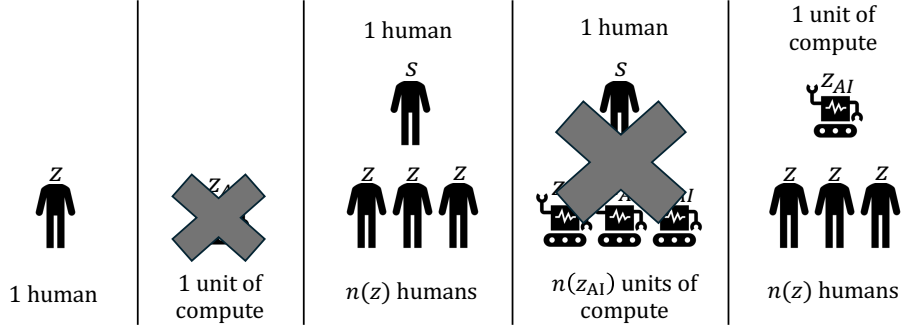


Figure 8: The Three Possible Firm Configurations with Non-Autonomous AI

creates agents that function both as co-workers (pursuing production opportunities) and as co-pilots (providing advice). We focus on this case first, as it generates the most anxiety and debate, driven by technology firms developing and deploying AI agents capable of independent cognitive work.

However, for technical or regulatory reasons, many existing AI systems lack the ability to operate autonomously. In this section, we analyze the impact of non-autonomous AI—a technology that creates agents that operate only as co-pilots. Our main finding is that non-autonomous AI tends to benefit the least knowledgeable individuals the most, whereas autonomous AI primarily benefits the most knowledgeable individuals. However, autonomous AI generates higher overall output than non-autonomous AI.

Formally, the model is the same as in Section 3.1, except that AI agents cannot pursue production opportunities. As a result, single-layer and bottom-automated firms do not emerge in equilibrium since they produce no output (see Figure 8). Thus,  $S_a = \emptyset$  and  $\mu_w = 0$ . In this context, we interpret the amount of compute rented for independent production,  $\mu_i$ , as the compute that remains idle (i.e., not productive).

To ensure a meaningful comparison of equilibrium outcomes between autonomous and non-autonomous AI, we maintain the assumption that firms have at most two layers. Additionally, we assume that compute availability is identical and abundant relative to time in both scenarios. In the Online Appendix, we extend this analysis to different compute levels. To distinguish between cases, we use the superscript “ $\star$ ” for equilibrium outcomes with non-autonomous AI and “ $\ast$ ” for those with autonomous AI.

## 6.1 The Impact of Non-Autonomous AI

Since compute is abundant relative to time and AI is non-autonomous, the equilibrium rental rate of compute is zero,  $r^\star = 0$ . The reason is that some compute must remain idle. However, in contrast to the case of an autonomous AI, the wage of an individual with knowledge  $z_{AI}$  is no longer equal to the equilibrium rental rate of compute ( $w^\star(z_{AI}) \neq r^\star = 0$ ). This difference occurs because an AI agent is no longer a perfect substitute for a human with knowledge  $z_{AI}$ .



For the following result, recall that  $w(z)$  denotes the wage schedule of the pre-AI equilibrium:

**Proposition 6.** *In the presence of a non-autonomous AI, there is a unique equilibrium. The equilibrium is efficient, maximizes labor income, and the rental rate of compute is  $r^* = 0$ . Moreover,*

- *If  $z_{\text{AI}} \leq w(0)$ , then AI is not used in equilibrium, i.e.,  $W_a^* = \emptyset$ . Hence, human occupations and wages coincide with the pre-AI ones.*
- *If  $z_{\text{AI}} > w(0)$ , then only the least knowledgeable individuals use AI as a solver, i.e.,  $W_a^* \subseteq (W_p^* \cup I^* \cup S_p^*)$ , where all these sets are not empty except possibly  $I^*$ .*

*In any case,*

1. *Overall output is strictly higher with autonomous than non-autonomous AI.*
2. *Non-autonomous AI creates losers:*  
 $\exists z \in (0, 1]$  such  $w^*(z) \leq w(z)$  (with strict inequality if  $z_{\text{AI}} > w(0)$ ).
3. *The least knowledgeable benefit more from non-autonomous AI than from no AI or autonomous AI:*  
 $\exists \epsilon > 0$  such that, for all  $z \in [0, \epsilon)$ ,  $w^*(z) \geq \max\{w(z), w^*(z)\}$  (with strict inequality if  $z_{\text{AI}} > w(0)$ ).
4. *The most knowledgeable benefit more from autonomous AI than from non-autonomous AI:*  
 $\exists \epsilon > 0$  such that, for all  $z \in (1 - \epsilon, 1]$ ,  $w^*(z) \leq w(z)$  (with strict inequality if  $z \neq 1$ ).

