
PIIE
PETERSON INSTITUTE FOR
INTERNATIONAL ECONOMICS
OECD
INTERNATIONAL MONETARY FUND
WORLD BANK GROUP
www.piie.com
@PIIE
linkedin.com/company/piie

AI Adoption: Widespread but shallow

Evidence from the Global Economic Survey on AI

A. Cancellieri, F. Filippucci, P. Gal, T. Kozluk, M. Schief, and **F. Unsal**

Restricted Use - À usage restreint

AI is expected to raise potential growth and productivity, but that requires broad-based adoption

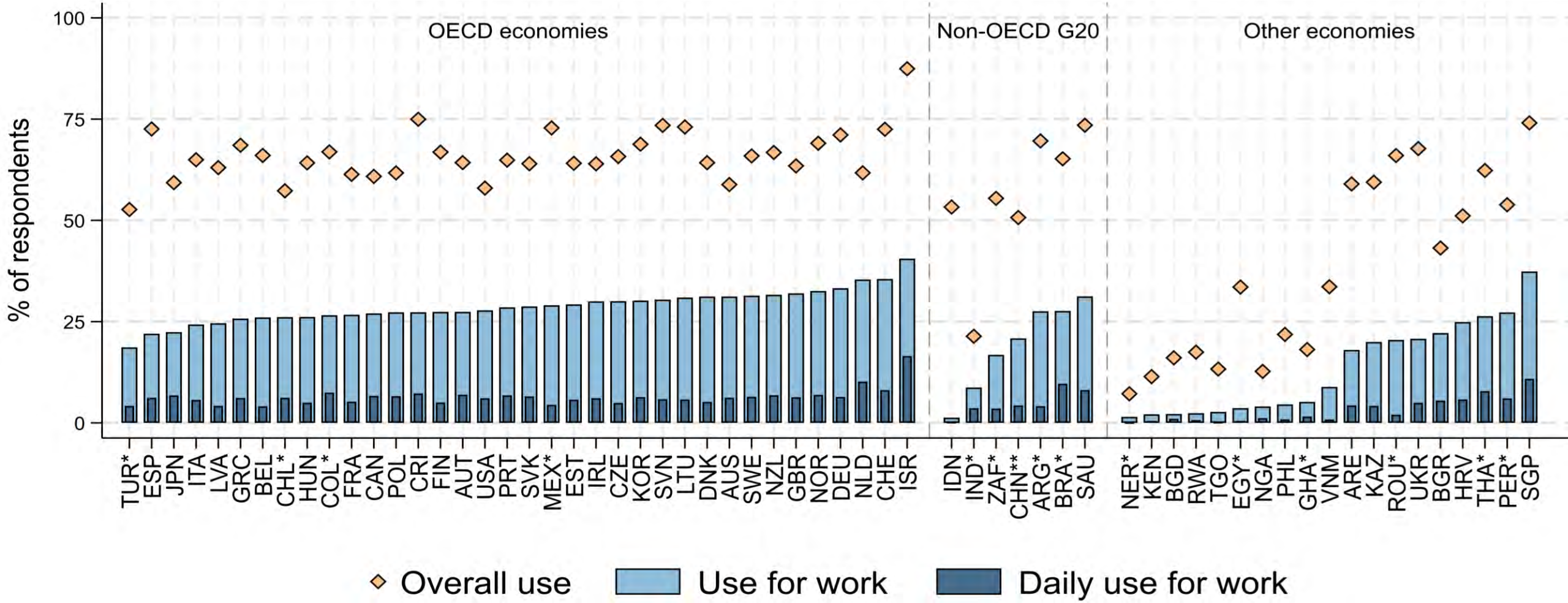
- **AI's productivity potential** (Acemoglu, 2025; Aghion and Bunel, 2026)
- At the OECD, we studied the impact of AI:
  - Across sectors: How a Baumol effect can limit aggregate growth? (Filippucci, Gal and Schief, 2025)
  - Across countries: How trade links may make AI benefits spillover across countries? (Filippucci et al., 2026)
    - In our projections, **intensive adoption** is a key driver of AI-driven productivity gains
- Currently, **global information on AI adoption is lacking**:
  - Academic surveys cover a subset of economies (Bick et al., 2026, Yotzov et al., 2026)
  - National statistical offices (US Census BTOS/HTOPS; Eurostat ICT in Household/in Enterprises surveys)
    - Ask different questions in different countries/regions and are available with delay
    - Do not uniformly investigate adoption intensity and barriers to adoption

The Global Economic Survey on AI (GES-AI)

- We provide **timely information and broad coverage (62 countries)** – OECD + G20 + lower-income countries
  - Focusing on individuals
  - Complementing official surveys by statistical offices
- In partnership with **Google and Ipsos**
- Outputs:
  - A **series of working papers,** including work using complementary OECD employers+employees surveys
  - **Data will be publicly released** to the research community **in mid-December**
  - May be repeated to provide **regular updates**

Overall AI use is widespread, but intensive use for work is still limited

Share of respondents using generative AI



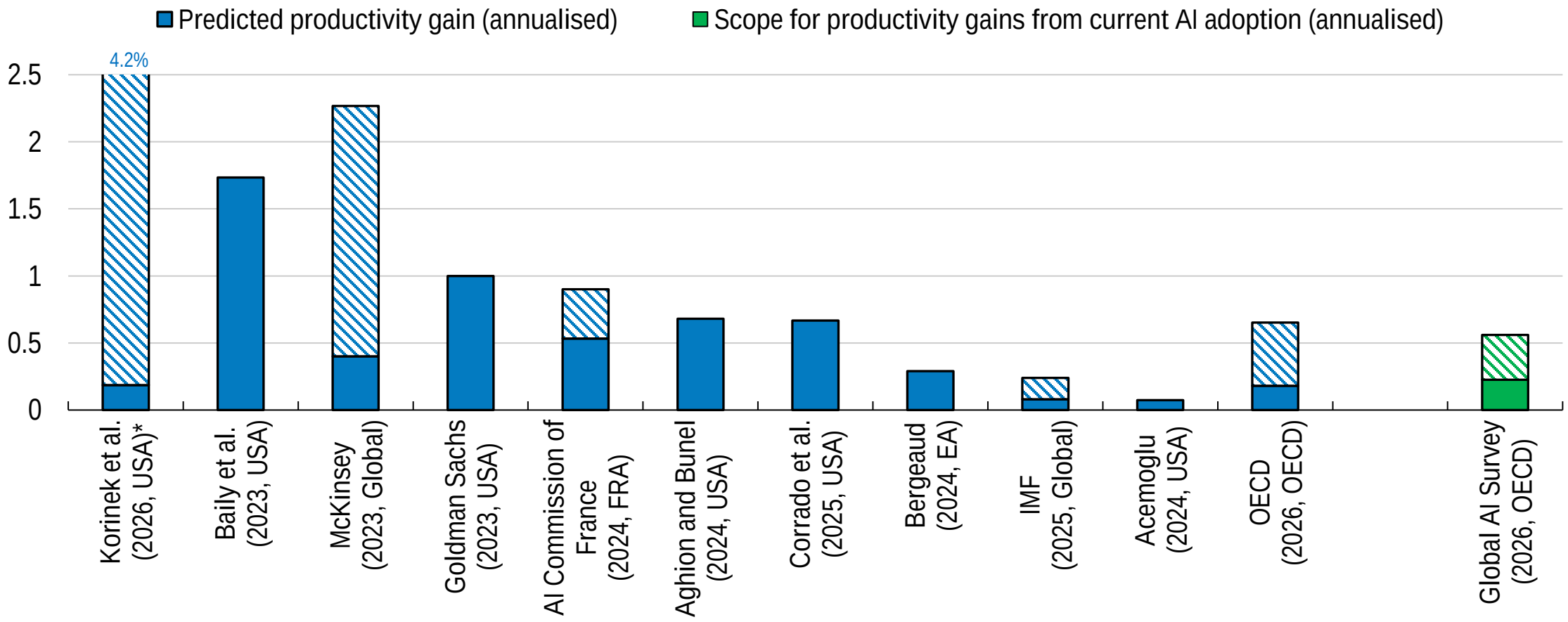
Notes: Overall use denotes current generative AI use for either work or personal purposes. Use for work denotes current generative AI use at work at any frequency, while daily use for work denotes use for work every day. For each country group, estimates are calculated using survey weights. Country groups are ordered by the share of respondents using AI for work within each group. In countries marked with an asterisk (*), the distribution of educational groups in the un-weighted sample differ more substantially from that in the population, leading to higher uncertainty of the estimates. A double asterisk (**) indicates an especially pronounced imbalance and therefore greater uncertainty around the estimates.

Source: Authors' calculations based on the Global Economic Survey on AI.

Survey's adoption rates imply potential productivity gains in line with moderately optimistic projections

AI-driven aggregate growth: projections vs. estimated scope given current adoption

MFP

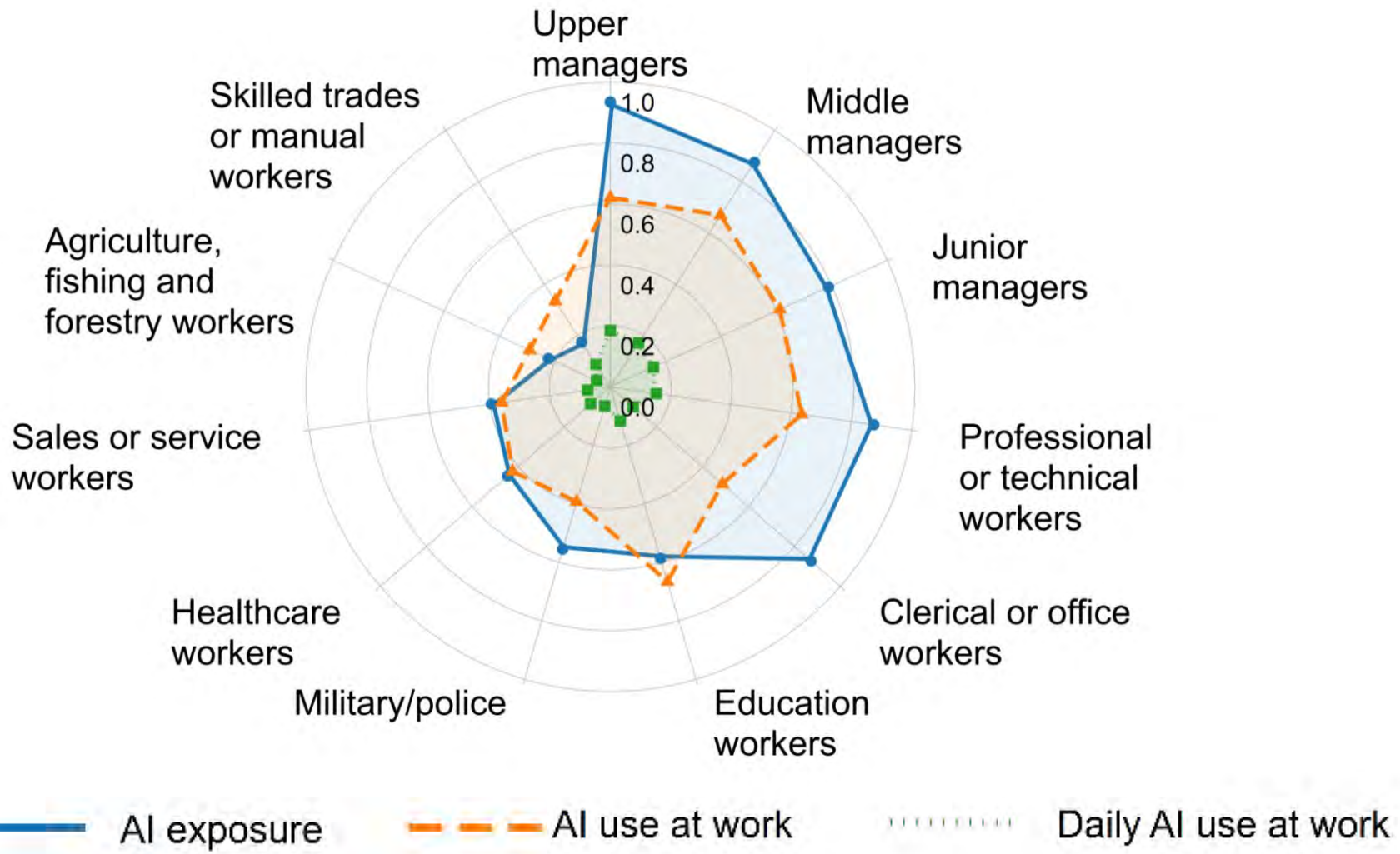


Notes: Most studies originally adopt a 10-year horizon, but Korinek et al. (*) refer to a 6-year horizon. For comparability, all estimates are harmonized as annualised contributions to multifactor productivity (MFP) growth over their respective horizons. Striped areas represent the range of estimates if the paper provides various projections. Where projections are expressed as labour productivity, implied MFP growth is obtained by applying a capital-deepening multiplier of 1.5. The country or area covered by each estimate is listed in parentheses. The green bar shows the estimated scope for MFP gains implied by the level of AI adoption observed in the survey, expressed in annualised terms. The upper (lower) scenario assumes that respondents saying they use AI "several times per week" use it on four (two) out of five working days, "weekly" use corresponds to two (one) out of five, and "less-than-weekly" use corresponds to half a day per week (no use). Daily use corresponds to all five working days in both scenarios. Source: Authors' compilation of relevant studies in the literature. For Goldman Sachs (2023), see Briggs and Kodnani (2023); for IMF (2024), see Rockall, Tavares and Pizzinelli (2025); For OECD (2026), see Filippucci et al. (2026); For IMF (2025), see Cerutti et al. (2025).

AI use is higher in more exposed occupations but lags exposure, suggesting scope for increasing AI use

AI exposure vs. AI adopti

OECD average, by occupatio



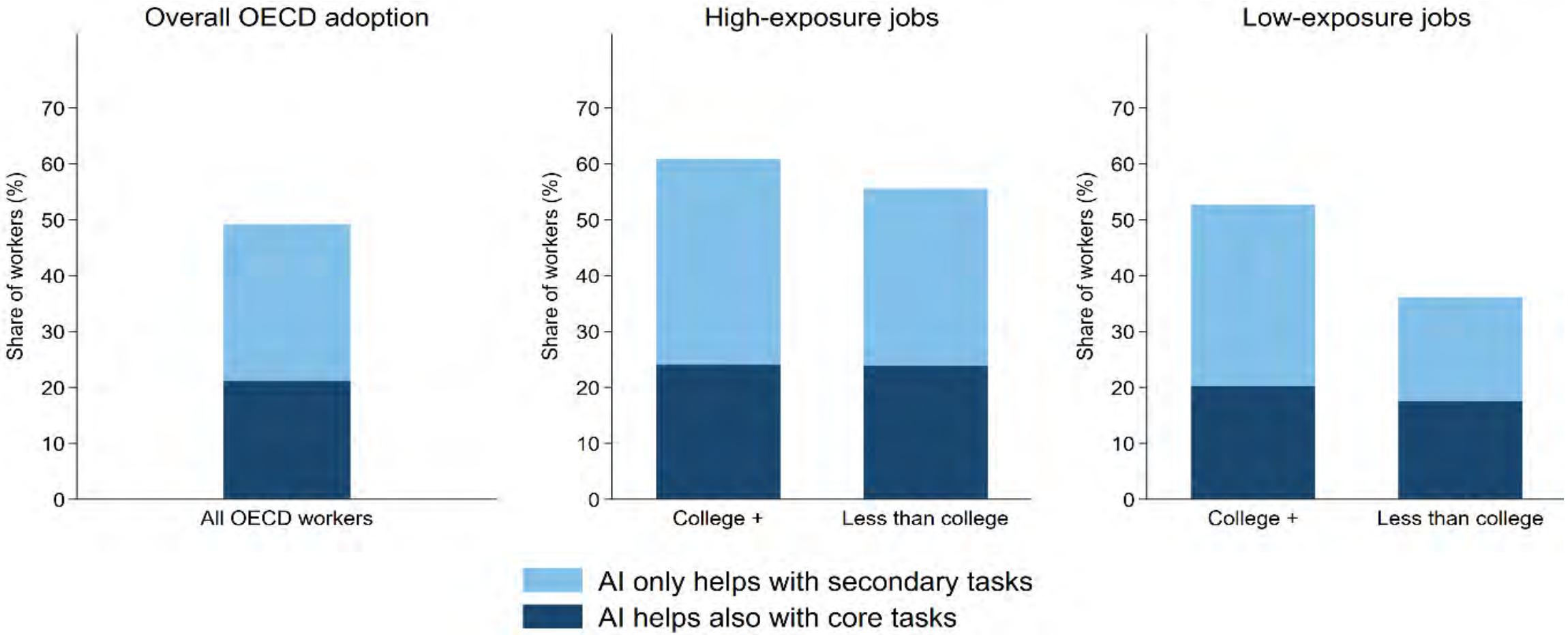
Notes: AI exposure measures the share of occupational tasks in which large language models, either alone or with complementary software, allow for significant time savings. The specific AI exposure measure used in this chart is the "gamma" measure in Eloundou et al. (2024). Task-level exposure scores are averaged within detailed occupations and aggregated to the occupation groups shown using US employment weights. The same occupation-level exposure estimates are used in both panels. Current and daily AI use refer to the shares of employed respondents in each occupation who use generative AI for work at any frequency or daily, respectively. Panel A reports estimates for the United States. Panels B, C and D report unweighted averages across surveyed countries in Europe, other OECD countries and non-OECD G20 countries, respectively.

Source: Authors' calculations based on survey responses, exposure measures from Eloundou et al. (2024), employment data from the US Bureau of Labor Statistics.

AI use is higher among college-educated workers, but not for use in *core tasks*

Share of workers using generative AI for work

By education and occupational AI exposure



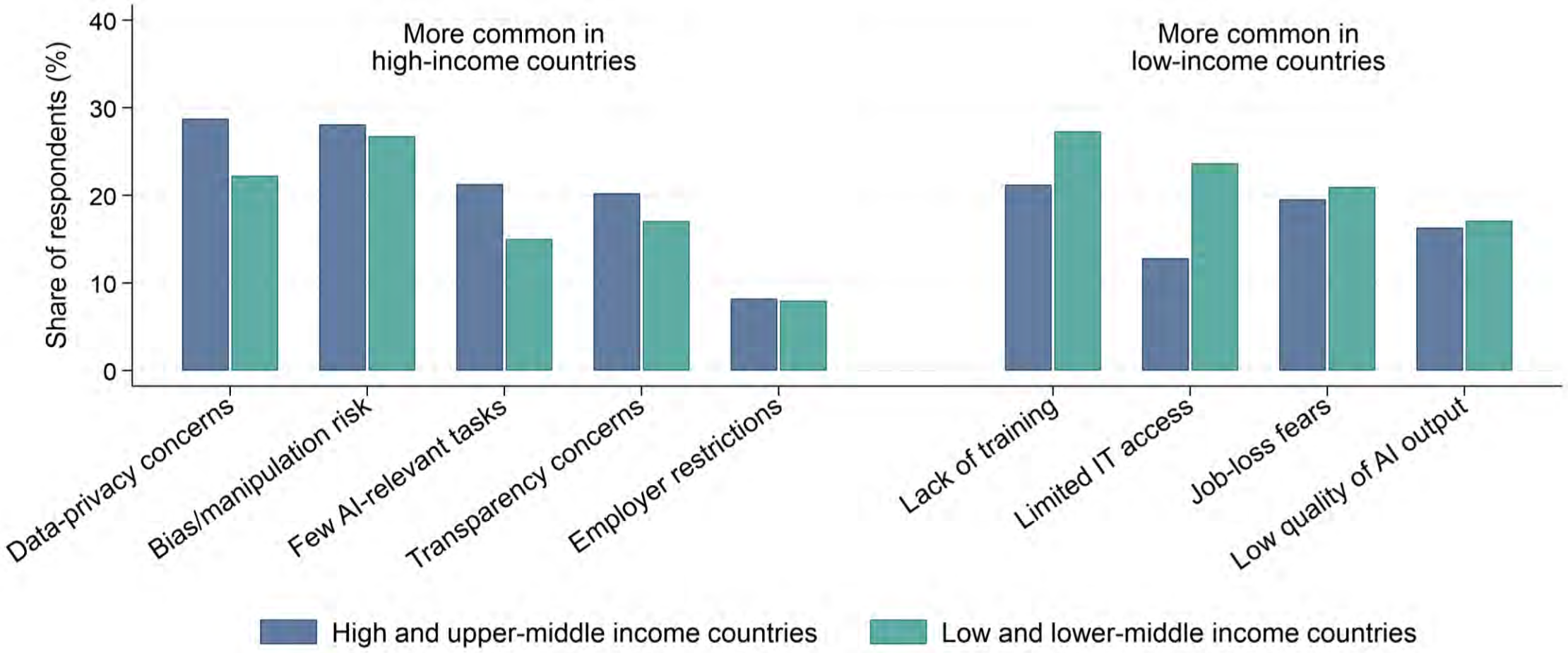
Notes: The bars show the share of employed respondents who currently use generative AI for work. The dark-blue component includes workers reporting that AI helps mainly with core tasks or with core and secondary tasks equally; the light-blue component includes workers reporting that AI helps mainly with secondary tasks. "College +" denotes tertiary education, while "Less than college" comprises workers with lower levels of education. High-exposure jobs comprise the five occupational groups with the highest AI exposure among the 11 mapped occupation groups; the remaining six groups are classified as low-exposure. Estimates use responses from surveyed OECD countries.

Source: Authors' calculations based on the Global Economic Survey on AI and occupational AI exposure measures from Eloundou et al. (2024).

Privacy and bias concerns are the main barriers to intensive AI use for work, while lack of training and IT access is more common in lower-income countries

Share of current users of AI at work reporting each barrier

by country income group

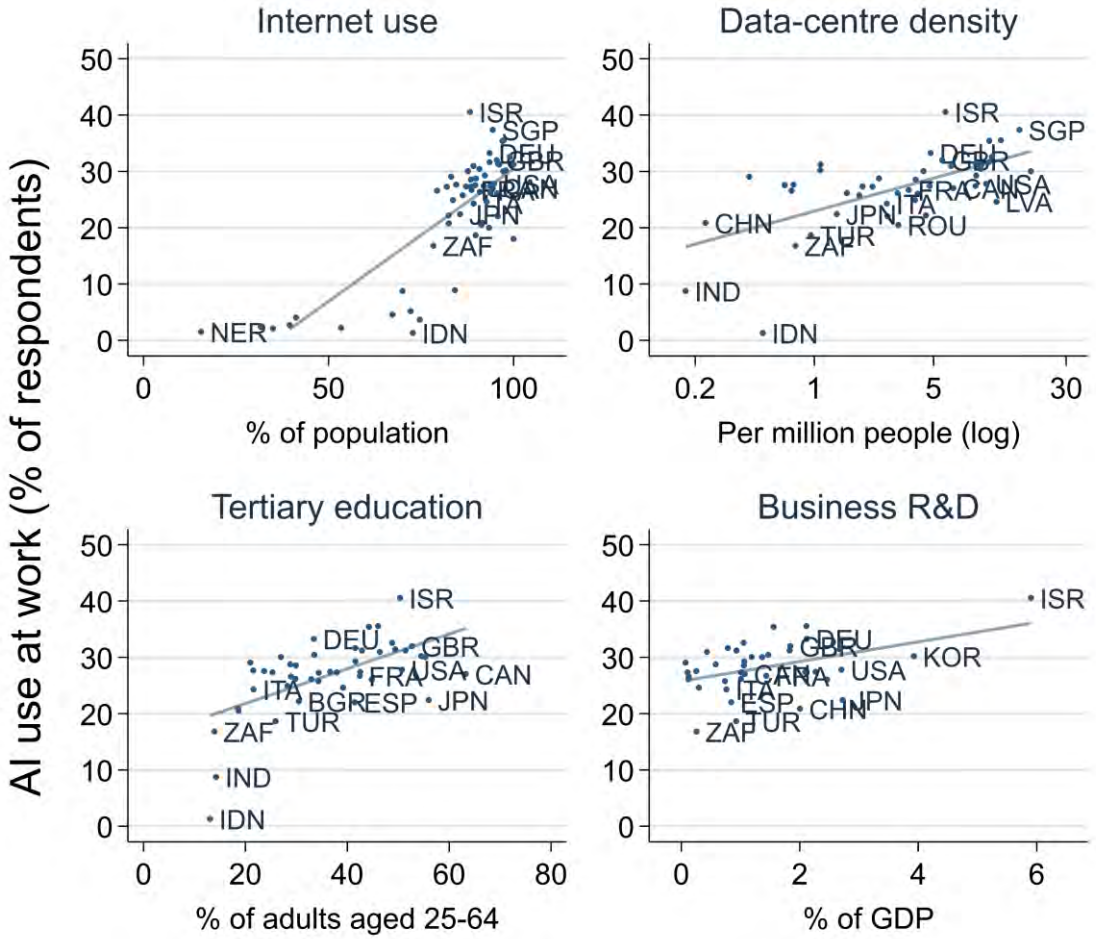


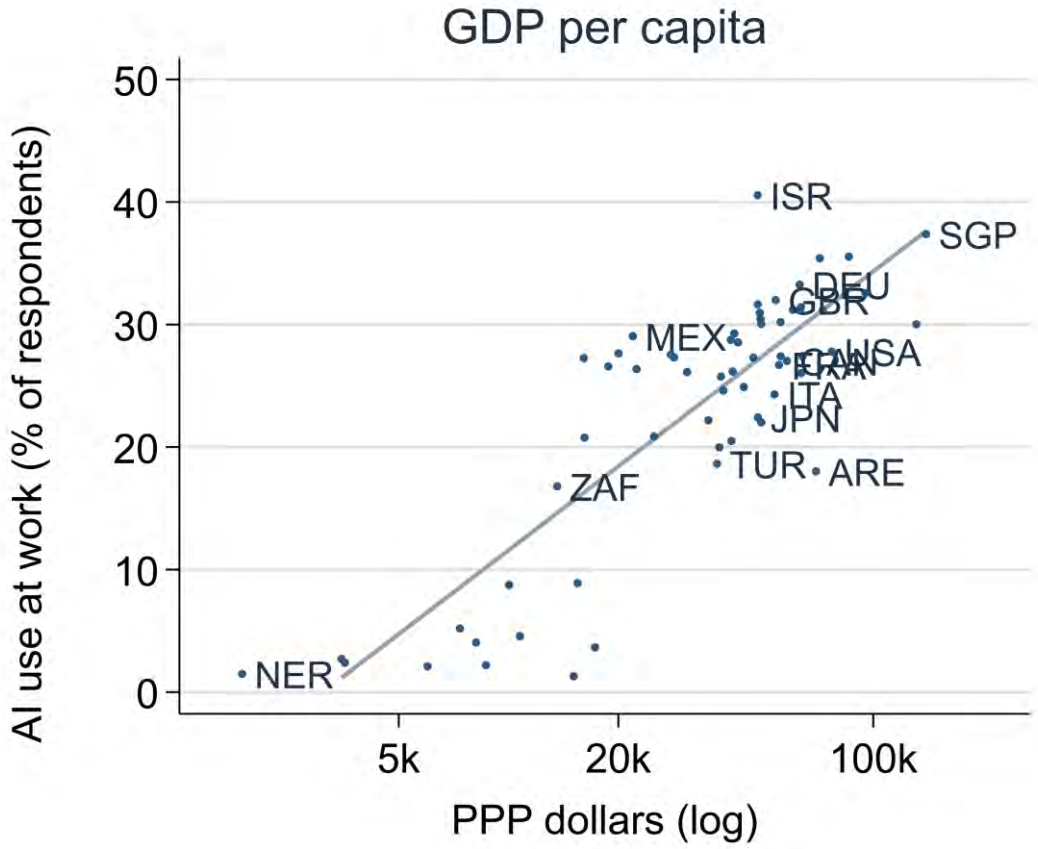
Notes: The figure reports cross-country averages of the share of current AI users at work reporting each barrier to greater AI use. Respondents are retained if they provided responses to the barrier questions and selected at least one response. Respondents selecting "I do not need to use AI more" as their only response are retained and counted as reporting none of the nine barriers shown; responses combining this option with another selection, as well as refusals, are excluded. Multiple responses were allowed, so the shares do not sum to 100%. Country-level estimates are calculated using survey weights and are then averaged across countries, giving each country equal weight. Barriers on the left are reported more frequently in high- and upper-middle-income countries, while those on the right are reported more frequently in low- and lower-middle-income countries.

Source: Authors' calculations based on the Global Economic Survey on AI.

AI use for work is positively related to digital infrastructure, skills, innovation and general economic development

Correlation between AI adoption at work and country characteristics



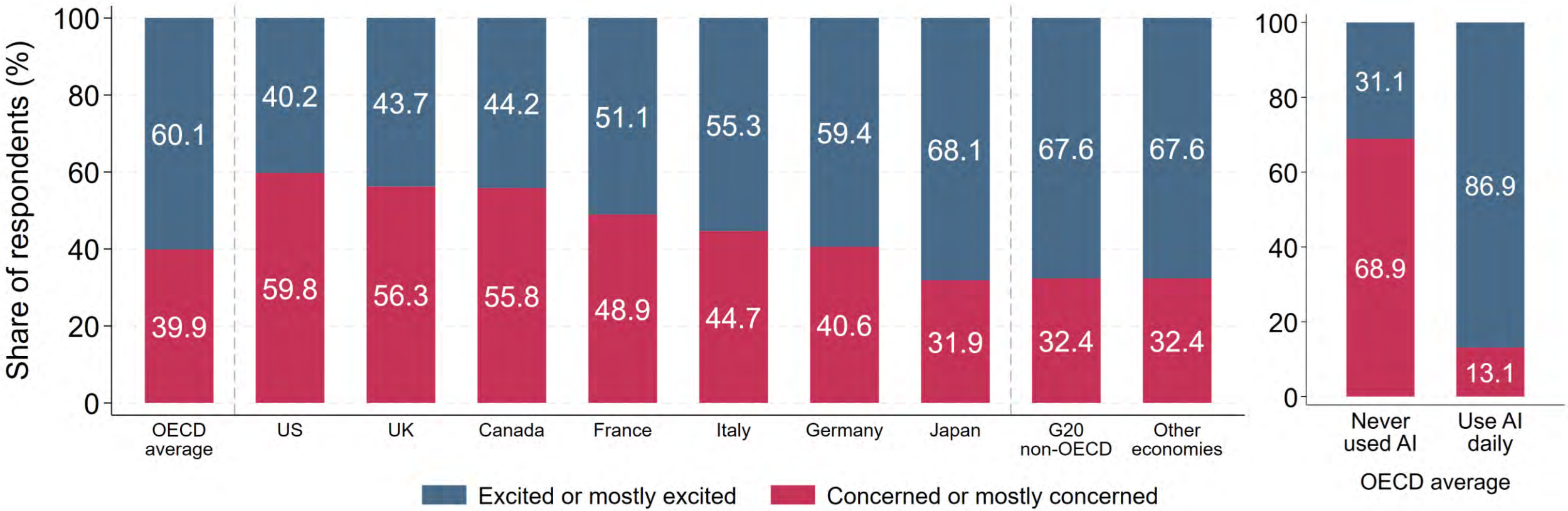


Notes: Each point represents a country. AI use for work is the share of all respondents who are employed and report currently using generative AI for work. Country estimates use survey weights and give equal weight. Internet use is the share of the population using the internet (2024–25); data-centre density is the number of data centres in 2025 per million people; tertiary education is the share of adults aged 25–64 with tertiary attainment; business R&D is expenditure on R&D as a percentage of GDP; and GDP per capita is expressed in constant 2021 PPP dollars. Education and R&D data refer to 2023 or the latest available year, while GDP data refer to 2024–25. Data-centre density and GDP per capita are displayed on logarithmic scales. Lines show unweighted linear fits; country coverage varies by data availability. Source: Authors' calculations based on the OECD Global AI Survey, World Bank World Development Indicators, Data Center Map and OECD databases.

Sentiment around AI varies across countries, with stark concentration of concerns among never-users

Excitement and concern about AI

By country group and G7 economies

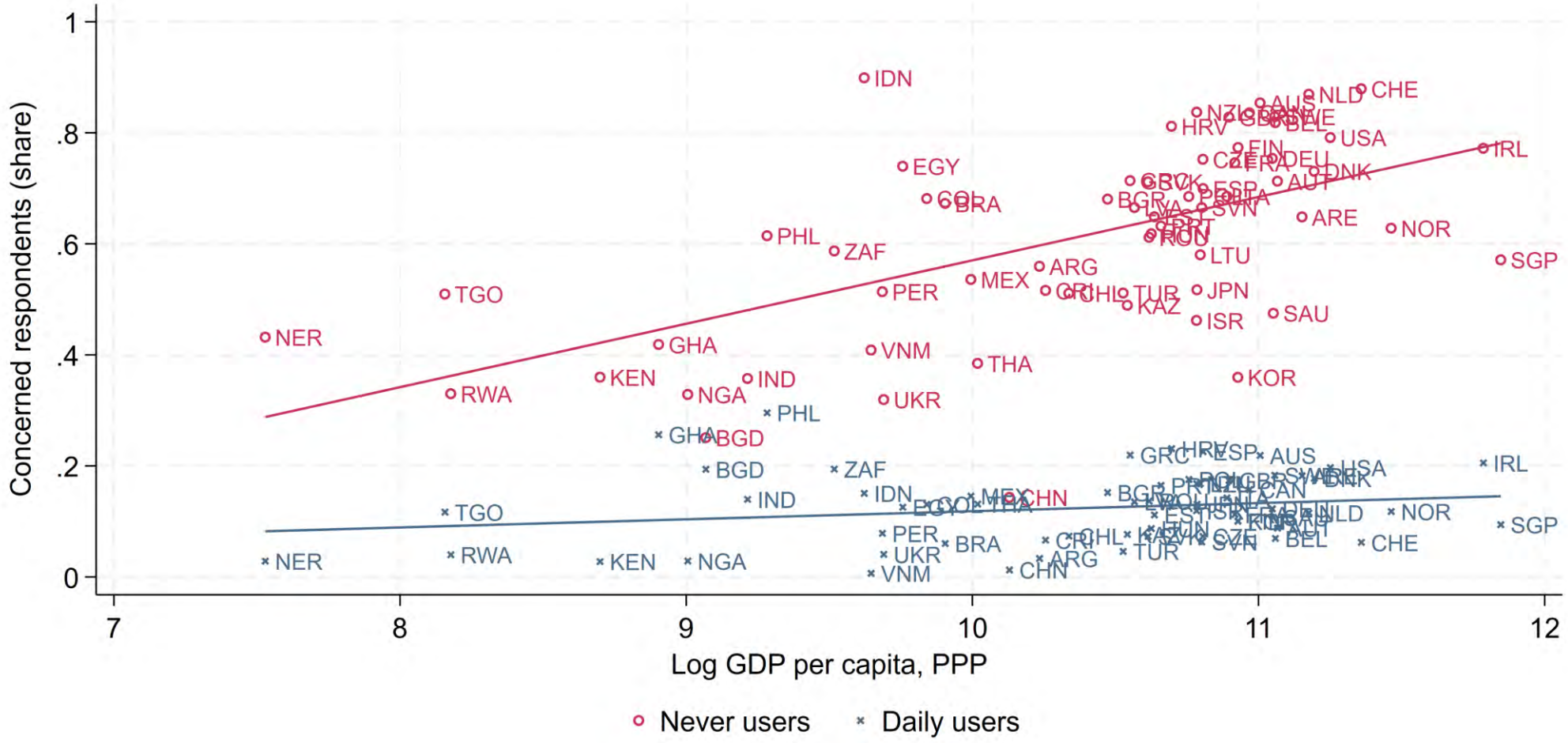


Notes: Respondents are classified as "excited or mostly excited" if they report being "mostly excited about the possibilities" or "excited about the possibilities", and as "concerned or mostly concerned" if they report being "concerned about the risks" or "mostly concerned about the risks, but somewhat excited about the possibilities". "Not sure" responses, refusals and non-responses are excluded. Panel includes all respondents with a substantive sentiment response, irrespective of employment status or AI use.

Source: Authors' calculations based on the Global Economic Survey on AI.

Among never-users, concerns about AI are more prevalent in higher- income economies

Correlation between attitudes towards AI and real GDP per capita, separately for never users and daily use



Notes: Note: Lines show unweighted linear fits across countries.

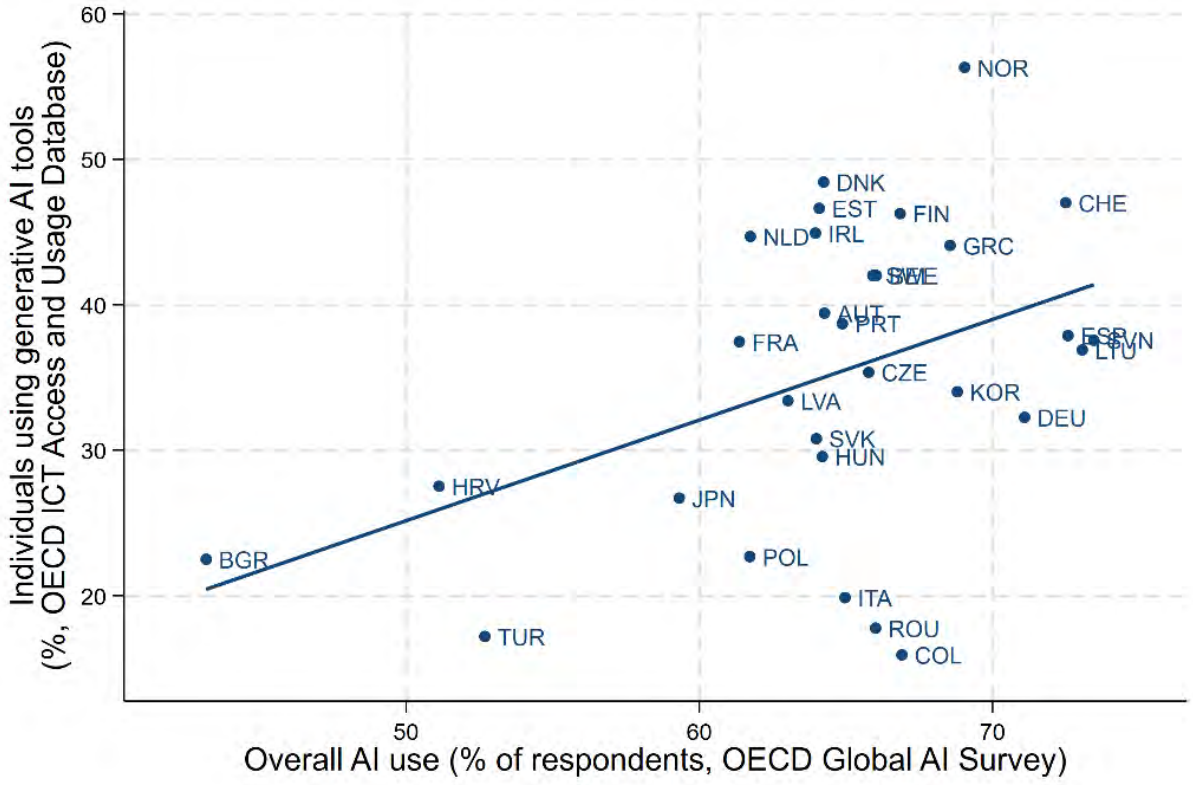
Source: Authors' calculations based on the Global Economic Survey on AI.

Thank you

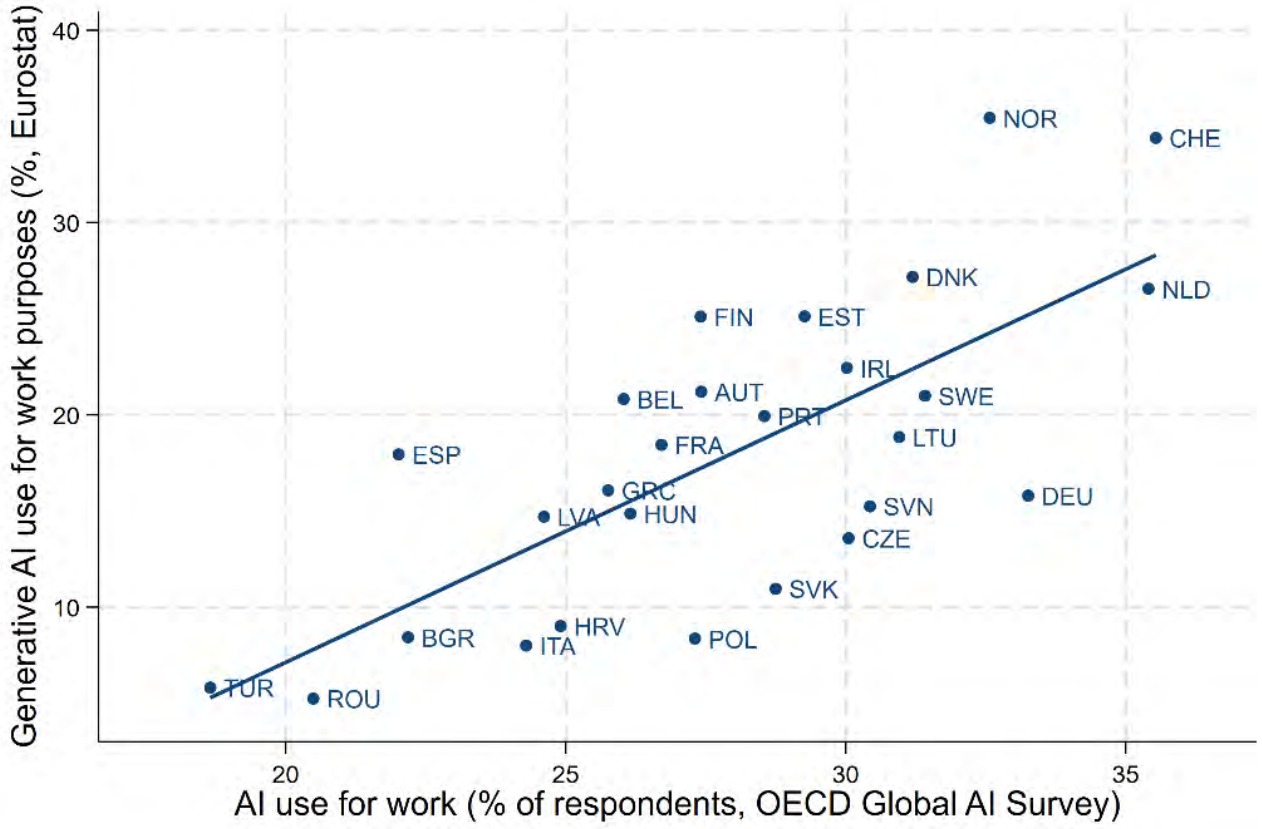
Restricted Use - À usage restreint

Comparison with other cross-country AI use data – individual surveys

Panel A: Statistical offices, overall AI use



Panel B: Statistical offices, AI use for work

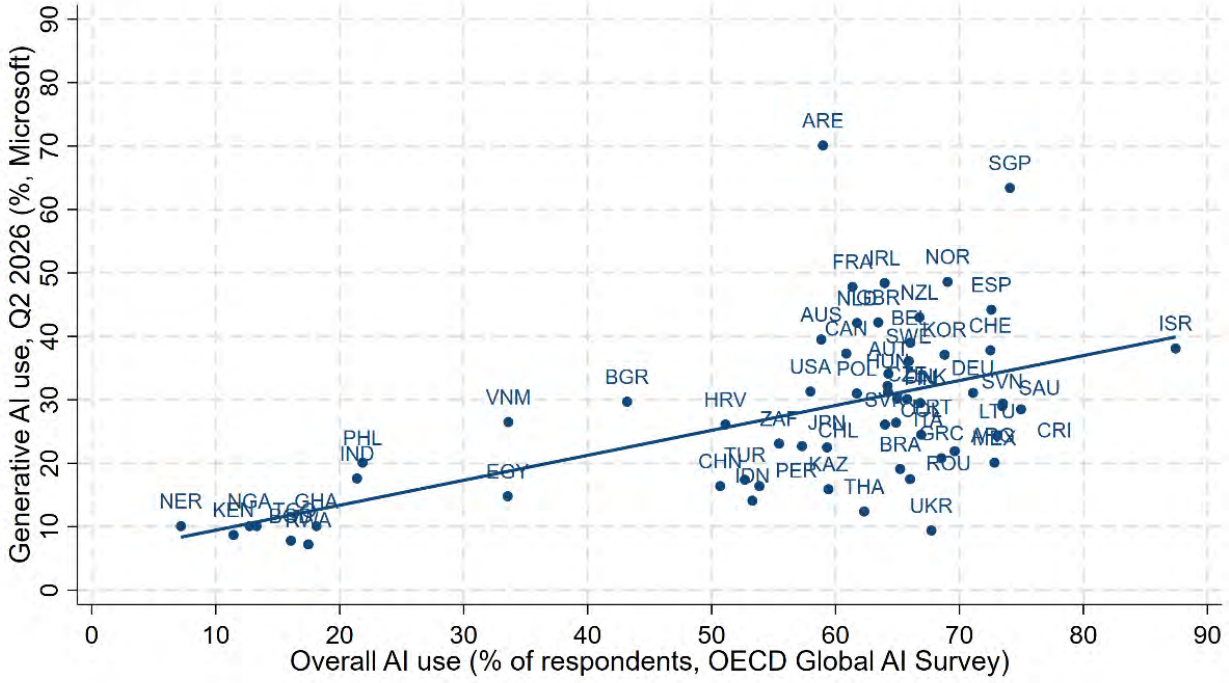


Notes: Panel A compares the share of respondents reporting having ever used generative AI in the OECD Global AI Survey with the share of individuals who used generative AI tools during the three months preceding the official ICT survey, as reported in the OECD ICT Access and Usage Database. Panel B compares AI use for work among employed respondents in the OECD Global AI Survey with the share of individuals who used generative AI for professional purposes during the preceding three months, as reported by Eurostat. Lines show linear fits across countries.

