14.662 spring 2026, Lecture slides 10 — Some (Labor) Economics of Artificial Intelligence

David Autor
MIT and NBER

March 11, 2026 (rev 2026/03/11)

The Jacquard loom of 1801

The first commercial symbolic processor

- Accesses, analyzes, and acts upon information
- Follows rules – carries out codifiable, 'routine' tasks, specified in programs

Jacquard loom of 1801

Seven trillion-fold decline in cost of computing between 1940 and 2006

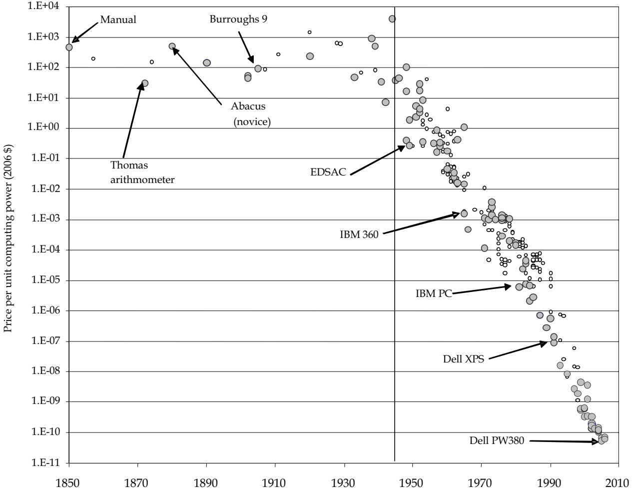


FIGURE 3
THE PROGRESS OF COMPUTING MEASURED IN COST PER COMPUTATION PER SECOND DEFLATED BY THE PRICE INDEX FOR GDP IN 2006 PRICES

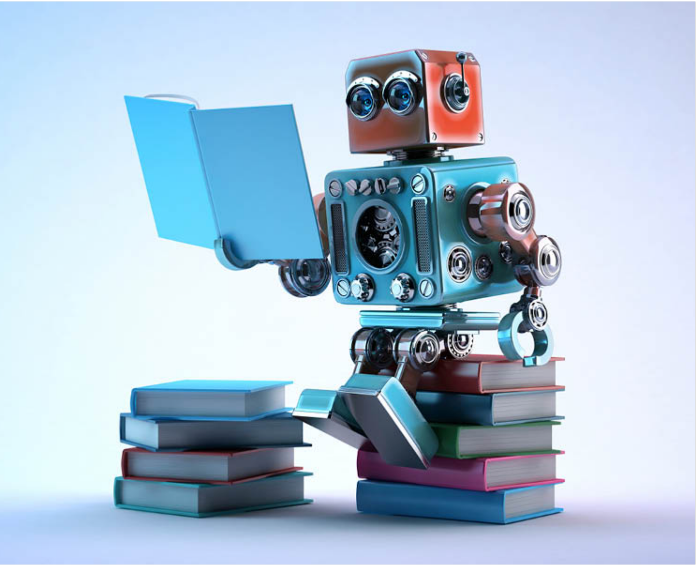
Nordaus 2007

"We know more than we can tell"
The Tacit Dimension, 1966

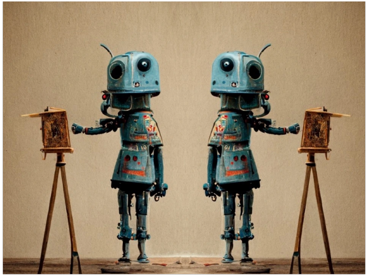
- Explicit knowledge – fully specified rules and procedures
- Tacit knowledge – not explicitly specified, often learned through immersion, practice

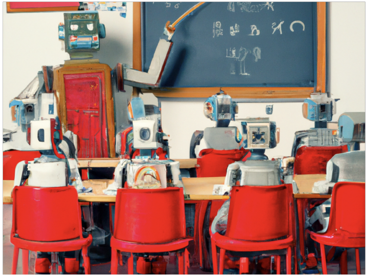
Michael Polanyi, 1891 - 1976

Autonomous learning



Creative tasks





Discovery and invention

Agenda

1. **The big questions**
2. **For what activities is AI a substitute for, or complement to, labor**
   - We won't be missed
   - Two views: Bottlenecks vs. feedback loops
3. **Is it sufficient to think of AI as automating or instantiating tasks?**
   - Automation in the task model
   - New task creation in the task model
   - Expertise leveling in the task model
   - Increasing the substitutability of indivisible factors (assignment model)
4. **Organizational design for AI**
   - Knowledge hierarchies view
   - Broader views on management, work design, and organizational structure
5. **Using AI to support learning (expertise acquisition) and performance (expertise deployment)**
6. **State of evidence on employment effects**
7. **Automation, collaboration, and expertise**

The big questions

1. For what activities is AI a substitute for, or complement to, labor?
2. Is it sufficient to think of AI as automating or instantiating tasks, or is a richer notion of jobs needed?
3. Expertise, automation, and collaboration: A potentially useful distinction?
4. Is organizational design and adoption of AI different from the worker/task-level frame?
5. How can AI be designed to support expertise acquisition (learning) and deployment (decision-making)?
6. What is the evidence on AI and job displacement?

For what activities is AI a substitute for, or complement to, labor?

- What *types* of tasks can be newly automated?
- What *types* of tasks are bottlenecks to automation?
  - NB: bottleneck $\rightarrow$ labor demand
- Is it correct to equate bottlenecks with rising scarcity?
- Will new tasks be created, how many, how quickly, demanding what skills?
- What affects the rate and locus of new task creation?

Tasks versus jobs: do we need a richer framework?

Is it sufficient to think of AI as automating or instantiating tasks, or is a richer notion of jobs needed?

- In atomistic task models with one type of labor, this distinction has no bite
- Distinction matters more when comparative advantage differs by skill/demographic group
- Matters *even more* when tasks are bundled in jobs, so the same task has a different role across occupations
- Matters again—and differently—in assignment models, where technology may change the distribution of b, i.e., the indivisible factors that workers are assigned to

Automation versus collaboration: A potentially useful distinction?

- **Automation:** Expertise removed, labor may remain (e.g., ride-hailing)
- **Collaboration:** Supporting tasks removed, expertise remains (e.g., accounting)
- The world is a bit more complicated than automation versus non-automation
- 'Automation exposure' is not a sufficient statistic for either

Does AI adoption require organizational change?

Is organizational design and adoption of AI different from the worker/task-level frame?

- We know almost nothing about this so far: research and theory are at the worker level
- But history suggests that organizations must change to adopt new technologies ('J-curve')

How can AI be used to support expertise acquisition (learning) and deployment (decision-making)?

- Can humans gain judgment without the sweat equity of grunt work, which AI increasingly does for them?
- Concern about learning ladders/job ladders: the experience Catch-22
- How should AIs interact with people to support both learning and decision-making — and is there a tradeoff?

What is the evidence so far on AI and job displacement?

- Not surprisingly, this has received a lot of attention
- Perhaps also no surprise, current conclusions are controversial

Agenda

1. **The big questions**
2. **For what activities is AI a substitute for, or complement to, labor**
   - We won't be missed
   - Two views: Bottlenecks vs. feedback loops
3. **Is it sufficient to think of AI as automating or instantiating tasks?**
   - Automation in the task model
   - New task creation in the task model
   - Expertise leveling in the task model
   - Increasing the substitutability of indivisible factors (assignment model)
4. **Organizational design for AI**
   - Knowledge hierarchies view
   - Broader views on management, work design, and organizational structure
5. **Using AI to support learning (expertise acquisition) and performance (expertise deployment)**
6. **State of evidence on employment effects**
7. **Automation, collaboration, and expertise**

Transformational Artificial Intelligence 'TAI' as phase transition

- **Before TAI**

$$Y = Q_C^{1-\sum \alpha_g} \prod_g l_g^{\alpha^g}$$

  - Q_C is the quantity of "compute"
  - Compute is growing exponentially over time (e.g., Moore's Law)
  - α_g is the share of 'bottlenecked' tasks performed by group g workers
  - *Why are bottlenecks important?*
  - All labor demand is **"derived demand."**

- **Imagine that we decided not to go to work today. Would we be missed?**
  - Marginal product of labor is $\partial Y / \partial l_g = \alpha_g y / l_g$
  - Labor's marginal product grows at same rate as output (bottleneck effect)
  - Workers retain importance: they would be missed if they "quietly quit"

Transformational Artificial Intelligence 'TAI' as phase transition

- **Transformative AI phase transition**

$$Y = A \cdot Q_C + \sum_g A \cdot \kappa_G \cdot l_g$$

  - **All possible tasks** can be completed by automated systems at computational cost κ
  - That is, $\alpha_g \to 0$

- **Would we be missed?**
  - Marginal product of labor is now $A \cdot \kappa$
  - Value marginal product of labor is equal to the computational resources that labor saves "the system"
  - This MP is constant, even though economy grows exponentially as it adds computational capacity

- **Good news: humans no longer a bottleneck**
  - Merely an additive term in a well functioning system
  - There is plenty of work
  - Bad news: it doesn't matter whether or not humans do this work

- **It machines became a perfect substitute for human labor, would profoundly alter the value of labor, dist'n of income**

OpenAI Charter

OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity.

OpenAI

A future without labor scarcity: Wall-E vs. Mad Max – Fury Road

WALL-E

Mad Max Fury Road

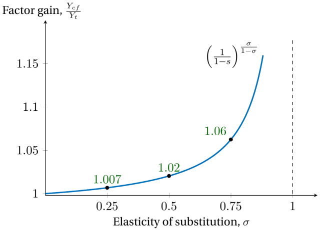
Jones & Tonetti (2025): What if AI fully automates some tasks with infinite productivity?

