Measuring Firms’ Technology Use with Employees’ Job Data: Application to Artificial Intelligence Technologies

Tania Babina	Anastassia Fedyk	Alex Xi He	James Hodson
Columbia	UC Berkeley	Maryland	Jozef Stefan Institute

NBER Innovation Information Initiative Technical Working Group Meeting, December 2021

Measurement of Innovation: Problem

- Innovation is important for economic growth
  - Previous waves of **general purpose technologies** (GPTs) led to increased growth and productivity, but also structural changes within industries (e.g., electricity, IT)
  - GPTs can be used across many industries and solve wide range of business problems

- Measurement of technological innovation
  - **Innovation-producing firms:** could use patents
    - E.g., for measuring innovation in artificial intelligence (AI) methods, Google has AI patents
  - **Innovation-using firms:** big problem measuring in literature – often do not observe firms **using** new technological innovations or how intensively new technologies used
    - Previous technologies: US data comes from surveys
    - Many problems with surveys (e.g., sampling frame, response rates, answers depend on framing of questions)
    - Current technologies like AI: no firm-level data

Measurement of Innovation: Our Solution

- **Idea:** can learn a lot about what firms do by observing what employees they hire (e.g., “machine learning engineer”)
  - Past: cannot see what firms’ employees do
  - Present: observe large share of firms’ employees on internet via online resumes (e.g., LinkedIn)
    - Digitization of economic activity has been dramatically increasing
    - Use firms’ “digital footprints” of their employees

- **This presentation:** use information on firms’ employees to measure their use of AI
  - Can capture AI-producing and AI-using firms

- **Ongoing work:** use information on firms’ employees to measure their use of 60 broad technologies and 500 sub-technologies

Motivation

- Explosion in Artificial Intelligence (AI) investment in recent years
  - Due to data accumulation, dropping computational costs, and methodological advances
  - Most of AI investments are in machine learning (ML), computer vision (CV), and natural language processing (NLP)
  - AI=algorithms allowing to learn and make predictions from huge amount of high-dimensional data (text, speech, and image data)

- How do AI investments affect firms and industries?
  - Huge potential to transform economies, yet sluggish growth recently (Mihet and Philippon 2019; Haltiwanger, 2019)

- Potential channels
  - Productivity improvements (e.g., Aghion, Jones, Jones 2017; Brynjolfsson, Rock, Syverson 2018)
  - Product innovation (Hottman, Redding, and Weinstein 2016; Cockburn, Henderson, Stern 2018; Argente, Baslandze, Hanley, and Moreira 2020)

This Paper

- Goals
  - Do firm AI investments affect firm growth? Which firms benefit most? Mechanisms?
  - Macro implications: do industries investing in AI grow and change concentration?
- Our approach
  - **Data:** detailed, firm-level information on hiring of AI talent (235 million resumes) + demand for AI talent (180 million job postings)
  - **Measure:** data-driven measure of AI-relatedness of each skill and job, without pre-specifying AI-related words
  - **Identification:** use differences in firms' ex-ante exposure to AI-skilled labor supply
    - Universities strong in AI (pre-AI shock) can train more AI-skilled grads (post-AI shock)
    - Use ex-ante firm hiring networks from universities strong in AI research

Preview of Findings

1. Rapid growth in AI technologies across a wide range of industries

2. Firm-level: AI investments increase sales, employment, market valuation
   - Growth mainly through product innovation: more trademarks, product patents, and product portfolios
   - No increase in cost cutting, productivity or process patents
   - Growth is concentrated in ex-ante largest firms, which accumulate more data

3. Industry-level: AI predicts increased industry growth and industry concentration

Main contribution. AI increases firm growth through product innovation:
our micro-level evidence helps to unpack the black box of where “new projects” and investment opportunities come from

Data

- Job postings: Burning Glass Technologies, aggregator of US job postings
  - 180 million job postings
  - Comprehensive coverage of US online job openings in 2007 and 2010–2018
  - Detailed taxonomy of required skills
- Employment profiles (resumes): Cognism, aggregator of public profile information
  - 535 million global profiles and 235 Million US profiles
  - Job histories, skills, education, publications, patents, awards, references

Resume Data Strengths

1. Provides good coverage
   1.1 235 million US resumes: in 2018 cover 64% employees and 3.8 million firms
2. Resume data captures actual hiring, not just demand
3. Resume data captures AI-skilled labor onboarded via acquisitions
4. Able to measure and control for use of other IT and data technologies
5. Able to incorporate external AI software into our internal AI-investments measure

How to Identify AI-related Jobs in 100+ Million Job Postings?