Figure 9 illustrates Proposition 6.<sup>20</sup> According to this proposition, a non-autonomous AI is used in equilibrium only if it is sufficiently advanced (i.e., if  $z_{\text{AI}} > w(0)$ ). In this case, individuals with knowledge below a certain threshold work in automated firms—using AI as a solver or co-pilot—while those above the threshold work in non-automated firms. Moreover, total output with non-autonomous AI is strictly lower than with autonomous AI because non-autonomy imposes a binding constraint on compute use, leaving some of it idle.

In terms of wages, the introduction of non-autonomous AI creates both winners and losers, similar to autonomous AI. However, the distribution of gains and losses changes dramatically. Non-autonomous AI always benefits the least knowledgeable individuals relative to the pre-AI equilibrium. Moreover, these benefits exceed those of autonomous AI. In contrast, non-autonomous AI may lead to gains or losses compared to the pre-AI scenario for the most knowledgeable individuals. In any case, their wages with non-autonomous AI are always lower than with autonomous AI.

Intuitively, non-autonomous AI benefits the least knowledgeable individuals the most because it functions solely as a solver. This eliminates competition between these individuals and AI for production work and enables them to solve complex problems without human assistance. Moreover, non-autonomous AI's lower opportunity cost—stemming from its inability to produce independently—enhances this advantage by allowing the least knowledgeable to capture a larger share of AI-generated benefits. For the most knowledgeable individuals, the opposite occurs because non-autonomous AI cannot handle routine work and may compete with them in advisory roles.

<sup>20</sup>In the Online Appendix, we also examine how non-autonomous AI affects (i) human occupational choices, (ii) the distribution of firm size, productivity, and decentralization, and (iii) the productivity and span of control of workers and solvers who remain in their positions, compared to the pre-AI equilibrium.

