NBER WORKING PAPER SERIES

THE ECONOMICS OF BICYCLES FOR THE MIND

Ajay K. Agrawal
Joshua S. Gans
Avi Goldfarb

Working Paper 34034
http://www.nber.org/papers/w34034

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2025

Funding provided by SSHRC and the Acceleration Consortium. We used GPT-o3 and Claude 4 Opus for research support. Thanks to Rene Thiang for research assistance and seminar participants at the University of Chicago and Amazon for useful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w34034

The Economics of Bicycles for the Mind
Ajay K. Agrawal, Joshua S. Gans, and Avi Goldfarb
NBER Working Paper No. 34034
July 2025
JEL No. D83, J24, L23, O33

ABSTRACT

Steve Jobs described computers as "bicycles for the mind," a tool that allowed people to dramatically leverage their capabilities. This paper presents a formal model of cognitive tools and technologies that enhance mental capabilities. We consider agents engaged in iterative task improvement, where cognitive tools are assumed to be substitutes for implementation skills and may or may not be complements to judgment, depending on their type. The ability to recognise opportunities to start or improve a process, which we term opportunity judgment, is shown to always complement cognitive tools. The ability to know which action to take in a given state, which we term payoff judgment, is not necessarily a complement to cognitive tools. Using these concepts, we can synthesise the empirical literature on the impact of computers and artificial intelligence (AI) on productivity and inequality. Specifically, while both computers and AI appear to increase productivity, computers have also contributed to increased inequality. Empirical work on the impact of AI on inequality has shown both increases and decreases, depending on the context. We also apply the model to understand how cognitive tools might influence incentives to automate processes and allocate decision-making authority within teams.

Ajay K. Agrawal
University of Toronto
Rotman School of Management
and NBER
ajay.agrawal@rotman.utoronto.ca

Avi Goldfarb
University of Toronto
Rotman School of Management
and NBER
agoldfarb@rotman.utoronto.ca

Joshua S. Gans
University of Toronto
Rotman School of Management
and NBER
joshua.gans@rotman.utoronto.ca

What a computer is to me is it's the most remarkable tool that we've ever come up with, and it's the equivalent of a bicycle for our minds. Steve Jobs, Library of Congress, 1990.

1 Introduction

A bicycle is a tool used by people to travel more efficiently. Indeed, a 1973 article in *Scientific American* calculated that it might be the lowest cost per kilogram of any means of transportation, both natural and artificial. Steve Jobs was influenced by this notion when he characterised computers as "bicycles for the mind."[1] The idea is that computers are tools that make minds more efficient in much the same way as a bicycle makes legs more efficient.

This paper is about computers and artificial intelligence (AI) as cognitive tools. A growing empirical literature explores how computers and AI affect work. We build a model that enables interpretation of many of the results in this literature, both the historical evidence on computers and the emerging work on AI. It provides a unifying framework for interpreting the following empirical results:[2]

- Computers increased productivity of adopting workers and firms, and at the macroeconomic level (e.g. Bresnahan et al. (2002); Jorgenson and Stiroh (2000)).
- AI adoption increases productivity of adopting workers and firms (e.g. McElheran et al. (2025); Cui et al. (2025)).
- Computers increased inequality, within adopting industries and economy-wide (e.g. Dranove et al. (2014); Autor et al. (2006)).
- AI often reduces inequality within adopting firms (e.g. Brynjolfsson et al. (2025); Kanazawa et al. (2022)).
- The task-based approach to anticipating AI's impact on the economy suggests high-income occupations will be most impacted (e.g. Brynjolfsson and Mitchell (2017); Felten et al. (2021); Eloundou et al. (2024)).
- Computer and AI adoption may not lead to full automation (e.g. Agrawal et al. (2018b); Scharre (2016)).
- For both computers and AI, team composition changes (e.g. Teodoridis (2018); Law and Shen (2025))

[1] This concept was articulated by Jobs as early as 1980.

[2] We provide a rich discussion of the evidence, with relevant citations, throughout the paper.

This framework examines how agents engage in iterative task improvement, where in each period, agents have the opportunity to enhance the quality of their work. The model has three key components: the agent's level of effort in implementing improvements (which affects both costs and probability of success), their ability to know how those improvements translate into action, and their ability to locate new improvement opportunities. Each successful improvement increases quality by a fixed amount, but the process eventually ends when the agent fails to identify a new opportunity for enhancement.

The model then examines how the introduction of a tool, defined as something that improves the ratio of success probability to cost for any given level of effort, affects this process. When a cognitive tool is adopted, agents reduce their level of direct effort since the tool makes it easier to achieve good outcomes. However, because the tool increases the probability of success relative to costs, the net benefit from improvements rises.

Our first motivating example is the computer. A knowledge worker is tasked with designing a product. Absent a computer, the worker performs calculations and creates visualisations by hand. With a computer (say, used for Computer-Aided Design or CAD), the worker can now iterate more quickly. In addition to the implementation skill of performing calculations and creating visualisations, the knowledge worker also identifies opportunities to improve the outcome, which we label 'opportunity judgment'. In this case, the introduction of the cognitive tool amplifies the value of opportunity judgment by increasing the net benefit of acting on opportunities.

Our second motivating example is AI prediction in the service of decision making (Agrawal et al., 2019), for example, as in a physician providing a diagnosis. Even when a diagnosis is made, there is uncertainty about whether a treatment is available to address that diagnosis. This uncertainty creates a need for a different kind of skill, which we label 'payoff judgment'. This refers to the decision maker's ability to translate the predicted knowledge of the state into an optimal decision. Unlike opportunity judgment, payoff judgment increases accuracy per unit of effort but decreases effort. Thus, payoff judgment is only a complement to a cognitive tool when the tool does not decrease effort too much.

We then build a general model that combines the insights from these motivating examples. The general model nests both computers and AI predictions, as well as generative AI, which provides aspects of both types of judgment. This general model yields several insights. First, cognitive tools lead to more output with less effort. They are productivity-enhancing.

Second, as with the motivating examples, cognitive tools are substitutes for implementation skill, are complements to opportunity judgment, and are generally complements to payoff judgment (as long as the reduction in effort is not so large as to reduce output). This framework builds on ideas in Autor and Thompson (2025), who also emphasise that the im-

pact of technological change on labour outcomes depends on what aspects of an occupation get automated.

Third, the impact of a cognitive tool on inequality depends on the distribution of human skills in relation to implementation and the various types of judgment. As a cognitive tool improves, it may exhibit an inverse skill bias: inequality may decline as the tool replaces the value created by highly skilled humans during implementation. Eventually, as the tool improves, inequality may begin to increase again through the amplification of differences in judgment, as the humans with high judgment benefit from the complementarity between judgment and the tool.

This third result helps interpret two distinct literatures on technology and inequality. For the empirical literature on whether computers and AI have increased or decreased labour market outcomes (Dranove et al., 2014; Autor et al., 2006; Brynjolfsson et al., 2025; Kanazawa et al., 2022; Humlum and Vestergaard, 2025a), the result suggests that the evidence that AI has decreased inequality may reflect the initial stage of inverse skill bias. For computers, the evidence may already reflect the longer run amplification of differences in judgment. A separate literature uses the tasks involved in current jobs to identify which workers are most likely to have their jobs affected by AI (Brynjolfsson and Mitchell, 2017; Felten et al., 2021; Eloundou et al., 2024). These papers suggest that high-wage jobs are likely to be most affected by AI. Because these papers reflect current workflows, they may also reflect the initial stage of inverse skill bias. As workflows change and new opportunities are identified, differences in judgment that do not reflect current workflows may become more important. Put differently, the model suggests that the early results on AI's likely impact on wage inequality may reflect an early stage where AI makes implementation skill less valuable rather than a longer run in which improvements in AI tools make judgment more valuable.

Fourth, because judgment is a complement to the tool, automation (which requires prespecifying judgment) may be less likely as the cognitive tool gets better. This highlights the brittleness of fully automated systems when the stakes are high (Scharre, 2016; Agrawal et al., 2018b).

Fifth, and finally, cognitive tools can change which teams are most productive (Teodoridis, 2018; Law and Shen, 2025). Implementation skills become less important than judgment skills. Depending on the ease of communication, this can change who determines what to do.

Thus, the model provides a structure for the existing empirical work on cognitive tools, whether computers or AI. We next turn to our motivating examples before describing the full model.

2 Motivating Examples

Before providing a general model of cognitive tools, it is instructive to build up from some motivating contexts that capture the two kinds of judgment separately. The first examines computers, which are cognitive tools that follow instructions directly. The second examines the role of AI prediction in a decision-making context.

2.1 Computers as Bicycles for the Mind

Before providing a more general model of cognitive tools, we begin with a simple motivating example that captures the essence of Steve Jobs' "bicycle for the mind" metaphor. This example illustrates how cognitive tools can enhance a fundamental aspect of human cognition: the ability to identify and implement improvements to a task over time.

The ability to identify opportunities is well understood as central to problem solving and creative work. Reiter-Palmon and Robinson (2009) and Mumford et al. (1994) emphasize that problem identification is a meaningful skill. Csikszentmihalyi and Getzels (1971) describe the importance of problem finding, quoting Einstein and Infeld (1938):

> The formation of a problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science

It is this skill to find problems and opportunities for improvement that we label 'opportunity judgment.'

Consider a knowledge worker tasked with designing a product. Without computational tools, the worker manually performs calculations and creates visualisations by hand. With computer-aided design, these operations become more efficient, allowing the worker to iterate more quickly and potentially achieve better results.

We model this scenario as an iterative improvement process where, in each period, the worker identifies an opportunity to enhance their analysis and implements it. The key components of this basic model are:

- **Opportunity Judgment (γ):** In each period t, the worker identifies an improvement opportunity with probability $\gamma \in [0, 1]$. This parameter represents what we call *opportunity judgment*—the cognitive ability to recognise potential enhancements to the task.

- **Implementation Effort (e_t):** When an opportunity is identified, the worker exerts effort $e_t \geq 0$ at cost $c(e_t)$ to implement the improvement. We assume $c'(e_t) > 0$ and $c''(e_t) \geq 0$, capturing increasing and convex costs of effort.

- **Implementation Skill (s):** The worker has implementation skill $s > 0$ that influences how effectively their effort translates into improvements.

- **Incremental Output Value ($v(se_t)$):** The implementation produces output value $v(se_t)$, where v is an increasing and concave function representing diminishing returns to skill-augmented effort, se_t.

The process continues until the worker fails to identify a new improvement opportunity (which happens with probability $1 - \gamma$ in each period). If the process lasts m periods, total output is $\sum_{t=1}^{m} v(se_t)$ or $mv(se)$ if $e_t = e$ for all t.

In each period, t, where, at the beginning of a period, an opportunity is identified, the worker chooses their effort level e_t to maximise the net benefit:

$$M(e) = v(se_t) - c(e_t) \tag{1}$$

The first term represents the output value, and the second term is the implementation cost. The worker's optimal effort level e_t^* satisfies the first-order condition:

$$v'(se_t^*) \cdot s = c'(e_t^*) \tag{2}$$

Given this set-up, as neither v nor c is time varying, $e_t^* = e^*$, the same value for all t (where opportunities are identified). Let $M^* \equiv M(e^*)$ denote the maximum expected net benefit per improvement opportunity.

The worker's continuation value - the expected benefit from the entire improvement process - can be expressed recursively:

$$V = \gamma(M^* + V) \tag{3}$$

Solving for V:

$$V = \frac{\gamma M^*}{1 - \gamma} = \frac{\gamma(v(se^*) - c(e^*))}{1 - \gamma} \tag{4}$$

The expected number of improvement iterations is $\frac{1}{1-\gamma}$, and the quality of the final output increases with each successful implementation.

Now, suppose the worker gains access to a cognitive tool; in this example, computer-aided design software. We model the tool through a parameter θ, where $\theta = 0$ represents working

without the tool and $\theta = 1$ represents working with the tool.

The tool affects both the output function and the cost function:

- $v(se_t; \theta)$ is the output value with the tool (with $v(se_t; 1) > v(se_t; 0)$ for all e_t)
- $c(e_t; \theta)$ is the implementation cost with the tool (with $c(e_t; 1) < c(e_t; 0)$ for all e_t)

Crucially, the tool functions as a substitute for cognitive effort in that:

$$\frac{v'(se_t; 1)}{c'(e_t; 1)} < \frac{v'(se_t; 0)}{c'(e_t; 0)} \tag{5}$$

This key assumption means the tool decreases the ratio of marginal benefit to marginal cost of effort, reflecting how computers reduce the incremental value of additional human exertion.

With the tool, the worker's problem becomes:

$$M(e_t; \theta) = v(se_t; \theta) - c(e_t; \theta) \tag{6}$$

The new first-order condition is:

$$v'(se^*(\theta); \theta) \cdot s = c'(e^*(\theta); \theta) \tag{7}$$

noting that $e_t^*(\theta) = e^*(\theta)$ for all relevant t. Comparing the optimal effort levels with and without the tool, we find:

$$e^*(1) < e^*(0) \quad \text{and} \quad M(e^*(1); 1) > M(e^*(0); 0) \tag{8}$$

In other words, the worker exerts less direct effort when using the computer but achieves a higher net benefit. That is, since $M(e^*(1); 1) \geq M(e^*(0); 1) = v(se^*(0); 1) - c(e^*(0); 1) > v(se^*(0); 0) - c(e^*(0); 0) = M(e^*(0); 0)$, where the first inequality follows from optimality and the second from the tool's direct benefits, we have $M(e^*(1); 1) > M(e^*(0); 0)$. This exemplifies how cognitive tools serve as "bicycles for the mind": they allow humans to achieve more with less mental exertion.

The continuation value with the tool becomes:

$$V(1) = \frac{\gamma M(e^*(1); 1)}{1 - \gamma} \tag{9}$$

Since $M(e^*(1); 1) > M(e^*(0); 0)$, it follows that $V(1) > V(0)$. The cognitive tool increases

the expected value of the entire improvement process.[3]

To illustrate with a concrete example, consider a worker designing a product. Without a computer, each calculation and drawing would require significant effort and yield limited insights. With Computer-Aided Design, the analyst can implement the same analysis with less effort, visualise results instantly, and achieve higher-quality outputs. While the analyst might invest less effort per iteration when using the software, the overall quality of their analysis improves significantly.

