The Turing Valley: How AI Capabilities Shape Labor Income*

Enrique Ide & Eduard Talamàs

January 15, 2026

Abstract

Current AI systems are significantly better than humans in some intelligence dimensions but remain weaker than humans in others. Guided by the long-standing vision of machine intelligence embodied by the Turing Test, AI developers increasingly seek to eliminate this "jagged" nature of AI by pursuing Artificial General Intelligence (AGI) with consistently high performance across domains. This pursuit has sparked a heated debate, with leading economists arguing that AGI risks eroding the value of human capital by turning machines into closer substitutes for labor. We contribute to this debate by characterizing how different paths of AI progress shape labor income in a multidimensional knowledge economy. We show that AI improvements in its strong dimensions always increase labor income, but the effects of AI progress in its weak dimensions fundamentally depends on the nature of human–AI communication. When communication is able to integrate partial solutions, AI improvements in its weak dimensions decrease labor income. In this case, as long as humans are sufficiently strong in one dimension, labor income is maximized by a deliberately jagged form of AI. In contrast, when communication is only able to share full solutions, AI improvements in its weak dimensions increase labor income provided that AI is sufficiently stronger than humans in its strong dimensions. In this case, if humans are sufficiently weak in one dimension, labor income is maximized when AI attains high performance across all dimensions. These results highlight the importance of empirically assessing the properties of human–AI communication for understanding the labor-market consequences of progress toward AGI.

*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 have benefited from the comments of Daron Acemoglu, Ricardo Alonso, Jeff Ely, Luis Garicano, Johannes Horner, Todd Lensman, Jin Li, Víctor Martínez de Albéniz, Cristóbal Otero, Gerard Padró-i-Miquel, Pascual Restrepo, Martin Rotemberg, Raffaella Sadun, Jaume Ventura and Xavier Vives. We acknowledge the financial support of IESE through the High Impact Initiative-course 2024/2026. We declare we have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

The development of Artificial Intelligence (AI) is largely guided by the ambition to create Artificial General Intelligence (AGI). Industry leaders describe this goal as eliminating the highly uneven, or "jagged," performance of current systems—AI that excels in some dimensions of intelligence while remaining weak in others—in favor of systems with consistently high performance across domains. As Demis Hassabis, CEO and co-founder of Google DeepMind, notes (Fridman and Hassabis, 2025):

> "[AGI] is quite a high bar: Can we match the cognitive functions that the brain has? [...] It isn't kind of a jagged intelligence where some things it's really good at but other things it's really flawed at. That's what we currently have with today's systems. They're not consistent. So you'd want consistency of intelligence across the board."

This vision echoes the long-standing objective embodied in the Turing Test of machine intelligence (Turing, 1950): building machines that do not fall short in any of the cognitive functions of the brain. While this aspiration is often framed as a technological milestone, leading economists argue that it risks eroding the value of human capital by turning machines into closer substitutes for labor (Brynjolfsson, 2022; Acemoglu and Johnson, 2024; Restrepo, 2025). In particular, Brynjolfsson (2022) argues that because human intelligence is also uneven across dimensions, AI can generate large gains by specializing in dimensions of knowledge where humans are comparatively weak:

> "Not all types of AI are human-like—in fact, many of the most powerful systems are very different from humans. We control the extent to which AI either expands human opportunity through augmentation or replaces humans through automation. We can work on challenges that are easy for machines and hard for humans, rather than hard for machines and easy for humans."

In this paper, we contribute to this debate by studying how advances in AI capabilities across different dimensions shape labor income in the knowledge economy. A natural framework to study this question is the canonical knowledge-economy framework developed by Garicano (2000) and Garicano and Rossi-Hansberg (2004, 2006).[1] This literature has been highly influential in clarifying how tacit knowledge, communication, and hierarchical organization jointly shape labor market outcomes (Garicano and Rossi-Hansberg, 2015). However, it treats knowledge as unidimensional. While this abstraction is appropriate for many questions—including AI's economic impact (Ide and Talamàs, 2025)—it is ill-suited to capture the sharply uneven, or "jagged," nature of current AI systems.

[1]Important contributions to this literature include Antràs et al. (2006), Garicano and Hubbard (2007, 2012), Caliendo and Rossi-Hansberg (2012), Garicano and Rossi-Hansberg (2012), Bloom et al. (2014), Fuchs et al. (2015), Caliendo et al. (2015), Caicedo et al. (2019), Caliendo et al. (2020) and Gumpert et al. (2022).

The innovation of this paper is to extend the knowledge-economy framework to a multidimensional setting to study how AI advances in different directions shape the marginal product of labor—or, equivalently in our setting, labor income. As in the canonical framework, production is organized through hierarchies in which "workers" pursue production opportunities and escalate unsolved problems to "solvers" who specialize in problem solving. As in Ide and Talamàs (2025), computational resources ("machines") can operate as workers or solvers, and they are abundant relative to human time. The novelty of this paper is that problems and knowledge are multidimensional.

The cost of communication is central in the knowledge economy (Garicano, 2000; Garicano and Rossi-Hansberg, 2006; Bloom et al., 2014). We show that, in a multidimensional knowledge economy, the properties of communication are also first-order. To illustrate this point as transparently as possible, we focus on two knowledge dimensions. Following Brynjolfsson (2022), we assume that AI systems are better than humans in one dimension but remain weaker than humans in the other. Our central question is how the equilibrium marginal product of labor changes when AI improves in its strong versus its weak dimension.

We focus on two polar forms of human–AI communication. *Additive communication* is only able to share full solutions: Asking the solver for help is useful when the solver has enough knowledge in all dimensions of the problem. *Superadditive communication* is also able to integrate partial solutions: Asking the solver for help is useful when the solver has enough knowledge in the dimensions of the problem the worker lacks. While the difference between these regimes is immaterial in the unidimensional knowledge economy, our results highlight that—in the multidimensional knowledge economy—it is a central determinant of the consequences of AI progress for labor income.

AI improvements in its strong dimension increase labor income regardless of the communication regime. The intuition is best grasped by focusing on the case in which humans use machines as workers. In this case, the marginal product of labor is proportional to the probability that a human is useful when machines require assistance. When AI improves in its strong dimension, machines only stop asking questions that humans could not help with, thus increasing the marginal product of labor.

The effects of AI improvements in its weak dimension fundamentally depend on the properties of human-AI communication. When communication is superadditive, such improvements unambiguously reduce labor income. Intuitively, in this case, AI improvements in its weak dimension imply that machines only stop asking questions that humans could help with (those that require AI's improvement and AI's knowledge in its strong dimension). Hence, the marginal product of labor unambiguously declines.

In contrast, when communication is additive, AI improvements in its weak dimension have an ambiguous effect on labor income. Intuitively, in this case, AI improvements in its weak dimension imply machines stop asking questions that humans could help with (those that require AI's improvement and knowledge in AI's strong dimension that *both humans and AI have*), but they also stop asking questions that humans could not help with (those that require AI's improvement and knowledge in AI's strong dimension that *AI has but humans don't have*). Hence, the effect on the marginal product of labor depends on the strength of these two forces. However, when AI is sufficiently stronger than humans in its strong dimension, the second effect dominates, so the marginal product of labor unambiguously increases.

The above local comparative statics have natural implications for the AI configuration that maximizes labor income. When communication is *additive* and humans are sufficiently weak in their weak dimension, labor income is maximized when AI attains the highest possible performance in all dimensions, corresponding to AGI. In contrast, when human–AI communication is *superadditive* and humans are sufficiently strong in their strong dimension, labor income is maximized by a deliberately jagged AI—one that performs minimally in the dimension in which humans are strong and maximally in the other dimension. Hence, the nature of human-AI communication is central to determine the implications of the direction of AI progress for the value of human capital.