Example of clearly AI-related job:

- Job title
  - Machine Learning Engineer
- Required Skills
  - Machine Learning, Artificial Intelligence, Computer Vision, Deep Learning, Python, C++, Research, Teamwork / Collaboration

How to Identify AI-related Jobs in 100+ Million Job Postings?

Example of clearly **not** AI-related job:

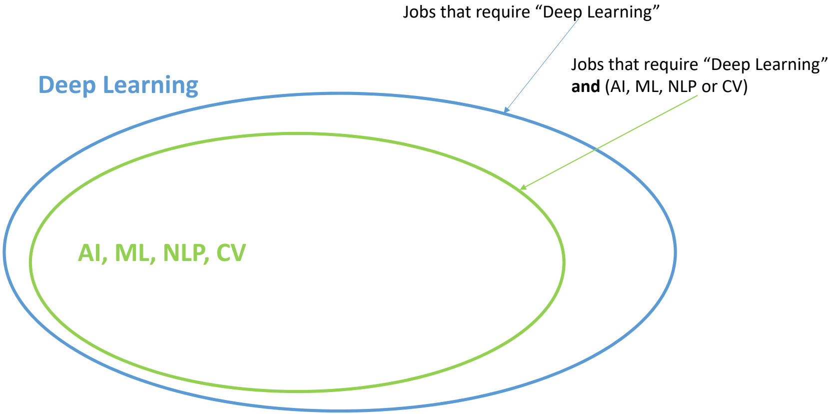
- Job title
  - Parking Attendant
- Required Skills
  - Teamwork / Collaboration, Communication Skills, Detail-Oriented, Scheduling, Heavy Lifting, Physical Abilities, Safety Codes, Snow Removal, Guest Services

How to Identify AI-related Jobs in 100+ Million Job Postings?

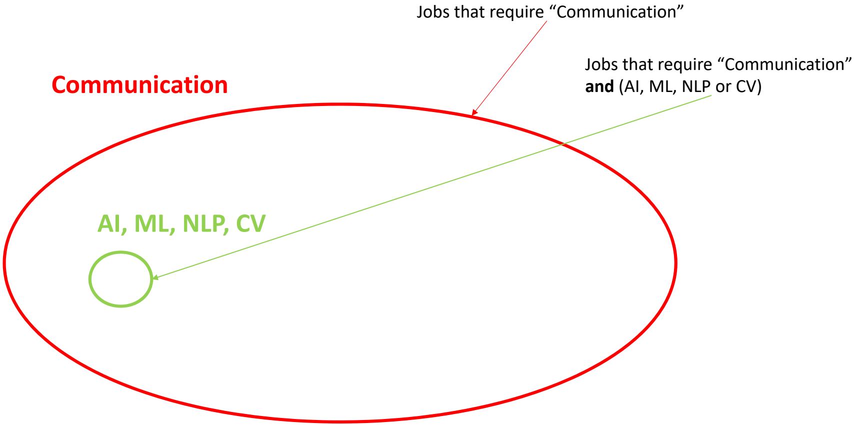
Example of job not obviously AI-related:

- Job title
  - Developer Programs Engineer
- Required Skills
  - TensorFlow, Kubernetes, Cloud Computing, Java, HTML5, Technical Writing / Editing, Teamwork / Collaboration, Writing, Troubleshooting

Some Skills **Have High Overlap** with AI-specific Skills



Some Skills **Do Not Have High Overlap** with AI-specific Skills



New Measure of AI Human Capital

- Identify relevant skills: Burning Glass
  - Four core AI skills: Artificial Intelligence, Machine Learning, Natural Language Processing, Computer Vision
  - AI-relatedness score of skill *s* =
    % of jobs requiring skill *s* that also require at least one core AI skill

Highly related (score>0.7) (N=68)	**Less related (0.05<score<0.7)** (N=533)	**Not related (score<0.05)** (N=13,577)
Tensorflow (0.90)	Information retrieval (0.37)	Communication skills (0.003)
Unsupervised learning (0.89)	Logistic regression (0.26)	Microsoft Office (0.001)
Deep learning (0.86)	Speech recognition (0.22)	Lawn mowing (0.000)
Random forests (0.84)	Python (0.12)	