This simple model captures the essence of cognitive tools as enhancers of human capability. The tool amplifies the value of opportunity judgment—the ability to recognise potential improvements—by reducing the cost and increasing the benefit of acting on those opportunities. In subsequent sections, we will extend this framework to incorporate an additional form of judgment and explore more complex interactions between humans and cognitive tools.

2.2 AI-Assisted Prediction for Decision-Making

As a second motivating example, consider a decision-maker facing a choice under uncertainty. The decision-maker can exert effort to improve the accuracy of their prediction about a relevant state, and then select an appropriate action based on that prediction. This ability to identify the appropriate action in a given state is a distinct skill from implementation and from the ability to identify opportunities for improvement. It is the process emphasised in our prior work as distinct from prediction (Agrawal et al., 2019, 2018b).[4] For instance, consider a physician diagnosing and treating a patient. The physician must first diagnose the patient's underlying condition (the state) and then select an appropriate treatment based on that diagnosis.

Let the patient's true state be $\omega \in \{\texttt{Disease A}, \texttt{Disease B}\}$ with equal prior probabilities. The physician always receives a diagnostic signal $\sigma \in \{\texttt{A}, \texttt{B}\}$, but the accuracy of this signal depends on both the physician's diagnostic skill $s \in (0, 1]$ and their effort $e \geq 0$.

We model the diagnostic accuracy as $p(se)$, where $p'(se) > 0$ and $p''(se) < 0$ (increasing and concave). Specifically, $p(se) = \Pr[\sigma = \texttt{A} | \omega = \texttt{Disease A}] = \Pr[\sigma = \texttt{B} | \omega = \texttt{Disease B}]$. Note that $p(0) = \frac{1}{2}$ and the signal is uninformative (random guess) while $\lim_{e \to \infty} p(se) \to 1$, a perfectly accurate diagnosis. The physician incurs a cost $c(e)$ for exerting effort e, where $c'(e) > 0$ and $c''(e) > 0$ (increasing and convex).

[3] While here we assume that γ, for example, is the same with and without the tool If the tool directly enhanced opportunity judgment (say, if this was endogenous and relied on effort from the agent), then it could easily be the case that even if per period effort was reduced by the tool, cumulative effort could be higher.

[4] For a full examination of the AI adoption from the lense of prediction and some formal implications of judgment in specific decision contexts see Gans (2025).

When the physician makes a diagnosis, the ability to identify treatment *for the signalled condition* is represented by parameter $\alpha \in [0, 1]$. The payoff structure is as follows:

- With probability α, the physician can identify a treatment:
  - If the diagnosis is correct (probability $p(se)$), the payoff is $\Delta_{\text{Correct}} > 0$
  - If the diagnosis is incorrect (probability $1 - p(se)$), the payoff is $\Delta_{\text{Wrong}} < 0$
- With probability $1 - \alpha$, the physician cannot identify a treatment, yielding payoff 0 regardless of diagnostic accuracy

The parameter α captures what we call *payoff judgment*; that is, the decision-maker's ability to translate a diagnosis into treatment decisions.[5] It is assumed that $\Delta_{\text{Correct}} > -\Delta_{\text{Wrong}}$ so that the physician chooses to apply an identified treatment for the signalled condition.

The physician's expected payoff is:

$$\begin{aligned}
\mathbb{E}[\text{Payoff}] &= p(se)\alpha\Delta_{\text{Correct}} + (1 - p(se))\alpha\Delta_{\text{Wrong}} && (10)\\
&= \alpha[p(se)\Delta_{\text{Correct}} + (1 - p(se))\Delta_{\text{Wrong}}] && (11)\\
&= \alpha[p(se)(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) + \Delta_{\text{Wrong}}] && (12)
\end{aligned}$$

The decision-maker's problem is to maximise expected net benefit:

$$\max_{e \geq 0} M(e) = \alpha[p(se)(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) + \Delta_{\text{Wrong}}] - c(e) \tag{13}$$

For an interior solution where $p(se^*) < 1$, the first-order condition yields the optimal effort level e^*:

$$p'(se^*)s\alpha(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) = c'(e^*) \tag{14}$$

This condition reveals that the physician increases effort until the marginal benefit equals the marginal cost, where the marginal benefit is proportional to α — the ability to act on the diagnosis. When α is low, there is little incentive to improve diagnostic accuracy since the physician cannot prescribe treatment anyway. Note that $p(se^*) > \frac{1}{2}$ as $\lim_{e \to 0} c'(e) = \infty$ and $p(se^*) < 1$ holds when the marginal cost of effort is sufficiently high relative to the benefits. [6]

[5] This payoff judgment was introduced by Agrawal et al. (2019) and the application of it is discussed extensively in Agrawal et al. (2018b).

[6] Note that payoff judgment is not all upside. Specifically, when a signal of the condition σ is received, payoff judgment prescribes a treatment with probability α even when the signal is incorrect. Without payoff judgment, the physician cannot act, but with payoff judgment, the physician can act even on an incorrect signal and finds it optimal to do so. Note that if $\Delta_{\text{Wrong}} << 0$, $M(e^*)$ might be negative. For expositional

Now, suppose the decision-maker gains access to an AI diagnostic tool. Assume, as we did in the first motivating example, that $\theta = 1(0)$ represents adopting and not adopting the AI-tool, respectively. Here, the AI tool transforms diagnostic accuracy from $p(se)$ to $p(se+\theta)$. The AI effectively provides an additive boost to the physician's skill-effort product in achieving diagnostic accuracy.[7]

With the AI tool, the decision-maker's maximisation problem becomes:

$$\max_{e \geq 0} M(e;1) = \alpha[p(se+1)(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) + \Delta_{\text{Wrong}}] - c(e) \tag{15}$$

The first order condition is: $p'(se^*+1)s\alpha(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) = c'(e^*)$. This leads to the first insight from this model. Comparing the first order conditions with and without AI, we find that:

$$e^*(1) < e^*(0) \quad \text{and} \quad M(e^*(1);1) > M(e^*(0);0) \tag{16}$$

That is, optimal effort typically decreases with the use of AI assistance, but the expected net benefit increases. Intuitively, AI handles the "low-hanging fruit" of diagnosis, making additional human effort less valuable at the margin; that is, by the concavity of $p(\cdot)$, $p'(se+1) < p'(se)$ for any $e > 0$. Despite the reduction in human effort, diagnostic accuracy improves: $p(se^*(1)+1) > p(se^*(0))$, showing that AI enables better diagnoses with less human effort.[8]

The second insight is that the magnitude of these effects depends directly on α. The increase in expected benefit from AI adoption is:

$$M(e^*(1);1) - M(e^*(0);0) = \alpha\{[p(se^*(1),1) - p(se^*(0),0)](\Delta_{\text{Correct}} - \Delta_{\text{Wrong}})\} - [c(e^*(1)) - c(e^*(0))] \tag{17}$$

simplicity, we assume that this is not the case here. In a more elaborate context, it could be imagined that the treatment option was only available for one disease, and this may generate some asymmetries in this situation (Agrawal et al., 2018a).

[7]It is also possible that adopting AI impacts the cost, $c(e)$, which is something we allow for in our more general model of cognitive tools below.

[8]In this motivating example, we can prove that $p(se^*(1)+1) > p(se^*(0))$. In the more general model, complementarity will require this condition to hold. To prove this, we show that $se^*(1)+1 > se^*(0)$. From the first-order conditions:

$$\text{Without AI: } p'(se^*(0))s\alpha(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) = c'(e^*(0))$$
$$\text{With AI: } p'(se^*(1)+1)s\alpha(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) = c'(e^*(1))$$

Since $e^*(1) < e^*(0)$ and c is convex, we have $c'(e^*(1)) < c'(e^*(0))$. Therefore:

$$p'(se^*(1)+1)s\alpha(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}}) < p'(se^*(0))s\alpha(\Delta_{\text{Correct}} - \Delta_{\text{Wrong}})$$

This implies $p'(se^*(1)+1) < p'(se^*(0))$. Since p is concave, p' is decreasing, which means $se^*(1)+1 > se^*(0)$. Since p is increasing, this gives us $p(se^*(1)+1) > p(se^*(0))$.

This shows that, when $p(se^*(1)+1) > p(se^*(0))$ as is the case here, AI diagnostic tools and payoff judgment are complements: the value of AI increases with α, though the relationship is generally nonlinear since optimal effort $e^*(\theta)$ depends on α. When α is low (the physician has limited ability to identify treatment), AI provides little value regardless of how much it improves diagnostic accuracy. When α is high, more value from improved diagnosis is realised. This complementarity arises because accurate diagnosis only creates value when the physician can act on it.

In our medical example, a physician who cannot identify treatments (low α) gains little from AI-enhanced diagnosis, while a physician with high payoff judgment (high α) benefits substantially. This suggests that AI diagnostic tools should be prioritised for deployment with physicians who have the authority and knowledge to act on diagnoses, and that expanding prescribing authority may be necessary to realise the full benefits of AI diagnostic technology.

3 General Model Setup

The motivating examples provide a foundation for a more general approach to modeling task improvement using cognitive tools. Interestingly, more recent cognitive tools such as those provided by generative AI, in many respects, combine aspects of both motivating examples. For instance, writing is an inherently iterative activity that relies on the ability to generate output (with skill s and ongoing effort, e_t), to apply opportunity judgment (γ) to identify areas to write about and to be able to read that output and determine whether it "does the job" in terms of the communication of ideas (α). Generative AI assists this process by allowing writing to take place with lower effort on the part of the author. The idea of the general model is to break down a task into considering how labor (human judgment and effort) and capital (the cognitive tools) are actually used in order to distinguish between alternative human capabilities and how they relate to tool adoption and availability.

3.1 Model Setup

In this regard, we model an agent working on a task in discrete time $t = 0, 1, 2, \ldots$, making choices to improve task output or quality with a discount factor of $\delta \in [0, 1]$. In each period, the agent identifies an opportunity to improve the task and then decides how much effort to exert in implementing this improvement. The key elements of the model are:

- **Implementation Effort (e_t):** When an opportunity is identified, the agent exercises effort $e_t \geq 0$ at cost $c(e_t; \theta)$ to implement the improvement. This generates a probability

$p(se_t;\theta) \in [0,1]$ that the implementation is successful. We assume that $c'(e_t;\theta) \geq 0$ and is (weakly) convex, while $p'(se_t;\theta) \geq 0$ and is (weakly) concave, capturing diminishing returns to effort.

- **Improvement Value (Δ):** If the implementation is successful, the task quality (potentially) increases by Δ, where $\Delta > 0$ or, alternatively, may not improve at all.
- **Implementation Skill (s):** The agent has skill, $s \in (0,1]$, in implementation as captured by its inclusion in $p(se_t;\theta)$. A higher s amplifies the productivity of their effort in implementation.

As introduced in the motivating examples, the agent applies two distinct types of judgment in this process:

- **Payoff Judgment** (α): The probability that the agent can correctly evaluate and extract value from a successful implementation. With probability α, the agent realises the full improvement value Δ; with probability $1-\alpha$, the value is not realised despite successful implementation. As illustrated in our prediction example, payoff judgment represents the agent's ability to match states to optimal actions or, in our writing example, to recognise and leverage high-quality content.
- **Opportunity judgment** ($\gamma(t)$): In each round, $t \in \{0,1,...\}$, the agent perceives an opportunity to produce task output with probability $\gamma(t)$. We suppose that $\gamma(t)$ is declining in t, which reflects the notion that new opportunities are increasingly hard to come by. One special case we might focus on is where $\gamma(0) = \gamma_0$ for the initial opportunity, followed by $\gamma(t) = \gamma(< \gamma_0)$ in each subsequent round. γ characterises the agent's ability to identify improvement opportunities. Higher values of γ_0 and γ reflect stronger opportunity judgment—a greater likelihood of spotting potential enhancements.

If the agent does not generate any opportunity within a period, which occurs with probability $1-\gamma(t)$, the process ends, and the task quality remains at the level achieved by previous successful implementations. Since $\gamma(t) < 1$, eventually the process will stop.

3.2 Equilibrium Outcome

Given this, in each round, conditional on perceiving an improvement opportunity, the agent obtains the following net benefit from implementing with effort e:

$$M(e_t;\theta) = p(se_t;\theta)\alpha\Delta - c(e_t;\theta) \tag{18}$$

The first term represents the expected benefit of implementation. Note that $p(se_t;\theta)\alpha\Delta = v(se_t;\theta)$ nests the payoff in the computer-assistant design example while $\delta = 0$ and $\Delta = \Delta_{\text{Correct}} - \Delta_{\text{Wrong}}$ nests the medical diagnosis example.

The agent's problem simplifies to choosing the optimal effort level e_t to maximise $M(e_t)$:

$$e_t^*(\theta) = \arg\max_{e_t \geq 0} M(e_t;\theta) \tag{19}$$

The first-order condition for optimal effort is:

$$p'(se_t^*(\theta);\theta)s\alpha\Delta = c'(e_t^*(\theta);\theta) \tag{20}$$

Given the optimal effort level $e_t^*(\theta)$, the present expected value of task quality from $t = 0$ is:

$$V_0(\theta) = \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \gamma(i) \right) \delta^t M(e_t^*(\theta);\theta). \tag{21}$$

If, however, $\gamma(t) = \gamma$ for $t > 0$, the agent's continuation value $V_t(\theta)$ satisfies the Bellman equation:

$$V_t(\theta) = \gamma(M(e^*(\theta);\theta) + \delta V_{t+1}(\theta)) \tag{22}$$

Solving for $V_0(\theta)$:

$$V_0(\theta) = \frac{\gamma_0 M(e^*(\theta);\theta)}{1 - \delta\gamma} \tag{23}$$

Notice that it is γ and not γ_0 that drives subsequent effort choice. γ_0 does impact the overall quality of the output.[9]

3.3 Impact of Tool Adoption

Of primary interest is what tool adoption does to effort and how it interacts with the various forms of human judgment. To this end, we need to formally define what a cognitive tool does.

Definition 1. *A cognitive tool is characterised by parameter $\theta \geq 0$ such that:*

1. *For all e and $\theta' > \theta$: $p(se;\theta') \geq p(se;\theta)$ and $c(e;\theta') \leq c(e;\theta)$;*
2. *The ratio $\frac{p'(se;\theta)}{c'(e;\theta)}$ is strictly decreasing in θ for all $e > 0$.*

These properties ensure that tools both enhance productivity directly and substitute for human effort at the margin.