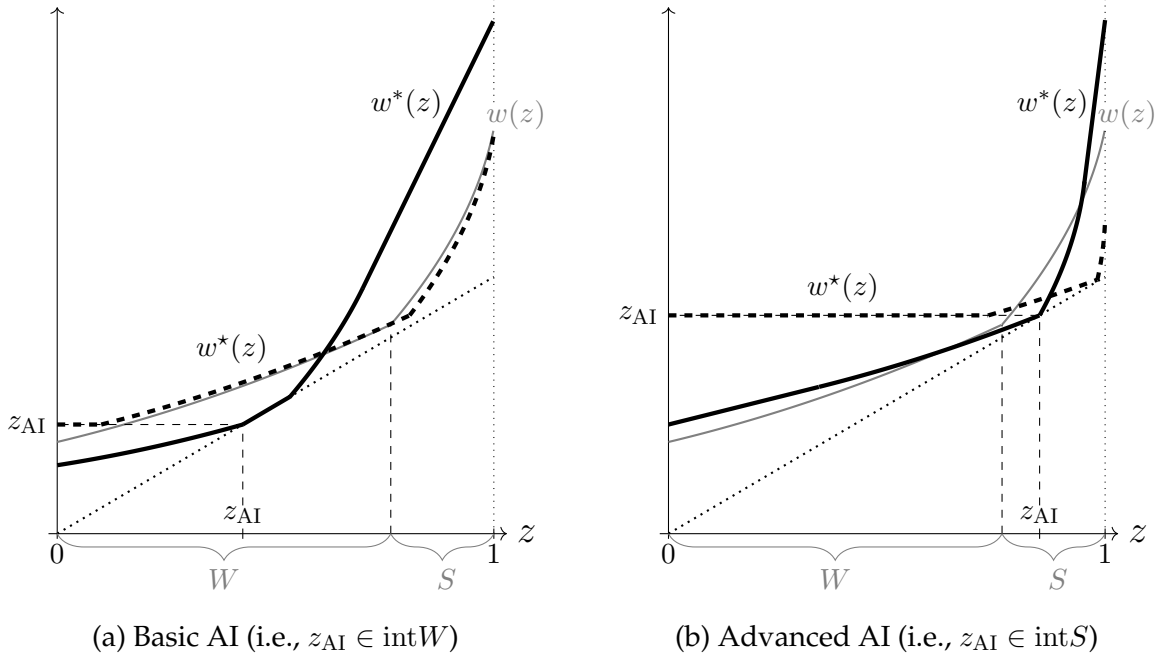


Figure 9: Non-Autonomous AI vs. Autonomous AI vs. No AI

Notes. Distribution of knowledge:  $G(z) = z$ . Parameter values:  $h = 1/2$ . Moreover, for panel (a),  $z_{AI} = 0.425$ , while for panel (b),  $z_{AI} = 0.85$ . The thick gray line represents the pre-AI equilibrium wage function  $w$ . The thick black dashed line depicts the equilibrium wage function with non-autonomous AI,  $w^*$ . The thick black line represents the equilibrium wage function with autonomous AI,  $w^*$ .

As noted earlier, we limit firms to a maximum of two layers to ensure a meaningful comparison of equilibrium outcomes between autonomous and non-autonomous AI. This restriction provides an additional benefit: it implies that problems within the organization can be escalated only once, effectively introducing a cost of seeking help beyond the helping cost  $h$  (which is borne solely by the recipient of the question). Notably, this implies that even with free compute, individuals with slightly less knowledge than AI prefer to seek help from another human rather than from AI (see Figure 9), as the probability that AI provides them with helpful assistance is low.<sup>21</sup>

We believe this is a reasonable outcome, as a similar situation would arise if problems could be escalated multiple times but the agents seeking assistance bore part of the time cost associated with requesting help. Without these constraints—such as with an arbitrary and endogenous number of layers and no time cost for the asking party—all individuals less knowledgeable than AI would always rely on the system for help. In this scenario, introducing AI effectively shifts the knowledge of all agents with  $z < z_{AI}$  towards  $z_{AI}$ . Nevertheless, the key result of this section—that the least knowledgeable individuals prefer non-autonomous AI while the most knowledgeable favor autonomous

<sup>21</sup>The equilibrium is continuous with respect to the availability of compute, so all our results hold when the price of compute is strictly positive but sufficiently close to zero. In the Online Appendix, we also examine the case where compute is significantly less abundant. The results of this section extend to that scenario as well, albeit with certain qualifications.

AI—remains valid even under this alternative case.<sup>22</sup>

## 7 Discussion and Final Remarks

To conclude, we discuss additional implications of our analysis and link our findings to emerging evidence and stylized facts about AI’s effects on the labor market.

*Co-pilots vs. Co-workers.*—Our analysis uncovers distinct labor market implications when comparing AI co-pilots—which provide problem-solving assistance—to AI co-workers—which independently execute production work. Broadly speaking, AI co-pilots primarily benefit less knowledgeable individuals, reducing inequality in performance and earnings. In contrast, AI co-workers offer the greatest advantages to highly knowledgeable individuals, amplifying the value of their expertise. Co-pilots emerge in equilibrium when AI is either (i) not autonomous or (ii) autonomous but sufficiently advanced, whereas co-workers can arise for any knowledge level but only in the case in which AI is autonomous.

These findings are particularly relevant given the emerging evidence on AI’s labor market effects. Several studies suggest that AI disproportionately benefits individuals with the least knowledge and reduces performance inequality (e.g., Dell’Acqua et al., 2023; Noy and Zhang, 2023; Peng et al., 2023; Wiles et al., 2023; Caplin et al., 2024; Brynjolfsson et al., 2025).<sup>23</sup> However, these studies focus on AI co-pilots. While our framework supports these conclusions, it also highlights that they may not extend to autonomous AI that can perform independent knowledge work, such as conducting research, drafting documents, or performing data analysis.

For instance, Toner-Rodgers (2024) shows that an AI that automates “idea-generation tasks” in material sciences, reallocate human researches towards evaluating AI-generated ideas—a more complex task. Consistent with our model’s predictions for basic autonomous AI, top scientists benefit the most because they can leverage their domain expertise to prioritize promising AI suggestions. Conversely, less experienced researchers waste resources testing false positives.