New Measure of AI Human Capital

Burning Glass job postings

A job posting is AI-related if average score of all required skills > 0.1

Example:
Required skills: Machine learning (1), Text mining (0.63), MapReduce (0.29), Apache Hadoop (0.21), Data mining (0.16), Software engineering (0.04), Research (0.01), Communication skills (0.00)
Average score: **0.29**

Cognism online resumes

A job in a resume is AI-related if it contains a highly related skill (score$>$0.7) in profile

Examples:
Job title: “Senior Machine Learning Developer"
Job description: “develop Chatbots using Python with scikit learn, tensorflow and deep learning models..."
Publications: “A New Cluster-Aware Regularization of Neural Networks"
Patents: “Systems and methods for prime design using machine learning"

- Match employers to Compustat and calculate % of AI jobs at firm level
- AI measures from two datasets highly correlated and yield consistent results

Validation. Our Narrow AI Continuous Measure Does Well!

Example of clearly AI-related job:

- Job title
  - Machine Learning Engineer
- Required Skills
  - Machine Learning, Artificial Intelligence, Computer Vision, Deep Learning, Python, C++, Research, Teamwork / Collaboration
- **AI measure: 0.52**

Validation. Our Narrow AI Continuous Measure Does Well!

Example of clearly **not** AI-related job:

- Job title
  - Parking Attendant
- Required Skills
  - Teamwork / Collaboration, Communication Skills, Detail-Oriented, Scheduling, Heavy Lifting, Physical Abilities, Safety Codes, Snow Removal, Guest Services
- **AI measure: 0**

Validation. Our Narrow AI Coninuous Measure Does Well!

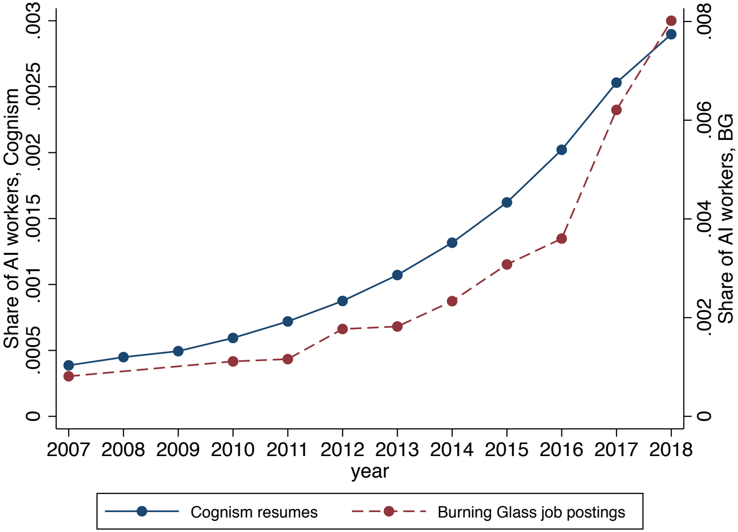
Example of job not obviously AI-related or not AI related:

- Job title
  - Developer Programs Engineer
- Required Skills
  - TensorFlow, Kubernetes, Cloud Computing, Java, HTML5, Technical Writing / Editing, Teamwork / Collaboration, Writing, Troubleshooting
- **AI measure: 0.15**

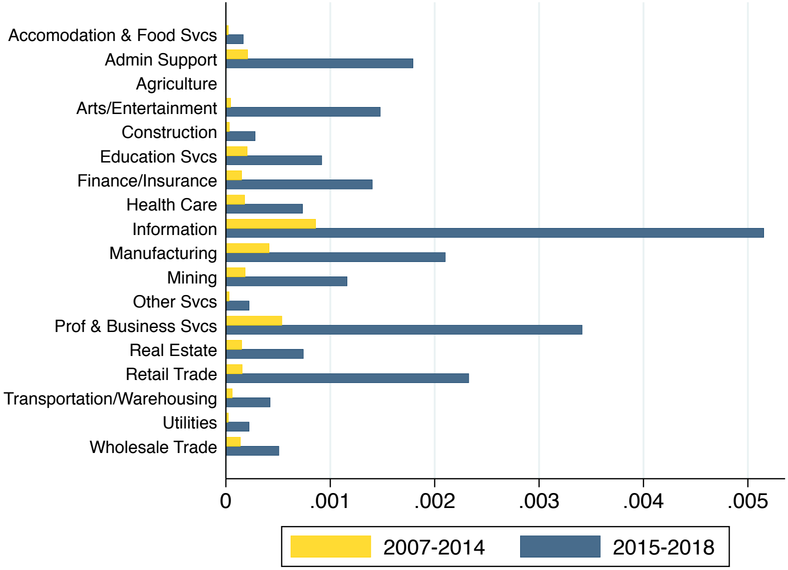
Validation. Job Titles with Top Values of Our AI Skill Measure