[9] Another possibility is that $\gamma(t) = e^{-\lambda t}$ giving rise to $V_0(\theta) = M(e^*(\theta);\theta) \sum_{t=0}^{\infty} \delta^t \exp\left(-\lambda \frac{t(t+1)}{2}\right)$.

That is, the tool increases the probability that an improvement is implemented and/or reduces the cost of implementing a task for any given effort level. Moreover, the task is a substitute for cognitive effort by the agent in implementation. While θ is initially introduced as a binary parameter (0 or 1) to represent the absence or presence of a cognitive tool, in subsequent analyses, we treat θ as a continuous parameter that can take values in $[0, \infty)$, with higher values representing more advanced or effective cognitive tools.

Given this, the following result characterises the impact of tool adoption on effort and output quality:

Proposition 1. *Let $e_t^*(\theta)$ be the optimal effort level in period t that maximises $M(e;\theta)$, and let $V_0(\theta)$ be the agent's expected task quality from $t = 0$ given opportunity sequence $\{\gamma(t)\}_{t=0}^{\infty}$. As an agent adopts a cognitive tool (θ increases from 0 to 1):*

1. *The optimal effort level in each period decreases: $e_t^*(1) < e_t^*(0)$ for all t*
2. *Moreover, $e_t^*(\theta) = e^*(\theta)$ for all t (effort is time-invariant)*
3. *The expected output quality increases: $V_0(\theta) > V_0(0)$*

The intuition for these results aligns with our motivating examples. The agent chooses the effort to balance the implementation cost against the probability of success. A cognitive tool (such as a computer or AI) decreases the marginal benefit per unit marginal cost of effort. Consequently, the agent can achieve a given probability of success with less effort; the optimal effort level e^* with the tool is lower than without it.

More output for less input means increased productivity. The cognitive tool is useful. This is consistent with evidence from the diffusion of computers and early research on the impact of AI on productivity. Computers increased productivity as shown in both macroeconomic data (Jorgenson and Stiroh, 2000; Jorgenson et al., 2005; Oliner et al., 2007) and firm-level data (Bresnahan et al., 2002). Firm-level deployment of AI also suggests increased productivity (Brynjolfsson et al., 2025; Cui et al., 2025; Dell'Acqua et al., 2023; Kanazawa et al., 2022; Noy and Zhang, 2023; McElheran et al., 2025; Czarnitzki et al., 2023; Peng et al., 2023), though there is some ambiguity about whether this has had a meaningful economy-wide impact (Humlum and Vestergaard, 2025a; Acemoglu et al., 2022).

Importantly, this optimal effort level is the same across all periods because the per-period maximisation problem is independent of time—the agent faces the same trade-off between implementation cost and success probability in each period, conditional on having identified an opportunity. However, because the tool also directly increases success probability and lowers cost, the net benefit function shifts upward. Even though the agent invests less effort, the overall net benefit increases, leading to a higher continuation value $V_0(\theta)$.

For the general opportunity sequence $\{\gamma(t)\}_{t=0}^{\infty}$, the continuation value can be expressed as:

$$V_0(\theta) = \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \gamma(i) \right) \delta^t M(e^*(\theta); \theta) \tag{24}$$

This formulation shows that while the opportunity sequence affects the expected number and timing of improvements, it does not affect the optimal effort level conditional on having an opportunity.

It is important to note, however, that unlike the earlier example where $p(se; \theta) = p(se+\theta)$ and $c(e)$ is independent of θ, it is not necessarily the case that $p(se^*(\theta); \theta)$ is increasing in θ. That is, tool use could reduce the implementation probability. To see this, let's compare the binary choice between $\theta \in \{0, 1\}$. From the first-order conditions:

$$\text{At } \theta = 0: \quad p'(se^*(0); 0)s\alpha\Delta = c'(e^*(0); 0) \tag{25}$$

$$\text{At } \theta = 1: \quad p'(se^*(1); 1)s\alpha\Delta = c'(e^*(1); 1) \tag{26}$$

From Proposition 1, we know that $e^*(1) < e^*(0)$. By Definition 1 of a cognitive tool, we have, for all e: $p(se; 1) \geq p(se; 0)$ and $c(e; 1) \leq c(e; 0)$ and that $\frac{p'(se;1)}{c'(e;1)} < \frac{p'(se;0)}{c'(e;0)}$. Note, however, at $e = e^*(0)$:

$$\frac{p'(se^*(0); 1)}{c'(e^*(0); 1)} < \frac{p'(se^*(0); 0)}{c'(e^*(0); 0)} = \frac{1}{s\alpha\Delta} \tag{27}$$

where the equality follows from (25). Since $e^*(1)$ is optimal for $\theta = 1$, we have from (26):

$$\frac{p'(se^*(1); 1)}{c'(e^*(1); 1)} = \frac{1}{s\alpha\Delta} \tag{28}$$

This means that we cannot directly compare $p(se^*(1); 1)$ and $p(se^*(0); 0)$ without additional structure. To see why, consider the decomposition:

$$p(se^*(1); 1) - p(se^*(0); 0) = \underbrace{[p(se^*(1); 1) - p(se^*(1); 0)]}_{\text{Tool effect at } e^*(1)} + \underbrace{[p(se^*(1); 0) - p(se^*(0); 0)]}_{\text{Effort effect at } \theta=0} \tag{29}$$

The first term of the RHS is non-negative by property 1 of Definition 1. The second term is negative since $e^*(1) < e^*(0)$ and p is increasing in its first argument. The net effect is ambiguous. Thus, for the inequality $p(se^*(1); 1) > p(se^*(0); 0)$ to hold, we need the tool's direct effect to dominate the effort reduction effect. This requires additional conditions on the functions p and c.[10]

[10]Below we use a functional form $p(se; \theta) = \sqrt{se + \theta}$ and $c(e) = e$ that implies that $p(se^*(1); 1) = p(se^*(0); 0)$.

3.4 Tool Adoption Drivers

Our model reveals how cognitive tools interact with different types of judgment and how the structure of opportunity arrival affects the value of tool adoption:

Proposition 2 (Tool Adoption Drivers)**.** *Let $V_0(\theta)$ be the agent's expected task quality from $t = 0$ given opportunity sequence $\{\gamma(t)\}_{t=0}^{\infty}$. Define $\Gamma = \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \gamma(i)\right) \delta^t$. Then:*

1. ***Opportunity judgment sequence (Γ)**: The value gain from tool adoption is:*

$$V_0(1) - V_0(0) = \Gamma \cdot [M(e^*(1); 1) - M(e^*(0); 0)] > 0$$

2. ***Payoff judgment (α)**: If and only if $p(se^*(1); 1) > p(se^*(0); 0)$, the value of tool adoption is non-decreasing with payoff judgment:*

$$\frac{\partial(V_0(1) - V_0(0))}{\partial \alpha} = \Gamma \cdot \Delta \cdot [p(se^*(1); 1) - p(se^*(0); 0)] \geq 0$$

3. ***Implementation skill (s)**: Under the condition that $p(se; \theta)$ has the form where $\frac{\partial^2 p}{\partial s \partial \theta} < 0$ (tool and skill are substitutes),[11] then:*

$$\frac{\partial(V_0(1) - V_0(0))}{\partial s} < 0$$

4. ***Early vs. late opportunities**: The relative importance of opportunities at different times for tool value is:*

$$\frac{\partial(V_0(1) - V_0(0))/\partial \gamma(t)}{\partial(V_0(1) - V_0(0))/\partial \gamma(t')} = \frac{\delta^t \prod_{i=0, i\neq t}^{t} \gamma(i)}{\delta^{t'} \prod_{i=0, i\neq t'}^{t'} \gamma(i)}$$

   for $t < t'$, showing that earlier opportunities contribute more to tool value due to discounting.

The signs of the parameter effects are clear: tool adoption value increases with the opportunity structure (Γ), while it decreases with implementation skill (s). Since γ is decreasing in t, the tool will be adopted early.[12] In the general model, it is ambiguous whether adoption is increasing in payoff judgment (α). This is because that requires $p(se^*(1); 1) > p(se^*(0); 0)$ that we noted earlier could not be established in the general model. Intuitively, increased

[11] This condition holds for common functional forms such as $p(se; \theta) = f(se + \theta)$ where f is concave.

[12] If, contrary to the assumption that γ is decreasing in t, $\gamma(t)$ initially increases (perhaps due to learning) before eventually declining, the optimal timing of tool adoption may be delayed.

payoff judgment is valuable when the implementation success probability is higher. However, as tool use might lead to a reduction in the realised success probability (through much lower effort), the incentives to adopt the tool may be lower when payoff judgment is high.[13]

Empirically, there is little analysis that separately examines implementation skill, opportunity judgment skill, and payoff judgment skill. An exception is Roldán-Monés (2024), who randomly provided an AI tool to participants in a university debate competition. They find that the tool helped the lower-skilled debaters differently than the higher-skilled debaters. While they do not explicitly separate implementation from judgment skills, they do have data on the different aspects of the debate score. The tool helped the lower-skilled debaters improve the clarity of their arguments. In contrast, the tool helped the high-skilled debaters with credibility, rhetoric, refutation, and the superiority of their arguments. One interpretation is that clarity is about implementation skill, and ChatGPT substituted for any advantage the highly skilled debaters had in this domain. The other categories relate to judgment, and so ChatGPT increased the advantage that the higher-skilled had in those domains. Several workplace specific studies focus on implementation skill alone, and find that providing an AI tool to workers appears to be a substitute for skill and experience (Brynjolfsson et al., 2025; Cui et al., 2025; Dell'Acqua et al., 2023; Kanazawa et al., 2022; Yiu et al., 2025). Next, we examine how differences in substitutability for different types of skills have implications for wage inequality.

4 Impact of Cognitive Tools on Wage Inequality

Thus far, we have examined how cognitive tools affect individual agent behaviour and task performance. We now examine how such tools might affect wage inequality across agents. Wage inequality is of particular interest because cognitive tools are often deployed with the goal of enhancing productivity across varying skill levels, but their distributional consequences remain underexplored. In labour markets, wages typically reflect marginal productivity, which in our framework corresponds to the expected value generated through task performance. Therefore, wage inequality can be understood through the distribution of the continuation value $V(\theta)$ across agents with heterogeneous skills.

For simplicity, the analysis in this section makes some additional assumptions. First, we make some functional form assumptions: (1) that the production function takes the form $p(se;\theta) = \sqrt{se+\theta}$ with linear costs $c(e) = e$, yielding tractable expressions for optimal effort and net benefits and (2) that $\gamma(0) = \gamma_0$, and $\gamma(t) = \gamma, \forall t > 0$. It is assumed that, for all

[13] A similar possibility of substitutability between payoff judgment and AI adoption was outlined in Agrawal et al. (2019).

agents, $\Delta > \frac{2}{s\sqrt{\alpha}}$ so that optimal effort is positive. Second, we assume agents are heterogeneous. Specifically, individuals vary in judgment and skill parameters $i \in \{\alpha, \gamma_0, \gamma, s\}$, distributed independently with positive support, means μ_i, and variances σ_i^2. Parameters satisfy $\mu_i > 3\sigma_i$ to ensure positive support. Furthermore, from this point forward, we will treat θ as a continuous parameter in $[0, \infty)$ representing the quality or sophistication of cognitive tools rather than a binary parameter (0 or 1). This allows us to analyse the effects of incremental improvements in tool capabilities on a variety of outcomes, including inequality. Finally, we introduce the opportunity multiplier $\Gamma = \frac{\gamma_0}{1-\delta\gamma}$. The following inequality will play a key role in the analysis:

$$\frac{E[\Gamma^2]}{(E[\Gamma])^2} < \mu_s E\left[\frac{1}{s}\right] \tag{30}$$

Equivalently, $1 + CV_\Gamma^2 < \mu_s E[\frac{1}{s}]$ where $CV_G = \sqrt{\frac{\text{Var}(\Gamma)}{E[\Gamma]}}$ is the coefficient of variation of Γ. When this inequality holds, implementation skill variance is large relative to opportunity judgment variance.

4.1 Tools and Inequality

The following proposition provides our main result in this section.

Proposition 3 (Cognitive Tools and Wage Inequality)**.** *Cognitive tools exhibit the following inequality effects:*

(a) ***Mean Effect:*** *Tools unambiguously increase average productivity:*

$$\frac{\partial E[V(\theta)]}{\partial \theta} = \mu_{\gamma_0} \cdot E\left[\frac{1}{1-\delta\gamma}\right] \cdot E\left[\frac{1}{s}\right] > 0 \tag{31}$$

(b) ***Variance Threshold:*** *When inequality* (30) *holds, wage variance follows a U-shaped pattern in tool quality θ. Specifically:*

$$\frac{\partial\, Var(V(\theta))}{\partial \theta}\bigg|_{\theta=0} < 0 \quad and \frac{\partial\, Var(V(\theta))}{\partial \theta}\bigg|_{\theta=1} > 0 \tag{32}$$

(c) ***Turning Point:*** *When inequality* (30) *holds, the variance-minimising tool quality is:*

$$\theta^* = \frac{\Delta^2(\mu_\alpha^2 + \sigma_\alpha^2)}{4} \cdot \frac{(E[\Gamma])^2 \mu_s E[1/s] - E[\Gamma^2]}{Var(\Gamma/s)} > 0 \tag{33}$$

The economic intuition of Proposition 3 centres on two competing forces. *Inverse skill bias* means tools help low-skill workers more through the θ/s term, creating an equalising effect. *Opportunity judgment amplification* means high opportunity judgment workers leverage tool benefits more effectively through the opportunity multiplier G, creating inequality. The U-shaped pattern emerges because inverse skill bias dominates initially, but opportunity judgment amplification eventually takes over as tools improve.[14]

The continuation value decomposes into three economically distinct components:

$$V(\theta) = \underbrace{\frac{\gamma_0}{1-\delta\gamma}}_{\text{Opportunity multiplier}} \cdot \left[\underbrace{\frac{\alpha^2\Delta^2 s}{4}}_{\text{Baseline productivity}} + \underbrace{\frac{\theta}{s}}_{\text{Tool boost}} \right] \tag{34}$$

This decomposition reveals the fundamental tension:

- **Baseline Productivity** ($\frac{\alpha^2\Delta^2 s}{4}$) reflects pre-tool inequality from complementarity between payoff judgment (α) and skill (s). High-skill workers benefit more from good payoff judgment, creating multiplicative inequality.
- **Tool Boost** ($\frac{\theta}{s}$) exhibits inverse skill bias. For example, workers with skill $s = 0.5$ receive twice the boost of workers with $s = 1$. This creates an equalising force that can counteract baseline inequality.
- **Opportunity Multiplier** ($\frac{\gamma_0}{1-\delta\gamma}$) amplifies all differences. Workers who identify more improvement opportunities (γ_0, γ) leverage both baseline skills and tool benefits more frequently, creating multiplicative effects on inequality.