- CES aggregation across tasks with $\sigma < 1 \Rightarrow$ weak links dominate
- Output gain from infinitely automating a factor with cost share s:

$$\frac{Y_{cf}}{Y_t} = \left(\frac{1}{1-s}\right)^{\frac{\sigma}{1-\sigma}}$$

- Software $\approx 2\%$ of GDP. Even with **infinite** AI productivity in software, GDP rises by just $\sim$2% when $\sigma = 1/2$
- Non-automated tasks become the binding constraint

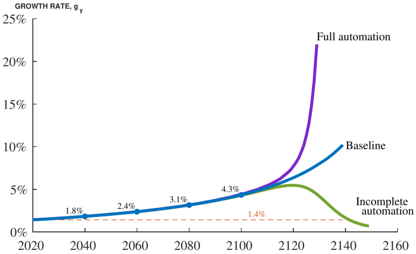
Figure 6: Automating Software, $s = 2\%$



Jones & Tonetti: Three scenarios for economic growth

Jones & Tonetti (2025), Figure 10

Figure 10: Simulating the Future: Economic Growth



Note: If the capital share reaches 100% in finite time, then growth explodes. If the capital share falls to zero, then growth falls to $\hat{\psi}_\ell = 0.5\%$. Surprisingly, even with a stable capital share, growth explodes in the Baseline case.

Three cases depending on whether some tasks **permanently** require labor:

- **Full automation:** Growth explodes—$Y_t = B_t K_t$
- **Incomplete automation:** Growth falls back to $\dot{\hat{\psi}}_\ell = 0.5\%$—the rate at which people get better. $Y_t = A_t L_t$
- **Baseline:** Automation rate rises over time $\Rightarrow$ growth eventually explodes, but slowly

Punchline: Under incomplete automation, labor share $\to 1$

When automating AI research produces explosive growth

Davidson, Halperin, Houlden & Korinek (2025): AI feedback loops in software & hardware

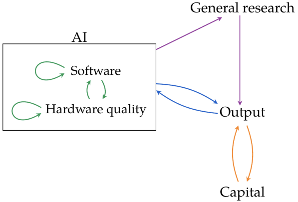


Figure 1: Model of AI automation with software & hardware feedback loops

Hyperbolic growth arises when:

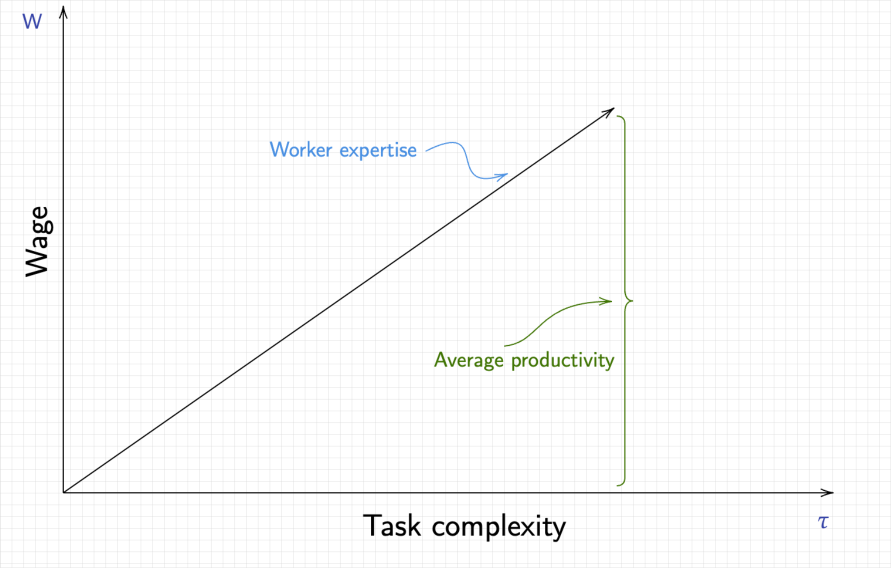
$$f_Y + f_S \left(\frac{1}{\beta_S} \right) + f_H \left(\frac{1}{\beta_H} \right) + \frac{f_A}{\alpha} \left(\frac{1}{\beta_A} \right) > 1$$

- f_i = fraction of tasks automated in sector i
- β_i = diminishing returns to research in sector i
- Hardware has **low** diminishing returns ($\beta_H = 0.2$) $\Rightarrow$ strong feedback
- Key assumption: Cobb-Douglas aggregation across tasks within sectors ($\sigma = 1$) — eliminates bottlenecks by construction

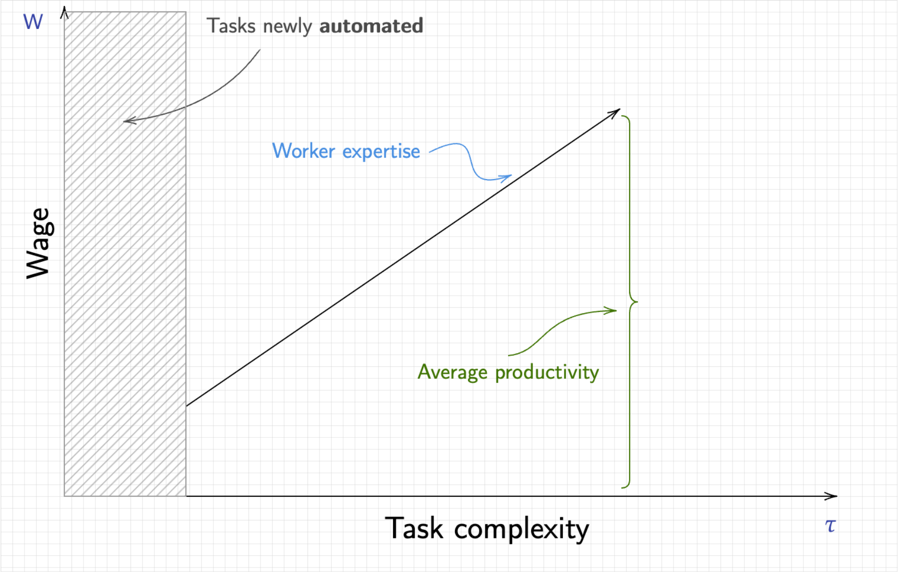
Agenda

1. **The big questions**
2. **For what activities is AI a substitute for, or complement to, labor**
   - We won't be missed
   - Two views: Bottlenecks vs. feedback loops
3. **Is it sufficient to think of AI as automating or instantiating tasks?**
   - Automation in the task model
   - New task creation in the task model
   - Expertise leveling in the task model
   - Increasing the substitutability of indivisible factors (assignment model)
4. **Organizational design for AI**
   - Knowledge hierarchies view
   - Broader views on management, work design, and organizational structure
5. **Using AI to support learning (expertise acquisition) and performance (expertise deployment)**
6. **State of evidence on employment effects**
7. **Automation, collaboration, and expertise**