Connections to Related Literature

This paper is most closely related to Ide and Talamàs (2025), which studies how AI reshapes occupational choice and labor income inequality in a knowledge economy with heterogeneous human expertise. While that framework focuses on human knowledge heterogeneity, it maintains the standard assumption that knowledge is unidimensional. In contrast, in this paper we allow knowledge to be multidimensional in order to capture the "jagged" nature of AI capabilities and to study the economic implications of the transition toward AGI. To keep the analysis tractable, we abstract from human heterogeneity and instead focus on how the effects of AI progress depend on the nature of human–AI communication—an issue that does not arise when knowledge is unidimensional.

Our multidimensional approach provides a unifying perspective on the marginal value of knowledge in organizations. In our setting, labor income rises whenever AI improves in a direction in which the marginal value of knowledge within the human–AI organization exceeds one. In the standard unidimensional knowledge economy, the marginal value of workers' knowledge is always below one, while the marginal value of solvers' knowledge is always above one—and this property drives many of the results and intuitions in this literature (Garicano and Rossi-Hansberg, 2015). In a multidimensional setting, however, this correspondence breaks down: the marginal value of knowledge in a given dimension may exceed one for a worker and be less than one for a solver. Hence, our results highlight that what matters for the marginal value of knowledge is not the agent's role, but how that knowledge affects the value of communication.

Recent studies document the aggregate performance of human–AI teams (e.g., Krakowski et al. 2024; Brynjolfsson et al. 2025; Dell'Acqua et al. 2025), but they typically do not distinguish between additive pooling and superadditive collaboration. Our analysis shows that this distinction is decisive: without it, existing evidence cannot tell apart environments in which the transition toward AGI augments the value of human capital from those in which it erodes it. Our analysis therefore underscores the importance of empirically assessing the additive properties of human–AI communication.

This paper also contributes to the broader literature on the labor-market effects of AI. In canonical task-based models of automation, production aggregates a continuum of tasks through a mechanical aggregator, and technological change operates by shifting unit costs across tasks (e.g., Zeira, 1998; Acemoglu and Restrepo, 2018, 2019, 2022; Graetz and Michaels, 2018; Moll et al., 2022; Acemoglu and Johnson, 2024; Ide, 2025; Althoff and Reichardt, 2026). This structure makes it useful to distinguish between two conceptually different directions of "machine progress." First, improvements that raise the productivity of machines in the tasks they already perform (an intensive-margin, narrowly capital-augmenting change) primarily generate a productivity effect and tend to augment labor in equilibrium. Second, improvements that make machines competitive in tasks previously performed by workers (an extensive-margin change that expands the set of tasks assigned to capital) reallocate tasks away from labor and therefore tend to substitute for labor.

Our framework delivers a similar dichotomy when communication is *superadditive*, so partial solutions can be combined seamlessly across dimensions in close analogy to how tasks are aggregated in the task-based framework. In this case, improvements in AI along dimensions where it is already strong behave like intensive-margin, narrowly capital-augmenting progress and they *augment* labor. In contrast, improvements along dimensions where AI is initially weak behave like a reallocation of tasks away from labor and therefore tend to substitute for labor. However, this analogy breaks down when communication is merely *additive*. In this case, even improvements in AI's weak dimensions can augment labor because limited complementarity in combining partial solutions can make these advances operate as productivity deepening for the human–AI team. This contrast underscores the importance of the nature of human–AI communication in the mapping from AI progress into augmentation or substitution of labor.

2 The Model

We begin by formally describing the model, followed by a discussion of its critical elements.

2.1 The Baseline Setting

The Basics.— There is a unit mass of identical humans, each endowed with one unit of time and exogenous knowledge $\boldsymbol{h} = (h_1, h_2) \in (0,1)^2$. Additionally, there are $\mu > 0$ units of compute, normalized such that one unit of compute is equivalent to one unit of time. An exogenous AI algorithm provides these units of compute with knowledge $\boldsymbol{m} = (m_1, m_2) \in [0,1]^2$. For simplicity, we refer to one unit of compute as a "machine," with each machine having one unit of "time."

There are many identical competitive firms. Production occurs within these firms, which are the residual claimants of output. Time and knowledge are the only inputs. Firms have no fixed costs and a maximum of two layers.

Single-layer firms can be automated or non-automated. Single-layer automated firms rent a single machine to produce. This machine dedicates its full unit of time to pursue a single production opportunity. Each production opportunity is associated with a problem with unknown difficulty $\boldsymbol{x} \equiv (x_1, x_2)$ drawn from a distribution with full support on $[0,1]^2$, cumulative distribution function F and probability density function f. If the knowledge of the machine exceeds the problem's difficulty in both dimensions, i.e., if $m_1 \geq x_1$ and $m_2 \geq x_2$, the machine solves the problem and produces one unit of output. Otherwise, no output is produced. Single-layer non-automated firms are identical to their automated counterparts, except that they hire humans for production instead of renting machines.

Two-layer organizations combine humans and machines, and can be bottom-automated or top-automated. Bottom-automated ("b") firms rent machines to pursue production opportunities and employ a single human to assist with problems the machines cannot solve. Production occurs when the combined knowledge of the human and machine exceeds the problem's difficulty; the specific probability of this event depends on the communication regime, as specified below. Otherwise, no production occurs. Assistance consumes $c \in (0,1)$ units of the human's time. Given that a machine cannot solve the problem it encounters with probability $1 - F(\boldsymbol{m})$, b-firms optimally employ $n(\boldsymbol{m})$ machines to fully utilize the human's time, where:

$$n(\boldsymbol{m}) \equiv \frac{1}{c\big(1 - F(\boldsymbol{m})\big)}$$

Top-automated ("t") firms are similar to b-firms, except the roles are reversed: Humans engage in production, while machines specialize in problem-solving. In this case, consulting with a machine consumes $c \in (0,1)$ units of the machine's time. Consequently, t-firms always employ $n(\boldsymbol{h})$ humans to fully exploit the time of the machine that assists these humans.

Figure 1 depicts the four firm configurations that can potentially arise in equilibrium. Note that a two-layer organization never uses only humans (or only machines) in both layers of the organization, as the human (or machine) engaged in production would solve the same set of problems as the one specialized in problem-solving. A similar reasoning implies that focusing on two-layer organizations is without loss of generality in this setting: organizations with more than two layers cannot solve any more problems than two-layer organizations.

For brevity, we refer to agents (humans or machines) in two-layer organizations as "workers" if they engage in production and as "solvers" if they specialize in problem-solving. We call agents in single-layer organizations "independent producers" as they engage in unassisted production work.

Wages, Prices and Profits.— Let w be the wage of humans and r be the rental rate of machines. All

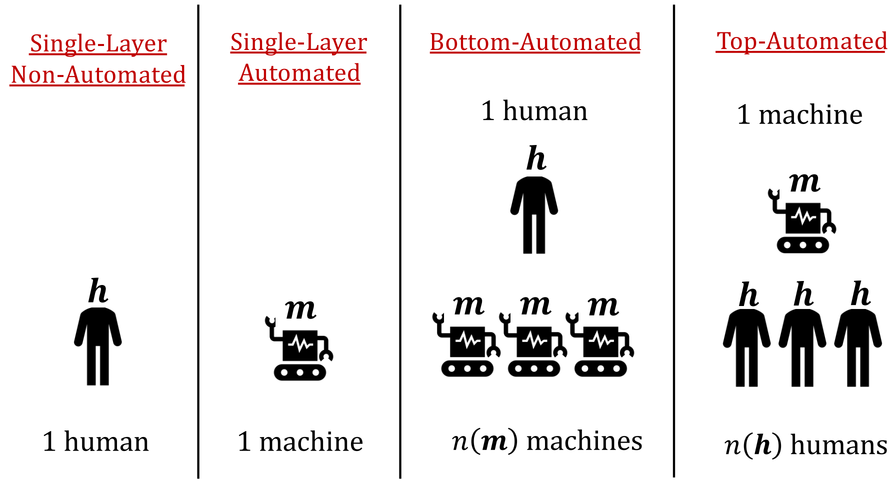


Figure 1: The Four Possible Firm Configurations

agents in this economy are risk neutral and maximize their income. This implies, in particular, that labor supply is inelastic. We normalize the price of each unit of output to one.