The relative strength of these forces determines whether tools increase or decrease inequality, with the transition occurring at the variance-minimising point θ^* from Proposition 3(c).

This becomes clearer if we focus on limiting cases where there is homogeneous judgment and homogeneous skills respectively:

- **No judgment variance** ($\sigma_\alpha = \sigma_\gamma = 0$)**:** All workers have identical judgment $\bar{\alpha}, \bar{\gamma}$. The tool provides a uniform boost θ/s to everyone, but this boost varies inversely with skill. Inequality follows the exact U-shaped pattern from Proposition 3, with:

$$\theta^* = \frac{\bar{\alpha}^2\Delta^2}{4} \cdot \frac{\mu_s E[1/s] - 1}{\text{Var}(1/s)}$$

[14] Under the functional form specified here, α is assumed to be independent of the tool. Therefore, payoff judgment α does not affect absolute inequality. Alternatively, if a functional form was chosen so that $p(se^*(1), 1) > p(se^*(0), 0)$, then the tool could also increase inequality through complementarity with payoff judgment.

- **No skill variance ($\sigma_s = 0$):** All workers have identical skill $\bar{s}$. The tool provides a uniform boost $\theta/\bar{s}$ multiplied by heterogeneous judgment. Variance increases monotonically:

$$\frac{\partial \text{Var}(V(\theta))}{\partial \theta} = \frac{2\text{Var}(\Gamma)}{\bar{s}} \cdot \left[\frac{\Delta^2(\mu_\alpha^2 + \sigma_\alpha^2)\bar{s}}{4} + \frac{\theta}{\bar{s}} \right] > 0$$

These limiting cases demonstrate that the U-shaped inequality pattern requires meaningful skill heterogeneity interacting with the inverse skill bias of cognitive tools. When only judgment varies, tools unambiguously increase inequality by amplifying existing differences in opportunity identification.

While Proposition 3 focuses on absolute inequality (variance), relative inequality is captured by the coefficient of variation of wages: $CV_{V(\theta)} = \sqrt{\text{Var}(V(\theta))}/E[V(\theta)]$. To understand when relative inequality decreases, we need to find when $\frac{d}{d\theta} CV_{V(\theta)} < 0$. Using the quotient rule:

$$\frac{d}{d\theta} CV_{V(\theta)} = \frac{\frac{d}{d\theta}\sqrt{\text{Var}(V(\theta))} \cdot E[V(\theta)] - \sqrt{\text{Var}(V(\theta))} \cdot \frac{d}{d\theta} E[V(\theta)]}{[E[V(\theta)]]^2}$$

Since $\frac{d}{d\theta}\sqrt{\text{Var}(V(\theta))} = \frac{1}{2\sqrt{\text{Var}(V(\theta))}} \cdot \frac{d}{d\theta}\text{Var}(V(\theta))$, the condition $\frac{d}{d\theta} CV_{V(\theta)} < 0$ becomes:

$$\frac{1}{2\sqrt{\text{Var}(V(\theta))}} \cdot \frac{d\text{Var}(V(\theta))}{d\theta} \cdot E[V(\theta)] < \sqrt{\text{Var}(V(\theta))} \cdot \frac{dE[V(\theta)]}{d\theta}$$

Dividing both sides by $\sqrt{\text{Var}(V(\theta))} \cdot E[V(\theta)]$ and using $\frac{1}{x}\frac{dx}{d\theta} = \frac{d}{d\theta}\log(x)$:

$$\frac{1}{2}\frac{d\log \text{Var}(V(\theta))}{d\theta} < \frac{d\log E[V(\theta)]}{d\theta}$$

This condition is most easily satisfied when skill variance is high, baseline productivity is low relative to tool benefits, and θ is near the variance-minimising point θ^*. Thus, tools can reduce relative inequality even when absolute inequality is increasing, particularly in the transition region around θ^*. Thus, the inequality implications depend critically on the measurement choice and the tool quality level.

4.2 Connection to the Literature on Information Technology and Inequality

This section has documented that the impact of a cognitive tool on inequality depends on the distribution of human skills with respect to implementation and opportunity judgment. As a cognitive tool gets better, inequality may decline as the tool replaces the value created by highly skilled humans at implementation. Eventually, if the tool improves enough,

inequality may begin to increase again as the humans with high judgment benefit from the complementarity between judgment and the tool.

This result helps interpret two distinct literatures: the empirical literature on whether computers and AI have increased or decreased labour market outcomes and the literature that examines current labour markets to anticipate how AI might impact jobs and inequality in the future.

4.2.1 Computers, AI, and Inequality

The empirical literature suggests that computers increased labour market inequality by automating routine tasks, because non-routine cognitive skills were rare relative to non-routine manual skills (Autor et al., 2006, 2008). Goldin and Katz (2008) argue that education did not keep up with increasing demand for non-routine skills, and therefore labour market inequality increased. The automation of human routine actions is analogous to implementation effect in our model, and may underlie the evidence that factory robots in the United States led to increased inequality through the automation of human actions (Acemoglu and Restrepo, 2018, 2020). The increased value of non-routine cognitive tasks includes the ability to recognise opportunities to increase the quality of work. Such opportunity judgment may underlie evidence that computer adoption was primarily valuable to those with access to skilled expertise (Dranove et al., 2014; Forman et al., 2012).

Overall, the empirical literature on computers and inequality, when interpreted through the lens of Proposition 3, suggests that the variance of judgment skills is sufficiently large relative to the variance of implementation skills that inequality rose as a consequence of the cognitive tool of computers.

In contrast, the empirical literature on AI suggests that AI has generally reduced productivity differences. Much of this literature examines what happens to different workers within the same workplace after the adoption of AI. For example, Brynjolfsson et al. (2025) show that newer and less productive call centre employees benefit most from AI recommendations, and Cui et al. (2025) show that less experienced developers benefit most from an AI-based coding assistant. Others with similar workplace results include Hui et al. (2024), Wiles et al. (2023), Dell'Acqua et al. (2023), Peng et al. (2023), and Kanazawa et al. (2022). Experimental results on knowledge workers (Dell'Acqua et al., 2023) and in an online sample (Noy and Zhang, 2023) also suggest decreasing inequality in conducting knowledge work tasks.[15] This literature is consistent with Proposition 2 part 3, demonstrating that implementation

[15]Data on adoption are ambiguous. Cui et al. (2025) show that the less experienced adopt earlier. In a study of ChatGPT adoption by thousands of Danish workers, Humlum and Vestergaard (2025b) also show that less experienced workers adopt first; however, women used it much less and lower-wage workers used it slightly less.

skill and cognitive tools are substitutes. This literature therefore suggests that AI reduces productivity gaps due to differences in implementation skill.

Three papers have found suggestive evidence that AI exacerbates existing labour market inequality. Otis et al. (2023) examine AI business advice for small and medium-sized enterprises in Kenya. They find that the lowest-performing businesses are not helped by AI advice, suggesting a lower bound on the skills needed to benefit from the AI tool. Tranchero (2024) shows that 'data-driven predictions' increase corporate innovation more in companies that have deeper domain knowledge. Roldán-Monés (2024) randomly provided an AI tool to participants in a college debate competition, finding that the tool helped the lower-skilled debaters differently than the higher-skilled debaters. While the analysis does not explicitly separate implementation from judgment skills, it does examine data on the different aspects of the debate score. The tool helped the lower-skilled debaters improve the clarity of their arguments. In contrast, the tool helped the high-skilled debaters with credibility, rhetoric, refutation, and the superiority of their arguments. To the extent that clarity involves the text of the speeches, while credibility and argument quality are about refinements, then this is suggestive evidence of both effects: the AI tool substituted for implementation skill and complemented judgment.

On balance, the empirical evidence on AI's impact thus far suggests decreased inequality, with exceptions for the very unskilled (Otis et al., 2023) and for some applications where judgment seems particularly valuable, such as corporate science (Tranchero, 2024) and intellectual competition (Roldán-Monés, 2024). AI adoption is still relatively new. Therefore, it may be that this is the first phase of reduced inequality that the model predicts as the value of high skills in implementation declines, but before the value of high skills in judgment arises.

4.2.2 Interpreting the task-based model

This difference between an earlier phase of reduced inequality, as cognitive tools substitute for those with high implementation skills, and a later phase of increased inequality, as those tools empower those with better judgment, also helps interpret a variety of papers that use a task-based model to anticipate how AI will affect jobs and inequality in the future. These papers build on the task-based approach to labour markets (Acemoglu and Autor, 2011), which decomposed jobs into tasks. Using data on the tasks involved in a variety of jobs, for example, through O*NET, these papers then identify which tasks are likely to be done by AI in the near future and therefore which jobs are most at risk of change. Building on Frey and Osborne (2017), this method has been refined by Brynjolfsson and Mitchell (2017), Felten et al. (2021), Eloundou et al. (2024), and Handa et al. (2025).

Overall, these papers suggest that high-wage occupations are at relatively high risk of AI-induced change. The papers are agnostic as to whether 'change' is good or bad for current workers in those occupations. Interpreted through the lens of Proposition 3, these results anticipate the first phase in which cognitive tools substitute for implementation skills in existing workflows. They do not anticipate the second phase in which variance in judgment may increase inequality because judgment (the ability to recognise opportunities and to identify payoffs of various actions) is not typically listed as a task.

Thus, this task-based approach to understanding AI's impact is, by design, focused on the phase when AI automates implementation and therefore inverse skill bias may apply. It does not examine the longer term, as implementation becomes automated and judgment differences become more important.

5 Cognitive Tools and Automation

Our analysis has focused on cognitive tools that augment human capabilities - technologies that enhance the effectiveness of human judgment and effort without replacing them. However, a natural question arises: when might full automation, where machines operate without human intervention, dominate the human-tool partnership we have analysed? This section develops a framework to address this question by examining the fundamental trade-off between judgment flexibility and implementation consistency.

The key insight is that automation requires judgment to be pre-specified - encoded into algorithms, training data, or decision rules before the task begins. This pre-specification creates an inherent disadvantage: automated systems cannot adapt their judgment parameters to match evolving task requirements the way humans can. We formalise this limitation and derive conditions under which automation's advantages in consistency and cost nonetheless make it preferable to human-operated tools.

5.1 The Flexibility Gap in Automated Judgment

Consider our motivating examples through the lens of automation. In the computer-aided design example, a human designer using CAD software can dynamically adjust their opportunity judgment, recognising when a design needs fundamental rethinking versus incremental refinement. An automated design system, by contrast, must operate with pre-programmed rules about when and how to modify designs.

Similarly, in medical diagnosis, a physician can adapt their judgment to unusual cases, considering context and nuance that may not have been anticipated. An automated diag-

nostic system, even one trained on vast datasets, operates with fixed parameters that cannot adjust judgment to novel situations. In content generation, human writers can recognise when their approach isn't working and pivot accordingly, while automated systems follow their prespecified judgment regardless of context.

To formalise this flexibility gap, we model automated systems as operating with pre-specified judgment parameters:

Definition 2 (Pre-Specified Judgment)**.** *An automated system operates with judgment parameters $\hat{\gamma}$ and $\hat{\alpha}$ that are determined before task execution. These relate to optimal human judgment parameters through flexibility constraints:*

$$\hat{\gamma} = \rho_\gamma \cdot \gamma, \quad \hat{\alpha} = \rho_\alpha \cdot \alpha \tag{35}$$

where $\rho_\gamma, \rho_\alpha \in (0, 1]$ capture the loss of flexibility, and γ, α represent the judgment levels a human would achieve for the specific task instance.

When $\rho_\gamma = \rho_\alpha = 1$, pre-specified judgment perfectly matches what a human is endowed with - an ideal but typically unattainable case. More realistically, $\rho_\gamma, \rho_\alpha < 1$, reflecting that automated judgment cannot perfectly anticipate the optimal parameters for every situation. Thus, the automated system misses some opportunities and also misses some optimal actions to match a given implementation outcome.

The automated system's value function parallels our human model, but with crucial differences:

$$V_{\text{auto}} = \frac{\hat{\gamma}}{1 - \delta\hat{\gamma}} \cdot [p(\hat{e}) \cdot \hat{\alpha} \cdot \Delta - c_a] \tag{36}$$

where $p(\hat{e})$ is the implementation success probability with automated effort $\hat{e}$, and c_a represents the (typically low) cost of automated operation.

5.2 Deriving the Automation Condition

The decision between automation and human-tool collaboration involves weighing automation's advantages against the flexibility gap. Automation offers three potential advantages:

1. **Implementation Consistency**: Automated systems can maintain consistent performance without fatigue or variation, potentially achieving $p(\hat{e}) > p(se^*; \theta)$ despite lower effort flexibility.

2. **Cost Efficiency**: The operational cost c_a may be much lower than human implementation costs $c(e^*; \theta)$.

3. **Scale**: Once developed, automated systems can operate continuously and in parallel, though this advantage is outside our single-agent model.

Against these advantages stands the fundamental limitation: automated systems cannot adapt their judgment to context. This inflexibility becomes more costly as task environments become more variable and novel situations more common.

We now derive precise conditions for when automation's advantages overcome the flexibility gap:

Proposition 4 (Automation versus Human-Tool Collaboration)**.** *Define the dynamic adjustment factor:*

$$\Phi(\rho_\gamma) = \frac{1 - \delta\gamma}{1 - \delta\rho_\gamma\gamma} \leq 1 \tag{37}$$

Automation is preferred when:

$$\rho_\gamma \cdot \Phi(\rho_\gamma) \cdot \frac{p(\hat{e}) \cdot \rho_\alpha\alpha \cdot \Delta - c_a}{p(se^*;\theta) \cdot \alpha \cdot \Delta - c(e^*;\theta)} > 1 \tag{38}$$

This condition is more likely to be satisfied when:

1. *The flexibility parameters ρ_γ, ρ_α are close to 1 (minimal judgment degradation)*
2. *The baseline human performance $p(se^*;\theta)\alpha\Delta - c(e^*;\theta)$ is low relative to automation performance $p(\hat{e}) \cdot \rho_\alpha\alpha \cdot \Delta - c_a$.*

This proposition reveals that automation faces a formidable challenge. To dominate human-tool collaboration, it must overcome three multiplicative penalties: the opportunity identification penalty ($\rho_\gamma < 1$), the dynamic adjustment penalty ($\Phi(\rho_\gamma) < 1$) which captures how inflexible opportunity identification compounds over time, and the payoff judgment penalty ($\rho_\alpha < 1$) embedded in the numerator. These penalties multiply together, creating a substantial hurdle that automation must compensate for through either superior implementation ($p(\hat{e}) > p(se^*;\theta)$) or dramatic cost savings ($c_a \ll c(e^*;\theta)$).