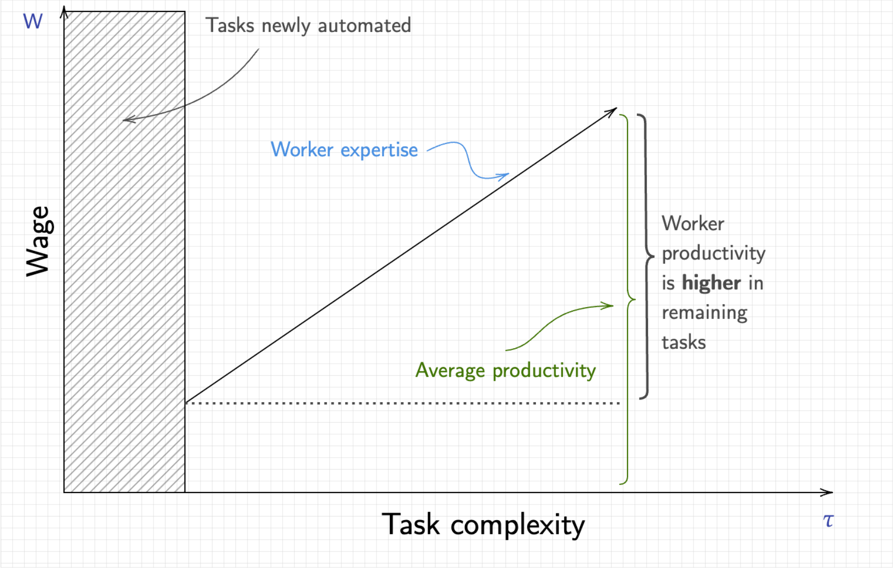
Baseline: Single index task model



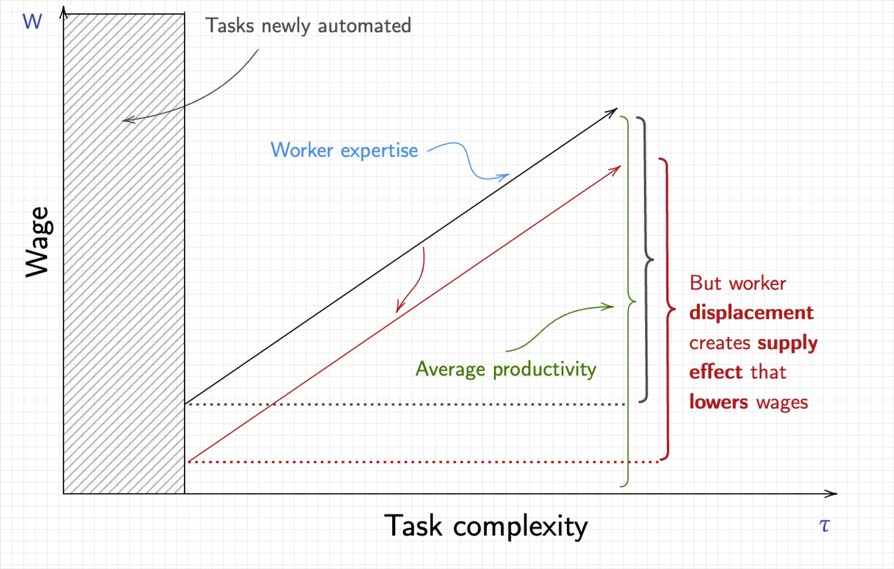
Automation in the task model



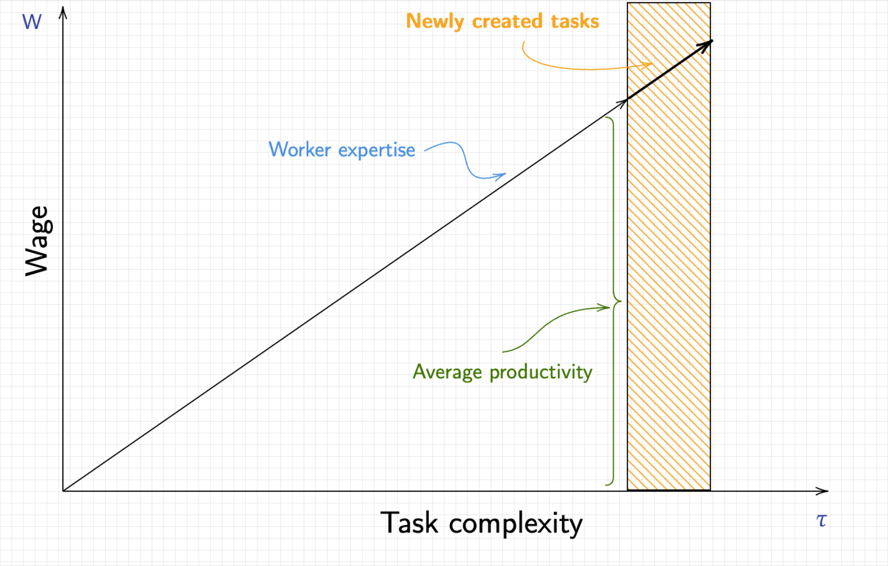
Automation in the task model



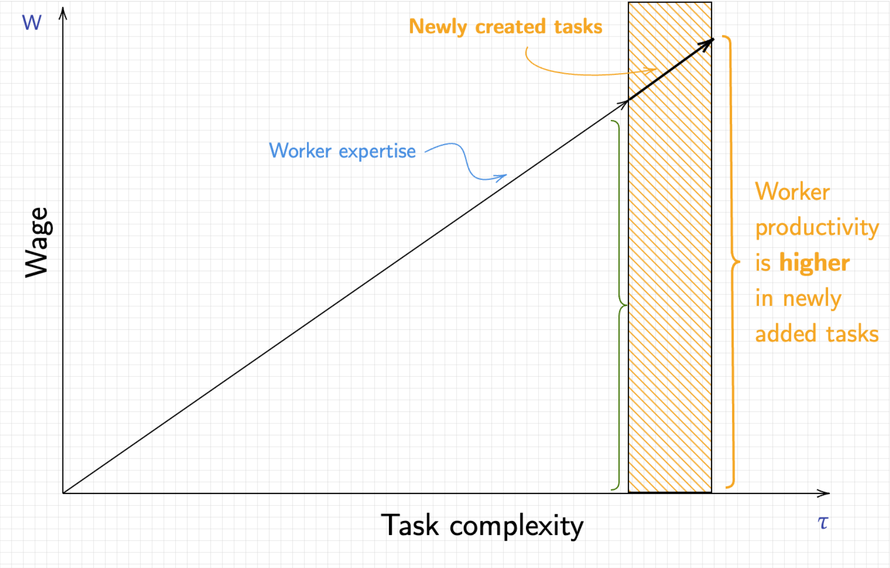
Automation in the task model



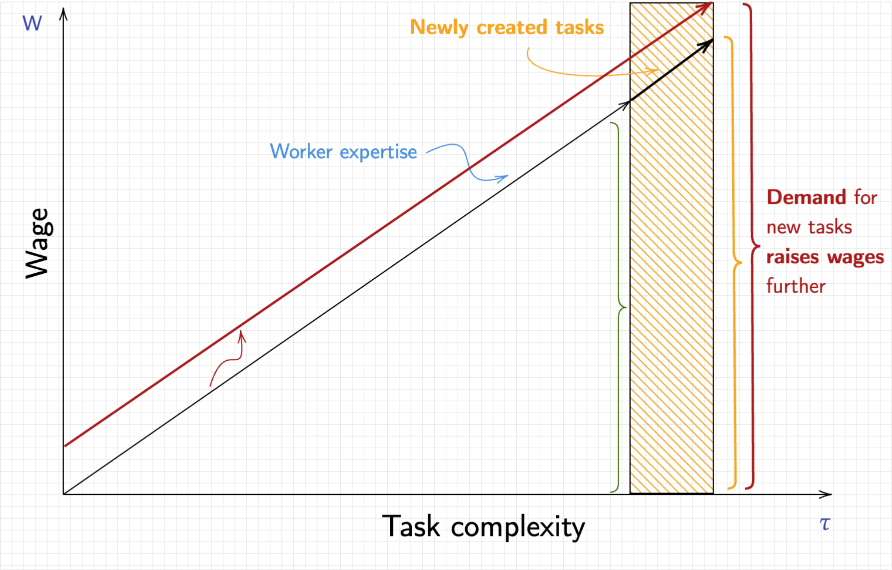
New task creation in the task model



New task creation in the task model



New task creation in the task model

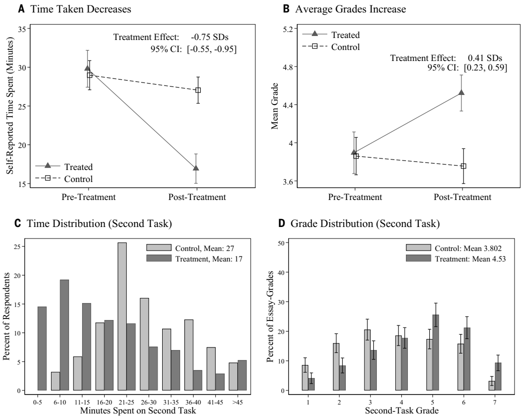


‘Leveling up’

Expertise leveling in the task model

Leveling up — AI raises productivity and compresses inequality in writing tasks

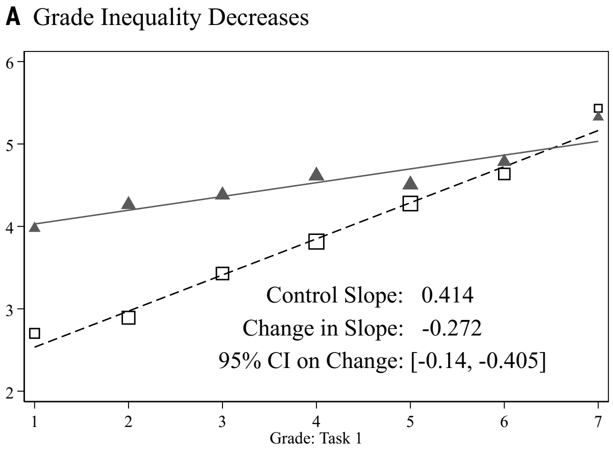
Noy & Zhang, *Science* 2023: 453 college-educated professionals; randomized access to ChatGPT for occupation-specific writing tasks



- **Time taken drops 40%** (27 → 17 min)
- **Output quality rises 18%** (grades up 0.45 SD)
- Effects are **not limited** to specific pockets of the time or grade distributions—entire distributions shift

Leveling up — ChatGPT disproportionately helps initially lower-performing workers

Noy & Zhang, *Science* 2023, Figure 2

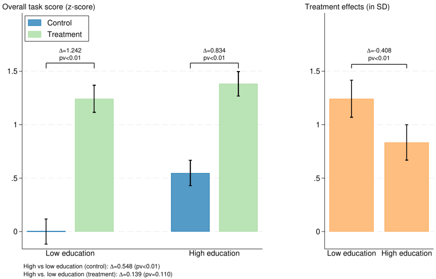


- **Grade inequality more than halved:**
  - Control: ρ(task 1, task 2 grades) $= 0.41$
  - Treatment: $\rho = 0.14$ ($p < 0.001$ for Δ in slopes)
- **Workers with low initial grades** gain 1–2 grade points and save 10 min
- **Workers with high initial grades** maintain their level while also saving $\sim$10 min

Leveling up — AI narrows education-based productivity gaps by $\approx$75%

Cruces, Fernández Meijide, Galiani, Gálvez & Lombardi, NBER WP 34851, 2026: N=1,174 adults (ages 25–45, Argentina); randomized AI access for incentivized business problem-solving task

Figure 2: Effect AI assistance on overall task score by education level



Notes: The left panel shows the average score on the task (average across 10 grading iterations), standardized relative to the low-education control group, along with 95% confidence intervals, separately for high- and low-education participants in the treatment and control groups. Above each education group, we report the difference between the treatment and control groups within that education category. The right panel presents the estimated treatment effects for high- and low-education participants. Above, we report the difference between the treatment effects of high- and low-education participants.