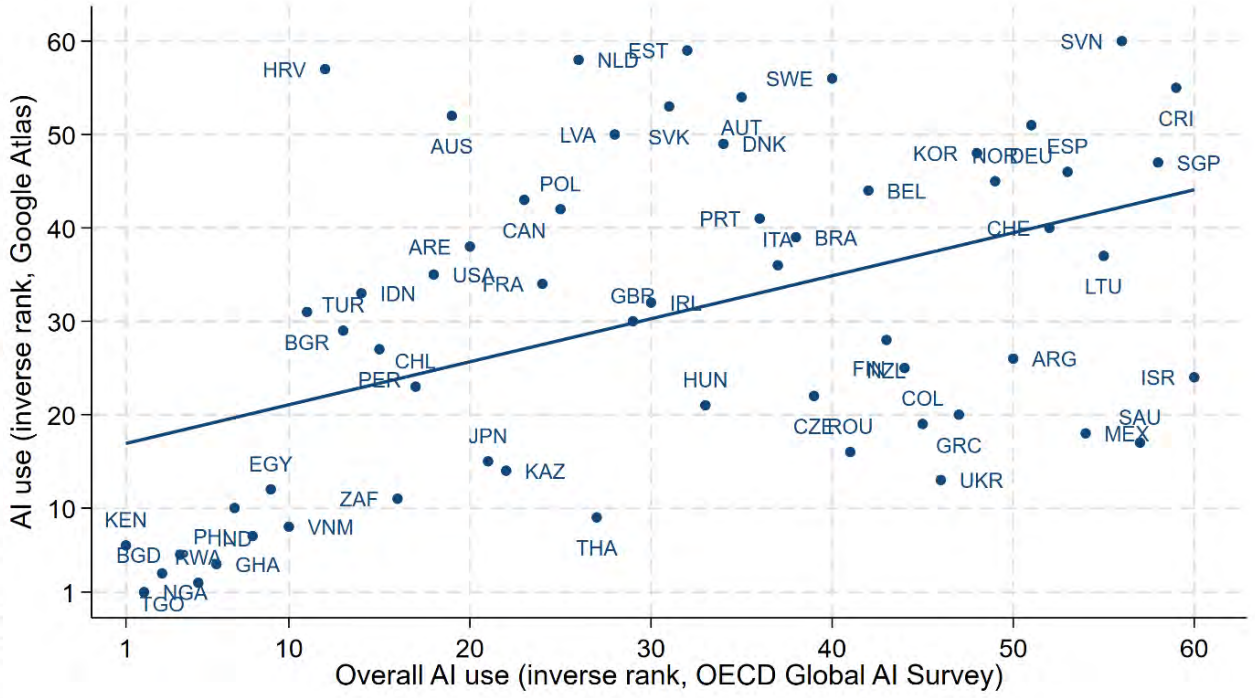
Source: Authors' calculations based on the OECD Global AI Survey, the OECD ICT Access and Usage Database (Panel A), and Eurostat (Panel B).

Comparison with other cross-country AI use data
Web-traffic based AI usage data

Panel A: Microsoft's Global AI Diffusion data, overall AI use



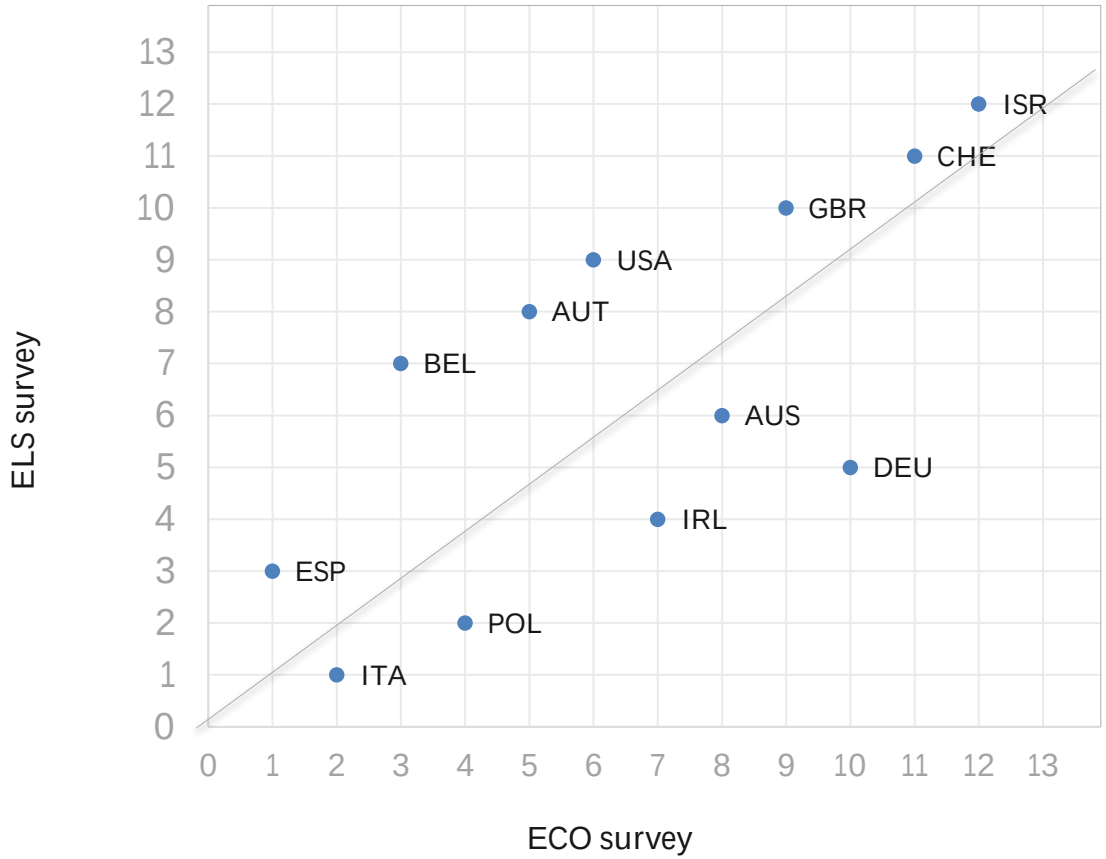
Panel B: Google's Atlas data, overall AI use (rank)



Notes: Each point represents a country. Panel A compares overall AI use in the OECD Global AI Survey with Microsoft's AI Diffusion data. The latter estimates the share of the population that used a generative AI product, based on aggregated and anonymised telemetry adjusted for device and operating-system market shares, internet penetration and country population. Panel B compares countries' rankings by overall AI use in the OECD Global AI Survey with their rankings by penetration-adjusted conversational AI use per capita in Google's ATLAS. In both rankings, higher values indicate greater AI use. The ATLAS measure is based on interactions with the Gemini App and Google AI Mode and is adjusted for differences in Gemini's market penetration using country-level web-referral data. Lines show linear fits across countries.
Source: Authors' calculations based on the OECD Global AI Survey; Microsoft AI Economy Institute, Global AI Diffusion Report, Q2 2026; and Google, AI & Economy ATLAS v1.0.

Comparison with ELS's survey among employees

Country ranking in ECO and ELS surveys (inverse ranks)

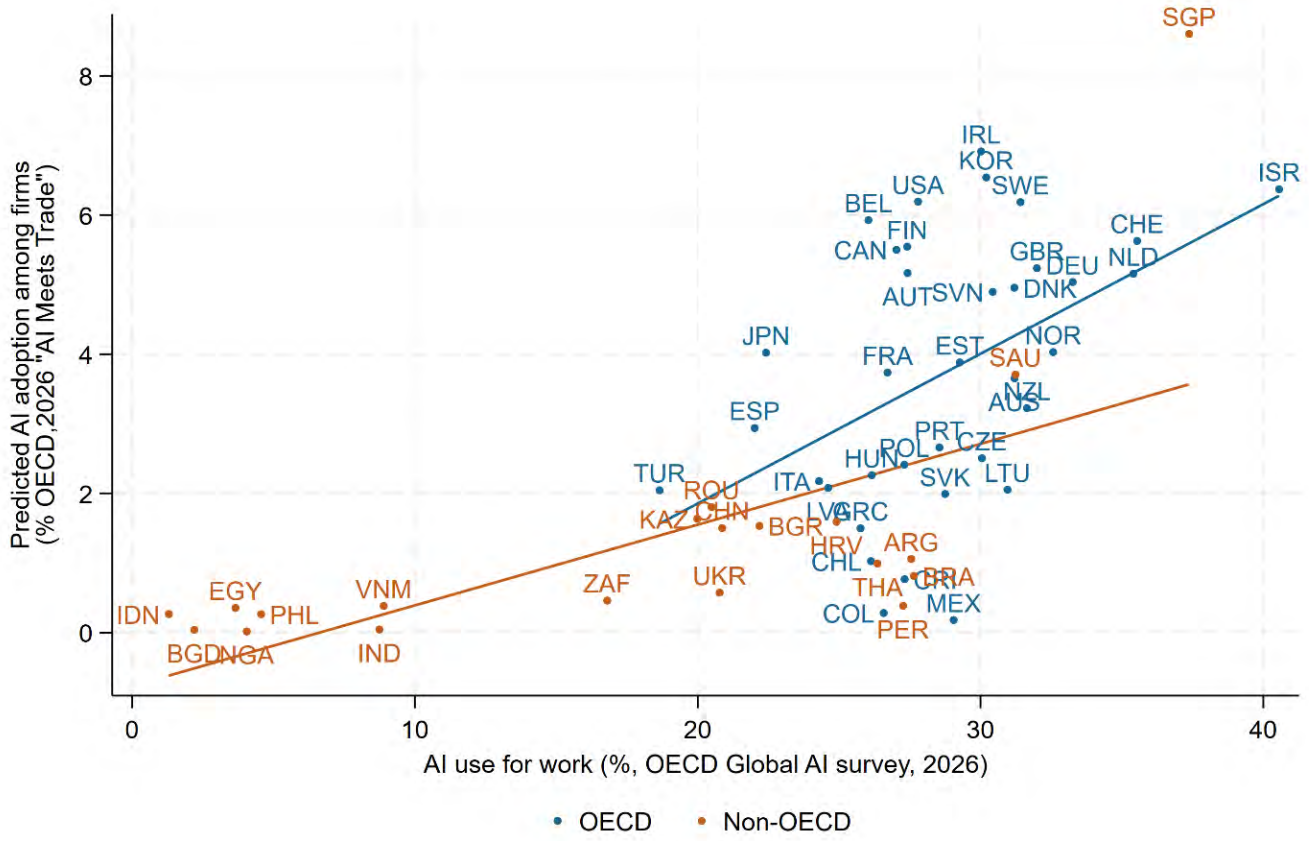


Notes:

ELS survey population: workers. ECO survey population:
Type of AI use in ELS survey is any AI, while in ECO survey it is Generative AI

Comparison with previous OECD/ECO estimates of firm-level adoption

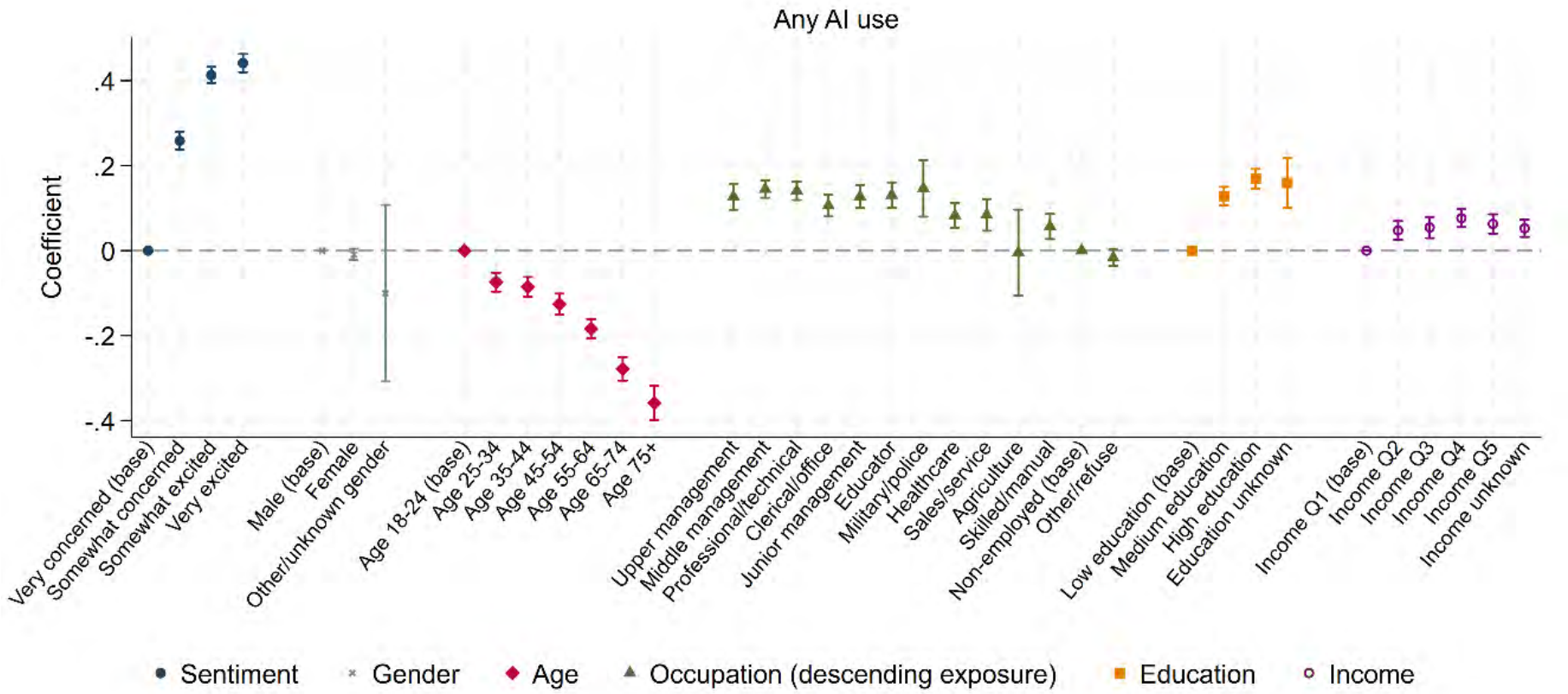
Previous OECD estimates on firm level adoption and AI use for work in the survey population from the Global AI survey



Notes: Each point represents a country. Panel A compares overall AI use in the OECD Global AI Survey with Microsoft's AI Diffusion data. The latter estimates the share of the population that used a generative AI product, based on aggregated and anonymised telemetry adjusted for device and operating-system market shares, internet penetration and country population. Panel B compares countries' rankings by overall AI use in the OECD Global AI Survey with their rankings by penetration-adjusted conversational AI use per capita in Google's ATLAS. In both rankings, higher values indicate greater AI use. The ATLAS measure is based on interactions with the Gemini App and Google AI Mode and is adjusted for differences in Gemini's market penetration using country-level web-referral data. Lines show linear fits across countries.
Source: Authors' calculations based on the OECD Global AI Survey; Microsoft AI Economy Institute, Global AI Diffusion Report, Q2 2026; and Google, AI & Economy ATLAS v1.0.

Positive views of AI are associated with adoption, even after accounting for individual characteristics

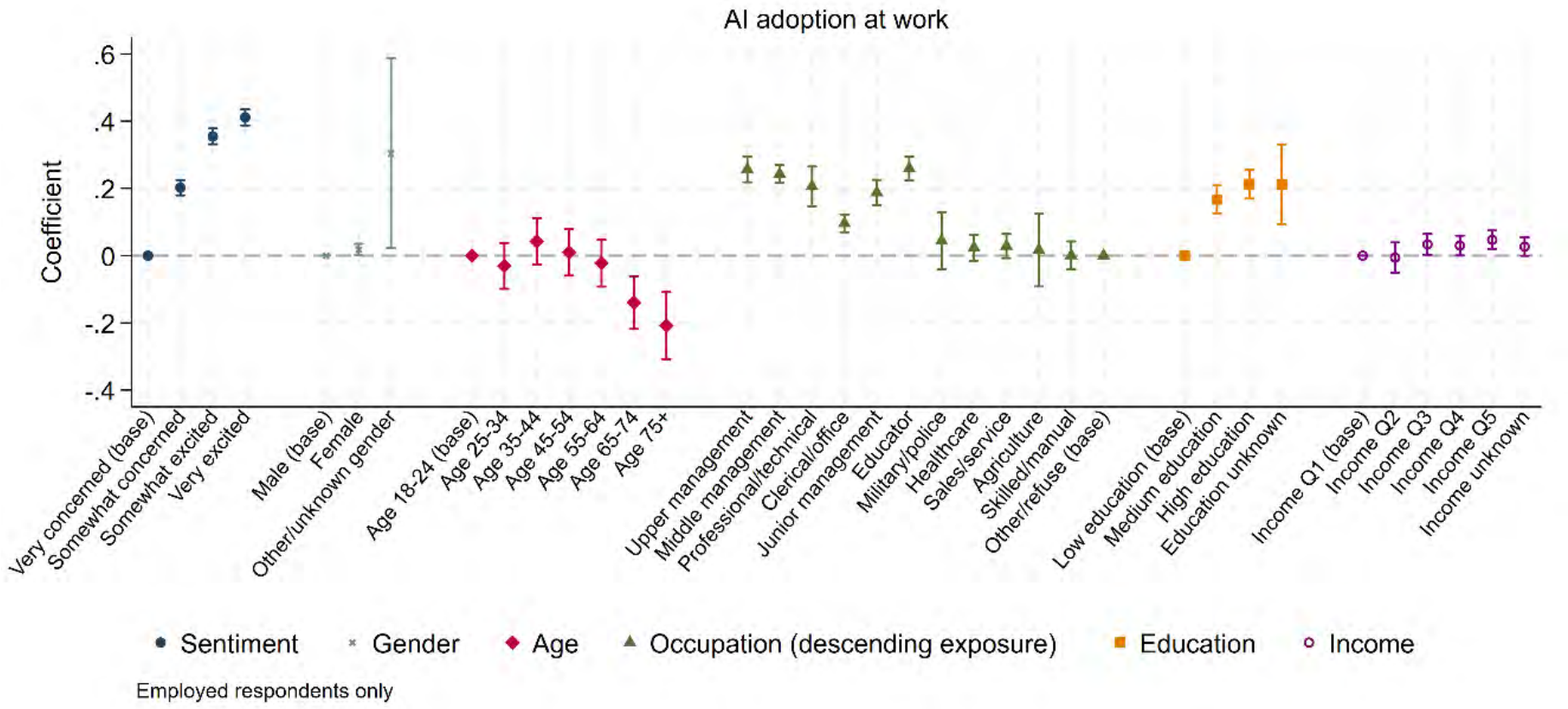
Estimated percentage-point differences in any AI use associated with sentiment and individual characteristics



Notes: Points report coefficients from pooled weighted linear probability regressions; vertical lines show robust 95% confidence intervals. All characteristics shown are included jointly. The upper panel includes all respondents and analyses any current AI use; the lower panel includes employed respondents and analyses current AI use for work. Reference categories are very concerned about AI, male, aged 18–24, low education and the bottom income quintile; management is the reference occupation in the employed sample. Missing or unknown responses are retained as separate categories.
Source: Authors' calculations based on the OECD Global AI Survey.

Positive views of AI are associated with adoption, even after accounting for individual characteristics

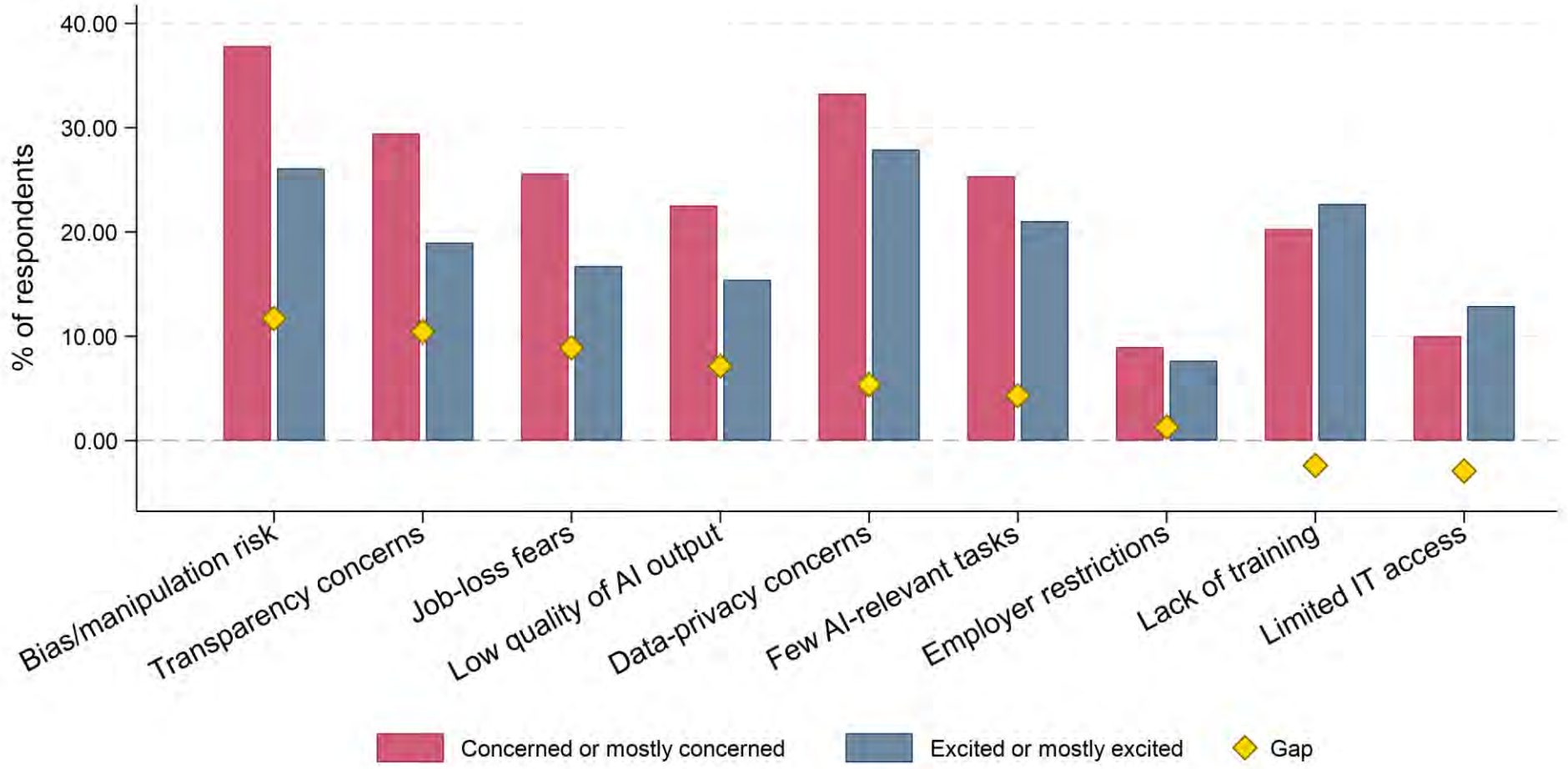
Estimated percentage-point differences in AI use for work associated with sentiment and individual characteristics



Notes: Points report coefficients from pooled weighted linear probability regressions; vertical lines show robust 95% confidence intervals. All characteristics shown are included jointly. The upper panel includes all respondents and analyses any current AI use; the lower panel includes employed respondents and analyses current AI use for work. Reference categories are very concerned about AI, male, aged 18–24, low education and the bottom income quintile; management is the reference occupation in the employed sample. Missing or unknown responses are retained as separate categories.
Source: Authors' calculations based on the OECD Global AI Survey.

Privacy and bias concerns are the main barriers to more intensive AI use at work in OECD countries

Share of current users of AI at work reporting barriers to greater use, by sentiment towards AI – OECD countries



Notes: The figure reports the share of current AI users for work identifying each factor that prevents using AI more. Respondents are retained if they provided responses to the barrier questions and selected at least one response. Respondents selecting "I do not need to use AI more" as their only response are retained and counted as reporting none of the nine barriers shown; responses combining this option with another selection, as well as refusals, are excluded. Multiple responses were allowed, so the shares do not sum to 100%. Country-level estimates are calculated using survey weights and are then averaged across countries, giving each country equal weight. It reports the unweighted average across the surveyed OECD countries.

Source: Authors' calculations based on the Global Economic Survey on AI.


PIIE
PETERSON INSTITUTE FOR
INTERNATIONAL ECONOMICS
OECD
INTERNATIONAL MONETARY FUND
WORLD BANK GROUP
www.piie.com
@PIIE
linkedin.com/company/piie



Summary of Results

AI, Productivity, and the Workforce: Evidence from Corporate Executives

Salomé Baslandze, Zach Edwards, John R. Graham, Ty McClure, Brent Myers, Michael Sparks, Sonya Ravindranath Waddell, Daniel Weitz

DUKE FUQUA SCHOOL OF BUSINESS

Federal Reserve Bank *of* Atlanta

FEDERAL RESERVE BANK OF RICHMOND
Richmond • Baltimore • Charlotte

Nasdaq

fei financial executives international

CFO Survey on Corporate AI Usage

Published in Journal of Finance: Insights and Perspectives

Disclaimer: The views expressed here are those of the authors and not of the Federal Reserve Banks of Atlanta, Richmond, or the Federal Reserve System.

Year-end 2025 asked CFOs about their firms **use AI**, **adoption** and how much they **spend** on AI, and the (expected) effects of AI on their companies' **productivity and labor force.**

734 executives responded to the AI survey.

Survey is about **2025 realizations and 2026 expectations** (plus, two questions about 2028).

Focus is on the ***adoption and use of AI at a typical company at the initial stage of adoption of a new technology*** more so than about developers of AI technology or data center construction.

Dashboard of open-ended responses:
https://people.duke.edu/~jgraham/CFO_Survey_AI_Dashboard.html

Outline

1. **AI Adoption, Spending, and Motivation**
   a. Widespread adoption but shallow (small $$)
   b. Per employee, small companies spend *more* than large companies
   c. Most spending is operational (not on hardware; more spending on subscriptions)

2. The Broad Effects of Using AI

3. Effects of Using AI – Productivity

4. Effects of Using AI – Workforce, Jobs, Tasks

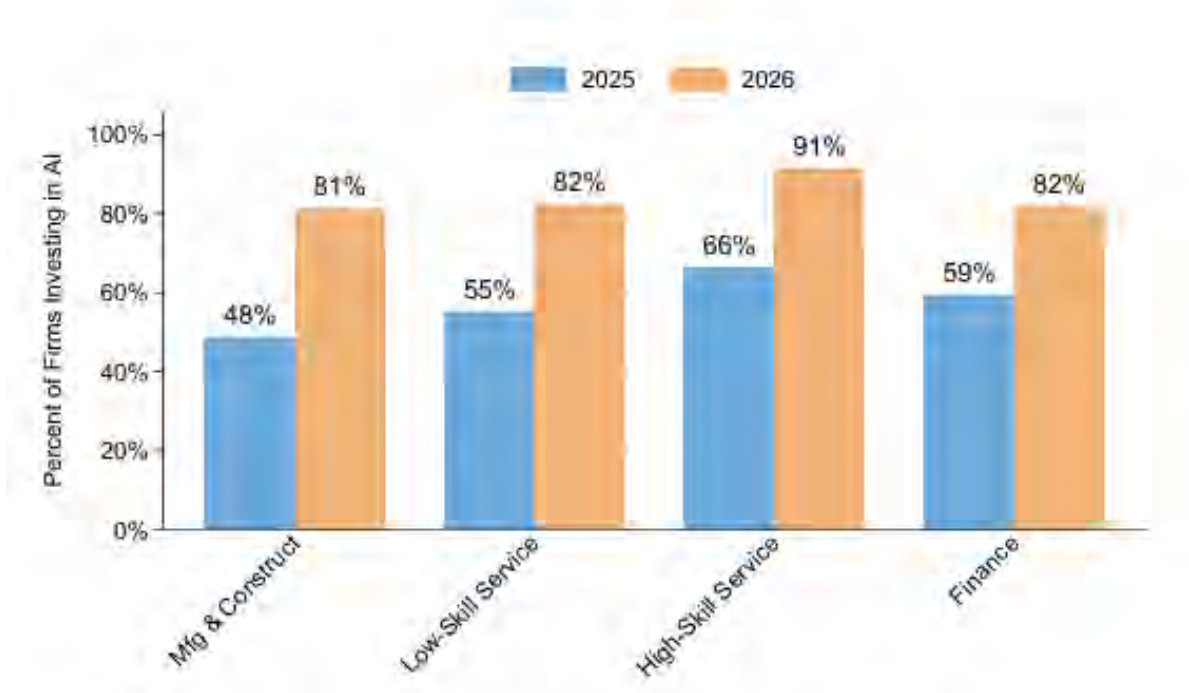
5. Takeaways and Future Research

Widespread AI Adoption – by Sector

While the majority of companies invested in AI during 2025, substantially more will invest in 2026. **Widespread**

- Small companies in particular will increasingly adopt AI in 2026.

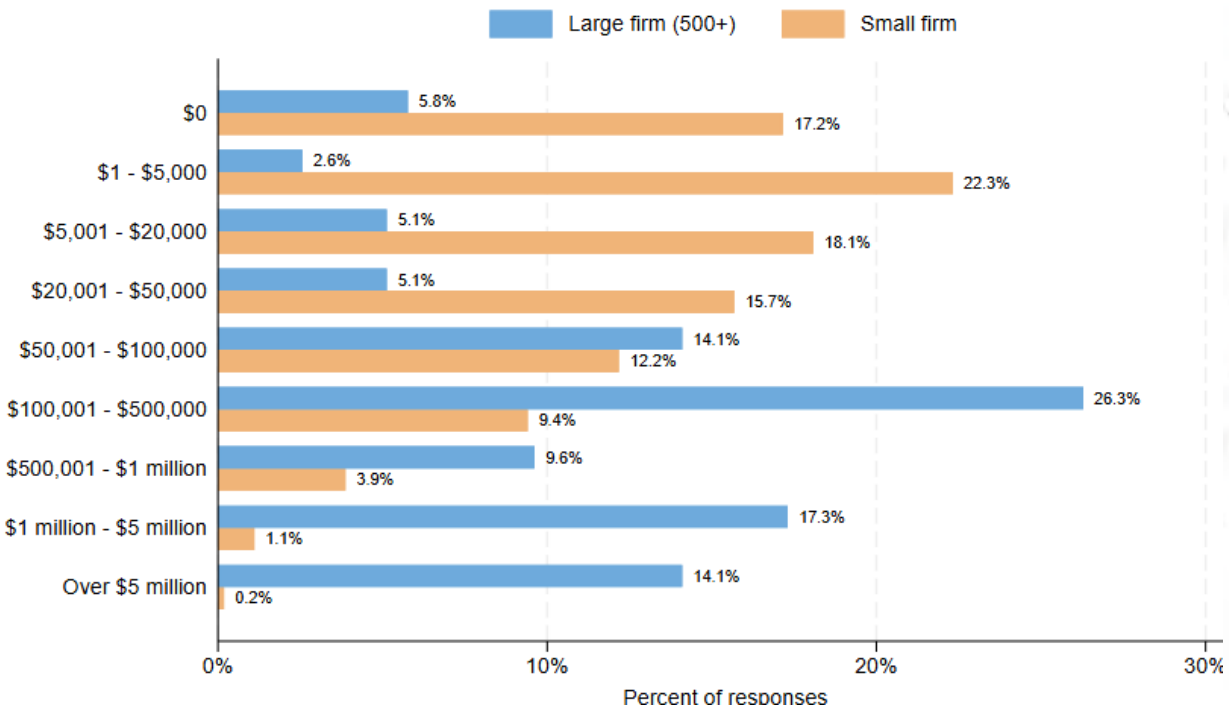
AI adoption is largest in high-skill service sector (see chart).



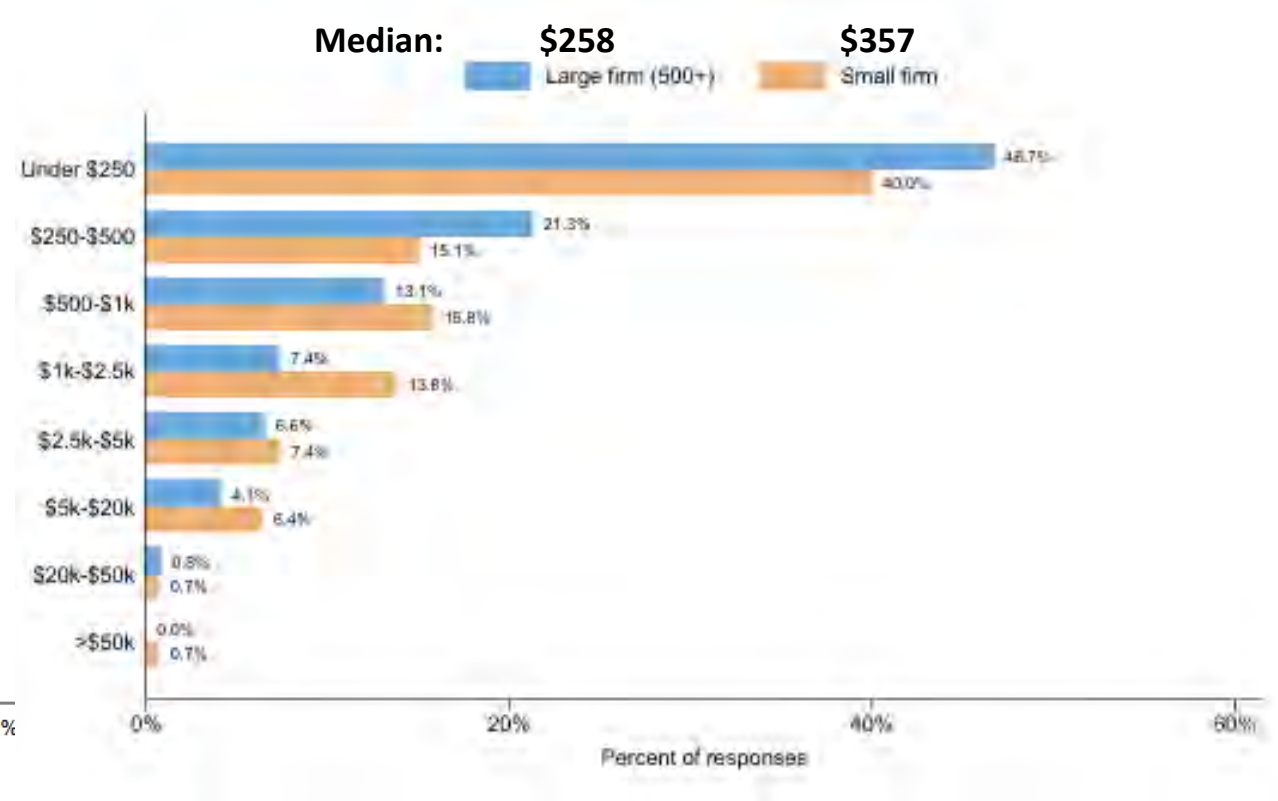
Percent of Firms Investing in AI

AI Spending – Total & per Worker

Large companies (>= 500 workers) will spend more total dollars on AI in 2026 (left chart). Small companies expect to spend slightly more per employee in 2026 (right chart). **Shallow**



Total AI Spending



AI Spending <u>per Employee</u>

AI Spending – by Spending Category

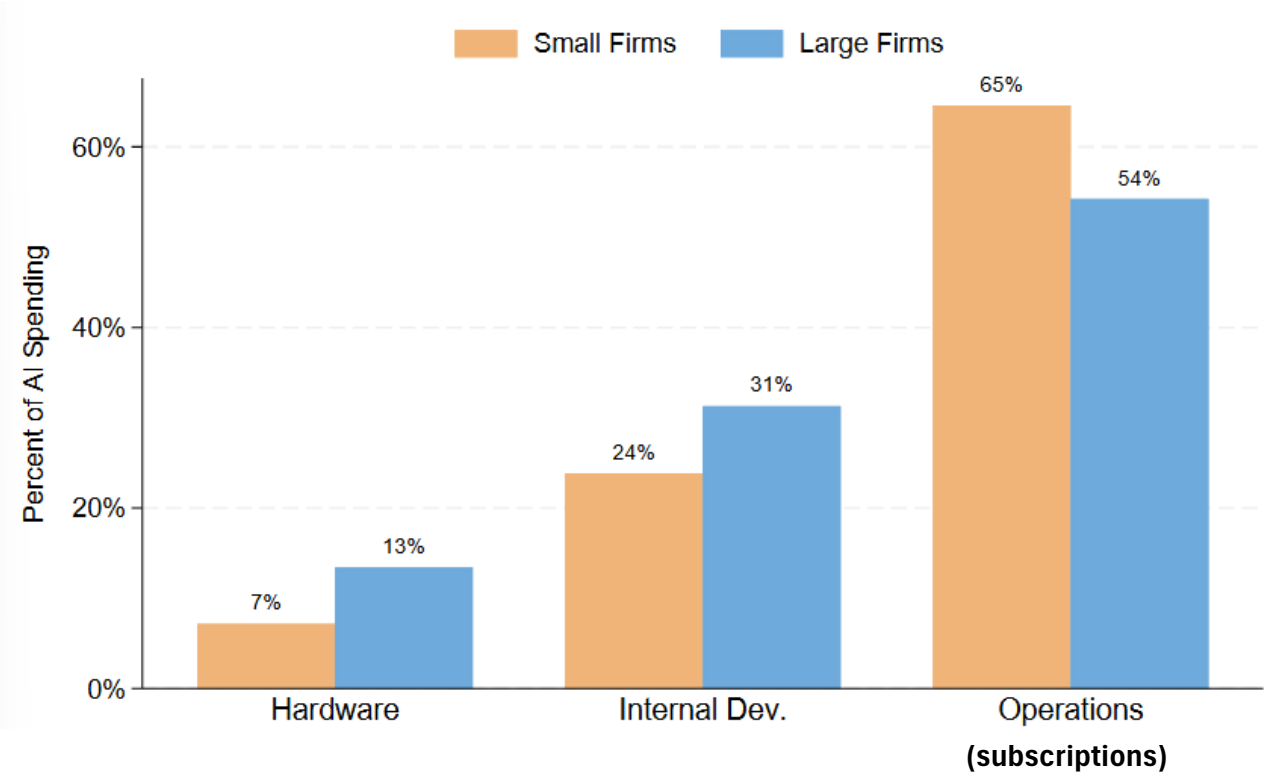
Typical company **primarily spends on AI via "subscription"**

- Large firms more likely to spend on hardware or internal development

In comparison to the "computer revolution" of the 1990s, the AI investment for a typical firm today is **less capital-intensive** (less hardware)

Spending categories:

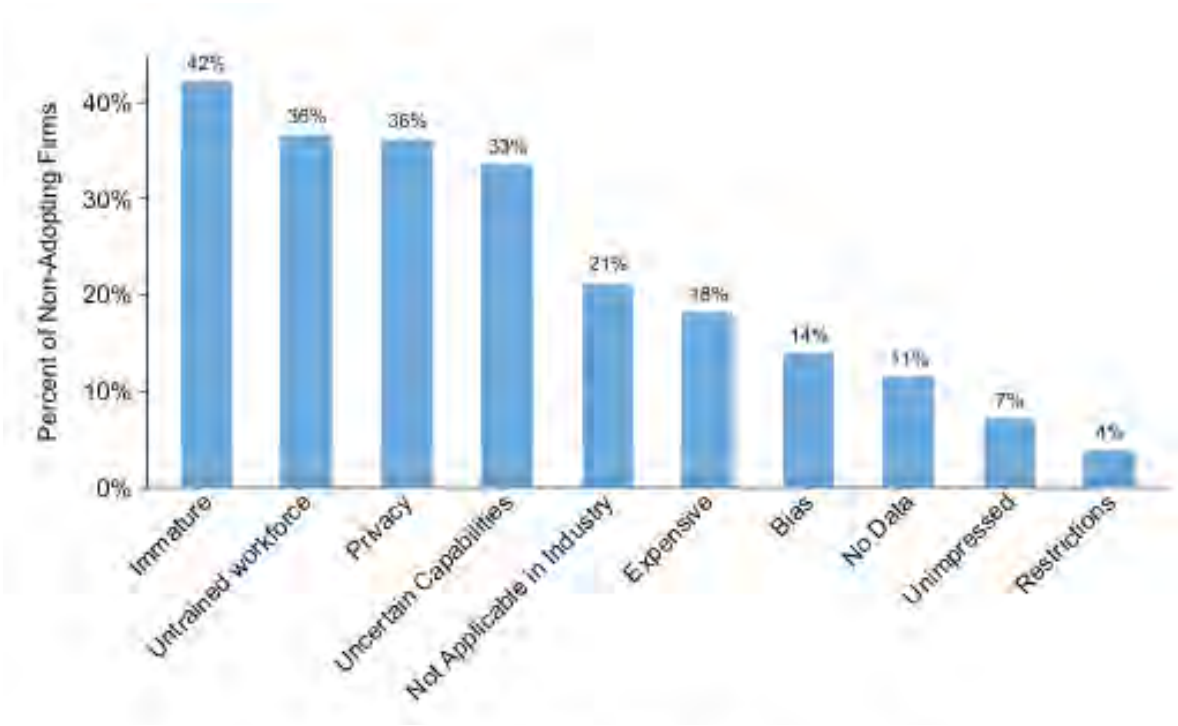
- Hardware for AI (e.g. servers/GPUs, devices)
- Developing or customizing internal AI systems
- Operational expenses: AI subscriptions, services, and training (cloud-based tools, software fees, consultants, employee training, data prep)



Why Did Firms Not Adopt AI?

Among companies that did not invest in AI in 2025, most common reasons were:

- AI technology is immature
- Workforce not trained to use AI
- Privacy concerns
- Unsure of AI capabilities



Reasons for Not Investing in AI

What Was Your Company's Motivation to Adopt AI?

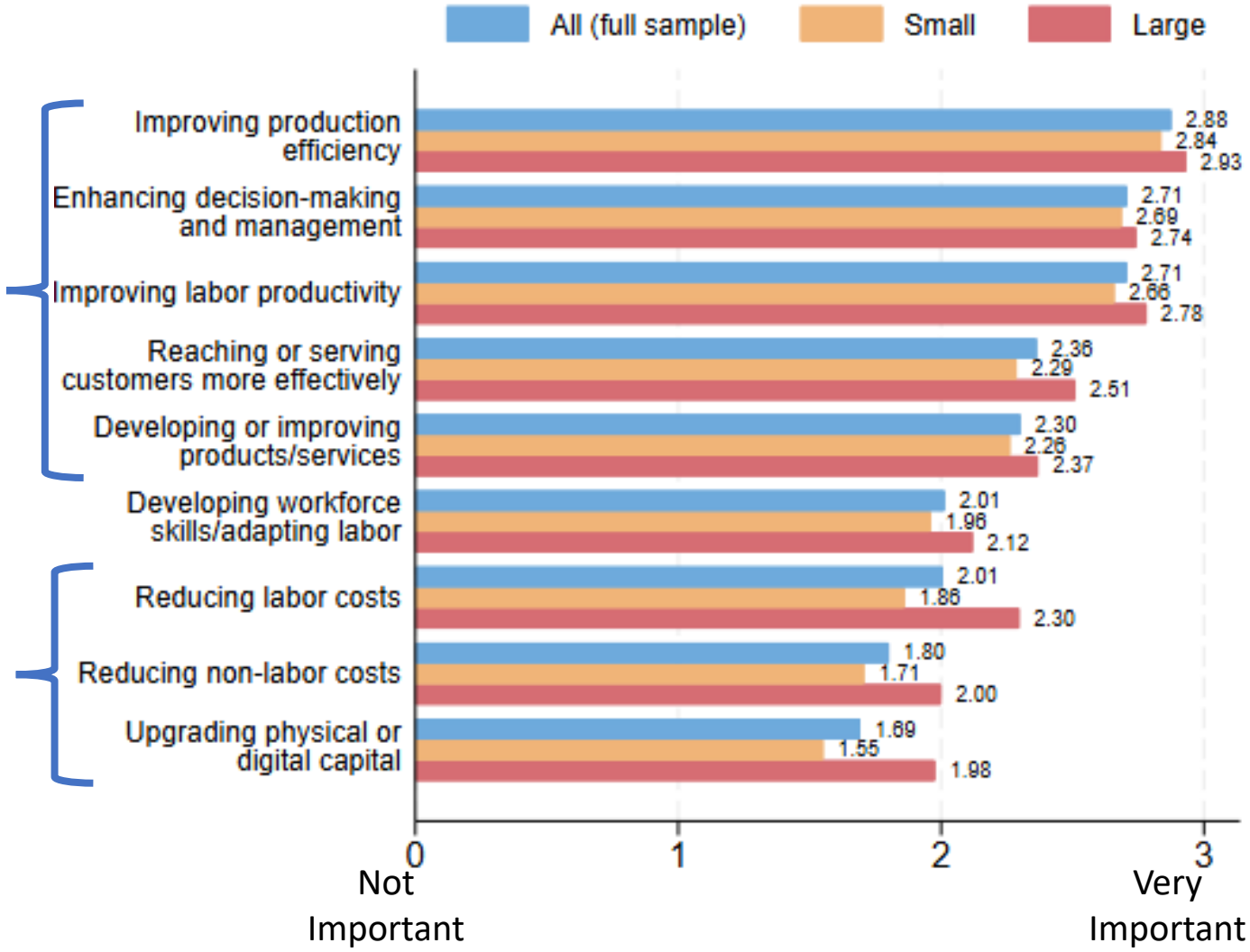
Motivations for investing in AI are shown for full sample, small firms, and large firms.