	Job Title	Avg. Continuous AI Measure
1	Artificial Intelligence Engineer	0.497
2	Senior Data Scientist - Machine Learning Engineer	0.394
3	Lead Machine Learning Scientist - Enterprise Products	0.369
4	AI Consultant	0.369
5	AI Senior Analyst	0.358
6	Machine Learning Engineer	0.315
7	Technician Architecture Delivery Senior Analyst AI	0.311
8	Artificial Intelligence Analyst	0.308
9	Software Engineer, Machine Learning	0.307
10	Artificial Intelligence Architect	0.303
11	Machine Learning Researcher	0.300
12	Computer Vision Engineer	0.293
13	Senior Machine Learning Engineer	0.286
14	Senior Machine Learning Scientist	0.281
15	Senior Software Engineer - Machine Learning	0.278
16	Senior Engineer II - Data Scientist	0.265
17	Senior Machine Learning Researcher	0.264
18	Artificial Intelligence Consultant	0.263
19	Computer Vision Scientist	0.256
20	Lead Machine Learning Researcher	0.255
21	Senior AI Engineer	0.248

Fast Growth in Share of AI Workers in the Last Decade



Increase in AI Workers in All Sectors



Main Results: Firm AI Investments and Firm Growth

- Following economics of technology literature (e.g., Acemoglu and Restrepo, 2020), use long-differences regression:

$$\Delta Y_i^{2010-2018} = \beta \Delta \mathit{ShareAIWorkers}_i^{2010-2018} + \gamma X_i^{2010} + \mathit{IndustryFE} + \varepsilon_i$$

	Δ Log Sales		Δ Log Employment		Δ Log Market Value	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Workers	0.204***	0.203***	0.240**	0.217***	0.232**	0.224***
	(0.070)	(0.060)	(0.097)	(0.078)	(0.094)	(0.078)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.225	0.428	0.239	0.417	0.222	0.364
Observations	1,051	1,051	1,051	1,051	1,009	1,009

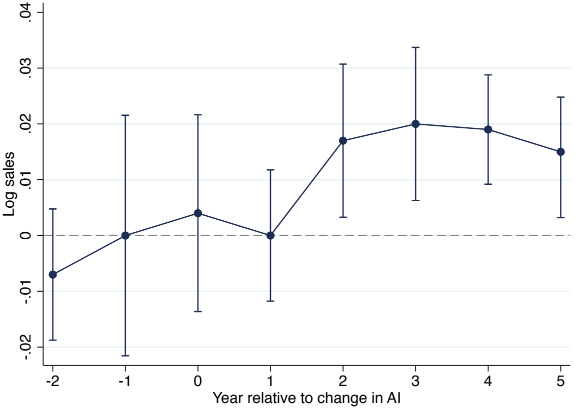
Controls. Firm-level: log employment, cash/assets, log sales, log industry wages, RD/Sales, and log markups.

Commuting-zone-level: average log wage, the share of college graduates, the share of routine workers, the share of workers in finance and manufacturing industries, the share of workers in IT-related occupations, and the share of female and foreign-born workers

Robustness: Dynamic Effects of AI on Firm Growth: Event Study Design

Distributed lead-lag model (Stock and Watson 2015; Aghion, Antonin, Budel, Jaravel 2020):

$$Y_{it} = \sum_{k=-2}^{5} \delta_k \Delta ShareAIWorkers_{i,t-k} + \mu_i + \lambda_{st} + \epsilon_{it}$$



Heterogeneity: Growth is Concentrated in Big Firms

- Large firms accumulate more data and more likely to benefit from AI
- Subsamples: firms ranked (terciles) based on employment as of 2010

	Δ Log Sales		Δ Log Employment		Δ Log Market Value	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Workers*Size Tercile 1	0.046**	0.001	0.041**	-0.013	0.059**	0.019
	(0.023)	(0.020)	(0.018)	(0.027)	(0.028)	(0.042)
Δ Share AI Workers*Size Tercile 2	0.209***	0.173***	0.208***	0.164***	0.192***	0.157***
	(0.051)	(0.047)	(0.045)	(0.056)	(0.042)	(0.050)
Δ Share AI Workers*Size Tercile 3	0.225***	0.215***	0.261**	0.227***	0.252**	0.236***
	(0.077)	(0.067)	(0.105)	(0.083)	(0.104)	(0.088)
NAICS2*Size tercile FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
T-test statistic	3.8	9.0	3.8	7.4	4.0	8.7
T-test p value	0.053	0.003	0.053	0.007	0.046	0.003