The profit of a single-layer firm as a function of its type is given by:

$$\Pi_s = \begin{cases} F(\boldsymbol{h}) - w & \text{if non-automated} \\ F(\boldsymbol{m}) - r & \text{if automated} \end{cases}$$

In other words, the profit of a single-layer non-automated firm is the expected fraction of problems that a human independent producer can solve, $F(\boldsymbol{h})$, minus her wage, w. The profit of a single-layer automated firm is obtained similarly, with the human replaced by a machine.

The profit of a two-layer organization as a function of its type (i.e., b or t) is given as follows:

$$\Pi_b = n(\boldsymbol{m})[F(\boldsymbol{m} \oplus \boldsymbol{h}) - r] - w$$
$$\Pi_t = n(\boldsymbol{h})[F(\boldsymbol{m} \oplus \boldsymbol{h}) - w] - r$$

where $\boldsymbol{m} \oplus \boldsymbol{h}$ denotes the effective combined knowledge of humans and machines. The precise meaning of this combination depends on the communication regime, as described below. In both 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 b-firm, its total expected output is $n(\boldsymbol{m})F(\boldsymbol{m} \oplus \boldsymbol{h})$, as such organization attempts $n(\boldsymbol{m})$ problems and successfully solves a fraction $F(\boldsymbol{m} \oplus \boldsymbol{h})$ of them. In turn, the cost of the resources is $w + n(\boldsymbol{m})r$, as the firm employs one human solver and $n(\boldsymbol{m})$ machine workers.

Human–AI communication.— A key modeling choice concerns how the knowledge of a worker and a solver combines when the worker consults the solver—that is, the nature of the effective combined knowledge $\boldsymbol{m} \oplus \boldsymbol{h}$. We consider two polar cases.

Additive communication only allows the solver to provide a *stand-alone* solution; partial solutions cannot be integrated across dimensions. In this case, a project succeeds if either the worker can solve the entire problem alone or the solver can solve the entire problem alone. Thus, the success probability of a two-layer organization on any given problem, denoted by $\Phi^A(\boldsymbol{h}, \boldsymbol{m})$, is:

$$\Phi^A(\boldsymbol{h}, \boldsymbol{m}) \equiv F(\boldsymbol{h}) + F(\boldsymbol{m}) - F(\boldsymbol{h} \wedge \boldsymbol{m}).$$

where $\boldsymbol{h} \wedge \boldsymbol{m}$ denotes the knowledge overlap between humans and machines $(\min\{h_1, m_1\}, \min\{h_2, m_2\})$.

Superadditive communication allows the worker and solver to combine partial solutions across knowledge dimensions. In this case, a project succeeds whenever each dimension can be covered by at least one of the two agents. Thus, the success probability of a two-layer organization on any given problem, denoted by $\Phi^S(\boldsymbol{h}, \boldsymbol{m})$, is given by:

$$\Phi^S(\boldsymbol{h}, \boldsymbol{m}) \equiv F(\boldsymbol{h} \vee \boldsymbol{m}).$$

where $\boldsymbol{h} \vee \boldsymbol{m}$ denotes the coordinate-wise maximum $(\max\{h_1, m_1\}, \max\{h_2, m_2\})$.

Naturally, superadditive communication always yields a weakly higher success probability than additive communication ($\Phi^S \geq \Phi^A$). The two communication technologies coincide if production is unidimensional or if one agent dominates the other in both dimensions.

Competitive Equilibrium.—Let α_s, α_b, and α_t be the mass of humans hired by single-layer non-automated, bottom-automated, and top-automated firms, respectively. Similarly, denote by μ_s, μ_b, and μ_t, the mass of machines rented by single-layer automated, bottom-automated, and top-automated firms, respectively.

Definition 1 (Competitive Equilibrium). An equilibrium consists of a set of non-negative allocations $(\alpha_s, \alpha_b, \alpha_t, \mu_s, \mu_b, \mu_t)$ and a set of non-negative prices (w, r), such that:

1. Firms optimally choose their structure (while earning zero profits).
2. Single-layer non-automated firms hire a mass α_s of humans.
3. Single-layer automated firms rent a mass μ_s of machines.
4. b-firms hire a mass α_b of humans and rent a mass $\mu_b = \alpha_b n(\boldsymbol{m})$ of machines.
5. t-firms hire a mass α_t of humans and rent a mass $\mu_t = \alpha_t n(\boldsymbol{h})^{-1}$ of machines.
6. Markets clear: $\alpha_s + \alpha_b + \alpha_t = 1$ and $\mu_s + \mu_b + \mu_t = \mu$.

Machines are "Abundant".— Following Ide and Talamàs (2025), we focus on the case where machines are sufficiently abundant so that the binding constraint in human-AI interactions is human time, not machine time. More precisely, we assume that:

$$\mu > \max\left\{ \overbrace{\frac{1}{c(1 - F(1, h_2))}}^{n(\boldsymbol{m})|_{\boldsymbol{m}=(1,h_2)}}, \overbrace{\frac{1}{c(1 - F(h_1, 1))}}^{n(\boldsymbol{m})|_{\boldsymbol{m}=(h_1,1)}} \right\} \tag{1}$$

This condition ensures that, regardless of machine knowledge $\boldsymbol{m} \in [0,1]^2$, there are always more machines available than those demanded in equilibrium by two-layer organizations, i.e., $\mu > \alpha_b^* n(\boldsymbol{m}) + \alpha_t^*/n(\boldsymbol{h})$.[2]

2.2 Discussion of the Model

Our framework makes several simplifying assumptions to illuminate, as transparently as possible, the importance of the nature of communication to determine how AI capabilities shape labor income. Here, we discuss the economic rationale behind these choices and map our stylized setting to real-world applications.

First, our analysis focuses on how AI affects labor outcomes, taking the technology itself as given. In this context, "compute" refers to the resources required during the *inference* phase—the generation of output by an already-trained model. Following Ide and Talamàs (2025), we assume these resources are abundant relative to human time (Equation 1). This assumption is motivated by the exponential growth in computational capacity over the last two centuries Nordhaus (2007), and implies that the binding constraint of human-AI interaction is the scarcity of human time, not the availability of compute.

Second, consistent with the knowledge-hierarchies literature, we assume that determining whether another agent can provide assistance requires communication (Garicano, 2000). This assumption captures a defining feature of non-codifiable knowledge: agents know whether a problem can be solved given their knowledge, but cannot identify the problem's inherent difficulty in each dimension. Because expertise is tacit, agents cannot precisely articulate the boundaries of their competence or assess the complexity of a problem without attempting it. As a result, learning whether someone else can be helpful inevitably requires communicating with them.

Third, we assume that both the human population and the AI technology are homogeneous. Regarding humans, we abstract from heterogeneity to isolate the novel interaction between multidimensional knowledge and communication technology (additive vs. superadditive). Regarding machines, we assume all units of compute are equipped with the same AI knowledge $\boldsymbol{m}$. This reflects the nonrival nature of digital information (Brynjolfsson and McAfee, 2016; Goldfarb and Tucker, 2019) and the tendency of the foundation model market toward commoditization (Navani, 2023; Stebbings, 2024), which drives compute toward the single, highest-performing efficient frontier. We also normalize the price of output to one, allowing us to measure labor income in terms of real production

[2] To see this, note that (i) bottom-automated firms use more machines per human than top-automated ones and (ii) bottom-automated firms cannot emerge when machines outperform humans in both dimensions of knowledge because human solvers would solve fewer problems than machine workers, rendering human assistance useless. Since the number of machines per human in bottom-automated firms is strictly increasing in $\boldsymbol{m}$, and these firms can only arise when either $m_1 \leq h_1$ or $m_2 \leq h_2$, then, if Condition (1) holds, there are always more machines available than those demanded in equilibrium by two-layer organizations.

and abstract from demand-side price effects.