Moreover, non-experimental evidence suggests that firms perceive current iterations of generative AI more as a basic co-worker than as a co-pilot. For instance, using high-frequency job posting data from LinkUp, Berger et al. (2024) document a marked decline in job postings for white-collar roles requiring low levels of knowledge, such as entry-level office jobs, following the introduction of ChatGPT in November 2022. At the same time, job postings for high-knowledge roles, such as

---

<sup>22</sup>With more than two layers, an advanced non-autonomous AI could further amplify the expertise of the most knowledgeable individuals by giving them access to workers that require less help (as workers would first seek AI assistance before turning to them). However, the key insight remains: the most knowledgeable individuals still prefer autonomous AI over non-autonomous AI, as the former can also perform routine work for them.

<sup>23</sup>For instance, Brynjolfsson et al. (2025) reports that deploying an AI-based conversational assistant significantly improved productivity for less-skilled agents in a Fortune 500 customer support center, with minimal impact on experienced ones. Similarly, Caplin et al. (2024) reported that AI assistance was more beneficial for lower-ability individuals than for higher-ability ones when tasked with estimating whether people in photographs appeared to be over 21 years old.

executive positions, increased. This trend indicates that firms view generative AI as a technology that complements high-skilled knowledge workers while substituting for their low-skilled counterparts.

Further empirical studies that disentangle and estimate the distinct labor market effects of AI co-workers and AI co-pilots represent an important avenue for future research.

*AI vs. Traditional Robots and Physical AI.*—Automation technologies that replace manual labor, such as traditional robots and physical AI, differ fundamentally from those that automate cognitive work, such as AI and enterprise software. This difference arises because physical machines must be tailored to their specific operating environments, making them best described as a “putty-clay” technology. In contrast, digital machines—such as AI agents—exhibit “putty-putty” characteristics because they rely on compute, which is general-purpose and can be rented on demand. This allows software improvements to be immediately deployed across all available hardware.<sup>24</sup>

A key implication of this distinction is that traditional robots and physical AI are likely to exist in vintages with varying capabilities, while compute is allocated to the most advanced and highest-performing digital machines.

*AI vs. Offshoring and Immigration.*—The shock induced by AI is different from classical labor shocks such as offshoring or immigration. Non-autonomous AI, for instance, has no clear analog in these frameworks. Moreover, even for autonomous AI, the introduction of the technology diverges significantly from offshoring or immigration. The key distinction lies in AI’s scalability. Once knowledge is encoded into an AI system, it can be deployed across all units of compute, which are widely available. This scalability amounts to the introduction of a large population of homogeneous agents.

In contrast, large human populations are inherently heterogeneous in terms of tacit knowledge—different individuals undoubtedly have different experiences. Consequently, offshoring or immigration can involve the introduction of a relatively small population of homogeneous agents or the introduction of a large population of heterogeneous agents, but not the introduction of a large population of homogeneous agents. As we formally show in the Online Appendix, this difference leads to qualitatively different outcomes.

*AI vs. ICTs.*—A significant implication of our model is that not all firms adopt AI to *automate knowledge work*. This characteristic makes AI, as an automation technology, fundamentally different from earlier waves of digitalization—such as the introduction of computers or email—which saw near-universal adoption across firms. Partial adoption is a common characteristic of automation technologies. Both empirical observations and standard models of automation (e.g., [Acemoglu and Loebbing, 2024](#)) suggest that it is often not cost-effective to automate all tasks that can potentially be automated.

For autonomous AI, partial adoption occurs because deploying AI involves a strictly positive opportunity cost: the foregone output that an autonomous AI agent could generate as an independent producer. As a result, adopting autonomous AI may not be economically viable for all firms, consis-

---

<sup>24</sup>Putty-clay models assume that capital and technology are flexible only prior to investment decisions, whereas putty-putty models allow for flexibility both before and after investments are made ([Atkeson and Kehoe, 1999](#); [Martinez, 2025](#)).

tent with the evidence presented by [Svanberg et al. \(2024\)](#). For non-autonomous AI, partial adoption arises even when the opportunity cost is zero. This is because AI's advice is useless in firms with very knowledgeable human workers.

However, the value of AI extends beyond automating knowledge work. Thus, firms may adopt AI technologies for purposes unrelated to automation, such as reducing communication costs or lowering knowledge acquisition costs. The effects of such technologies have been extensively studied (see, e.g., [Garicano and Rossi-Hansberg, 2006](#); [Antràs et al., 2008](#); [Bloom et al., 2014](#); [Garicano and Rossi-Hansberg, 2015](#)), and are fundamentally different from the ones of AI as an automation technology.<sup>25</sup>

*AI in Advanced vs. Developing Economies.*— An important implication of our analysis is that AI's effects depend on its capabilities relative to human pre-AI occupations. Since these occupations are endogenous—shaped by the knowledge of the population and the level of communication technologies—the same AI technology can lead to different outcomes in advanced and developing economies.

