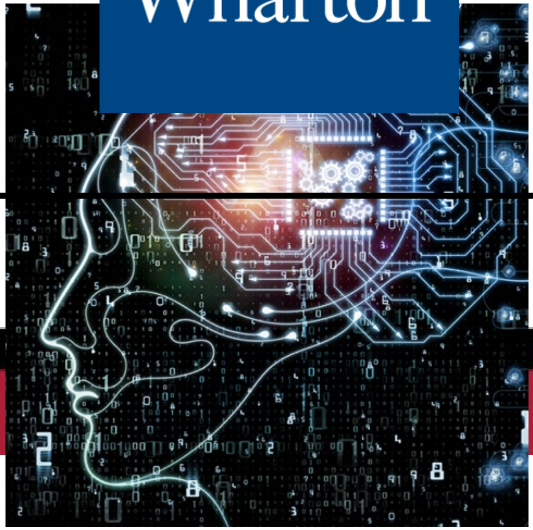
Quantifying the impact of AI invention on productivity

Sagar Baviskar, Lee Branstetter, Cameron Drayton, Brian Fujiy, Eduard Hovy, Prasanna Tambe, Liujie Wu

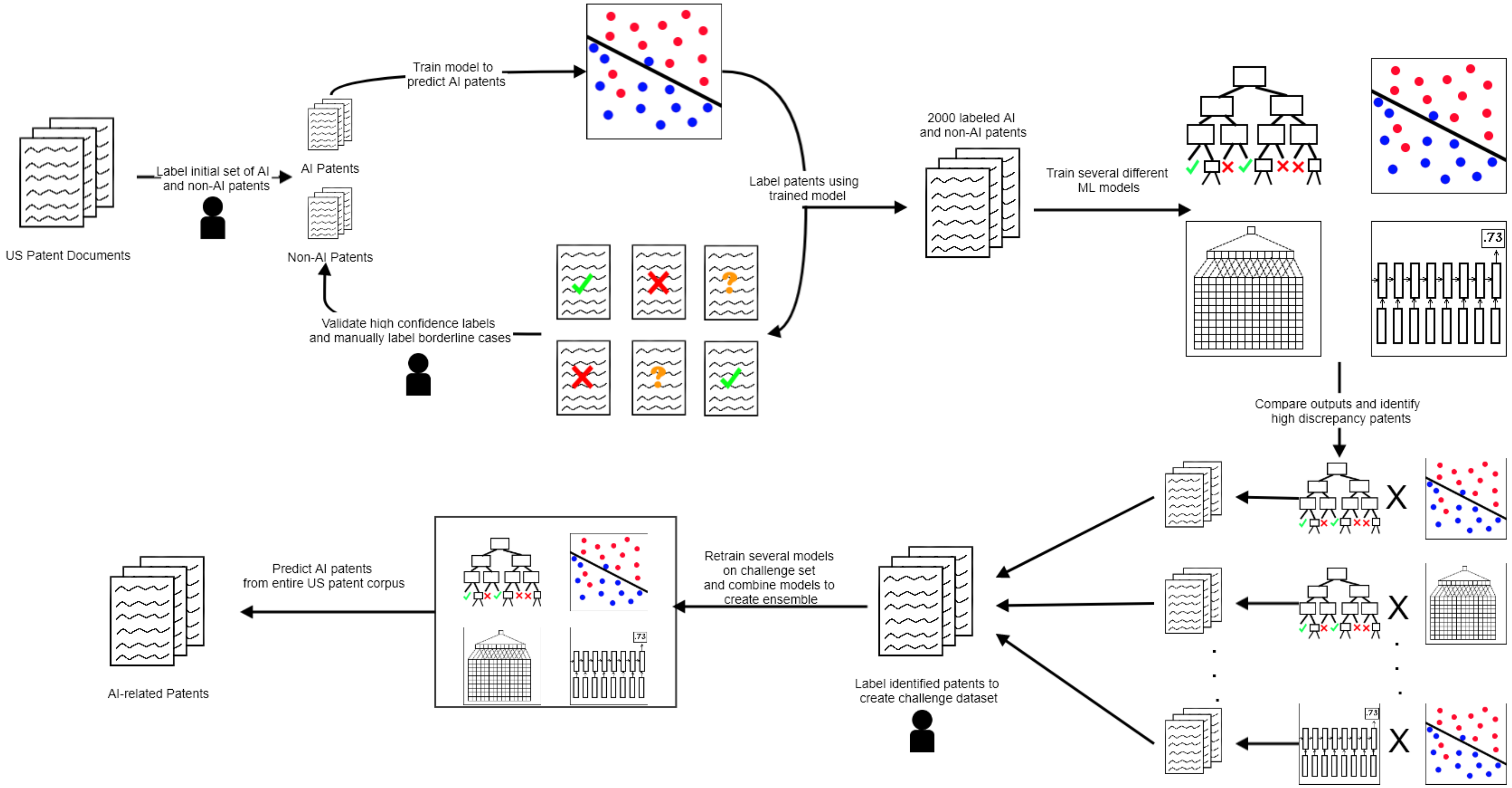


Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau or the National Bureau of Economic Research. All results have been reviewed to ensure that no confidential information is disclosed. The DRB codes for this project are: DRB-B0027-CED-20190205, CBDRB-FY19-414, and CBDRB-FY20-105.

This project provides new measures of AI-related innovation that complement existing studies

- We build on and extend the efforts of Cockburn et al. (2019), Webb et al. (2019), and Giczy et al. (2021) to identify AI-related patents…
- And we examine the impact of these AI inventions on the AI-inventing firms, using Compustat and Census micro-data for U.S.-based firms.
- This complements efforts to directly measure AI adoption (McElheran et al., 2024; Zolas et al, 2019)…
- And efforts to measure the impact of the hiring of AI workers as reflected in online job ad data (Babina et al., 2022, 2024)…
- And more recent experimental studies of the impact of particular kinds of AI in particular work contexts (Brynjolfsson et al., 2023; Noy and Zhang, 2023)

We began this project years ago, taking a machine learning approach….



Which was completely upended by the rise of LLMs and GenAI…

ChatGPT

Claude

Gemini

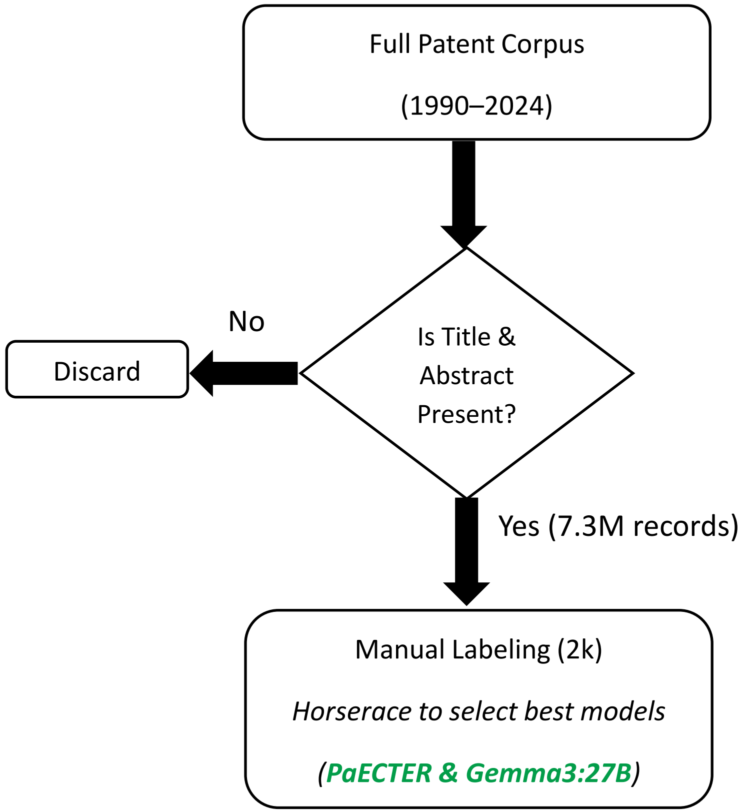
LLaMA by Meta

deepseek

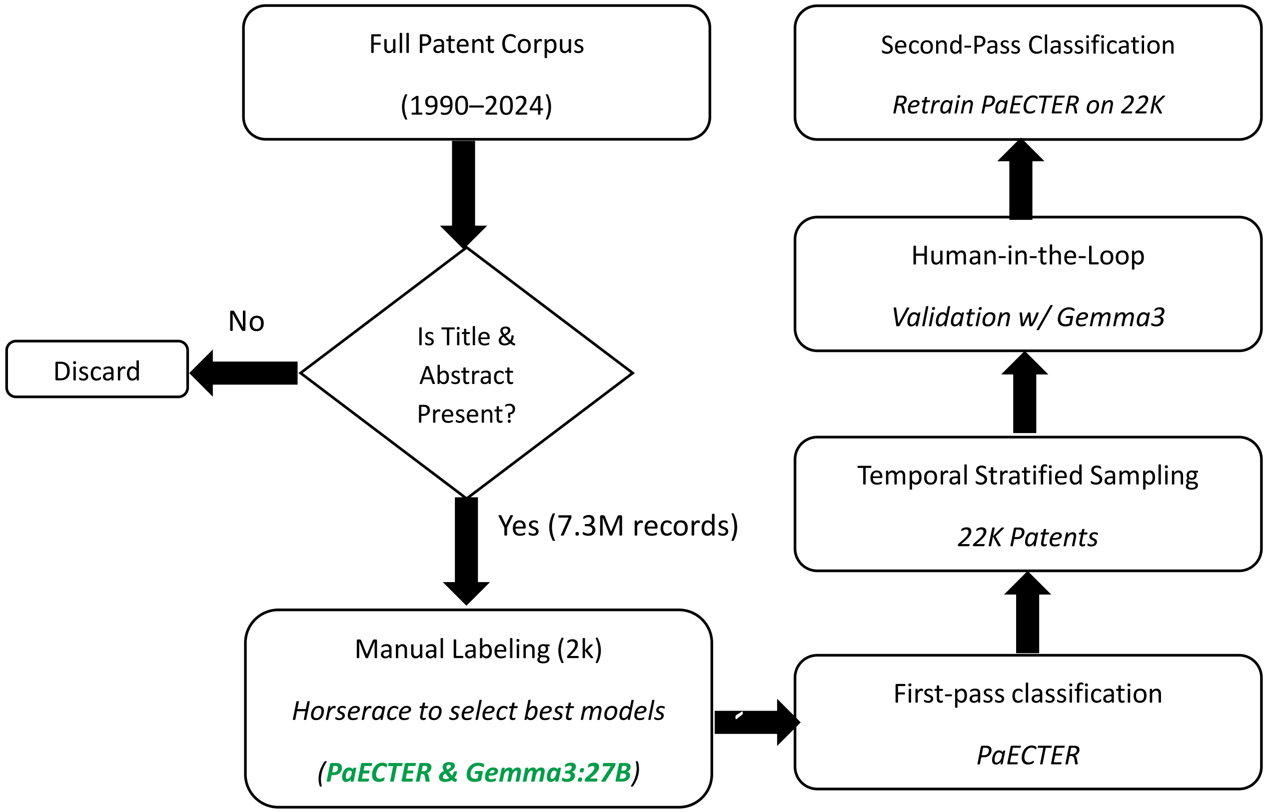
Using the PaLLaFi framework to find AI patents…

Patent Labeling via Language Models and Fine-Tuned Inference

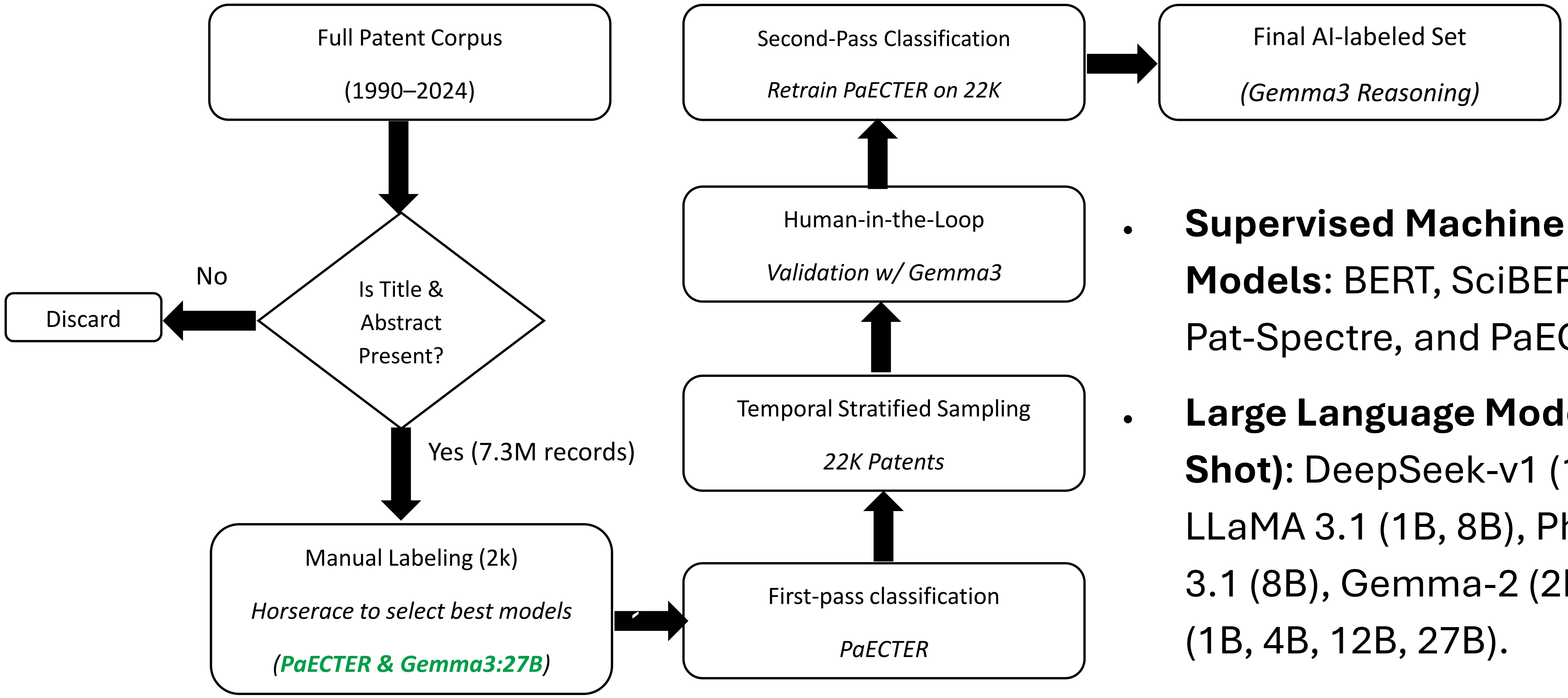
We download U.S. patents, hand-label a small sample as AI or not AI, and train and compare leading models…