Finally, the organizational forms depicted in Figure 1 correspond to distinct modes of AI deployment. In automated single-layer and bottom-automated firms, machines operate as *AI co-workers* or autonomous AI agents capable of producing independently. This aligns with recent deployments such as Klarna's automation of customer service roles. In contrast, top-automated firms use machines as *AI co-pilots*, where the technology assists humans with complex problem solving without engaging in production—similar to tools like GitHub Copilot.

3 Equilibrium Characterization

In this section, we characterize how the equilibrium labor income depends on machine knowledge, and how this relationship differs under *additive* versus *superadditive* human–AI communication. We index the communication technology by $\tau \in \{S, A\}$, using the success probabilities $\Phi^\tau(\boldsymbol{h}, \boldsymbol{m})$ defined in Section 2.

Given that machines are abundant relative to humans, a positive mass of machines must operate as independent producers in single-layer automated firms. By the zero-profit condition of these firms, the equilibrium rental rate of machines equals their expected output in independent production:

$$r^* = F(\boldsymbol{m}). \tag{2}$$

Notably, r^* does *not* depend on τ: communication affects the value of human–machine *matching*, but not the outside option of a stand-alone machine.

Let $\bar{w}_i^\tau$ denote the highest wage that a firm of type $i \in \{s, b, t\}$ can offer under communication regime τ, taking r^* as given.

Single-layer (non-automated) firms. A single human produces one unit of expected output with probability $F(\boldsymbol{h})$. Zero profits imply

$$\bar{w}_s \equiv F(\boldsymbol{h}),$$

which is independent of τ.

Bottom-automated firms (b). A human solver supervises $n(\boldsymbol{m})$ machines. Each machine–human pair succeeds with probability $\Phi^\tau(\boldsymbol{h}, \boldsymbol{m})$, so expected output is $n(\boldsymbol{m})\Phi^\tau(\boldsymbol{h}, \boldsymbol{m})$. The firm must rent $n(\boldsymbol{m})$ machines at price $r^* = F(\boldsymbol{m})$, and pay wage w to the human. Zero profits therefore imply the maximal feasible wage:

$$\bar{w}_b^\tau \equiv n(\boldsymbol{m})\big(\Phi^\tau(\boldsymbol{h}, \boldsymbol{m}) - F(\boldsymbol{m})\big).$$

This expression makes clear that a bottom-automated firm can pay labor exactly the *incremental value* created by matching each machine with a human (relative to the machine's independent production), aggregated across the $n(\boldsymbol{m})$ machines supervised by one human.

Top-automated firms (t). A machine solver supervises $n(\boldsymbol{h})$ human producers. Equivalently, employing one additional human in a top-automated organization requires renting $1/n(\boldsymbol{h})$ machines in the top layer. A human–machine pair succeeds with probability $\Phi^{\tau}(\boldsymbol{h},\boldsymbol{m})$, so expected output per human is $\Phi^{\tau}(\boldsymbol{h},\boldsymbol{m})$, while the associated machine rental cost per human is $F(\boldsymbol{m})/n(\boldsymbol{h})$. Zero profits give:

$$\bar{w}_t^{\tau} \;\equiv\; \Phi^{\tau}(\boldsymbol{h},\boldsymbol{m}) \;-\; \frac{F(\boldsymbol{m})}{n(\boldsymbol{h})}.$$

Proposition 1 (Equilibrium characterization). *Fix $\tau \in \{S, A\}$. The competitive equilibrium maximizes total output. The equilibrium rental rate is $r^* = F(\boldsymbol{m})$ and the equilibrium wage is*

$$w^{*,\tau} \;=\; \max\{\bar{w}_s, \bar{w}_b^{\tau}, \bar{w}_t^{\tau}\}.$$

Moreover, in equilibrium all humans are allocated to the organizational form with the highest marginal product of labor.

Proof. See Appendix A.1. □

The equilibrium maximizes total output because the First Welfare Theorem applies in this economy (perfect competition, complete information, and only pecuniary externalities). Consequently, each input is paid its marginal product, defined as the increase in total output generated by an additional unit of that input.

Because machines are abundant, an additional machine must be employed as an independent producer. Hence the marginal product of capital (MPK) satisfies

$$\text{MPK} \;=\; F(\boldsymbol{m}) \;=\; r^*.$$

Similarly, introducing an additional human and placing her in the type of firm in which her net contribution is largest increases total output by exactly $\bar{w}_i^{\tau}$, so the marginal product of labor (MPL) satisfies

$$\text{MPL}^{\tau} \;=\; \max\{\bar{w}_s, \bar{w}_b^{\tau}, \bar{w}_t^{\tau}\} \;=\; w^{*,\tau}.$$

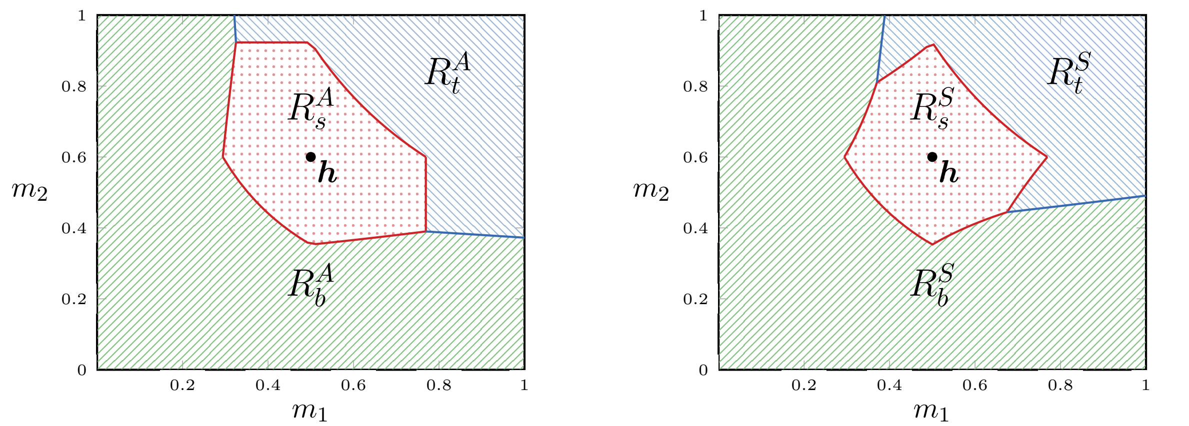
Define the regions in which each organizational form offers the highest wage:

$$R_s^{\tau} \equiv \{\boldsymbol{m} \in [0,1]^2 : \bar{w}_s \geq \max\{\bar{w}_b^{\tau}, \bar{w}_t^{\tau}\}\},$$

$$R_b^{\tau} \equiv \{\boldsymbol{m} \in [0,1]^2 : \bar{w}_b^{\tau} > \max\{\bar{w}_s, \bar{w}_t^{\tau}\}\}, \qquad R_t^{\tau} \equiv [0,1]^2 \setminus (R_s^{\tau} \cup R_b^{\tau}).$$

Figure 2 illustrates these regions in the special case in which problems are uniformly distributed.

The logic determining which two-layer form is used is common across communication regimes τ. Two-layer organizations are valuable when (i) the top layer can leverage its time by handling many exceptions (captured by $c < 1$), and (ii) the top layer is likely to succeed on the pool of exceptions it



(a) Additive Communication ($\tau = A$)

(b) Superadditive Communication ($\tau = S$)

Figure 2: Equilibrium Regions: Additive vs. Superadditive Communication

Notes. Parameter values: $h_1 = 1/2$, $h_2 = 3/5$, and $c = 1/2$, assuming $F(\boldsymbol{x}) = x_1 x_2$.

receives. Whether machines are placed at the bottom or the top layer depends on comparative advantage in solving the problems that remain after the first attempt: when humans dominate machines in both dimensions, machines cannot profitably occupy the solver role; when machines dominate humans in both dimensions, humans cannot profitably occupy the solver role; and when advantages are mixed across dimensions, both placements are in principle viable, with equilibrium selecting the one that yields the highest net surplus once the opportunity cost of machines (embedded in $r^* = F(\boldsymbol{m})$) is taken into account.