29

- **Without AI:** High-ed outperforms low-ed by 0.548 SD
- **With AI:** Gap falls to 0.139 SD (not significant, $p = 0.110$)
- **Treatment effect:** +1.24 SD for low-ed vs +0.83 SD for high-ed
- **Difference is significant** ($\Delta = -0.41$ SD, $p < 0.01$)
- Equalizing effects in both content and writing quality

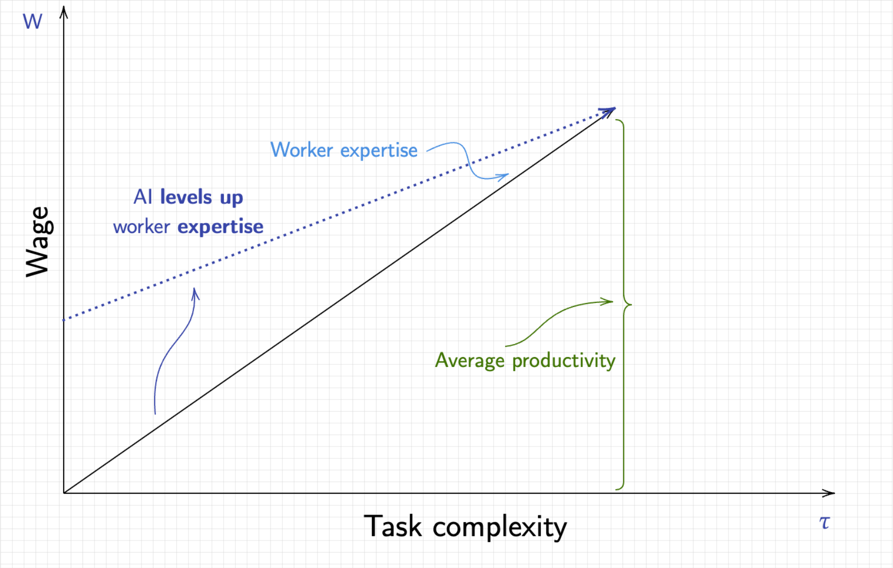
Leveling up — How low- and high-education workers use AI differently

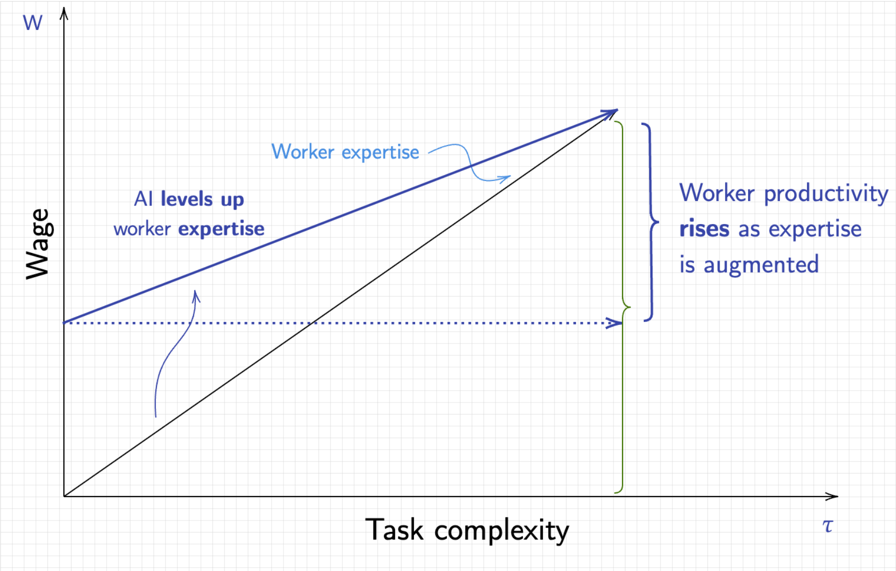
Cruces et al. '26, Table 2: AI use by education level among 471 treated participants who used the assistant

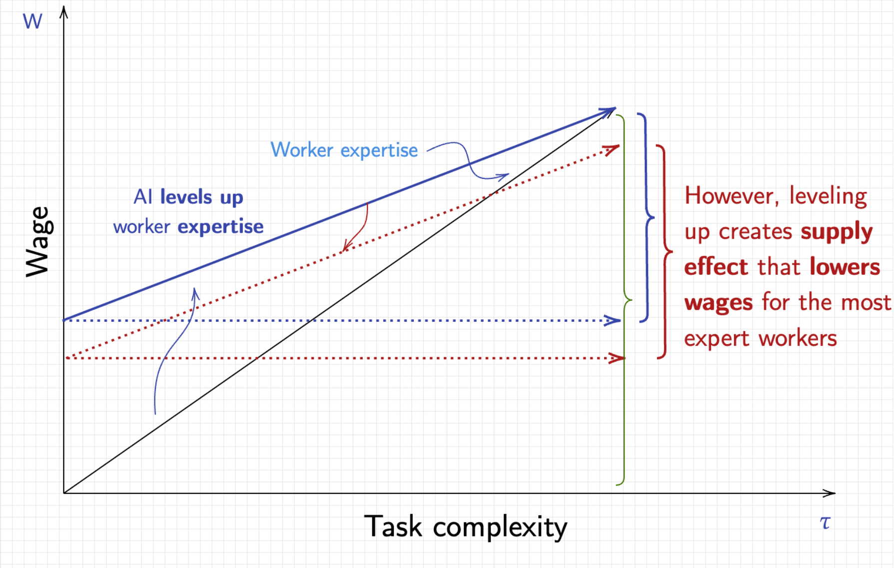
Table 2: AI use by education level

	Low education	High education	Difference
Number and length of messages sent			
Number of messages sent to the AI assistant	2.813	2.971	0.158
Total characters sent to the AI assistant	794.172	847.396	53.224
Assistance with working through the task			
Share of task components with any AI assistance	0.608	0.615	0.007
Share of task components with AI assistance – generate answer	0.595	0.601	0.006
Share of task components with AI assistance – validate own response	0.026	0.026	-0.001
Detailed reasoning instructions (standardized index)	0.000	0.183	0.183*
Provided contextual information (standardized index)	0.000	0.033	0.033
Use of the AI-generated output			
Final output full copy/paste of AI-generated content	0.702	0.601	-0.101**
Final output partial copy/paste of AI-generated content	0.242	0.308	0.065
Final output paraphrases AI-generated content	0.045	0.073	0.028
Final output does not rely on AI-generated content	0.010	0.018	0.008
Workflow and orchestration of the conversation			
High initial specificity	0.288	0.355	0.067
Iterative engagement	0.465	0.473	0.008
Structured workflow	0.141	0.216	0.075**
Observations	198	273	

- **Intensity of use is similar:** ∼3 messages, ∼800 characters, ∼61% of task components
- **Differences are qualitative:**
  - High-ed: more **content-specific** instructions (+0.145 SD) and **disciplined prompting** (+0.18 SD)
  - High-ed: more **structured workflow** (21.6% vs 14.1%**)
  - Low-ed: more likely to **fully copy-paste** AI output (70% vs 60%**)
  - High-ed: more likely to **partially incorporate** AI text with own writing
- Residual gap driven by *how* tool is used, not *how much*


W
Wage
Worker expertise
AI **levels up**
worker **expertise**
Average productivity
Task complexity
τ


W
Wage
Worker expertise
AI **levels up**
worker **expertise**
Worker productivity
rises as expertise
is augmented
Task complexity
τ


W
Wage
Worker expertise
AI **levels up**
worker **expertise**
However, leveling up creates **supply effect** that **lowers wages** for the most expert workers
Task complexity
τ

‘Leveling up’: AI is a tool that is more complementary to M workers than to either L or H workers

- Recall that all factors are perfect substitutes in task production
- All tasks i can be performed by low, medium or high skill workers

$$y(i) = A_L \alpha_L(i)\, l(i) + A_M \alpha_M(i)\, m(i) + A_H \alpha_H(i)\, h(i) + A_K \alpha_K(i)\, k(i)$$

- **But comparative advantage differs among skill groups** $\{\alpha_L(i), \alpha_M(i), \alpha_H(i)\}$
  - Comparative advantage of H relative to M, and M relative to L, monotonically rising in task index
  - $\alpha_L(i)/\alpha_M(i)$ *and* $\alpha_M(i)/\alpha_H(i)$ are continuously differentiable and strictly decreasing
- **Defining ‘leveling up’**
  - $\alpha'_M(i) > \alpha_M(i)$ for all $i < 1$, and $\alpha'_M(1) = \alpha_M(1)$. Hence:

$$\begin{aligned} \alpha_L(i)/\alpha'_M(i) &< \alpha_L(i)/\alpha_M(i) \quad \text{for } i < 1 \\ \alpha'_M(i)/\alpha_H(i) &> \alpha_M(i)/\alpha_H(i) \quad \text{for } i < 1 \end{aligned}$$

What are the implications of 'leveling up' in the three-skill task model?

1. Aggregate productivity
2. Task shares $\{I_L, I_H - I_L, 1 - I_H\}$
3. Wages $\{W_L, WM, W_H\}$
4. Wage inequality $\frac{W_H}{W_M}, \frac{W_M}{W_L}, \frac{W_H}{W_L}$
5. How does 'leveling up' differ from a rise in A_M?
6. What is the likelihood that the 'supply' effect overwhelms the 'productivity' effect for each group L, M, H?

Return to wage-setting in the three-skills task model

- **Recall:** Three conditions — (1) law of one price; (2) equal factor shares; (3) no arbitrage — imply that relative wages are solely a function of labor supplies and task thresholds

$$w_J = w_J \left[I_H, I_L | H, M, L, A_H, A_M, A_L, \alpha_H(\cdot), \alpha_M(\cdot), \alpha_L(\cdot)\right] \text{ for } J \in [H, M, L]:$$

$$\frac{w_H}{w_M} = \left(\frac{1 - I_H}{I_H - I_L}\right)\left(\frac{H}{M}\right)^{-1},$$

$$\frac{w_M}{w_L} = \left(\frac{I_H - I_L}{I_L}\right)\left(\frac{M}{L}\right)^{-1}$$

- So, labor supplies L, M, H plus compare adv. $\alpha(L)$, $\alpha(M)$, $\alpha(L)$ determine task allocation, I_L and I_H, and hence wages

Recall equilibrium wages, factor shares, TFP in this model

- Output is

$$Y = B \times \underbrace{L^{I_L} M^{I_H - I_L} H^{1-I_H}}_{\text{Cobb-Douglas labor aggregate}}$$

where

$$B = \underbrace{\exp\left(\int_0^{I_L} \ln A_l \alpha(i)\, d_i + \int_{I_L}^{I_H} \ln A_m \alpha(i)\, d_i + \int_{I_H}^{1} \ln A_h \alpha(i)\, d_i\right)}_{\text{TFP}}$$

- Wages equal

$$\begin{aligned} W_L = \frac{\partial Y}{\partial L} &= B \times I_L \times L^{I_L - 1} M^{I_H - I_L} H^{1-I_H} \\ &= \frac{I_L}{L} \times \left(B L^{I_L} M^{I_H - I_L} H^{1-I_H}\right) \\ &= I_L \times \frac{Y}{L} \end{aligned}$$

and similarly, $W_M = Y(I_H - I_L)/M$ and $W_H = Y(1 - I_H)/H$.