We use the best models to expand our training data set…

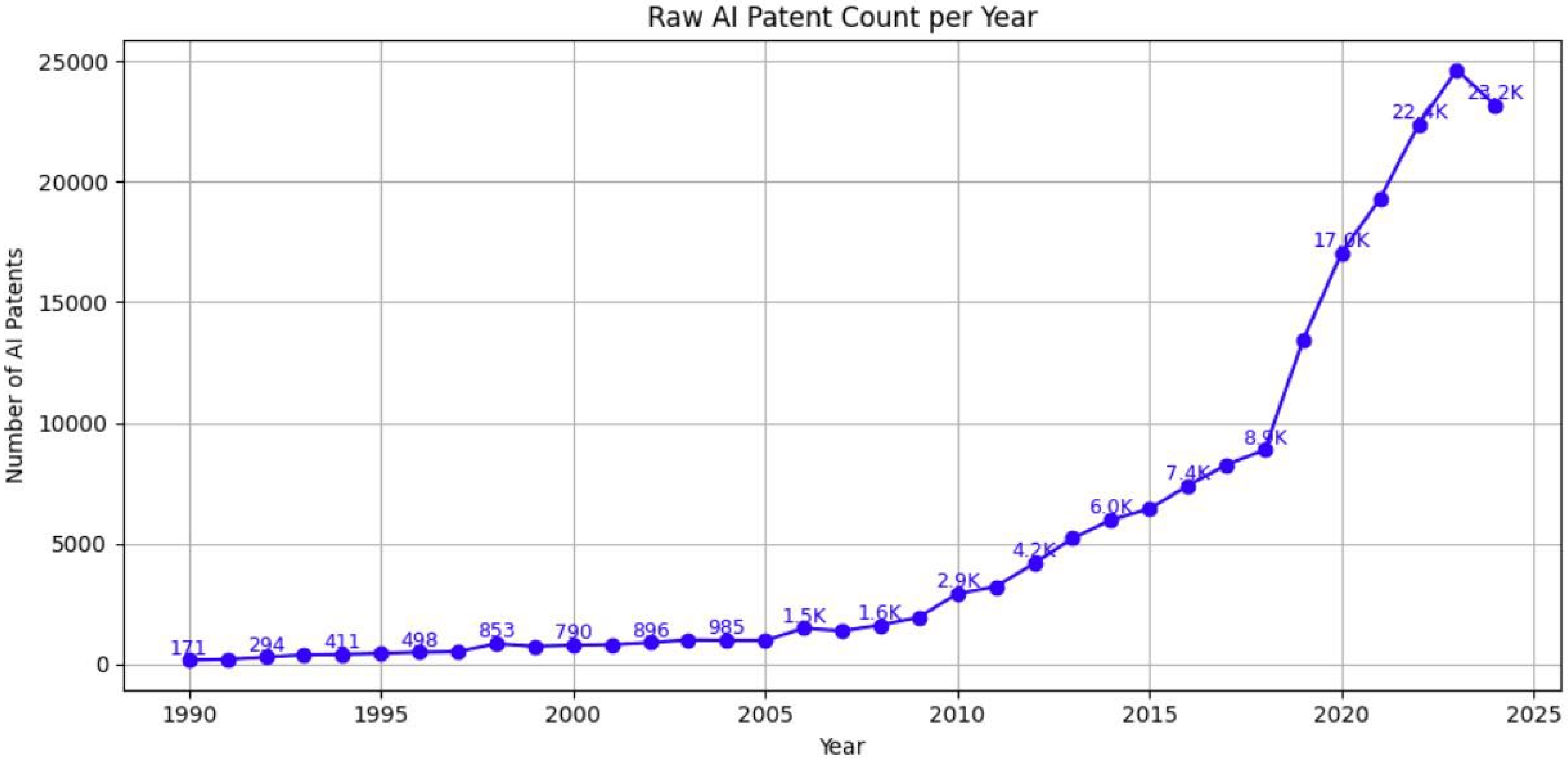


And then we use our best models, trained on expanded data, to label the entire USPTO utility patent corpus from 1990-2024

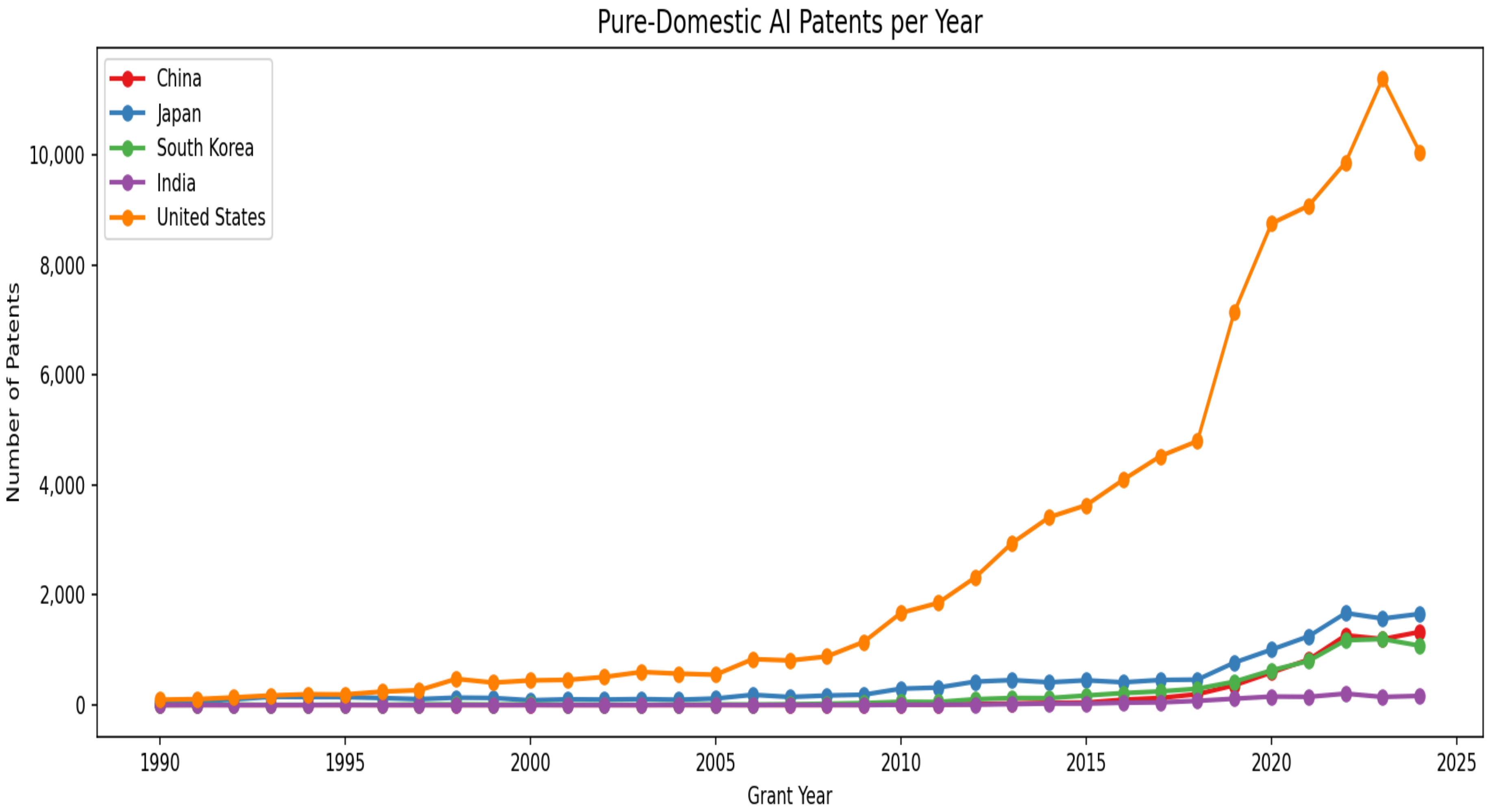


- **Supervised Machine Learning Models**: BERT, SciBERT, Longformer, Pat-Spectre, and PaECTER.
- **Large Language Models (Zero/Few-Shot)**: DeepSeek-v1 (1.5B, 8B, 32B), LLaMA 3.1 (1B, 8B), Phi-3 (8B), Granite 3.1 (8B), Gemma-2 (2B, 9B), Gemma-3 (1B, 4B, 12B, 27B).

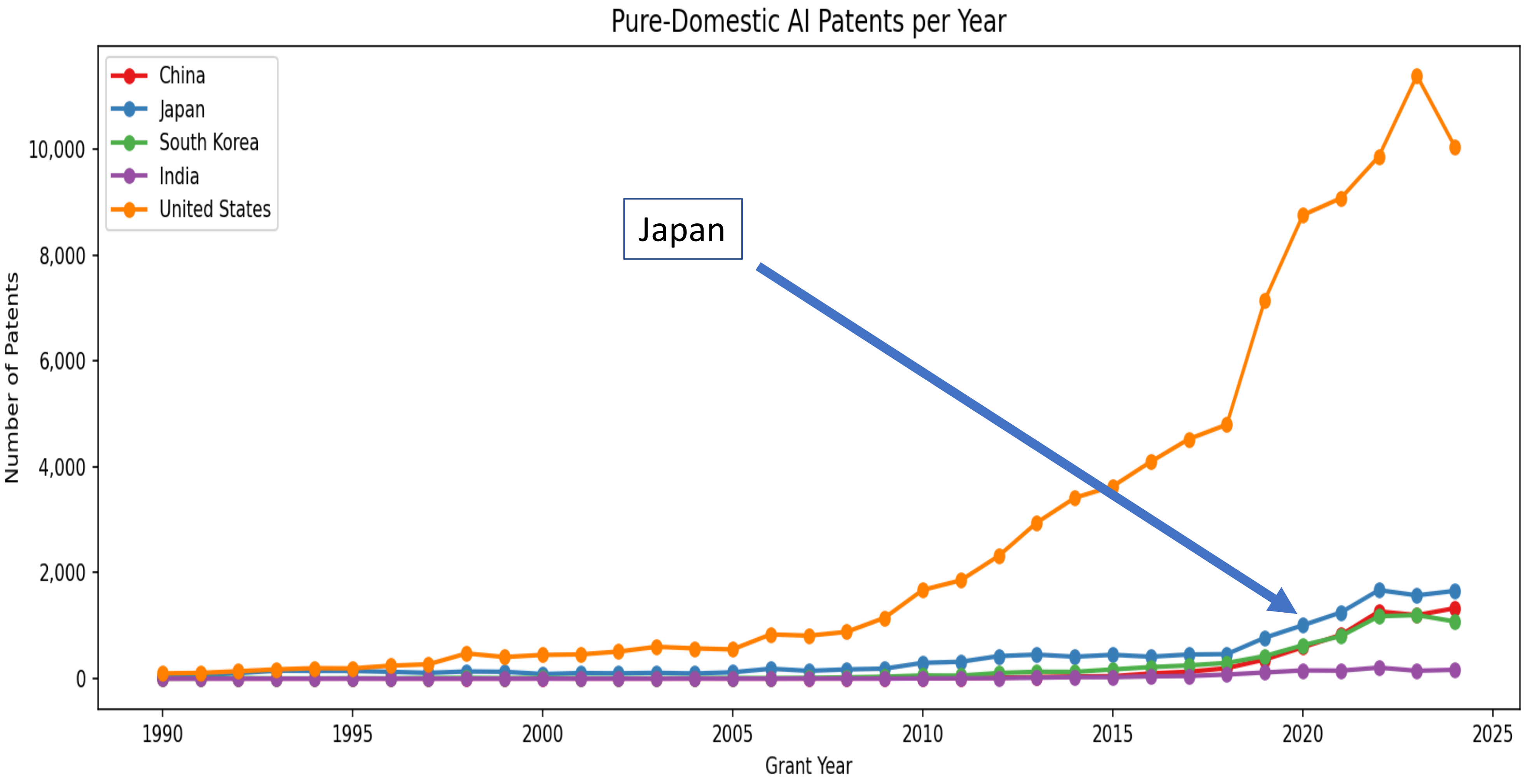
AI patenting has grown rapidly over our sample period