Since firms earn zero profits, equilibrium total output can be written as the sum of capital and labor income:

$$Y^{*,\tau} = \mu r^* + w^{*,\tau} = \overbrace{\mu \, \mathrm{MPK}}^{\text{capital income}} + \overbrace{\mathrm{MPL}^{\tau}}^{\text{labor income}} .$$

A more capable AI algorithm—i.e., higher machine knowledge $\boldsymbol{m}$—always increases equilibrium total output and capital income, $\mu r^* = \mu F(\boldsymbol{m})$, under either communication regime. We now turn to showing how equilibrium *labor income* varies with machine knowledge, and how this relationship differs under *superadditive* versus *additive* human–AI communication.

4 The Turing Valley

We define the *Turing Valley* under regime $\tau \in \{S, A\}$ as the equilibrium wage map

$$w^{*,\tau} : [0,1]^2 \to \mathbb{R}_{\geq 0},$$

which assigns to each machine knowledge profile $\boldsymbol{m}$ the marginal product of labor (or, equivalently, labor income) under that regime (Proposition 1).

4.1 Jagged AI

A defining feature of contemporary AI is the unevenness of its capabilities across dimensions of knowledge. This "jagged" profile—characterized by systems that are "really good" at some tasks yet "really flawed" at others—stands in sharp contrast to the ambition of Artificial General Intelligence (AGI). Motivated by this "jaggedness," we focus on the case in which machines outperform humans in one dimension but underperform in the other:

$$m_1 > h_1 \qquad \text{and} \qquad m_2 < h_2. \tag{3}$$

In the terminology suggested by the Turing's (1950) test of machine intelligence, machines "pass" in dimension 1 but "fail" in dimension 2.[3]

Our comparative statics results below hold globally (for all $\boldsymbol{m}$), but the jagged region (3) is where the contrast between superadditive and additive communication is important, and it is precisely where "de-jagging" AI (improving it in its weak dimension) corresponds to moving toward AGI.

4.2 How AI capabilities shape labor income

We now characterize how equilibrium labor income responds to marginal improvements in machine knowledge. We focus specifically on the consequences of "de-jagging" AI—improving machine capabilities in dimensions where they currently lag behind humans (as defined in (3)). The following result establishes that the labor-market impact of such progress depends fundamentally on the communication regime.

Proposition 2 (Local comparative statics in the Turing Valley)**.** *Fix $\tau \in \{S, A\}$. The equilibrium wage $w^{*,\tau}$ is continuous and almost everywhere differentiable in $\boldsymbol{m}$. Moreover, letting $\mathrm{int}A$ denote the interior of A:*

1. ***No matching region.** If $\boldsymbol{m} \in \mathrm{int}R_s^\tau$, then $\partial w^{*,\tau}/\partial m_1 = \partial w^{*,\tau}/\partial m_2 = 0$.*
2. ***Strong AI dimension (m_1).** If $\boldsymbol{m} \in \mathrm{int}\,(R_b^\tau \cup R_t^\tau)$, then labor income rises with AI improvements:*

$$\frac{\partial w^{*,\tau}}{\partial m_1} \geq 0 \quad \textit{(with strict inequality if } m_2 > 0\textit{)}.$$

3. ***Weak AI dimension (m_2).** If $\boldsymbol{m} \in \mathrm{int}\,(R_b^\tau \cup R_t^\tau)$, the effect of AI improvements depends on the nature of communication:*
   - *Under **additive** communication ($\tau = A$), the sign of $\partial w^{*,A}/\partial m_2$ is ambiguous. However, there exist $U \in \mathbb{R}$ such that, if $m_1/h_1 > U$, then $\partial w^{*,A}/\partial m_2 > 0$.*
   - *Under **superadditive** communication ($\tau = S$), labor income strictly falls: $\partial w^{*,S}/\partial m_2 < 0$.*

[3] Turing (1950) proposed testing machine intelligence via an imitation game: can machines imitate humans so well that their responses are indistinguishable from those of a human? In our context, we apply a similar criterion separately to each dimension of knowledge. Formally, we say that machines pass the (dimension-specific) Turing Test in dimension i if $m_i \geq h_i$.

Proof. See Appendix B.1. □

The contrast between the two communication regimes is most transparent in bottom-automated organizations. When $\boldsymbol{m} \in R_b^\tau$, equilibrium labor income is

$$w^{*,\tau} = \bar{w}_b^\tau = \frac{\Phi^\tau(\boldsymbol{h},\boldsymbol{m}) - F(\boldsymbol{m})}{c\big(1 - F(\boldsymbol{m})\big)} = \frac{1}{c}\,\mathbb{P}(\text{assistance is useful} \mid \text{assistance is requested}), \tag{4}$$

Equation (4) mirrors the standard "management-by-exception" logic (Garicano, 2000; Garicano and Rossi-Hansberg, 2015): each human can provide assistance $1/c$ times, and the value of that assistance is the probability that humans can help resolve a problem that machines cannot solve.

First, consider an increase in m_1—an improvement in the dimension where machines are initially strong. Under additive communication, raising m_1 expands the set of problems that mac¡hines can solve autonomously (area α in Figure 3(a)). These are problems for which human assistance was not useful to begin with. As a result, the set of problems reaching humans becomes more favorable, increasing the probability that human assistance is useful.

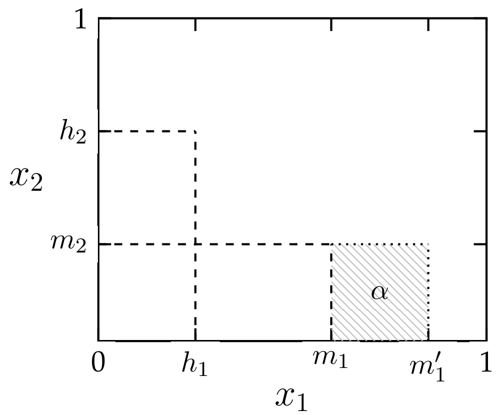
Under superadditive communication, raising m_1 again expands the set of problems solved autonomously by machines (area α in Figure 3(b)), but it also enables human–AI communication to solve new problems by combining machine knowledge in dimension 1 with human knowledge in dimension 2 (area β in Figure 3(b)). This second effect strictly increases the set of problems for which human assistance is valuable. Both effects therefore raise the usefulness of human assistance.

Second, consider second an increase in m_2—an improvement in the dimension where machines are initially weak. Under additive communication, raising m_2 creates two opposing forces visible in Figure 3(c). On one hand, machines are now able to solve problems that humans previously helped with (those whose difficulty lies in area γ), and this reduces the probability that humans' assistance is useful. On the other hand, machines can also now solve problems that humans could not help with (those whose difficulty lies in area η), and this increases the probability that humans' assistance is useful. The net effect on the usefulness of human assistance therefore depends on the size of γ relative to η. However, when AI is sufficiently stronger than humans in its strong dimension (i.e., m_1/h_1 is sufficiently high), the second effect dominates, so such AI improvement increases labor income.

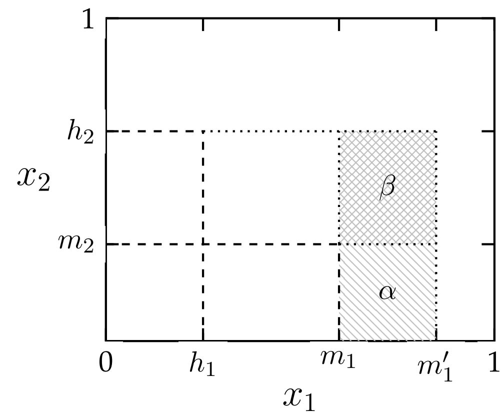
Under superadditive communication, raising m_2 creates two forces visible in Figure 3(d), but in this case both forces go in the same direction. Machines are now able to solve problems that humans could previously solve (those whose difficulty lies in area γ) or help solve (those whose difficulty lies in area η). Both effects reduce the probability that the humans' assistance is useful, thus leading to a reduction in labor income.