Popular motivations

- **Improve production efficiency**
- **Enhance decision-making**
- **Improve productivity**

Less popular motivations

- **Reducing costs**
- **Upgrading physical capital**



Outline

1. AI Adoption, Spending, and Motivation

2. **The Broad Effects of Using AI**
    a. Efficiency, Productivity
        i. NOT cost reduction; NOT large employment reduction
    b. AI is useful for tasks with pattern recognition and text generation
        i. NOT high stakes; NOT case-specific judgment

3. Effects of Using AI – Productivity

4. Effects of Using AI – Workforce, Jobs, Tasks

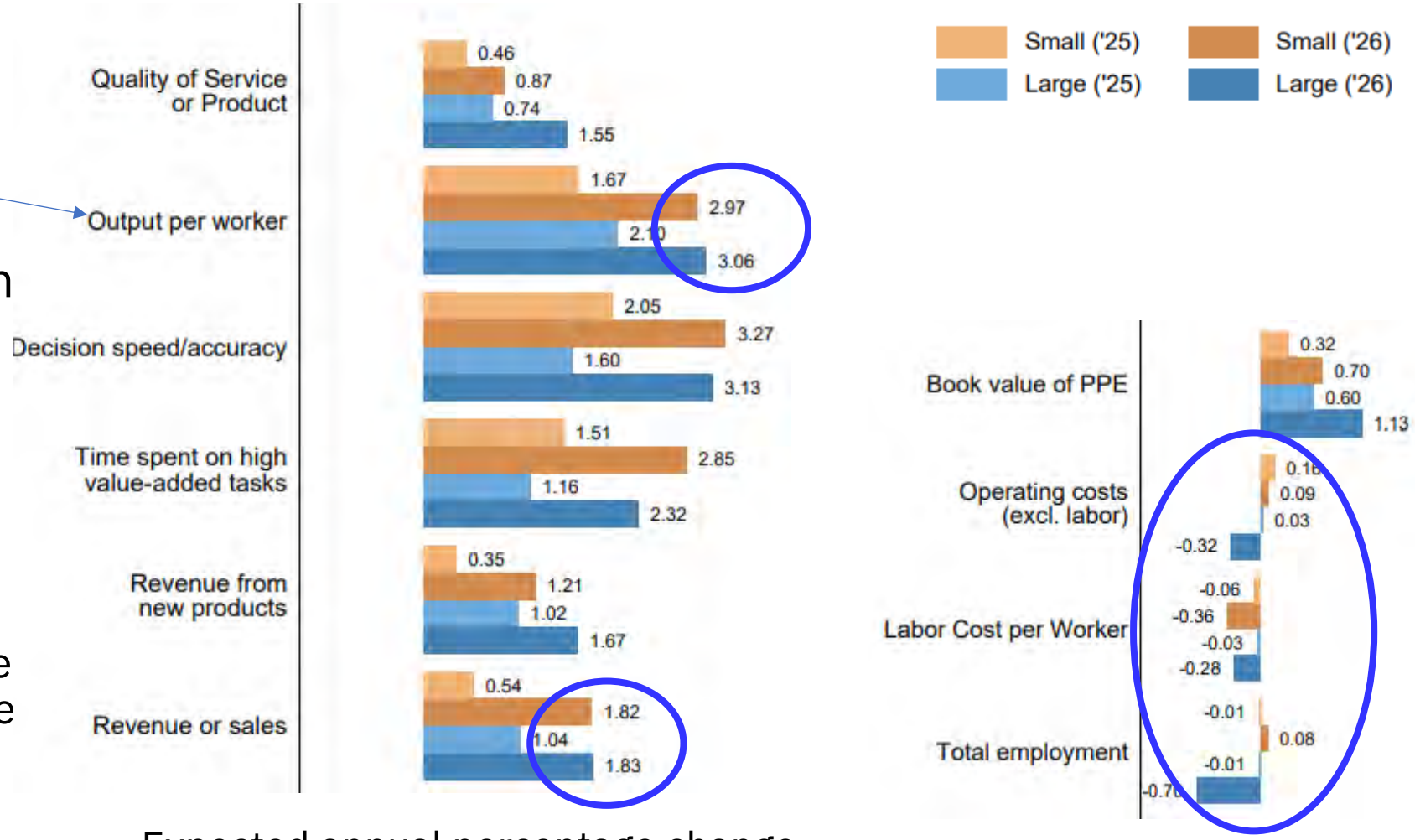
5. Takeaways and Future Research

The Broad Effects of Using AI

In 2026, *due to AI*, CFOs expect output per worker to increase by 2.97% at small companies and by 3.06% at large companies – and decision speed/accuracy to increase by more than 3% overall.

Firms expect small changes in costs and total employment.

- Large companies expect to reduce employment by 0.7% in 2026 (see bottom right corner).

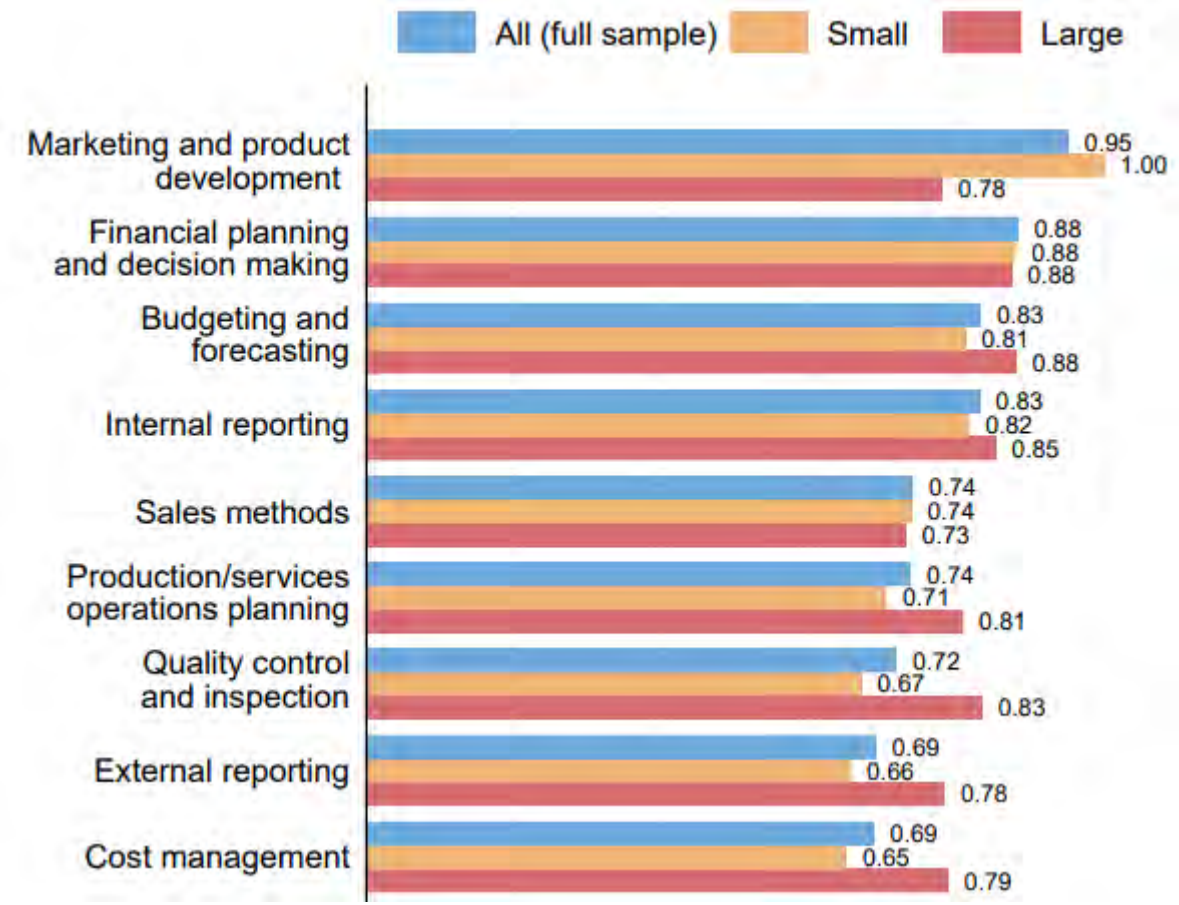


Expected annual percentage change

How Helpful are AI Tools for Various Tasks?

CFOs report that AI has been helpful for tasks with **routine**, **pattern recognition, and text generation**

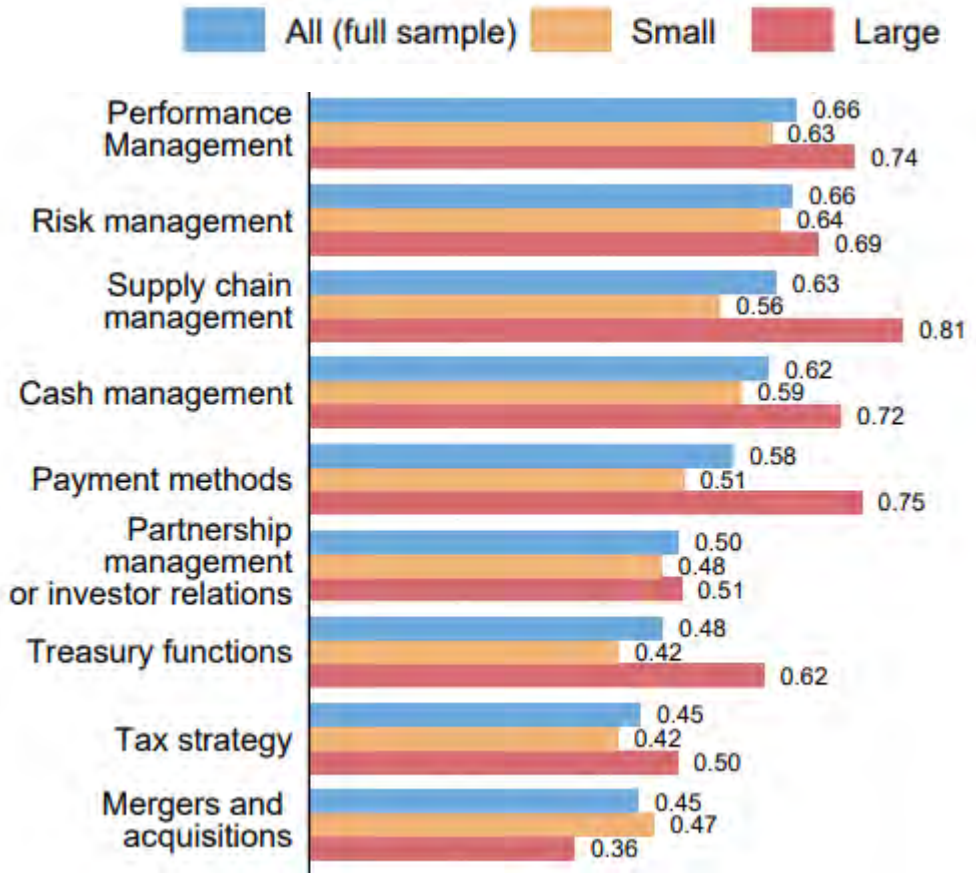
(Continued on next page)



0 = Not helpful, 1 = Moderately helpful, 2 = Extremely helpful

How Helpful are AI Tools for Various Tasks? (cont'd)

CFOs report that AI has **NOT** been very helpful in tasks that are **high-stakes, case-specific judgment.**



0 = Not helpful, 1 = Moderately helpful, 2 = Extremely helpful

Outline

1. AI Adoption, Spending, and Motivation

2. The Broad Effects of Using AI

3. **Effects of Using AI – Productivity**
   a. Productivity increase. Why? demand channel, innovation
      i. NOT capital deepening *at typical firm* (productivity increases not tied to increasing capital)
   b. Productivity Paradox (reported productivity effects > calculated productivity)
      i. Likely explanation: delay in AI-driven revenue growth

4. Effects of Using AI – Workforce, Jobs, Tasks

5. Takeaways and Future Research

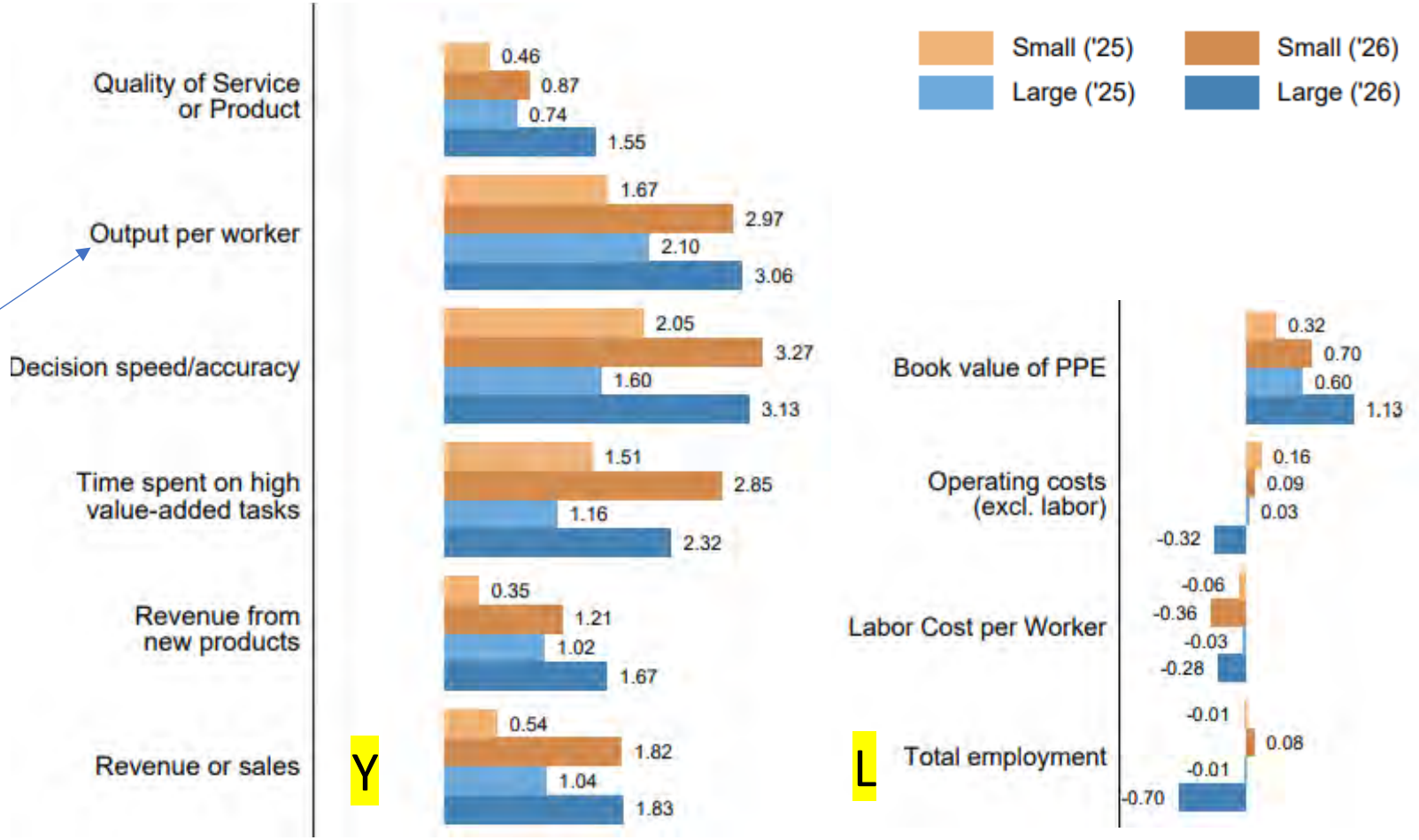
Measuring Productivity

The survey has two measures of the effects of AI on productivity (output per worker).

1) **Reported productivity**, based on this survey response.

2) **"Implied" productivity**, based the change in revenues (Y) relative to change in employment (L).

- That is, change in Y/L



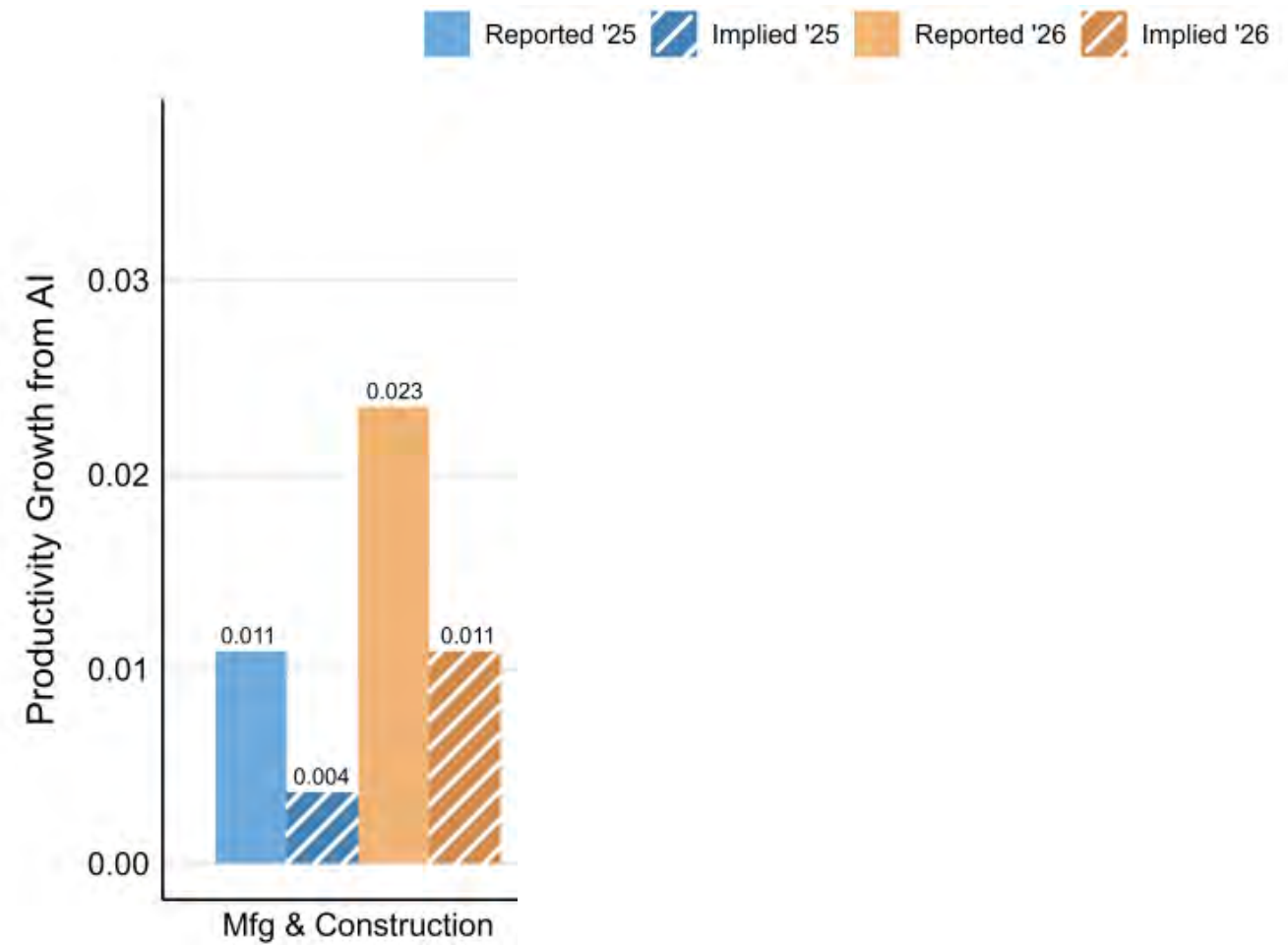
Effects of AI on Productivity – Key Takeaways

1. Productivity improvements expected to increase ‘25 → ‘26

2. Greater productivity improvement in finance, high-skill industries

3. **Productivity paradox**: In both 2025 and 2026, Reported > Implied productivity effects.
   With transformative technologies, there is often a delay in when effects appear in output (revenues).

4. Reported 2025 approx. equals Implied 2026.
   Effects of AI-induced productivity improvement expected to **increase sales with a 1-year delay**.



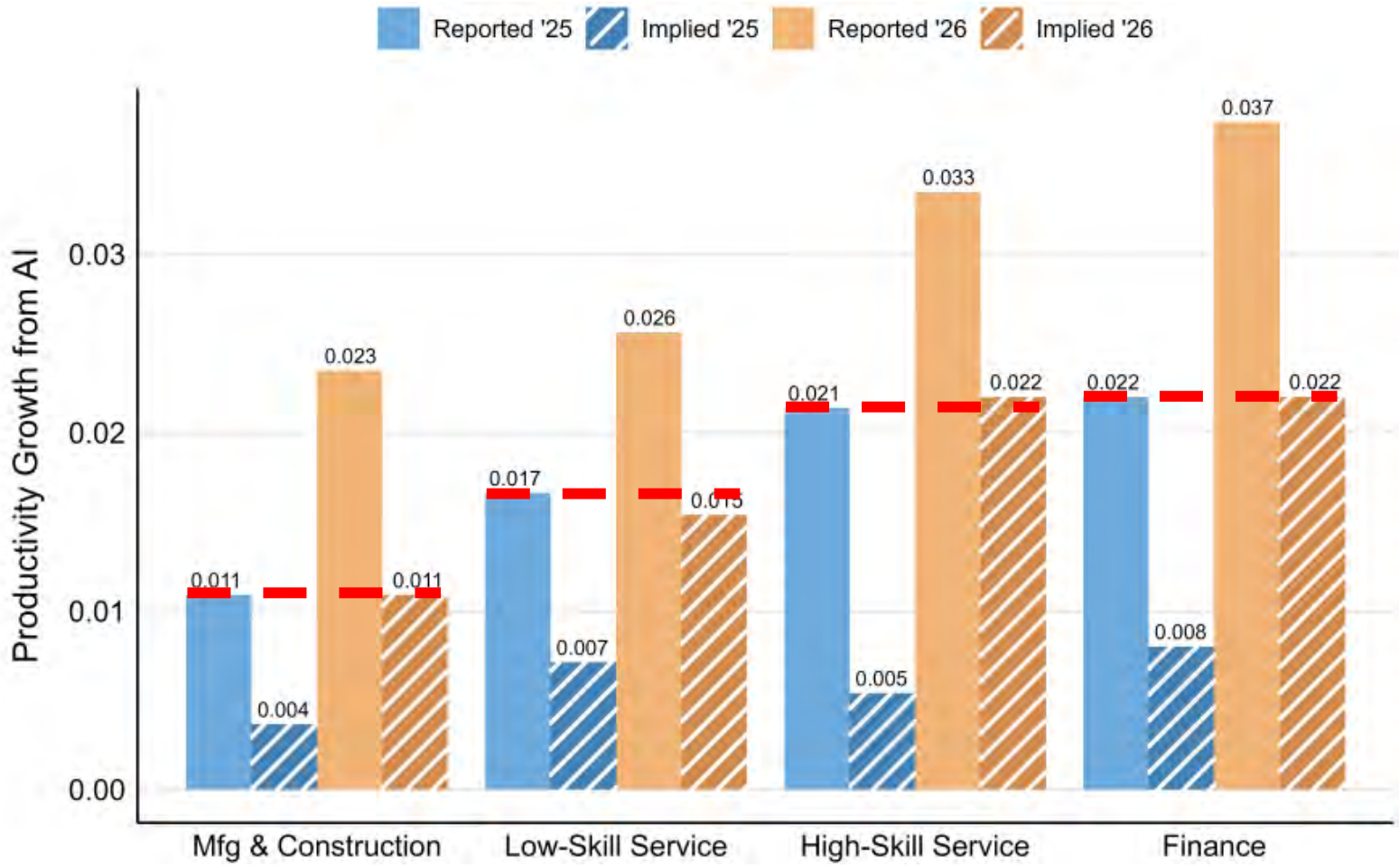
Effects of AI on Productivity – Key Takeaways

1. Productivity improvements expected to increase ‘25 → ‘26

2. Greater productivity improvement in finance, high-skill industries

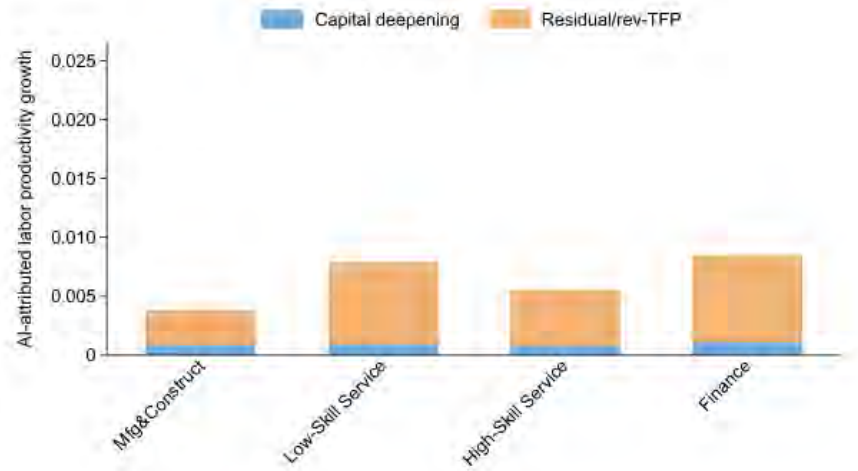
3. **Productivity paradox**: In both 2025 and 2026, Reported > Implied productivity effects.
   With transformative technologies, there is often a delay in when effects appear in output (revenues).

4. Reported 2025 approx. equals Implied 2026.
   Effects of AI-induced productivity improvement expected to **increase sales with a 1-year delay**.

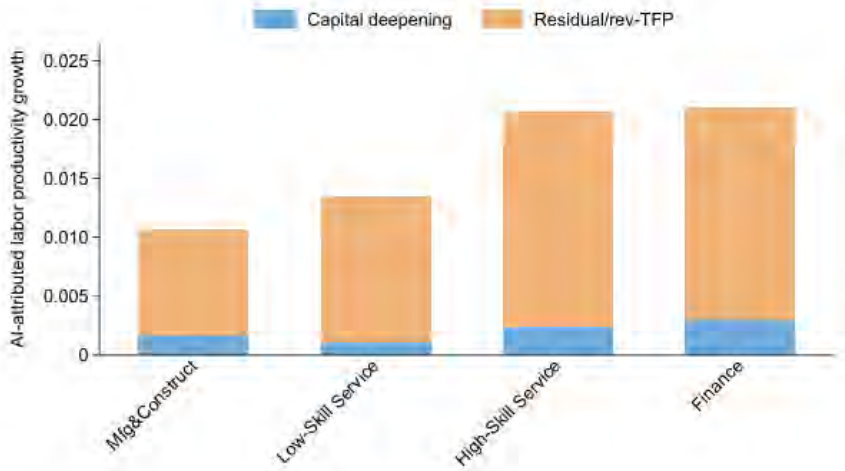


"Why?"
At the Typical Firm, Is Productivity Improvement Due to Capital Investment?

- **In previous IT surges, such as the computer revolution, productivity gains were tied closely to how much physical capital** a company would purchase (e.g., purchase a computer or server). We use a Cobb–Douglas growth-accounting decomposition (Jorgenson et al., 2008).
- With AI, because a company can use AI via a subscription (without purchasing hardware), it is possible that improvements in AI-driven productivity will not be tied to what economists call "**Capital Deepening**" (increase in the capital/labor ratio).
- Our analysis confirms this: At the typical company, very little of the improvements in AI-driven productivity is driven by capital deepening (blue in graphs below).
- Instead, the improvements are tied to "everything else;" that is, the residual or "**total factor productivity**" (orange in graphs below). **TFP** captures improved production efficiency, product or service quality, pricing power, quality or effectiveness of intermediate inputs and labor due to AI adoption.
    - Specifically, the improvements in TFP productivity reflect **innovation, product development, and better serving customers and demand**.



(a) 2025



(b) 2026

Outline

1. AI Adoption, Spending, and Motivation

2. The Broad Effects of Using AI

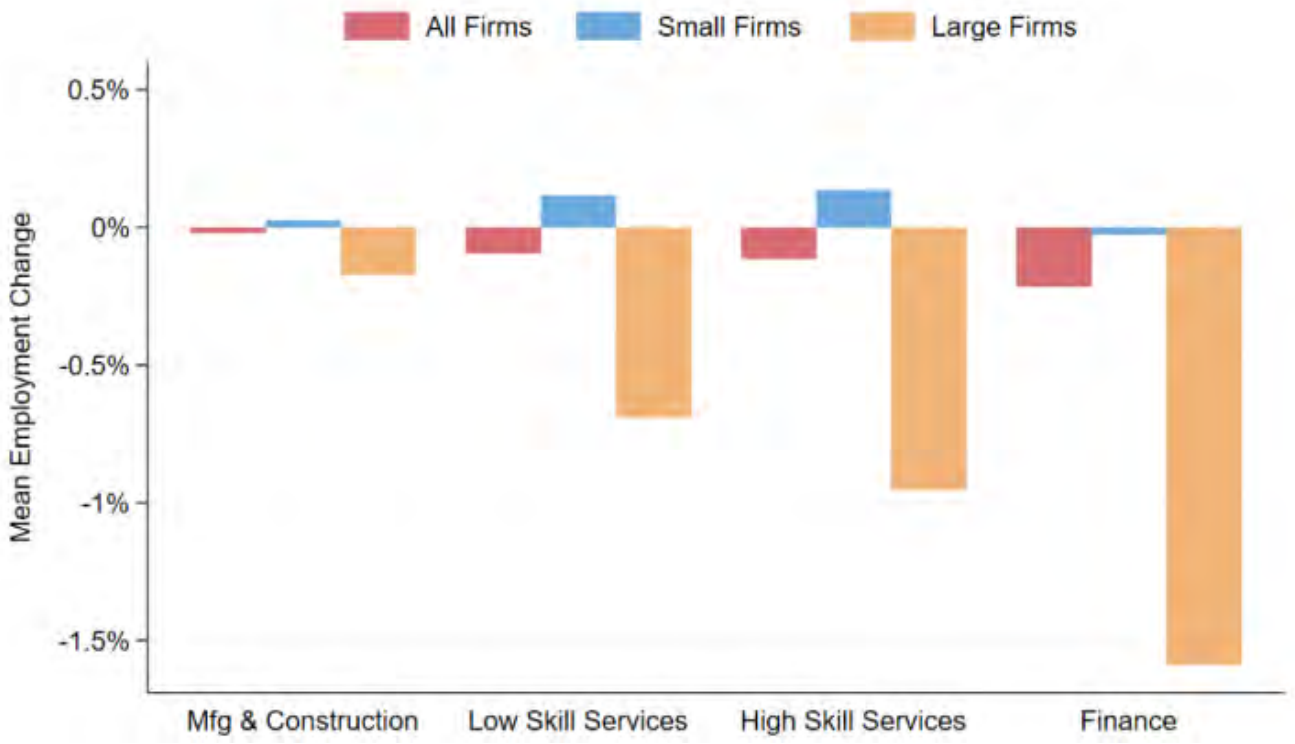
3. Effects of Using AI – Productivity

4. **Effects of Using AI – Workforce, Jobs, Tasks**
   a. Modest Aggregate Effects (primarily large firms; in finance and high-skill services)
   b. Routine/Clerical: Large firms *reduce* (small firms hold steady)
   c. Technical: Small firms *increase* (large firms hold steady)
   d. Exposure Index ranks specific jobs most enhanced and most likely to be replaced by AI

5. Takeaways and Future Research

Modest Aggregate Employment Effects

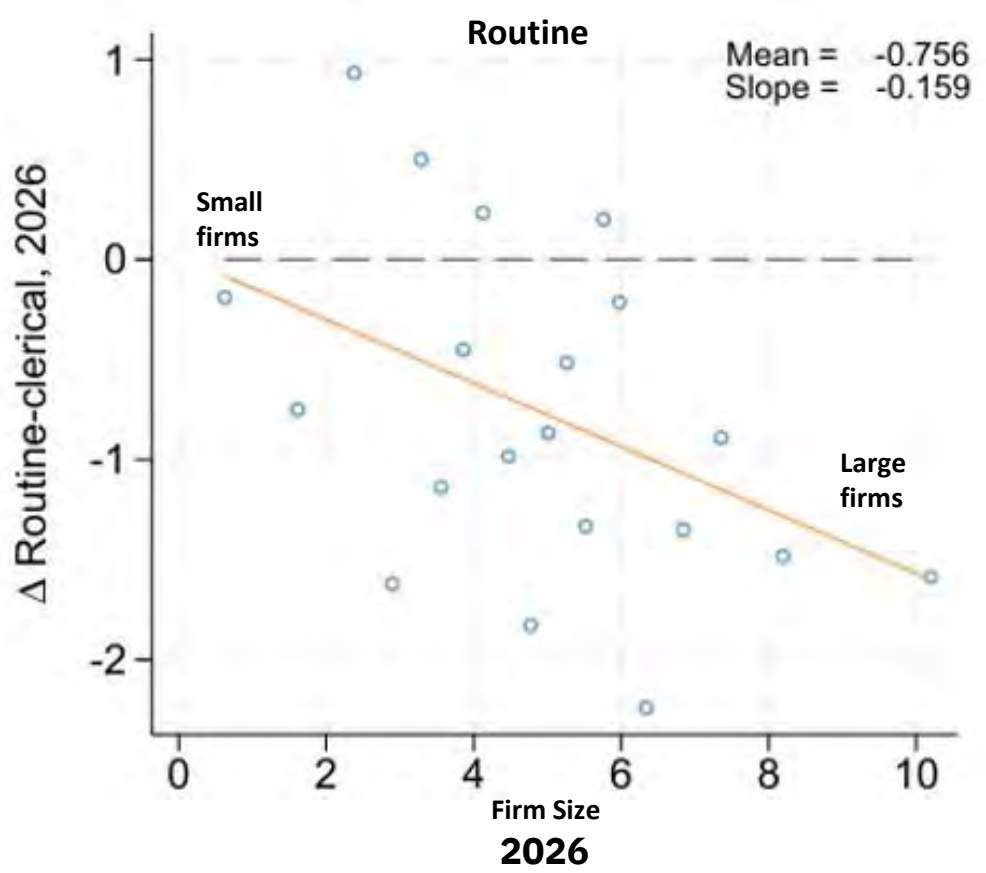
- In an open-ended question that asked “Which jobs/roles and responsibilities will be replaced by AI?”, **half of the respondents wrote in “no jobs replaced”** in the near-term.
- Overall, CFOs state expect **500,000 jobs to be lost in 2026** due to the adoption of AI (0.4% of workforce)

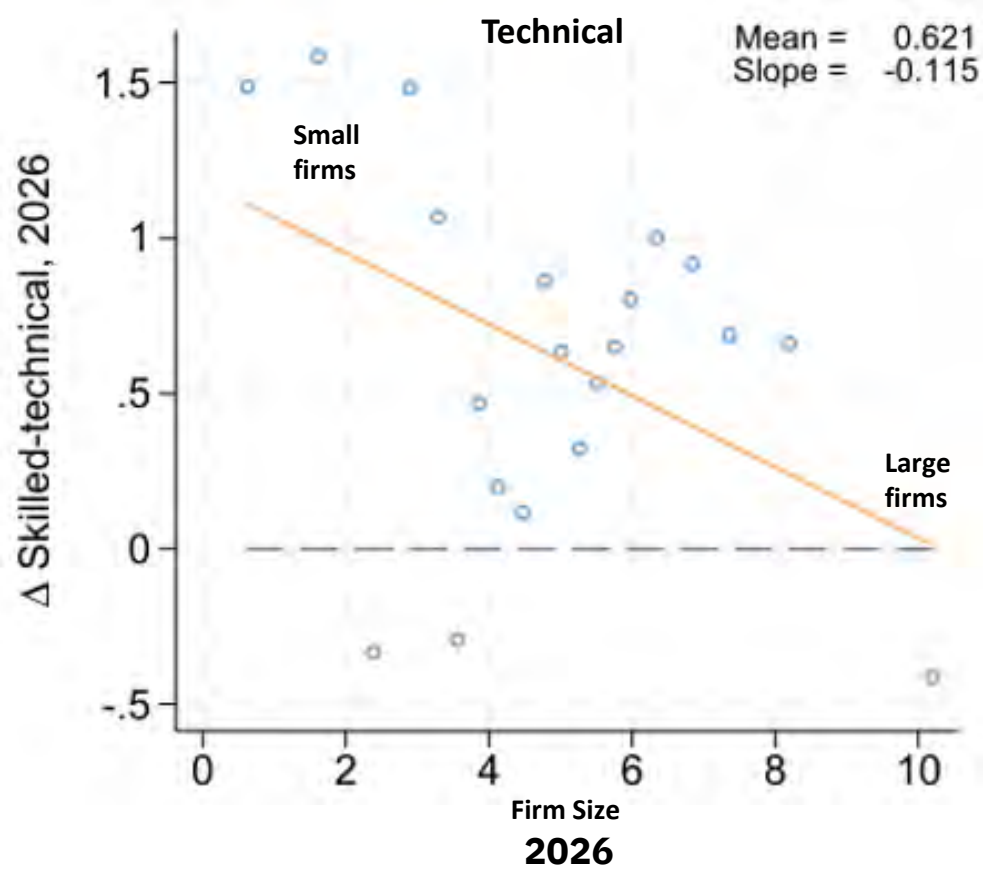


Composition: Expected Change in Types of Jobs (Conditional on Firm Size)

Small firms expect to increase number of *technical* employees due to using AI (right graph); large companies expect to reduce number of *routine* employees due to using AI (left graph).

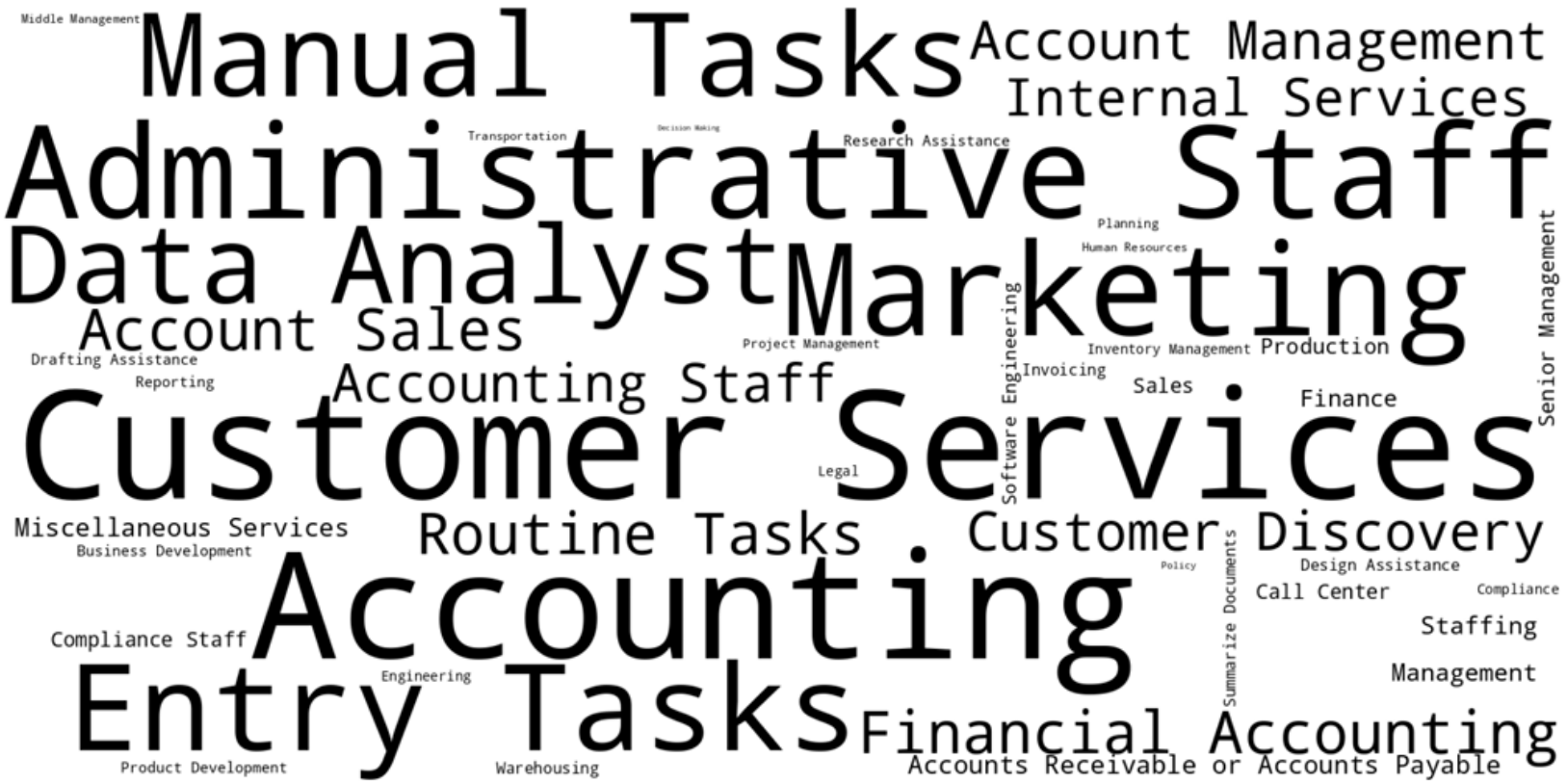
Across the economy, this would imply a reallocation of jobs from routine to technical roles.





Jobs/Roles Replaced by AI

Based on open-ended responses, the jobs/roles that CFOs most often mention as being replaced by AI including administrative staff, customer services, and accounting.



Jobs/Roles Enhanced by AI

Based on open-ended responses, the jobs/roles that CFOs most often mention as being enhanced/complemented by AI are marketing, accounting, and finance.