This result relates to research on the brittleness of fully automated systems when the stakes are high (Agrawal et al., 2024). As Crootof et al. (2023) emphasises in the context of safety-critical systems, "an algorithm's inherent brittleness and possible ineptness in addressing long tail events may result in inaccurate determinations" (p. 476). Similarly, Scharre (2016) discusses ways to embed human judgment, as moral agents or as fail-safes, on a continuous basis into military AI applications.

5.3 The Paradox of Tool Improvement

A striking implication emerges: as cognitive tools improve (increasing θ), the automation condition becomes *harder* to satisfy. Better tools increase $V_{\text{human}}(\theta)$ (the payoff from human-tool collaboration) by making human judgment more effective, widening the gap that automation must overcome. Specifically, improved tools reduce human effort costs while maintaining judgment flexibility; the value of human judgment (both γ and α) gets amplified by better tools, and automation's fixed judgment parameters become an increasingly binding constraint.

This suggests a paradox: *advances in cognitive tools may actually reduce rather than increase automation adoption*. As tools become more powerful at augmenting human judgment, the flexibility advantage of human-tool partnerships becomes more valuable, making full automation less attractive even as the underlying technology improves.

This insight helps explain patterns in technology adoption. Full automation has proven difficult in a variety of settings because technology substitutes for labour in some tasks but complements labour in others (Autor, 2015; Autor and Thompson, 2025). Bessen (2015) showed that automated teller machines led to an increase in the number of humans involved in retail banking. McCullough et al. (2016) emphasises the complementary role of labour in the successful deployment of healthcare information technology. Feigenbaum and Gross (2024) showed that fully automating call switching at AT&T took nearly 100 years.

Early automation through AI has succeeded in some routine, predictable tasks where judgment requirements were minimal and stable, such as financial fraud detection (Agrawal et al., 2022). But in complex, judgment-intensive domains - from medical diagnosis to strategic planning- the progression has been toward increasingly sophisticated human-tool partnerships rather than full automation. Our framework suggests this pattern may persist: the better cognitive tools become at amplifying human judgment, the more valuable that judgment flexibility becomes relative to automation's implementation advantages.

6 Team Production with Specialised Judgment

Thus far, we have analysed how a cognitive tool impacts a single agent's iterative task improvement process. In practice, however, many complex tasks are performed by teams of specialists rather than individuals. This raises important organisational questions: How should teams allocate control over the implementation process? How do cognitive tools affect the optimal team structure and decision rights?

In this section, we extend our model to a team setting where different members specialise

in different types of judgment. This extension allows us to examine the organisational implications of tool adoption, particularly how tools reshape the optimal allocation of decision rights and team composition.

6.1 Model Extension: Judgment Specialisation in Teams

Consider a team with two specialists, each excelling in one of the judgment types identified in our baseline model:

- *Opportunity Specialist (OS)*: Has superior ability to identify improvement opportunities, represented by a higher $\gamma_{OS} > \gamma_{PS}$
- *Payoff Specialist (PS)*: Has superior ability to evaluate and extract value from improvements, represented by a higher $\alpha_{PS} > \alpha_{OS}$

Each specialist $i \in \{OS, PS\}$ has implementation capability represented by a cost function $c_i(e)$ and success probability $p_i(e)$. Communication between specialists incurs a cost κ per message, reflecting coordination overhead.

A critical organisational design question is: which specialist should control the implementation process (i.e., choose e)? This decision affects both the quality of implementation and the pattern of communication costs within the team. The communication structure depends on who controls implementation:

- When the PS controls implementation, the OS must communicate each identified opportunity, incurring communication cost κ with probability γ_{OS}
- When the OS controls implementation, communication to the PS is needed only when implementation succeeds (to evaluate payoffs), incurring cost κ with probability $p_{OS}(e_{OS})$

Crucially, this asymmetric communication structure affects not only direct costs but also effort incentives. When the OS controls implementation, the communication cost creates a "tax" on successful implementation, reducing the net value of success and thereby diminishing effort incentives.

6.2 Optimal Assignment of Implementation Control

To determine optimal control, we must account for both direct effects (implementation capabilities and communication costs) and indirect effects (altered effort incentives). When

the PS controls implementation, their optimal effort e^*_{PS} satisfies:

$$p'_{PS}(e^*_{PS})\alpha_{PS}\Delta = c'_{PS}(e^*_{PS})$$

The communication cost κ is fixed per opportunity and does not affect marginal effort incentives. The team's expected value is:

$$V_{PS} = \frac{\gamma_{OS}}{1 - \delta\gamma_{OS}} \cdot [p_{PS}(e^*_{PS})\alpha_{PS}\Delta - c_{PS}(e^*_{PS}) - \kappa]$$

When the OS controls implementation, the communication cost is incurred only upon successful implementation, effectively reducing the payoff from success. The optimal effort e^*_{OS} satisfies:

$$p'_{OS}(e^*_{OS})(\alpha_{PS}\Delta - \kappa) = c'_{OS}(e^*_{OS})$$

The team's expected value is:

$$V_{OS} = \frac{\gamma_{OS}}{1 - \delta\gamma_{OS}} \cdot [p_{OS}(e^*_{OS})\alpha_{PS}\Delta - c_{OS}(e^*_{OS}) - p_{OS}(e^*_{OS})\kappa]$$

Proposition 5 (Optimal Control with Communication and Effort Effects)**.** *The Opportunity Specialist should control implementation when:*

$$p_{OS}(e^*_{OS})\alpha_{PS}\Delta - c_{OS}(e^*_{OS}) - p_{OS}(e^*_{OS})\kappa > p_{PS}(e^*_{PS})\alpha_{PS}\Delta - c_{PS}(e^*_{PS}) - \kappa$$

When both specialists have identical implementation technologies ($p_{OS}(\cdot) = p_{PS}(\cdot)$ and $c_{OS}(\cdot) = c_{PS}(\cdot)$), effort levels satisfy $e^*_{OS} < e^*_{PS}$ due to the communication tax reducing the OS's marginal benefit of effort.

The proposition reveals a fundamental tension in organisational design. Even when the OS has superior implementation capabilities, giving them control creates two countervailing forces: they save on communication when implementations fail (which favours OS control when success rates are low), but they also exert less effort due to the reduced net value of success (which favours PS control).

To illustrate the effort effect, consider the case where both specialists have identical implementation technologies. Under PS control, effort maximises $p(e)\alpha_{PS}\Delta - c(e)$. Under OS control, effort maximises $p(e)(\alpha_{PS}\Delta - \kappa) - c(e)$. The communication tax κ reduces the marginal benefit of effort, leading to strictly lower effort under OS control. This effort penalty must be weighed against any implementation advantages the OS might possess.

6.3 Impact of Tool Quality on Implementation Control

The analysis thus far has examined how team structure affects performance for a given level of tool quality. We now turn to the core question: how does the optimal allocation of control evolve as cognitive tools improve? This question has important implications for organisational design in an era of rapidly advancing AI capabilities.

Cognitive tools impact the two primary determinants of optimal control—implementation capabilities and communication costs—in fundamentally different ways. Understanding these competing forces is crucial for predicting how organisations might restructure as tools advance.

First, consider how tools affect implementation capabilities. As we have seen throughout this paper, cognitive tools act as substitutes for human implementation effort. This substitution tends to compress differences across workers. A specialist who excels at implementation when tools are primitive may find their advantage eroded as tools handle more of the technical complexity. Meanwhile, a specialist with weaker implementation skills benefits disproportionately, as the tool compensates for their limitations. This convergence effect suggests that as tools improve, control should shift toward specialists with superior judgment rather than superior implementation skills.

Second, consider how tools affect communication costs. Recall that in our team model, communication patterns depend on who controls implementation. When the Payoff Specialist controls, the Opportunity Specialist must communicate every identified opportunity, incurring cost κ with probability γ_{OS}. When the Opportunity Specialist controls, communication occurs only after successful implementation, incurring cost κ with probability $p_{OS}(e_{OS})$. As tools improve and success rates increase, this asymmetry becomes less important—when p_{OS} approaches one, both control structures require nearly the same amount of communication. This suggests that at very high tool quality, even small residual advantages in implementation or coordination might determine optimal control.

These two forces—capability convergence and communication cost convergence—can work in opposite directions. Early in a technology's lifecycle, when implementation is difficult and failures common, control typically resides with whoever has the strongest implementation capabilities. As tools improve, the convergence of implementation capabilities may shift control to those with better judgment. But further tool improvements that push success rates toward certainty can again reverse this shift, as communication costs become less relevant and other factors dominate.

6.3.1 A Specific Example

To make these competing forces concrete, let us work through our model with specific functional forms. Consider a team where both specialists use the technology $p(e;\theta) = \sqrt{se+\theta}$ with cost $c(e) = e$. To capture some heterogeneity while maintaining tractability, suppose the Opportunity Specialist has a small cost advantage: $c_{OS}(e) = e - \epsilon$ where $\epsilon > 0$ is small. This could reflect familiarity with the task domain or slightly more efficient work habits.

Under this specification, we can solve for optimal effort levels. When the Payoff Specialist controls implementation, they maximize:

$$M_{PS}(e) = \alpha_{PS}\Delta\sqrt{se+\theta} - e - \kappa \tag{39}$$

yielding optimal effort:

$$e^*_{PS}(\theta) = \frac{\alpha^2_{PS}\Delta^2 s}{4} - \frac{\theta}{s} \tag{40}$$

When the Opportunity Specialist controls, they face a communication tax on successful implementation, maximising:

$$M_{OS}(e) = (\alpha_{PS}\Delta - \kappa)\sqrt{se+\theta} - (e - \epsilon) \tag{41}$$

yielding:

$$e^*_{OS}(\theta) = \frac{(\alpha_{PS}\Delta - \kappa)^2 s}{4} - \frac{\theta}{s} \tag{42}$$

Several insights emerge from this example. First, the communication tax creates a persistent effort wedge: $e^*_{OS}(\theta) < e^*_{PS}(\theta)$ for all tool qualities. The Opportunity Specialist, knowing they must pay κ upon success, exerts less effort than the Payoff Specialist who faces a fixed communication cost. This organisational friction cannot be eliminated by better tools.

Second, with these functional forms, success probabilities take particularly simple forms:

$$p_{PS}(e^*_{PS}(\theta);\theta) = \frac{\alpha_{PS}\Delta s}{2} \tag{43}$$

$$p_{OS}(e^*_{OS}(\theta);\theta) = \frac{(\alpha_{PS}\Delta - \kappa)s}{2} \tag{44}$$

Interestingly, these probabilities are constant in θ — the tool's direct positive effect on success exactly offsets the reduction in optimal effort. While this specific result is an artifact of our functional form, it illustrates an important principle: tool improvements need not monotonically increase success rates when effort adjusts endogenously.

The net advantage of Opportunity Specialist control is:

$$\Omega(\theta) = M^*_{OS} - M^*_{PS} = -\frac{s\alpha_{PS}\Delta\kappa}{2} + \frac{s\kappa^2}{4} + \epsilon + \kappa \tag{45}$$

To understand this expression, note that with our functional forms:

- Success probability under PS control: $p^*_{PS} = \frac{\alpha_{PS}\Delta s}{2}$
- Success probability under OS control: $p^*_{OS} = \frac{(\alpha_{PS}\Delta - \kappa)s}{2}$

The net advantage can be decomposed as:

$$\Omega(\theta) = \underbrace{\epsilon - \frac{s\alpha_{PS}\Delta\kappa}{2} + \frac{s\kappa^2}{4}}_{\text{Implementation effect including communication tax}} + \underbrace{\kappa}_{\text{Fixed communication cost saved}} \tag{46}$$

The first term captures two effects: (i) the OS's direct cost advantage ϵ, and (ii) the efficiency loss from the communication tax, which reduces the OS's effective payoff per success from $\alpha_{PS}\Delta$ to $(\alpha_{PS}\Delta - \kappa)$, leading to lower effort and success probability. The second term κ represents the fixed communication cost that PS must pay per opportunity but OS avoids when implementation fails. For OS control to be optimal, we need $\Omega(\theta) > 0$, which requires:

$$\epsilon + \kappa > \frac{s\alpha_{PS}\Delta\kappa}{2} - \frac{s\kappa^2}{4} \tag{47}$$

With our specific functional forms, $\Omega(\theta)$ is independent of θ, suggesting that control assignment would not change as tools improve under these assumptions.

However, this invariance result should not be interpreted too literally. It emerges from the knife-edge property that $p(e;\theta) = \sqrt{se + \theta}$ exhibits constant returns to the combined input $se + \theta$. More realistic production functions would likely exhibit either (i) diminishing returns that become more pronounced at high θ, causing success probabilities to increase toward one; (ii) threshold effects where very high tool quality enables qualitatively different approaches and (iii) interaction effects where tools complement some skills more than others.

6.3.2 Broader Implications

This simplified model is useful because it reveals several important principles for organisational design. First, even with perfect tools, the structure of communication costs creates lasting differences in incentives. The specialist who must communicate upon success will always exert less effort than one who faces fixed communication costs.

Second, while our example yields constant control, it illustrates how different forces pull in opposite directions. Implementation advantages, judgment quality, and communication patterns interact in complex ways. Small changes in functional forms or parameter values could easily generate situations where control shifts from OS to PS and potentially back again as tools improve.

Finally, our analysis highlights how the timing of communication, before or after implementation, affects optimal effort and control. Tasks where coordination is needed upfront differ fundamentally from those where it occurs after results are known. Tool improvements affect these task types differently.

Real-world examples abound. In journalism, early desktop publishing tools shifted control from typesetters (implementation specialists) to editors (judgment specialists). But modern content management systems are so streamlined that individual writers often control the entire publication process, reducing coordination needs. Similar patterns appear in architecture (from drafters to designers to AI-assisted generalists), finance (from quants to strategists to automated systems with human oversight), and many other fields.