Robustness. Firm AI Investments (Internal + External) and Firm Growth

- Our AI measure captures internal AI investments (i.g., AI-using firms)
- Other ways to leverage AI: license external AI solutions (e.g., IPSoft Amelia)
- Measurement concern: internal and external AI investments are substitutes
- We construct AI measure including external AI solutions: search individual employees' job descriptions for mention of external AI software solutions

	Δ Log Sales		Δ Log Employment		Δ Log Market Value	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Workers	0.242***	0.228***	0.289***	0.265***	0.285***	0.279***
	(0.056)	(0.050)	(0.081)	(0.062)	(0.075)	(0.063)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.245	0.445	0.268	0.448	0.253	0.397
Observations	1,023	1,023	1,023	1,023	994	994

Robustness. Control for Other Technologies: IT, Robotics or Data Analytics

	Δ Log Sales	Δ Log Employment	Δ Log Sales	Δ Log Employment	Δ Log Sales	Δ Log Employment	Δ Log Sales	Δ Log Employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Share AI Workers	0.207***	0.217***	0.208***	0.219**	0.192***	0.203***	0.195***	0.207***
	(0.053)	(0.075)	(0.063)	(0.084)	(0.049)	(0.070)	(0.053)	(0.074)
Δ Share Non-AI IT Workers	0.138***	0.114**						
	(0.051)	(0.045)						
Δ Share Robot Workers			-0.016	-0.016				
			(0.032)	(0.040)				
Δ Share Non-AI Data Workers					0.136***	0.131***		
					(0.037)	(0.034)		
Δ Share Non-AI Data Analysis Workers							0.073**	0.068**
							(0.030)	(0.031)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.456	0.430	0.434	0.419	0.465	0.441	0.443	0.425
Observations	970	970	970	970	970	970	970	970

Mechanisms. How Do Firm AI Investments Affect Firm Growth?

1. **Product innovation**: AI allows firms to expand via product innovation

   - **Trademarks** are registered whenever new products or services are ready for commercialization (Hsu et al 2021)
   - **Product patents** protect products from competitors (Ganglmair et al 2021)
   - **Product fluidity** reflects updates to firms' product portfolios (Hoberg and Phillips 2016)

	Δ Trademarks		Δ Product Patents		Δ Product Fluidity	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Workers	0.144**	0.152**	0.221***	0.227***	0.148***	0.114***
	(0.065)	(0.077)	(0.035)	(0.039)	(0.036)	(0.035)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y

Mechanisms. How Do Firm AI Investments Affect Firm Growth?

2. **Productivity improvements**: so far, no association between AI and productivity

- **Sales/Worker** is model-free measure of productivity
- **(Revenue) TFP** is a standard measure of productivity based on Cobb-Douglas production function
- **Process patents** protect new processes from competitors (Ganglmair et al 2021)

	Δ Log Sales/Worker		Δ Revenue TFP		Δ Log Process Patents	
Δ Share AI Workers	-0.028	-0.006	-0.015	0.004	0.002	-0.009
	(0.038)	(0.035)	(0.033)	(0.035)	(0.064)	(0.075)
Ind FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y

Robustness: Instrumental Variable (IV) Analysis. Logic

- We offer new instrument to further bolster causal interpretation of AI on firm growth: instrument AI with differences in firms' ex-ante exposure to AI-skilled labor supply
  - Scarcity of AI-skilled labor is a large constraint on AI investment (CorrelationOne 2019)
  - Universities are key source of skilled labor: universities strong in AI research (pre-AI shock) can train more AI-skilled graduates (post-AI shock)
    - Academic research in AI has flourished for decades since John McCarthy coined the term in 1956, but commercial interest in AI was not big prior to the 2012 ImageNet challenge
  - Measure firms' exposure to AI-strong universities using firm-university hiring networks
    - Hiring networks are highly persistent, and firms with (ex-ante) strong ties to (ex-ante) AI-focused universities will have better access to AI-trained students

Robustness: Instrumental Variable (IV) Analysis. Concerns

- Key concern #1: ex-ante AI-strong universities also strong in computer science (CS)
  - Solution: measure and control for firms' exposure to CS-strong universities in all specifications
- Key concern #2: firms build hiring networks in expectation of AI-labor need
  - Solution: find that this is not true in the data