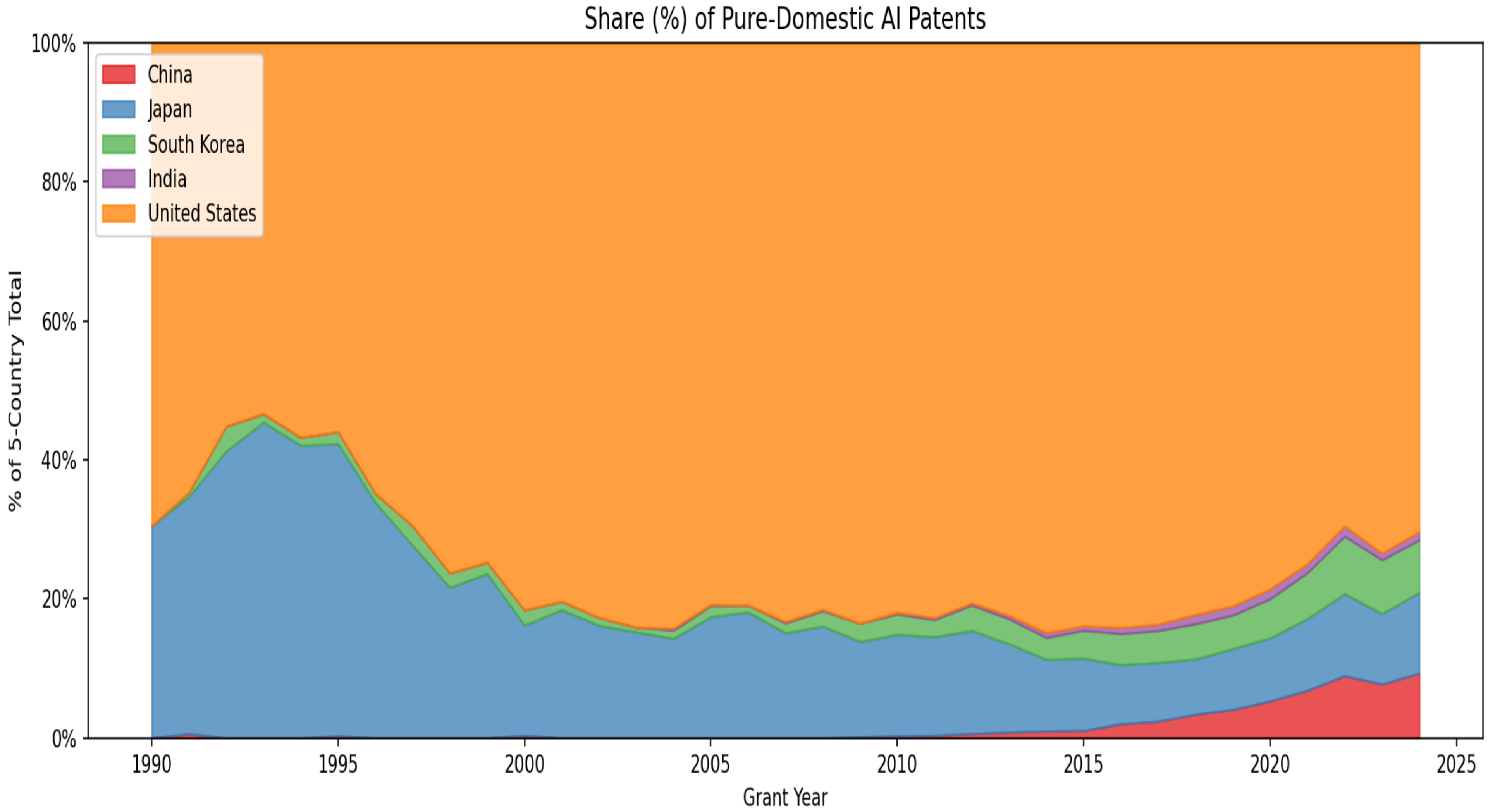
With especially rapid growth in patents awarded to U.S. firms and inventors…



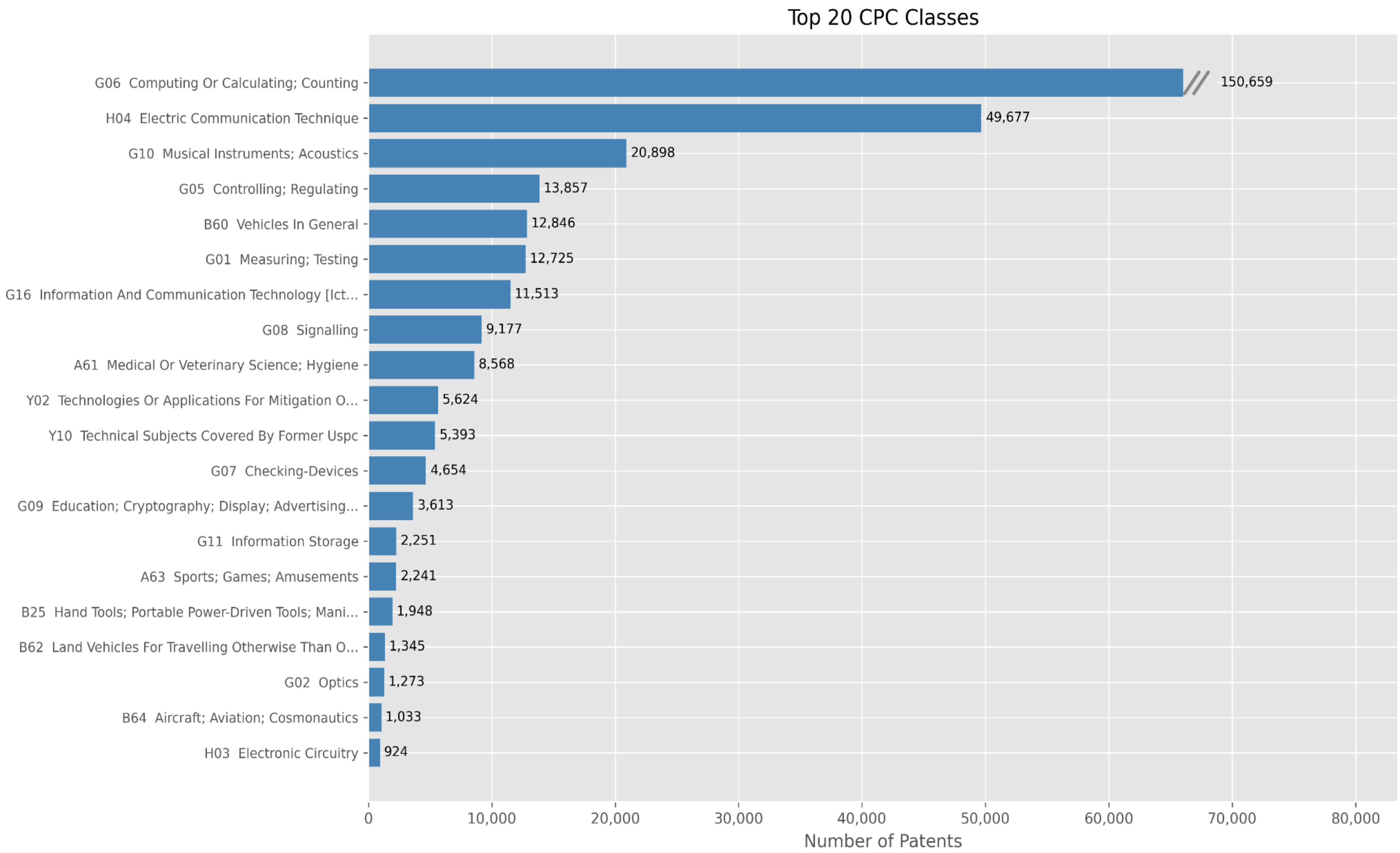
But note that Japan comes in second (ahead of China and South Korea…)



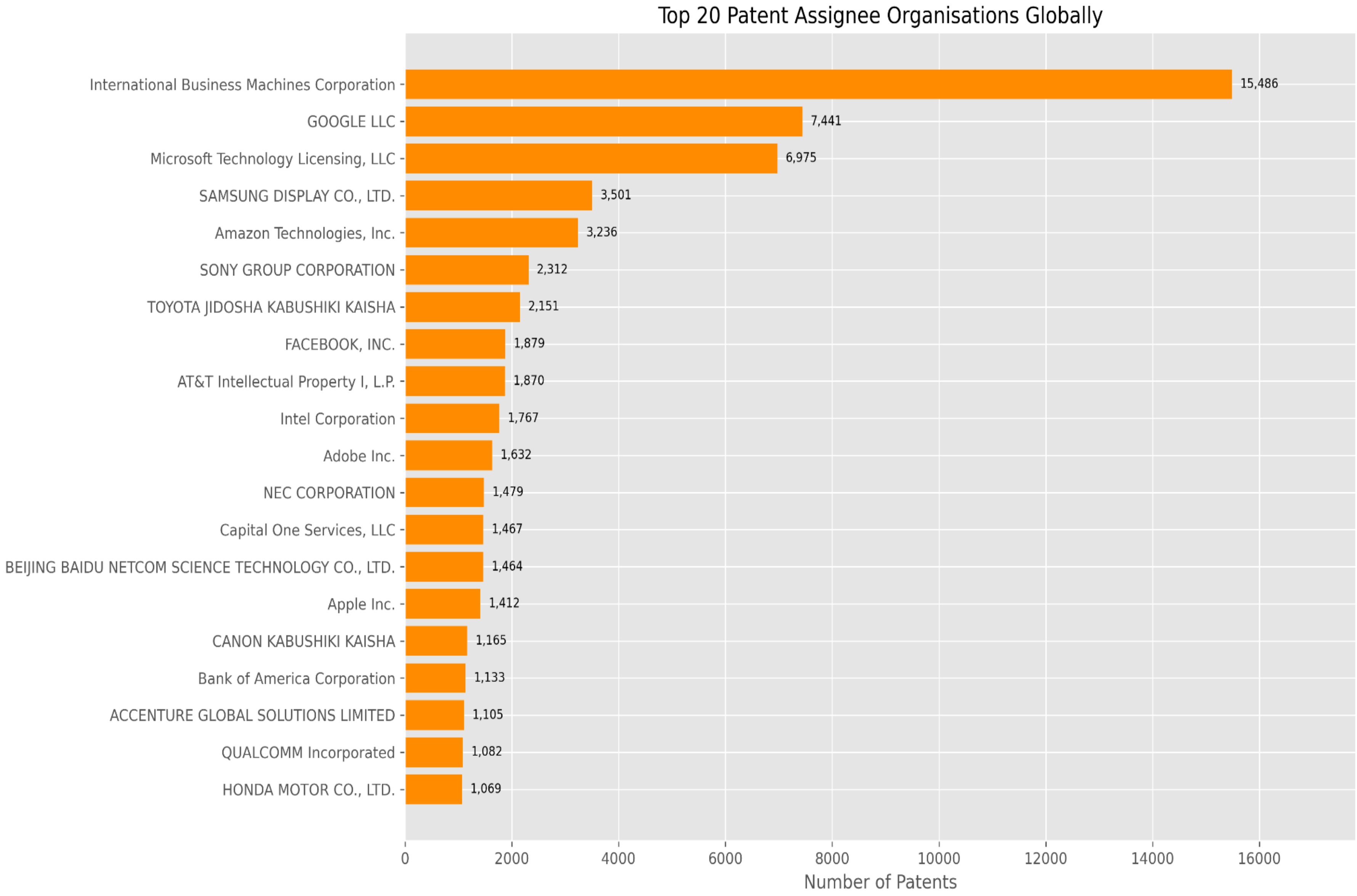
And in the early-to-mid 1990s, Japan was a much closer second to the U.S….



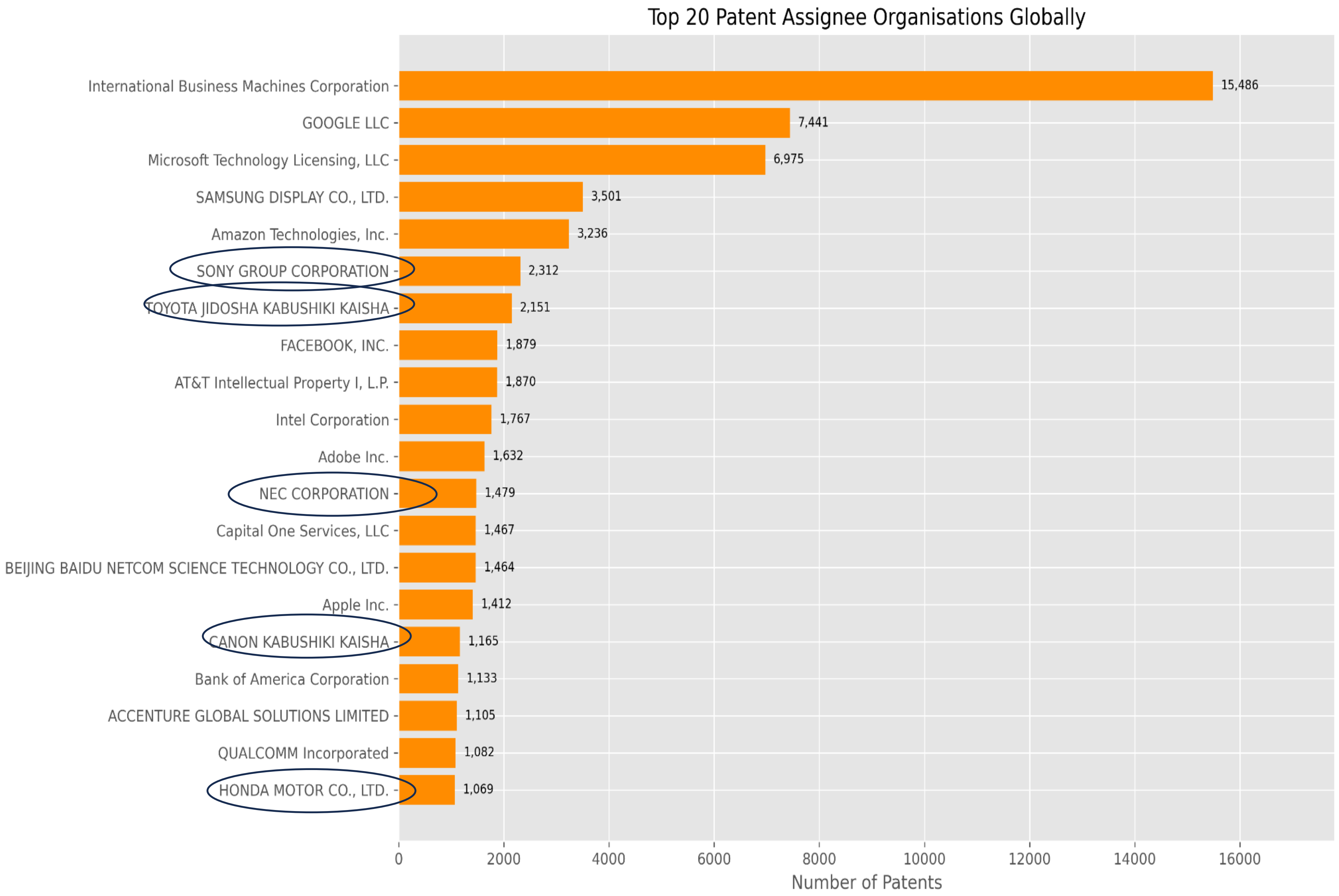
AI patenting is widely distributed across patent classes…