Jobs Most Affected by AI (Negative Exposure Index)

Based on open-ended questions about whether 1) AI enhances jobs, 2) AI will replace jobs, or 3) both

- **Negative Exposure Index (NEI)** is a simple ratio of Replace/Enhance (RM/EM)

Occupation	Job Titles	NEI (RM/EM)
Office and Administrative Support	Bookkeeping, Accounting, and Auditing Clerks; Office Clerks Customer Service Representatives;	2.025
Business and Financial	Accountants, and Auditors; Financial Analysts; Management Analysts;	0.829
Computer and Information Technology	Software Developers, Quality Assurance Analysts, & Testers; Computer Programmers; Computer Systems Analysts;	0.596
Legal	Lawyers; Paralegals& Legal Assistants;	0.471
Production	Assemblers & Fabricators; Metal & Plastic Machine Workers; Quality Control Inspectors; Machinists;	0.308
Sales	Advertising Sales Agents; Wholesale & Manufacturing Sales Representatives; Sales Engineers;	0.298
Management	Top Executives; Financial Managers; Advertising, Promotions, and Marketing Managers;	0.138
Architecture and Engineering	Civil/Mechanical/Industrial Engineers; Architectural & Engineering Managers; Engineering Technologists & Technicians;	0.100

Index value of 2.0 means twice as many replaced mentions as enhanced

Index value of 0.1 means ten times as many mentions of enhanced with AI, relative to replaced by AI

Sectors Most Affected by AI (Negative Exposure Index)

Based on open-ended questions about whether 1) AI enhances jobs, 2) AI will replace jobs, or 3) both

- **Negative Exposure Index (NEI)** is a simple ratio of Replace/Enhance (RM/EM)

Table 7: Firm Exposure and Negative Exposure Index by Sector

Sector	Share of Firms Exposed	NEI (RM/EM)
Information	0.704	0.929
Transportation and Warehousing	0.677	0.765
Professional and Business Services	0.636	0.731
Health Care and Social Assistance	0.625	0.688
Other Services Except Government	0.622	1.308
Construction	0.609	0.520
Leisure and Hospitality	0.609	1.000
Real Estate and Rental and Leasing	0.603	1.227
Manufacturing	0.596	0.580
Finance and Insurance	0.582	1.100
Retail and Wholesale Trade	0.575	0.758
Mining and Utilities	0.524	0.800
Educational Services	0.409	0.714

NEI value of 1.0 means same number of jobs in this sector are expected to be replaced as enhanced by AI

- >1.0 means more jobs replaced
- <1.0 means more jobs enhanced

Middle Column: Share of Firms Exposed indicates the percentage of firms in a sector with jobs that were mentioned as being enhanced by AI, replaced by AI, or both

Outline

1. AI Adoption, Spending, and Motivation
2. The Broad Effects of Using AI
3. Effects of Using AI – Productivity
4. Effects of Using AI – Workforce, Jobs, Tasks
5. **Takeaways and Future Research**

Comparing AI to Previous Technology Waves

- **Similarities**
  - Adoption uneven, led by larger companies
  - Productivity Paradox (measured productivity gains lag perceived gains from new tech)
    - Likely reflects time needed to reorganize workflows and build complementary skills
  - Early labor market adjustments concentrated in routine clerical work

- **Differences**
  - Initial AI spending is broad but shallow (small $'s; subscriptions)
  - AI adoption does not require capital deepening *at typical firm*
    - AI: small number of firms making capital investment for whole economy
  - Initial AI productivity tied to innovation and demand channels, at this stage more so than cost-cutting (which was part of automation/robotics)
  - Employment: AI service industries affected most; Automation wave: manufacturing

Table 1 (outline)

Table 1: Initial Stages of AI Adoption, Workforce, and Productivity

	Small Firms	Large Firms	Evidence
Panel A: Initial Stages of AI Adoption			
Adoption	Moderate in 2025, growing sharply in 2026.	Already widespread in 2025; expected even higher in 2026.	Figures A4, A6
Why not investing in AI?	Top reasons (similar for small and large firms): immature technology, untrained workforce, privacy concerns, and uncertain AI capabilities.		Figures A4, A5
Dollar investment	Median expected 2026 spending: $12,500; most companies plan to spend $20,000 or less.	Median expected 2026 spending: $300,000; about 30% plan to spend over $1 million.	Table A1; Figure A7
Investment intensity	Median per employee AI spending in 2026: $357.	Median per employee AI spending in 2026: $258.	Figures A8, A9
Type of spending	Operational expenses (subscription, services) dominate (65%).	Operational expenses dominate (54%); more internal development/hardware.	Figure A10
Use cases	Highest for marketing/product development, planning, forecasting, and reporting.	Similar top uses, with some edge in supply chain, quality control, and payment methods.	Figure A14
Takeaways: AI adoption	AI investment is broad but shallow: many firms are adopting or planning to adopt, but typical spending remains modest and largely subscription-based, especially among smaller firms.		
Panel B: Initial Stage AI Effects on Workforce and Productivity			
Number of Employees	Small increase	Modest decrease	Figures 2, A13
Workforce Composition	Reduction in routine-clerical (bigger for large firms); Increase in skilled-technical (bigger for small firms)		Figures 3, A12
Implied Productivity	1.62% increase in 2026	2.41% increase in 2026	Figure 4
Productivity Mechanism	Product innovation; reach/serve customers	Product innovation; cost reduction	Figure 4; Table A8

Takeaways and Future Research

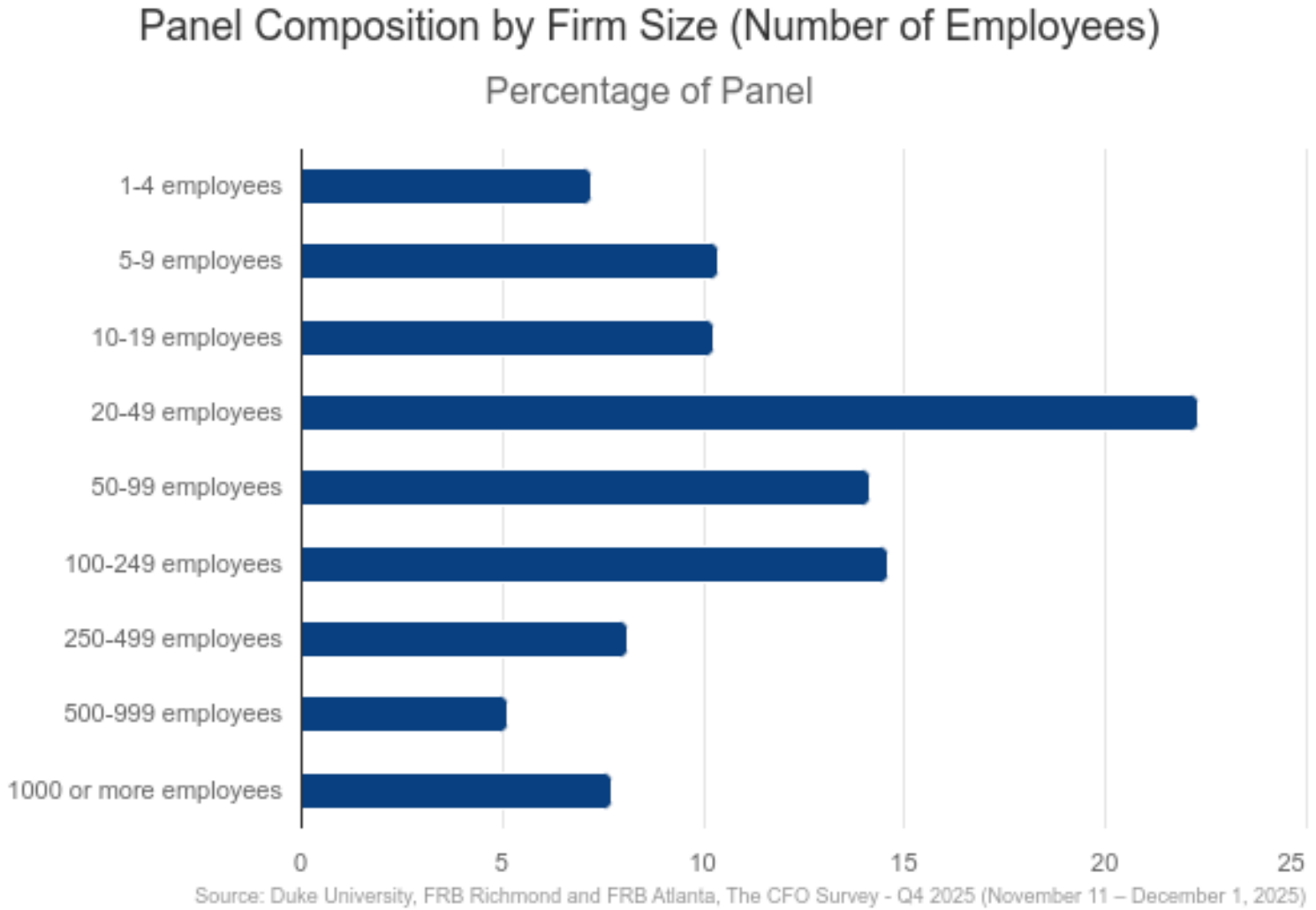
- **Adoption**
  - Widespread but shallow

- **Productivity**
  - Increase in productivity due to using AI
  - Productivity Paradox (reported > implied productivity – delayed effect of AI on revenue/worker)
  - Largest gains expected via innovation and better serving demand

- **Workforce**
  - Modest aggregate effects; Large firms expect to reduce employment
  - Composition: shift from routine (large firms -) to technical jobs (small firms +)
  - Biggest effects in routine accounting, clerk, customer service jobs

- **Considerations and Future Research**
  - AI is fast changing -- should repeat this survey again very soon!
  - We focus on the short-term, given difficulty anticipating long-term. Need to repeat the survey regularly to monitor ever-changing effects of AI!
  - We focus on typical firm usage of AI. More research needed to also consider important effects of companies that develop AI technologies and build data centers.
  - We find evidence that AI will increase innovation (and hence growth). More research is needed on the possibility that AI will create new, job-creating firms that are not currently in our survey.

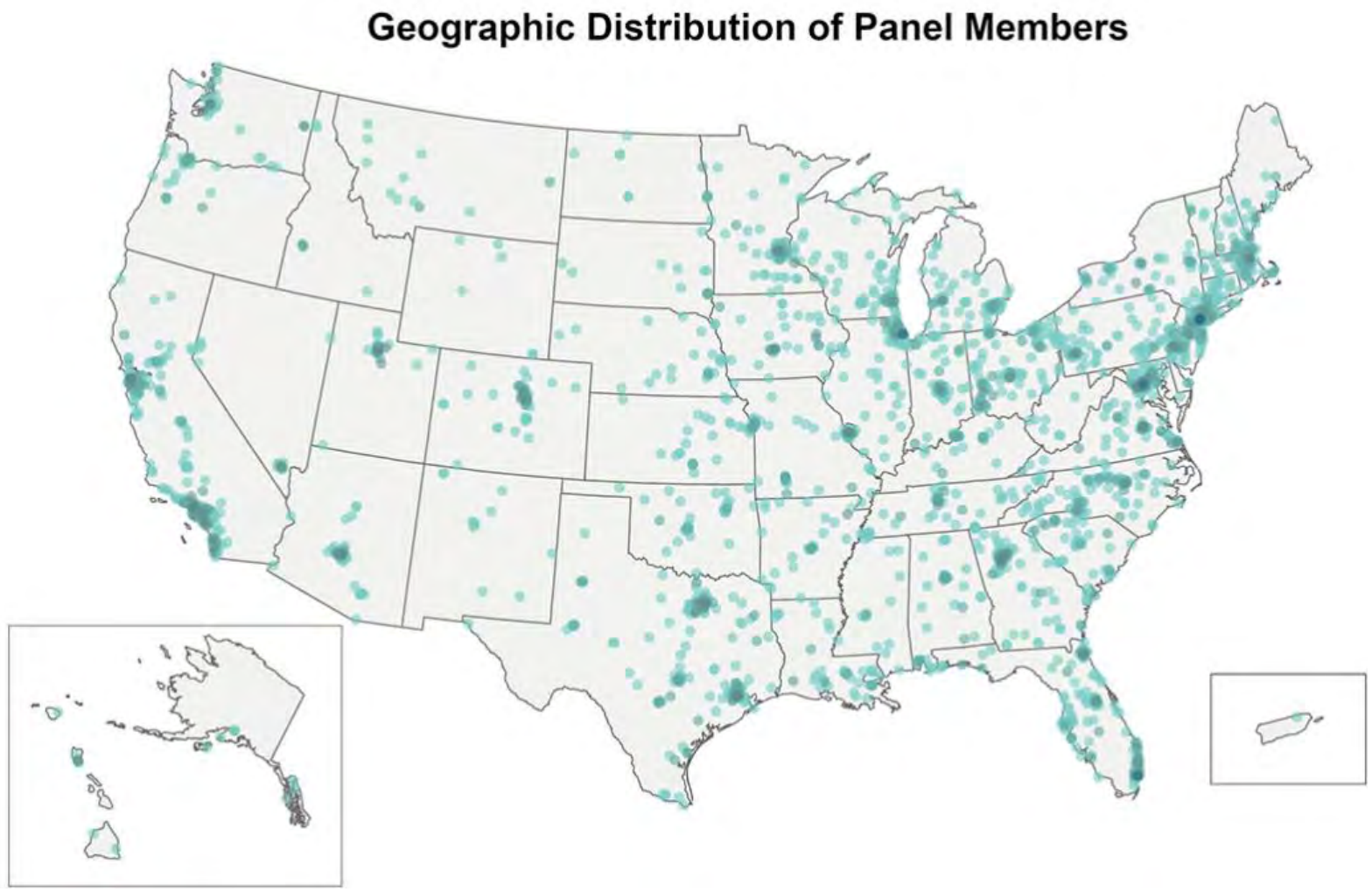
Who Participates in the CFO Survey?



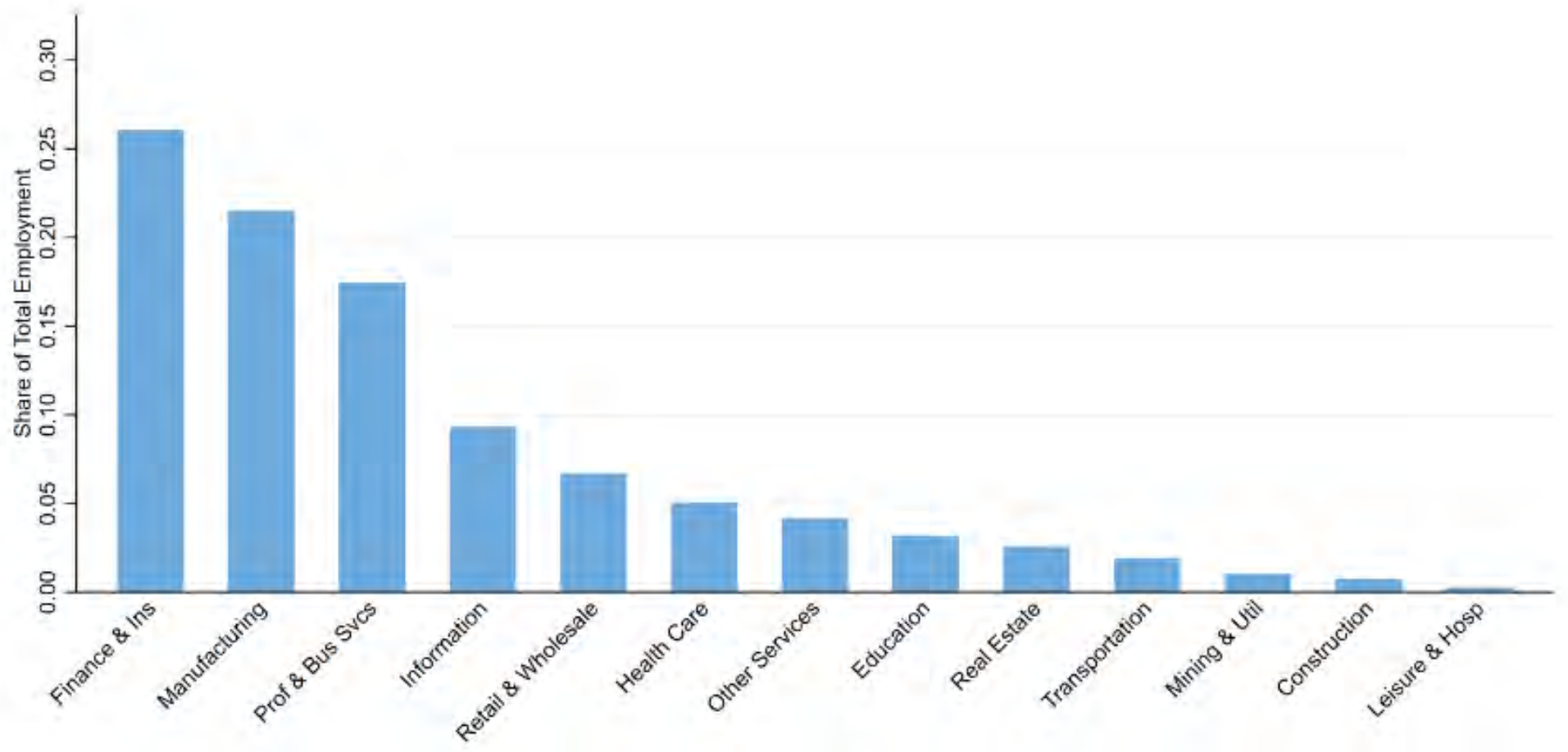
Note: this graph shows the distribution of number of employees for the full CFO Survey panel. For the AI survey project, 22% of responding firms have at least 500 employees.

Note: In the statistical analysis, a large firm is defined as one with at least 500 employees.

Who Participates in the CFO Survey?



Who Participates in the CFO Survey?



Notes: This figure shows share of total employment by sector among firms in the full AI survey sample.


PIIE
PETERSON INSTITUTE FOR
INTERNATIONAL ECONOMICS
OECD
INTERNATIONAL MONETARY FUND
WORLD BANK GROUP
www.piie.com
@PIIE
linkedin.com/company/piie

Emerging Trends in Global Technology Diffusion: Is Artificial Intelligence Different?

Paulo Bastos[1], Katherine Stapleton[1], Daria Taglioni[1], Hannah Yi Wei[2]

[1]The World Bank

[2]International Monetary Fund

November 18, 2026

The views expressed are those of the authors and do not necessarily represent the views of the World Bank or the IMF, their Executive Boards, or their management.

Roadmap

1. Motivation
2. Data
3. AI hiring at a glance
4. Stylized facts about AI diffusion
5. AI diffusion through multinational supply chains
6. Conclusion

Motivation

Why it matters

- AI is widely seen as the defining technology of our time – yet how it spreads is poorly understood (Agrawal et al., 2019; McElheran et al., 2023)
- Earlier frontier technologies spread through proximity and firm networks (Comin and Mestieri, 2013; Bloom et al., 2021; Bastos et al., 2026)
- Policy leans on these channels: value chain integration to catch up; export controls to limit competitors' access

This paper asks

How does AI diffuse differently from other technologies?

This paper

Studies how AI diffuses compared with 28 other frontier technologies – through firms' hiring for AI skills

Built on granular micro data

573 million job postings, 17 advanced economies, 2014–22; ∼100k firms matched to supply-chain data

▸ Details

Finds that, relative to other technologies, AI-related hiring is:

1. **More dispersed**: less concentrated across countries and regions
2. **More tied to income**: richer economies tilt hiring more towards AI
3. **Shaped by distance within countries**, not across them
4. **Less transmitted through firm networks**, especially early on

AI seems to spread more through local labor markets than firm networks

Roadmap

1. Motivation
2. Data
3. AI hiring at a glance
4. Stylized facts about AI diffusion
5. AI diffusion through multinational supply chains
6. Conclusion

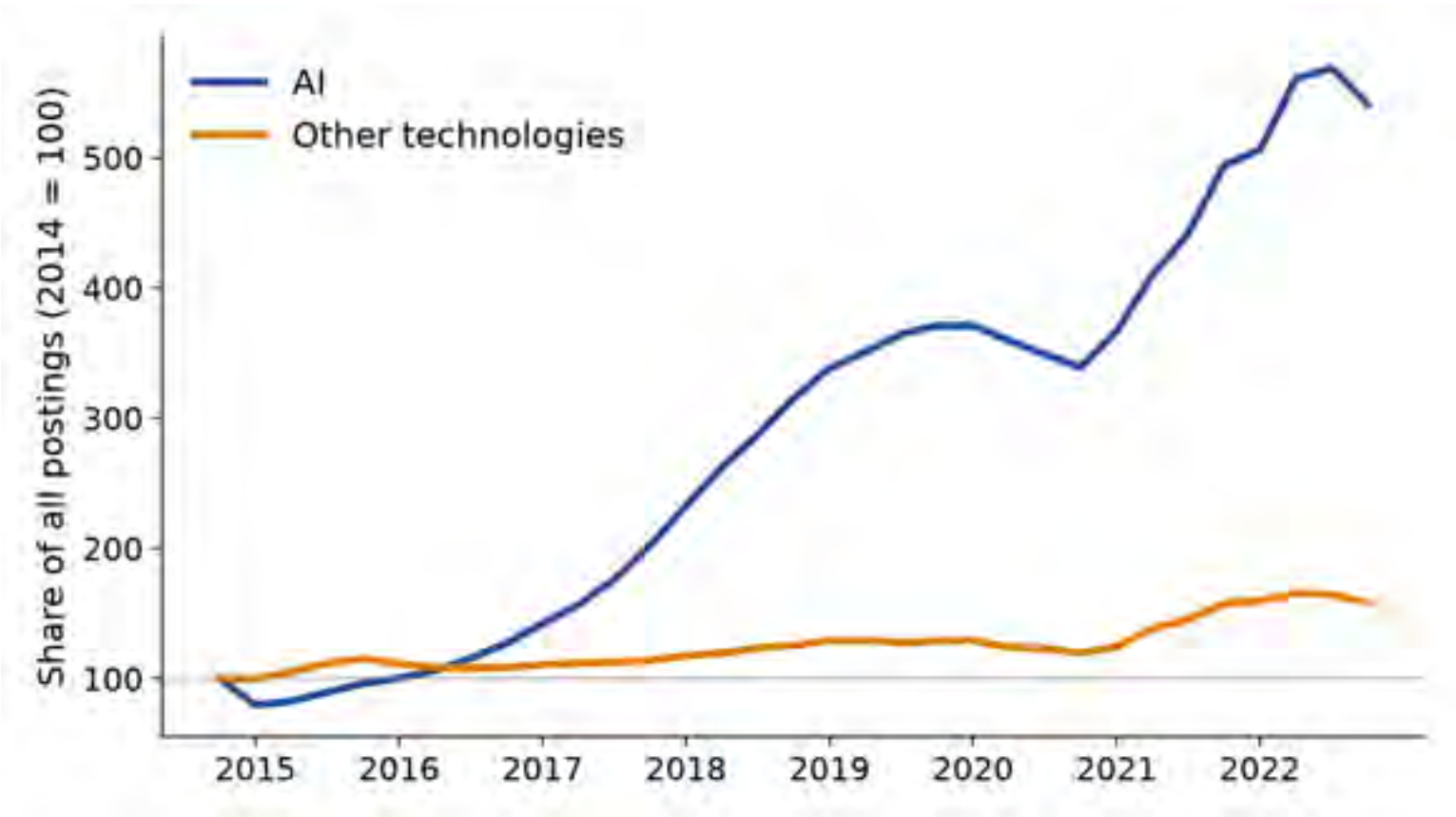
Measuring AI diffusion through firms' demand for skills

- **573 million job postings** (Lightcast) across 17 advanced economies (∼60% of world GDP), 2014–2022
  - Covers 85–90% of all vacancies in the US and UK (Hershbein and Kahn, 2018; Javorcik et al., 2019; Acemoglu et al., 2022)
- **29 technologies** from Bloom et al. (2021): a posting signals adoption if it mentions a technology's key phrases ▸ Details
  - Key phrases identified from patents and earnings calls
  - AI bigrams: e.g. "machine learning", "natural language", "neural network"
  - Searched in all local languages; data end in 2022 (pre-generative AI)
- **Firm linkages** from FactSet: 107,981 firms matched to their buyers, suppliers, partners and affiliates (Huang et al., 2023)

Roadmap

1. Motivation
2. Data
3. AI hiring at a glance
4. Stylized facts about AI diffusion
5. AI diffusion through multinational supply chains
6. Conclusion

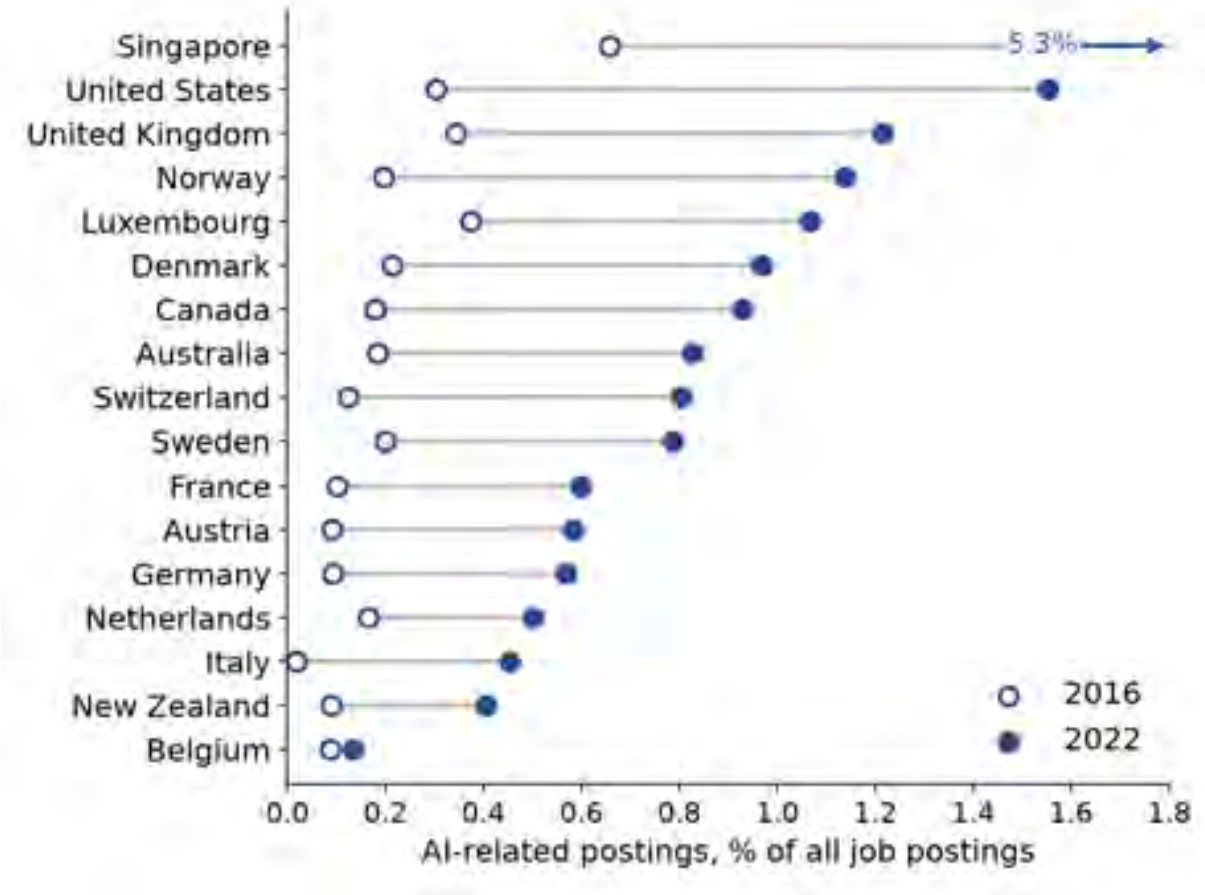
Demand for AI in job postings rose 5.4-fold in 8 years



- From 2014 to 2022: AI 5.4×, other technologies 1.6×
- 20.5% a year – the second fastest of 29 technologies

▸ Details

AI demand rose in every country – from very different levels



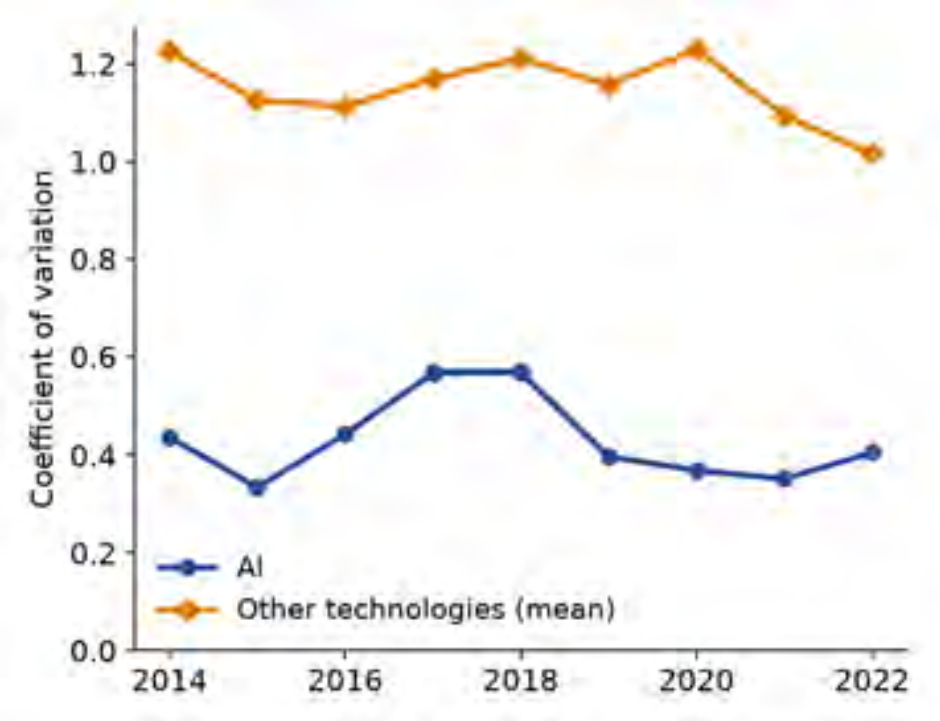
Notes: AI-related postings as % of all job postings, annual averages. Singapore's 2022 share is off the scale.

Roadmap

1. Motivation
2. Data
3. AI hiring at a glance
4. Stylized facts about AI diffusion
5. AI diffusion through multinational supply chains
6. Conclusion

Fact 1: AI demand is more evenly spread than other tech

Across 17 countries



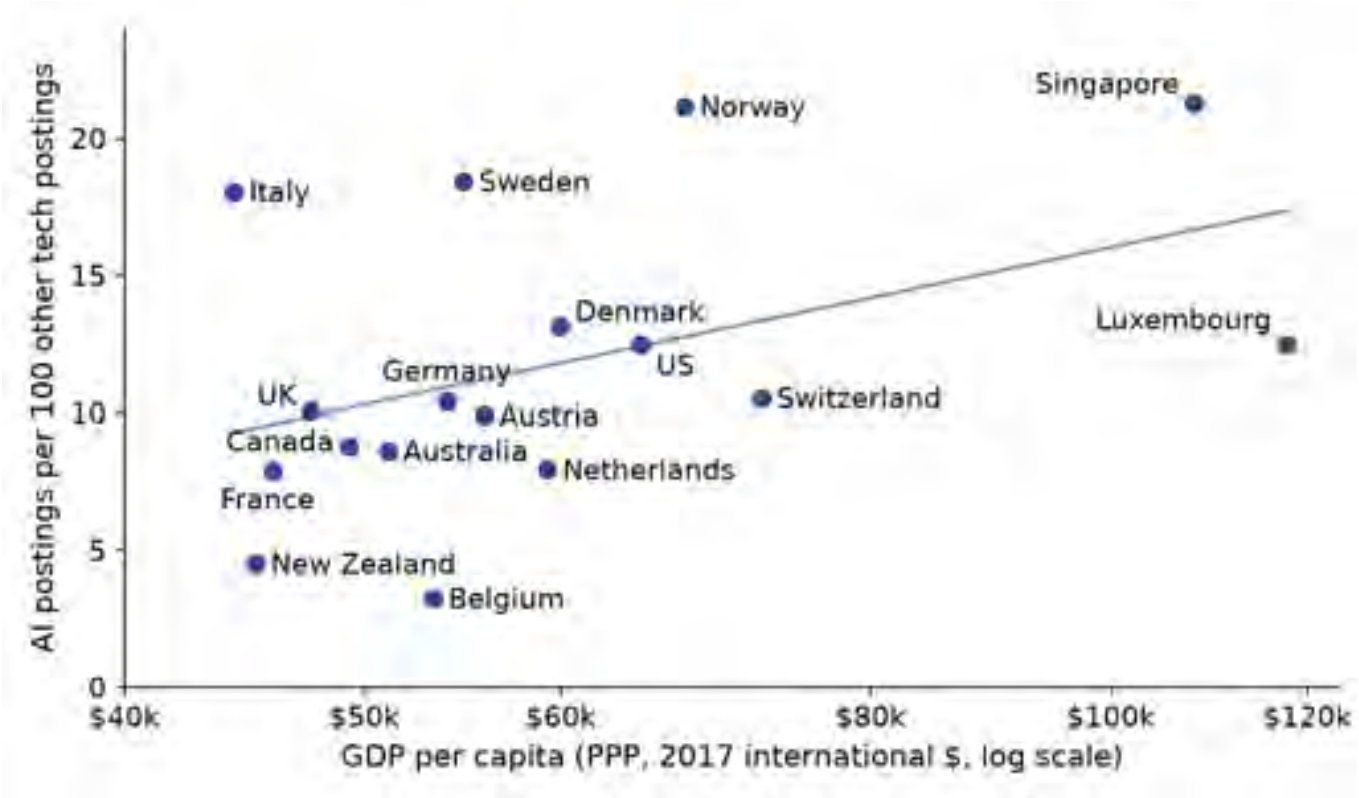
Across 156 regions



- Lower dispersion than the average technology in every year (coefficient of variation: lower = more even)
- Across regions, AI keeps spreading out over time

▸ Details

Fact 2: Richer countries tend to tilt tech hiring towards AI



- AI's share of tech hiring rises with income (correlation 0.44)
- Income matters more for AI than other tech (next page)
- Not driven by structure: controlling for finance (e.g. Luxembourg) or professional jobs, the correlation stays at about 0.4

Notes: All data for 2022.

Fact 3: Distance holds AI back within countries, not across

Definitions

Pioneers: regions that together hold half of a technology's citation-weighted patents in the ten years before it emerged (Bloom et al., 2021; Bastos et al., 2026)

- AI: mainly the San Francisco Bay Area

Distance: km from a country, state or firm to the nearest pioneer region

Specification

$$Y_{i,t}^{T} = \alpha + \beta_1 Dist_i^{T} + \beta_2 (Dist \times AI)_i^{T} + \gamma_1 Inc_{i,t} + \gamma_2 (Inc \times AI)_{i,t} + \mu_T + \mu_t + \epsilon_{i,t}^{T}$$

- Y: share of unit i's postings for technology T, relative to T's average share
- Units i: 17 countries, US states, or firms; technology and year fixed effects
- β_2, γ_2: how AI differs from other technologies

Fact 3 (cont.): Distance holds AI back within countries, not across

	Interaction with AI	
	Distance	Income
Across 17 countries	0.12**	0.58**
Across US states	-0.09***	0.69**
Across firms	-0.20***	-0.02**

Effect of distance or income on AI hiring, *relative to* other technologies

Notes: Coefficients on log distance and log income (GDP per capita; household income for states; sales for firms).

▸ Details

- **Across countries**: distance matters less for AI; income more
- **Within countries**: AI hiring falls off faster away from pioneer regions
- **Across firms**: distance matters more for AI; firm size less

Roadmap

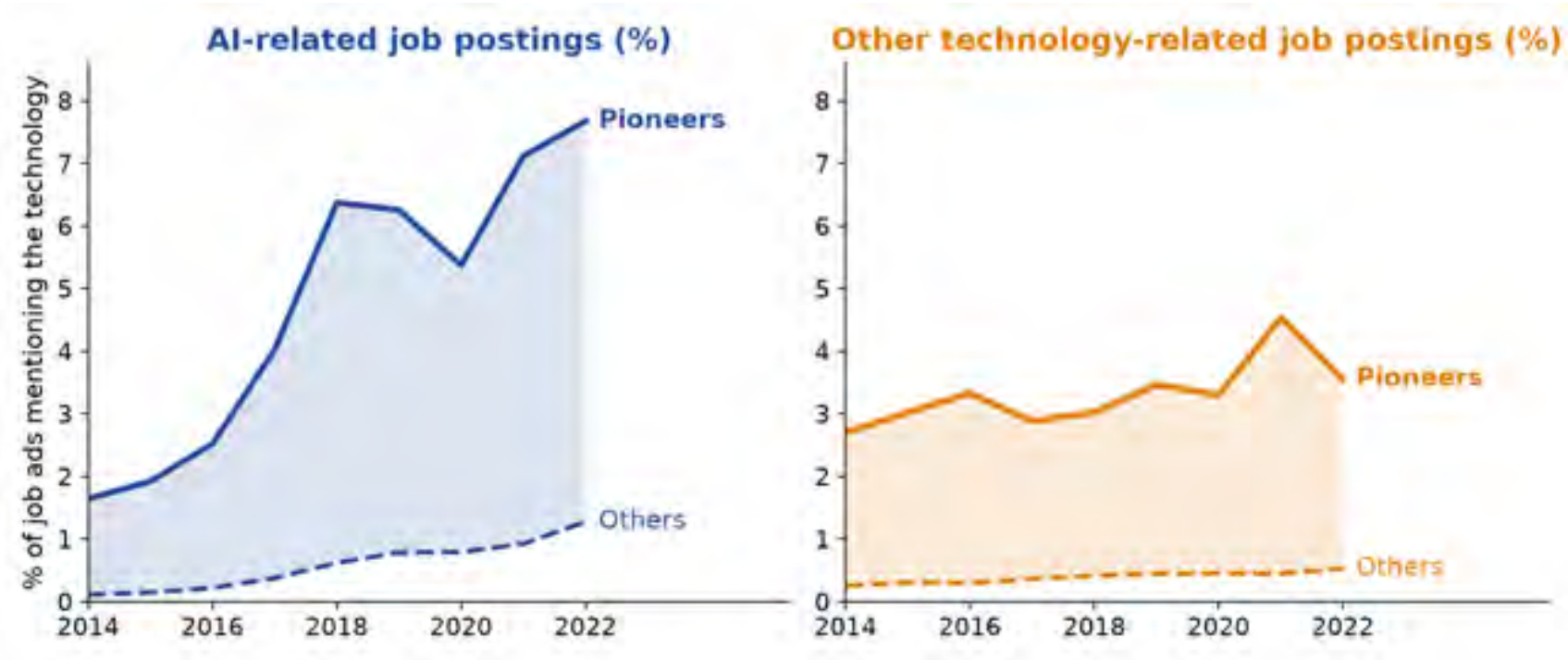
1. Motivation
2. Data
3. AI hiring at a glance
4. Stylized facts about AI diffusion
5. AI diffusion through multinational supply chains
6. Conclusion

Fact 4: AI pioneers race far ahead of other firms

Definitions

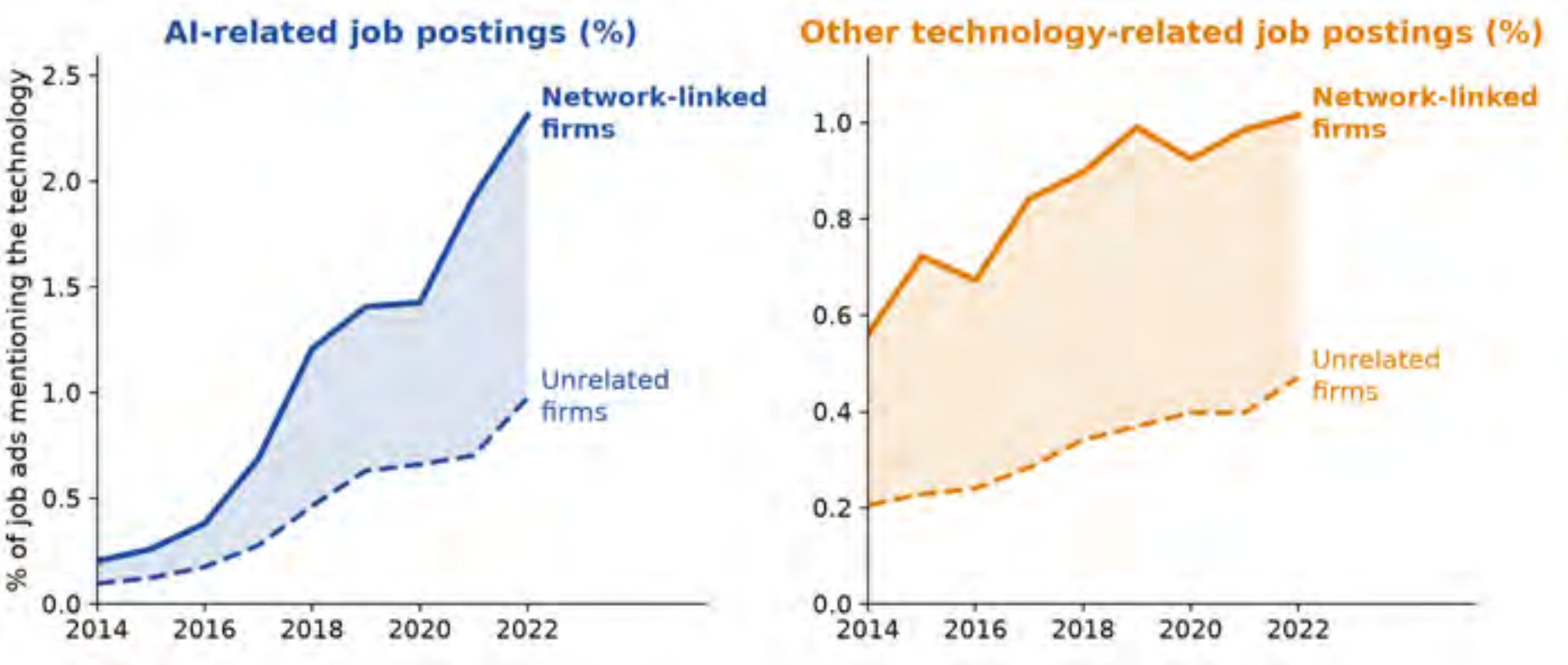
Pioneer firms: 5,031 firms in a technology's pioneer regions and lead industries before it emerged (Bastos et al., 2026)

Network-linked firms: pre-existing buyers, suppliers, innovation partners or affiliates of a pioneer firm (FactSet)



Notes: Share of job ads mentioning the technology, pioneer firms vs. all other firms. Raw data; same scale in both panels.

Fact 4 (cont.): Links to AI pioneers matter less early on



- In 2014, network-linked firms' edge was 0.1 pp for AI vs. 0.4 pp for other tech; it then widened for AI

Notes: Share of job ads mentioning the technology, network-linked vs. unrelated firms. Raw data. Note the different scales.

Fact 4 (cont.): Testing the network channel for AI diffusion

Specification

$$Y_{i,t}^{T} = \beta_1 Pioneer_i^T + \beta_2 (Pioneer \times AI)_i^T + \gamma_1 Link_i^T + \gamma_2 (Link \times AI)_i^T + \text{FE} + \epsilon_{i,t}^T$$

- Y: firm i's share of postings for T, relative to T's average
- Dummies (0/1):
  - $Pioneer_i^T$: firm i is a pioneer for T
  - $Link_i^T$: i is a pioneer's buyer, supplier or innovation partner (by type)
  - AI^T: technology T is AI
- 29 technologies pooled; firm, technology and year fixed effects (FE)
- β_2, γ_2: how AI differs from other technologies
- Each term also interacted with years since emergence to measure whether the effects change over time

Fact 4 (cont.): Links to AI pioneers matter less early on

	Dep. var.: normalized share of tech-related postings				
	Pioneer only	+ Buyer	+ Supplier	+ Innov. partner	All links
Pioneer	2.95***	2.67***	2.84***	2.68***	2.27**
Buyer		2.42***			1.87***
Supplier			1.08***		0.23
Innovation partner				2.19***	1.32***
1 Pioneer × AI	5.63***	5.94***	5.75***	5.93***	5.74***
Buyer × AI		-2.14***			-1.53***
2 Supplier × AI			-1.06***		-0.20
Innovation partner × AI				-2.02***	-1.13**

1. AI pioneers start far ahead of other technologies' pioneers
2. Links to AI pioneers add much less than for other technologies

*Notes: Effects in the year each technology emerged. Firm, technology, year FE; affiliates and s.e. omitted. *** p<0.01, ** p<0.05, * p<0.1*

Fact 4 (cont.): AI pioneers' lead fades; linkages catch up

	Dep. var.: normalized share of tech-related postings				
	Pioneer only	+ Buyer	+ Supplier	+ Innov. partner	All links
Pioneer × Years	-0.05	-0.02	-0.04	-0.03	-0.04
Buyer × Years		-0.10***			-0.09**
Supplier × Years			-0.04*		0.01
Innov. partner × Years				-0.09***	-0.04
3 Pioneer × AI × Years	-0.77***	-0.80***	-0.78***	-0.80***	-0.79***
Buyer × AI × Years		0.21***			0.14**
4 Supplier × AI × Years			0.08*		-0.03
Innov. partner × AI × Years				0.22***	0.17***

3. AI pioneers' lead shrinks about 0.8 a year faster
4. Links to AI pioneers gain about 0.2 a year relative to other tech

Notes: Same regressions as previous slide. Years = years since emergence.

Why do firm networks transmit less AI?

- **AI is embodied in workers**: it cannot be shipped along a value chain; firms must hire or train AI talent (Babina et al., 2024; Robertson et al., 2025)
- **Talent is pooled locally**: AI workers cluster in a few labor markets, and multinationals keep AI teams in core R&D hubs (Muro and Liu, 2021; McElheran et al., 2023)
- **Skills, not suppliers**: open-source tools spread the knowledge; what is scarce is people who can use it (Shrestha et al., 2023)
- **Early days**: network channels strengthen as AI matures and AI skills become more standard (Brynjolfsson et al., 2021)

Roadmap

1. Motivation
2. Data
3. AI hiring at a glance
4. Stylized facts about AI diffusion
5. AI diffusion through multinational supply chains
6. Conclusion

Conclusions: AI spreads more through local labor markets

- **AI diffuses differently**: more dispersed, more tied to income, held back by distance only within countries, and weakly carried by firm networks at first
- **Value chains alone are not enough**: links to AI pioneers add little early on – local skills, training and talent pipelines matter
- **Risk of an AI divide**: richer countries tilt hiring towards AI
- **Export controls may have limited reach** where domestic AI talent is strong
- **Next**: which labor market conditions – skills, demographics, mobility – enable AI uptake?

Thank you

Hannah Yi Wei
International Monetary Fund
hwei2@imf.org

Backup: Fact 4 is robust across controls and samples

	Baseline	Controlling for distance	Excluding Silicon Valley	Technologies emerging since 2014
Pioneer × AI	5.63***	5.63***	5.14***	5.70***
Buyer × AI	-2.14***	-2.15***	-1.91***	-3.50***
Supplier × AI	-1.06***	-1.07***	-0.93***	-2.01***
Innovation partner × AI	-2.02***	-2.03***	-1.77***	-3.60***

Notes: AI interactions at emergence, each from a separate regression. Last column: AI vs. autonomous vehicles and extremely high frequency technologies, all observed from emergence.