Empirical research has documented how cognitive tools change the nature of teams (Hoffmann et al., 2024; Dell'Acqua et al., 2025), favouring some workers over others. Teodoridis (2018) showed that a new tool can change team structure in the sciences. When Microsoft's X-Box Kinect machine vision tool became available to researchers, specialists in machine vision published less and generalists who were better able to identify opportunities to use the new tool benefited more. Put differently, opportunity judgment became more valuable while implementation skill became less valuable. The central role of communication costs is consistent with Deming (2017), who documented the increasing importance of social skills over the time period that computers became a longer part of the workforce, and Law and Shen (2025), who show that AI adoption increased the value of soft skills such as customer service relative to hard skills such as database skills.

The broader lesson is that cognitive tools do not simply make existing organisational structures more efficient—they fundamentally alter the comparative advantages that determine optimal control. Organisations must therefore view structure as dynamic, adapting not just to current tool capabilities but anticipating how future improvements will reshape the balance between implementation skills, judgment capabilities, and coordination costs.

7 Conclusion

This paper provides a unifying economic framework for understanding how cognitive tools, specifically computers and artificial intelligence, interact with human capabilities in iterative

task improvement. By decomposing human skills into implementation ability, opportunity judgment, and payoff judgment, we demonstrate that cognitive tools consistently substitute for implementation skills while complementing opportunity judgment. The relationship with payoff judgment proves more nuanced, depending on whether tools reduce human effort sufficiently to offset their productivity benefits.

These insights help reconcile seemingly contradictory empirical findings: computers increased both productivity and inequality by amplifying the value of non-routine cognitive skills, while AI often reduces inequality within adopting firms by disproportionately helping workers with weaker implementation skills. Our model predicts that this equalising effect may reverse as AI tools become more sophisticated, eventually favouring workers with superior judgment capabilities and creating a U-shaped pattern in wage inequality.

Our approach complements Autor and Thompson (2025) and responds to the call in Agarwal et al. (2024) for economists to specify the ways in which AI is interdependent with human decision-making skills. Our framework highlights how human judgment complements AI's shortcomings (Loaiza and Rigobon, 2024). This complementarity explains why full automation remains elusive despite rapid technological progress: The benefit of responsive human judgment rises as AI improves. By explicitly modelling the interaction between human judgment and machine capabilities, our approach offers a foundation for analysing the economic mechanisms through which technological advances reshape production, organisational design, and labour market outcomes across a variety of contexts.

A Appendix: Proofs of Propositions

A.1 Proof of Proposition 1

Part 1: The optimal effort level decreases with tool adoption. We need to show that $e^*(\theta) < e^*(0)$ for $\theta > 0$. From the first-order conditions:

$$\text{At } \theta: \quad p'(se^*(\theta);\theta)s\alpha\Delta = c'(e^*(\theta);\theta) \tag{48}$$

$$\text{At } 0: \quad p'(se^*(0);0)s\alpha\Delta = c'(e^*(0);0) \tag{49}$$

Dividing both sides by $s\alpha\Delta > 0$:

$$\frac{p'(se^*(\theta);\theta)}{c'(e^*(\theta);\theta)} = \frac{1}{s\alpha\Delta} \tag{50}$$

$$\frac{p'(se^*(0);0)}{c'(e^*(0);0)} = \frac{1}{s\alpha\Delta} \tag{51}$$

Now, by the definition of a cognitive tool, for any given effort level e:

$$\frac{p'(se;\theta)}{c'(e;\theta)} < \frac{p'(se;0)}{c'(e;0)}$$

In particular, this holds at $e = e^*(0)$:

$$\frac{p'(se^*(0);\theta)}{c'(e^*(0);\theta)} < \frac{p'(se^*(0);0)}{c'(e^*(0);0)} = \frac{1}{s\alpha\Delta}$$

Since we also know that:

$$\frac{p'(se^*(\theta);\theta)}{c'(e^*(\theta);\theta)} = \frac{1}{s\alpha\Delta} > \frac{p'(se^*(0);\theta)}{c'(e^*(0);\theta)}$$

This implies:

$$\frac{p'(se^*(\theta);\theta)}{c'(e^*(\theta);\theta)} > \frac{p'(se^*(0);\theta)}{c'(e^*(0);\theta)}$$

Now observe that the ratio $\frac{p'(se;\theta)}{c'(e;\theta)}$ is decreasing in e because:

- $p'(se;\theta)$ is decreasing in e (since p is concave)
- $c'(e;\theta)$ is increasing in e (since c is convex)

Since the ratio is decreasing in e and we have:

$$\frac{p'(se^*(\theta);\theta)}{c'(e^*(\theta);\theta)} > \frac{p'(se^*(0);\theta)}{c'(e^*(0);\theta)}$$

We must have $e^*(\theta) < e^*(0)$.

Part 2: Optimal effort is time-invariant. In each period t, conditional on having identified an opportunity, the agent's problem is:

$$\max_{e_t \geq 0} M(e_t;\theta) = p(se_t;\theta)\alpha\Delta - c(e_t;\theta)$$

The first-order condition is:

$$p'(se_t^*(\theta);\theta)s\alpha\Delta = c'(e_t^*(\theta);\theta)$$

This optimization problem is independent of the time period t, the opportunity identification probability $\gamma(t)$ and any continuation values. Since the agent faces the same static trade-off in each period conditional on having an opportunity, the optimal effort is the same in every period: $e_t^*(\theta) = e^*(\theta)$ for all t.

Part 3: Expected output quality increases. The continuation value from $t = 0$ is:

$$V_0(\theta) = \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \gamma(i) \right) \delta^t M(e_t^*(\theta);\theta)$$

Since $e_t^*(\theta) = e^*(\theta)$ for all t (from Part 2), we can factor out $M(e^*(\theta);\theta)$:

$$V_0(\theta) = M(e^*(\theta);\theta) \cdot \underbrace{\sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \gamma(i) \right) \delta^t}_{=\Gamma > 0}$$

To show $V_0(\theta) > V_0(0)$, it suffices to show $M(e^*(\theta);\theta) > M(e^*(0);0)$. Since $e^*(\theta)$ maximizes $M(e;\theta)$, we have:

$$M(e^*(\theta);\theta) \geq M(e^*(0);\theta)$$

Now compare $M(e^*(0);\theta)$ with $M(e^*(0);0)$:

$$M(e^*(0);\theta) = p(se^*(0);\theta)\alpha\Delta - c(e^*(0);\theta) \tag{52}$$

$$M(e^*(0);0) = p(se^*(0);0)\alpha\Delta - c(e^*(0);0) \tag{53}$$

By the definition of a cognitive tool: $p(se^*(0);\theta) \geq p(se^*(0);0)$ and $c(e^*(0);\theta) \leq c(e^*(0);0)$ with at least one inequality being strict. Therefore: $M(e^*(0);\theta) > M(e^*(0);0)$

Combining these results:

$$M(e^*(\theta);\theta) \geq M(e^*(0);\theta) > M(e^*(0);0)$$

Thus $M(e^*(\theta);\theta) > M(e^*(0);0)$, which implies:

$$V_0(\theta) = M(e^*(\theta);\theta) \cdot \Gamma > M(e^*(0);0) \cdot \Gamma = V_0(0)$$

A.2 Proof of Proposition 2

Part 1: Opportunity sequence structure. From Proposition 1, we established that $e_t^*(\theta) = e^*(\theta)$ for all t. Therefore:

$$V_0(\theta) = \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \gamma(i) \right) \delta^t M(e^*(\theta);\theta) \tag{54}$$

$$= M(e^*(\theta);\theta) \cdot \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \gamma(i) \right) \delta^t \tag{55}$$

$$= M(e^*(\theta);\theta) \cdot \Gamma \tag{56}$$

Thus:

$$V_0(1) - V_0(0) = \Gamma \cdot [M(e^*(1);1) - M(e^*(0);0)]$$

Since $M(e^*(1);1) > M(e^*(0);0)$ (from Proposition 1) and $\Gamma > 0$, the value gain is positive.

Part 2: Payoff judgment. Taking the derivative with respect to α:

$$\frac{\partial(V_0(1) - V_0(0))}{\partial\alpha} = \Gamma \cdot \frac{\partial}{\partial\alpha}[M(e^*(1);1) - M(e^*(0);0)]$$

For any θ, the net benefit is $M(e^*(\theta);\theta) = p(se^*(\theta);\theta)\alpha\Delta - c(e^*(\theta);\theta)$.

By the envelope theorem, at the optimal effort level:

$$\frac{\partial M(e^*(\theta);\theta)}{\partial\alpha} = p(se^*(\theta);\theta)\Delta$$

This is because the first-order condition $p'(se^*(\theta);\theta)s\alpha\Delta = c'(e^*(\theta);\theta)$ ensures that the

indirect effects through $\frac{\partial e^*}{\partial \alpha}$ vanish at the optimum. Therefore:

$$\frac{\partial(V_0(1) - V_0(0))}{\partial \alpha} = \Gamma \cdot \Delta \cdot [p(se^*(1); 1) - p(se^*(0); 0)]$$

Part 3: Implementation skill. We want to show $\frac{\partial(V_0(1)-V_0(0))}{\partial s} < 0$ under appropriate conditions. Using the envelope theorem:

$$\frac{\partial(V_0(1) - V_0(0))}{\partial s} = \Gamma \cdot \left[\frac{\partial M(e^*(1); 1)}{\partial s} - \frac{\partial M(e^*(0); 0)}{\partial s}\right]$$

At the optimal effort:

$$\frac{\partial M(e^*(\theta); \theta)}{\partial s} = \alpha\Delta \cdot \frac{\partial p(se^*(\theta); \theta)}{\partial s}$$

Therefore:

$$\frac{\partial(V_0(1) - V_0(0))}{\partial s} = \Gamma \cdot \alpha\Delta \cdot \left[\frac{\partial p(se^*(1); 1)}{\partial s} - \frac{\partial p(se^*(0); 0)}{\partial s}\right]$$

For this difference to be negative, we need:

$$\frac{\partial p(se^*(1); 1)}{\partial s} < \frac{\partial p(se^*(0); 0)}{\partial s}$$

This inequality holds when the production function satisfies $\frac{\partial^2 p(se;\theta)}{\partial s \partial \theta} < 0$, meaning that increases in tool quality reduce the marginal productivity of skill.

To see why this condition is natural, consider the class of production functions $p(se; \theta) = f(se + \theta)$ where f is increasing and concave. For these functions:

$$\frac{\partial p}{\partial s} = e \cdot f'(se + \theta)$$

The cross-partial is:

$$\frac{\partial^2 p}{\partial s \partial \theta} = e \cdot f''(se + \theta) < 0$$

The inequality follows from the concavity of f (i.e., $f'' < 0$). This captures the idea that as the tool improves (higher θ), it substitutes for the skill-effort combination, reducing the marginal impact of skill on success probability.

Examples of functions satisfying this condition include:

- $p(se; \theta) = \sqrt{se + \theta}$
- $p(se; \theta) = 1 - e^{-(se+\theta)}$
- $p(se; \theta) = \frac{se+\theta}{1+se+\theta}$

Under this cross-partial condition, we obtain $\frac{\partial(V_0(1)-V_0(0))}{\partial s} < 0$, confirming that the value of tool adoption decreases with implementation skill.

Part 4: Early vs. late opportunities. The value difference is:

$$V_0(1) - V_0(0) = [M(e^*(1);1) - M(e^*(0);0)] \cdot \Gamma$$

Taking the derivative with respect to $\gamma(t)$:

$$\frac{\partial(V_0(1) - V_0(0))}{\partial\gamma(t)} = [M(e^*(1);1) - M(e^*(0);0)] \cdot \frac{\partial\Gamma}{\partial\gamma(t)}$$

For $\Gamma = \sum_{k=0}^{\infty} \left(\prod_{i=0}^{k} \gamma(i)\right) \delta^k$:

1. When $t = 0$:
$$\frac{\partial\Gamma}{\partial\gamma(0)} = \sum_{k=0}^{\infty} \delta^k \prod_{i=1}^{k} \gamma(i) = 1 + \sum_{k=1}^{\infty} \delta^k \prod_{i=1}^{k} \gamma(i)$$

2. When $t > 0$:
$$\frac{\partial\Gamma}{\partial\gamma(t)} = \sum_{k=t}^{\infty} \delta^k \left(\prod_{i=0, i\neq t}^{k} \gamma(i) \right)$$

For any $t < t'$, the ratio of marginal impacts is:

$$\frac{\frac{\partial(V_0(1)-V_0(0))}{\partial\gamma(t)}}{\frac{\partial(V_0(1)-V_0(0))}{\partial\gamma(t')}} = \frac{\frac{\partial\Gamma}{\partial\gamma(t)}}{\frac{\partial\Gamma}{\partial\gamma(t')}}$$

This ratio reflects how the discount factor δ and the structure of the opportunity sequence determine the relative importance of opportunities at different times. When $\gamma(t) = \gamma$ for $t \geq 1$ and $\delta\gamma < 1$, earlier opportunities ($t < t'$) have larger marginal impact due to discounting. For general opportunity sequences $\{\gamma(t)\}$, the relative importance depends on both discounting and the specific values of $\gamma(t)$.