And across firms…

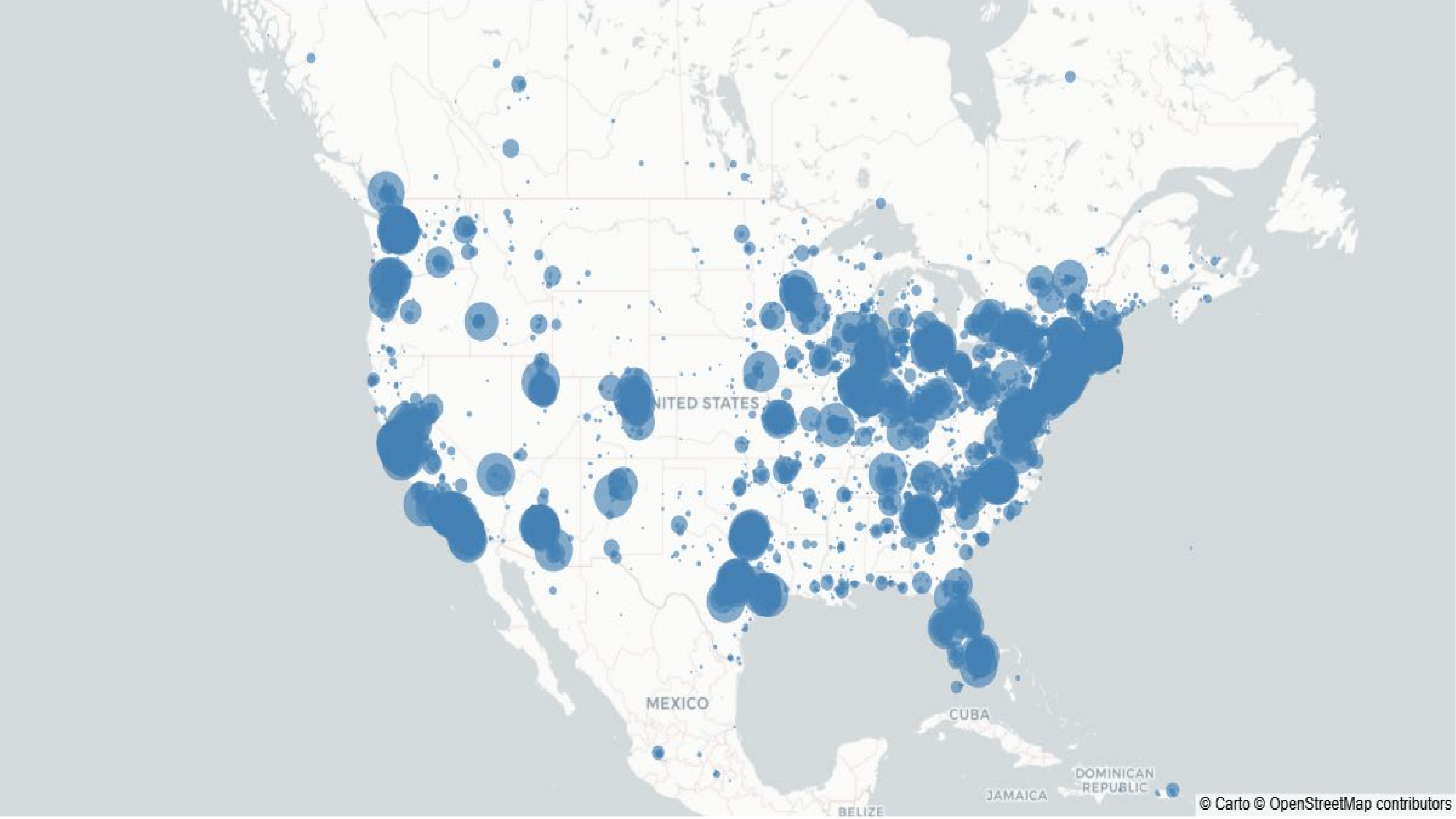


Including some Japanese firms!



A map of the location of US AI inventors

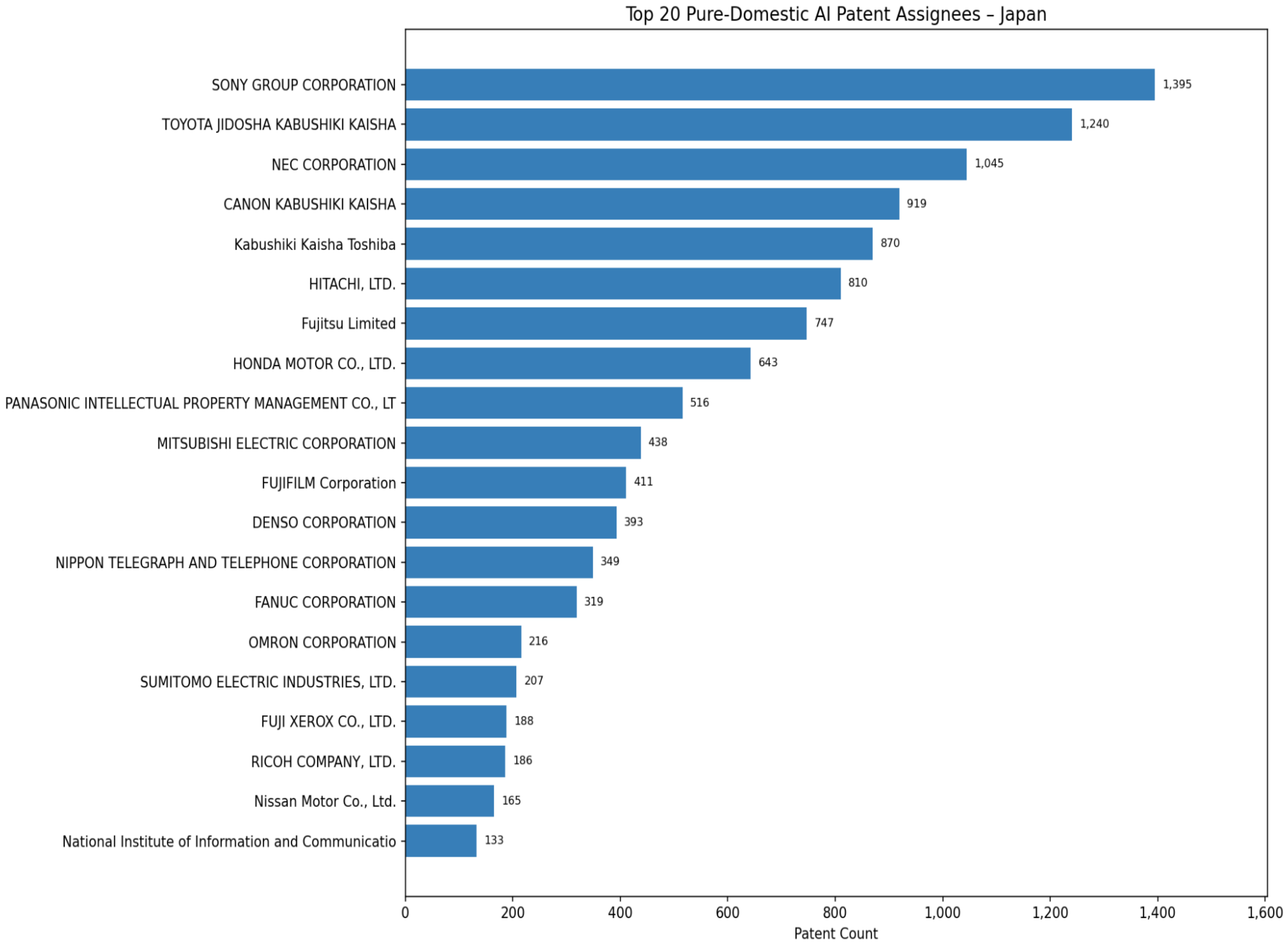
Inventor Locations – United States



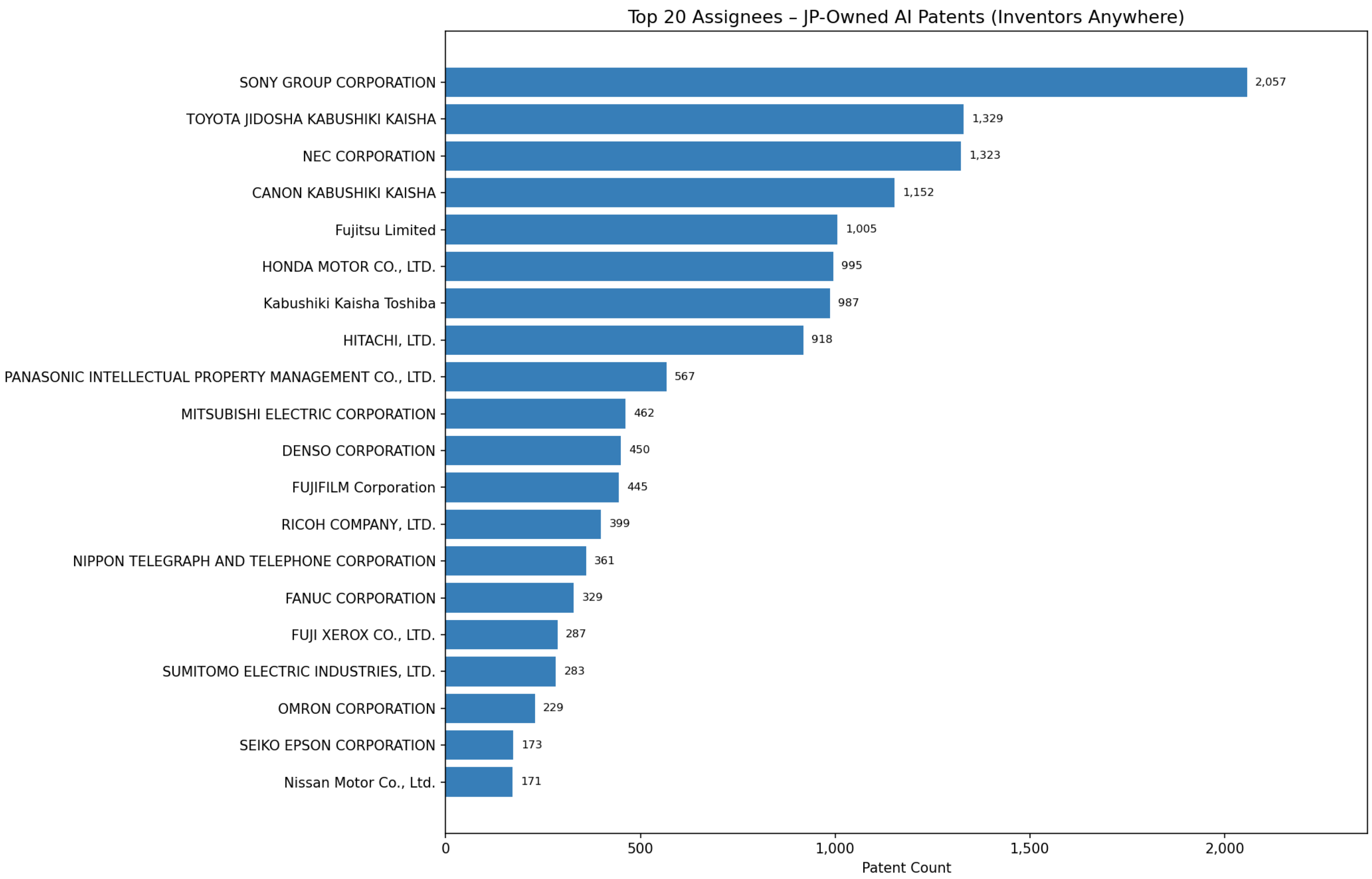
A map of the location of Japanese AI inventors

Inventor Locations – Japan

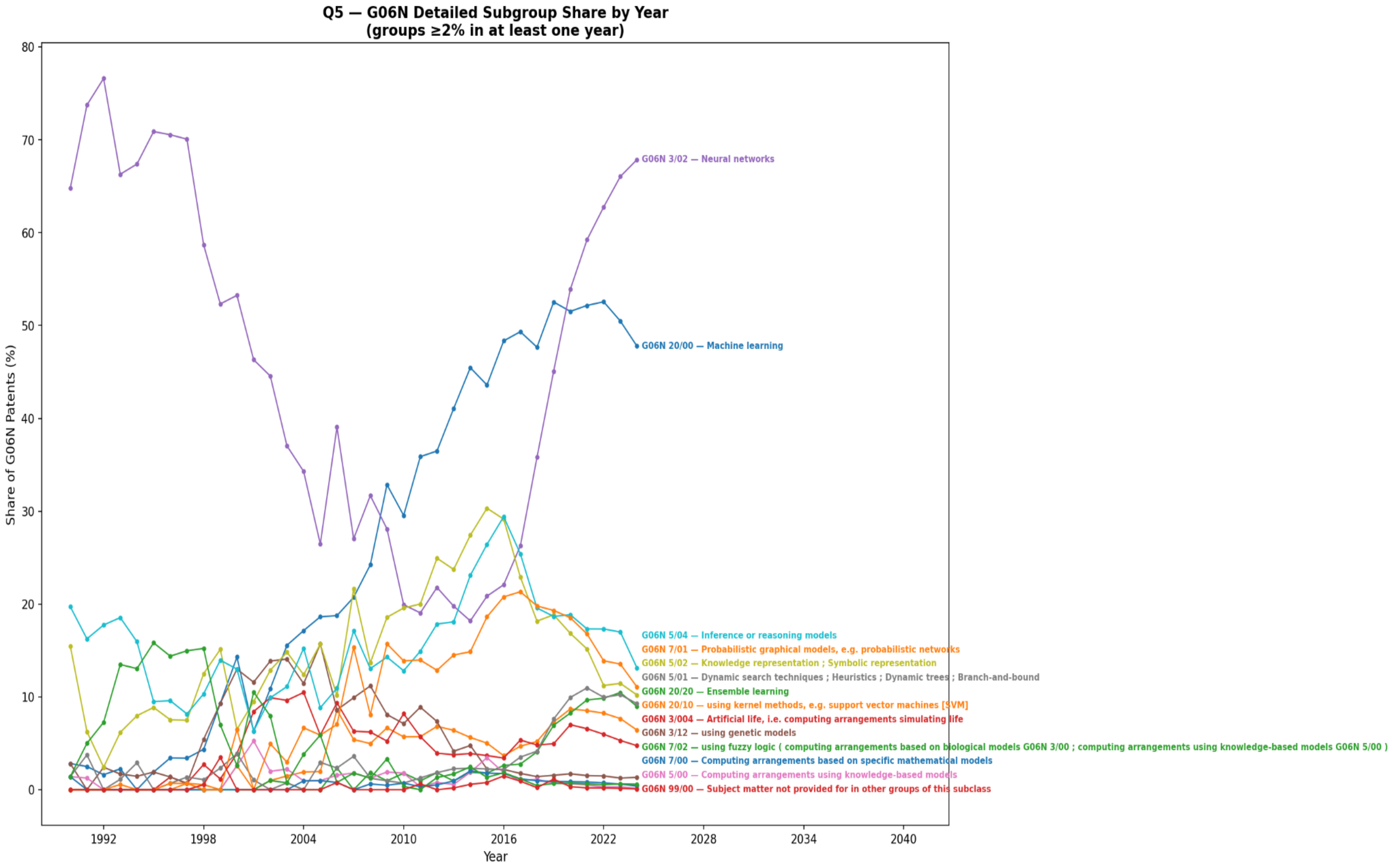
One in-depth look at leading Japanese AI assignees



Another in-depth look at leading Japanese AI assignees



The technological content of AI is changing over time



Methodology for impact assessment

- Use existing USPTO-Census patent-to-firm crosswalk (Graham et al. 2018), augmented by our own work to link the AI patents with a Census FirmID
  - Incorporate various outcome measures including:
    - Employment, Revenue and Revenue per Employee (Revenue-enhanced LBD)
    - Value-added, Total Factor Productivity, Production Worker Share (ASM and CMF)
    - 90-10, 90-50 and 50-10 earnings ratio by firm (LEHD)

- Measure within-firm changes from AI innovations at the *extensive* and *intensive* margins.