A parallel intuition applies in top-automated organizations. When $\boldsymbol{m} \in R_t^\tau$, equilibrium labor income is

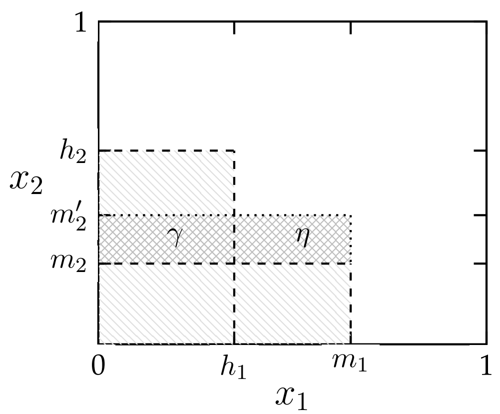
$$w^{*,\tau} = \bar{w}_t^\tau = \Phi^\tau(\boldsymbol{h},\boldsymbol{m}) - \frac{F(\boldsymbol{m})}{n(\boldsymbol{h})}. \tag{5}$$



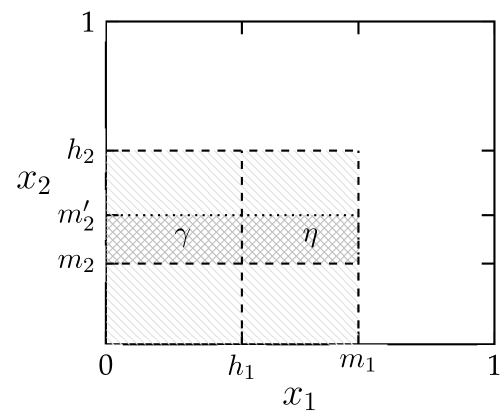
(a) Additive: Improving Strength (m_1)



(b) Superadditive: Improving Strength (m_1)



(c) Additive: Improving Weakness (m_2)



(d) Superadditive: Improving Weakness (m_2)

Figure 3: Comparative Effects of Improving AI Capabilities

Notes. The figure illustrates the effects of improving AI capabilities. Panels (a) and (b) show the effects of improving AI in its strong dimension m_1. Under additive communication (a), the additional problems that machines now solve (area α) are precisely problems for which human assistance was not useful. Under superadditive communication (b), there is an additional effect: human assistance becomes more useful as the organization solves new problems (area β). Panels (c) and (d) show the effects of improving AI in its weak dimension m_2. In panel (c) (Additive), the effect is ambiguous: the improvement allows machines to solve problems humans could previously solve (area γ) but also enables machines to solve problems neither humans nor human-AI communication could solve before (area η). In panel (d) (Superadditive), the increase unambiguously reduces the value of human assistance, because the organization could already solve the new problems machines can solve.

The first term is labor productivity (humans' expected output), while the second term is the machine cost per human.

Consider first an increase in m_1. Under additive communication, the improvement enables the machine to solve problems that previously went unsolved (area α in Figure 3(a)). In this case, Φ^A increases by the same amount as machine capabilities $F(\boldsymbol{m})$. As evidenced by (5), the productivity gain Φ^A accrues fully to each human while the machine cost is shared across $n(\boldsymbol{h}) > 1$ humans, so labor income rises. Under superadditive communication, the increase in Φ^S is larger than the increase in $F(\boldsymbol{m})$, so the effect is stronger than in the additive communication case. This is because the improvement not only allows machines to solve problems whose difficulty lie in area α in Figure 3(b), but also allows the human–AI team to solve new complementary problems (area β) that require human-AI communication.

Consider second an increase in m_2. Under additive communication, the effect is ambiguous because total output Φ^A increases due to the expansion into problems that humans could not help with (area η in Figure 3(c)), but by a smaller amount than machine's output in autarky $F(\boldsymbol{m})$, which also includes problems that human-AI communication was already solving (area γ). However, when AI is sufficiently stronger than humans in its strong dimension (i.e., m_1/h_1 is sufficiently high), the increase in Φ^A is sufficiently large relative to the increase in $F(\boldsymbol{m})$, so such improvement increases labor income. Under superadditive communication, Φ^S does not increase because the problems in area η and γ (in Figure 3(d)) were already solved by human-AI communication. Consequently, the only change in (5) is a higher machine cost $F(\boldsymbol{m})$, which unambiguously reduces labor income.

4.3 The Labor-Income Maximizing AI

We now characterize the machine knowledge $\boldsymbol{m}$ that maximizes labor income. The following result establishes that when human knowledge is itself jagged—characterized by high proficiency in one dimension and significant limitations in the other—the optimal AI configuration depends fundamentally on the communication regime.

Proposition 3 (Global Optimality: Jagged AI vs. AGI)**.** *The machine knowledge profile $\boldsymbol{m}^* \in [0,1]^2$ that maximizes equilibrium labor income satisfies:*

*(i) Under **additive communication**, there exists $\underline{h} > 0$ such that, if $h_1 < \underline{h}$, labor income is maximized by **AGI** that is as good as possible in every dimension ($\boldsymbol{m}^* = (1,1)$).*

*(ii) Under **superadditive communication**, there exists $\overline{h} < 1$ such that, if $h_2 > \overline{h}$, labor income is maximized by a **Jagged AI** that is perfect in one dimension but minimal in the other ($\boldsymbol{m}^* = (1,0)$).*

Proof. See Appendix B.2. □

The intuition for Proposition 3 is parallel to that of the local comparative statics in Proposition 2. When humans knowledge in dimension 1 is weak and communication is additive, the probability that human assistance is useful is small. As a result, humans are better off using machines as solvers. In that case, all AI improvements increase the price $F(\boldsymbol{m})$ of machines by roughly the same amount as labor productivity Φ^A. But, as evidenced by (5), the higher price $F(\boldsymbol{m})$ of machines is shared across $n(\boldsymbol{h}) > 1/c$ humans. Hence, all AI improvements increase labor income, and the AI configuration that maximizes labor income is $\boldsymbol{m}^* = (1, 1)$.

When humans knowledge in dimension 2 is strong and communication is superadditive, the probability that human assistance is useful when $\boldsymbol{m}^* = (1, 0)$ is almost one. As a result, the marginal product of labor in bottom-automated firms is close to the number of problems solvers can tackle ($1/c > 1$). In contrast, their marginal product in top-automated firms is naturally bounded by one. Hence, their marginal product is determined by the probability that human assistance is useful, and this is achieved when machines know as much as possible in the dimension in which humans are weak, but as least as possible in the dimension in which humans are strong.

5 Conclusion

AI's rapid progress is intensifying the debate over the economic consequences of achieving Artificial General Intelligence (AGI)—a form of AI that eliminates the "jagged" nature of current systems by attaining high performance across all domains. Motivated by this debate, we develop a model of the knowledge economy where production requires solving multidimensional problems, and where the impact of AI depends not just on its capabilities, but on the nature of human–AI communication.

Our analysis reveals that the economic effects of smoothing AI's jaggedness are fundamentally determined by the nature of human-AI communication. When communication is *additive* and AI is sufficiently stronger than humans in one dimension, all AI improvements increase labor income. As a result, when human expertise is narrow (sufficiently weak in one dimension), labor income is maximized when AI achieves uniformly high performance across all dimensions, corresponding to AGI.

The picture changes qualitatively under *superadditive* communication. In this case, improvements in dimensions where machines already outperform humans raise labor income, but improvements in dimensions where machines fail the Turing Test reduce it. As a result, when human expertise is deep (sufficiently good in one dimension) labor income is maximized by a deliberately jagged form of AI that complements human strengths rather than by AGI.