Indeed, in economies with superior communication technologies or a more knowledgeable population, the knowledge threshold required to become a solver in the pre-AI equilibrium is higher (see the Online Appendix for details). As a result, the same autonomous AI technology might be considered “basic” in an advanced economy—displacing humans from routine knowledge work to complex problem-solving—while being regarded as “advanced” in a developing economy, thereby displacing humans in the opposite direction.

*Autonomy vs. No Autonomy: Policy Implications.*— The rise of AI and its potential to transform labor markets has captured policymakers' attention, as evidenced by the recent report of [Council of Economic Advisers \(2024\)](#) and the [EU Artificial Intelligence Act \(2024\)](#). Our findings reveal key tradeoffs in regulating AI autonomy. While autonomous AI generates higher overall output than non-autonomous AI, it also exacerbates labor income inequality.

Using this framework to quantify the tradeoff between output and inequality associated with autonomous and non-autonomous AI is an important direction for future research.

## References

- ACEMOGLU, D. (2024): “The Simple Macroeconomics of AI,” Working Paper.
- ACEMOGLU, D. AND D. AUTOR (2011): “Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings,” Elsevier, vol. 4 of *Handbook of Labor Economics*, 1043–1171.
- ACEMOGLU, D., F. KONG, AND P. RESTREPO (2024): “Tasks At Work: Comparative Advantage, Technology and Labor Demand,” NBER Working Paper No. 32872.
- ACEMOGLU, D. AND J. LOEBBING (2024): “Automation and Polarization,” Working Paper.

---

<sup>25</sup>For instance, better communication technologies—modeled as a reduction in the communication cost  $h$ —always displace humans from specialized problem-solving towards routine work as the very best solvers can now supervise larger teams. In contrast, the displacement generated by AI as an automation technology depends on its autonomy and its problem-solving capabilities.

- ACEMOGLU, D. AND P. RESTREPO (2018): “The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment,” *American Economic Review*, 108, 1488–1542.
- (2022): “Tasks, Automation, and the Rise in U.S. Wage Inequality,” *Econometrica*, 90, 1973–2016.
- ADHVARYU, A., V. BASSI, A. NYSHADHAM, J. A. TAMAYO, AND N. TORRES (2023): “Organizational Responses to Product Cycles,” NBER Working Paper No. 31582.
- AGHION, P., B. F. JONES, AND C. I. JONES (2017): “Artificial Intelligence and Economic Growth,” NBER Working Paper No. 23928.
- AGRAWAL, A., J. GANS, AND A. GOLDFARB (2019): “Introduction,” in *The Economics of Artificial Intelligence: An Agenda*, ed. by A. Agrawal, J. Gans, and A. Goldfarb, University of Chicago Press, 1–19.
- ALEKSEEVA, L., J. AZAR, M. GINÈ, AND S. SAMILA (2024): “AI Adoption and the Demand for Managerial Expertise,” SSRN Working Paper No. 4986379.
- ALFRED SLOAN (1924): “The Most Important Thing I Ever Learned about Management,” *System: The Magazine of Business*, 194, 140–141.
- AMODEI, D. (2024): “Machines of Loving Grace: How AI Could Transform the World for the Better,” October, 2024. Available at <https://darioamodei.com/machines-of-loving-grace> (accessed October 14, 2024).
- ANTRÀS, P., L. GARICANO, AND E. ROSSI-HANSBERG (2006): “Offshoring in a Knowledge Economy,” *The Quarterly Journal of Economics*, 121, 31–77.
- (2008): “Organizing Offshoring: Middle Managers and Communication Costs,” in *The Organization of Firms in a Global Economy*, ed. by E. Helpman, D. Marin, and T. Verdier, Harvard University Press., 311–339.
- ATKESON, A. AND P. J. KEHOE (1999): “Models of Energy Use: Putty-Putty versus Putty-Xlay,” *American Economic Review*, 89, 1028–1043.
- AUTOR, D. (2024): “Applying AI to Rebuild Middle Class Jobs,” NBER Working Paper No. 32140.
- AUTOR, D. H., F. LEVY, AND R. J. MURNANE (2003): “The Skill Content of Recent Technological Change: An Empirical Exploration\*,” *The Quarterly Journal of Economics*, 118, 1279–1333.
- BEANE, M. AND C. ANTHONY (2024): “Inverted Apprenticeship: How Senior Occupational Members Develop Practical Expertise and Preserve Their Position When New Technologies Arrive,” *Organization Science*, 35, 405–431.
- BERGER, P. G., W. CAI, L. QIU, AND C. XINYI SHEN (2024): “Employer and Employee Responses to Generative AI: Early Evidence,” SSRN Working Paper No. 4874061.
- BLOOM, N., L. GARICANO, R. SADUN, AND J. VAN REENEN (2014): “The Distinct Effects of Information Technology and Communication Technology on Firm Organization,” *Management Science*, 60, 2859–2885.
- BOMMASANI, R., D. A. HUDSON, E. ADELI, R. ALTMAN, S. ARORA, S. VON ARX, M. S. BERNSTEIN, J. BOHG, A. BOSSELUT, E. BRUNSKILL, ET AL. (2022): “On the Opportunities and Risks of Foundation Models,” ArXiv preprint arXiv:2108.07258.
- BRYNJOLFSSON, E. (2022): “The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence,” *Daedalus*, 151, 272–287.
- BRYNJOLFSSON, E., D. LI, AND L. RAYMOND (2025): “Generative AI at Work,” *The Quarterly Journal of Economics*.
- BRYNJOLFSSON, E. AND A. MCAFEE (2016): *The Second Machine Age*, WW Norton.
- CAICEDO, S., J. LUCAS, ROBERT E., AND E. ROSSI-HANSBERG (2019): “Learning, Career Paths, and the Distribution of Wages,” *American Economic Journal: Macroeconomics*, 11, 49–88.
- CALIENDO, L., G. MION, L. D. OPROMOLLA, AND E. ROSSI-HANSBERG (2020): “Productivity and Organiza-