- Construct a comparable subgroup of firms to which AI-inventing firms can be compared and perform an event-study analysis
  - Use propensity score/exact matching to construct a control group using size, age, industry (4-digit NAICS), payroll, and patenting activity as predictors
  - Event study centered around the timing of the first AI patent filed by the AI-inventing firm.

AI Invention and firm productivity

Table 5: Impact of AI Innovations on Firm Productivity, 1997-2018 (manufacturing only)

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln Total Value of Shipments per Employee		Ln Value Added Per Employee		Ln Total Factor Productivity (TFP)	
AI Treatment (1/0)	0.272***		0.225***		0.0830***	
	(0.0322)		(0.0337)		(0.0225)	
IHS AI Patents		0.148***		0.104***		0.0565***
		(0.0161)		(0.0166)		(0.0119)
Ln Capital Stock	0.310***	0.310***	0.301***	0.301***	-0.0633***	-0.0633***
	(0.00195)	(0.00195)	(0.00203)	(0.00203)	(0.000888)	(0.000888)
Age Bins	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,124,000	1,124,000	1,124,000	1,124,000	1,124,000	1,124,000
R-squared	0.674	0.674	0.921	0.921	0.706	0.706

Robust Standard Errors clustered at the 4-digit NAICS industry level. *, ** and *** denotes significance at the 5%, 1% and 0.1% respectively. Each regression includes a constant, age bins, multi-unit and multi-national indicator controls, which are not displayed here. Note that our Multi-Unit regressor drops out from the within-firm specification as the firm-identifier for multi-unit status does not change. Across firm effects are listed in the appendix.

Event study on employment and labor productivity (all firms, not just manufacturing)

Table 10: Impact of AI Innovations on Employment and Revenue, 1997-2018 (matched only)

	(1)	(2)	(3)	(4)
	Ln Employment		Ln Revenue per Employee	
AI Treatment (1/0)	Dropped	Dropped	Dropped	Dropped
Post AI Year	0.0206		-0.124***	
	(0.0144)		(0.0192)	
AI Treatment x Post AI Year	0.138***		0.164***	
	(0.0194)		(0.0263)	
Age Bins	Yes	Yes	Yes	Yes
Industry-Year Fixed Effects	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Observations	36,000	36,000	36,000	36,000
R-squared	0.975	0.976	0.787	0.787

Robust Standard Errors clustered at the 4-digit NAICS industry level. *, ** and *** denotes significance at the 5%, 1% and 0.1% respectively. Each regression includes a constant, which is not displayed here.

Does AI widen income inequality?

- Earlier generations of IT innovation dramatically expanded income inequality by raising demand for skill in the U.S. labor market.
- Will AI continue – or even worsen – this longstanding trend?
- We can examine the impact of AI on demand for production workers.
- By linking data on our AI-inventing firms to the LEHD, we can also examine the the association between AI invention and actual changes in within-firm earnings inequality.
- We measure earnings inequality using the 90^{th} percentile – 10^{th} percentile, 90^{th}-50^{th}, and 50^{th}-10^{th} income ratios.
- Earlier research on this important question has been ***predictive*** rather than observational.
- To the best of our knowledge, our study provides the first firm-level observational evidence on the relationship between AI invention and within-firm wage inequality.

Does AI invention increase within-firm income inequality (in event study models)?

Table 12: Impact of AI Innovations on 90-10, 90-50 and 50-10 Earnings Ratio, 1997-2019 (full matched set of firms)

	(1)	(2)	(3)	(4)	(5)	(6)
	90-10 Earnings Ratio		90-50 Earnings Ratio		50-10 Earnings Ratio	
AI Firm (1/0)	Dropped	Dropped	Dropped	Dropped	Dropped	Dropped
Post AI (1/0)	-0.0170		-0.00563		-0.0101	
	(0.0159)		(0.00825)		(0.0117)	
AI x Post (1/0)	-0.00464		-0.0135		0.00920	
	(0.0211)		(0.0109)		(0.0149)	
AI x Year = -2		-0.00633		-0.00442		0.00107
		(0.0274)		(0.0149)		(0.0193)
AI x Year = -1		0.00795		0.00647		-0.00394
		(0.0216)		(0.0114)		(0.0160)
AI x Year = 0	Dropped	Dropped	Dropped	Dropped	Dropped	Dropped
AI x Year = +1		-0.000185		0.00146		-0.00348
		(0.0181)		(0.00978)		(0.0134)
AI x Year = +2		0.00475		-0.0143		0.0161
		(0.0223)		(0.0123)		(0.0161)
AI x Year = +3		-0.0249		-0.0236		-0.00268
		(0.0242)		(0.0137)		(0.0171)
AI x Year = +4		-0.00885		-0.0306*		0.0177
		(0.0267)		(0.0148)		(0.0187)
AI x Year = +5		0.000276		-0.0217		0.0249
		(0.0290)		(0.0160)		(0.0201)
ln (Emp)	0.104***	0.105***	0.0422***	0.0432***	0.0646***	0.0646***
	(0.0130)	(0.0131)	(0.00743)	(0.00746)	(0.00784)	(0.00787)
Age Bins	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,500	33,500	33,500	33,500	33,500	33,500
R-squared	0.727	0.727	0.728	0.729	0.679	0.679

Robust Standard Errors clustered at the 4-digit NAICS industry level. *, ** and *** denotes significance at the 5%, 1% and 0.1% respectively. Each regression includes a constant, which is not displayed here.

Matching to Compustat data…

- **Compustat**:
  - Database of financial and market data on publicly traded companies provided by S&P Global Market Intelligence
  - Same period as patent years - **1990 to 2024**

- **Compustat-Patents Firm Matching:**
  - 70 Years of Patents Matched to Compustat Firms: Methodology and Insights About Firm Heterogeneity by Dyèvre, A. & Seager, O. (2023)
  - Amended with our firm matching
  - After matching - Total # of publicly traded firms: **~6,000**

Estimating standard firm-level production functions

- We estimate a log-linear Cobb–Douglas production function to quantify the role of AI innovation and AI human capital on productivity:

$$\log(Y_{it}) = \alpha.\log(K_{it}) + \gamma\log(L_{it}) + \delta\log(T_{it}) + Z`\theta + \mu_i + \lambda_t + \eta_s + \epsilon_{it}$$

Y_{it}: Firm output (Value added)
K_{it}: Capital input (estimated using Perpetual Inventory Method)
L_{it}: Labor input (employees)
T_{it}: Technological input — *AI patents* or *AI human capital*
Z : Controls (R&D expenses)
μ_i, λ_t, η_s: Firm, Year, and Industry (NAICS-2) fixed effects
ϵ_{it}: Error term

Taking multiple approaches to estimation of AI invention's impact on productivity

- **Intensive vs. Extensive Margin Analysis –** Examined both the growth in AI patent stock and the transition into AI patenting and
- **Firm Fixed Effects Models** – Measured within-firm changes in productivity
- **Propensity Score Matching –** Matched AI-inventing firms with similar non-AI firms based on size (employee, capital, R&D) and industry
- **Event Study Design –** Analyzed firm outcomes before and after first AI patent to assess impacts on productivity

AI invention has a positive and significant effect on firm TFP

Table 1: Impact of Cumulative AI Patents on value added

VARIABLES	(1) value_added	(2) value_added	(3) value_added	(4) value_added	(5) value_added	(6) value_added	(7) 333-336	(8) 333-336	(9) 51	(10) 51	(11) 51
log_cum_aipatent	0.150***		0.0574***		0.0562***		0.0665***		0.0800***		
	(0.00967)		(0.00588)		(0.00680)		(0.00920)		(0.0184)		
L.log_cum_aipatent		0.122***		0.0429***		0.0430***		0.0495***		0.0657***	
		(0.00927)		(0.00600)		(0.00691)		(0.00937)		(0.0183)	
L2.log_cum_aipatent											0.0602***
											(0.0176)
log_k_real			0.269***	0.309***	0.179***	0.209***	0.160***	0.191***	0.149***	0.166***	0.154***
			(0.00764)	(0.00886)	(0.0100)	(0.0120)	(0.0155)	(0.0185)	(0.0289)	(0.0349)	(0.0405)
log_emp			0.637***	0.617***	0.632***	0.624***	0.648***	0.634***	0.702***	0.688***	0.693***
			(0.00912)	(0.00970)	(0.0136)	(0.0143)	(0.0204)	(0.0212)	(0.0502)	(0.0548)	(0.0597)
log_xrd_real					0.134***	0.119***	0.141***	0.132***	0.106***	0.105***	0.103***
					(0.00914)	(0.00955)	(0.0150)	(0.0157)	(0.0310)	(0.0332)	(0.0356)
Constant	3.990***	4.065***	2.219***	2.060***	2.241***	2.166***	2.097***	2.010***	2.321***	2.281***	2.352***
	(0.00425)	(0.00416)	(0.0272)	(0.0329)	(0.0363)	(0.0439)	(0.0536)	(0.0638)	(0.121)	(0.142)	(0.165)
Observations	65,996	61,724	63,581	59,399	41,955	39,267	19,300	18,065	5,216	4,833	4,461
R-squared	0.893	0.898	0.930	0.932	0.934	0.936	0.925	0.927	0.922	0.924	0.927

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Sharpening causal inference with an event study, using propensity score matching

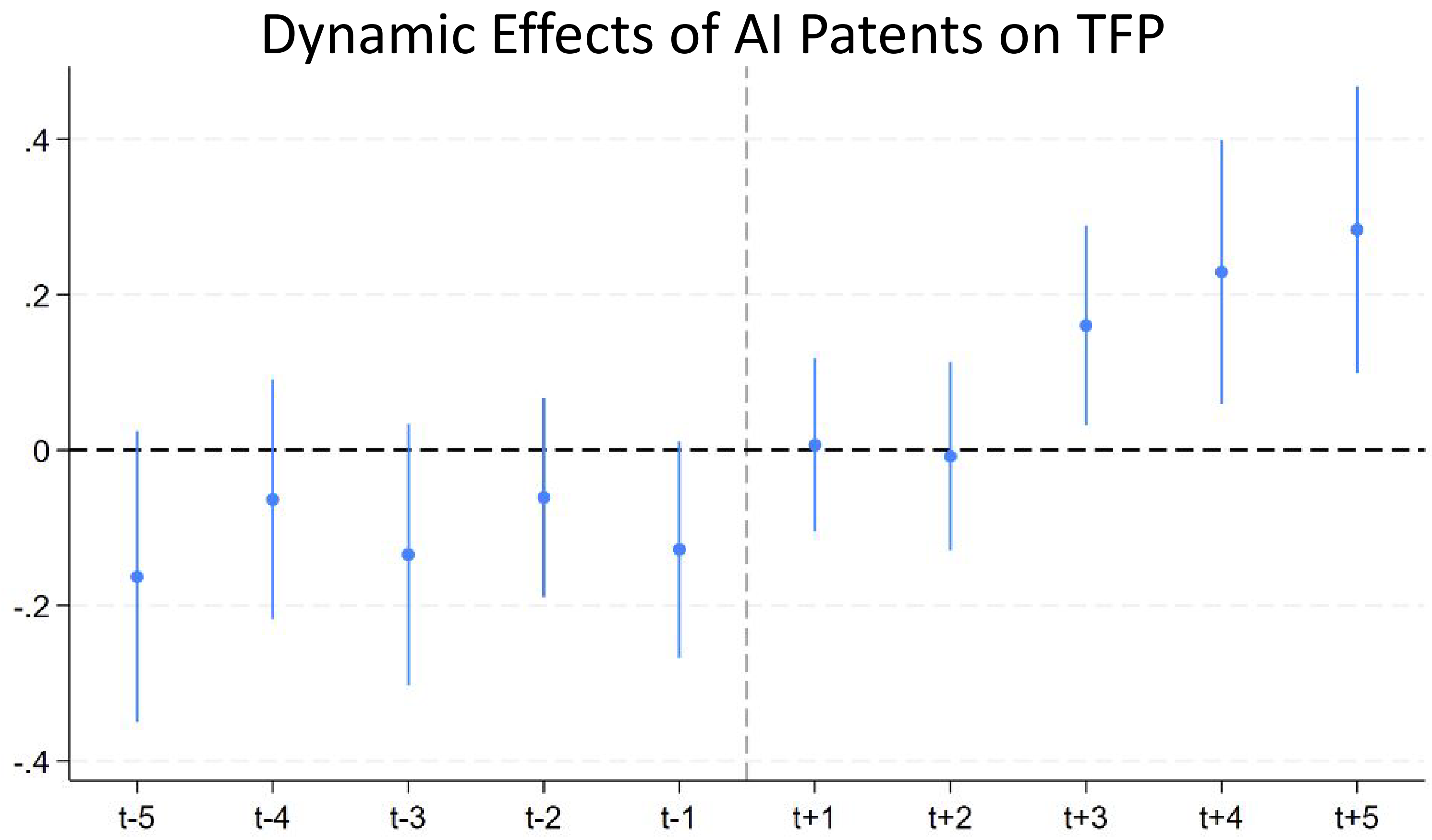
- **We estimate the following regression model over a symmetric 11-year window, with the year of AI patent adoption omitted as the baseline period:**

$$Y_{it} = \sum_{k=-5, k\neq 0}^{+5} \beta_k D_{it}^k + \alpha. \log(K_{it}) + \gamma \log(L_{it}) + \delta \log(XRD_{it}) + \mu_i + \lambda_t + \eta_s + \epsilon_{it}$$