A.3 Proof of Proposition 3

Given the functional forms $p(se;\theta) = \sqrt{se + \theta}$ and $c(e) = e$, the first-order condition for optimal effort is:

$$\alpha\Delta \cdot \frac{s}{2\sqrt{se^* + \theta}} = 1 \tag{57}$$

Solving this yields:

$$e^*(\theta) = \frac{\alpha^2\Delta^2 s}{4} - \frac{\theta}{s} \tag{58}$$

Recall that it is assumed that $\Delta > \frac{2}{s\sqrt{\alpha}}$ so that $e^* > 0$. The net benefit per improvement is:

$$M(e^*(\theta);\theta) = \alpha\Delta\sqrt{se^* + \theta} - e^* = \frac{\alpha^2\Delta^2 s}{4} + \frac{\theta}{s} \tag{59}$$

The continuation value is:

$$V(\theta) = \Gamma \cdot M(\theta) = \frac{\gamma_0}{1-\delta\gamma} \cdot \left(\frac{\alpha^2\Delta^2 s}{4} + \frac{\theta}{s}\right) \tag{60}$$

Part (a): Mean Effect. Since α, γ_0, γ, and s are independent:

$$E[V(\theta)] = E[\Gamma] \cdot E[M(\theta)] \tag{61}$$

$$= \mu_{\gamma_0} \cdot E\left[\frac{1}{1-\delta\gamma}\right] \cdot \left(\frac{\Delta^2}{4}(\mu_\alpha^2 + \sigma_\alpha^2)\mu_s + \theta E\left[\frac{1}{s}\right]\right) \tag{62}$$

The expectation $E\left[\frac{1}{1-\delta\gamma}\right]$ exists and is positive since $\delta\gamma < 1$. Taking the derivative:

$$\frac{\partial E[V(\theta)]}{\partial\theta} = \mu_{\gamma_0} \cdot E\left[\frac{1}{1-\delta\gamma}\right] \cdot E\left[\frac{1}{s}\right] > 0 \tag{63}$$

Part (b): Variance Threshold. Since Γ and $M(\theta)$ are independent:

$$\text{Var}(V(\theta)) = E[\Gamma^2]\text{Var}(M(\theta)) + \text{Var}(\Gamma)(E[M(\theta)])^2 \tag{64}$$

Computing the variance of $M(\theta)$:

$$E[M(\theta)] = \frac{\Delta^2}{4}(\mu_\alpha^2 + \sigma_\alpha^2)\mu_s + \theta E\left[\frac{1}{s}\right] \tag{65}$$

$$E[M(\theta)^2] = \frac{\Delta^4}{16}E[\alpha^4]E[s^2] + \frac{\Delta^2\theta}{2}E[\alpha^2] + \theta^2 E\left[\frac{1}{s^2}\right] \tag{66}$$

Therefore:

$$\text{Var}(M(\theta)) = E[M(\theta)^2] - (E[M(\theta)])^2 = A + B\theta + C\theta^2 \tag{67}$$

where:

$$A = \frac{\Delta^4}{16} E[\alpha^4] E[s^2] - \left(\frac{\Delta^2}{4} (\mu_\alpha^2 + \sigma_\alpha^2) \mu_s \right)^2 = \text{Var}\left(\frac{\alpha^2 \Delta^2 s}{4} \right) \geq 0 \tag{68}$$

$$B = \frac{\Delta^2}{2} (\mu_\alpha^2 + \sigma_\alpha^2) \left[1 - \mu_s E\left[\frac{1}{s} \right] \right] \tag{69}$$

$$C = E\left[\frac{1}{s^2} \right] - \left(E\left[\frac{1}{s} \right] \right)^2 = \text{Var}\left(\frac{1}{s} \right) > 0 \tag{70}$$

Taking the derivative of variance:

$$\frac{\partial \text{Var}(V(\theta))}{\partial \theta} = E[\Gamma^2](B + 2C\theta) + \text{Var}(\Gamma) \cdot 2E[M(\theta)] \cdot E\left[\frac{1}{s} \right] \tag{71}$$

At $\theta = 0$:

$$\left. \frac{\partial \text{Var}(V(\theta))}{\partial \theta} \right|_{\theta=0} = E[\Gamma^2] B + \text{Var}(\Gamma) \cdot \frac{\Delta^2}{2} (\mu_\alpha^2 + \sigma_\alpha^2) \mu_s E\left[\frac{1}{s} \right] \tag{72}$$

$$= \frac{\Delta^2}{2} (\mu_\alpha^2 + \sigma_\alpha^2) \left[E[\Gamma^2] - (E[\Gamma])^2 \mu_s E\left[\frac{1}{s} \right] \right] \tag{73}$$

This is negative when $E[\Gamma^2] < (E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right]$, which is equivalent to $\frac{E[\Gamma^2]}{(E[\Gamma])^2} < \mu_s E\left[\frac{1}{s}\right]$.

Therefore, we have $B < 0$ and thus:

$$\left. \frac{\partial \text{Var}(V(\theta))}{\partial \theta} \right|_{\theta=0} < 0 \tag{74}$$

Now, let's examine the exact form of the derivative:

$$\frac{\partial \text{Var}(V(\theta))}{\partial \theta} = E[\Gamma^2](B + 2C\theta) + \text{Var}(\Gamma) \cdot 2 \left(\frac{\Delta^2}{4} (\mu_\alpha^2 + \sigma_\alpha^2) \mu_s + \theta E\left[\frac{1}{s} \right] \right) \cdot E\left[\frac{1}{s} \right] \tag{75}$$

Expanding and collecting terms:

$$\frac{\partial \text{Var}(V(\theta))}{\partial \theta} = \left[E[\Gamma^2] B + \text{Var}(\Gamma) \cdot \frac{\Delta^2}{2} (\mu_\alpha^2 + \sigma_\alpha^2) \mu_s E\left[\frac{1}{s} \right] \right] \tag{76}$$

$$+ 2\theta \left[E[\Gamma^2] C + \text{Var}(\Gamma) \left(E\left[\frac{1}{s} \right] \right)^2 \right] \tag{77}$$

Since $C = \text{Var}(1/s)$ and using the fact that $E[\Gamma^2] = (E[\Gamma])^2 + \text{Var}(\Gamma)$:

$$E[\Gamma^2]C + \text{Var}(\Gamma)\left(E\left[\frac{1}{s}\right]\right)^2 = E[\Gamma^2] \cdot E\left[\frac{1}{s^2}\right] - E[\Gamma^2] \cdot \left(E\left[\frac{1}{s}\right]\right)^2 + \text{Var}(\Gamma) \cdot \left(E\left[\frac{1}{s}\right]\right)^2 \tag{78}$$

$$= E[\Gamma^2] \cdot E\left[\frac{1}{s^2}\right] - (E[\Gamma])^2 \cdot \left(E\left[\frac{1}{s}\right]\right)^2 \tag{79}$$

$$= \text{Var}\left(\frac{\Gamma}{s}\right) > 0 \tag{80}$$

Therefore, the derivative has the exact linear form:

$$\frac{\partial \text{Var}(V(\theta))}{\partial \theta} = a_0 + 2\theta \cdot \text{Var}\left(\frac{\Gamma}{s}\right) \tag{81}$$

where:

$$a_0 = E[\Gamma^2]B + \text{Var}(\Gamma) \cdot \frac{\Delta^2}{2}(\mu_\alpha^2 + \sigma_\alpha^2)\mu_s E\left[\frac{1}{s}\right] \tag{82}$$

$$= \frac{\Delta^2}{2}(\mu_\alpha^2 + \sigma_\alpha^2)\left[E[\Gamma^2] - (E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right]\right] \tag{83}$$

From part (b), when the condition $\frac{E[\Gamma^2]}{(E[\Gamma])^2} < \mu_s E\left[\frac{1}{s}\right]$ holds, we have $a_0 < 0$. Also, $\text{Var}(G/s) > 0$ always (unless G or s is constant).

Since this is a linear function with negative intercept and positive slope:

- There exists a unique $\theta^* = -\frac{a_0}{2 \cdot \text{Var}(\Gamma/s)} > 0$ where the derivative equals zero (θ^* is derived below).
- For $\theta < \theta^*$, the derivative is negative (variance decreases)
- For $\theta > \theta^*$, the derivative is positive (variance increases)

The second derivative confirms convexity:

$$\frac{\partial^2 \text{Var}(V(\theta))}{\partial \theta^2} = 2\text{Var}\left(\frac{\Gamma}{s}\right) > 0 \tag{84}$$

Part (c): Turning Point. **Derivation of θ^***. From the proof, we have shown that:

$$\frac{\partial \text{Var}(V(\theta))}{\partial \theta} = a_0 + 2\theta \cdot \text{Var}\left(\frac{\Gamma}{s}\right) \tag{85}$$

where:

$$a_0 = \frac{\Delta^2}{2}(\mu_\alpha^2 + \sigma_\alpha^2)\left[E[\Gamma^2] - (E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right]\right] \tag{86}$$

To find the minimum of the variance, we set the derivative equal to zero:

$$a_0 + 2\theta^* \cdot \mathrm{Var}\left(\frac{\Gamma}{s}\right) = 0 \tag{87}$$

Solving for θ^*:

$$2\theta^* \cdot \mathrm{Var}\left(\frac{\Gamma}{s}\right) = -a_0 \Leftrightarrow \theta^* = -\frac{a_0}{2 \cdot \mathrm{Var}\left(\frac{\Gamma}{s}\right)} \tag{88}$$

From part (b), when the condition $\frac{E[\Gamma^2]}{(E[\Gamma])^2} < \mu_s E\left[\frac{1}{s}\right]$ holds, we have:

$$E[\Gamma^2] - (E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right] < 0 \tag{89}$$

Therefore:

$$a_0 = \frac{\Delta^2}{2}(\mu_\alpha^2 + \sigma_\alpha^2) \times \underbrace{\left[E[\Gamma^2] - (E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right]\right]}_{<0} < 0 \tag{90}$$

Since:

- $a_0 < 0$ (negative)
- $\mathrm{Var}\left(\frac{\Gamma}{s}\right) > 0$ (positive, unless G or s is constant)

We have:

$$\theta^* = -\frac{a_0}{2 \cdot \mathrm{Var}\left(\frac{\Gamma}{s}\right)} = -\frac{(\text{negative})}{2 \times (\text{positive})} = \frac{(\text{positive})}{(\text{positive})} > 0 \tag{91}$$

Substituting the expression for a_0:

$$\theta^* = -\frac{\frac{\Delta^2}{2}(\mu_\alpha^2 + \sigma_\alpha^2)\left[E[\Gamma^2] - (E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right]\right]}{2 \cdot \mathrm{Var}\left(\frac{\Gamma}{s}\right)} \tag{92}$$

$$= \frac{\frac{\Delta^2}{2}(\mu_\alpha^2 + \sigma_\alpha^2)\left[(E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right] - E[\Gamma^2]\right]}{2 \cdot \mathrm{Var}\left(\frac{\Gamma}{s}\right)} \tag{93}$$

$$= \frac{\Delta^2(\mu_\alpha^2 + \sigma_\alpha^2)}{4} \cdot \frac{(E[\Gamma])^2 \mu_s E\left[\frac{1}{s}\right] - E[\Gamma^2]}{\mathrm{Var}\left(\frac{\Gamma}{s}\right)} \tag{94}$$

In addition, since $\text{Var}\left(\frac{\Gamma}{s}\right) = E[\Gamma^2]E\left[\frac{1}{s^2}\right] - (E[\Gamma])^2\left(E\left[\frac{1}{s}\right]\right)^2$, we can write:

$$\theta^* = \frac{\Delta^2(\mu_\alpha^2 + \sigma_\alpha^2)}{4} \cdot \frac{(E[\Gamma])^2\left[\mu_s E\left[\frac{1}{s}\right] - 1\right] - \text{Var}(\Gamma)}{E[\Gamma^2]E\left[\frac{1}{s^2}\right] - (E[\Gamma])^2\left(E\left[\frac{1}{s}\right]\right)^2} \tag{95}$$

A.4 Proof of Proposition 4

To determine when automation dominates human-tool collaboration, we compare their respective value functions.

The human-tool collaboration value is:

$$V_{\text{human}}(\theta) = \frac{\gamma}{1 - \delta\gamma} \cdot [p(se^*; \theta) \cdot \alpha \cdot \Delta - c(e^*; \theta)] \tag{96}$$

The automation value is:

$$V_{\text{auto}} = \frac{\hat{\gamma}}{1 - \delta\hat{\gamma}} \cdot [p(\hat{e}) \cdot \hat{\alpha} \cdot \Delta - c_a] \tag{97}$$

Substituting the flexibility constraints $\hat{\gamma} = \rho_\gamma \gamma$ and $\hat{\alpha} = \rho_\alpha \alpha$:

$$V_{\text{auto}} = \frac{\rho_\gamma \gamma}{1 - \delta\rho_\gamma\gamma} \cdot [p(\hat{e}) \cdot \rho_\alpha \alpha \cdot \Delta - c_a] \tag{98}$$

Let us define:

$$M_{\text{auto}} = p(\hat{e}) \cdot \rho_\alpha \alpha \cdot \Delta - c_a \tag{99}$$

$$M_{\text{human}} = p(se^*; \theta) \cdot \alpha \cdot \Delta - c(e^*; \theta) \tag{100}$$

For automation to dominate, we need $V_{\text{auto}} > V_{\text{human}}(\theta)$:

$$\frac{\rho_\gamma \gamma}{1 - \delta\rho_\gamma\gamma} \cdot M_{\text{auto}} > \frac{\gamma}{1 - \delta\gamma} \cdot M_{\text{human}} \tag{101}$$

Dividing both sides by γ and rearranging:

$$\frac{\rho_\gamma}{1 - \delta\rho_\gamma\gamma} \cdot M_{\text{auto}} > \frac{1}{1 - \delta\gamma} \cdot M_{\text{human}} \tag{102}$$

Cross-multiplying:

$$\rho_\gamma \cdot (1 - \delta\gamma) \cdot M_{\text{auto}} > (1 - \delta\rho_\gamma\gamma) \cdot M_{\text{human}} \tag{103}$$

Expanding the right side:

$$\rho_\gamma \cdot (1 - \delta\gamma) \cdot M_{\text{auto}} > M_{\text{human}} - \delta\rho_\gamma\gamma \cdot M_{\text{human}} \tag{104}$$

Rearranging:

$$\rho_\gamma \cdot (1 - \delta\gamma) \cdot M_{\text{auto}} + \delta\rho_\gamma\gamma \cdot M_{\text{human}} > M_{\text{human}} \tag{105}$$

Factoring out ρ_γ:

$$\rho_\gamma \cdot [(1 - \delta\gamma) \cdot M_{\text{auto}} + \delta\gamma \cdot M_{\text{human}}] > M_{\text{human}} \tag{106}$$

Define the dynamic adjustment factor:

$$\Phi(\rho_\gamma) = \frac{1 - \delta\gamma}{1 - \delta\rho_\gamma\gamma} \tag{107}$$

We can verify that our condition is equivalent to:

$$\rho_\gamma \cdot \Phi(\rho_\gamma) \cdot M_{\text{auto}} > M_{\text{human}} \tag{108}$$

To see that $\Phi(\rho_\gamma) \leq 1$, note that:

$$\Phi(\rho_\gamma) = \frac{1 - \delta\gamma}{1 - \delta\rho_\gamma\gamma} = \frac{1 - \delta\gamma}{1 - \delta\gamma + \delta\gamma(1 - \rho_\gamma)} \tag{109}$$

Since $\delta\gamma(1 - \rho_\gamma) \geq 0$, the denominator is at least as large as the numerator, so $\Phi(\rho_\gamma) \leq 1$, with equality only when $\rho_\gamma = 1$.