This paper opens several avenues for future research. Empirically, distinguishing domains with superadditive versus additive communication is critical for predicting where AGI will enhance rather than erode labor income. Theoretically, extending our model to include heterogeneous human capital could shed light on how the transition from jagged AI to AGI will reshape labor income inequality.

APPENDIX

A Proofs Omitted From Section 3

A.1 Proof of Proposition 1

First, the abundance of machines (Equation 1) ensures that there is always a surplus of machines employed as independent producers. The zero-profit condition for these single-layer automated firms pins down the equilibrium rental rate at their marginal product:

$$r^* = F(\boldsymbol{m}).$$

Second, perfect competition for labor ensures that humans are hired by the organizational form that can pay the highest wage. Since firms earn zero profits, the maximum wage a firm of type $i \in \{s, b, t\}$ can pay is exactly the marginal product of labor in that form, denoted $\bar{w}_i^\tau$ (as derived in the text). Consequently, the equilibrium wage is:

$$w^{*,\tau} = \max\{\bar{w}_s, \bar{w}_b^\tau, \bar{w}_t^\tau\}.$$

In equilibrium, labor is allocated to the organizational form(s) that achieve this maximum.

Finally, to see that this equilibrium maximizes total output, note that aggregate profits are zero. Thus, total output Y equals total factor income:

$$Y = \text{Capital Income} + \text{Labor Income} = \mu r^* + 1 \cdot w^{*,\tau}.$$

Substituting $r^* = F(\boldsymbol{m})$, total output is $Y = \mu F(\boldsymbol{m}) + w^{*,\tau}$. Since the capital income term $\mu F(\boldsymbol{m})$ is constant regardless of how labor is organized, maximizing total output is equivalent to maximizing the equilibrium wage $w^{*,\tau}$. Since the competitive equilibrium allocates labor to the use with the highest return (maximizing $w^{*,\tau}$), it necessarily maximizes total output. $\square$

B Proofs Omitted From Section 4

B.1 Proof of Proposition 2

The continuity and almost-everywhere differentiability of $w^{*,\tau}$ follow because $w^{*,\tau}$ is the upper envelope of continuous functions ($\bar{w}_s, \bar{w}_b^\tau, \bar{w}_t^\tau$). We characterize the derivatives in each region assuming the jagged configuration ($m_1 > h_1$ and $m_2 < h_2$).

1. No matching region ($\boldsymbol{m} \in \text{int}R_s^\tau$). $w^{*,\tau} = F(\boldsymbol{h})$, which is independent of $\boldsymbol{m}$. Thus $\partial w^{*,\tau}/\partial m_i = 0$.

2. Strong AI dimension (m_1). Consider $\boldsymbol{m} \in \text{int}(R_b^\tau \cup R_t^\tau)$. The marginal increase in machine capabilities is given by $\frac{\partial F(\boldsymbol{m})}{\partial m_1} = \int_0^{m_2} f(m_1, x_2)\, dx_2$. Because f has full support, this term is strictly positive if and only if $m_2 > 0$.

- **Top-automated firms (t).** The wage is $w^* = \Phi^\tau - c(1-F(\boldsymbol{h}))F(\boldsymbol{m})$. Differentiating with respect to m_1:

$$\frac{\partial w^*}{\partial m_1} = \frac{\partial \Phi^\tau}{\partial m_1} - c(1-F(\boldsymbol{h}))\frac{\partial F(\boldsymbol{m})}{\partial m_1}.$$

  - Under **Additive** communication, $\frac{\partial \Phi^A}{\partial m_1} = \frac{\partial F(\boldsymbol{m})}{\partial m_1}$. The derivative becomes $[1-c(1-F(\boldsymbol{h}))]\,\frac{\partial F(\boldsymbol{m})}{\partial m_1}$. Since the bracket is positive, the derivative is strictly positive if and only if $m_2 > 0$.
  - Under **Superadditive** communication, $\Phi^S = F(m_1, h_2)$. Since $h_2 > m_2$, the marginal gain $\frac{\partial \Phi^S}{\partial m_1}$ strictly exceeds $\frac{\partial F(\boldsymbol{m})}{\partial m_1}$. Thus, the derivative is strictly positive even if $m_2 = 0$.

  Combining cases, $\frac{\partial w^*}{\partial m_1} \geq 0$, with strict inequality guaranteed if $m_2 > 0$.

- **Bottom-automated firms (b).** The wage is $w^* = \frac{\Phi^\tau - F(\boldsymbol{m})}{c(1-F(\boldsymbol{m}))}$.

  - Under **Additive** communication, the numerator $\Phi^A - F(\boldsymbol{m})$ is constant in m_1. The wage rises solely because the denominator shrinks. This increase is strict if and only if $\frac{\partial F(\boldsymbol{m})}{\partial m_1} > 0$, i.e., $m_2 > 0$.
  - Under **Superadditive** communication, the numerator strictly increases (since $\frac{\partial \Phi^S}{\partial m_1} > \frac{\partial F(\boldsymbol{m})}{\partial m_1}$) while the denominator decreases. Both forces strictly raise wages.

Thus, in all cases, $\partial w^{*,\tau}/\partial m_1 \geq 0$, and $m_2 > 0$ is a sufficient condition for strict inequality.

3. Weak AI dimension (m_2). Consider $\boldsymbol{m} \in \mathrm{int}(R_b^\tau \cup R_t^\tau)$. The effect of increasing m_2 depends on how Φ^τ responds relative to the rising machine cost $F(\boldsymbol{m})$.

- **Top-automated firms (t).** The wage is $w^* = \Phi^\tau - c(1-F(\boldsymbol{h}))F(\boldsymbol{m})$. Differentiating with respect to m_2:

$$\frac{\partial w^*}{\partial m_2} = \frac{\partial \Phi^\tau}{\partial m_2} - \frac{F(\boldsymbol{m})}{n(\boldsymbol{h})}\frac{\partial F(\boldsymbol{m})}{\partial m_2}.$$

  - Under **Additive** communication, $\frac{\partial \Phi^A}{\partial m_2} = \frac{\partial F(\boldsymbol{m})}{\partial m_2} - \frac{\partial F(h_1, m_2)}{\partial m_2}$. The wage derivative contains a positive term (from machine capability growth) minus negative terms (overlap and cost). The sign is ambiguous.
  - Under **Superadditive** communication, $\Phi^S = F(m_1, h_2)$. Since $m_2 < h_2$, the joint success probability is locally independent of m_2 ($\frac{\partial \Phi^S}{\partial m_2} = 0$). The wage derivative is strictly negative due to the rising cost term.

- **Bottom-automated firms (b).** When $\boldsymbol{m} \in R_b^\tau$, equilibrium labor income is determined by the value of human assistance on the pool of problems the machine cannot solve. Using the definition of Φ^τ, we can rewrite the wage as:

$$w^{*,\tau} = \frac{\Phi^\tau(\boldsymbol{h}, \boldsymbol{m}) - F(\boldsymbol{m})}{c(1-F(\boldsymbol{m}))} = \frac{1}{c}\cdot \mathbb{P}(\text{assistance is useful} \mid \text{assistance is requested}). \tag{6}$$

The term $\Phi^\tau - F(\boldsymbol{m})$ represents the probability that the human-AI team succeeds on a problem that the machine alone would fail. An improvement in the weak AI dimension ($m_2 < h_2$) increases the machine's standalone success probability $F(\boldsymbol{m})$, shrinking the pool of escalations (the denominator). The effect on the wage depends on how the set of *useful* escalations (the numerator) responds relative to the total pool:

– **Additive communication ($\tau = A$).** Increasing m_2 generates two opposing forces. First, the machine automates some problems that the human could previously solve alone (Area γ in Figure 4). This effect reduces the numerator, lowering the wage. Second, the machine successfully solves some problems that the human alone would have failed (Area η). By filtering out these problems from the escalation pool, the machine increases the conditional probability that the remaining escalations are within the human's competence. The net effect on labor income is therefore ambiguous.