Instrumental Variable (IV) Analysis. Data

- (i) **Data on AI-related research**: Microsoft Academic Graph — publications data
  - Identify AI journals/conference proceedings and their authors as AI researchers
  - Get their university affiliation, and aggregate AI and non-AI researches within university
  - Identify AI-strong universities as those with AI researches above 95th percentile 2005-2009
  - For robustness, also identify (non-AI) CS researches to make sure the effects are specific to AI, and not CS, faculty

- (ii) **Data on firm-university hiring networks**: Cognism — resume data
  - Firms' employee composition in 2010: number of STEM graduates from each university
    - Use STEM to construct network likely more relevant for AI connections
  - High-quality data
    - On average 51% of the university graduates (from IPEDS) are captured by Cognism data
    - Cognism number of graduates is 73% correlated with number of graduates in IPEDS

IV Analysis. First-stage Results

- Firms' ex-ante connections to AI-research-strong universities predict firms' future hiring of AI-skilled workers
  - All columns controls for firms' exposure to (non-AI) Computer Science faculty via (ex-ante) hiring networks
  - Column 2 adds baseline controls (the share of college graduates, the share of workers in IT-related occupations, etc.)
  - Column 3 controls for pre-trend in firm growth, and column 4 adds for state FE

	Δ Share of AI Workers			
	(1)	(2)	(3)	(4)
Instrument	0.594***	0.392***	0.420***	0.439***
	(0.155)	(0.089)	(0.093)	(0.082)
Industry FE	Y	Y	Y	Y
CS Control	Y	Y	Y	Y
Additional Controls	N	Y	Y	Y
Control Pre-trend	N	N	Y	Y
State FE	Y	N	N	Y
F Statistic	14.7	19.4	20.5	28.6
Observations	1,001	1,001	777	773

IV Analysis. Second-stage Results

- Instrumented firm AI investments predict firm growth
  - All columns controls for firms' exposure to (non-AI) Computer Science faculty via (ex-ante) hiring networks
  - Column 2 adds baseline controls (the share of college graduates, the share of workers in IT-related occupations, etc.)
  - Column 3 controls for pre-trend in firm growth, and column 4 adds for state FE

	Δ Log Sales				Δ Log Employment				Δ Log Market Value			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Δ Share AI Workers	0.311***	0.446***	0.407***	0.261*	0.426***	0.690***	0.556***	0.270*	0.345**	0.391**	0.319**	0.180
	(0.097)	(0.123)	(0.127)	(0.149)	(0.140)	(0.214)	(0.177)	(0.160)	(0.136)	(0.162)	(0.161)	(0.185)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
CS Control	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Control Pre-trend	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
State FE	N	N	N	Y	N	N	N	Y	N	N	N	Y
F Statistic	14.7	19.4	20.4	29.2	14.7	19.4	20.4	29.2	14.8	19.6	20.3	30.2
Observations	1,001	1,001	777	773	1,001	1,001	777	773	963	963	753	746

Conclusion

- Data on firms' employees allow to measure use of technologies based on human capital
- AI investments predict higher firm growth
  - AI-investing firms grow through increased product innovation
  - AI allows large firms to grow larger, consistent with large firms accumulating more data
  - No significant impact on productivity
- Industry-level: AI predicts increased industry growth and industry concentration
  - The impact of AI is most pronounced among the largest firms
  - Our results highlight the role of AI technology in shaping industry concentration and reinforcing winner-take-most dynamics

Main contribution. AI technology increases firm growth through product innovation: our micro-level evidence helps to unpack the black box of where "new projects" and investment opportunities come from

Appendix

Other Application: Measures of Organizational Structure

1. **Departmental composition:** % of firms' workforce in each department
   - Does ex-ante department composition predict AI adoption?
   - How did the composition change with AI adoption? Any organizational functions (e.g., R&D) complementary to AI? Any organizational functions (e.g., HR) substitutable by AI?

2. **Employee seniority:** classify employees into levels of seniority ranging from lowest (entry-level job) to highest (C-suite executive)
   - Were junior or senior employees cut down with the advent of AI?

3. **Hierarchical depth and width:** use the reconstructed firm hierarchies to measure the total number of layers (depth) and the average # of employees per layer (width)
   - Did less hierarchical firms (with fewer layers, more employees per layer) see fewer barriers to AI adoption?
   - Did AI adoption make firms more or less hierarchical?