Substituting back the definitions of M_{auto} and M_{human} and dividing both sides by $M_{\text{human}} = p(se^*; \theta) \cdot \alpha \cdot \Delta - c(e^*; \theta)$ (assumed positive):

$$\rho_\gamma \cdot \Phi(\rho_\gamma) \cdot \frac{p(\hat{e}) \cdot \rho_\alpha\alpha \cdot \Delta - c_a}{p(se^*; \theta) \cdot \alpha \cdot \Delta - c(e^*; \theta)} > 1 \tag{110}$$

A.5 Proof of Proposition 5

We first establish the optimal effort levels under each control structure, then compare the resulting values.

Step 1: Optimal effort under PS control. When PS controls implementation, the net benefit per opportunity is:

$$M_{PS}(e_{PS}) = p_{PS}(e_{PS})\alpha_{PS}\Delta - c_{PS}(e_{PS}) - \kappa$$

The optimal effort e^*_{PS} maximizes $M_{PS}(e_{PS})$. Since κ is a fixed cost per opportunity, it does not affect the first-order condition:

$$p'_{PS}(e^*_{PS})\alpha_{PS}\Delta = c'_{PS}(e^*_{PS})$$

Step 2: Optimal effort under OS control. When OS controls implementation, the net benefit per opportunity is:

$$M_{OS}(e_{OS}) = p_{OS}(e_{OS})\alpha_{PS}\Delta - c_{OS}(e_{OS}) - p_{OS}(e_{OS})\kappa$$

This can be rewritten as:

$$M_{OS}(e_{OS}) = p_{OS}(e_{OS})(\alpha_{PS}\Delta - \kappa) - c_{OS}(e_{OS})$$

The optimal effort e^*_{OS} satisfies:

$$p'_{OS}(e^*_{OS})(\alpha_{PS}\Delta - \kappa) = c'_{OS}(e^*_{OS})$$

Step 3: Effort comparison with identical technologies. When both specialists have identical implementation technologies ($p_{OS}(\cdot) = p_{PS}(\cdot) = p(\cdot)$ and $c_{OS}(\cdot) = c_{PS}(\cdot) = c(\cdot)$), we can directly compare effort levels.

- Under PS control: $p'(e^*_{PS})\alpha_{PS}\Delta = c'(e^*_{PS})$
- Under OS control: $p'(e^*_{OS})(\alpha_{PS}\Delta - \kappa) = c'(e^*_{OS})$

Since $\kappa > 0$, we have $(\alpha_{PS}\Delta - \kappa) < \alpha_{PS}\Delta$. Given that $p'(e)$ is decreasing (by concavity of p) and $c'(e)$ is increasing (by convexity of c), the effort level that equates a smaller marginal benefit to marginal cost must be lower. Therefore, $e^*_{OS} < e^*_{PS}$.

Step 4: Value comparison. The expected values under each control structure are:

$$V_{PS} = \frac{\gamma_{OS}}{1 - \delta\gamma_{OS}} \cdot M_{PS}(e^*_{PS})$$

$$V_{OS} = \frac{\gamma_{OS}}{1 - \delta\gamma_{OS}} \cdot M_{OS}(e^*_{OS})$$

Since both share the factor $\frac{\gamma_{OS}}{1-\delta\gamma_{OS}} > 0$, OS should control when $M_{OS}(e^*_{OS}) > M_{PS}(e^*_{PS})$:

$$p_{OS}(e^*_{OS})\alpha_{PS}\Delta - c_{OS}(e^*_{OS}) - p_{OS}(e^*_{OS})\kappa > p_{PS}(e^*_{PS})\alpha_{PS}\Delta - c_{PS}(e^*_{PS}) - \kappa$$

Step 5: Decomposition of effects. Rearranging the inequality:

$$[p_{OS}(e^*_{OS}) - p_{PS}(e^*_{PS})]\alpha_{PS}\Delta - [c_{OS}(e^*_{OS}) - c_{PS}(e^*_{PS})] > \kappa[p_{OS}(e^*_{OS}) - 1] \tag{111}$$

This decomposition reveals three effects:

1. Direct implementation differences: $[p_{OS}(e^*_{OS}) - p_{PS}(e^*_{PS})]\alpha_{PS}\Delta - [c_{OS}(e^*_{OS}) - c_{PS}(e^*_{PS})]$
2. Communication cost savings: $-(1 - p_{OS}(e^*_{OS}))\kappa$ saved by PS control
3. Effort distortion: Embedded in $e^*_{OS} < e^*_{PS}$, which affects $p_{OS}(e^*_{OS})$ and $c_{OS}(e^*_{OS})$

References

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

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

_ **and** _ , "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

_ **, David Autor, Jonathan Hazell, and Pascual Restrepo**, "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, 2022, *40* (s1), S293–S340.

Agarwal, Ritu, Michelle Dugas, and Guodong (Gordon) Gao, "Augmenting physicians with artificial intelligence to transform healthcare: Challenges and opportunities," *Journal of Economics & Management Strategy*, March 2024, *33* (2), 360–374. RePEc handle: RePEc:bla:jemstr:v:33:y:2024:i:2:p:360-374.

Agrawal, Ajay, Joshua S Gans, and Avi Goldfarb, "Human Judgment and AI Pricing," *AEA Papers and Proceedings*, 2018, *108*, 58–63.

_ **,** _ **, and** _ , *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Press, 2018.

_ **,** _ **, and** _ , "Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence," in Ajay Agrawal, Joshua S Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, Chicago: University of Chicago Press, 2019, pp. 89–110.

_ **,** _ **, and** _ , *Power and Prediction: The Disruptive Economics of Artificial Intelligence*, Boston, MA: Harvard Business Review Press, 2022.

_ **,** _ **, and** _ , "Prediction Machines, Insurance, and Protection: An Alternative Perspective on AI's Role in Production," *Journal of the Japanese and International Economies*, 2024, *72.*

Autor, David H., "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," *Journal of Economic Perspectives*, September 2015, *29* (3), 3–30.

_ **and Neil Thompson**, "Expertise," NBER Working Paper Series Working Paper No. 33941, National Bureau of Economic Research June 2025.

_ **, Lawrence F. Katz, and Melissa S. Kearney**, "The Polarization of the U.S. Labor Market," *American Economic Review*, 2006, *96* (2), 189–194.

_ **,** _ **, and** _ , "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, 2008, *90* (2), 300–323.

Bessen, James, "Toil and Technology," *Finance & Development*, 2015, *52* (1), 16–19.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, February 2002, *117* (1), 339–376.

Brynjolfsson, Erik and Tom Mitchell, "What Can Machine Learning Do? Workforce Implications," *Science*, December 2017, *358* (6370), 1530–1534.

_ **, Danielle Li, and Lindsey Raymond**, "Generative AI at Work*," *The Quarterly Journal of Economics*, 02 2025, p. qjae044.

Crootof, Rebecca, Margot E. Kaminski, and W. Nicholson Price II, "Humans in the Loop," *Vanderbilt Law Review*, 2023, *76* (2), 429–510.

Csikszentmihalyi, Mihaly and Jacob W. Getzels, "Discovery-oriented Behavior and the Originality of Creative Products: A Study with Artists," *Journal of Personality and Social Psychology*, January 1971, *19* (1), 47–52.

Cui, Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz, "The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers," *SSRN Electronic Journal*, February 10 2025.

Czarnitzki, Dirk, Gastón P. Fernández, and Christian Rammer, "Artificial intelligence and firm-level productivity," *Journal of Economic Behavior Organization*, 2023, *211*, 188–205.

Dell'Acqua, Fabrizio, Charles Ayoubi, Hila Lifshitz, Raffaella Sadun, Ethan Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, and Karim R. Lakhani, "The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise," Technical Report 25-043, Harvard Business School Working Paper March 2025. Field experiment with 776 professionals at Procter & Gamble.

Dell'Acqua, Fabrizio, Saran Rajendran, Edward McFowland III, Lisa Krayer, Ethan Mollick, François Candelon, Hila Lifshitz-Assaf, Karim R. Lakhani, and Katherine C. Kellogg, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Technical Report, Harvard Business School Working Paper 24-013 2023.

Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, October 2017, *132* (4), 1593–1640.

Dranove, David, Chris Forman, Avi Goldfarb, and Shane M. Greenstein, "The Trillion Dollar Conundrum: Complementarities and Health Information Technology," *American Economic Journal: Economic Policy*, 2014, *6* (4), 239–270.

Einstein, Albert and Leopold Infeld, *The Evolution of Physics: The Growth of Ideas from Early Concepts to Relativity and Quanta*, Cambridge, UK: Cambridge University Press, 1938. Quote on p.92 (1966 Simon Schuster edition).

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308.

Feigenbaum, James J. and Daniel P. Gross, "Organizational and Economic Obstacles to Automation: A Cautionary Tale from AT&T in the Twentieth Century," *Management Science*, December 2024, *70* (12), 8520–8540.

Felten, Edward, Manav Raj, and Robert Seamans, "Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses," *Strategic Management Journal*, December 2021, *42* (12), 2195–2217.

Forman, Chris, Avi Goldfarb, and Shane Greenstein, "The Internet and Local Wages: A Puzzle," *American Economic Review*, 2012, *102* (1), 556–575.

Frey, Carl Benedikt and Michael A. Osborne, "The Future of Employment: How Susceptible Are Jobs to Computerisation?," *Technological Forecasting and Social Change*, January 2017, *114*, 254–280.

Gans, Joshua S, *The Microeconomics of Artificial Intelligence*, MIT Press, 2025.

Goldin, Claudia Dale and Lawrence F. Katz, *The Race Between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005*, Cambridge, MA: Belknap Press of Harvard University Press, 2008.

Handa, Kunal, Alexander Tamkin, Michael McCain, Shengjia Huang, Emine Yilmaz Durmus, Spencer Heck, Jared Mueller, Jason Hong, Spencer Ritchie, Thomas Belonax, Kristina K. Troy, Dario Amodei, Jared Kaplan, Jack Clark, and Deep Ganguli, "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," arXiv preprint arXiv:2503.04761 February 2025. Anthropic Working Paper.

Hoffmann, Manuel, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu, "Generative AI and the Nature of Work," Technical Report, Harvard Business School Working Paper 25-021 2024.

Hui, Xiang, Oren Reshef, and Luofeng Zhou, "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market," *Organization Science*, 2024, *35* (6), 1977–1989.

Humlum, Anders and Emilie Vestergaard, "Large Language Models, Small Labor Market Effects," Technical Report 33777, National Bureau of Economic Research May 2025.

_ **and** _ , "The unequal adoption of ChatGPT exacerbates existing inequalities among workers," *Proceedings of the National Academy of Sciences*, 2025, *122* (1), e2414972121.

Jorgenson, Dale W. and Kevin J. Stiroh, "Raising the Speed Limit: U.S. Economic Growth in the Information Age," *Brookings Papers on Economic Activity*, 2000, *31* (1), 125–236.

_ **, Mun S. Ho, and Kevin J. Stiroh**, "Growth of US Industries and Investments in Information Technology and Higher Education," in "Measuring Capital in the New Economy" NBER Chapters, National Bureau of Economic Research, Inc, 2005, pp. 403–478.

Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe, "AI, Skill, and Productivity: The Case of Taxi Drivers," Working Paper 30612, National Bureau of Economic Research October 2022.

Law, Kelvin K. F. and Michael Shen, "How Does Artificial Intelligence Shape Audit Firms?," *Management Science*, 2025, *71* (5), 3641–3666.

Loaiza, Isabella and Roberto Rigobon, "The EPOCH of AI: Human-Machine Complementarities at Work," 2024. MIT Sloan School of Management, November 2024. Forthcoming.

McCullough, Jeffrey S., Stephen T. Parente, and Robert Town, "Health information technology and patient outcomes: the role of information and labor coordination," *The RAND Journal of Economics*, 2016, *47* (1), 207–236.

McElheran, Kristina, Mu-Jeung Yang, Zachary Kroff, and Erik Brynjolfsson, "The Rise of Industrial AI in America: Microfoundations of the Productivity J-curve(s)," Working Paper CES-WP-25-27, Center for Economic Studies, U.S. Census Bureau, Washington, D.C. April 2025.

Mumford, Michael D., Roni Reiter-Palmon, and Matthew R. Redmond, "Problem Construction and Cognition: Applying Problem Representations in Ill-Defined Domains," in Mark A. Runco, ed., *Problem Finding, Problem Solving, and Creativity*, Norwood, NJ: Ablex, 1994, pp. 3–39.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

Oliner, Steven D., Daniel E. Sichel, and Kevin Stiroh, "Explaining a Productive Decade," *Brookings Papers on Economic Activity*, 2007, (1), 81–152.

Otis, Nicholas G., Rowan Philip Clarke, Solene Delecourt, David Holtz, and Rembrand Koning, "The Uneven Impact of Generative AI on Entrepreneurial Performance," OSF Preprints hdjpk, Center for Open Science December 2023.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer, "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," 2023.

Reiter-Palmon, Roni and Erika J. Robinson, "Problem Identification and Construction: What Do We Know, What Is the Future?," *Psychology of Aesthetics, Creativity, and the Arts*, 2009, *3* (1), 43–47.

Roldán-Monés, Toni, "When GenAI Increases Inequality: Evidence from a University Debating Competition," Working Paper, ESADE – Esade Centre for Economic Policy September 2024. Updated version published Apr 2025.

Scharre, Paul, "Centaur Warfighting: The False Choice of Humans vs. Automation," *Temple International & Comparative Law Journal*, 2016, *30* (1), 151–165.

Teodoridis, Florenta, "Understanding Team Knowledge Production: The Interrelated Roles of Technology and Expertise," *Management Science*, 2018, *64* (8), 3625–3648.

Tranchero, Matteo, "Finding diamonds in the rough: Data-driven opportunities and pharmaceutical innovation," December 2024. Job Market Paper.

Wiles, Emma, Zanele T Munyikwa, and John J Horton, "Algorithmic Writing Assistance on Jobseekers' Resumes Increases Hires," Working Paper 30886, National Bureau of Economic Research January 2023.

Yiu, Shun, Rob Seamans, Manav Raj, and Teng Liu, "Strategic Responses to Technological Change: Evidence from Online Labor Markets," March 2025. Working paper.