Recall **subtle** effects of changes in supply of M or productivity of A_M

What happens when either M or A_M rises?

$$\frac{d\ln\left(w_H/w_L\right)}{d\ln M} \gtrless 0 \quad \frac{d\ln\left(w_H/w_L\right)}{d\ln A_M} \gtrless 0$$

The answer depends critically on this term:

$$\beta_H\left(I\right) \equiv \ln\alpha_M\left(I\right) - \ln\alpha_H\left(I\right), \beta_L\left(I\right) \equiv \ln\alpha_L\left(I\right) - \ln\alpha_M\left(I\right)$$

- β are comp. advantage of L versus H workers in M tasks
- $\beta_L'\left(I_L\right) I_L = \partial\beta_L/\partial I_L$ and $\beta_H'\left(I_H\right)\left(1 - I_H\right) = \partial\beta_H/\partial I_H$
- If $\beta_L'\left(I_L\right)$ low relative to $\beta_H'\left(I_H\right)$, high skill workers have *strong comparative advantage* for tasks above I_H

Hence, rise in M displaces L workers more than H iff

$$\frac{d\ln\left(w_H/w_L\right)}{d\ln M} > 0 \text{ iff } \left|\beta_L'\left(I_L\right) I_L\right| < \left|\beta_H'\left(I_H\right)\left(1 - I_H\right)\right|$$

This occurs because I_L falls more than I_H rises

Instead of rise in A_M, consider the consequences of 'leveling up'

'Leveling up'

- $\alpha'_M(i) > \alpha_M(i)$ for all $i < 1$, and $\alpha'_M(1) = \alpha_M(1)$

$$\begin{aligned} \alpha_L(i)/\alpha'_M(i) &< \alpha_L(i)/\alpha_M(i) \quad \text{for } i < 1 \\ \alpha'_M(i)/\alpha_H(i) &> \alpha_M(i)/\alpha_H(i) \quad \text{for } i < 1 \end{aligned}$$

'Leveling up:' $\alpha_M(\cdot)$ schedule gets steeper. How does this differ from ΔA_M?

$$\frac{d \ln (w_H/w_L)}{d \ln A_M} \gtrless 0$$

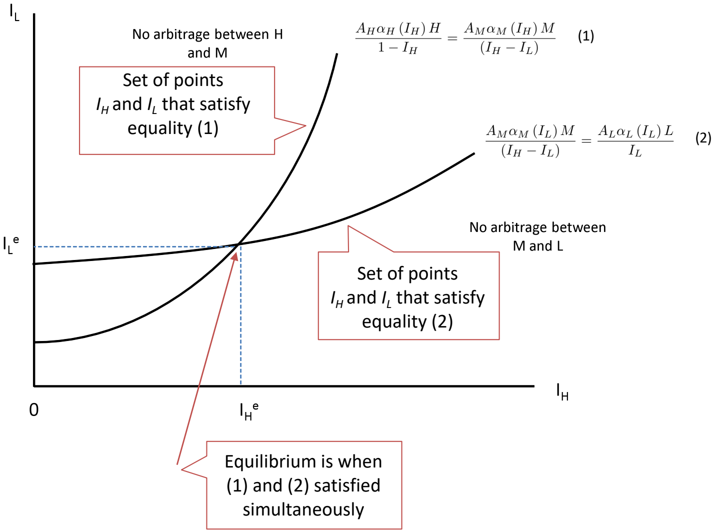
As before, answer depends on this term

- Original expression $\beta_H(I) \equiv \ln \alpha_M(I) - \ln \alpha_H(I)$, $\beta_L(I) \equiv \ln \alpha_L(I) - \ln \alpha_M(I)$
- New expression $\tilde{\beta}_H(I) \equiv \ln \alpha'_M(I) - \ln \alpha_H(I)$, $\tilde{\beta}_L(I) \equiv \ln \alpha_L(I) - \ln \alpha'_M(I)$
- Notice that $\tilde{\beta}_H(I) \geq \beta_H(I)$, $\tilde{\beta}_L(I) \leq \beta_L(I)$

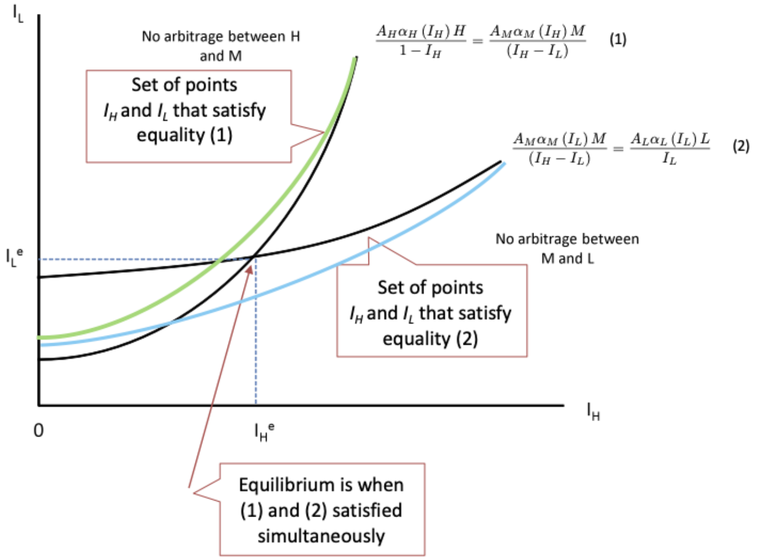
This shift increases comparative advantage of M relative to L workers – but also increases comparative disadvantage of M rel to H

- Because the 'leveling up' upward shift is greater for lower than higher-index tasks, it's going to create more competition for L than H workers.
- Leveling up is will unambiguously expand the interval $[I_L, I_H]$ but mostly at the expense of L workers. It's possible (even likely) that I_H will fall (though by as much as I_L)
- This is *less advantageous* for M workers than an analogous shift in the slope of the comparative advantage schedule that increased M productivity relatively more at higher than lower-index tasks.
- Why don't M workers benefit more? Because leveling up does not make M workers more substitutable for H workers.
- This logic differs in the assignment model (next), where all competition is downward not upward.

Revisiting that no arbitrage figure

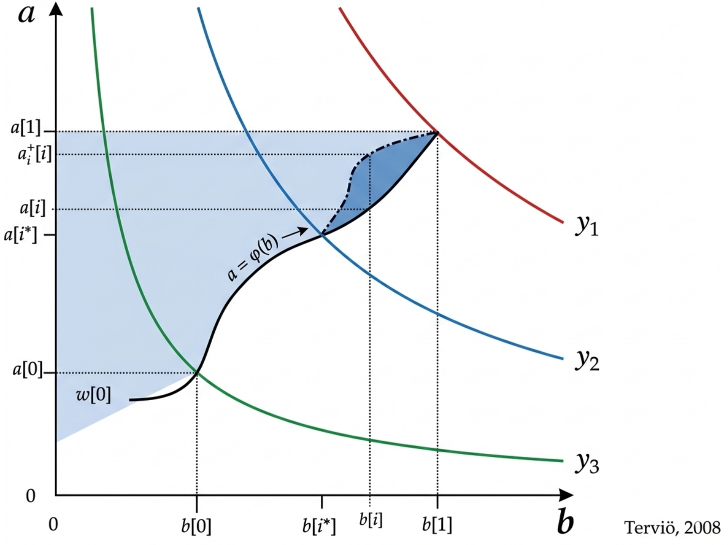


Revisiting that no arbitrage figure (note: does not show new eq'm)



‘Leveling up’

Expertise leveling in the assignment model


a
a[1]
$a_i^+[i]$
a[i]
a[i*]
a[0]
w[0]
a = φ(b)
y1
y2
y3
0
0
b[0]
b[i*]
b[i]
b[1]
b
Terviö, 2008

Expertise leveling in the assignment model

1. Which workers benefit – which are made worse off?
2. Are lower-index workers harmed if highest index workers are (also) augmented?
3. In what ways is this setting qualitatively different from the task setting?
   - Task model: leveling up creates more downward than upward competition (if M levels up)
   - Superstars model: leveling up creates upward competition
4. Difference comes from indivisibiliies
   - No task reassignment
   - Reduces the scarcity value of higher index workers by making lower index workers closer substitutes
   - Output is rising, though this is not Pareto-improving
5. Assignment model captures idea for which people have strong intuition but lack formal reasoning (imho)

Agenda

1. **The big questions**
2. **For what activities is AI a substitute for, or complement to, labor**
   - We won't be missed
   - Two views: Bottlenecks vs. feedback loops
3. **Is it sufficient to think of AI as automating or instantiating tasks?**
   - Automation in the task model
   - New task creation in the task model
   - Expertise leveling in the task model
   - Increasing the substitutability of indivisible factors (assignment model)
4. **Organizational design for AI**
   - Knowledge hierarchies view
   - Broader views on management, work design, and organizational structure
5. **Using AI to support learning (expertise acquisition) and performance (expertise deployment)**
6. **State of evidence on employment effects**
7. **Automation, collaboration, and expertise**

Motivation: AI can automate noncodifiable work, overcoming Polanyi's paradox (1966)

- Earlier automation technologies (robots, enterprise software) automate codifiable tasks
- AI learns from examples, acquiring tacit knowledge — patterns that cannot be reduced to explicit rules
- This allows AI to perform cognitive, noncodifiable work: drafting contracts, diagnosing problems, advising on strategy
- Key question: How does this reshape the organization of knowledge work and the distribution of labor income?