- The shortfall is in arm's-length links; links within the same organization matter for no technology
- Being linked to a pioneer in several ways does not help AI either

Backup: AI keywords

A job posting is classified as AI-related if it contains at least one of the following bigrams, in English or the relevant local language:

- ai machine
- ai technology
- artificial intelligence
- deep learning
- language processing
- learning algorithms
- machine intelligence
- machine learning
- natural language
- neural network
- reinforcement learning
- supervised learning
- unsupervised learning

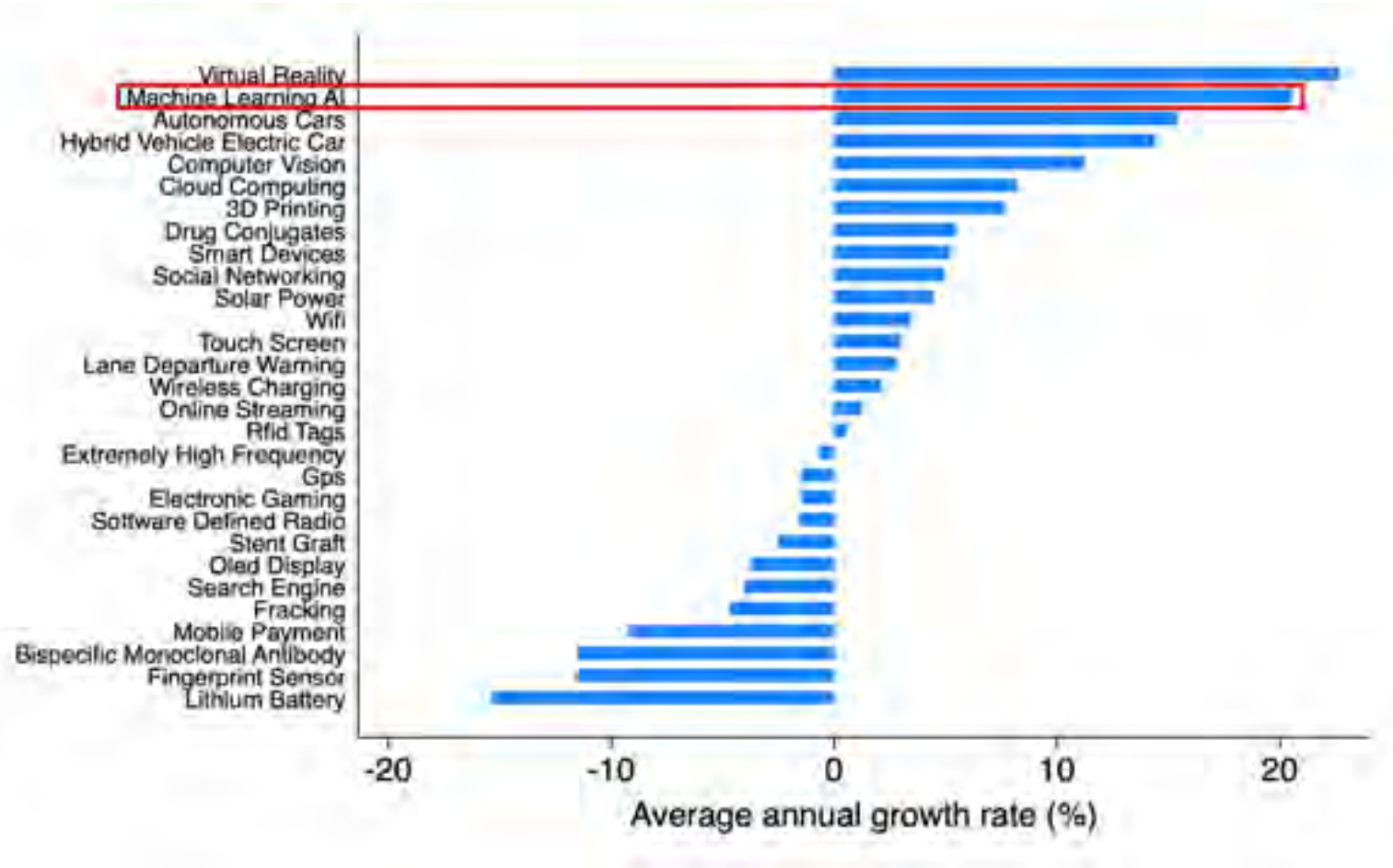
◂ Back

Backup: Sample

- **17 countries** (about 60% of world GDP): Australia, Austria, Belgium, Canada, Denmark, France, Germany, Italy, Luxembourg, Netherlands, New Zealand, Norway, Singapore, Sweden, Switzerland, United Kingdom, United States
- **Period**: 2014Q1–2022Q4 for all countries except Sweden (2016Q1–2022Q3)
- **Postings**: 573 million, all occupations and sectors, as scraped by Lightcast from company websites and job boards
- **Firms**: 107,981 posting firms matched to FactSet, in all 17 countries; 5,031 pioneer firms, all in the US
- **Technologies**: 29 from Bloom et al. (2021), emerging between 2002 and 2015; AI is the latest

◂ Back

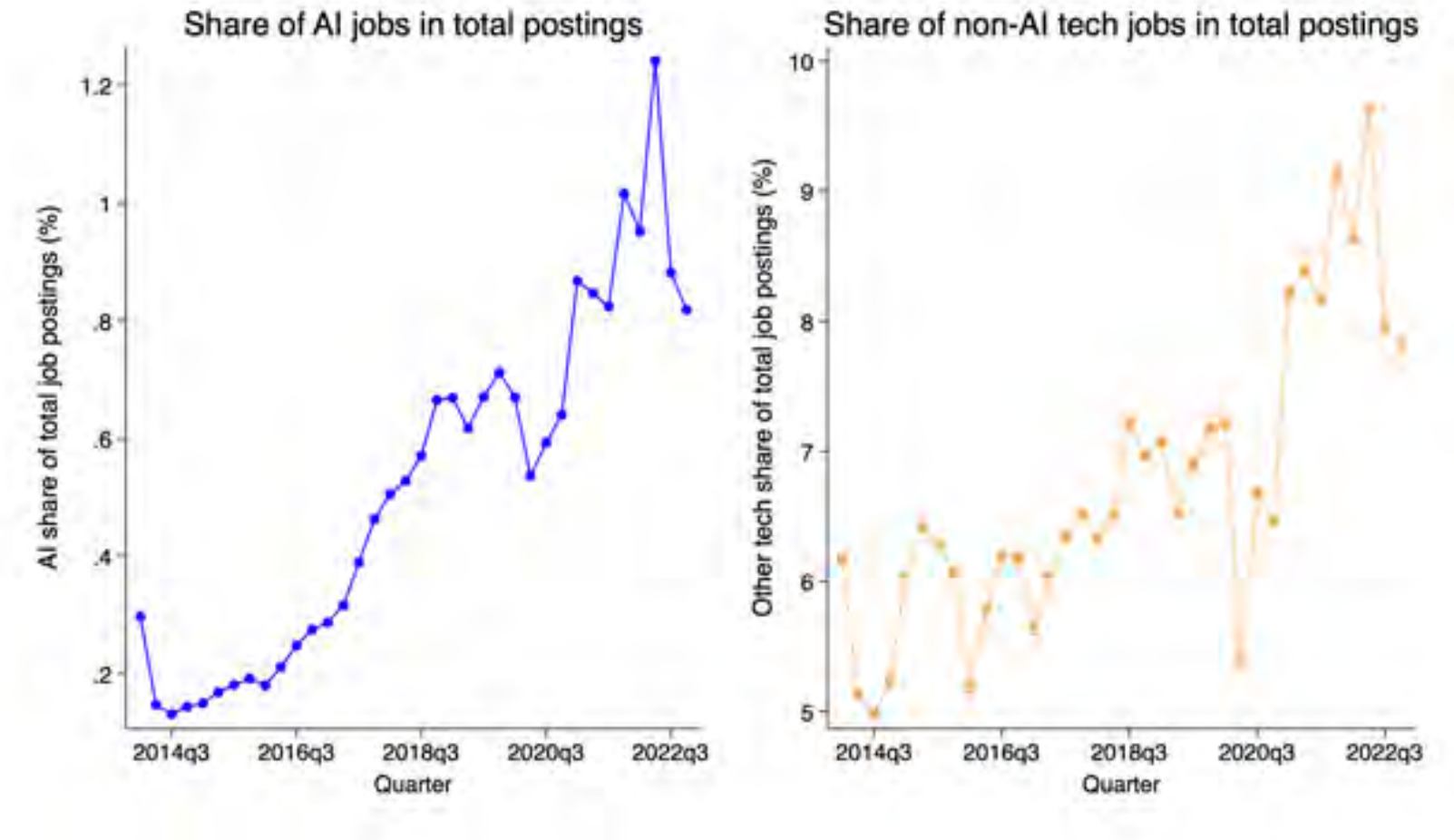
Backup: AI grew 20.5% a year – second fastest of 29 technologies



Notes: Compound annual growth rate of each technology's share in total job postings, 2014Q1–2022Q4. Average across the 29 technologies 2.3%.

◂ Back

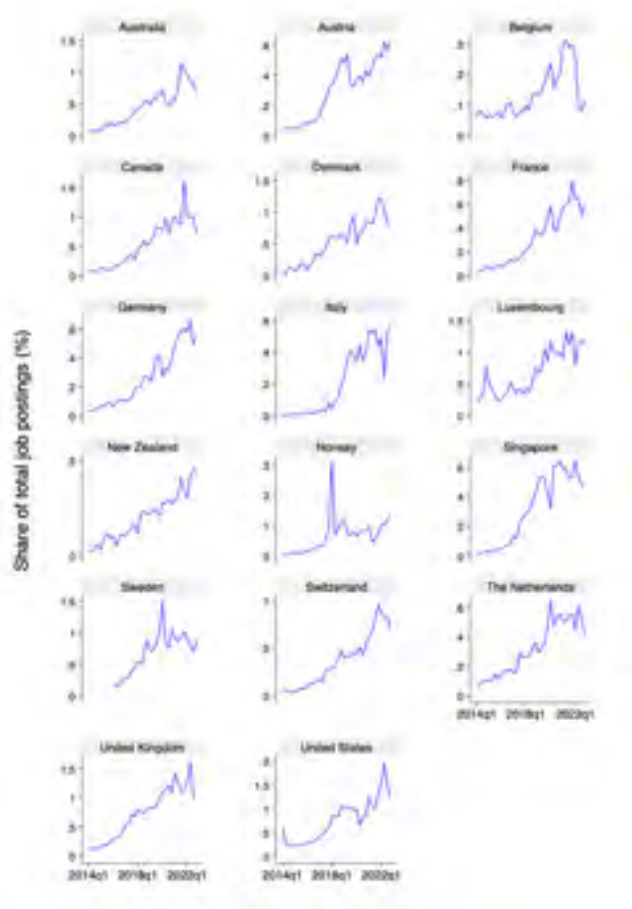
Backup: AI and other technology postings, levels



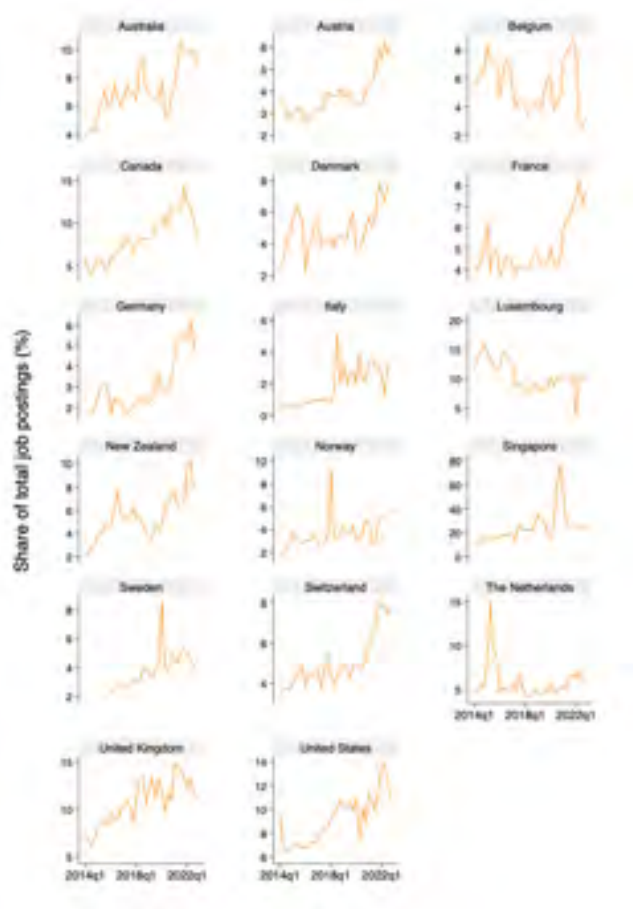
Notes: AI-related postings (left) and non-AI technology postings (right) as shares of total job postings, 17 countries, 2014Q1–2022Q4.

Backup: Posting shares by country

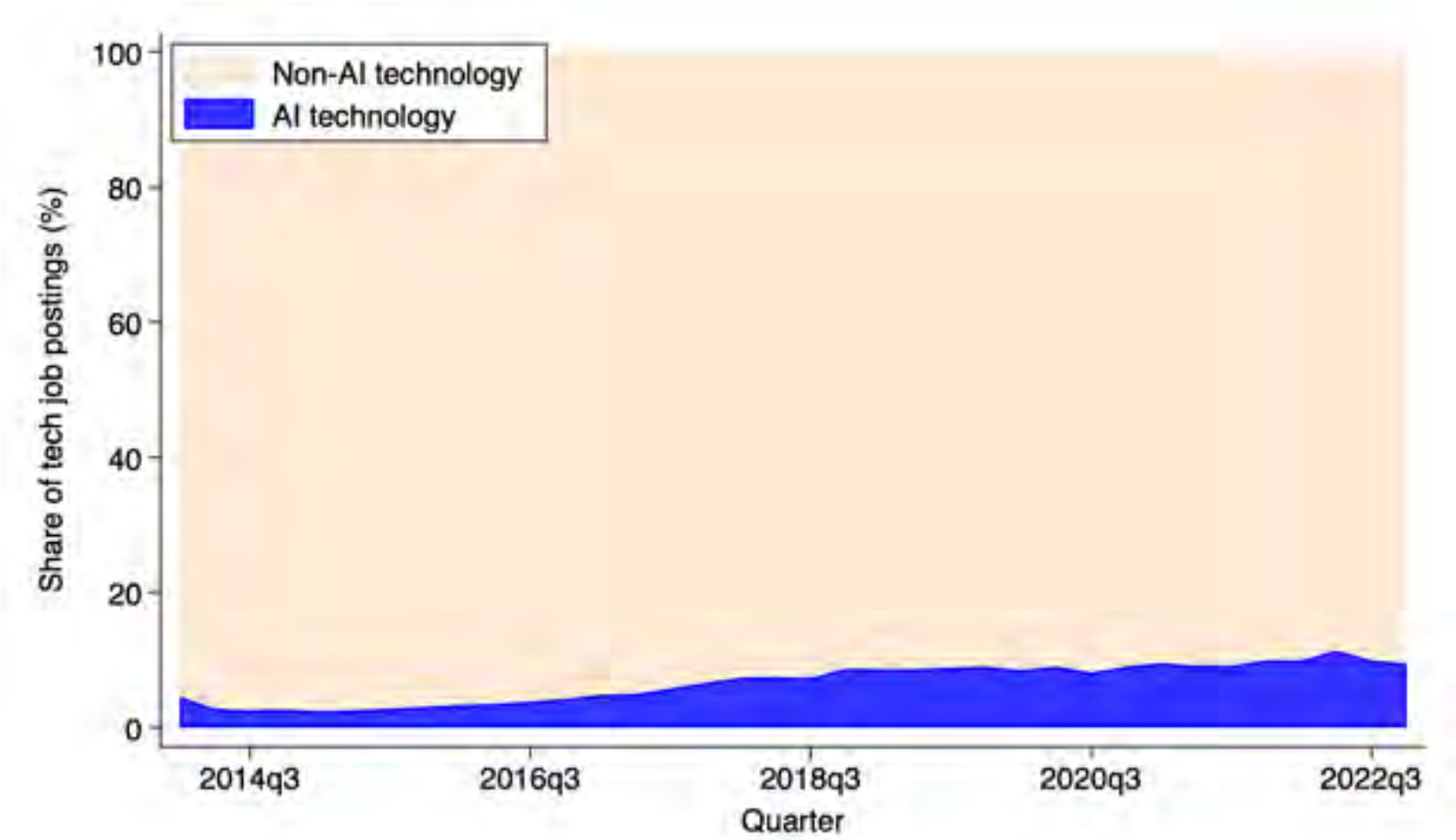
AI postings, % of total, by country



Non-AI tech postings, % of total, by country



Backup: AI takes a growing share of all technology hiring



Notes: AI-related and non-AI technology postings as shares of total technology-related postings, 17 countries, 2014Q1–2022Q4.

Backup: Dispersion over time

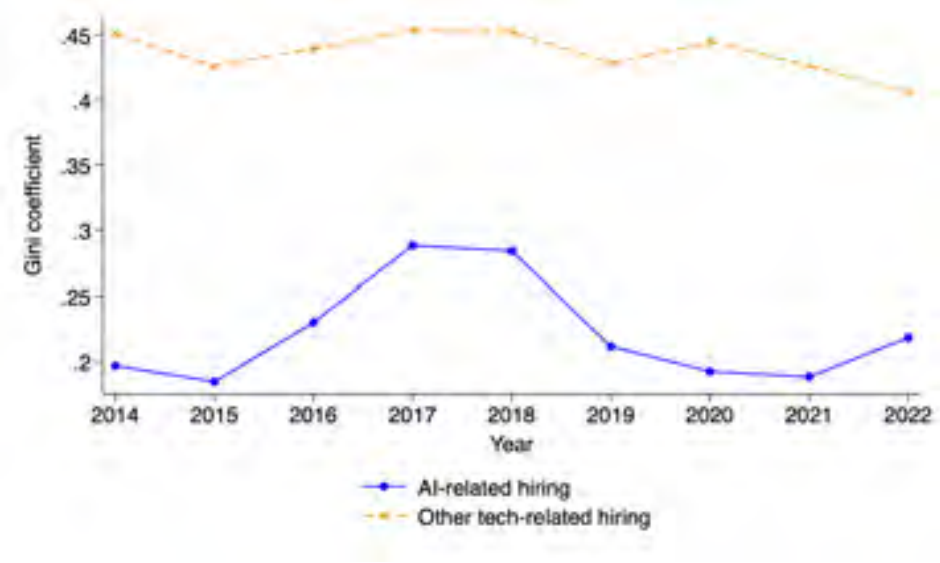
	Coefficient of Variation (1)	Gini Coefficient (2)
	Panel (a): Across 17 countries	
Years since emergence	-0.026*	-0.006
	(0.014)	(0.003)
Years since emergence × AI	0.005	0.002
	(0.012)	(0.004)
Mean (non-AI)	1.149	0.437
Mean (AI)	0.431	0.221
	Panel (b): Across 156 regions	
Years since emergence	-0.035**	-0.001
	(0.014)	(0.003)
Years since emergence × AI	-0.007	-0.007***
	(0.010)	(0.002)
Mean (non-AI)	1.754	0.517
Mean (AI)	1.021	0.453

*Notes: 261 technology-year observations per panel. Technology and year fixed effects. Standard errors clustered by technology. *** $p<0.01$, ** $p<0.05$, * $p<0.1$*

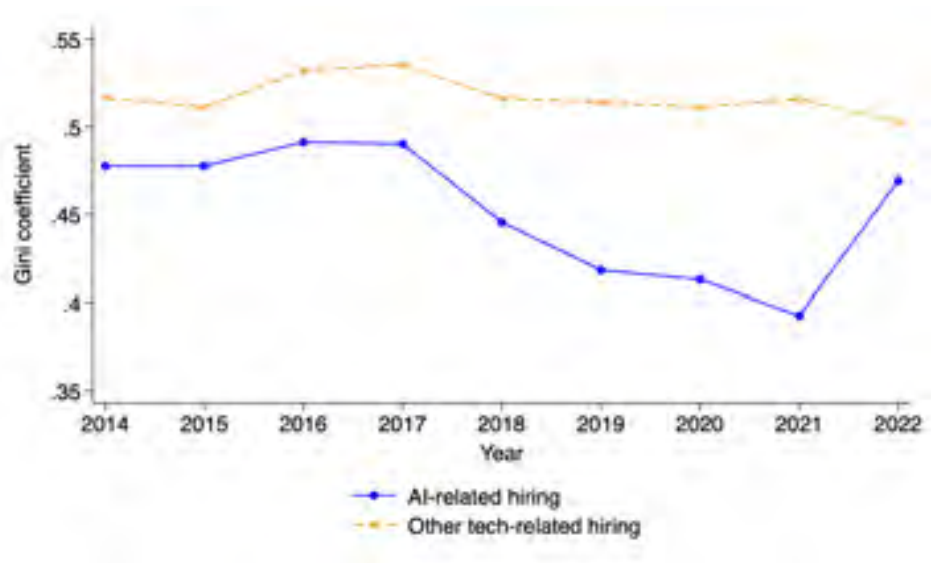
◂ Back

Backup: Gini coefficients show the same dispersion pattern

Gini coefficient across 17 countries



Gini coefficient across 156 regions



Notes: Gini coefficient of AI-related job postings (blue) and of other technology-related job postings (orange).

Backup: Distance and income across countries

	Dep var: Normalized share of tech-related postings	
	(1)	(2)
Log distance from technology pioneer	-0.122**	-0.147***
	(0.054)	(0.052)
Log distance from technology pioneer × AI	0.123**	0.099*
	(0.054)	(0.052)
Log income per capita		0.919***
		(0.250)
Log income per capita × AI		0.575**
		(0.245)
Adjusted R-squared	0.063	0.112
Observations	4,350	4,350

*Notes: Country-technology-year level, 17 countries, 2014–2022. Technology and year fixed effects. Standard errors clustered by technology. *** p<0.01, ** p<0.05, * p<0.1*

◂ Back

Backup: Distance and income across US states

	Dep var: Normalized share of tech-related postings	
	(1)	(2)
Log distance from technology pioneer	-0.118***	-0.107***
	(0.029)	(0.027)
Log distance from technology pioneer × AI	-0.087***	-0.062**
	(0.029)	(0.028)
Log median household income		0.430
		(0.300)
Log median household income × AI		0.685**
		(0.270)
Adjusted R-squared	0.110	0.113
Observations	13,311	13,311

*Notes: State-technology-year level, US states. Technology and year fixed effects. Standard errors clustered by technology. *** $p<0.01$, ** $p<0.05$, * $p<0.1$*

Backup: Distance and size across firms

	Dep var: Normalized share of tech-related postings			
	(1)	(2)	(3)	(4)
Log distance from technology pioneer	-0.073***	-0.099***	-0.127***	-0.159***
	(0.016)	(0.020)	(0.024)	(0.031)
Log distance from technology pioneer × AI	-0.197***	-0.196***	-0.066**	-0.064**
	(0.016)	(0.016)	(0.024)	(0.025)
Log firm sales in USD			0.045***	0.042***
			(0.013)	(0.014)
Log firm sales in USD × AI			-0.022**	-0.022
			(0.013)	(0.013)
Country FE		Yes		Yes
Adjusted R-squared	0.002	0.002	0.002	0.002
Observations	19,135,505	19,135,505	3,963,082	3,963,082

*Notes: Firm-technology-year level, 107,981 firms, years after each technology's emergence. Technology and year fixed effects. Standard errors clustered by technology. *** p<0.01, ** p<0.05, * p<0.1*

Backup: The shortfall is in linkages outside the firm boundary

	Dep var: Normalized share of tech-related postings				
	(1)	(2)	(3)	(4)	(5)
Linkages of pioneer firms					
Buyer outside org.		2.413***			1.873***
		(0.379)			(0.403)
Buyer outside org. × AI		-2.146***			-1.613***
		(0.461)			(0.459)
Supplier outside org.			1.064***		0.232
			(0.232)		(0.231)
Supplier outside org. × AI			-1.059***		-0.196
			(0.270)		(0.262)
Innovation partner outside org.				2.167***	1.287***
				(0.356)	(0.379)
Innovation partner outside org. × AI				-2.051***	-1.111**
				(0.432)	(0.444)
Pioneer terms, within-org. linkages and time interactions omitted for brevity					
Adjusted R-squared	0.041	0.041	0.041	0.041	0.041
Observations	10,072,709	10,072,709	10,072,709	10,072,709	10,072,709

*Notes: Technology, year and firm fixed effects. Standard errors clustered by firm. Within-organization links: no significant effect for any technology. *** p<0.01, ** p<0.05, * p<0.1*

Backup: Multiple links to pioneers

	Dep var: Normalized share of tech-related postings				
	(2)	(3)	(4)	(5)	(6)
Linkages of pioneer firms, outside org.					
Buyer and supplier	2.789***				2.033*
	(0.627)				(1.050)
Buyer and supplier × AI	-2.636***				-2.436**
	(0.733)				(1.132)
Supplier and innovation partner		2.351***			1.511**
		(0.485)			(0.596)
Supplier and innovation partner × AI		-2.196***			-1.659**
		(0.585)			(0.669)
Innovation partner and buyer			2.817***		2.226**
			(0.597)		(0.966)
Innovation partner and buyer × AI			-2.639***		-2.943***
			(0.736)		(1.103)
Innovation partner and buyer and supplier				3.083***	-2.420
				(0.758)	(1.708)
Innovation partner and buyer and supplier × AI				-2.682***	4.381**
				(0.925)	(1.946)
Pioneer terms and time interactions omitted for brevity					
Adjusted R-squared	0.041	0.041	0.041	0.041	0.041
Observations	10,072,709	10,072,709	10,072,709	10,072,709	10,072,709

*Notes: Technology, year and firm fixed effects. Standard errors clustered by firm. *** p<0.01, ** p<0.05, * p<0.1*


PIIE
PETERSON INSTITUTE FOR
INTERNATIONAL ECONOMICS
OECD
INTERNATIONAL MONETARY FUND
WORLD BANK GROUP
www.piie.com
@PIIE
linkedin.com/company/piie

The Race between AI Costs and Capabilities: Task-Based Automation When Machines Improve

Danial Lashkari*+, Omeed Maghzian+, Matthias Mertens+, Christina Qiu†, Neil Thompson+

*LSE, +MIT FutureTech, †University of Miami

November 18, 2026

Two Views of AI Automation

> *"We are on course for AGI by 2027. These AI systems will basically be able to automate basically all cognitive jobs (think: all jobs that could be done remotely)."*
>
> *Aschenbrenner (2024), Situational Awareness*

Capabilities view: automation follows the trend in what AI can do

Evidence: Kwa et al. (METR, 2025); Altman (2025)

Two Views of AI Automation

> *"We are on course for AGI by 2027. These AI systems will basically be able to automate basically all cognitive jobs (think: all jobs that could be done remotely)."*
>
> *Aschenbrenner (2024), Situational Awareness*

Capabilities view: automation follows the trend in what AI can do

Evidence: Kwa et al. (METR, 2025); Altman (2025)

> *"What $1tn problem will AI solve? Replacing low-wage jobs with tremendously costly technology is basically the polar opposite of the prior technology transitions."*
>
> *Covello, Goldman Sachs Top of Mind (2024)*

Cost view: automation happens when AI computing is cheaper than labor

The Race between AI Costs and Capabilities

Dynamics of Capabilities: growth in the scale of frontier AI models

- New models expand the capability set but require more compute

The Race between AI Costs and Capabilities

Dynamics of Capabilities: growth in the scale of frontier AI models

- New models expand the capability set but require more compute

Dynamics of Costs: innovations in computing/model design and organizational adaptation

- Cheaper compute and better architectures lower computing costs on tasks AI can perform
- Easing frictions (regulatory, organizational) lowers the non-computing costs of adopting AI

The Race between AI Costs and Capabilities

Dynamics of Capabilities: growth in the scale of frontier AI models

- New models expand the capability set but require more compute

Dynamics of Costs: innovations in computing/model design and organizational adaptation

- Cheaper compute and better architectures lower computing costs on tasks AI can perform
- Easing frictions (regulatory, organizational) lowers the non-computing costs of adopting AI

How do we understand these dynamics through the lens of the task-based automation theories?

Zeira (1998); Autor, Levy & Murnane (2003); Acemoglu & Restrepo (2018, 2019, 2022); Jones & Tonetti (2026)

This Paper

Quantitative task-based theory of automation with **menu of AI models:**

- Model choice: adoption trades off capability and costs subject to (labor) frictions
- Multi-occupation GE environment: AI progress + decline in compute price & frictions

This Paper

Quantitative task-based theory of automation with **menu of AI models:**

- Model choice: adoption trades off capability and costs subject to (labor) frictions
- Multi-occupation GE environment: AI progress + decline in compute price & frictions

Data and quantification:

- MIT FutureTech Survey: worker evaluations of LLM performance for models of varying size
- Estimating AI cost structure: cross-task/model relation bet. capability and compute requirement
- Calibration: match frontier-model usage + agg. LLM expenditure + SOC-2-level AI adoption

This Paper

Quantitative task-based theory of automation with **menu of AI models:**

- Model choice: adoption trades off capability and costs subject to (labor) frictions
- Multi-occupation GE environment: AI progress + decline in compute price & frictions

Data and quantification:

- MIT FutureTech Survey: worker evaluations of LLM performance for models of varying size
- Estimating AI cost structure: cross-task/model relation bet. capability and compute requirement
- Calibration: match frontier-model usage + agg. LLM expenditure + SOC-2-level AI adoption

Quantitative Exercise:

- Today, compute costs limit automation mainly through firms' choice of model
- Algorithmic progress drives automation and growth; cheaper compute alone adds little
- Growth comes mainly from TFP, with a small contribution from AI capital accumulation

Tasks and Production

Output requires production in a continuum of task varieties ν:

$$y = f(k, \ell) = \left(\int_0^1 y(\nu)^{\frac{\lambda-1}{\lambda}} \, d\nu \right)^{\frac{\lambda}{\lambda-1}} \qquad \lambda \in (0, 1)$$

Variety ν output with machine m:

$$y(\nu) = \frac{\ell_L(\nu)}{\tau}$$

Labor ℓ_L: task length τ of labor time per unit

Tasks and Production

Output requires production in a continuum of task varieties ν:

$$y = f(k, \ell) = \left(\int_0^1 y(\nu)^{\frac{\lambda-1}{\lambda}} \, d\nu \right)^{\frac{\lambda}{\lambda-1}} \qquad \lambda \in (0, 1)$$

Variety ν output with machine m:

$$y(\nu) = \begin{cases} \frac{\ell_L(\nu)}{\tau}, & q(\nu; m) < q^* \quad \text{AI is incapable} \\ \frac{\ell_L(\nu)}{\tau} + \min\left\{ \frac{k(\nu)}{\gamma(\nu;m)}, \frac{\ell_K(\nu)}{\omega\,\tau} \right\}, & q(\nu; m) \geq q^* \quad \text{AI is capable} \end{cases}$$

Labor ℓ_L: task length τ of labor time per unit

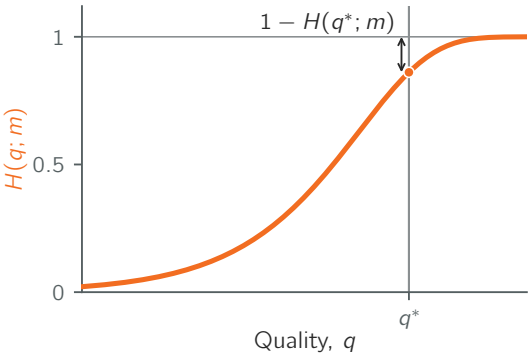
Machine m:

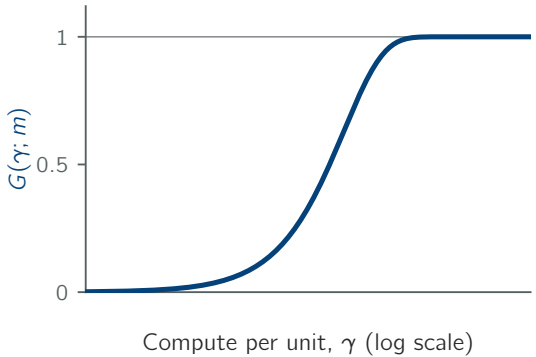
- Capability: quality $q(\nu; m)$ of machine's work should exceed benchmark q^*
- Cost: compute $\gamma(\nu; m)$ per unit of output
- Overhead: labor time $\omega\,\tau$ per unit a machine produces

Capabilities and Costs Across Varieties

Capability distribution $H(q; m)$: share of varieties with quality $q(\nu; m) < q$

Cost distribution $G(\gamma; m)$: share of capable varieties with compute $\gamma(\nu; m) \leq \gamma$

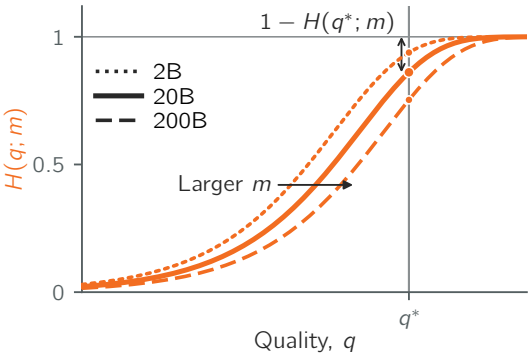


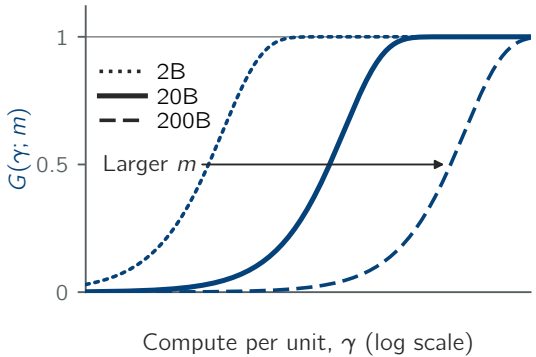


Capabilities and Costs Across Varieties

Capability distribution $H(q; m)$: share of varieties with quality $q(\nu; m) < q$

Cost distribution $G(\gamma; m)$: share of capable varieties with compute $\gamma(\nu; m) \leq \gamma$





Larger m: more capable varieties, higher compute requirements

Automation within the Task

Automate variety ν iff the machine bundle is cheaper than labor:

Details

$$R\,\gamma(\nu; m) + \omega\,\tau\,w \;\leq\; \tau\,w \qquad \Longleftrightarrow \qquad \gamma(\nu; m) \;\leq\; \gamma^* \equiv \tau\,(1-\omega)\,\frac{w}{R}$$

Automation within the Task

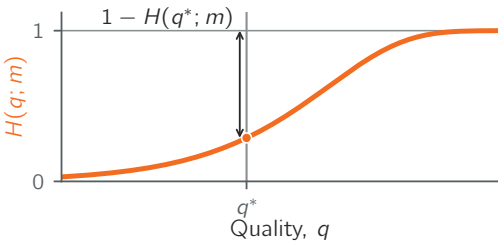
Automate variety ν iff the machine bundle is cheaper than labor:

Details

$$R\,\gamma(\nu; m) + \omega\,\tau\,w \;\leq\; \tau\,w \qquad \Longleftrightarrow \qquad \gamma(\nu; m) \;\leq\; \gamma^* \equiv \tau\,(1-\omega)\,\frac{w}{R}$$

Automated share = capabilities × conditional comparative advantage:

$$\varphi_K(m) = \underbrace{(1 - H(q^*; m))}_{\text{capability share}} \times \underbrace{G(\gamma^*; m)}_{\text{conditional comparative advantage}}$$



Automation within the Task

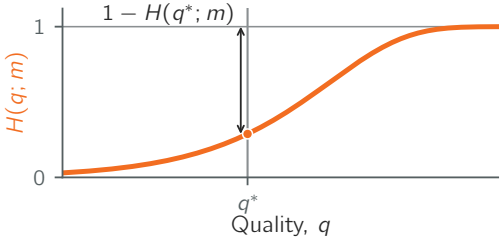
Automate variety ν iff the machine bundle is cheaper than labor:

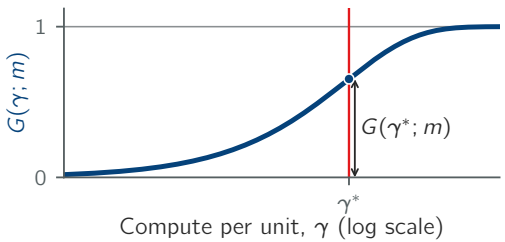
Details

$$R\,\gamma(\nu; m) + \omega\,\tau\,w \;\leq\; \tau\,w \qquad \Longleftrightarrow \qquad \gamma(\nu; m) \;\leq\; \gamma^* \equiv \tau\,(1-\omega)\,\frac{w}{R}$$

Automated share = capabilities × conditional comparative advantage:

$$\varphi_K(m) = \underbrace{(1 - H(q^*; m))}_{\text{capability share}} \times \underbrace{G(\gamma^*; m)}_{\text{conditional comparative advantage}}$$





Machine Choice: Capabilities vs. Costs

Unit cost: labor's cost, reduced by capability × saving

$$c(w, R; m) = \tau\, w \left[1 - \big(1 - H(q^*; m)\big)\, \zeta(\gamma^*; m)\right]^{\frac{1}{1-\lambda}}, \qquad m^* = \arg\min_{m \in M} c(w, R; m)$$

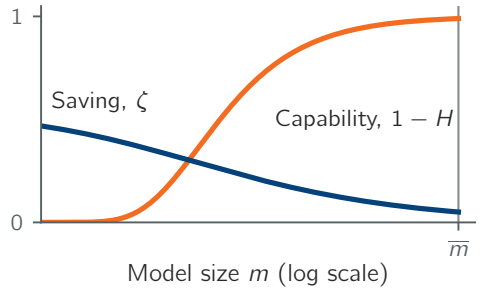
- Saving per capable variety: $\zeta(\gamma^*; m) \equiv \int_0^{\gamma^*} \left[1 - \left((1-\omega)\frac{\gamma}{\gamma^*} + \omega\right)^{1-\lambda}\right] dG(\gamma; m) \leq 1 - \omega^{1-\lambda}$

Machine Choice: Capabilities vs. Costs

Unit cost: labor's cost, reduced by capability × saving

$$c(w, R; m) = \tau\, w \left[1 - \big(1 - H(q^*; m)\big)\, \zeta(\gamma^*; m)\right]^{\frac{1}{1-\lambda}}, \qquad m^* = \arg\min_{m \in M} c(w, R; m)$$

- Saving per capable variety: $\zeta(\gamma^*; m) \equiv \int_0^{\gamma^*} \left[1 - \left((1-\omega)\frac{\gamma}{\gamma^*} + \omega\right)^{1-\lambda}\right] dG(\gamma; m) \leq 1 - \omega^{1-\lambda}$

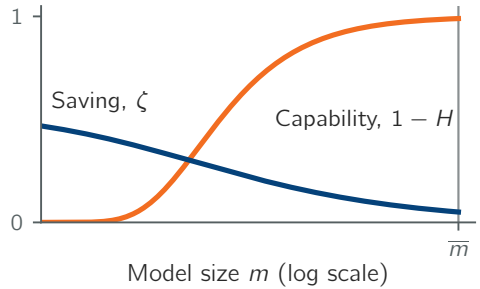


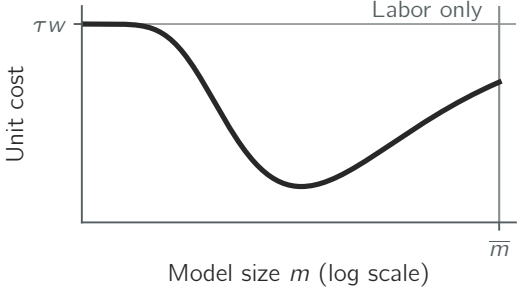
Machine Choice: Capabilities vs. Costs

Unit cost: labor's cost, reduced by capability × saving

$$c(w, R; m) = \tau\, w \left[1 - \big(1 - H(q^*; m)\big)\, \zeta(\gamma^*; m)\right]^{\frac{1}{1-\lambda}}, \qquad m^* = \arg\min_{m \in M} c(w, R; m)$$

- Saving per capable variety: $\zeta(\gamma^*; m) \equiv \int_0^{\gamma^*} \left[1 - \left((1-\omega)\frac{\gamma}{\gamma^*} + \omega\right)^{1-\lambda}\right] dG(\gamma; m) \leq 1 - \omega^{1-\lambda}$



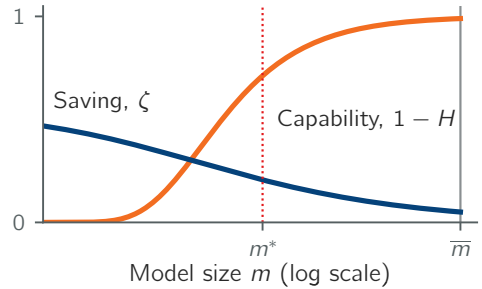


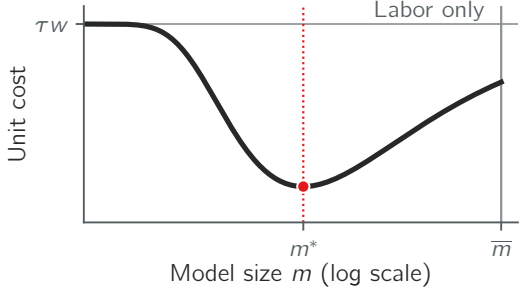
Machine Choice: Capabilities vs. Costs

Unit cost: labor's cost, reduced by capability × saving

$$c(w, R; m) = \tau\, w \left[1 - \big(1 - H(q^*; m)\big)\, \zeta(\gamma^*; m)\right]^{\frac{1}{1-\lambda}}, \qquad m^* = \arg\min_{m \in M} c(w, R; m)$$

- Saving per capable variety: $\zeta(\gamma^*; m) \equiv \int_0^{\gamma^*} \left[1 - \left((1-\omega)\frac{\gamma}{\gamma^*} + \omega\right)^{1-\lambda}\right] dG(\gamma; m) \leq 1 - \omega^{1-\lambda}$





As costs change

Quantitative Model: Production Given the Menu of Models

Menu of AI models $M_t \equiv [0, \overline{m}_t]$: parameter size m trained with data $d = D(m) = m^{\beta_D}$

Final output from occupational output $y_{o,t}$ combining tasks and non-AI capital $X_{o,t}$:

$$Y_t = Z_t \left(\sum_o \xi_o^{\frac{1}{\varepsilon}} y_{o,t}^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}} \qquad y_{o,t} = Z_o X_{o,t}^{\alpha_o} \left(\sum_{j \in \mathcal{J}_o} \mu_{j,o}^{\frac{1}{\lambda}} y_{j,t}^{\frac{\lambda-1}{\lambda}} \right)^{\frac{\lambda}{\lambda-1}(1-\alpha_o)}$$

O*NET task group j: choose model $m_{j,t}^*$ maximizing $y_{j,t} = f(\ell_{j,t}, k_{j,t}; m)$

- Capability distribution $H_j(q; m)$: $q = Q(m, \tau_j) + u_q$ with $u_q \sim \text{Gumbel}_{\min}(\kappa_H)$
- Cost distribution $G_j(\gamma; m)$: $\log \gamma = \Gamma(m, \tau_j) + u_\gamma$ with $u_\gamma \sim \text{Gumbel}_{\min}(\kappa_G)$
- Subject to adoption frictions $\omega_{o,t}$

Aggregate non-AI capital X_t rented at $p_{X,t}$ and accumulated as $\dot{X}_t = A_{X,t} Y_{X,t} - \delta_X X_t$

Quantitative Model: the AI Sector and Equilibrium

Computing sector: produces AI capital services $K_t = X_{K,t}$ from computing capital $X_{K,t}$

$$\dot{X}_{K,t} = A_t Y_{K,t} - \delta_K X_{K,t} \qquad R_t = \frac{r_t + \delta_K + \eta_{A,t}}{A_t}$$

- Firms pay $\omega_{R,t} R_t$ and computing capital receives R_t (tax $\omega_{R,t} - 1$ rebated lump sum)

AI (foundation-model) sector: training frontier model uses compute $\chi \overline{m}_t D(\overline{m}_t)$ (tax-financed)

Market clearing:

$$L_t = \sum_o \ell_{o,t} \qquad X_t = \sum_o X_{o,t} \qquad K_t = \sum_o k_{o,t} + \chi \overline{m}_t D(\overline{m}_t) \qquad Y_t = C_t + Y_{X,t} + Y_{K,t}$$

Data

Survey data: LLM capabilities and task length in LLM-addressable tasks via Prolific

Mertens et al. (2026)

- Currently: 801 occupations, 7,404 tasks, 10,503 task examples, 70,055 evaluations, 41 models.

Data

Survey data: LLM capabilities and task length in LLM-addressable tasks via Prolific
Mertens et al. (2026)

- Currently: 801 occupations, 7,404 tasks, 10,503 task examples, 70,055 evaluations, 41 models.

LLM-addressable tasks: Query GPT-4 if LLMs can generate a 10% time saving on task

Performance: Participants see LLM task example solution and rate it on a 1–9 scale

- *Overall, how would a manager view this response to the task?*
- *Not useful (1–3), useful with edits (4–6), useful as is (7–9); 8 = average worker quality*

Data

Survey data: LLM capabilities and task length in LLM-addressable tasks via Prolific
Mertens et al. (2026)

- Currently: 801 occupations, 7,404 tasks, 10,503 task examples, 70,055 evaluations, 41 models.