D_{it}^k: Indicator for event year k relative to the AI adoption year
Y_{it}: Firm output (Value added)

- **Propensity Score Matching (PSM)**: Based on pre-treatment firm characteristics such as size (employment, capital stock, R&D expenditure), industry

AI invention raises firm productivity at the extensive margin



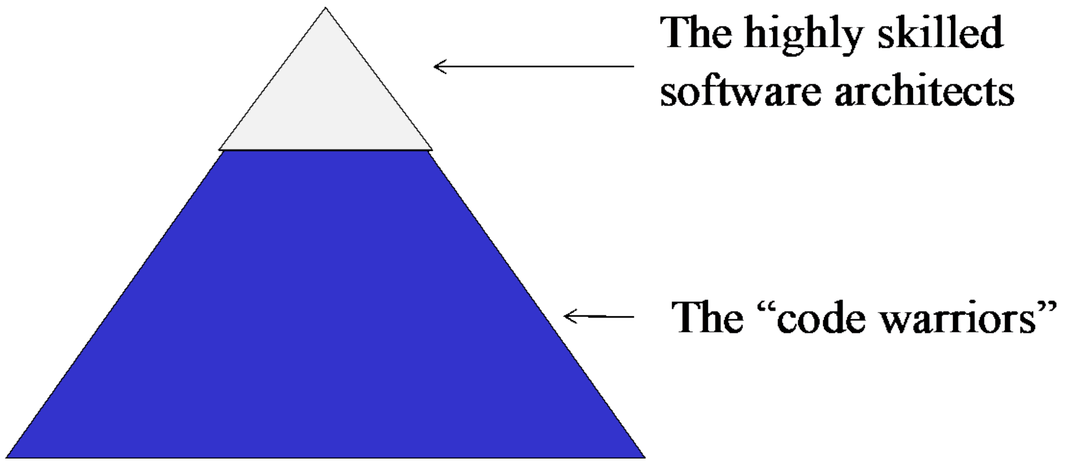
(Preliminary) Conclusions

- AI-based approaches can identify AI-related patents.
- We find strong relationships between AI invention and the subsequent productivity growth of AI-inventing firms!
- These effects are observed at the extensive and intensive margins.
- No apparent impact (yet) on income inequality.
- Japanese firms are also taking out significant numbers of AI-related patents in the U.S.

Future work

- Complete updated analysis of AI-related invention on productivity and income inequality, using Census microdata.
- Examine impact heterogeneity across types of firms and types of AI inventions.
- Extend these analyses to Japanese firms and conduct comparative analysis of the nature and impact of AI invention across the two countries.
- Could we use these techniques to examine the impact of other technologies on the productivity of inventing firms?

Related Work: Measuring AI transfer to industry by tracking the movement of experts



- Any firm seeking to apply frontier AI technology to a particular problem requires a "pyramid" consisting of workers with different levels of AI skills.
- At the moment, the high-level "software architects" who can guide the application of frontier technology are in especially short supply.
- The allocation of these elite software architects across firms, industries, and projects may be especially predictive of success.
- Can one track the high-level architects as they move across firms that employ them?
- We find the leading AI scientists using Elsevier publication data. We use faculty websites and other sources to identify their (doctoral/post-doc) students.
- We are then using Revelio/LinkedIn data and other sources to trace their movements across organizational boundaries over time (many thanks to Sonny Tambe!).
- We can use U.S. Compustat and Census data to test the hypothesis that the emergence of a "critical mass" of frontier AI researchers within a firm leads to increases in output, employment, and productivity.

Thanks!

Tracing the impact generated by AI experts across time and organizational boundaries

Carnegie Mellon University School of Computer Science

↓

Google DeepMind

↓

SIEMENS Healthineers

↓

ARTERYS

Dr. Who

CMU Ph.D.

We can measure the impact of an accumulating stock of AI experts on firm productivity and other outcomes

Event study models

$$y_{it} = \alpha + \beta_1 AI_{it}(1|0) + \beta_2 TIME + \beta_3 AI_{it} \times TIME + X_{it} + \varepsilon$$

Production functions (with fixed effects)

$$q_i - l_i = \alpha k_i + \varphi a_i + \varepsilon_i$$

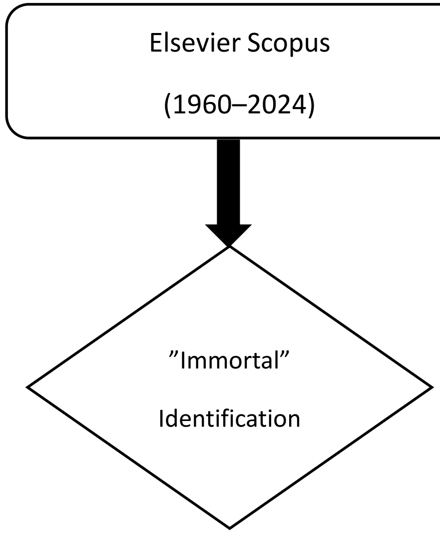
Identifying Elite AI Talent

Elsevier Scopus

(1960–2024)

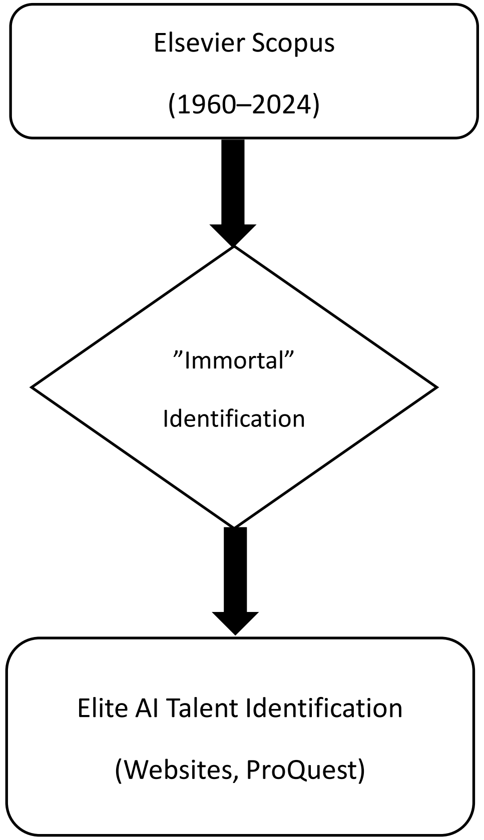
- All scientific publications in **Elsevier database – 1960 to 2024**
- **(10) Sub-domains of AI**: ML, NLP, Robotics, Speech, Agents, Information Retrieval, Computer Vision, Knowledge Representation, Human Computer Interaction, General-AI
- **Top venues (conferences)**: In each sub-domains of AI - **Example. ML** – ICML, NeurIPS, AAAI, ICLR…
- **1200+ top venues (conferences)** covered across 10 sub-domains of AI

Identifying Elite AI Talent



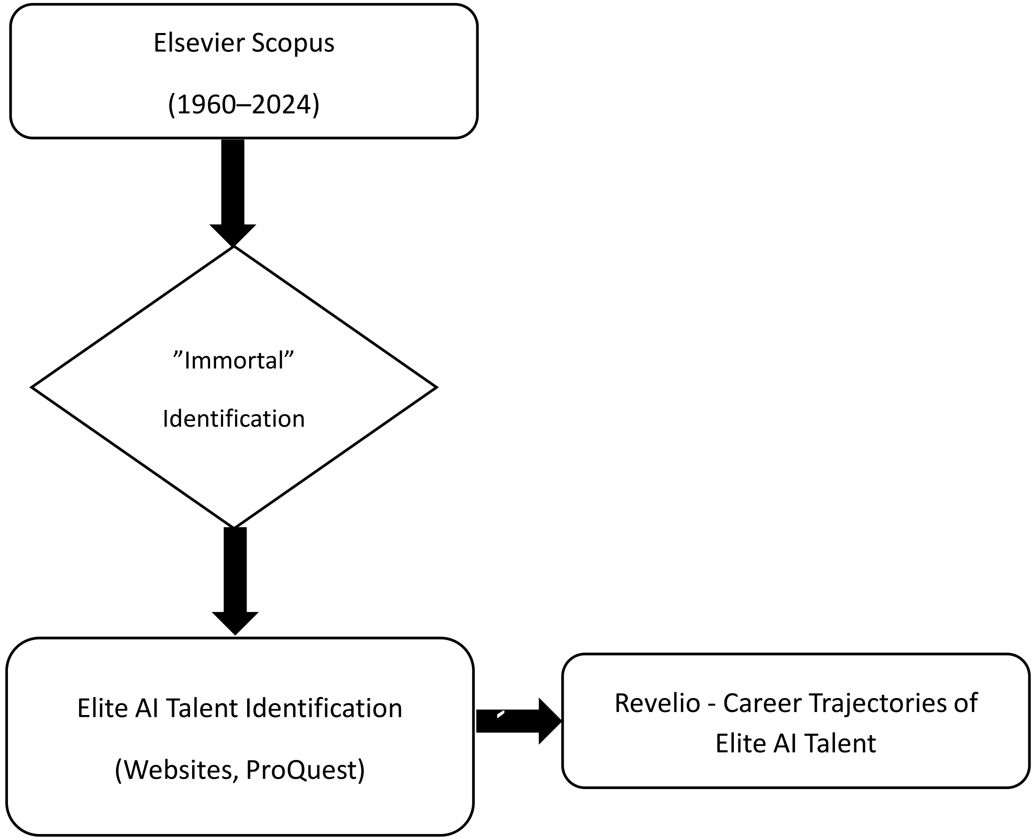
- Calculated our own **H-index** for all authors
- Top **~0.2 percentile** scholars are identified as **“Leading/Immortal” AI scientists** in each sub-domain
- **Identified (~2000) “Immortals”**
- **460,000+ research papers associated with immortals**

Identifying Elite AI Talent



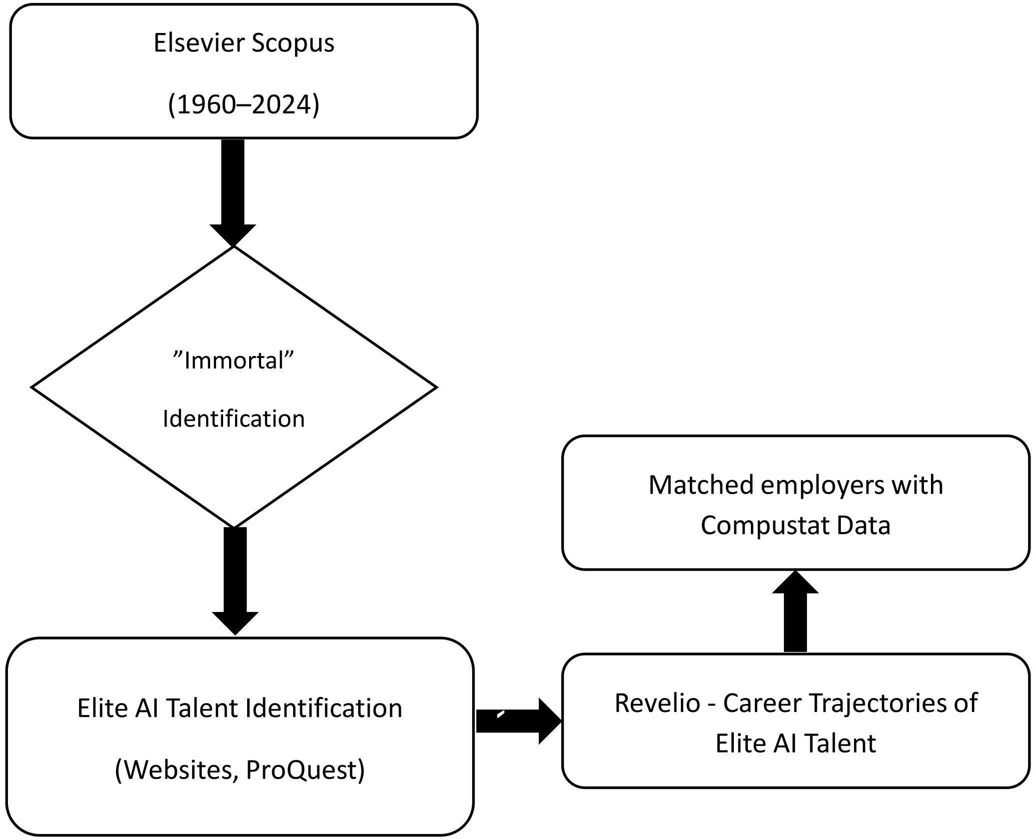
- (**Manually**) Identified academic children **(elite AI talent)** of all immortals
    - Using university and/or personal **websites, CVs**, etc.
    - **ProQuest** dissertation database to supplement the above approach
    - (So far) **32,000+ academic children** of immortals; **~22,000** US based