Ide, Enrique and Eduard Talamàs. "Artificial Intelligence in the Knowledge Economy." *Journal of Political Economy*, 133(12), December 2025.

Pre-AI economy: Humans form hierarchical organizations

- Individuals are heterogeneous in knowledge $z \in [0, 1]$
- Production requires solving problems of varying difficulty; knowledge is tacit — solvers cannot prescribe a plan of action in advance
- **Workers** (less knowledgeable): pursue production opportunities, handle routine problems
- **Solvers** (more knowledgeable): handle exceptions that workers cannot solve
- Communication is costly: each request for help consumes h units of the solver's time
- Equilibrium features occupational stratification ($W \preceq I \preceq S$) and positive assortative matching: more knowledgeable workers matched with more knowledgeable solvers

AI as a technology that converts computational resources into "AI agents"

- AI agents have a fixed knowledge level z_{AI} and require 1 unit of compute each
- Unlike humans, AI is scalable: once knowledge is encoded, it can be deployed across all available compute simultaneously
- Compute is abundant relative to human time $\Rightarrow$ the binding constraint in human-AI interaction is human time, not compute

Two modes of AI deployment:

1. **Autonomous AI**: agents operate as both coworkers (pursuing production) and copilots (advising) — can work independently
2. **Nonautonomous AI**: agents operate solely as copilots — can only provide advice, cannot pursue production independently

Five possible firm configurations after AI

1. **Single-layer human**: independent producer
2. **Single-layer automated**: AI as independent producer
3. **Two-layer human**: human solver + human workers
4. **Bottom-automated**: human solver + AI workers
5. **Top-automated**: AI solver + human workers

A firm never uses AI in *both* layers — an AI solver already knows the solutions its AI workers would need

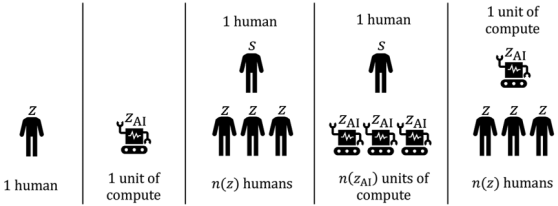


Fig. 2.—Five possible firm configurations after AI.

Autonomous AI primarily benefits the most knowledgeable individuals

- The most knowledgeable humans shift to supporting AI-driven production (bottom-automated firms), leveraging their expertise at low cost across many AI workers
- Less knowledgeable humans face competition from AI in production work and must rely on lower-quality human solvers
- The least knowledgeable individuals are harmed: they face direct competition from AI coworkers and lose access to the best human solvers

But: overall output is higher with autonomous AI

- AI agents can work independently, expanding the economy's productive capacity
- The most knowledgeable humans can apply their scarce expertise at unprecedented scale

Impact of autonomous AI on wages: Basic vs. advanced AI

Ide & Talamàs, Fig. 5: Bold = post-AI wage $w^*(z)$; gray = pre-AI wage $w(z)$; dotted = 45° line

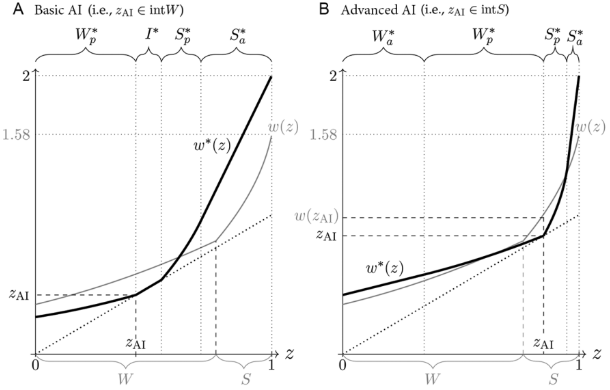


FIG. 5.—Comparison of pre- and post-AI equilibria. Distribution of knowledge: $G(z) = z$.

Panel A — Basic AI ($z_{AI} \in$ int W):
Top earners gain; lowest earners harmed

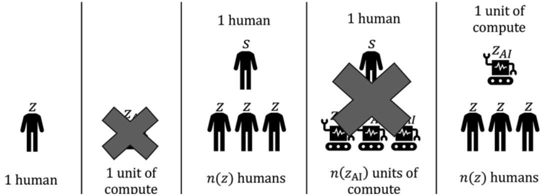
Panel B — Advanced AI ($z_{AI} \in$ int S):
Both extremes gain; those near z_{AI} harmed

Nonautonomous AI primarily benefits the least knowledgeable

- AI acts only as a copilot — cannot compete with humans for production work
- Least knowledgeable gain access to cheap AI advisors (top-automated firms)
- Most knowledgeable are less well served: AI competes for advisory roles but can't handle routine work for them

But: overall output is lower

- AI copilots cannot independently pursue production opportunities



Ide & Talamàs, Fig. 8: Nonautonomous AI eliminates independent AI production and bottom-automated firms

The autonomy trade-off and reconciling the evidence

Paper posits/conjectures a trade-off: output vs. equality

- **Autonomous AI**: higher total output, but benefits concentrated among the most knowledgeable — increases inequality
- **Nonautonomous AI**: lower total output, but benefits concentrated among the least knowledgeable — reduces inequality

Paper argues that it reconciles contradictory empirical evidence:

- Evidence that AI levels up less-skilled workers (Dell'Acqua et al. 2023; Noy & Zhang 2023; Brynjolfsson, Li & Raymond 2025) $\Rightarrow$ consistent with *nonautonomous* AI (copilots)
- Evidence that AI complements high-skilled workers and substitutes for low-skilled (Alekseeva et al. 2024; Berger et al. 2024) $\Rightarrow$ consistent with *autonomous* AI (agents)

Their takeway: Regulating AI autonomy involves choosing on this frontier

My takeway: Formally elegant, profoundly reductive / impoverished, relevance unclear

Broader views on management, work design, and organizational structure

- **Managing workers and agents**
  - Papers by Weidmann and Deming, as discussed in earlier lectures
  - Dessain and Santos, "Adaptive Organizations" *JPE* 2006
- **Transmission of knowledge across experience groups (e.g., experts to novices)**
  - See Ide (2025), "Automation, AI, and the Intergenerational Transmission of Knowledge"
- **Far too little research in this domain**

Agenda

Does AI use support or erode expertise acquisition?

- AI decision-support tools are proliferating in professional settings: medicine, law, finance, engineering. . .
- Competing hypotheses about how AI exposure affects unaided human performance:
  - **Contextual learning:** AI teaches users to recognize patterns they would otherwise miss $\Rightarrow$ performance improves even without AI
  - **Overeliance / de-skilling:** users offload cognition to AI, reducing attention and effort $\Rightarrow$ unaided performance deteriorates
- Which effect dominate—and under what circumstances?

Budzyń et al. (2025, Lancet): Multicenter observational study, 4 endoscopy centers in Poland

Setting

- Colonoscopy: key tool for detecting precancerous polyps (adenomas)
- Adenoma detection rate (ADR) = fraction of colonoscopies finding ≥ 1 adenoma — the standard performance metric
- AI polyp-detection tools boost ADR by 5–20 pp in randomized trials

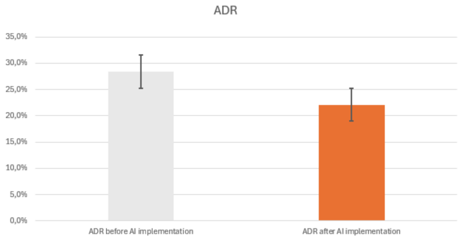
Design

- Four centers implemented AI tools for polyp detection, end of 2021
- Colonoscopies randomly conducted with or without AI by date
- Compare ADR of standard (non-AI) colonoscopies in 3 months *before* vs. 3 months *after* AI implementation
- 1,443 non-AI-assisted procedures (795 before, 648 after); 19 experienced endoscopists

Result: Ademonma Detection Rate (ADR) falls after AI usage—perhaps due to deskilling

Budzyń et al., Fig. 1: ADR of standard colonoscopy before and after AI implementation (95% CI)

Figure 1. Change in adenoma detection rate of standard, non-AI assisted colonoscopy before and after implementation of artificial intelligence for polyp detection. Each plot represents a 95% confidence interval.



Main finding

- ADR of non-AI colonoscopy fell from **28.4%** to **22.4%** (-6 pp, $p = 0.009$)
- Effect persists after controlling for patient age, sex, operator specialty

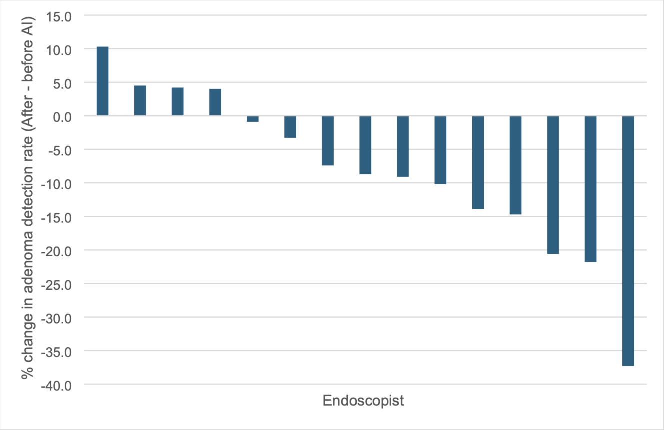
Flaming caveat

- No external ground truth for adenoma prevalence — ADR is itself a detection measure.
- Before/after comparison; no control group of centers without AI
- Patient observables are similar across periods, but this is not that powerful

Fall in ADR is 'pervasive' across endoscopists

Budzyń et al., Fig. 2: Change in ADR by endoscopist after AI implementation

Figure 2: Endoscopist-level absolute percentage change of adenoma detection rate of standard, non-artificial intelligence (AI) assisted colonoscopy after implementation of AI in colonoscopy at each center. Four endoscopists who did less than 10 colonoscopies for either before or after AI implementation are removed from the presentation.