- tion in Portuguese Firms,” *Journal of Political Economy*, 128, 4211–4257.
- CALIENDO, L., F. MONTE, AND E. ROSSI-HANSBERG (2015): “The Anatomy of French Production Hierarchies,” *Journal of Political Economy*, 123, 809–852.
- CALIENDO, L. AND E. ROSSI-HANSBERG (2012): “The Impact of Trade on Organization and Productivity,” *The Quarterly Journal of Economics*, 127, 1393–1467.
- CAPLIN, A., D. J. DEMING, S. LI, D. J. MARTIN, P. MARX, B. WEIDMANN, AND K. J. YE (2024): “The ABC’s of Who Benefits from Working with AI: Ability, Beliefs, and Calibration,” NBER Working Paper No. 33021.
- CARMONA, G. AND K. LAOHAKUNAKORN (2024): “Improving the Organization of Knowledge in Production by Screening Problems,” *Journal of Political Economy*, 132, 1290–1326.
- COUNCIL OF ECONOMIC ADVISERS (2024): “Chapter 7: An Economic Framework for Understanding Artificial Intelligence,” in *The 2024 Economic Report of the President*, Executive Office of the President of the United States, available at <https://www.whitehouse.gov/cea/written-materials/2024/03/21/the-2024-economic-report-of-the-president/> (accessed October 8, 2024).
- DELL’ACQUA, F., E. MCFOWLAND III, E. R. MOLICK, H. LIFSHITZ-ASSAF, K. KELLOGG, S. RAJENDRAN, L. KRAYER, F. CANDELON, AND K. R. LAKHANI (2023): “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality,” Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013.
- ELOUNDOU, T., S. MANNING, P. MISHKIN, AND D. ROCK (2024): “GPTs are GPTs: Labor market impact potential of LLMs,” *Science*, 384, 1306–1308.
- EU ARTIFICIAL INTELLIGENCE ACT (2024): “Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence,” *Official Journal of the European Union*, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>.
- FARBOODI, M. AND L. VELDKAMP (2023): “Data and markets,” *Annual Review of Economics*, 15, 23–40.
- (2024): “A Model of the Data Economy.” Working Paper.
- FUCHS, W., L. GARICANO, AND L. RAYO (2015): “Optimal Contracting and the Organization of Knowledge,” *The Review of Economic Studies*, 82, 632–658.
- GARICANO, L. (2000): “Hierarchies and the Organization of Knowledge in Production,” *Journal of Political Economy*, 108, 874–904.
- GARICANO, L. AND T. HUBBARD (2007): “Managerial Leverage Is Limited by the Extent of the Market: Hierarchies, Specialization, and the Utilization of Lawyers’ Human Capital,” *The Journal of Law & Economics*, 50, 1–43.
- GARICANO, L. AND T. N. HUBBARD (2009): “Specialization, Firms, and Markets: The Division of Labor Within and Between Law Firms,” *The Journal of Law, Economics, & Organization*, 25, 339–371.
- (2012): “Learning About the Nature of Production from Equilibrium Assignment Patterns,” *Journal of Economic Behavior & Organization*, 84, 136–153.
- (2018): “Earnings Inequality and Coordination Costs: Evidence from US Law Firms,” *The Journal of Law, Economics, and Organization*, 34, 196–229.
- GARICANO, L. AND E. ROSSI-HANSBERG (2004): “Inequality and the Organization of Knowledge,” *American Economic Review Papers and Proceedings*, 94, 197–202.
- (2006): “Organization and Inequality in a Knowledge Economy,” *The Quarterly Journal of Economics*, 121, 1383–1435.
- (2012): “Organizing Growth,” *Journal of Economic Theory*, 147, 623–656.