Identifying Elite AI Talent



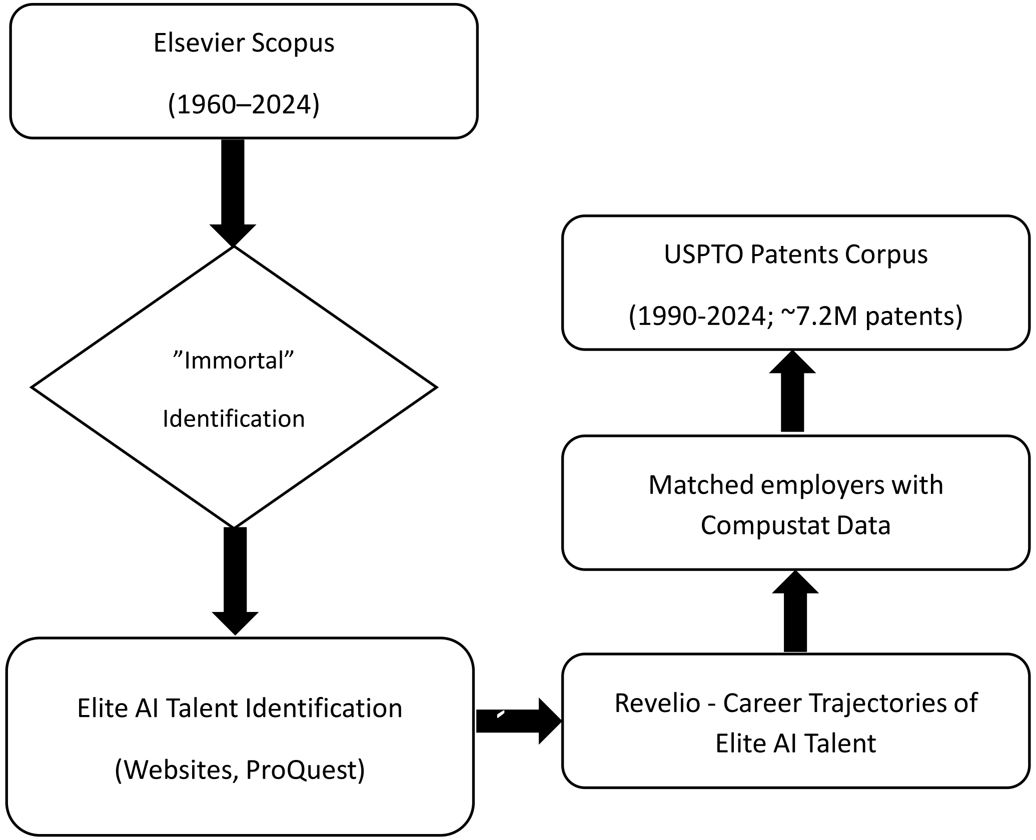
- **Revelio (LinkedIn) data**:
  - **Matched** identified **academic children to Revelio Data**
  - Using **career trajectory data**, curated a **dataset of employers**

Identifying Elite AI Talent



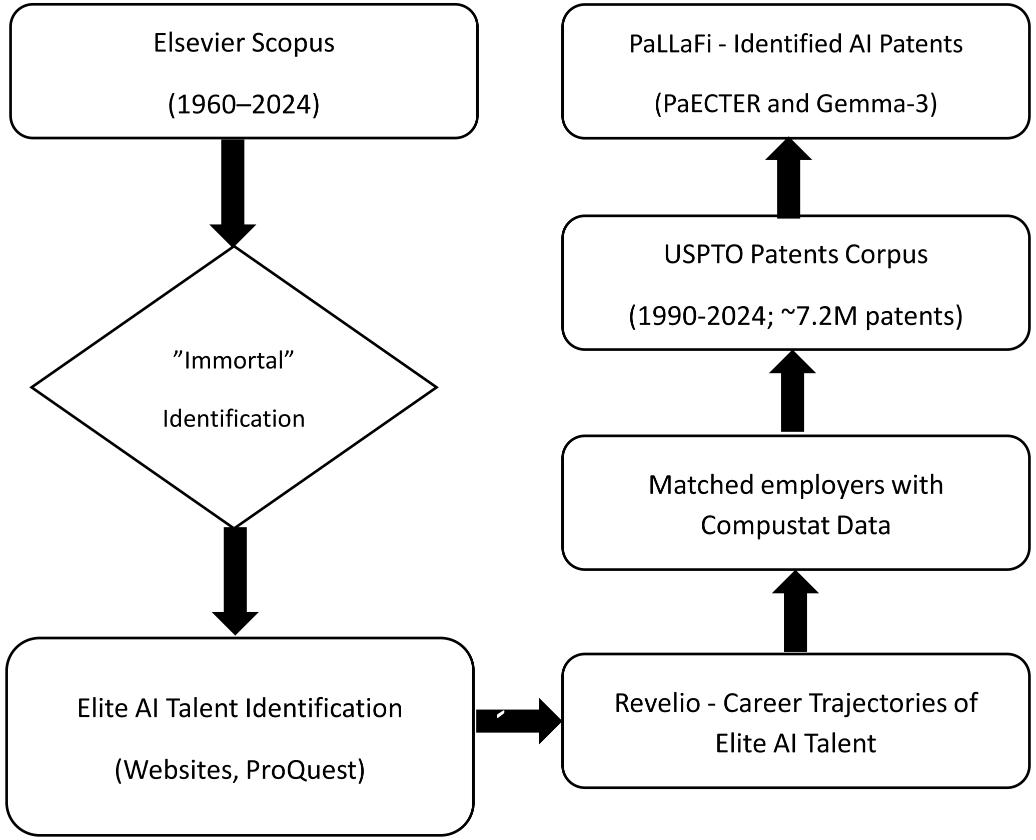
- **Compustat Dataset**:
  - Database of **financial and market data** on **publicly traded companies**
  - **Mapped employers** identified in previous step to **Compustat firms**

Identifying Elite AI Talent



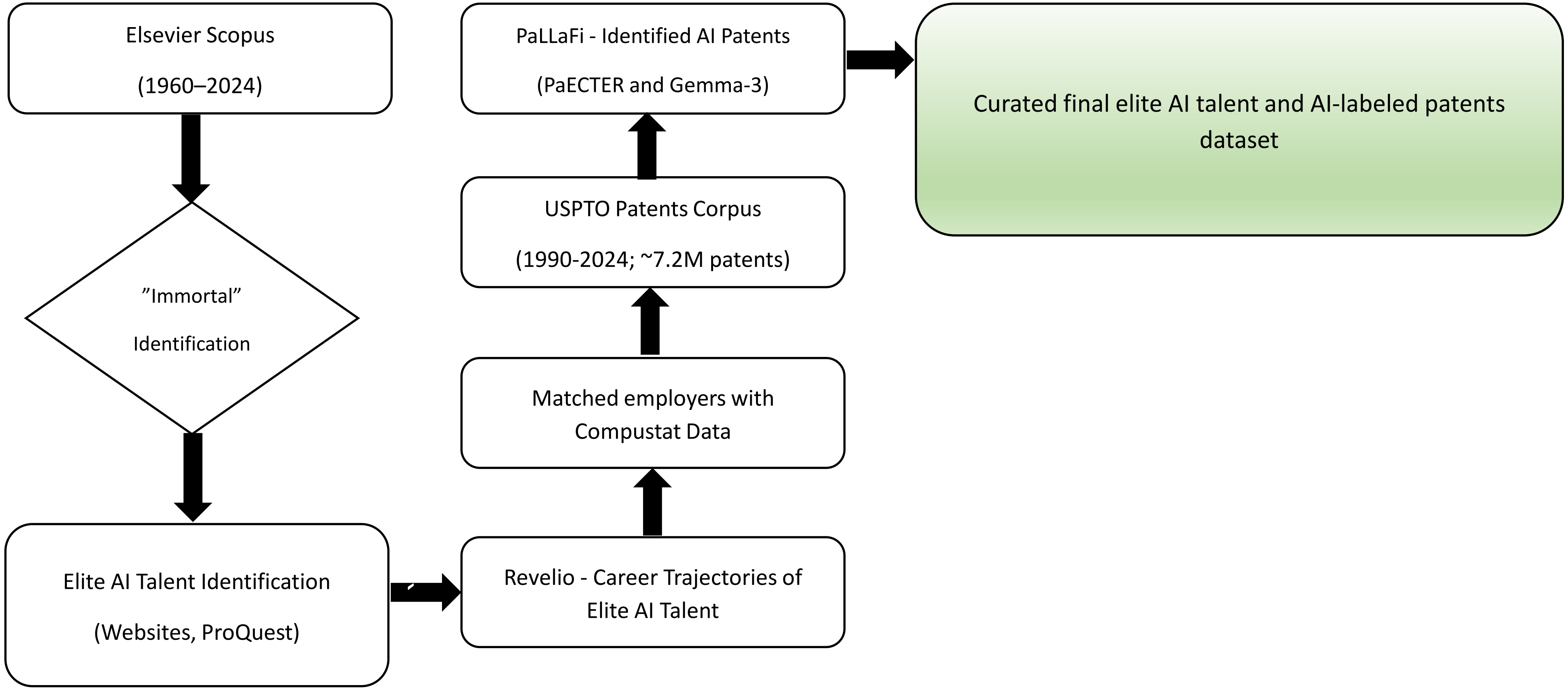
- **USPTO Patent Corpus**:
  - Downloaded all patents granted by USPTO spanning from **1990 to 2024**
  - After cleanup – total # of patents: **7,256,235**

Identifying Elite AI Talent

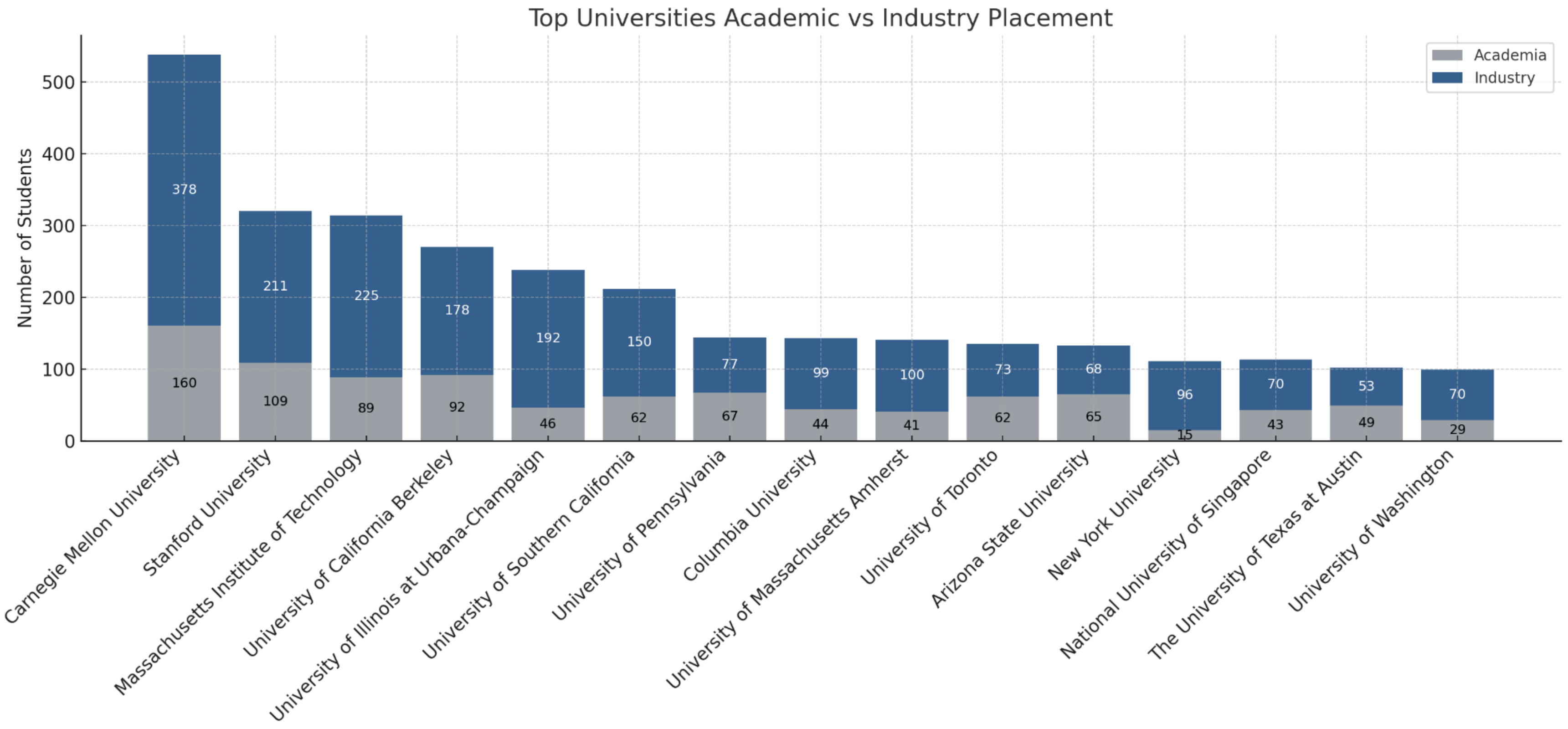


- **PaLLaFi:**
  - Patent Labeling via Language Models and Fine-Tuned Inference
  - **Supervised Machine Learning Models:** BERT, SciBERT, Longformer, Pat-Spectre, and **PaECTER**
  - **Large Language Models (Zero/Few-Shot):** DeepSeek-v1 (1.5B, 8B, 32B), LLaMA 3.1 (1B, 8B), Phi-3 (8B), Granite 3.1 (8B), Gemma-2 (2B, 9B), **Gemma-3** (1B, 4B, 12B, 27B)