Implications for AI deployment in expert work?

- AI as copilot may generate dependency: when the tool is unavailable, performance is worse than before the tool existed
  - Cognitive offloading
  - Poor skill retention
  - Over-reliance
- Where this might show up
  - Aviation (autopilot and pilot de-skilling)
  - Medical practice
  - Creativity collapse: "Generative AI enhances individual creativity but reduces the collective diversity of novel content" (Doshi, Hauser *Science Advances* 2024)
- Design question: building AI tools that support both learning and performance
- Connects to Ide & Talamàs: distinction between autonomous and nonautonomous AI matters not statically (performance), but dynamically (human capital accumulation)

Can AI explanations foster learning instead of dependency?

Buçinca, Swaroop, Paluch, Doshi-Velez & Gajos (2024): "Contrastive Explanations That Anticipate Human Misconceptions Can Improve Human Decision-Making Skills"

- Most AI systems offer unilateral explanations: justify the AI's choice by listing all supporting features
  - "The AI suggests Exercise A *because of*. . ." (one-sided justification)
- Humans may benefit from contrastive explanations: "Why P rather than Q?"
  - "Why Exercise A *instead of* Exercise B?" (addresses the user's knowledge gap)
- Contrastive explanations require a model of the human — predicting what the user would have chosen, and explaining only where AI and human reasoning *diverge*
- That's a pretty challenging problem

Unilateral vs. contrastive explanations

Buçinca et al., Fig. 1: (a) Unilateral explanation lists all features supporting AI's choice; (b) Contrastive explanation highlights where AI and human reasoning diverge

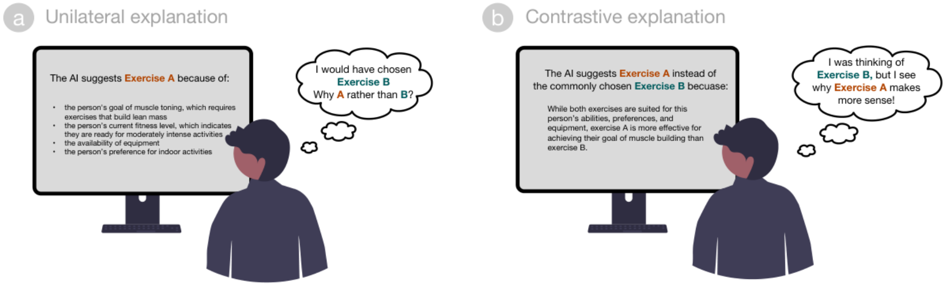
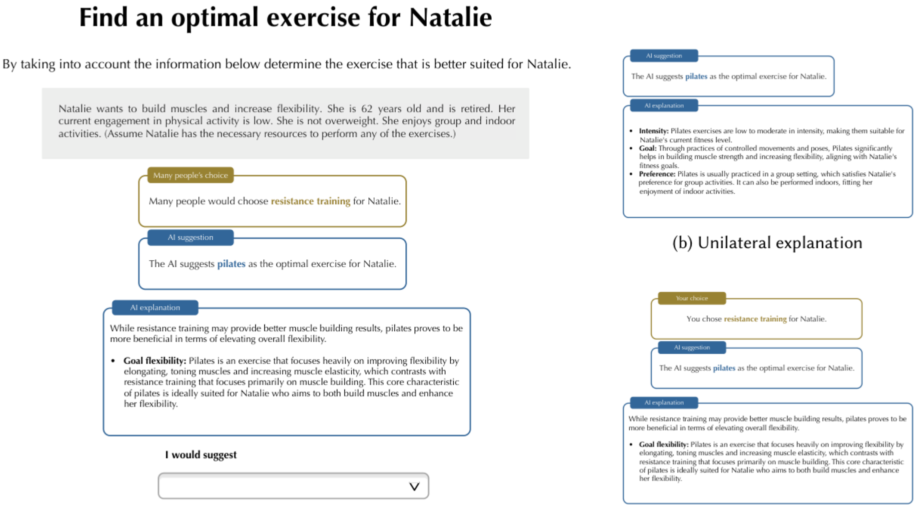


Fig. 1. A simplified illustration of (a) unilateral explanations, which list all the features contributing to the AI's decision, and (b) contrastive explanations, which highlight the differences between the AI's choice and a likely human response for an exercise recommendation task.

- **Unilateral**: "AI suggests A because..." $\Rightarrow$ user passively receives justification
- **Contrastive**: "AI suggests A *instead of* B because..." $\Rightarrow$ user engages with the discrepancy between their reasoning and the AI's

Buçinca et al., Fig. 3: Participants choose optimal exercises for fictional characters; AI provides recommendations with different explanation types



(a) Sample task with contrastive explanation

(b) Unilateral explanation

(c) Contrastive explanation after

Experimental design: Five conditions ($N = 628$)

Five treatment conditions

1. **No AI**: Participants decide without any AI support (baseline)
2. **Unilateral**: AI recommendation + explanation listing all supporting features
3. **Contrastive (predicted foil)**: AI recommendation + explanation contrasting with the choice a *human model* predicts the user would make
4. **Contrastive (random foil)**: Same format, but the alternative is randomly chosen (not predicted)
5. **Contrastive after**: Contrastive explanation shown *after* user makes initial decision

Three outcome measures

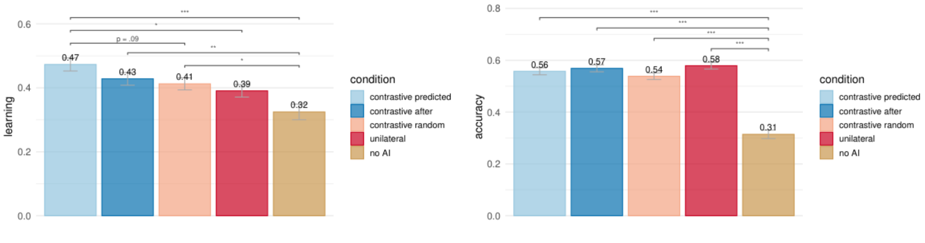
1. Learning: post-test accuracy (controlling for pre-test) — do skills *transfer* to unaided decisions?
2. Accuracy: correctness during AI-assisted phase
3. Overreliance: following AI when AI is wrong

Unstated postulate: Humans have value to add if engaged effectively

1. This is certainly not true in all cases
2. And it may be a transitory phenomenon in many applications (e.g., driving, flight, diagnosis)

Main result: Contrastive explanations improve human learning

Buçinca et al., Fig. 4: Marginal means of learning (post-test performance, controlled for pre-test) and accuracy across conditions. Error bars = 1 SE.



(a) Learning (b) Accuracy

Fig. 4. Main results. Marginal means of human learning (post-intervention performance, controlled for pre-intervention performance) and accuracy accross different conditions. Error bars represent one standard error. Significance levels after Holm-Bonferroni correction are presented only for significant (or marginally significant) differences, indicated by: * p < 0.05, ** p < 0.01, *** p < 0.001.

- **Learning** (Panel a): contrastive explanations with predicted foil significantly improve post-test performance vs. no AI ($d = 0.65$, $p < 0.001$) and vs. unilateral ($d = 0.35$, $p = 0.02$)
- **Accuracy** (Panel b): no significant accuracy loss — contrastive explanations do not sacrifice decision quality for learning gains
- Unilateral explanations do not improve learning over the no-AI baseline

Unstated postulates and tradeoffs

Unstated postulate

- Humans have value to add if engaged effectively
- Certainly not true in all cases
- Could be a transitory phenomenon in many applications (e.g., driving, flight, diagnosis)

Potential tradeoff: Is there a performance cost to keeping people in the loop?

- Tradeoff 1: People do slightly worse than autonomous AI, but their ongoing training is needed for backup
- Tradeoff 2: Machines will make mistakes without human oversight, but perhaps that's a cost worth paying

Agenda

Headcount by age group across all occupations (Brynjolfsson, Chandar, and Chen ’25)