LLM-addressable tasks: Query GPT-4 if LLMs can generate a 10% time saving on task

Performance: Participants see LLM task example solution and rate it on a 1–9 scale

- *Overall, how would a manager view this response to the task?*
- *Not useful (1–3), useful with edits (4–6), useful as is (7–9); 8 = average worker quality*

Task length: Participants estimate the time it takes to complete the task

- *How much time would it take a typical employee working this job to complete this task from scratch?*

Token costs: Costs for generating task-example-level LLM output

Collection scheme | Performance/duration distribution | Performance-duration scatter | Performance by SOC2 | Survey models

Estimating Capabilities

Capabilities: project responses on the model's Epoch Capabilities Index ECI

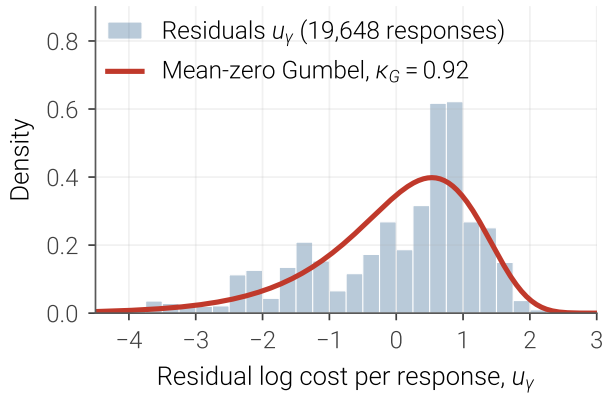
$$\text{quality grade} \;=\; Q\Big(\text{ECI},\, \tau\Big) \;+\; u_q$$

	Grade (1–9): $Q(\text{ECI}, \tau)$, 30 models				
	(1) Baseline	(2)	(3)	(4)	(5)
ECI	+0.027	+0.027	+0.029	+0.027	+0.027
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\log \tau$	−0.055	−0.054	−0.122	−0.033	absorbed
	(0.023)	(0.023)	(0.010)	(0.024)	
ECI × $\log \tau$	+0.002		+0.002	+0.002	+0.003
	(0.001)		(0.001)	(0.001)	(0.001)
Wage	✓	✓		✓	absorbed
IWA family, family × $\log \tau$	✓	✓		✓	absorbed
IWA family × ECI	✓	✓		✓	✓
Occupation (SOC 2-digit) effects				✓	
Rater effects (one task each)					✓
N	38,069	38,069	38,069	38,069	38,069
R^2 (with fixed effects)	0.032	0.032	0.028	0.036	0.456

Estimating Compute Requirements

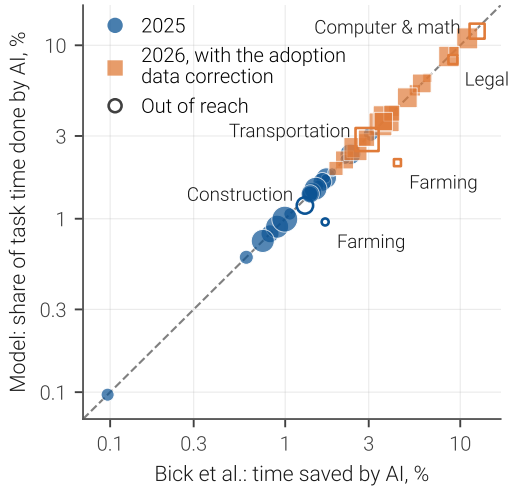
$$\underbrace{\log \gamma}_{\log(\text{tokens} \times \text{price per token})} = \underbrace{\log T\left(\text{ECI}, \tau\right)}_{\text{mean log tokens per answer}} + \underbrace{\log P(m)}_{\text{mean log price per token}} + u_\gamma$$

	$\log T$	$\log P$
ECI	+0.017	
	(0.006)	
$\log \tau$	+0.019	
	(0.003)	
$\log m$		+1.184
		(0.170)
Controls		
Reasoning	✓	
Wage	✓	
Models	30	30
N	19,648	19,648
R^2	0.588	0.626



Calibration Targets and AI Adoption Fit

	Data	Model
Targeted		
Frontier cost ratio		
2025	1.466	1.374
2026	1.657	1.777
AI spending share		
2025	0.048%	0.047%
2026	0.229%	0.237%
Time saved by AI		
2025	1.35%	1.34%
2026	4.88%	4.84%
2026, uncorrected	1.69%	
Not targeted		
Δ compute growth	×4.87	×5.21
Frontier compute share		
2025		17.1%
2026		12.3%



Adoption Frictions by Occupation

Model Choice and Automation (calibrated, 2026)

In the aggregate:

capability: 6.2%

conditional comparative advantage: 0.99

$$\underbrace{\varphi_K(m^*)}_{6.1\%} = \overbrace{\underbrace{(1 - H(q^*; \overline{m}))}_{\substack{\text{frontier} \\ 17.4\%}} \times \underbrace{\frac{1 - H(q^*; m^*)}{1 - H(q^*; \overline{m})}}_{\substack{\text{model choice} \\ \times 0.35}}}^{} \times \overbrace{\underbrace{G(\gamma^*; \overline{m})}_{\substack{\text{frontier} \\ 0.15}} \times \underbrace{\frac{G(\gamma^*; m^*)}{G(\gamma^*; \overline{m})}}_{\substack{\text{model choice} \\ \times 6.7}}}^{}$$

Model Choice and Automation (calibrated, 2026)

In the aggregate:

$$\underbrace{\varphi_K(m^*)}_{6.1\%} = \overbrace{\underbrace{(1 - H(q^*; \overline{m}))}_{\substack{\text{frontier} \\ 17.4\%}} \times \underbrace{\frac{1 - H(q^*; m^*)}{1 - H(q^*; \overline{m})}}_{\substack{\text{model choice} \\ \times 0.35}}}^{\text{capability: } 6.2\%} \times \overbrace{\underbrace{G(\gamma^*; \overline{m})}_{\substack{\text{frontier} \\ 0.15}} \times \underbrace{\frac{G(\gamma^*; m^*)}{G(\gamma^*; \overline{m})}}_{\substack{\text{model choice} \\ \times 6.7}}}^{\text{conditional comparative advantage: } 0.99}$$

Across tasks (covariance shares): model choice accounts for most variation

$$\log \varphi_K(m^*) = \underbrace{\log (1 - H(q^*; \overline{m}))}_{21.3\%} + \underbrace{\log \frac{1 - H(q^*; m^*)}{1 - H(q^*; \overline{m})}}_{75.2\%} + \underbrace{\log G(\gamma^*; m^*)}_{3.5\%}$$

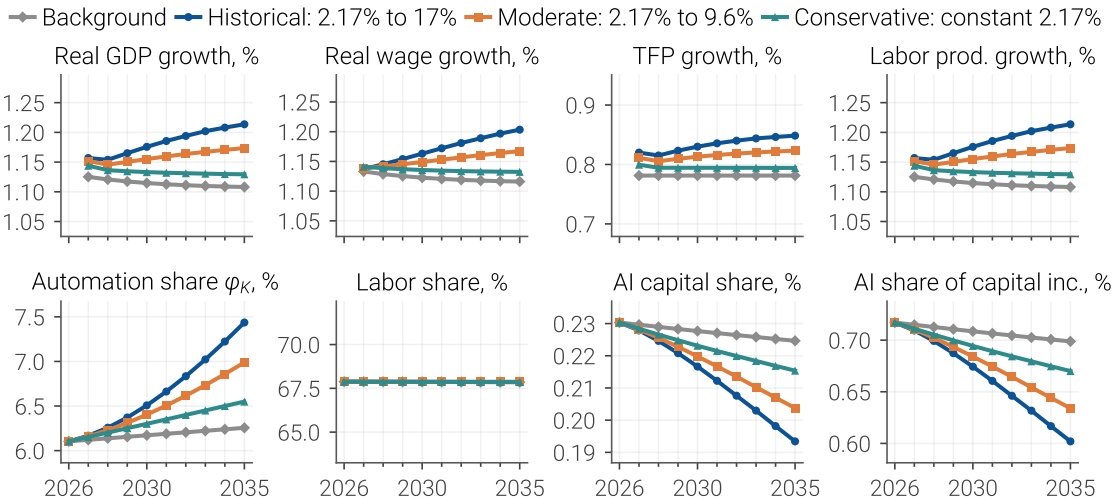
Small direct price effect: cheaper AI capital helps mainly through larger models

Projections: Four Drivers

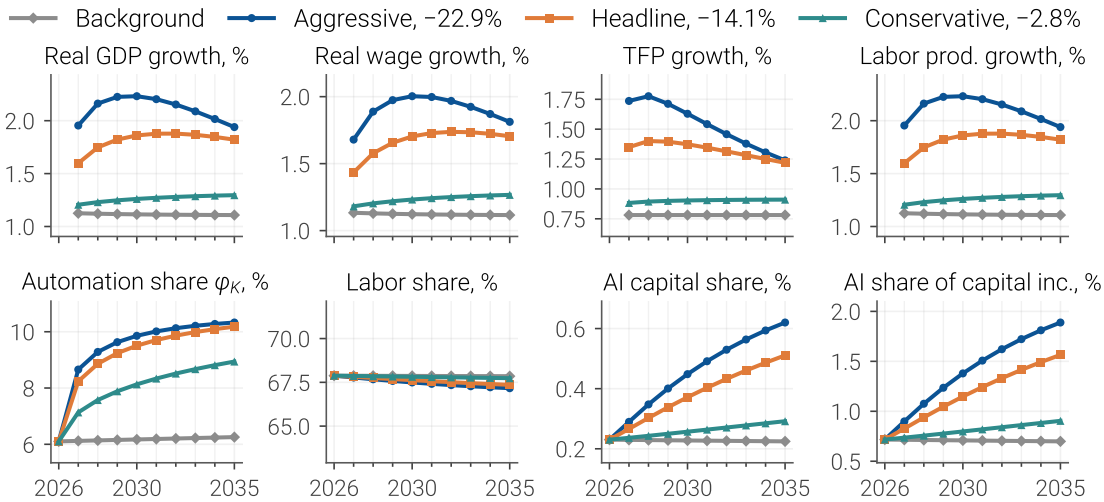
	Data	Estimate	Projection assumption, 2026–35
Compute prices	Our chip prices, 2008–25; token prices, 2025–26	Chips −17% a year (95% CI 7–25%); tokens −2.17% (nominal)	From 2.17% toward a long-run rate, halfway every 5 years: historical 17%, moderate 9.6%, conservative 2.17%
Adoption frictions	Calibrated frictions by occupation, 2025–26	−14.1% a year (s.e. 4.7%)	Headline 14.1%; conservative 2.8%, aggressive 22.9% (80% band)
Algorithmic progress	ECI law: 93 models, 2023–26	+10.4 ECI points a year at a given size	Benchmark 10.4; moderate 8.4; conservative 5.9
Scaling	Monthly ECI leaders, 2024–26	Size +41% a year (95% CI −32% to +191%): about +1 ECI point a year	Starts at +50% a year, slows with a 10-year half-life

Compute-price and adoption-friction rates are annual log changes; scaling rates are percent changes.
Compute-price paths are set at the 2026 interest rate; in the runs the price also moves with the interest rate.
Scenario ranges, not forecast intervals. Projections provisional.

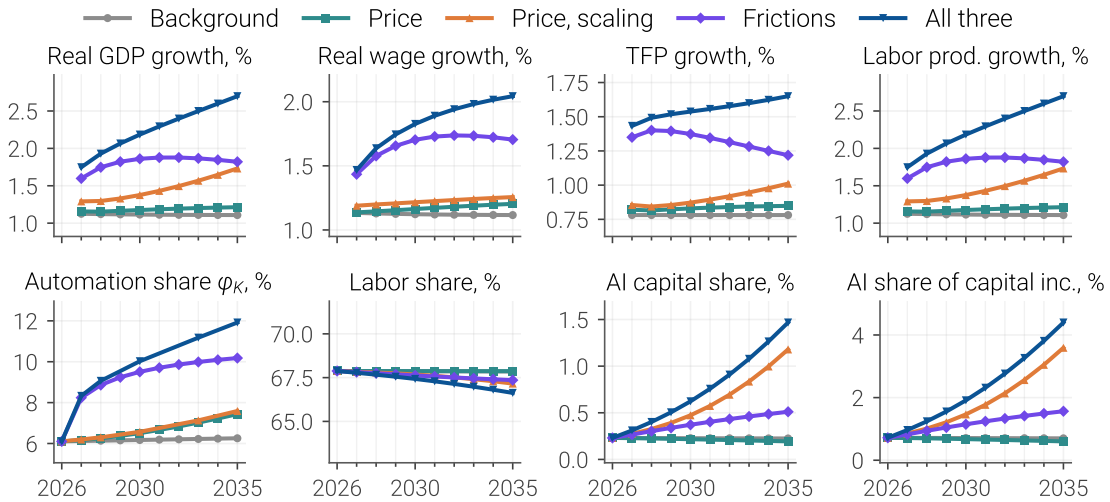
Compute Price Scenarios (fixed AI technology and adoption frictions)



Scenarios for Adoption Frictions (fixed compute prices and AI technology)

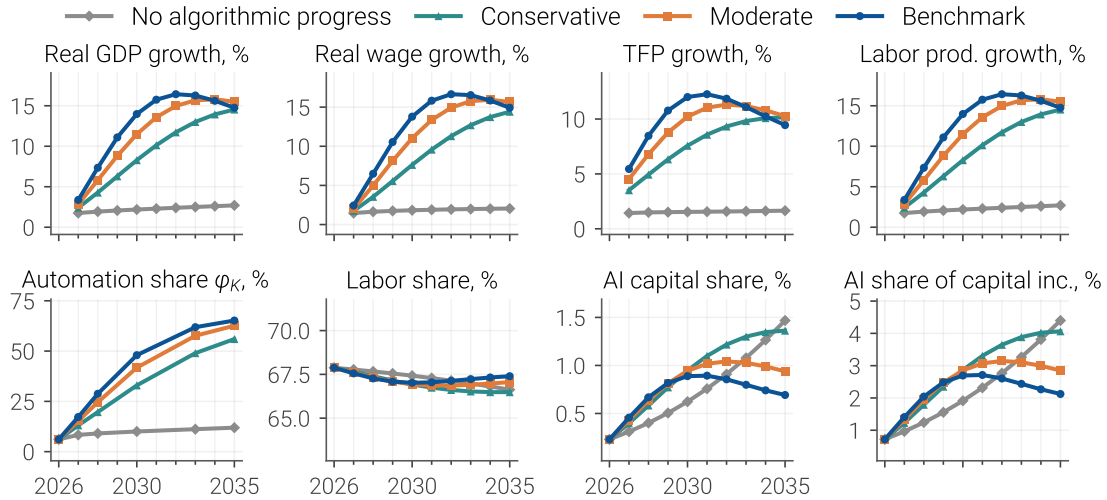


Comparing the Channels with Scaling (historical price decline, headline frictions)

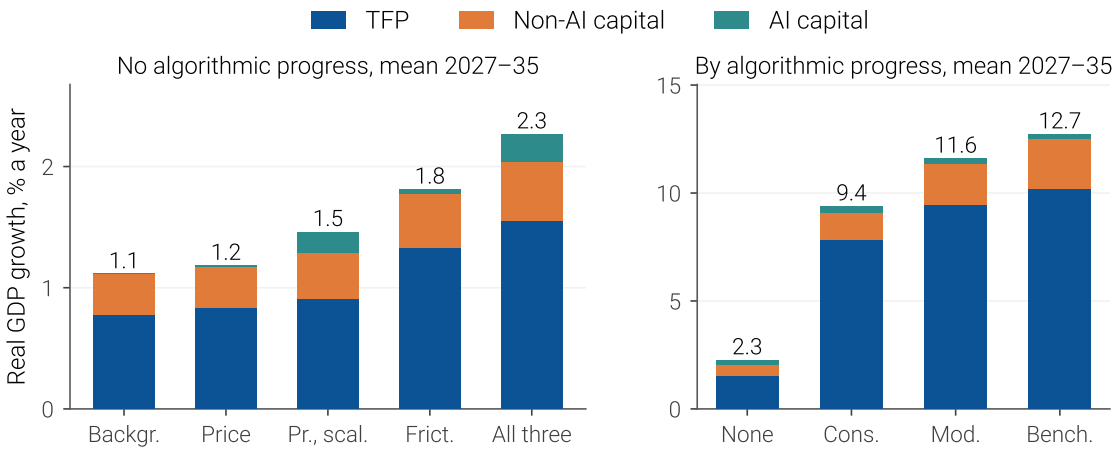


Different price scenarios | Adoption frictions | Real GDP gains by 2035

Algorithmic Progress Scenarios (historical price decline, headline frictions)



Growth Accounting (historical price decline, headline frictions)



Why AI capital's share stays small

Next: AI and Work in Developing Countries

Countries differ in three ways our framework captures:

- Task mix, set by sectoral productivity and demand
- Labor costs task by task, set by the supply of skills
- Cost of computing, which sets the cost of AI

Next: AI and Work in Developing Countries

Countries differ in three ways our framework captures:

- Task mix, set by sectoral productivity and demand
- Labor costs task by task, set by the supply of skills
- Cost of computing, which sets the cost of AI

By country: automation paths and the AI deployment gap: impact without vs. with frictions

Conclusion

Theory: firms choose which AI model to use for each task

- Larger models can perform more tasks, but require more compute
- Firms may choose smaller models to reduce the cost of automation

Data: worker evaluations of LLM outputs on workplace tasks

- Measure task duration, output quality, and the compute used by each model
- Match AI adoption across occupations to estimate adoption costs

Quantitative exercise:

- Today, compute costs limit automation mainly through firms' choice of model
- Algorithmic progress drives automation and growth; cheaper compute alone adds little
- Growth comes mainly from TFP, with a small contribution from AI capital accumulation

Next: countries that differ in task mix, labor costs and the cost of computing

The Dynamics of Comparative Advantage?

"[I]n the short term I agree with arguments that comparative advantage will continue to keep humans relevant […]. In fact, even if AI can do 100% of things better than humans, but it remains inefficient or expensive at some tasks, [...] then the logic of comparative advantage continues to apply. [...] However, I do think in the long run AI will become so broadly effective and so cheap that this will no longer apply."

Amodei (2024), *Machines of Loving Grace*

"I thought when we got to GPT-4, which was back in 2023, that very quickly after that there was going to be much more disruption, software businesses up for grabs right away, than it turned out to be. [...] I think I was wrong about a few things, but one in terms of the speed: the economy just has so much inertia."

Sam Altman (2026), quoted in *The New Yorker*

Prior Work

Task-based automation: comparative advantage and aggregate dynamics

Zeira (1998); Autor et al. (2003); Acemoglu & Autor (2011); Aghion et al. (2017);
Acemoglu & Restrepo (2018, 2019); Humlum (2019); Acemoglu & Restrepo (2022); Korinek & Suh (2024);
Acemoglu (2025); Hampole et al. (2025); Althoff & Reichardt (2026); Freund & Mann (2026);
Jones & Tonetti (2026); Lashkari, Li, Qiu & Thompson (2026); Lindenlaub et al. (2026)

Micro evidence on AI: task exposure, workplace adoption, and productivity

Brynjolfsson & Mitchell (2017); Brynjolfsson et al. (2018); Felten et al. (2018); Brynjolfsson, Rock & Syverson (2021);
Felten et al. (2021); Dell'Acqua et al. (2023); Peng et al. (2023); Cui et al. (2024); Eloundou et al. (2024); Fleming et al. (2024);
McElheran et al. (2024); Svanberg et al. (2024); Brynjolfsson, Li & Raymond (2025); Cunningham et al. (METR, 2026);
Mertens et al. (2026), Crashing Waves vs. Rising Tides

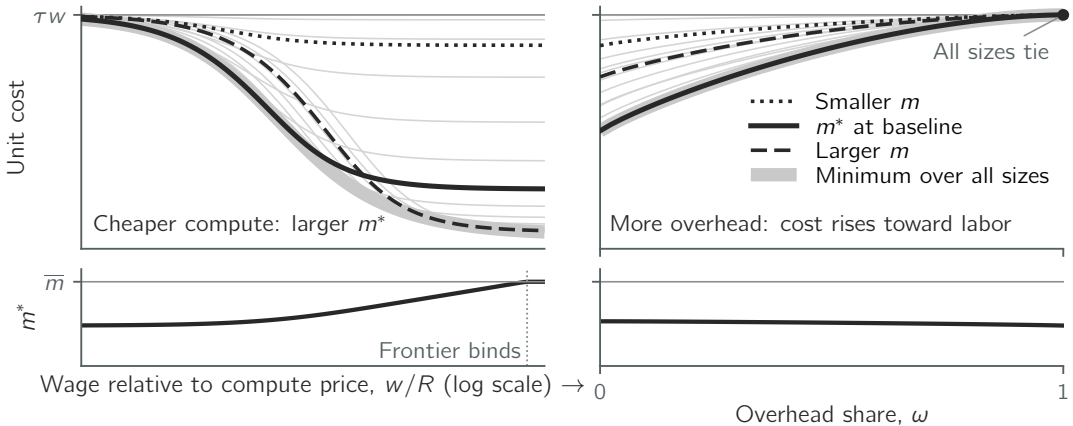
AI forecasts: capabilities, compute costs, and economic scenarios

Kaplan et al. (OpenAI, 2020); Hoffmann et al. (DeepMind, 2022); Ho et al. (Epoch AI, 2024);
Erdil et al. (Epoch AI, 2025); Gundlach et al. (2025); Kwa et al. (METR, 2025); Korinek et al. (Anthropic, 2026)

Machine Choice as Costs Change

Lower envelope: $c(w, R) \equiv \min_{m \in M} c(w, R; m)$

By size



Back

Model Choice and Automation

Automated share with the chosen machine m^* vs. the frontier $\overline{m}$, at the same prices:

$$\varphi_K(m^*) = \overbrace{\underbrace{(1 - H(q^*; \overline{m}))}_{\text{frontier}} \times \underbrace{\frac{1 - H(q^*; m^*)}{1 - H(q^*; \overline{m})}}_{\text{model choice}}}^{\text{capability}} \times \overbrace{\underbrace{G(\gamma^*; \overline{m})}_{\text{frontier}} \times \underbrace{\frac{G(\gamma^*; m^*)}{G(\gamma^*; \overline{m})}}_{\text{model choice}}}^{\text{conditional comparative advantage}}$$

- Smaller models lose capability but gain conditional comparative advantage
- The net effect on automation depends on which change is larger
- Cheaper compute weakly raises G at any fixed model

A Closed-Form Benchmark

Log-linear benchmark: mean quality and mean log compute requirement:

$$Q(m,\tau) \;=\; b_0^q + b_m^q \log\big(m\, D(m)\big) - b_\tau^q \log \tau \qquad \Gamma(m,\tau) \;=\; b_0^\gamma + b_m^\gamma \log m$$

- Step capability: no rating dispersion, so every variety is capable iff $Q(m,\tau) \geq q^*$

Model choice: the smallest capable size $\underline{m}_j$, if the frontier offers it

Smooth capability | Balanced growth

$$\underline{m}_j = e^{(q^* - b_0^q)/((1+\beta_D)b_m^q)}\; \tau_j^{\,b_\tau^q/((1+\beta_D)b_m^q)}$$

Automation share and the race between costs and capabilities:

G: c.d.f. of u_γ

$$\varphi_{Kj,t} = \mathbb{I}\left\{\underline{m}_j \leq \overline{m}_t\right\} \times G\left(\log\left((1-\omega_{o,t})\,\tfrac{w_t}{R_t}\right) + \log\tau_j - b_m^\gamma \log \underline{m}_j - b_0^\gamma\right)$$

- Cheaper compute ($w_t/R_t \uparrow$) raises automation on every frontier-feasible task
- Expanded frontier ($\overline{m}_t \uparrow$) only automates the frontier-constrained tasks
- Feasible tasks within an occupation: automation rises with τ_j iff $1 > b_m^\gamma b_\tau^q/((1+\beta_D)b_m^q)$

Calibrated Parameters

Parameter	Symbol	Value	Target / estimation	Data source	Details
From Literature					
Task EoS	λ	0.50	—	Humlum (2019)	—
Occupation EoS	ε	1.34	—	Caunedo & Keller (2021)	—
Frontier model size, bn. active (2025, 2026)	$\overline{m}$	(135, 202)	—	Epoch AI, ECI data	—
Calibrated					
Task weights	$\boldsymbol{\mu}_o$	Vector	Within-occupation task frequencies	Lashkari et al. (2026)	—
Occupation weights	$\boldsymbol{\xi}_o$	Vector	Occupational wage-bill shares	OEWS	—
Adoption frictions	$\boldsymbol{\omega}_o$	Vector	Time saved by occupation group (2025, 2026)	Bick et al. (2026)	Fit
Adoption data correction (2026)	h	49.5	Level of time saved (2026)	Bick et al. (2026)	Fit
Adoption friction trend	$\Delta \log \boldsymbol{\omega}$	−0.14	Change in frictions by group (2025, 2026)	Bick et al. (2026)	Visual
Compute price (2025)	$\log(R/w)$	4.15	AI spending share (2025, 2026)	Various Details	(0.048%, 0.229%)
Compute-price wedge (2025, 2026)	ω_R	(3.50, 3.32)	Adoption and model choice (2025, 2026)	Bick et al., OpenRouter	—
Spread of capability	κ_H	0.58	Frontier cost ratio (2025, 2026)	OpenRouter Details	(×1.47, ×1.66)
Estimated					
Mean grade and compute	$Q(\cdot), T(\cdot), P(\cdot)$	Prior slides	Grade and cost laws	MIT FutureTech survey	Grade Cost
Spread of compute cost	κ_G	0.92	Cost-law residuals, maximum likelihood	MIT FutureTech survey	Fit

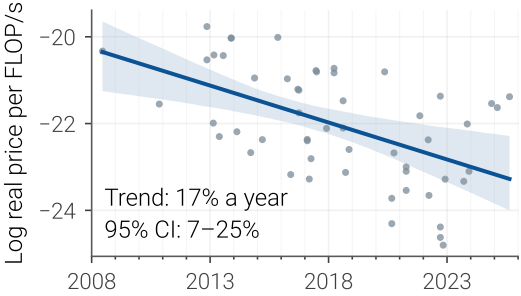
Scenarios for Compute Prices

Historical trend: compute prices fell fast, over 20% a year

Del Sozzo et al. (2026); Lashkari, Li, Qiu & Thompson (2026)

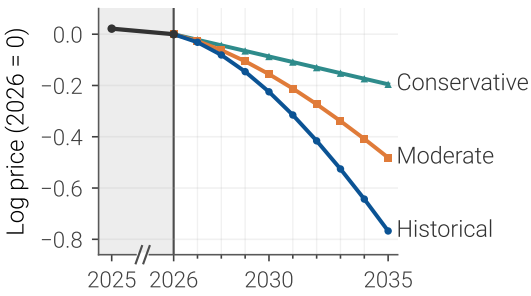
Slowdown in 2025–26: an OpenRouter token price index fell 2.17% a year

Our chip prices, 2008–25 (vector FP32)



Details

Rental price of compute, 2025–35



Details

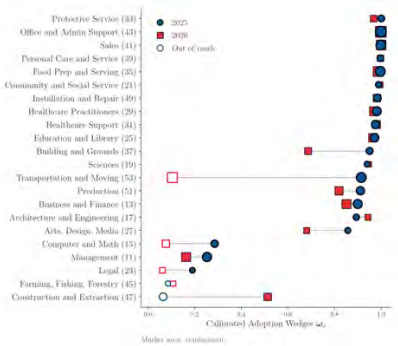
Back

Scenarios for Adoption Frictions

2025–26: an annual decline of 14.1% in adoption frictions

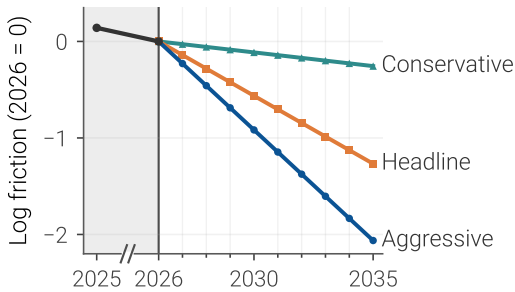
Three scenarios: headline 14.1%; conservative 2.8%, aggressive 22.9% (80% confidence interval)

Calibrated frictions, 2025 and 2026



Trend estimate

Friction scenarios, 2025–35



Back

Model Scaling and Scenarios for Algorithmic Progress

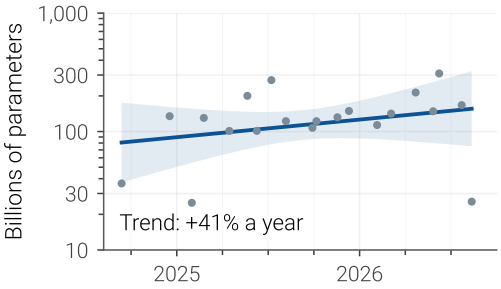
Frontier models: the highest-ECI model each month grew 41% a year in size, 2024–26

Capability gain: $\beta_M \times \log(1.41) = 2.85 \times 0.34 \approx 1.0$ ECI points a year

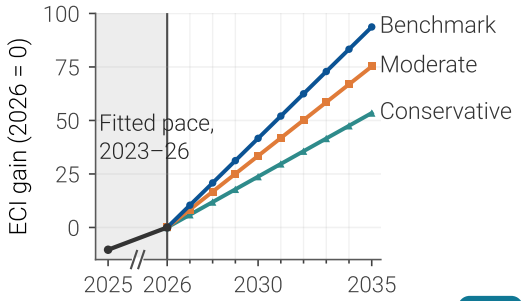
Algorithmic progress: Benchmark 10.4 ECI points a year, the 2023–26 pace at a given size

- Moderate 8.4, its 90% lower bound; Conservative 5.9, 75% lower bound of the slowest two-year pace

Monthly ECI leaders, 2024–26 Details



ECI gain at a given size, 2025–35 Details



Back

The ECI Law: Model Size and Release Date

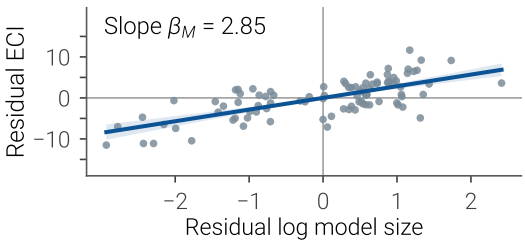
ECI law: OLS, one row per model (93, 2023–26); m: active parameters, t: release date

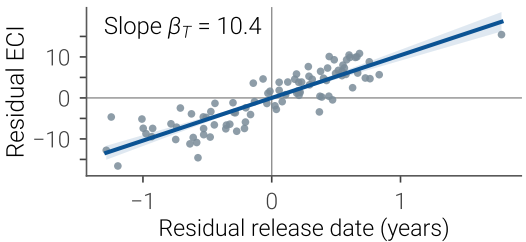
Epoch AI (2026)

$$\mathrm{ECI}(m, t) = \beta_M \log m + \beta_T\, t + \text{reasoning effect} + \text{constant} + \text{error}$$

Model scaling: at a given date, doubling model size adds $\beta_M \log 2 = 2.0$ points

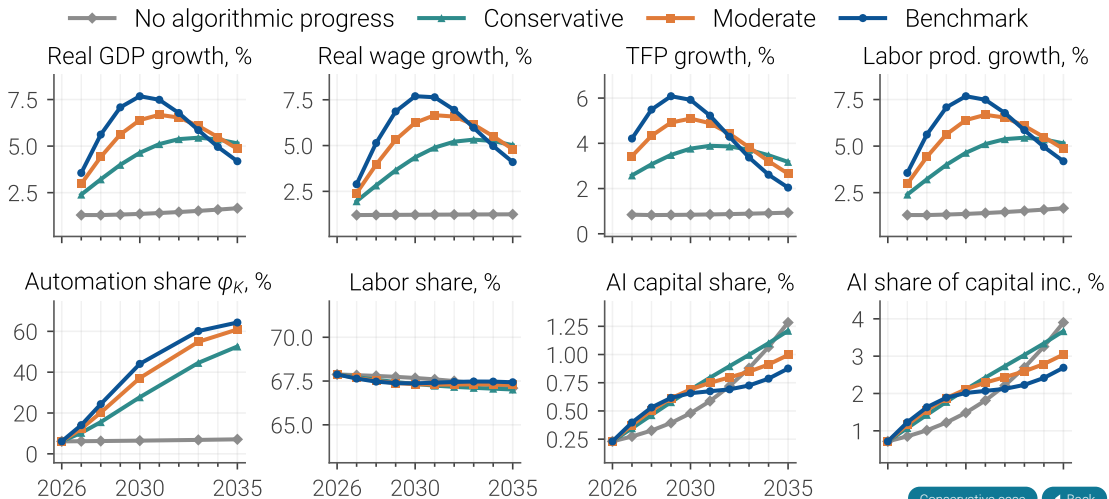
Algorithmic progress: at a given size, ECI rises $\beta_T = 10.4$ points a year





Details Back

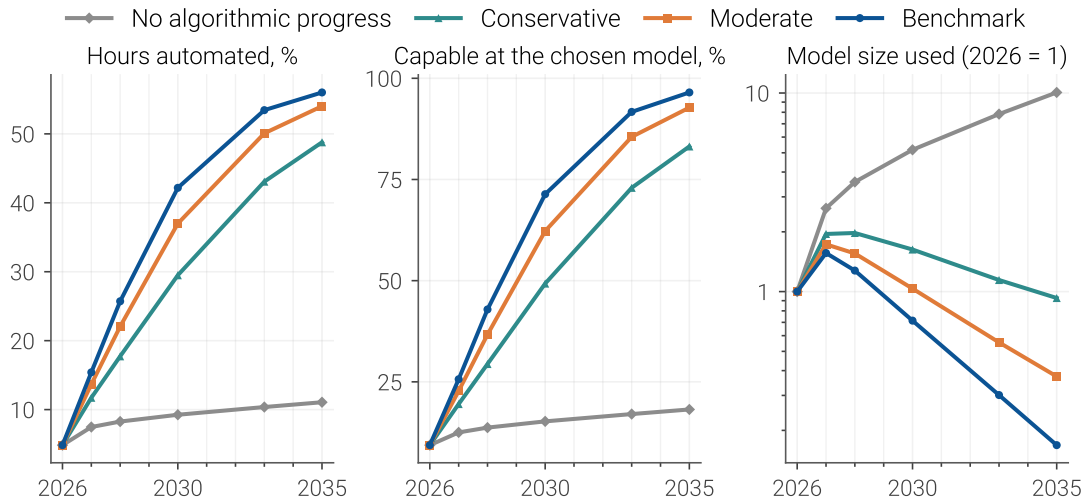
Algorithmic Progress Scenarios (moderate price decline, fixed frictions)



Conservative case

Back

How Algorithmic Progress Automates (historical price decline, headline frictions)



Machine Choice: Details

Saving per capable variety:

$$\zeta(\gamma^*; m) \equiv \int_0^{\gamma^*} \left[1 - \left((1-\omega)\frac{\gamma}{\gamma^*} + \omega\right)^{1-\lambda}\right] dG(\gamma; m)$$

Same product, per automated variety:

$$\left(1 - H(q^*; m)\right) \zeta(\gamma^*; m) = \varphi_K(m)\left[1 - \left(\frac{\widetilde{c}_K}{\tau w}\right)^{1-\lambda}\right]$$

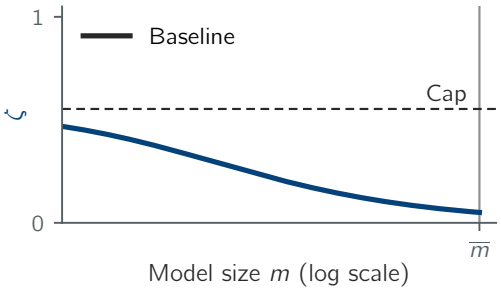
- $\widetilde{c}_K$: CES mean cost over automated varieties; bracket: saving per automated variety

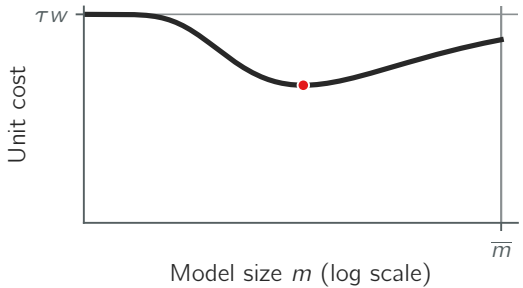
Firm level | Bundle cost

◂ Back

Machine Choice as Compute Gets Cheaper: By Model Size

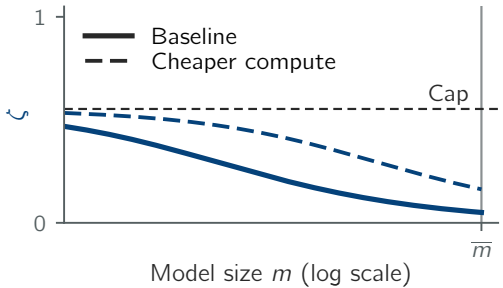
Cheaper compute favors larger models

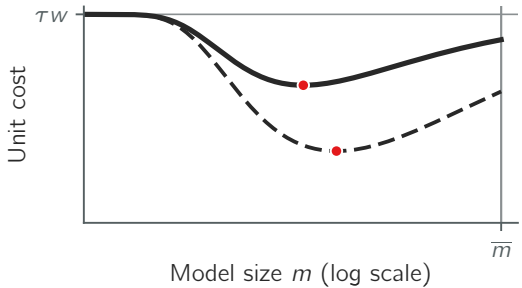




Machine Choice as Compute Gets Cheaper: By Model Size

Cheaper compute favors larger models



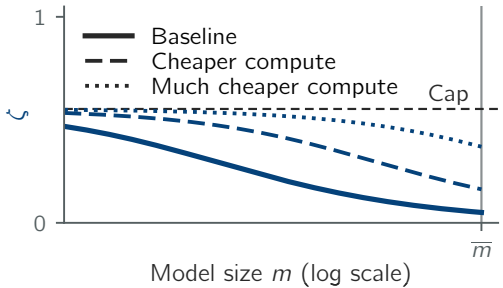


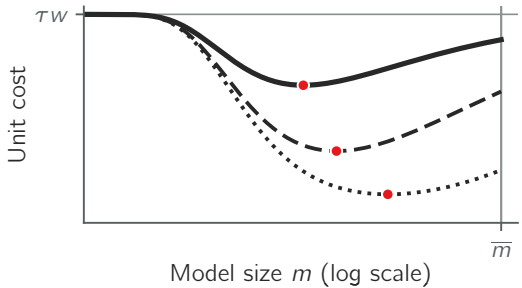
- ζ rises toward its cap at every size; capability unchanged

Back

Machine Choice as Compute Gets Cheaper: By Model Size

Cheaper compute favors larger models





- ζ rises toward its cap at every size; capability unchanged

Back

Selected Machine-Bundle Cost, Overhead Share, and Task Cost

Selected CES average of the machine bundle below a threshold z:

$$\widetilde{c}_{K,j}(z; w, R, m, \omega_o) \equiv \left(\frac{\int_0^z (u\,R + \omega_o \tau_j\, w)^{1-\lambda}\; dG_j(u; m)}{G_j(z; m)} \right)^{\frac{1}{1-\lambda}}$$

- Weakly rises with z: more compute per added variety, same overhead wage bill $\omega_o \tau_j w$ on every one
- Defined if $(1 - H_j(q^*; m))\, G_j(z; m) > 0$

Selected Machine-Bundle Cost, Overhead Share, and Task Cost

Selected CES average of the machine bundle below a threshold z:

$$\widetilde{c}_{K,j}(z; w, R, m, \omega_o) \equiv \left(\frac{\int_0^z (u R + \omega_o \tau_j w)^{1-\lambda} \, dG_j(u; m)}{G_j(z; m)} \right)^{\frac{1}{1-\lambda}}$$

- Weakly rises with z: more compute per added variety, same overhead wage bill $\omega_o \tau_j w$ on every one
- Defined if $(1 - H_j(q^*; m)) \, G_j(z; m) > 0$

Overhead's share of the machine bundle and the resulting cost shares:

$$\psi_j \equiv \frac{\omega_o \tau_j \, w \int_0^{\gamma_j^*} (uR + \omega_o \tau_j w)^{-\lambda} \, dG_j}{\int_0^{\gamma_j^*} (uR + \omega_o \tau_j w)^{1-\lambda} \, dG_j}, \qquad \alpha_{L,j} = \varphi_{L,j} \left(\frac{\tau_j w}{c_j} \right)^{1-\lambda} + \psi_j \varphi_{K,j} \left(\frac{\widetilde{c}_{K,j}}{c_j} \right)^{1-\lambda}, \quad \alpha_{K,j} = 1 - \alpha_{L,j}$$

- Task share vs. cost share: automated varieties still use overhead labor; $\omega_o = 0 \Rightarrow \psi_j = 0$
- Free compute still leaves the overhead: $\omega_o \tau_j \leq c_j / w \leq \tau_j$

back

Firm-Level Cost Function

Aggregation over tasks j of occupation o: each task picks its own model:

$$c_o\left(w, R; M, \omega_o\right)^{1-\lambda} = \sum_j \mu_{j,o} \min_{m \in M} c_j\left(w, R; m, \omega_o\right)^{1-\lambda} = \varphi_{L,o}\left(\widetilde{\tau}_o\, w\right)^{1-\lambda} + \varphi_{K,o}\, \widetilde{c}_{K,o}{}^{1-\lambda}$$

- Task shares $\varphi_{a,o} \equiv \sum_j \mu_{j,o}\varphi_{a,j}$, $a \in \{L, K\}$; $\widetilde{\tau}_o$, $\widetilde{c}_{K,o}$: task-share-weighted CES means of τ_j and $\widetilde{c}_{K,j}$

Firm-Level Cost Function

Aggregation over tasks j of occupation o: each task picks its own model:

$$c_o\,(w, R; M, \omega_o)^{1-\lambda} = \sum_j \mu_{j,o} \min_{m \in M} c_j\,(w, R; m, \omega_o)^{1-\lambda} = \varphi_{L,o}\,(\widetilde{\tau}_o\, w)^{1-\lambda} + \varphi_{K,o}\,\widetilde{c}_{K,o}{}^{1-\lambda}$$

- Task shares $\varphi_{a,o} \equiv \sum_j \mu_{j,o}\varphi_{a,j}$, $a \in \{L, K\}$; $\widetilde{\tau}_o$, $\widetilde{c}_{K,o}$: task-share-weighted CES means of τ_j and $\widetilde{c}_{K,j}$

Cost shares (Shephard): labor is paid for its own tasks *and* for the overhead of automated ones

$$\alpha_{L,o} = \varphi_{L,o}\left(\frac{\widetilde{\tau}_o\, w}{c_o}\right)^{1-\lambda} + \psi_o\,\varphi_{K,o}\left(\frac{\widetilde{c}_{K,o}}{c_o}\right)^{1-\lambda}, \qquad \alpha_{K,o} = (1-\psi_o)\,\varphi_{K,o}\left(\frac{\widetilde{c}_{K,o}}{c_o}\right)^{1-\lambda}$$

- ψ_o: cost-weighted average of the task shares ψ_j; within a task, $\alpha_{K,j} < \varphi_{K,j}$ whenever $\psi_j > 0$

Firm-Level Cost Function

Aggregation over tasks j of occupation o: each task picks its own model:

$$c_o\,(w, R; M, \omega_o)^{1-\lambda} = \sum_j \mu_{j,o} \min_{m \in M} c_j\,(w, R; m, \omega_o)^{1-\lambda} = \varphi_{L,o}\,(\widetilde{\tau}_o\, w)^{1-\lambda} + \varphi_{K,o}\,\widetilde{c}_{K,o}{}^{1-\lambda}$$

- Task shares $\varphi_{a,o} \equiv \sum_j \mu_{j,o}\varphi_{a,j}$, $a \in \{L, K\}$; $\widetilde{\tau}_o$, $\widetilde{c}_{K,o}$: task-share-weighted CES means of τ_j and $\widetilde{c}_{K,j}$

Cost shares (Shephard): labor is paid for its own tasks *and* for the overhead of automated ones

$$\alpha_{L,o} = \varphi_{L,o}\left(\frac{\widetilde{\tau}_o\, w}{c_o}\right)^{1-\lambda} + \psi_o\,\varphi_{K,o}\left(\frac{\widetilde{c}_{K,o}}{c_o}\right)^{1-\lambda}, \qquad \alpha_{K,o} = (1-\psi_o)\,\varphi_{K,o}\left(\frac{\widetilde{c}_{K,o}}{c_o}\right)^{1-\lambda}$$

- ψ_o: cost-weighted average of the task shares ψ_j; within a task, $\alpha_{K,j} < \varphi_{K,j}$ whenever $\psi_j > 0$
- CES-like identity in task units: direct labor vs. the automated composite, not primal

CES form Elasticity

back

Factor Payments and the CES-Like Identity

Factor payments at the optimal model choices $\boldsymbol{m}_o^*(w, R; \omega_o)$ (Shephard + envelope):

$w\,\ell_o = \alpha_{L,o}\,c_o\,y_o$ $\qquad$ $R\,k_o = \alpha_{K,o}\,c_o\,y_o$ $\qquad$ $\ell_o = \ell_{L,o} + \ell_{K,o}$ (direct task labor + overhead)

Factor Payments and the CES-Like Identity

Factor payments at the optimal model choices $\boldsymbol{m}_o^*(w, R; \omega_o)$ (Shephard + envelope):

$$w\,\ell_o = \alpha_{L,o}\,c_o\,y_o \qquad R\,k_o = \alpha_{K,o}\,c_o\,y_o \qquad \ell_o = \ell_{L,o} + \ell_{K,o} \text{ (direct task labor + overhead)}$$

Two-term conditional cost and the aggregate overhead share:

$$c_o\,(w, R; \boldsymbol{m}_o, \omega_o)^{1-\lambda} = \varphi_{L,o}\,(\widetilde{\tau}_o\,w)^{1-\lambda} + \varphi_{K,o}\,\widetilde{c}_{K,o}{}^{1-\lambda}, \qquad \psi_o \equiv \frac{\sum_j \mu_{j,o}\varphi_{K,j}\psi_j\widetilde{c}_{K,j}{}^{1-\lambda}}{\sum_j \mu_{j,o}\varphi_{K,j}\widetilde{c}_{K,j}{}^{1-\lambda}}$$

- Interior Hicksian system $\Rightarrow$ CES-like identity in task units: direct labor vs. the automated composite
- Composite: fixed proportions per task, not in aggregate; prices shift the mix across tasks

Factor Payments and the CES-Like Identity

Factor payments at the optimal model choices $\boldsymbol{m}_o^*(w, R; \omega_o)$ (Shephard + envelope):

$$w\,\ell_o = \alpha_{L,o}\,c_o\,y_o \qquad R\,k_o = \alpha_{K,o}\,c_o\,y_o \qquad \ell_o = \ell_{L,o} + \ell_{K,o} \text{ (direct task labor + overhead)}$$

Two-term conditional cost and the aggregate overhead share:

$$c_o\,(w, R; \boldsymbol{m}_o, \omega_o)^{1-\lambda} = \varphi_{L,o}\,(\widetilde{\tau}_o\,w)^{1-\lambda} + \varphi_{K,o}\,\widetilde{c}_{K,o}{}^{1-\lambda}, \qquad \psi_o \equiv \frac{\sum_j \mu_{j,o}\varphi_{K,j}\psi_j\widetilde{c}_{K,j}{}^{1-\lambda}}{\sum_j \mu_{j,o}\varphi_{K,j}\widetilde{c}_{K,j}{}^{1-\lambda}}$$

- Interior Hicksian system ⇒ CES-like identity in task units: direct labor vs. the automated composite
- Composite: fixed proportions per task, not in aggregate; prices shift the mix across tasks
- Not primal: $\varphi_{a,o}$, $\widetilde{\tau}_o$, $\widetilde{c}_{K,o}$ move with (w, R) through the cutoffs γ_j^* and the model choices $\boldsymbol{m}_o$

back

Elasticity of Substitution: Details

Elasticity of substitution of occupation o (o suppressed), three nonnegative margins:

$$\sigma = \underbrace{\lambda \frac{\mathrm{Var}_\Pi(s_K)}{\alpha_K \alpha_L}}_{\text{intensive}} + \underbrace{\frac{\sum_j (1-\omega_o)\,\pi_j^*}{\alpha_K \alpha_L}}_{\text{extensive (cutoffs)}} + \underbrace{\frac{w}{\alpha_K c_w}\, c_{wm} c_{mm}^{-1} c_{mw}}_{\text{model choice}} \;\geq\; 0$$

Objects: Π weights pairs (j, ν) by expenditure; $s_K(j,\nu) \in [0, 1-\omega_o]$ is their AI capital share

- $\pi_j^* \equiv \mu_{j,o}(\tau_j w/c)^{1-\lambda}\, \varphi_{K,j}\, \mathcal{E}_{G,j}(\gamma_j^*; m_j^*)$: expenditure density at the automation margin

Elasticity of Substitution: Details

Elasticity of substitution of occupation o (o suppressed), three nonnegative margins:

$$\sigma = \underbrace{\lambda \frac{\mathrm{Var}_\Pi(s_K)}{\alpha_K \alpha_L}}_{\text{intensive}} + \underbrace{\frac{\sum_j (1-\omega_o)\,\pi_j^*}{\alpha_K \alpha_L}}_{\text{extensive (cutoffs)}} + \underbrace{\frac{w}{\alpha_K c_w} c_{wm} c_{mm}^{-1} c_{mw}}_{\text{model choice}} \geq 0$$