Identifying Elite AI Talent

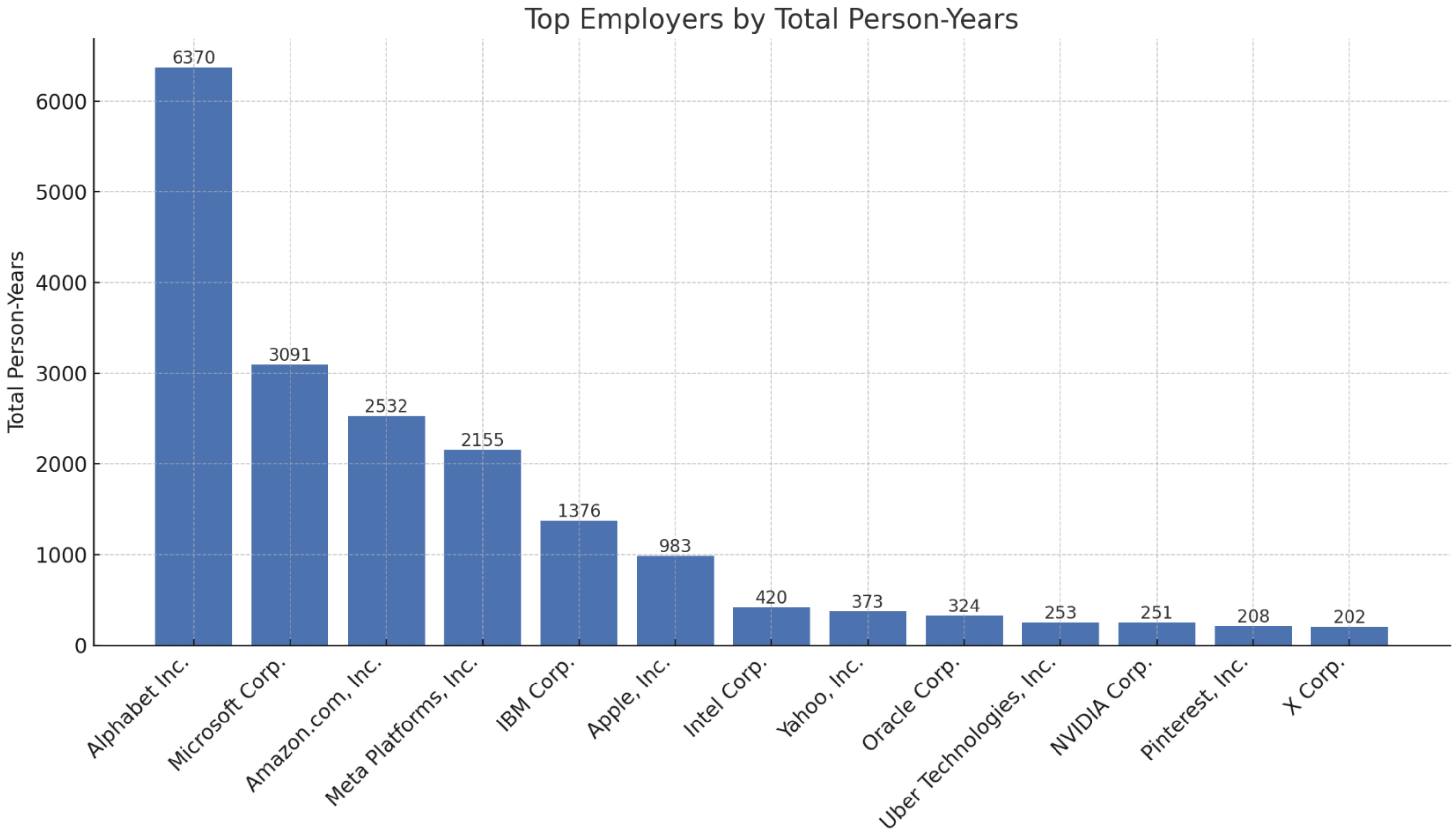


Elite AI Talent: Top Universities – Placement

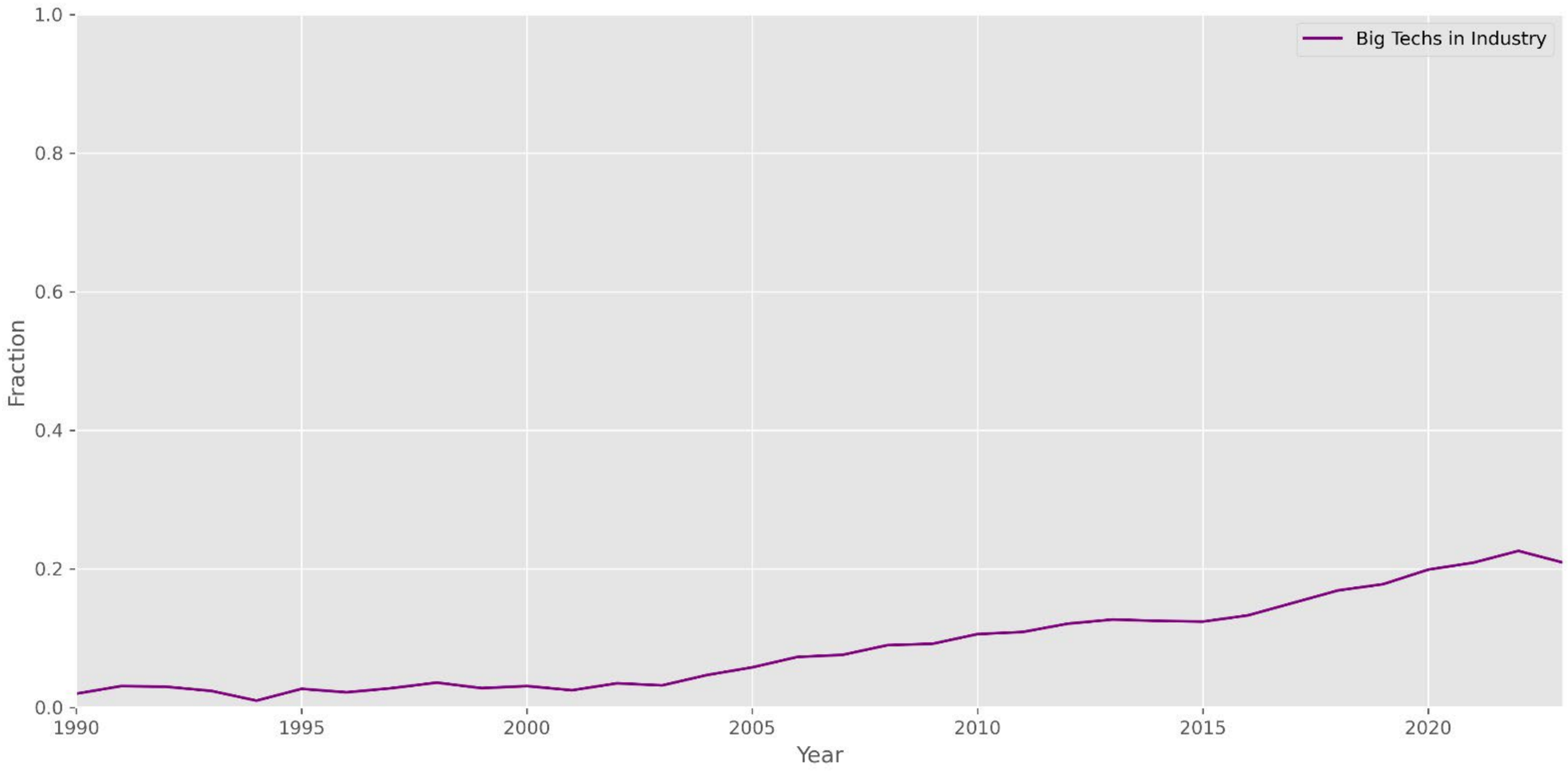


Carnegie Mellon University

Elite AI Talent: Top Employers by Total Person-Years



Elite AI Talent: “Big Tech” firm Employment



Elite AI Talent Increases TFP by 7-13%

VARIABLES	(1) TFP	(2) TFP	(3) TFP (Manufacturing)	(4) TFP (Manufacturing)	(5) TFP (Information)	(6) TFP (Information)
log_elite_talent_stock	0.102***		0.130***		0.0859***	
	(0.0108)		(0.0167)		(0.0294)	
L.log_elite_talent_stock		0.0869***		0.110***		0.0721**
		(0.0111)		(0.0174)		(0.0298)
log_capital	0.181***	0.209***	0.160***	0.191***	0.158***	0.174***
	(0.0100)	(0.0120)	(0.0155)	(0.0185)	(0.0291)	(0.0351)
log_employment	0.631***	0.623***	0.648***	0.634***	0.705***	0.689***
	(0.0136)	(0.0143)	(0.0204)	(0.0212)	(0.0505)	(0.0550)
log_r&d_expenditure	0.132***	0.118***	0.141***	0.131***	0.0967***	0.0979***
	(0.00914)	(0.00954)	(0.0150)	(0.0157)	(0.0311)	(0.0332)
Constant	2.239***	2.164***	2.099***	2.011***	2.336***	2.290***
	(0.0363)	(0.0438)	(0.0534)	(0.0636)	(0.121)	(0.141)
Observations	41,955	39,267	19,300	18,065	5,216	4,833
R-squared	0.934	0.936	0.925	0.927	0.921	0.924

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Elite AI Talent Increases Labor Productivity by 7-12%

VARIABLES	(1) Labor Prod.	(2) Labor Prod.	(3) Labor Prod. (Mfg)	(4) Labor Prod. (Mfg)	(5) Labor Prod. (Info)	(6) Labor Prod. (Info)
log_elite_talent_stock	0.0902***		0.124***		0.0742**	
	(0.0107)		(0.0166)		(0.0290)	
L.log_elite_talent_stock		0.0781***		0.107***		0.0629**
		(0.0111)		(0.0174)		(0.0298)
log_capital_per_employee	0.191***	0.225***	0.170***	0.205***	0.160***	0.180***
	(0.00995)	(0.0116)	(0.0154)	(0.0179)	(0.0291)	(0.0348)
log_r&d_expenditure_per_employee	0.139***	0.122***	0.147***	0.136***	0.102***	0.102***
	(0.00917)	(0.00957)	(0.0150)	(0.0157)	(0.0319)	(0.0341)
Constant	2.146***	2.057***	2.014***	1.918***	2.307***	2.244***
	(0.0335)	(0.0387)	(0.0495)	(0.0561)	(0.124)	(0.144)
Observations	41,955	39,267	19,300	18,065	5,216	4,833
R-squared	0.655	0.665	0.564	0.575	0.653	0.666

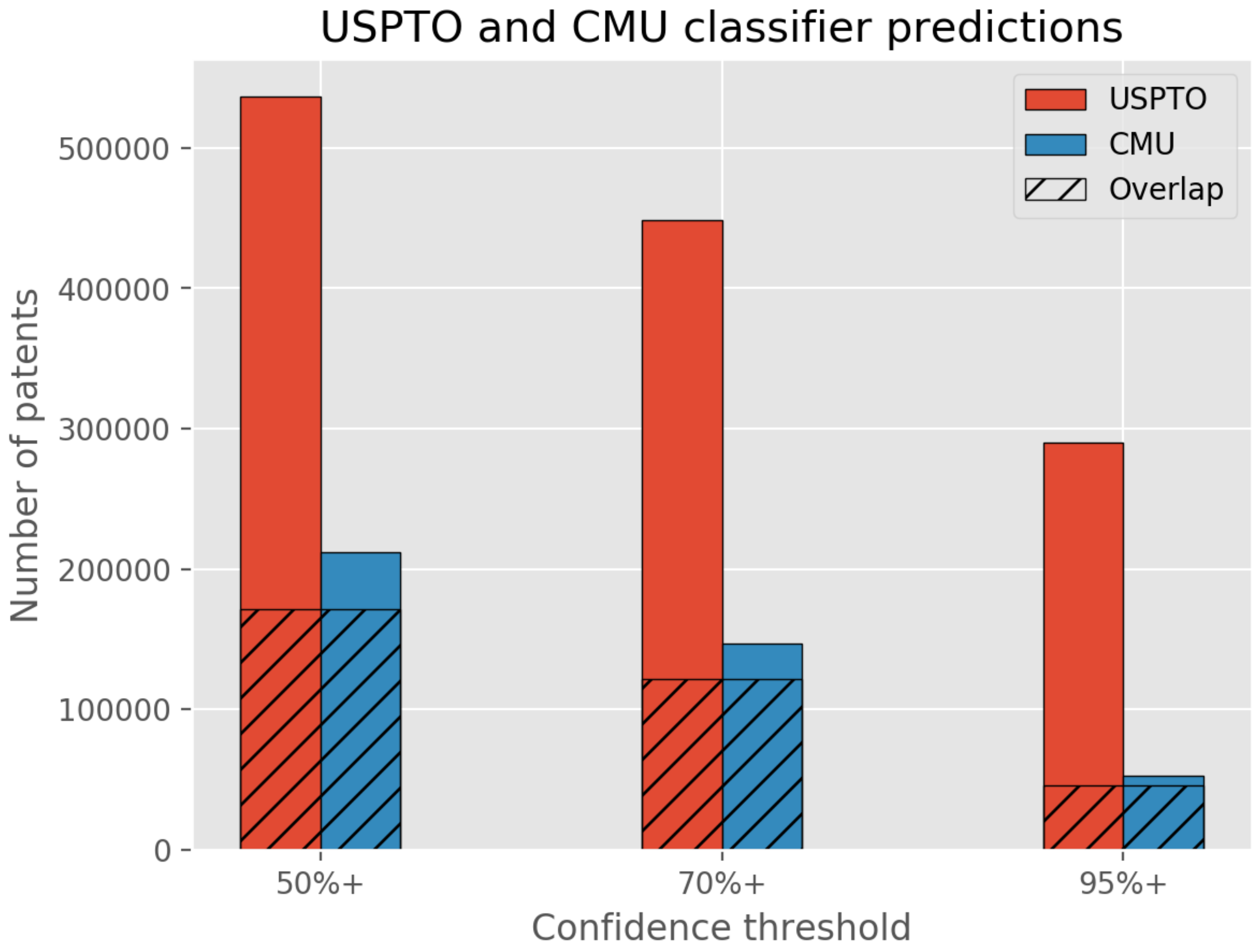
Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Thanks!

Our methods find far more AI patents than previous approaches taken by some economists

- Cockburn et al. (2019) take a “standard approach,” focusing on a relatively small set of key words and patent classes.
- This approach identifies fewer than 14,000 patents between 1990 and 2014, and it includes large numbers of ”robotics hardware” patents.
- Webb et al. (2019) take a similar, more focused approach, identifying 2,000+ patents related to ”machine learning” and 4,000+ related to “neural networks.”
- Our approach identifies 52,896 patents that are AI related with 95% confidence and 146,952 patents that are AI related with 70% confidence (through 2018).
- We identify most of the AI patents tagged by other economists as ”AI patents” but also capture a very large number that traditional techniques omit.
- However, our methods find far *fewer* patents than do recent efforts by the USPTO to apply machine learning methods to patent data.

But our methods find far fewer AI patents than do comparable efforts by the USPTO



Next steps

- Continue to update AI patents and the analysis of AI invention on firm productivity
- Obtain new data on the movement of the students of elite AI scientist into U.S. firms
- Measure direct collaboration between elite AI scientists and firms
- Examine the impact of this elite human capital (if any) on U.S. firm output, employment, and productivity
- Track the movement of elite human capital outside of the U.S.