- (2015): “Knowledge-Based Hierarchies: Using Organizations to Understand the Economy,” *Annual Review of Economics*, 7, 1–30.
- GARICANO, L. AND Y. WU (2012): “Knowledge, communication, and organizational capabilities,” *Organization science*, 23, 1382–1397.
- GOLDFARB, A. AND C. TUCKER (2019): “Digital Economics,” *Journal of Economic Literature*, 57, 3–43.
- GOLDMAN SACHS (2023): “Generative AI: Hype, or Truly Transformative?” July 5, 2023. Available at <https://www.goldmansachs.com/pdfs/insights/pages/top-of-mind/generative-ai-hype-or-truly-transformative/report.pdf> (accessed November 30, 2024).
- GRANT, R. M. (1996): “Toward a Knowledge-Based Theory of the Firm,” *Strategic Management Journal*, 17, 109–122.
- GUMPERT, A., H. STEIMER, AND M. ANTONI (2022): “Firm Organization with Multiple Establishments,” *The Quarterly Journal of Economics*, 137, 1091–1138.
- HADFIELD, G. (2025): “ASSA 2025 Panel: AI and the Future of Work,” ASSA 2025 Webcasts of Select Sessions. Available at <https://www.aeaweb.org/webcasts/2025/ai-future-of-work> (accessed January 27, 2025).
- HUMLUM, A. AND E. VESTERGAARD (2024): “The Adoption of ChatGPT,” University of Chicago, Becker Friedman Institute for Economics Working Paper.
- IDE, E. AND E. TALAMÀS (2024): “Artificial Intelligence in the Knowledge Economy,” ArXiv preprint arXiv:2312.05481 (version 6).
- (2025): “Artificial Intelligence in the Knowledge Economy,” ArXiv preprint arXiv:2312.05481 (version 10).
- JOHNSON, S. AND D. ACEMOGLU (2023): *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*, Hachette UK.
- JONES, C. I. (2005): “Chapter 16 - Growth and Ideas,” Elsevier, vol. 1 of *Handbook of Economic Growth*, 1063–1111.
- (2024): “The AI Dilemma: Growth versus Existential Risk,” *American Economic Review: Insights*, 6, 575–90.
- JONES, C. I. AND C. TONETTI (2020): “Nonrivalry and the Economics of Data,” *American Economic Review*, 110, 2819–2858.
- KORINEK, A. AND D. SUH (2024): “Scenarios for the Transition to AGI,” NBER Working Paper No. 32255.
- LI, X., S. CHAN, X. ZHU, Y. PEI, Z. MA, X. LIU, AND S. SHAH (2023): “Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? A Study on Several Typical Tasks,” ArXiv preprint arXiv:2305.05862.
- LITERA (2022): “Litera Technology in M&A Report,” Available at <https://info.litera.com/report-2022-ma-technology-survey.html> (accessed September 5, 2024).
- MARTINEZ, J. (2025): “Putty-Clay Automation,” CEPR Working Paper No. DP16022.
- MCÉLHERAN, K., J. F. LI, E. BRYNJOLFSSON, Z. KROFF, E. DINLERSOZ, L. S. FOSTER, AND N. ZOLAS (2023): “AI Adoption in America: Who, What, and Where,” NBER Working Paper No. 31788.
- MOLL, B., L. RACHEL, AND P. RESTREPO (2022): “Uneven Growth: Automation’s Impact on Income and Wealth Inequality,” *Econometrica*, 90, 2645–2683.
- MURO, M., J. WHITON, AND R. MAXIM (2019): “What Jobs are Affected by AI?” Tech. rep., Brookings Institution.
- NAVANI, D. A. (2023): “The Commoditization of LLMs,” *Communications of the ACM*, September 12,