Objects: Π weights pairs (j, ν) by expenditure; $s_K(j,\nu) \in [0, 1-\omega_o]$ is their AI capital share

- $\pi_j^* \equiv \mu_{j,o}(\tau_j w/c)^{1-\lambda}\, \varphi_{K,j}\, \mathcal{E}_{G,j}(\gamma_j^*; m_j^*)$: expenditure density at the automation margin

Zero frictions: $s_K \in \{0,1\}$ with $\mathbb{E}_\Pi[s_K] = \alpha_K \Rightarrow \mathrm{Var}_\Pi(s_K) = \alpha_K \alpha_L$: the floor $\sigma \geq \lambda$ returns

Elasticity of Substitution: Details

Elasticity of substitution of occupation o (o suppressed), three nonnegative margins:

$$\sigma = \lambda \underbrace{\frac{\mathrm{Var}_\Pi\left(s_K\right)}{\alpha_K \alpha_L}}_{\text{intensive}} + \underbrace{\frac{\sum_j \left(1-\omega_o\right)\pi_j^*}{\alpha_K \alpha_L}}_{\text{extensive (cutoffs)}} + \underbrace{\frac{w}{\alpha_K c_w}\, c_{wm} c_{mm}^{-1} c_{mw}}_{\text{model choice}} \;\geq\; 0$$

Objects: Π weights pairs (j,ν) by expenditure; $s_K(j,\nu) \in [0, 1-\omega_o]$ is their AI capital share

- $\pi_j^* \equiv \mu_{j,o}(\tau_j w/c)^{1-\lambda}\, \varphi_{K,j}\, \mathcal{E}_{G,j}(\gamma_j^*; m_j^*)$: expenditure density at the automation margin

Zero frictions: $s_K \in \{0,1\}$ with $\mathbb{E}_\Pi[s_K] = \alpha_K \Rightarrow \mathrm{Var}_\Pi(s_K) = \alpha_K \alpha_L$: the floor $\sigma \geq \lambda$ returns

Overhead attenuates both margins: σ can fall below λ

- Automated tasks use both factors: $\mathrm{Var}_\Pi(s_K) < \alpha_K \alpha_L$
- A task switching to the bundle moves its AI capital share by only $1-\omega_o$

Elasticity of Substitution: Details

Elasticity of substitution of occupation o (o suppressed), three nonnegative margins:

$$\sigma = \underbrace{\lambda \frac{\mathrm{Var}_\Pi(s_K)}{\alpha_K \alpha_L}}_{\text{intensive}} + \underbrace{\frac{\sum_j (1-\omega_o)\,\pi_j^*}{\alpha_K \alpha_L}}_{\text{extensive (cutoffs)}} + \underbrace{\frac{w}{\alpha_K c_w}\, c_{wm} c_{mm}^{-1} c_{mw}}_{\text{model choice}} \;\geq\; 0$$

Objects: Π weights pairs (j,ν) by expenditure; $s_K(j,\nu) \in [0, 1-\omega_o]$ is their AI capital share

- $\pi_j^* \equiv \mu_{j,o}(\tau_j w/c)^{1-\lambda}\, \varphi_{K,j}\, \mathcal{E}_{G,j}(\gamma_j^*; m_j^*)$: expenditure density at the automation margin

Zero frictions: $s_K \in \{0,1\}$ with $\mathbb{E}_\Pi[s_K] = \alpha_K \Rightarrow \mathrm{Var}_\Pi(s_K) = \alpha_K \alpha_L$: the floor $\sigma \geq \lambda$ returns

Overhead attenuates both margins: σ can fall below λ

- Automated tasks use both factors: $\mathrm{Var}_\Pi(s_K) < \alpha_K \alpha_L$
- A task switching to the bundle moves its AI capital share by only $1-\omega_o$

Cheap compute: at a fixed model $\sigma \to 0$ as bundles converge to their overhead wages

- On a rich menu, model re-optimization alone keeps σ away from zero

back

Model Choice: the Race between Costs and Capabilities

The saving is the product of the two conditions for automation:

$$(1 - H_j(q^*; m)) \int_0^{\gamma_j^*} \left[1 - \left((1-\omega_o)\frac{\gamma}{\gamma_j^*} + \omega_o\right)^{1-\lambda}\right] dG_j(\gamma; m) = \underbrace{\varphi_{K,j}}_{\text{automated share}} \times \underbrace{\left[1 - \left(\widetilde{c}_{K,j}/\tau_j w\right)^{1-\lambda}\right]}_{\text{saving per automated variety}}$$

Model Choice: the Race between Costs and Capabilities

The saving is the product of the two conditions for automation:

$$(1 - H_j(q^*; m)) \int_0^{\gamma_j^*} \left[1 - \left((1-\omega_o)\frac{\gamma}{\gamma_j^*} + \omega_o\right)^{1-\lambda}\right] dG_j(\gamma; m) \;=\; \underbrace{\varphi_{K,j}}_{\text{automated share}} \;\times\; \underbrace{\left[1 - \left(\widetilde{c}_{K,j}/\tau_j w\right)^{1-\lambda}\right]}_{\text{saving per automated variety}}$$

A larger model moves the two factors in opposite directions

- Capability share $1 - H_j(q^*; m)$ $\uparrow$: more of the task's varieties become automatable
- Requirements $G_j(\cdot; m)$ shifts up: fewer capable varieties clear γ_j^*, and those that do save less

Model Choice: the Race between Costs and Capabilities

The saving is the product of the two conditions for automation:

$$(1 - H_j(q^*; m)) \int_0^{\gamma_j^*} \left[1 - \left((1-\omega_o)\frac{\gamma}{\gamma_j^*} + \omega_o\right)^{1-\lambda}\right] dG_j(\gamma; m) \;=\; \underbrace{\varphi_{K,j}}_{\text{automated share}} \;\times\; \underbrace{\left[1 - \left(\widetilde{c}_{K,j}/\tau_j w\right)^{1-\lambda}\right]}_{\text{saving per automated variety}}$$

A larger model moves the two factors in opposite directions

- Capability share $1 - H_j(q^*; m)$ ↑: more of the task's varieties become automatable
- Requirements $G_j(\cdot; m)$ shifts up: fewer capable varieties clear γ_j^*, and those that do save less

Interior optimum along the size menu: proportional capability gain = proportional loss of expected saving

- Low-requirement varieties carry the saving; varieties near the cutoff γ_j^* add almost nothing

Model Choice: the Race between Costs and Capabilities

The saving is the product of the two conditions for automation:

$$(1 - H_j(q^*; m)) \int_0^{\gamma_j^*} \left[1 - \left((1-\omega_o)\frac{\gamma}{\gamma_j^*} + \omega_o\right)^{1-\lambda}\right] dG_j(\gamma; m) \;=\; \underbrace{\varphi_{K,j}}_{\text{automated share}} \;\times\; \underbrace{\left[1 - \left(\tilde{c}_{K,j}/\tau_j w\right)^{1-\lambda}\right]}_{\text{saving per automated variety}}$$

A larger model moves the two factors in opposite directions

- Capability share $1 - H_j(q^*; m) \uparrow$: more of the task's varieties become automatable
- Requirements $G_j(\cdot; m)$ shifts up: fewer capable varieties clear γ_j^*, and those that do save less

Interior optimum along the size menu: proportional capability gain = proportional loss of expected saving

- Low-requirement varieties carry the saving; varieties near the cutoff γ_j^* add almost nothing
- Neither capability nor the automated share identifies the chosen model

what picks the model

Model Choice: the Race between Costs and Capabilities

The saving is the product of the two conditions for automation:

$$(1 - H_j(q^*; m)) \int_0^{\gamma_j^*} \left[1 - \left((1-\omega_o)\frac{\gamma}{\gamma_j^*} + \omega_o\right)^{1-\lambda}\right] dG_j(\gamma; m) \;=\; \underbrace{\varphi_{K,j}}_{\text{automated share}} \;\times\; \underbrace{\left[1 - \left(\tilde{c}_{K,j}/\tau_j w\right)^{1-\lambda}\right]}_{\text{saving per automated variety}}$$

A larger model moves the two factors in opposite directions

- Capability share $1 - H_j(q^*; m)$ ↑: more of the task's varieties become automatable
- Requirements $G_j(\cdot; m)$ shifts up: fewer capable varieties clear γ_j^*, and those that do save less

Interior optimum along the size menu: proportional capability gain = proportional loss of expected saving

- Low-requirement varieties carry the saving; varieties near the cutoff γ_j^* add almost nothing
- Neither capability nor the automated share identifies the chosen model

what picks the model

⇒ the cost-minimizing model can sit below the frontier: firms rationally lag the most capable model

back

Fine-Graining the Benchmark: What Picks the Model

Smooth capability: the gap $\delta_j(m) \equiv H_j(q^*; m) - H_j(q^*; \infty)$ now closes gradually

$$\beta(m) \equiv -\frac{d \log \delta_j(m)}{d \log m} = \underbrace{\frac{h(q^* - Q)}{\delta_j(m)}}_{\text{steepness of capability near the bar}} \times \underbrace{\frac{dQ}{d \log m}}_{\text{grade gain per log size}}$$

Fine-Graining the Benchmark: What Picks the Model

Smooth capability: the gap $\delta_j(m) \equiv H_j(q^*; m) - H_j(q^*; \infty)$ now closes gradually

$$\beta(m) \equiv -\frac{d \log \delta_j(m)}{d \log m} = \underbrace{\frac{h(q^* - Q)}{\delta_j(m)}}_{\text{steepness of capability near the bar}} \times \underbrace{\frac{dQ}{d \log m}}_{\text{grade gain per log size}}$$

Interior optimum: capability gain balances compute inflation

$$\underbrace{\beta(m)\,\delta_j(m)}_{\text{capability gain}} \approx \underbrace{(1-\lambda)\, b_m^{\gamma}\, e^{-(1-\lambda)v} \int^{v} e^{(1-\lambda)u} dG(u)}_{\text{compute loss } (\omega = 0;\ \text{with frictions } \int \to J_P,\ \text{see proof sketch})}$$

Fine-Graining the Benchmark: What Picks the Model

Smooth capability: the gap $\delta_j(m) \equiv H_j(q^*; m) - H_j(q^*; \infty)$ now closes gradually

$$\beta(m) \equiv -\frac{d \log \delta_j(m)}{d \log m} = \underbrace{\frac{h(q^* - Q)}{\delta_j(m)}}_{\text{steepness of capability near the bar}} \times \underbrace{\frac{dQ}{d \log m}}_{\text{grade gain per log size}}$$

Interior optimum: capability gain balances compute inflation

$$\underbrace{\beta(m)\,\delta_j(m)}_{\text{capability gain}} \approx \underbrace{(1-\lambda)\, b_m^{\gamma}\, e^{-(1-\lambda)v} \int^{v} e^{(1-\lambda)u} dG(u)}_{\text{compute loss } (\omega = 0;\ \text{with frictions } \int \to J_P,\ \text{see proof sketch})}$$

- Step H (benchmark): β spikes at the threshold $\Rightarrow$ smallest capable model

Fine-Graining the Benchmark: What Picks the Model

Smooth capability: the gap $\delta_j(m) \equiv H_j(q^*; m) - H_j(q^*; \infty)$ now closes gradually

$$\beta(m) \equiv -\frac{d \log \delta_j(m)}{d \log m} = \underbrace{\frac{h(q^* - Q)}{\delta_j(m)}}_{\text{steepness of capability near the bar}} \times \underbrace{\frac{dQ}{d \log m}}_{\text{grade gain per log size}}$$

Interior optimum: capability gain balances compute inflation

$$\underbrace{\beta(m)\,\delta_j(m)}_{\text{capability gain}} \approx \underbrace{(1-\lambda)\, b_m^{\gamma}\, e^{-(1-\lambda)v} \int^{v} e^{(1-\lambda)u} dG(u)}_{\text{compute loss } (\omega = 0;\ \text{with frictions } \int \to J_P,\ \text{see proof sketch})}$$

- Step H (benchmark): β spikes at the threshold $\Rightarrow$ smallest capable model
- Smooth H: cheap variance below the bar, so firms may sit below it

Fine-Graining the Benchmark: What Picks the Model

Smooth capability: the gap $\delta_j(m) \equiv H_j(q^*; m) - H_j(q^*; \infty)$ now closes gradually

$$\beta(m) \equiv -\frac{d \log \delta_j(m)}{d \log m} = \underbrace{\frac{h(q^* - Q)}{\delta_j(m)}}_{\text{steepness of capability near the bar}} \times \underbrace{\frac{dQ}{d \log m}}_{\text{grade gain per log size}}$$

Interior optimum: capability gain balances compute inflation

$$\underbrace{\beta(m)\,\delta_j(m)}_{\text{capability gain}} \approx \underbrace{(1-\lambda)\, b_m^{\gamma}\, e^{-(1-\lambda)v} \int^{v} e^{(1-\lambda)u} dG(u)}_{\text{compute loss } (\omega = 0;\ \text{with frictions } \int \to J_P,\ \text{see proof sketch})}$$

- Step H (benchmark): β spikes at the threshold $\Rightarrow$ smallest capable model
- Smooth H: cheap variance below the bar, so firms may sit below it

Ahead: the gap closes at an asymptotically constant elasticity β, the engine of growth

back

Market Clearing and the Government Budget

Market clearing in labor, AI capital, and final output:

$$w_t L_t = \overline{\alpha}_L \, Y_t \qquad K_t = \frac{\overline{\alpha}_K \, Y_t}{R_t} + \chi \, \overline{m}_t \, D\left(\overline{m}_t\right) \qquad Y_t = C_t + \frac{K_t}{A_t}$$

- $\overline{\alpha}_K = \sum_o s_o \, \alpha_{K,o}\left(w, R; M_t, \omega_o\right)$ with expenditure weights $s_o \propto \xi_o p_o^{1-\epsilon}$; $\overline{\alpha}_L \equiv 1 - \overline{\alpha}_K$

Market Clearing and the Government Budget

Market clearing in labor, AI capital, and final output:

$$w_t L_t = \overline{\alpha}_L \, Y_t \qquad K_t = \frac{\overline{\alpha}_K \, Y_t}{R_t} + \chi \, \overline{m}_t \, D\left(\overline{m}_t\right) \qquad Y_t = C_t + \frac{K_t}{A_t}$$

- $\overline{\alpha}_K = \sum_o s_o \, \alpha_{K,o}\left(w, R; M_t, \omega_o\right)$ with expenditure weights $s_o \propto \xi_o p_o^{1-\epsilon}$; $\overline{\alpha}_L \equiv 1 - \overline{\alpha}_K$
- Overhead wages are labor income, inside $\overline{\alpha}_L$; only AI capital income goes to the computing sector
- Overhead's share of total cost: $\overline{\alpha}_O \equiv \sum_o s_o \, \psi_o \, \varphi_{K,o} \left(\widetilde{c}_{K,o} / c_o\right)^{1-\lambda}$; no waste term, no wedge revenue

Market Clearing and the Government Budget

Market clearing in labor, AI capital, and final output:

$$w_t L_t = \overline{\alpha}_L Y_t \qquad K_t = \frac{\overline{\alpha}_K Y_t}{R_t} + \chi \overline{m}_t D(\overline{m}_t) \qquad Y_t = C_t + \frac{K_t}{A_t}$$

- $\overline{\alpha}_K = \sum_o s_o \alpha_{K,o}(w, R; M_t, \omega_o)$ with expenditure weights $s_o \propto \xi_o p_o^{1-\epsilon}$; $\overline{\alpha}_L \equiv 1 - \overline{\alpha}_K$
- Overhead wages are labor income, inside $\overline{\alpha}_L$; only AI capital income goes to the computing sector
- Overhead's share of total cost: $\overline{\alpha}_O \equiv \sum_o s_o \psi_o \varphi_{K,o} (\widetilde{c}_{K,o}/c_o)^{1-\lambda}$; no waste term, no wedge revenue

Government taxes lump-sum to fund training; consumption = labor income net of that burden:

$$T_t = R_t \chi \overline{m}_t D(\overline{m}_t) \qquad C_t = w_t L_t - T_t \qquad \frac{\dot{c}_t}{c_t} = r_t - \rho$$

back

Equilibrium Dynamics

Wage growth from the numeraire condition (Shephard within occupations):

$$\frac{\dot{w}_t}{w_t} = \frac{\eta_Z + \overline{\alpha}_{K,t}\,\eta_A + \Phi_t\,\eta_M - \overline{\alpha}_{O,t}\,\eta_{\omega,t}}{\overline{\alpha}_{L,t}} \qquad \Phi_t \equiv -\sum_o s_{o,t}\frac{\partial \log c_o}{\partial \log \overline{m}_t} \geq 0$$

- Φ_t: gain from the growing frontier; only frontier-constrained tasks contribute
- Declining frictions ($\eta_{\omega,t} < 0$) raise wage growth like cheaper compute, weighted by $\overline{\alpha}_{O,t}$

Equilibrium Dynamics

Wage growth from the numeraire condition (Shephard within occupations):

$$\frac{\dot{w}_t}{w_t} = \frac{\eta_Z + \overline{\alpha}_{K,t}\,\eta_A + \Phi_t\,\eta_M - \overline{\alpha}_{O,t}\,\eta_{\omega,t}}{\overline{\alpha}_{L,t}} \qquad \Phi_t \equiv -\sum_o s_{o,t}\frac{\partial \log c_o}{\partial \log \overline{m}_t} \geq 0$$

- Φ_t: gain from the growing frontier; only frontier-constrained tasks contribute
- Declining frictions ($\eta_{\omega,t} < 0$) raise wage growth like cheaper compute, weighted by $\overline{\alpha}_{O,t}$

The race within each occupation:

$$\frac{\alpha_{K,o,t}}{\alpha_{L,o,t}} = \frac{(1-\psi_{o,t})\,\varphi_{K,o,t}\,\widetilde{c}_{K,o,t}^{\,1-\lambda}}{\varphi_{L,o,t}\,(\widetilde{\tau}_{o,t}\,w_t)^{1-\lambda} + \psi_{o,t}\,\varphi_{K,o,t}\,\widetilde{c}_{K,o,t}^{\,1-\lambda}}$$

Equilibrium Dynamics

Wage growth from the numeraire condition (Shephard within occupations):

$$\frac{\dot{w}_t}{w_t} = \frac{\eta_Z + \overline{\alpha}_{K,t}\,\eta_A + \Phi_t\,\eta_M - \overline{\alpha}_{O,t}\,\eta_{\omega,t}}{\overline{\alpha}_{L,t}} \qquad \Phi_t \equiv -\sum_o s_{o,t}\frac{\partial \log c_o}{\partial \log \overline{m}_t} \geq 0$$

- Φ_t: gain from the growing frontier; only frontier-constrained tasks contribute
- Declining frictions ($\eta_{\omega,t} < 0$) raise wage growth like cheaper compute, weighted by $\overline{\alpha}_{O,t}$

The race within each occupation:

$$\frac{\alpha_{K,o,t}}{\alpha_{L,o,t}} = \frac{(1-\psi_{o,t})\,\varphi_{K,o,t}\,\widetilde{c}_{K,o,t}^{\,1-\lambda}}{\varphi_{L,o,t}\,(\widetilde{\tau}_{o,t}\,w_t)^{1-\lambda} + \psi_{o,t}\,\varphi_{K,o,t}\,\widetilde{c}_{K,o,t}^{\,1-\lambda}}$$

- Automation raises the bundle terms; cheaper compute lowers $\widetilde{c}_K$, but not below its overhead wages
- $\lambda < 1$: spending shifts away from what becomes cheap; cheaper compute also raises $\psi_{o,t}$

back

Balanced Growth: Four Regimes

Constant growth $(\eta_L, \eta_Z, \eta_A, \eta_M)$; frictions fall at common rate $\eta_\omega \leq 0$; gap elasticity $\to \beta$ β

Critical rates: frontier speed $\eta_M^c \equiv (\eta_Z + \eta_A)/b_m^\gamma$; friction decline $\eta_\omega^c \equiv -\beta \min\left\{\eta_M, \eta_M^c\right\}/(1-\lambda)$

Balanced Growth: Four Regimes

Constant growth $(\eta_L, \eta_Z, \eta_A, \eta_M)$; frictions fall at common rate $\eta_\omega \leq 0$; gap elasticity $\to \beta$ β

Critical rates: frontier speed $\eta_M^c \equiv (\eta_Z + \eta_A)/b_m^\gamma$; friction decline $\eta_\omega^c \equiv -\beta \min\{\eta_M, \eta_M^c\}/(1-\lambda)$

	Bounded capabilities	**Persistent frictions**	**Scarce frontier**	**Abundant frontier**
fully in reach	none	at least one	at least one	at least one
friction decline	any	$\eta_\omega > \eta_\omega^c$	$\eta_\omega < \eta_\omega^c$	$\eta_\omega < \eta_\omega^c$
frontier speed	any	any	$\eta_M < \eta_M^c$	$\eta_M > \eta_M^c$
growth η	η_Z	$\eta_Z - \eta_\omega$	$\eta_Z + \frac{\beta \eta_M}{1-\lambda}$	$\eta_Z + \frac{\beta \eta_M^c}{1-\lambda}$
AI capital share $\overline{\alpha}_K^\infty$	0	0	0	$\beta/(\beta + (1-\lambda)b_m^\gamma)$

Balanced Growth: Four Regimes

Constant growth $(\eta_L, \eta_Z, \eta_A, \eta_M)$; frictions fall at common rate $\eta_\omega \leq 0$; gap elasticity $\to \beta$ β

Critical rates: frontier speed $\eta_M^c \equiv (\eta_Z + \eta_A)/b_m^\gamma$; friction decline $\eta_\omega^c \equiv -\beta \min\{\eta_M, \eta_M^c\}/(1-\lambda)$

	Bounded capabilities	**Persistent frictions**	**Scarce frontier**	**Abundant frontier**
fully in reach	none	at least one	at least one	at least one
friction decline	any	$\eta_\omega > \eta_\omega^c$	$\eta_\omega < \eta_\omega^c$	$\eta_\omega < \eta_\omega^c$
frontier speed	any	any	$\eta_M < \eta_M^c$	$\eta_M > \eta_M^c$
growth η	η_Z	$\eta_Z - \eta_\omega$	$\eta_Z + \frac{\beta \eta_M}{1-\lambda}$	$\eta_Z + \frac{\beta \eta_M^c}{1-\lambda}$
AI capital share $\overline{\alpha}_K^\infty$	0	0	0	$\beta/(\beta + (1-\lambda)b_m^\gamma)$

Gross substitutes ($\varepsilon > 1$; $\omega > 0$ where fully in reach): one occupation within reach suffices

Proof sketch

- Regimes 2–4: $\eta = \eta_Z + \min\{-\eta_\omega, \beta \min\{\eta_M, \eta_M^c\}/(1-\lambda)\}$: η_Z plus the slowest-declining cost

back

The Race, Resolved

Persistent frictions: all tasks in reach automated, yet overhead wages dominate the bundle

- Pace-setters cheapen only as frictions fall: $\eta = \eta_Z - \eta_\omega$; adoption frictions, not compute, bind

The Race, Resolved

Persistent frictions: all tasks in reach automated, yet overhead wages dominate the bundle

- Pace-setters cheapen only as frictions fall: $\eta = \eta_Z - \eta_\omega$; adoption frictions, not compute, bind

Scarce frontier: desired model outruns the available one $\Rightarrow$ adopt it

- Each improvement closes the gap $\Rightarrow$ wages grow faster than η_Z, by $\beta\eta_M/(1-\lambda)$

The Race, Resolved

Persistent frictions: all tasks in reach automated, yet overhead wages dominate the bundle

- Pace-setters cheapen only as frictions fall: $\eta = \eta_Z - \eta_\omega$; adoption frictions, not compute, bind

Scarce frontier: desired model outruns the available one $\Rightarrow$ adopt it

- Each improvement closes the gap $\Rightarrow$ wages grow faster than η_Z, by $\beta\eta_M/(1-\lambda)$

Abstract frontier: firms rationally lag the frontier

$$\overline{\alpha}_K^{\infty} = \frac{\beta}{\beta + (1-\lambda)\, b_m^{\gamma}}$$

- Capability gain just offsets the compute penalty: the race's summary

The Race, Resolved

Persistent frictions: all tasks in reach automated, yet overhead wages dominate the bundle

- Pace-setters cheapen only as frictions fall: $\eta = \eta_Z - \eta_\omega$; adoption frictions, not compute, bind

Scarce frontier: desired model outruns the available one $\Rightarrow$ adopt it

- Each improvement closes the gap $\Rightarrow$ wages grow faster than η_Z, by $\beta\eta_M/(1-\lambda)$

Abundant frontier: firms rationally lag the frontier

$$\overline{\alpha}_K^{\infty} = \frac{\beta}{\beta + (1-\lambda)\, b_m^{\gamma}}$$

- Capability gain just offsets the compute penalty: the race's summary

Automation is not its reward: $\varphi_{K,j} \to 1 - H_j(q^*;\infty)$ in all regimes

- Persistent frictions, scarce frontier: automation completes while $\overline{\alpha}_{K,t} \to 0$ (Baumol, $\lambda < 1$)

The Race, Resolved

Persistent frictions: all tasks in reach automated, yet overhead wages dominate the bundle

- Pace-setters cheapen only as frictions fall: $\eta = \eta_Z - \eta_\omega$; adoption frictions, not compute, bind

Scarce frontier: desired model outruns the available one $\Rightarrow$ adopt it

- Each improvement closes the gap $\Rightarrow$ wages grow faster than η_Z, by $\beta\eta_M/(1-\lambda)$

Abundant frontier: firms rationally lag the frontier

$$\overline{\alpha}_K^{\infty} = \frac{\beta}{\beta + (1-\lambda)\, b_m^{\gamma}}$$

- Capability gain just offsets the compute penalty: the race's summary

Automation is not its reward: $\varphi_{K,j} \to 1 - H_j(q^*;\infty)$ in all regimes

- Persistent frictions, scarce frontier: automation completes while $\overline{\alpha}_{K,t} \to 0$ (Baumol, $\lambda < 1$)
- The race, not the extent of automation, decides AI capital's share

back

Balanced Growth: Proof Sketch

Large-size model choice: max $(1 - H_j(q^*; m))B(v; \omega)$; v: log cutoff/requirement gap

- Interior optimum: $\delta_j(m) \asymp e^{-(1-\lambda)v} J_P(v, s)$, $J_P \equiv \int_{-\infty}^{v} e^u (e^u + e^s)^{-\lambda} \, dG(u)$; s locates the overhead

Balanced Growth: Proof Sketch

Large-size model choice: max $(1 - H_j(q^*; m))B(v; \omega)$; v: log cutoff/requirement gap

- Interior optimum: $\delta_j(m) \asymp e^{-(1-\lambda)v} J_P(v, s)$, $J_P \equiv \int_{-\infty}^{v} e^u (e^u + e^s)^{-\lambda}\, dG(u)$; s locates the overhead

Two faces of the bundle set the unconstrained model's growth g_m^{int}:

$$\underbrace{g_m^{\text{int}} = \frac{(1-\lambda)(g_w + \eta_A)}{\beta + (1-\lambda)b_m^{\gamma}}}_{\text{compute dominates } (s \to -\infty)} \qquad \underbrace{g_m^{\text{int}} = \frac{g_w + \eta_A + \lambda\eta_\omega}{\beta + b_m^{\gamma}}}_{\text{overhead dominates } (s \to +\infty)}$$

overhead dominates ($s \to +\infty$): bundle → its overhead wages, $\psi_j \to 1$

- Frontier binds iff $\eta_M < g_m^{\text{int}}$; η_M vs. η_M^c splits scarce from abundant

Balanced Growth: Proof Sketch

Large-size model choice: max $(1 - H_j(q^*; m))B(v; \omega)$; v: log cutoff/requirement gap

- Interior optimum: $\delta_j(m) \asymp e^{-(1-\lambda)v} J_P(v, s)$, $J_P \equiv \int_{-\infty}^{v} e^u (e^u + e^s)^{-\lambda}\, dG(u)$; s locates the overhead

Two faces of the bundle set the unconstrained model's growth g_m^{int}:

$$\underbrace{g_m^{\text{int}} = \frac{(1-\lambda)(g_w + \eta_A)}{\beta + (1-\lambda) b_m^{\gamma}}}_{\text{compute dominates } (s \to -\infty)} \qquad \underbrace{g_m^{\text{int}} = \frac{g_w + \eta_A + \lambda \eta_\omega}{\beta + b_m^{\gamma}}}_{\text{overhead dominates } (s \to +\infty)}$$

bundle → its overhead wages, $\psi_j \to 1$

- Frontier binds iff $\eta_M < g_m^{\text{int}}$; η_M vs. η_M^c splits scarce from abundant

Prices follow the slowest-declining component (cost-index units):

- Residual task labor falls at βg_m, overhead wages at $(1 - \lambda)|\eta_\omega|$
- Comparing them gives $\eta_\omega^c = -\beta \min\{\eta_M, \eta_M^c\}/(1 - \lambda)$ and the persistent-frictions regime

Balanced Growth: Proof Sketch

Large-size model choice: max $(1 - H_j(q^*; m))B(v; \omega)$; v: log cutoff/requirement gap

- Interior optimum: $\delta_j(m) \asymp e^{-(1-\lambda)v} J_P(v, s)$, $J_P \equiv \int_{-\infty}^{v} e^u (e^u + e^s)^{-\lambda}\, dG(u)$; s locates the overhead

Two faces of the bundle set the unconstrained model's growth g_m^{int}:

$$\underbrace{g_m^{\text{int}} = \frac{(1-\lambda)(g_w + \eta_A)}{\beta + (1-\lambda) b_m^{\gamma}}}_{\text{compute dominates } (s \to -\infty)} \qquad \underbrace{g_m^{\text{int}} = \frac{g_w + \eta_A + \lambda \eta_\omega}{\beta + b_m^{\gamma}}}_{\text{overhead dominates } (s \to +\infty)}$$

bundle → its overhead wages, $\psi_j \to 1$

- Frontier binds iff $\eta_M < g_m^{\text{int}}$; η_M vs. η_M^c splits scarce from abundant

Prices follow the slowest-declining component (cost-index units):

- Residual task labor falls at βg_m, overhead wages at $(1-\lambda)|\eta_\omega|$
- Comparing them gives $\eta_\omega^c = -\beta \min\{\eta_M, \eta_M^c\}/(1-\lambda)$ and the persistent-frictions regime

Shares: abundant frontier fixes AI capital/labor at $\beta/[(1-\lambda) b_m^{\gamma}]$; frontier-bound tasks capped

Balanced Growth: Proof Sketch

Large-size model choice: max $(1 - H_j(q^*; m))B(v; \omega)$; v: log cutoff/requirement gap

- Interior optimum: $\delta_j(m) \asymp e^{-(1-\lambda)v} J_P(v, s)$, $J_P \equiv \int_{-\infty}^{v} e^u (e^u + e^s)^{-\lambda}\, dG(u)$; s locates the overhead

Two faces of the bundle set the unconstrained model's growth g_m^{int}:

$$\underbrace{g_m^{\text{int}} = \frac{(1-\lambda)(g_w + \eta_A)}{\beta + (1-\lambda) b_m^{\gamma}}}_{\text{compute dominates } (s \to -\infty)} \qquad \underbrace{g_m^{\text{int}} = \frac{g_w + \eta_A + \lambda \eta_\omega}{\beta + b_m^{\gamma}}}_{\text{overhead dominates } (s \to +\infty)}$$

bundle → its overhead wages, $\psi_j \to 1$

- Frontier binds iff $\eta_M < g_m^{\text{int}}$; η_M vs. η_M^c splits scarce from abundant

Prices follow the slowest-declining component (cost-index units):

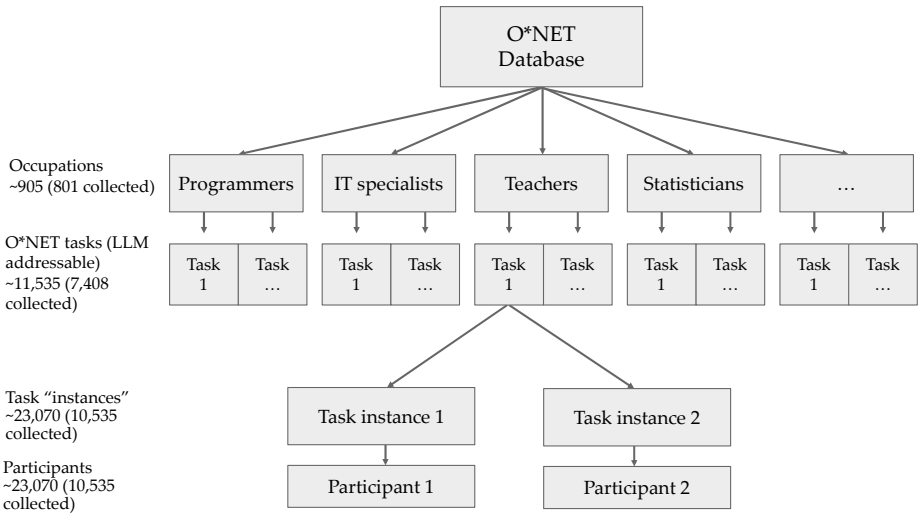
- Residual task labor falls at βg_m, overhead wages at $(1-\lambda)|\eta_\omega|$
- Comparing them gives $\eta_\omega^c = -\beta \min\{\eta_M, \eta_M^c\}/(1-\lambda)$ and the persistent-frictions regime

Shares: abundant frontier fixes AI capital/labor at $\beta/[(1-\lambda) b_m^{\gamma}]$; frontier-bound tasks capped

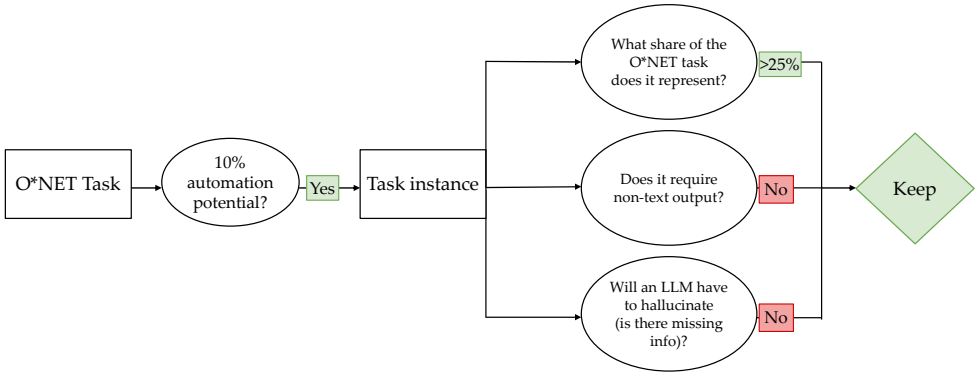
Wages: numeraire pins η; output per worker $= w_t / (1 - \overline{\alpha}_{K,t})$

back

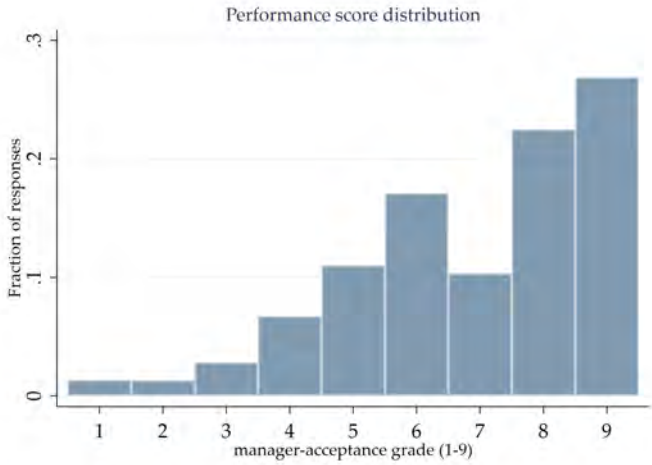
O*NET Task Selection



Task Filtering

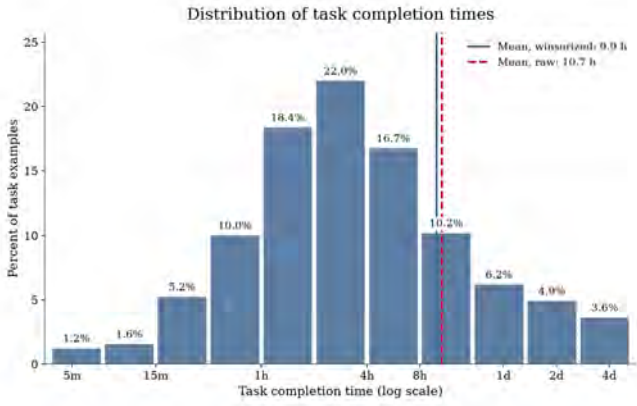


Survey descriptive statistics - performance scores



back

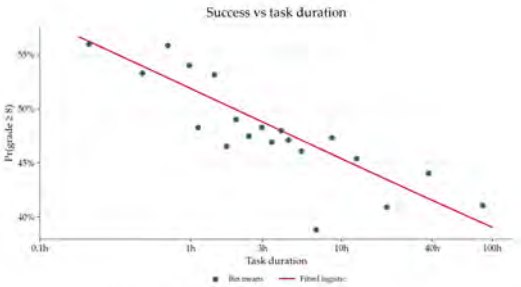
Survey descriptive statistics - task duration distribution

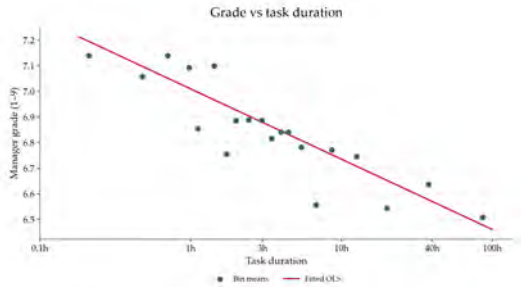


back

Notes: One elicited task per respondent; N = 13,986. Median 3 hours; bars give the share of task examples per doubling of completion time. The winsorized series clips completion time at the 1st and 99th percentiles (0.10 and 125 hours); means are drawn on the log scale.

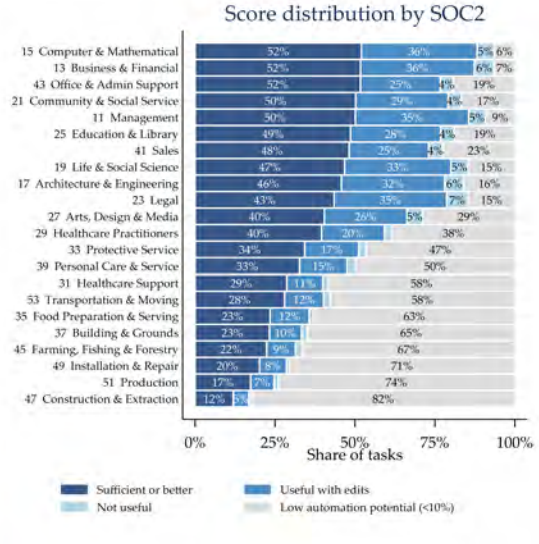
Survey descriptive statistics - performance vs. task duration





back

Survey descriptive statistics - performance by SOC2



back

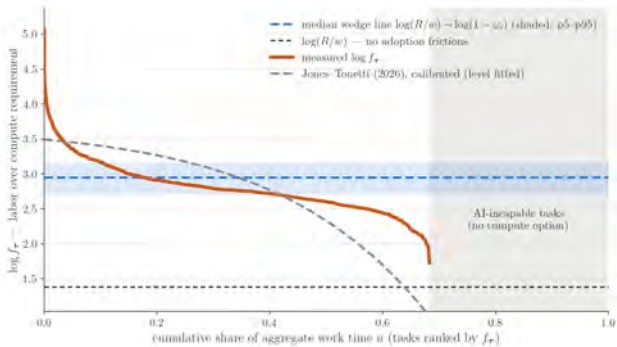
Survey model coverage

		Billions (best available estimate)	
Model	Publication date	Parameters	Training tokens
GPT-4	2023-03-15	1,800.0	5,417
GPT-3.5-Turbo	2023-06-13	20.0	–
Llama-2-70B	2023-07-18	70.0	2,000
Llama-2-7B	2023-07-18	7.0	2,000
Claude-Haiku-3	2024-03-04	–	–
Claude-Opus-3	2024-03-04	2,500.0	40,000
GPT-4o	2024-05-13	200.0	20,000
Gemini-1.5-Pro	2024-05-14	200.0	30,000
Qwen-2-72B-Instruct	2024-06-06	72.0	7,000
Qwen-2-7B-Instruct	2024-06-06	7.0	7,000
GPT-4o-mini	2024-07-18	8.0	13,000
Llama-3.1-405B-Instruct	2024-07-23	405.0	15,600
Llama-3.1-70B-Instruct	2024-07-23	70.0	15,600
Llama-3.1-8B-Instruct	2024-07-23	8.0	15,600
Claude-Haiku-3.5	2024-10-22	–	–
DeepSeek-V3	2024-12-24	671.0	14,800
DeepSeek-R1	2025-01-20	671.0	14,800
Claude-Sonnet-3.7	2025-02-24	175.0	20,000
Granite-3.3-2B	2025-02-26	2.5	12,000
QwQ-32B	2025-03-06	32.5	18,000
Gemma-3-1B-it	2025-03-12	1.0	2,000
Gemini-2.5-Pro	2025-03-25	400.0	80,000
Llama-4-Maverick-400B-A17B	2025-04-05	400.0	30,000
Llama-4-Scout-109B-A17B	2025-04-05	109.0	30,000
Granite-3.3-8B	2025-04-16	8.1	12,000
o3	2025-04-16	600.0	100,000
o4-mini	2025-04-16	200.0	40,000
Qwen-3-14B	2025-04-29	14.0	36,000
Qwen-3-235B	2025-04-29	235.0	36,000
Qwen-3-32B	2025-04-29	32.0	36,000
Mistral-Medium	2025-05-07	50.0	12,000
Claude-Sonnet-4	2025-05-22	–	–
Gemini-2.5-Flash	2025-06-17	80.0	20,000
Gemini-2.5-Flash-Lite	2025-07-22	–	–
GPT-5-Thinking	2025-08-01	1,700.0	–
Claude-Opus-4.1	2025-08-05	2,000.0	100,000
GPT-OSS-120B-A5.1B	2025-08-05	116.8	30,000
GPT-OSS-20B-A3.6B	2025-08-05	20.9	13,000
GPT-5	2025-08-07	1,700.0	114,000
GPT-5-mini	2025-08-07	85.0	–
GPT-5-nano	2025-08-07	15.0	–

back

Shape of Comparative Advantage Across Tasks

$\vartheta_j \equiv \tau_j/\bar{\gamma}_j$: labor over mean compute requirement; automated on average iff $\vartheta_j \geq \frac{R/w}{1-\omega_o}$



Flat where AI is capable, a cliff at the edge: adoption frictions set the automation margin

Task-level counterpart of the schedule Jones & Tonetti (2026) calibrate to aggregate capital shares; τ wage-adjusted.

back

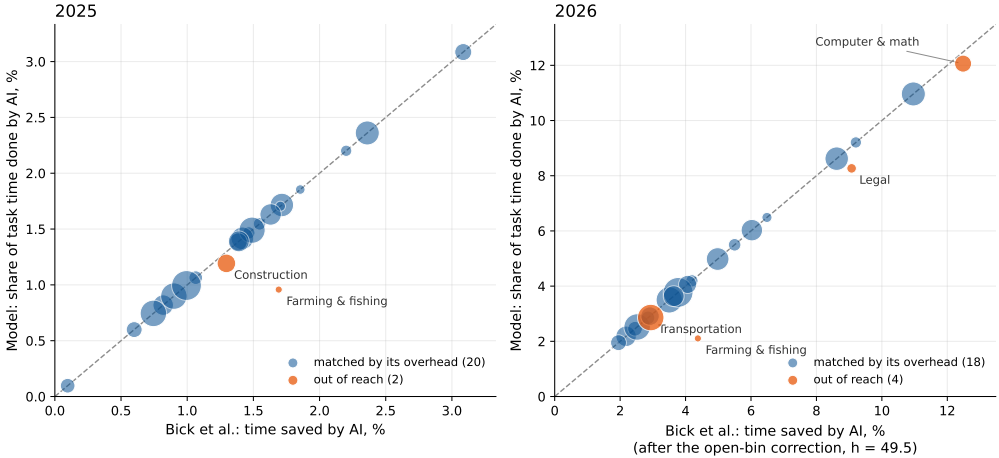
AI Spending Targets

- US spending on AI services over AI spending plus labour costs (wages × BEA compensation-to-wages, about 1.21), July–June years.
- 2025 anchor: calendar-2025 spending of $13.19bn = (OpenAI $13.07bn + Anthropic $4.60bn) × 50% US / 67% coverage; half-year revenues set its timing.