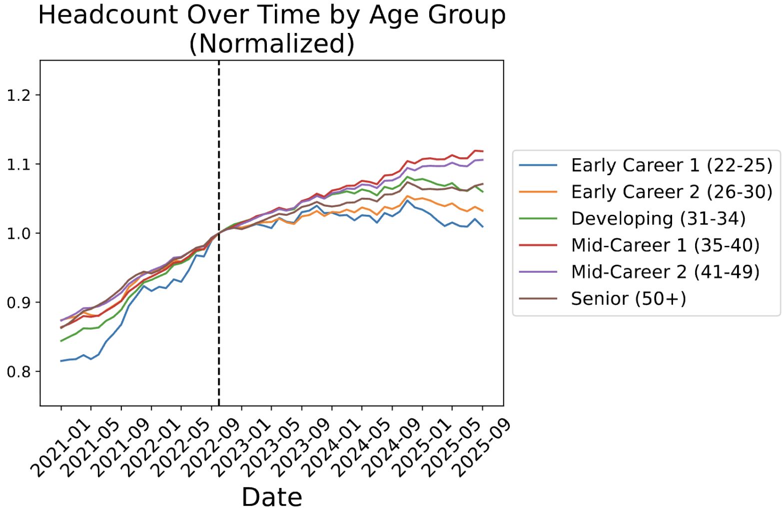
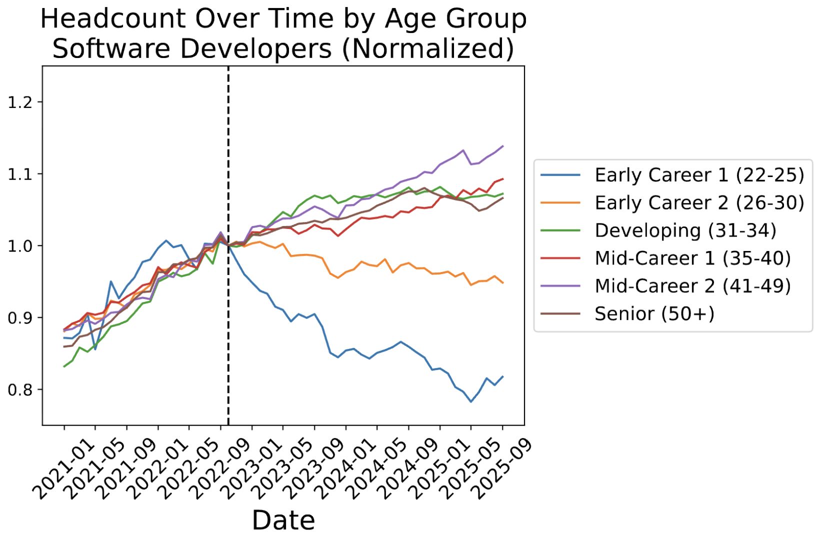
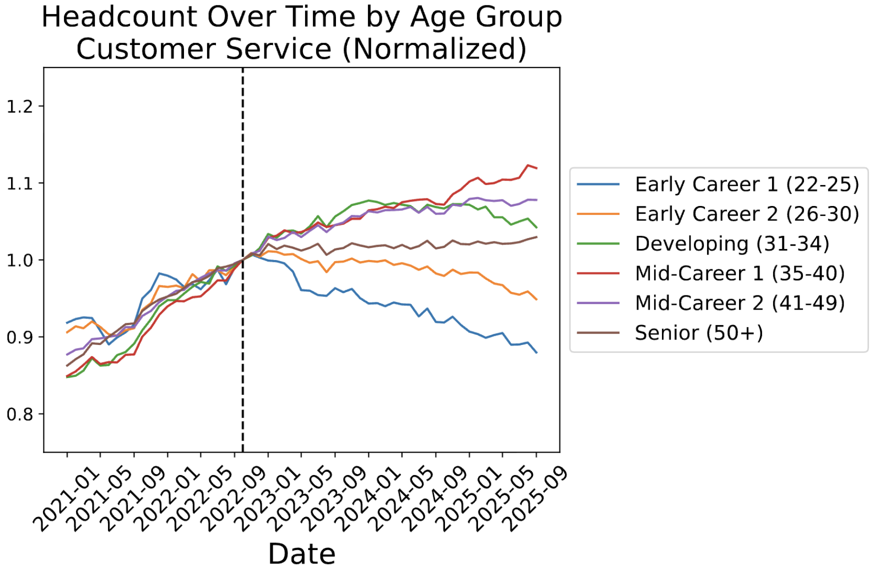


Figure A4: Employment changes by age. Including all occupations.

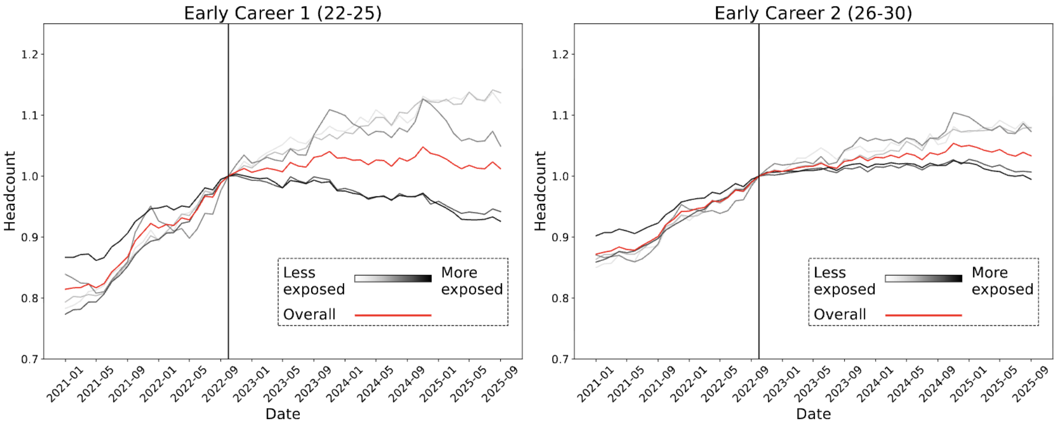
Software developers: Headcount by age group (Brynjolfsson, Chandar, and Chen ’25)



Customer service: Headcount by age group (Brynjolfsson, Chandar, and Chen '25)



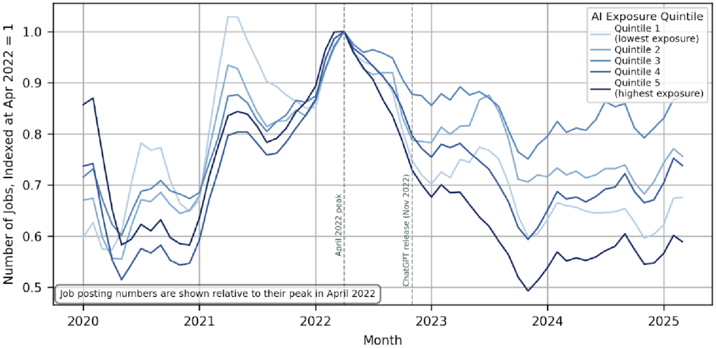
Early-career headcount by AI exposure (Brynjolfsson, Chandar, and Chen ’25)



Job postings by AI exposure quintile, Jan 2022–Mar 2025 (Iscenko and Millet ’26)

Chart 1: The downturn in hiring of AI-exposed occupations starts six months before the release of ChatGPT

Evolution of Job Postings by AI Exposure Quintile, Jan 2022 - Mar 2025

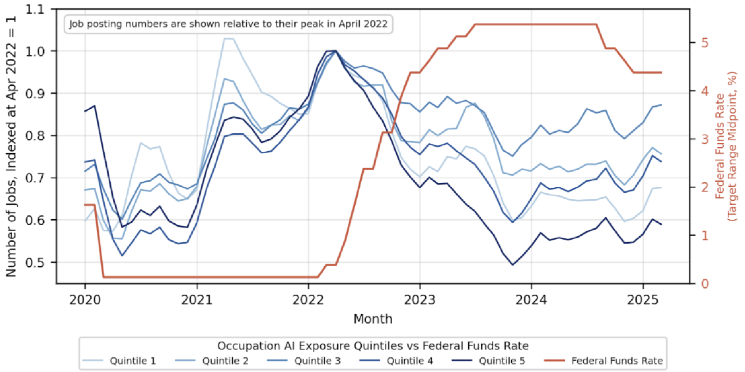


Sources: Lightcast, Eloundou et al. (2024)

This chart illustrates that job postings for the most AI-exposed quintile (Q5, dark blue line) peaked in Spring 2022 and began a steep decline well before the November 2022 launch of ChatGPT, indicating a pre-existing shock.

Job postings by AI exposure vs. federal funds rate (Iscenko and Millet '26)

Chart 3: Federal Funds Rate vs. Job Postings by AI Exposure Quintile (Jan 2022 - Oct 2023)

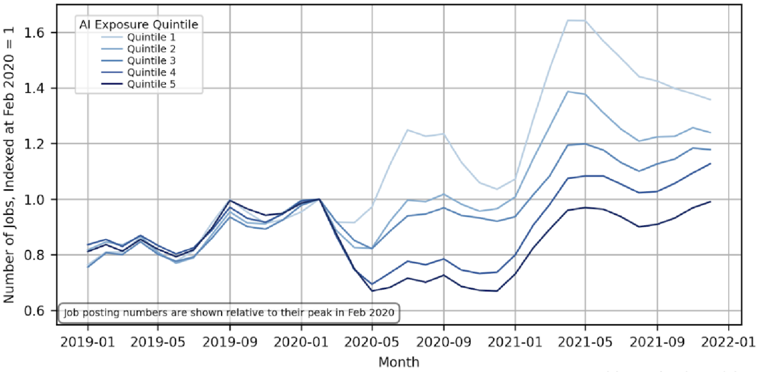


Sources: Lightcast, Eloundou et al. (2024), FRED

This chart overlays the Federal Funds Effective Rate on the job postings index for the most AI-exposed occupations. The strong inverse relationship, with postings declining as interest rates rise from March 2022 onward, points to monetary policy as a primary driver.

Job postings by AI exposure during the COVID-19 shock (Iscenko and Millet ’26)

Chart 4: Job Postings by AI Exposure Quintile During The COVID-19 Shock (Jan 2019 - Dec 2021)



Sources: Lightcast, Eloundou et al. (2024)

This chart looks at the evolution of job postings by AI-exposure quintile around the reduction in job postings observed during 2020, as the COVID pandemic hit. Postings for jobs in the highest AI-exposure quintile experienced a more marked decline, pointing to their more pronounced cyclicality.

Agenda

1. **The big questions**
2. **For what activities is AI a substitute for, or complement to, labor**
   - We won't be missed
   - Two views: Bottlenecks vs. feedback loops
3. **Is it sufficient to think of AI as automating or instantiating tasks?**
   - Automation in the task model
   - New task creation in the task model
   - Expertise leveling in the task model
   - Increasing the substitutability of indivisible factors (assignment model)
4. **Organizational design for AI**
   - Knowledge hierarchies view
   - Broader views on management, work design, and organizational structure
5. **Using AI to support learning (expertise acquisition) and performance (expertise deployment)**
6. **State of evidence on employment effects**
7. **Automation, collaboration, and expertise**