– **Superadditive communication ($\tau = S$).** Increasing m_2 unambiguously reduces labor income. Because the machine is weak in this dimension ($m_2 < h_2$), the team's joint capability is constrained by the human's strength, so $\Phi^S = F(h \vee m)$ remains locally constant. Consequently, the machine does not filter out any problems the team could not already solve. Instead, the machine merely automates problems that were previously solved by the human (Area γ) or by human-AI collaboration (Area η). Both effects reduce the numerator ("useful assistance") without a compensating improvement in the "quality" of the escalation pool.

$\square$

B.2 Proof of Proposition 3

We analyze the labor-income maximizing machine knowledge $\boldsymbol{m}^*$ when human expertise is itself jagged (specifically, for h_2 sufficiently close to 1 and h_1 sufficiently close to 0).

Part (i): Additive Communication ($\tau = A$). We show that if human expertise is sufficiently narrow in dimension 1 (i.e., for sufficiently low h_1), labor income is strictly maximized at the AGI corner $\boldsymbol{m} = (1, 1)$.

Consider the limit $h_1 \to 0$. In this limit, $F(\boldsymbol{h}) \to 0$ and the overlap term $F(\boldsymbol{h} \wedge \boldsymbol{m}) \to 0$. Consequently, the additive success probability converges to the machine's standalone probability: $\Phi^A(\boldsymbol{h}, \boldsymbol{m}) \to F(\boldsymbol{m})$.

- **Bottom-automated firms.** The wage converges to zero:

$$\bar{w}_b^A \to \frac{F(\boldsymbol{m}) - F(\boldsymbol{m})}{c(1 - F(\boldsymbol{m}))} = 0.$$

  Intuitively, without the ability to recombine partial solutions, a human with negligible knowledge in dimension 1 cannot effectively assist a machine that fails.

- **Top-automated firms.** The wage converges to:

$$\bar{w}_t^A \to F(\boldsymbol{m}) - cF(\boldsymbol{m}) = (1-c)F(\boldsymbol{m}).$$

  Since $c < 1$, this asymptotic wage is strictly increasing in machine output $F(\boldsymbol{m})$. Because f has full support, $F(\boldsymbol{m})$ is strictly increasing in both m_1 and m_2. Therefore, labor income is uniquely maximized at the upper bound of the machine knowledge space, $\boldsymbol{m}^* = (1, 1)$.

Part (ii): Superadditive Communication ($\tau = S$). Under superadditivity, $\Phi^S(\boldsymbol{h}, \boldsymbol{m}) = F(\boldsymbol{h} \vee \boldsymbol{m})$. Consider the limit where humans are sufficiently deep in dimension 2 (i.e., $h_2 \to 1$). As argued in the text, in this case the marginal product of labor is determined in bottom-automated firms. The bottom-automated wage is $w^* = \frac{K(m_1)-F(\boldsymbol{m})}{c(1-F(\boldsymbol{m}))}$, where the joint success probability converges to $K(m_1) \equiv F(\max\{h_1, m_1\}, 1)$ is independent of m_2. Differentiating with respect to m_2:

$$\frac{\partial w^*}{\partial m_2} = \frac{\partial w^*}{\partial F(\boldsymbol{m})} \frac{\partial F(\boldsymbol{m})}{\partial m_2}.$$

The term $\frac{\partial w^*}{\partial F(\boldsymbol{m})} = \frac{-(1-F)-(K-F)(-1)}{c(1-F)^2} = \frac{-(1-K)}{c(1-F)^2}$. Since $K < 1$, this is strictly negative. Since $\frac{\partial F}{\partial m_2} > 0$, we have $\frac{\partial w^*}{\partial m_2} < 0$. Thus, labor income is maximized by setting $m_2 = 0$ and $m_1 = 1$.

$\square$

References

Acemoglu, D. and S. Johnson (2024): "Learning From Ricardo and Thompson: Machinery and Labor in the Early Industrial Revolution and in the Age of Artificial Intelligence," *Annual Review of Economics*, 16, 597–621.

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.

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

——— (2022): "Tasks, automation, and the rise in US wage inequality," *Econometrica*, 90, 1973–2016.

Althoff, L. and H. Reichardt (2026): "Task-Specific Technical Change and Comparative Advantage," Working paper.

Antràs, P., L. Garicano, and E. Rossi-Hansberg (2006): "Offshoring in a Knowledge Economy," *The Quarterly Journal of Economics*, 121, 31–77.

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.

BRYNJOLFSSON, E. (2022): "The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence," *Dædalus*, 151, 272–282.

BRYNJOLFSSON, E., D. LI, AND L. RAYMOND (2025): "Generative AI at work," *The Quarterly Journal of Economics*, 140, 889–942.

BRYNJOLFSSON, E. AND A. MCAFEE (2016): *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W. W. Norton & Company.

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 Organization 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.

DELL'ACQUA, F., C. AYOUBI, H. LIFSHITZ, R. SADUN, E. MOLLICK, L. MOLLICK, Y. HAN, J. GOLDMAN, H. NAIR, S. TAUB, ET AL. (2025): "The cybernetic teammate: A field experiment on generative AI reshaping teamwork and expertise," Tech. rep., National Bureau of Economic Research.

FRIDMAN, L. AND D. HASSABIS (2025): "Demis Hassabis: Future of AI, Simulating Reality, Physics and Video Games — Lex Fridman Podcast #475," Lex Fridman Podcast (video interview), episode webpage: https://lexfridman.com/demis-hassabis-2/; published July 23, 2025.

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 (2012): "Learning About the Nature of Production from Equilibrium Assignment Patterns," *Journal of Economic Behavior & Organization*, 84, 136–153.

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.

GOLDFARB, A. AND C. TUCKER (2019): "Digital Economics," *Journal of Economic Literature*, 57, 3–43.

GRAETZ, G. AND G. MICHAELS (2018): "Robots at work," *Review of Economics and Statistics*, 100, 753–768.

GUMPERT, A., H. STEIMER, AND M. ANTONI (2022): "Firm Organization with Multiple Establishments," *The Quarterly Journal of Economics*, 137, 1091–1138.

IDE, E. (2025): "Automation, AI, and the Intergenerational Transmission of Knowledge," *arXiv preprint arXiv:2507.16078*.

IDE, E. AND E. TALAMÀS (2025): "Artificial Intelligence in the Knowledge Economy," *Journal of Political Economy*, 133.

KRAKOWSKI, I., J. KIM, Z. R. CAI, R. DANESHJOU, J. LAPINS, H. ERIKSSON, A. LYKOU, AND E. LINOS (2024): "Human-AI interaction in skin cancer diagnosis: a systematic review and meta-analysis," *NPJ digital medicine*, 7, 78.

MOLL, B., L. RACHEL, AND P. RESTREPO (2022): "Uneven Growth: Automation's Impact on Income and Wealth Inequality," *Econometrica*, 90, 2645–2683.

NAVANI, D. A. (2023): "The Commoditization of LLMs," *Communications of the ACM*, available at https://cacm.acm.org/blogcacm/the-commoditization-of-llms/.

NORDHAUS, W. D. (2007): "Two Centuries of Productivity Growth in Computing," *The Journal of Economic History*, 67, 128–159.

RESTREPO, P. (2025): "We Won't Be Missed: Work and Growth in the AGI World," in *The Economics of Transformative AI*, ed. by A. Agrawal, A. Korinek, and E. Brynjolfsson, forthcoming.

STEBBINGS, H. (2024): "Aravind Srinivas: Will Foundation Models Commoditise?" 20VC with Harry Stebbings, Episode #1161, available at https://www.thetwentyminutevc.com/aravind-srinivas.

TURING, A. M. (1950): "Computing Machinery and Intelligence," *Mind*, LIX, 433–460.

ZEIRA, J. (1998): "Workers, Machines, and Economic Growth," *The Quarterly Journal of Economics*, 113, 1091–1117.