Year (July–June)	AI services ($bn)	Labour costs ($bn)	AI spending share (%)
2025: July 2024–June 2025	6.13	12,710	0.0482
2026: July 2025–June 2026	30.49	13,274	0.2292
2026 relative to 2025	×4.98	×1.044	odds ×4.767

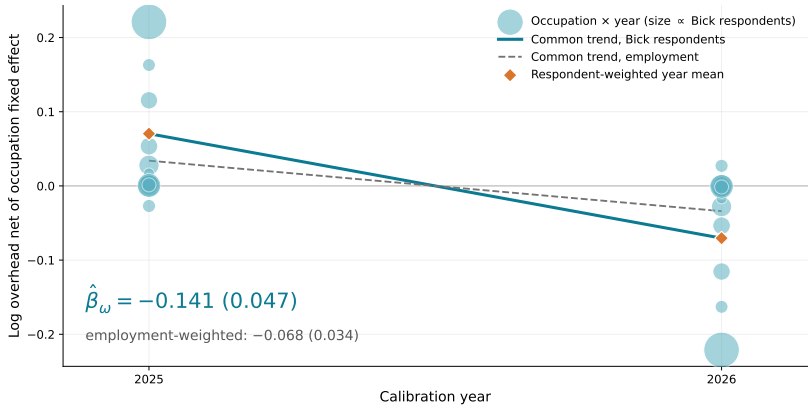
back

Fit of Adoption Rate



back

Change in Adoption Frictions



$\log \omega_{o,t} = \mathrm{FE}_o + \beta_\omega (t - 2025.5) + \varepsilon_{o,t}$; the 17 occupations matched in both years. Standard errors in parentheses.

back

How the Friction Trend Is Estimated

Regression: log frictions on the calibration year, with occupation fixed effects

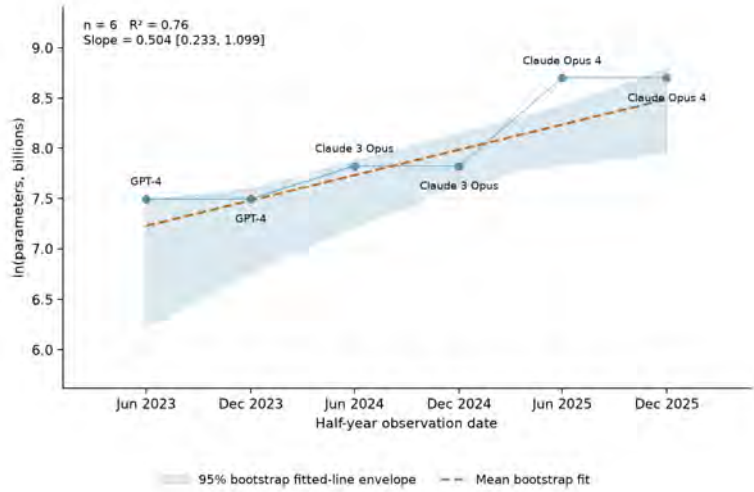
- The 17 occupation groups matched in both the 2025 and the 2026 calibrations
- Each group is weighted by its survey respondents in the two years

Estimate: an annual decline of 14.1% (standard error 4.7%)

- In a bootstrap across the 17 groups, the middle 80% spans annual declines of 2.8% to 22.9%
- Weighting groups by employment instead gives 6.8% (standard error 3.4%)

Back

Trend in Parameter Size of Frontier Models



back

Trend in the Price of Compute: Chip Purchase Prices

Data: release prices of ML accelerator chips, 2008–25, in 2024 dollars

- We measure each chip's price per unit of throughput; the throughput measure varies by row
- The intervals resample release years, so chips released in the same year can move together

Throughput measure	Annual decline	95% interval	Chips (years)
FP32, all chips (headline)	0.170	[0.068, 0.251]	54 (15)
Tensor, FP16/BF16	0.231	[0.108, 0.348]	20 (8)
FP32, the cheapest chip each year	0.127	[−0.012, 0.279]	15 (15)

back

Back

The Rental Price of Compute: Data and Scenarios

Data: OpenRouter list prices per output token, August 2025 to July 2026

- The same models appear in every month, weighted by their fixed shares of spending
- Model and month fixed effects give the common change in token prices

Change: token prices fell 1.09% from Aug–Dec 2025 to Jan–Jul 2026

- The two periods are six months apart, so the annual pace is 2.17% (in nominal terms)
- This tracks the rental price if serving efficiency and markups stay constant
- The calibration reads it as the fall in the rental price relative to wages
- The forecast applies the scenario rates to the rental price in units of output

Scenarios: the decline moves halfway to its long-run rate every 5 years

- The paths use the 2026 interest rate; in the scenarios the price also moves with the rate

Back

Automation Without a Rising Capital Share

Capability share rises by $d(1-H)$ at given w, R, m

$$d\log c = -\frac{1}{1-\lambda}\underbrace{\left(\frac{\tau w}{c}\right)^{1-\lambda} d(1-H)}_{\text{spending the new varieties carried}}\underbrace{\zeta(\gamma^*;m)}_{\text{cost saving}}$$

$$d\alpha_K = \left(\frac{\tau w}{c}\right)^{1-\lambda} d(1-H)\int_0^{\gamma^*}\underbrace{(1-\omega)\frac{\gamma}{\gamma^*}\left((1-\omega)\frac{\gamma}{\gamma^*}+\omega\right)^{-\lambda}}_{\text{compute's part: 0 at } \gamma=0,\ 1-\omega \text{ at } \gamma^*} dG(\gamma;m)\ \underbrace{-(1-\lambda)\,\alpha_K\, d\log c}_{\text{reallocation}}$$

Automation Without a Rising Capital Share

Capability share rises by $d(1-H)$ at given w, R, m

$$d\log c = -\frac{1}{1-\lambda}\underbrace{\left(\frac{\tau\, w}{c}\right)^{1-\lambda} d(1-H)}_{\text{spending the new varieties carried}}\ \underbrace{\zeta(\gamma^*; m)}_{\text{cost saving}}$$

$$d\alpha_K = \left(\frac{\tau\, w}{c}\right)^{1-\lambda} d(1-H)\int_0^{\gamma^*} \underbrace{(1-\omega)\frac{\gamma}{\gamma^*}\left((1-\omega)\frac{\gamma}{\gamma^*}+\omega\right)^{-\lambda}}_{\text{compute's part: 0 at } \gamma=0,\ 1-\omega \text{ at } \gamma^*} dG(\gamma; m)\ \underbrace{-(1-\lambda)\,\alpha_K\, d\log c}_{\text{reallocation}}$$

- Acemoglu–Restrepo ($\omega = 0$): whole spending moves to capital; the variety at the cutoff saves nothing

Automation Without a Rising Capital Share

Capability share rises by $d(1-H)$ at given w, R, m

$$d\log c = -\frac{1}{1-\lambda}\underbrace{\left(\frac{\tau\, w}{c}\right)^{1-\lambda} d(1-H)}_{\text{spending the new varieties carried}}\ \underbrace{\zeta(\gamma^*; m)}_{\text{cost saving}}$$

$$d\alpha_K = \left(\frac{\tau\, w}{c}\right)^{1-\lambda} d(1-H)\int_0^{\gamma^*} \underbrace{(1-\omega)\frac{\gamma}{\gamma^*}\left((1-\omega)\frac{\gamma}{\gamma^*}+\omega\right)^{-\lambda}}_{\text{compute's part: 0 at } \gamma=0,\ 1-\omega \text{ at } \gamma^*} dG(\gamma; m)\ \underbrace{-(1-\lambda)\,\alpha_K\, d\log c}_{\text{reallocation}}$$

- Acemoglu–Restrepo ($\omega = 0$): whole spending moves to capital; the variety at the cutoff saves nothing
- This model: capability binds, $\gamma \ll \gamma^* \Rightarrow \zeta$ near its cap $1-\omega^{1-\lambda}$, AI capital's part near 0

Automation Without a Rising Capital Share

Capability share rises by $d(1-H)$ at given w, R, m

$$d\log c = -\frac{1}{1-\lambda}\underbrace{\left(\frac{\tau\, w}{c}\right)^{1-\lambda} d(1-H)}_{\text{spending the new varieties carried}}\ \underbrace{\zeta(\gamma^*;m)}_{\text{cost saving}}$$

$$d\alpha_K = \left(\frac{\tau\, w}{c}\right)^{1-\lambda} d(1-H)\int_0^{\gamma^*}\underbrace{(1-\omega)\frac{\gamma}{\gamma^*}\left((1-\omega)\frac{\gamma}{\gamma^*}+\omega\right)^{-\lambda}}_{\text{compute's part: 0 at } \gamma=0,\ 1-\omega \text{ at } \gamma^*} dG(\gamma;m)\ \underbrace{-(1-\lambda)\,\alpha_K\, d\log c}_{\text{reallocation}}$$

- Acemoglu–Restrepo ($\omega = 0$): whole spending moves to capital; the variety at the cutoff saves nothing
- This model: capability binds, $\gamma \ll \gamma^* \Rightarrow \zeta$ near its cap $1-\omega^{1-\lambda}$, AI capital's part near 0
- Overhead: at most $1-\omega$ moves to AI capital; reallocation scales with small α_K

Why the Machine Cost Share Stays Small

Acemoglu–Restrepo ($\omega_o = 0$): an automated task's whole spending moves to capital

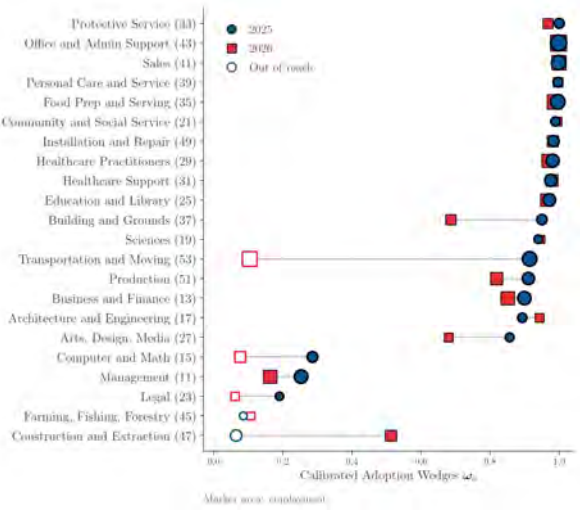
- At the cost margin: displacement with no productivity gain

This model: capability, not cost, limits automation

- Compute costs a fraction of the labor it replaces
- Tasks are complements: a task whose cost collapses loses spending share
- Overhead labor ω_o stays in automated tasks
- Algorithmic progress makes smaller models capable: typical model size falls by 2035

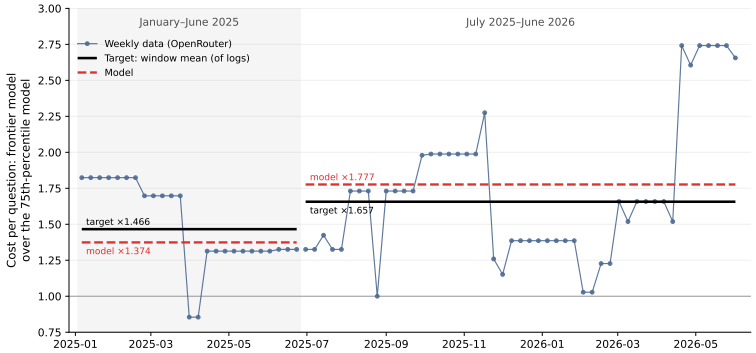
Back Theory

Adoption Frictions ω_o, 2025 and 2026



Back

Usage Target: the Frontier Cost Ratio



Window	Weeks	Target ratio	s.e. (logs)	Model	Ramp (robustness)
January–June 2025	25	1.466	0.106	1.374	1.470
July 2025–June 2026	49	1.657	0.106	1.777	1.575

The ECI Law: Estimates

Data: 93 models released 2023–26, with ECI and active parameters; robust s.e. in parentheses

$$\text{ECI} = 137.3 + 2.85\,(\log m - 4.12) + 10.41\,(t - 2025.5) + 7.8 \times \text{reasoning}$$

	ECI law	8 developer effects	Size × date
Log size, β_M	2.85 (0.32)	2.89 (0.30)	2.85 (0.32)
Date per year, β_T	10.41 (0.68)	10.81 (0.74)	10.28 (0.66)
Log size × date	—	—	0.47 (0.48)
Reasoning model	7.83 (1.10)	7.08 (1.27)	7.94 (1.07)
R^2	0.936	0.956	0.937

Figures: ECI and one regressor net of the others; slope = coefficient, with a 95% band

Size × date: not significant; its 2026 size slope is $2.85 + 0.47 \times 0.49 = 3.08$

Shading on the pace figure: a pointwise 80% band, leaving out one developer at a time

Back

Model Scaling: the Estimate

Sample: the highest-ECI model released each month, reasoning models, 2024–26

- 20 models; size is the number of active parameters

Estimate: size grew 41% a year; the 95% interval runs from −32% to +191%

Capability gain: $\beta_M \times \log(1.41) = 2.85 \times 0.3425 = 0.97$ ECI points a year

In the scenarios: frontier growth starts at 50% a year and slows, with a half-life of 10 years

- Long run: training spending stays a constant share of output that grows 1.16% a year
- That gives 3%, 10% or 17% a year under the three price scenarios
- A 1% larger model uses 0.14% more training data

Back

Three Paces of Algorithmic Progress

Benchmark, 10.4 a year: the 2023–26 pace of a model of a given size

Moderate, 8.4 a year: the pace below which the 2023–26 data reject at 10%

- Leaving out one developer at a time (8 developers) gives a standard error of 1.45

Conservative, 5.9 a year: the slowest two-year stretch, March 2023 to March 2025

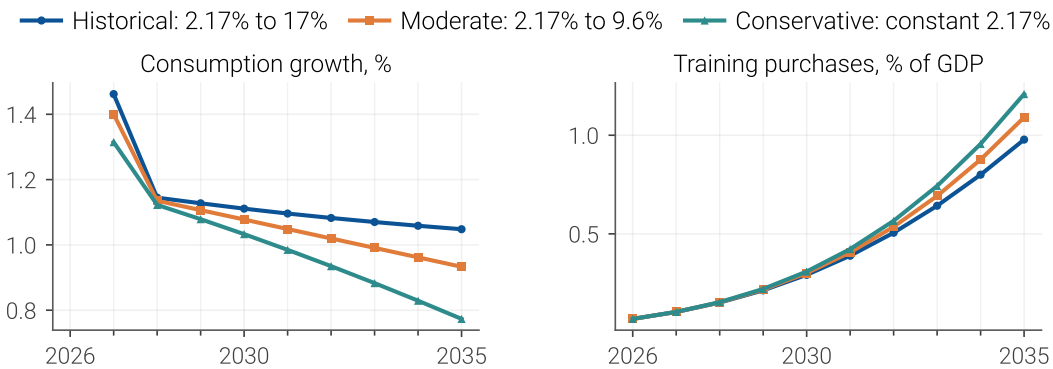
- Its 29 models gained 7.9 points a year; 5.9 is the 75% lower bound of that pace
- A stress case: a lasting return to the slowest pace seen so far

Back

Model Scaling: Where the Extra Output Goes

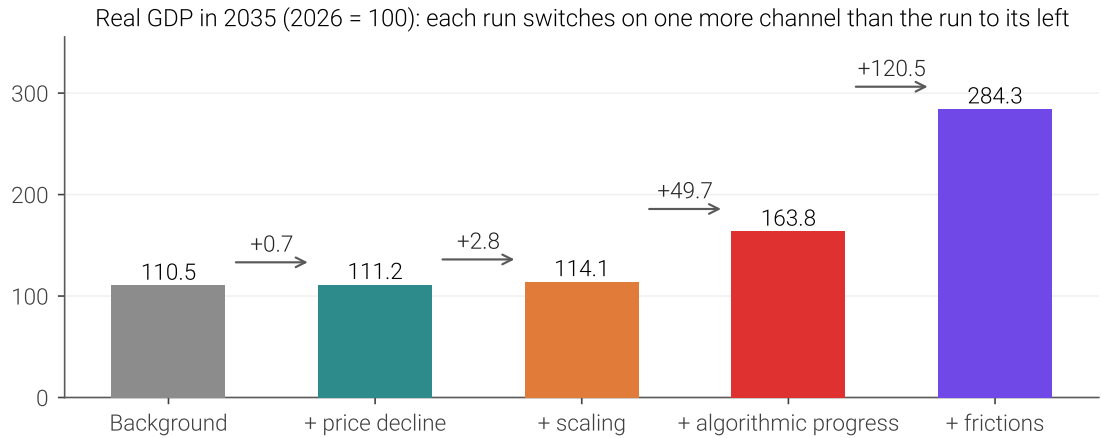
Real GDP counts training: its share of GDP rises from 0.07% to 1.0–1.2% by 2035

Consumption: from 2028, it grows more slowly than real GDP

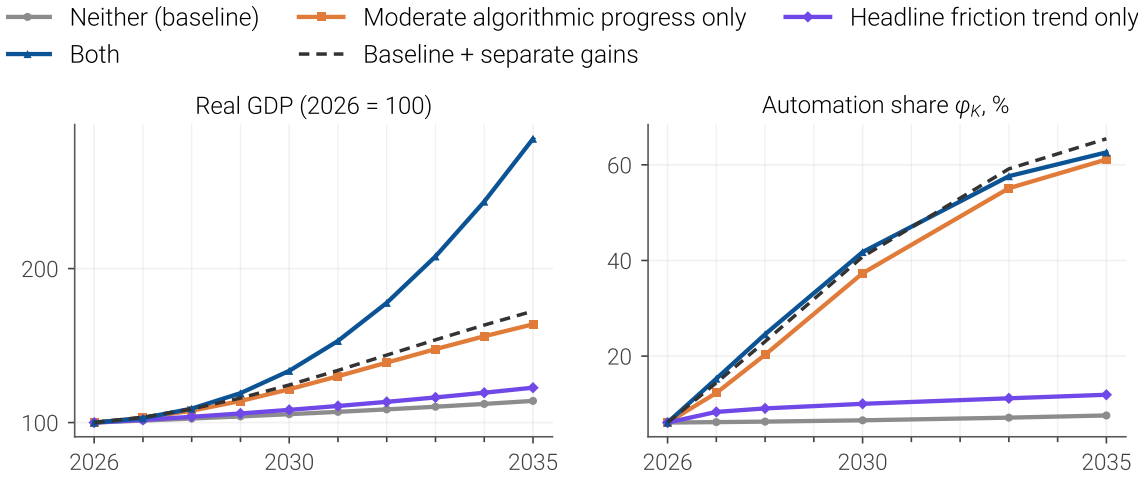


Back

Sequential Gains by 2035 (moderate algorithmic progress)

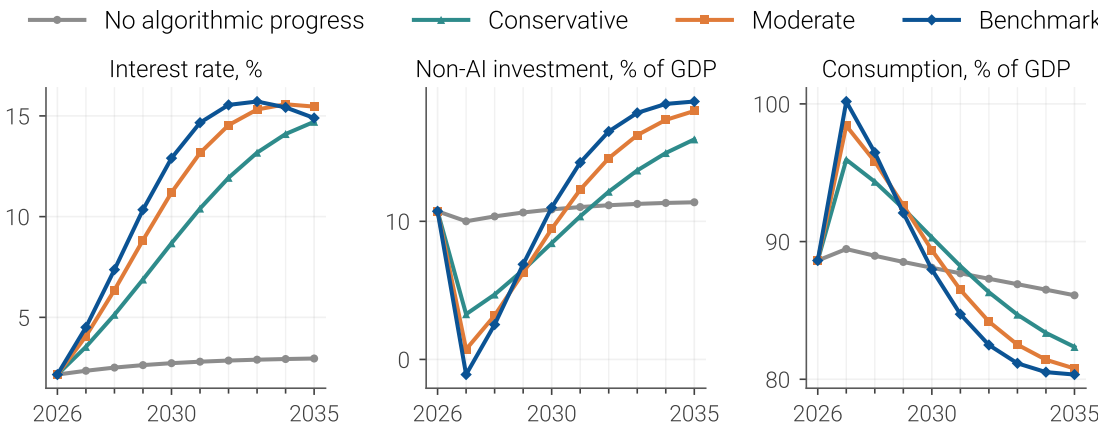


Output and Automation with Both Channels (historical price decline, scaling)



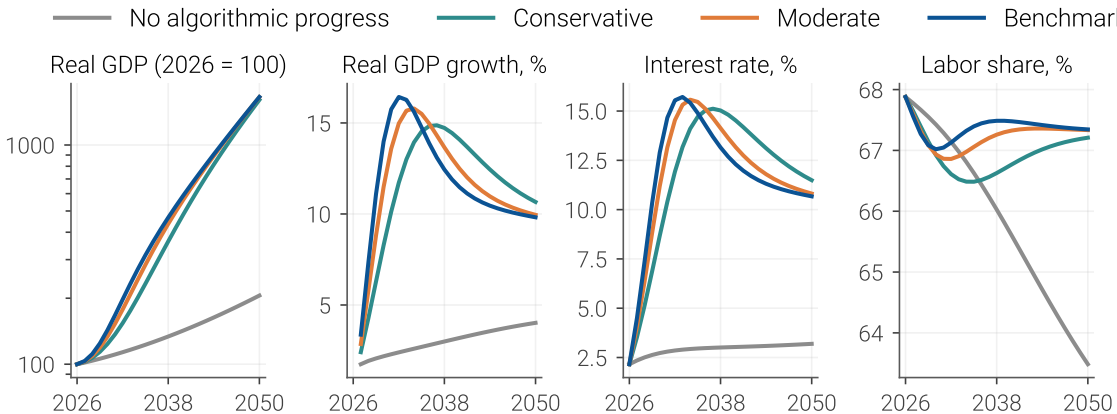
Back

Investment and the Interest Rate (historical price decline, headline frictions)



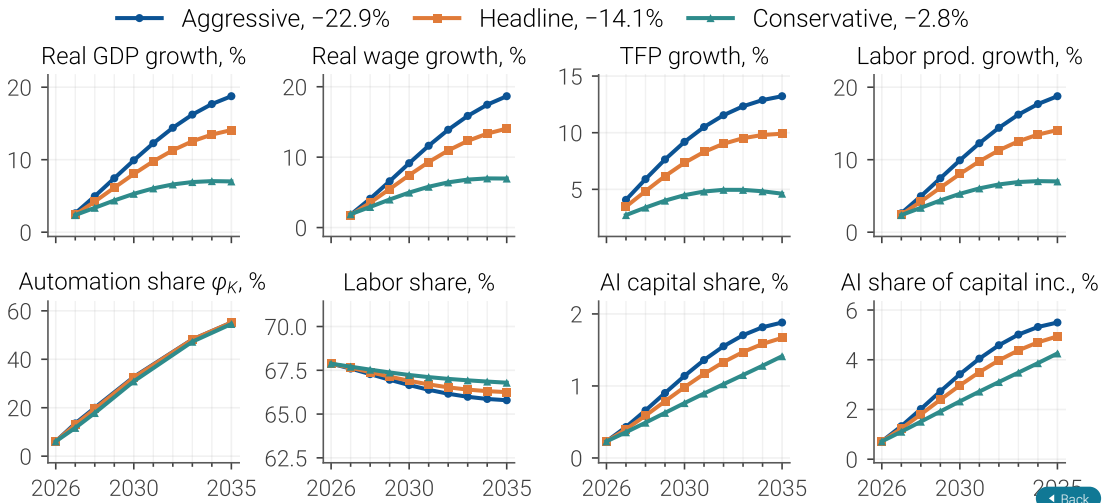
Back

Paths to 2050 (historical price decline, headline frictions)



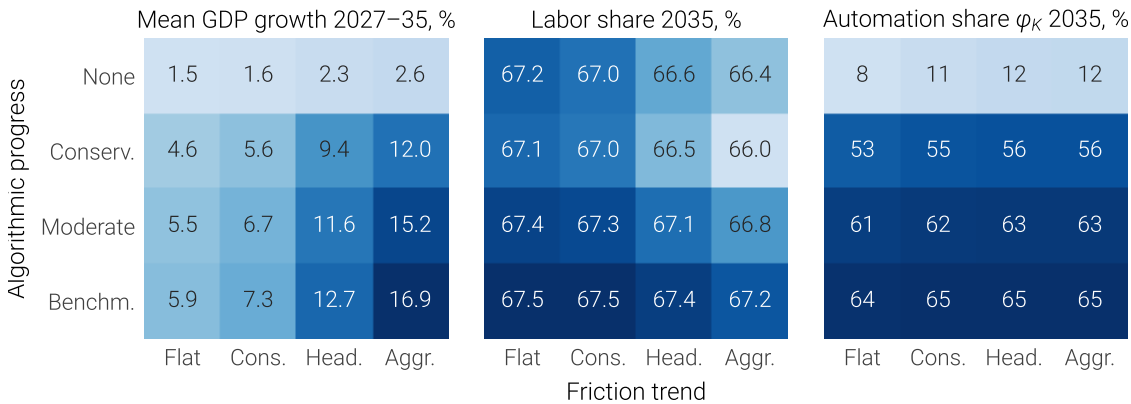
Back

Scenarios for Adoption Frictions (conservative AI progress and price decline)



Back

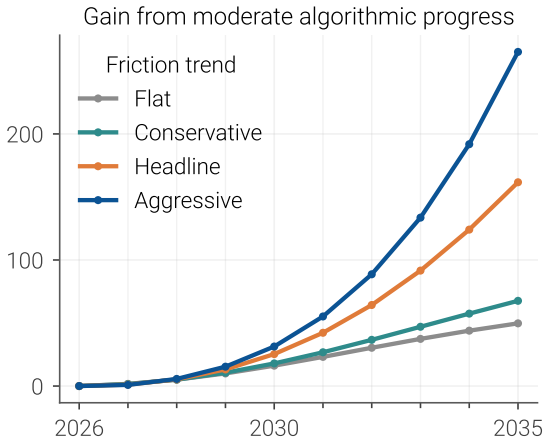
The Scenarios in One View (historical price decline)

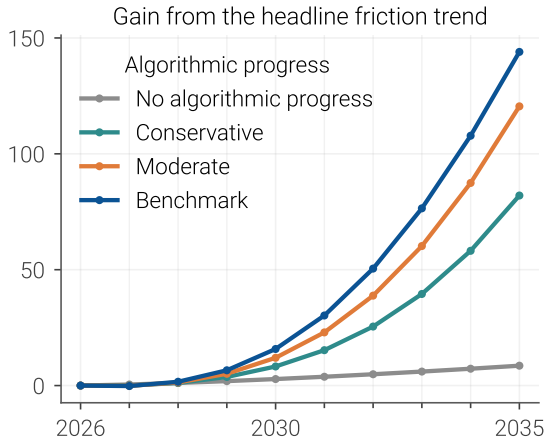


◂ Back | Investment and interest rate | Paths to 2050

Algorithmic Progress and Frictions Are Complements

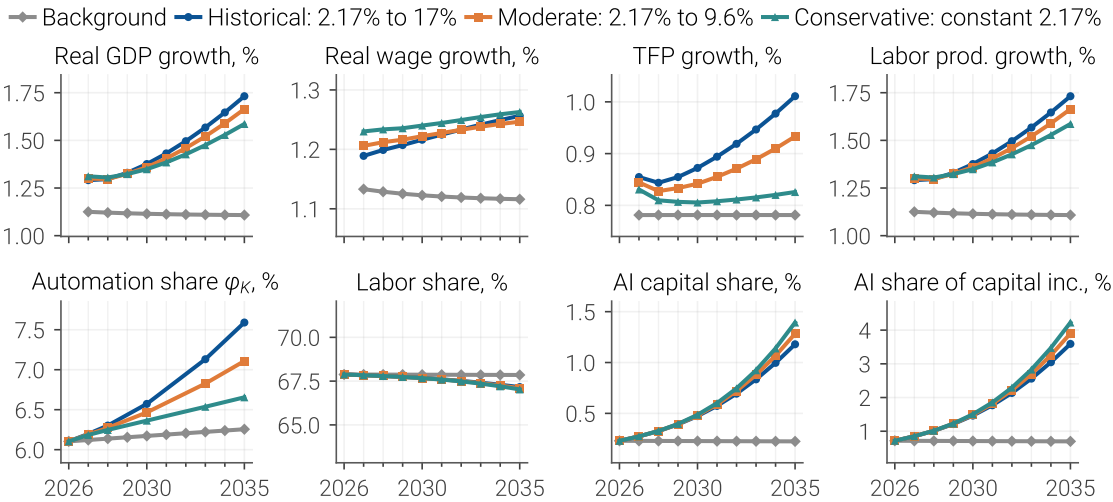
Gain in real GDP, index points (2026 = 100); historical price decline and scaling in all runs



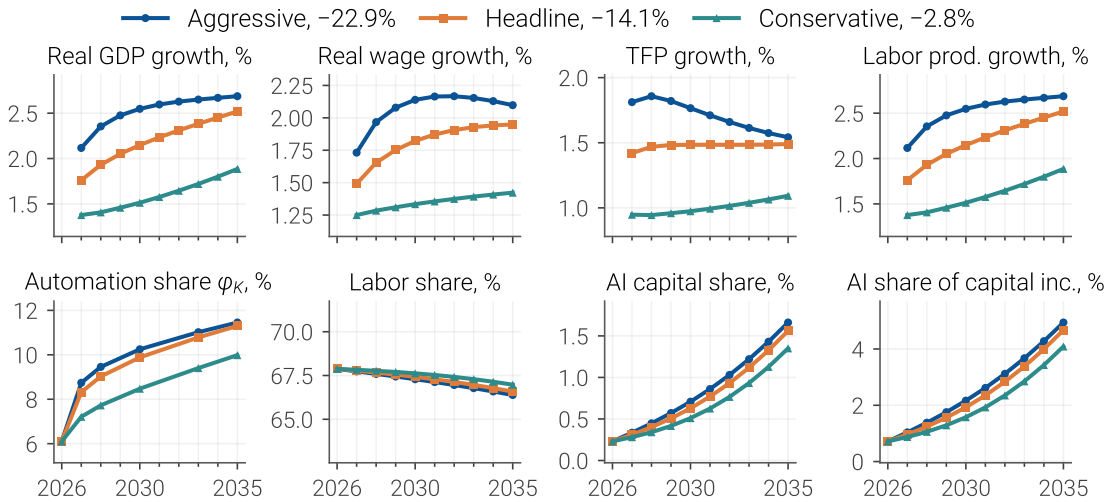


◂ Back | Output and automation paths | Step-by-step gains

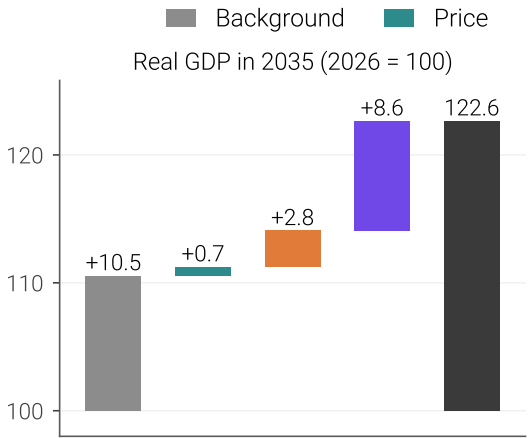
Model Scaling (under three price scenarios)

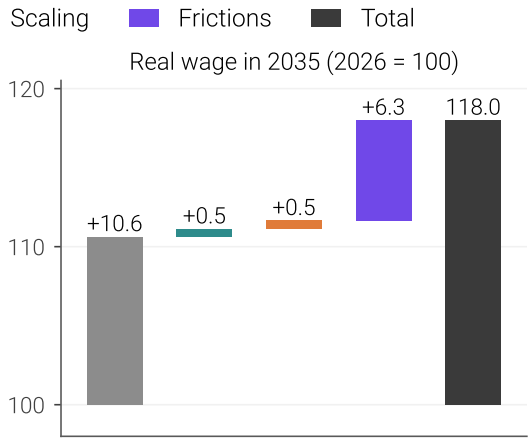


Scenarios for Adoption Frictions (scaling, moderate price decline)



Real GDP Gains by 2035 through Scaling (historical price decline, headline frictions)






PIIE
PETERSON INSTITUTE FOR
INTERNATIONAL ECONOMICS
OECD
INTERNATIONAL MONETARY FUND
WORLD BANK GROUP
www.piie.com
@PIIE
linkedin.com/company/piie

Discussion: AI Diffusion

Three lenses: the firm, the task, and the network

Baslandze, Edwards, **Graham**, McClure, Sparks, Meyer, Waddell & Weitz — *AI, Productivity, and the Workforce*
Lashkari, Maghzian, Mertens, Qiu & Thompson — *The Race between AI Costs and Capabilities*
Bastos, Stapleton, Taglioni & **Wei** — *Emerging Trends in Global Technology Diffusion*

Rodimiro Rodrigo
George Washington University

IMF–OECD–PIIE–World Bank Conference · November 18, 2026

Three lenses on how AI diffuses

The firm (Graham et al.)

Who invests, what returns?

CFO survey, ≈750 U.S. firms, 2025–26

⇒ Adoption is broad; gains are modest; employment effects are small, with occupational reallocation.

The task (Lashkari et al.)

What limits automation?

Worker-rated LLM tasks + task-based GE

⇒ Capabilities and adoption frictions, not compute prices, set the pace.

The network (Wei et al.)

Where, and through whom?

573M job postings, 17 countries, 29 techs

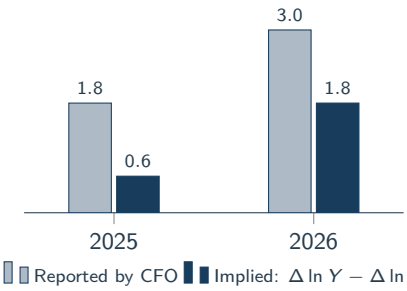
⇒ AI spreads early and widely, follows income, and travels weakly along value chains.

What binds AI diffusion: the technology, or the capacity to adopt it?

Baslandze, Graham et al.: AI at the typical firm

- **Adoption:** 59% invested in 2025, 85% plan to in 2026. Small firms invest **more per worker**.
- **AI arrives as a service:** 55–64% of spending is opex (subscriptions, training).
- **Sources of gains:** capital deepening ≈15%; the rest is revenue TFP, tied to **product and customer** motives, not cost-cutting.
- **Jobs:** −0.37% in 2026 (≈500k). Large adopters cut, small adopters hire.
- **Composition:** routine-clerical share −0.76% in 2026, office/admin is most exposed (NEI = 2.0); offset by +0.62% in skilled technical.

AI-attributed labor productivity growth, %



The wedge is read as a firm-level *productivity paradox*.

Graham et al.: attribution and the productivity wedge

1. Can CFOs identify the effect of AI?

Identification

Why it matters

- **"AI-attributed"** growth is not an observed counterfactual.
- Adopters may rationalize an investment they approved.
- Gains respond to the extensive margin but barely to the intensive one.

Suggestions

- Link past CFO Survey waves to actual revenue, employment; compare adopters/non with pre-2025 trends.
- **Next wave:** Re-ask the *2026 realized* outcomes; measure forecast errors and ex-ante optimism by adoption status.

Graham et al.: attribution and the productivity wedge

1. Can CFOs identify the effect of AI?

Identification

Why it matters

- **"AI-attributed"** growth is not an observed counterfactual.
- Adopters may rationalize an investment they approved.
- Gains respond to the extensive margin but barely to the intensive one.

Suggestions

- Link past CFO Survey waves to actual revenue, employment; compare adopters/non with pre-2025 trends.
- **Next wave:** Re-ask the *2026 realized* outcomes; measure forecast errors and ex-ante optimism by adoption status.

2. Is the wedge really a productivity J-curve?

Interpretation

Why it matters

- AI reported $\Delta LP >$ implied gains ΔLP. Delayed returns are *one* explanation.
- Price declines, hours saved without layoffs, quality improvements, can all widen the wedge.

Suggestions

- Relate AI use to firm price growth.
- **Next wave:** Ask hours saved, redeployment vs. slack, quality/output units, expected date benefits reach sales.

Graham et al.: measurement and mechanisms

3. AI capital deepening is understated – National Accounts

Growth accounting

Why it matters

- Capital deepening explains only $\sim$ 15% of implied gains.
- **55–64%** of AI spending is operating exp. (users' subscriptions are providers' capital).
- Y = Revenue (not VA = R - M); AI services (M) can inflate the residual "revenue TFP."

Suggestions

- Account for purchased AI services contributions in revenue-TFP.
- Construct *VA* using non-labor operating cost (Fig. 5).
- **Next wave:** Separately measure AI subscriptions, training, and internal intangible investment.

Graham et al.: measurement and mechanisms

3. AI capital deepening is understated – National Accounts

Growth accounting

Why it matters

- Capital deepening explains only $\sim$ 15% of implied gains.
- **55–64%** of AI spending is operating exp. (users' subscriptions are providers' capital).
- Y = Revenue (not VA = R - M); AI services (M) can inflate the residual "revenue TFP."

Suggestions

- Account for purchased AI services contributions in revenue-TFP.
- Construct *VA* using non-labor operating cost (Fig. 5).
- **Next wave:** Separately measure AI subscriptions, training, and internal intangible investment.

4. Are the proposed channels identified?

Interpretation / externalities

Why it matters

- Table 3 relates **motives to gains**, but growth opportunities may drive *both* adoption motives and reported results.
- "Reaching customers" could shift sales from rivals: private $\neq$ aggregate productivity gains.

Suggestions

- Add firm size, age and pre-2025 growth.
- **Next wave:** Ask market share, total market expansion, competitors' AI use, products/customers generate AI-sales.

Baslandze, Graham et al.: policy implications

1. **Measurement: rented AI shows up as TFP, not investment**
   - With potential implications for National Accounts.
2. **Labor: redeploy and retrain clerical workers; talent/training top adoption barrier**
   - EMDEs: AI may lower the cost of scaling up (Felix-Rodrigo 26) but erode an entry-level job ladder.
3. **Macro: productivity gains of AI might come from slow maturity efforts like innovation or client acquisition**
   - Other drivers of the productivity paradox of AI: measurement (Brynjolfsson et al. 21), prices, reorganization (Babina 26, Rodrigo 23)?

Lashkari et al.: the race between AI costs and capabilities

Each task variety ν uses labor or an AI model m from a menu (larger m: more capable, more compute):

$$y(\nu) = \begin{cases} \dfrac{\ell_L(\nu)}{\tau}, & q(\nu; m) < q^* \qquad \text{AI is incapable} \\ \dfrac{\ell_L(\nu)}{\tau} + \min\left\{ \dfrac{\kappa(\nu)}{\gamma(\nu; m)}, \dfrac{\ell_K(\nu)}{\omega\,\tau} \right\}, & q(\nu; m) \geq q^* \qquad \text{AI is capable} \end{cases}$$

A capable variety is automated iff the machine bundle (compute + overhead labor) is cheaper:

$$R\,\gamma(\nu; m) + \omega\,\tau\,w \;\leq\; \tau\,w \quad\Longleftrightarrow\quad \gamma(\nu; m) \;\leq\; \gamma^* \equiv \tau\,(1-\omega)\,\frac{w}{R}$$

Automated share = capabilities × cost advantage; firms choose the model that minimizes unit cost:

$$\varphi_K(m) = \underbrace{\big(1 - H(q^*; m)\big)}_{\text{capability share}} \times \underbrace{G(\gamma^*; m)}_{\text{cost advantage}}, \qquad m^* = \arg\min_{m \in M} c(w, R; m)$$

Lashkari et al.: quantification and results

Capabilities and compute measured from worker evaluations of LLM outputs, indexed by model capability (ECI):

$$\underbrace{\text{grade}_{j,m} = Q(\text{ECI}_m, \tau_j) + u_q}_{\text{capability}} \qquad \underbrace{\log \gamma_{j,m} = \log T(\text{ECI}_m, \tau_j) + \log P(m) + u_\gamma}_{\text{compute per answer}}$$

Lashkari et al.: quantification and results

Capabilities and compute measured from worker evaluations of LLM outputs, indexed by model capability (ECI):

$$\underbrace{\text{grade}_{j,m} = Q(\text{ECI}_m, \tau_j) + u_q}_{\text{capability}} \qquad \underbrace{\log \gamma_{j,m} = \log T(\text{ECI}_m, \tau_j) + \log P(m) + u_\gamma}_{\text{compute per answer}}$$

From survey to projections

MIT FutureTech survey: 70k evaluations, 7,404 tasks, 41 LLMs (grade 1–9, task length)

↓

Calibrate frictions ω_o to time saved by AI by occupation (Bick et al.), 2025–26

↓

Project 2026–35: compute prices, frictions, scaling, algorithmic progress (+10.4 ECI/yr)

Results

- 2026: AI does **6.1%** of task time; firms use smaller models than the frontier — **model choice** explains 75% of cross-task variation.
- Cheaper compute adds little (growth 1.1 → 1.2%/yr); falling frictions and scaling, **2.3%**.
- Algorithmic progress dominates: **9–13%/yr** growth; automation ≈60% by 2035.
- Labor share ≈ flat; AI capital share <1.5%; growth is mostly TFP.

Lashkari et al.: four questions to the quantitative model

1. **How much rides on the capability trend?**
   - Algorithmic progress (+10.4 ECI/yr, 93 models, 2023–26) lifts growth from 2.3% to 9–13%/yr: a three-year benchmark trend extended for a decade.
   - → Report outcomes as functions of capability path; backtest (fit 2023–24, predict 2026 adoption).
2. **Do worker ratings measure capability?**
   - Prolific raters grade one example per task; $R^2 = 0.03$ (0.46 with rater effects).
   - → Occupation-insider raters, inter-rater reliability, validation against measured productivity gains.
3. **What are the adoption frictions?**
   - ω is a calibrated residual; its -14.1%/yr trend comes from 17 occupations over two years (6.8% employment-weighted).
   - → Unpack it (organization, skills, finance, regulation), lot to learn over here!
4. **Who gains?**
   - One wage, frictionless mobility across occupations, no firm heterogeneity: flat LSh with no losers(!)
   - → Occupation-specific labor with switching costs; firm size; AI-supplier markups in compute prices.

Lashkari et al.: policy implications

1. **Frictions, not compute, are the policy margin**
   - Cheaper compute alone adds little; falling frictions and better models drive automation.
   - Policy: Invest in complementary capabilities (skills, managerial know-how, finance for intangibles, regulatory clarity) rather than subsidize compute (Rodrigo 23; Ayyagari et al. 26).
2. **Automation without a rising capital share**
   - Automation can reach $\approx$60% of task time while AI capital's share stays small: gains accrue as TFP and wages.
   - Policy: The "no losers" result rests on frictionless mobility; support retraining and job transitions.
3. **Developing economies: a deployment gap**
   - Lower wages shrink AI's cost advantage ($\gamma^* \propto w/R$), and frictions are likely higher.
   - Policy: Measure frictions locally (firm censuses) and lower SME adoption barriers; watch exposure through services trade (Kugler et al. 26).

Will AI's pace be set by the technology, or by the capacity to adopt it?

Bastos, Stapleton, Taglioni & Wei: is AI different?

573M Lightcast postings · 17 advanced economies · 2014–22 · 29 technologies · FactSet firm networks

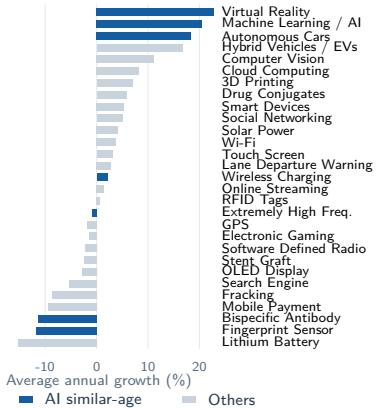
- **Rapid AI hiring:** AI-related postings grew 20.5%/yr, among fastest-growing technologies.
- **Spatial diffusion:** AI hiring is more dispersed across countries from the start and spreads faster within countries.
- **Income and distance:** Stronger GDP-pc gradient; distance matters less across countries, more within U.S. and across firms.
- **Networks:** AI pioneers have a larger initial hiring advantage. Buyer spillovers look weaker early on, but catch up over time.

Bastos, Stapleton, Taglioni & Wei: is AI different?

573M Lightcast postings · 17 advanced economies · 2014–22 · 29 technologies · FactSet firm networks

- **Rapid AI hiring:** AI-related postings grew 20.5%/yr, among fastest-growing technologies.
- **Spatial diffusion:** AI hiring is more dispersed across countries from the start and spreads faster within countries.
- **Income and distance:** Stronger GDP-pc gradient; distance matters less across countries, more within U.S. and across firms.
- **Networks:** AI pioneers have a larger initial hiring advantage. Buyer spillovers look weaker early on, but catch up over time.

Technology posting growth (Fig. 1)



Bastos, Wei et al.: what I would push on

1. **Different ages.** AI is observed at ages 0–7; most other technologies at ages up to 20.
   → **Align in event time** using the longer U.S. series; compare technologies at the same age.

2. **One treated technology.** Every "×AI" coefficient compares 1 technology with 28; SEs clustered by technology (or firm) may be unreliable for inference.
   → **Randomization inference**: treat each technology as "AI" in turn and show where AI ranks.

3. **What is measured?** Postings for AI-*producing* skills ≠ AI *use*; the data end in 2022 (**pre-ChatGPT**).
   → Extend to 2023–25; split tech vs. non-tech occupations.

4. **Pioneers defined by historical U.S. patents.** Do these locations capture the relevant sources of AI capabilities during 2015–22?
   → Test alternative pioneer definitions using recent publications, AI patents, and leading research centers.

What we learn from the three papers together

1. What binds?

Frictions and capabilities, limit automation (Lashkari).

Skills and maturity are top barriers (Graham).

Absorptive capacity, not value chains (Bastos).

⇒ **The adopter, not the technology, binds today.**

2. Adoption to gains

Graham: Perceived productivity gains > revenue-based gains.

Lashkari: Firms use smaller models; gains hinge on capability growth.

Bastos: AI hiring spreads widely among developed countries.

⇒ **Diffusion ≠ immediately** $\Delta Prod$**.**

3. Reallocation, so far

Expected: clerical ↓, technical ↑ (Graham).

Projected: flat labor share, small AI capital share (Lashkari).

Observed: AI hiring spreads across regions (Bastos).

⇒ **Effects shift across tasks, firms, and places.**

Emerging economies: who can build these capabilities—and who captures the gains?

Open questions for the research community

1. **Will capability gains become productivity gains?** Benchmarks and ratings vs. measured output at work.
2. **What are adoption frictions, and how fast do they fall?** Organization, skills, finance; the length of the J-curve.
3. **How is AI financed?** Subscriptions vs. intangibles; liquidity constraints and monetary transmission.
4. **Does AI create or reallocate talent?** New capabilities vs. competition for scarce skilled workers.
5. **Will AI foster convergence in EMDEs?** Low wages, small firms, and disruption of comparative advantage in services.

For policymakers

- Lower **adoption frictions** (skills, know-how, finance for intangibles, regulation), not compute subsidies.
- Rethink how AI-as-a-service is accounted for in **national accounts**.
- Prepare for **reallocation**: clerical and back-office ladders first.
- Capability, not connectivity: GVC links alone will not deliver AI.

Thanks