2024. Available at <https://cacm.acm.org/blogcacm/the-commoditization-of-llms/> (accessed November 25, 2024).
- NONAKA, I. (1991): “The Knowledge-Creating Company,” *Harvard Business Review*, 69, 96–104.
- NORDHAUS, W. D. (2007): “Two Centuries of Productivity Growth in Computing,” *The Journal of Economic History*, 67, 128–159.
- NORI, H., Y. T. LEE, S. ZHANG, D. CARIGNAN, R. EDGAR, N. FUSI, N. KING, J. LARSON, Y. LI, W. LIU, ET AL. (2023): “Can Generalist Foundation Models Outcompete Special-Purpose Tuning? Case Study in Medicine,” ArXiv preprint arXiv:2311.16452.
- NOY, S. AND W. ZHANG (2023): “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence,” *Science*, 381, 187–192.
- NVIDIA (2025): “What is Physical AI?” Available at <https://www.nvidia.com/en-us/glossary/generative-physical-ai/> (accessed December 27, 2025).
- PANCS, R. (2018): *Lectures on Microeconomics: The Big Questions Approach*, MIT Press.
- PENG, S., E. KALLIAMVAKOU, P. CIHON, AND M. DEMIRER (2023): “The Impact of AI on Developer Productivity: Evidence from GitHub Copilot,” ArXiv preprint arXiv:2302.06590.
- POLANYI, M. (1966): *The Tacit Dimension*, Doubleday & Co.
- ROMER, P. M. (1990): “Endogenous Technological Change,” *Journal of Political Economy*, 98, S71–S102.
- (1992): “Two Strategies for Economic Development: Using Ideas and Producing Ideas,” *The World Bank Economic Review*, 6, 63–91.
- ROSENBUSH, S. (2024): “AI Agents Can Do More Than Answer Queries. That Raises a Few Questions,” *The Wall Street Journal*, October 16, 2024. Available at <https://www.wsj.com/articles/ai-agents-can-do-more-than-answer-queries-that-raises-a-few-questions-15009853> (accessed October 17, 2024).
- ST. LOUIS FED (2024): “Job Polarization,” *The FRED Blog*, April 28, 2016 (Data updated through December 2024). Available at <https://fredblog.stlouisfed.org/2016/04/job-polarization/> (accessed January 24, 2025).
- STEBBINGS, H. (2024a): “Aravind Srinivas: Will Foundation Models Commoditise?” *20VC with Harry Stebbings*, Episode #1161. Available at <https://www.thetwentyminutevc.com/aravind-srinivas> (accessed October 8, 2024).
- (2024b): “Sarah Tavel: Will Foundation Models Be Commoditised?” *20VC with Harry Stebbings*, Episode #1149. Available at <https://www.thetwentyminutevc.com/sarah-tavel> (accessed October 8, 2024).
- SVANBERG, M., W. LI, M. FLEMING, B. GOEHRING, AND N. THOMPSON (2024): “Beyond AI Exposure: Which Tasks are Cost-Effective to Automate with Computer Vision?” SSRN Working Paper No. 4700751.
- TONER-RODGERS, A. (2024): “Artificial Intelligence, Scientific Discovery, and Product Innovation,” ArXiv preprint arXiv:2412.17866.
- WAISBERG, N. AND A. HUDEK (2021): *AI for Lawyers: How Artificial Intelligence is Adding Value, Amplifying Expertise, and Transforming Careers*, John Wiley & Sons.
- WANG, S. AND S. XU (2024): “16 Changes to the Way Enterprises Are Building and Buying Generative AI,” *Andreessen Horowitz*, March 21, 2024. Available at <https://a16z.com/generative-ai-enterprise-2024/> (accessed December 13, 2024).
- WILES, E., Z. T. MUNYIKWA, AND J. J. HORTON (2023): “Algorithmic Writing Assistance on Jobseekers’ Resumes Increases Hires,” NBER Working Paper No. 30886.
- ZEIRA, J. (1998): “Workers, Machines, and Economic Growth,” *The Quarterly Journal of Economics*, 113, 1091–1117.