AI Economics / 人工智能经济学

徐兴成

2024年2月

OUTLINE

1 Machine Learning and LLM AI

2 AI for Economics

3 Economics of AI

引言：AI Economics

Why AI Economics

一时代之学术，必有其新材料与新问题。取用此材料，以研求问题，则为此时代学术之新潮流。治学之士，得预于此潮流者，谓之预流。

- 陈寅恪：《陈垣敦煌劫余录序》

AI Economics-双层涵义

- “新材料”——新材料“大数据”需要新工具“人工智能”！
- AI for Economics，AI赋能的经济学，甚至整个社会科学！
- “新问题”——在数字经济时代，人工智能将如何改变经济与社会?
- Economics of AI，问题导向，AI与经济社会的关系！

The Jobs AI Can Do: AI for Economics

- Quantification of Unstructured Data
- New Data and New Measurement
- Prediction/Forecasting
- Causal Inference
- Computation
- Simulation
- Hypothesis Generation

The Jobs AI Can Do: AI for Finance

- Return Prediction
- Risk-Return Tradeoff
- Optimal Portfolio
- LLM as an expert
- Answering the queen: ML and Financial Crises Prediction

The Economics of AI

- AI and Economic Growth

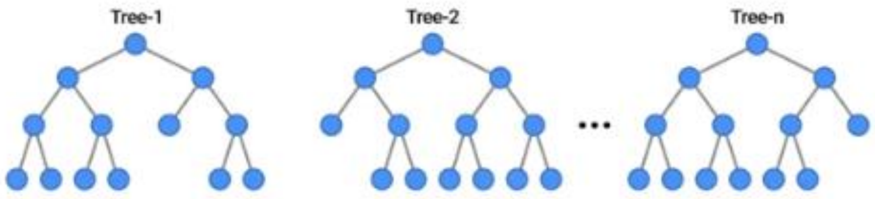
- AI and Innovation - AI for Science

- Labor Market

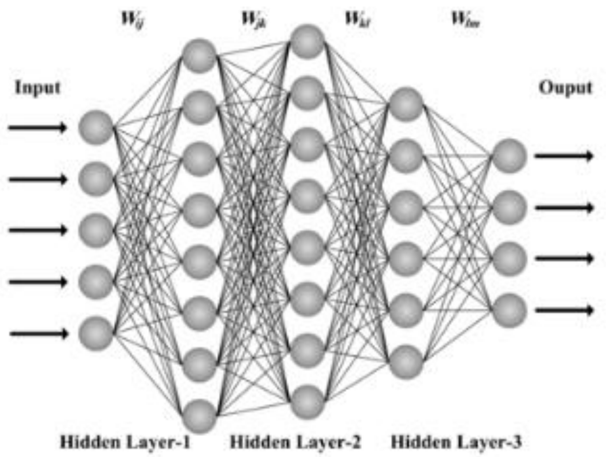
01

Machine Learning and LLM AI

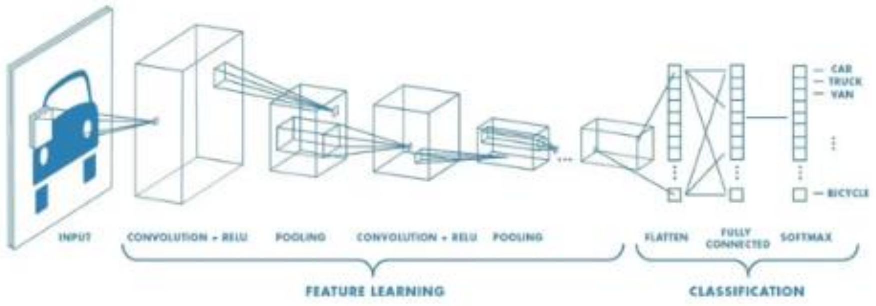
Machine Learning and Deep Learning



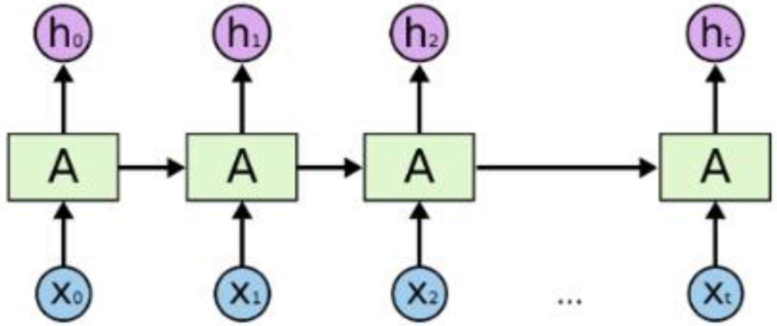
RF



MLP



CNN



RNN

Transformer

Attention Is All You Need

Adept · character.ai · Adept · Inceptive

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

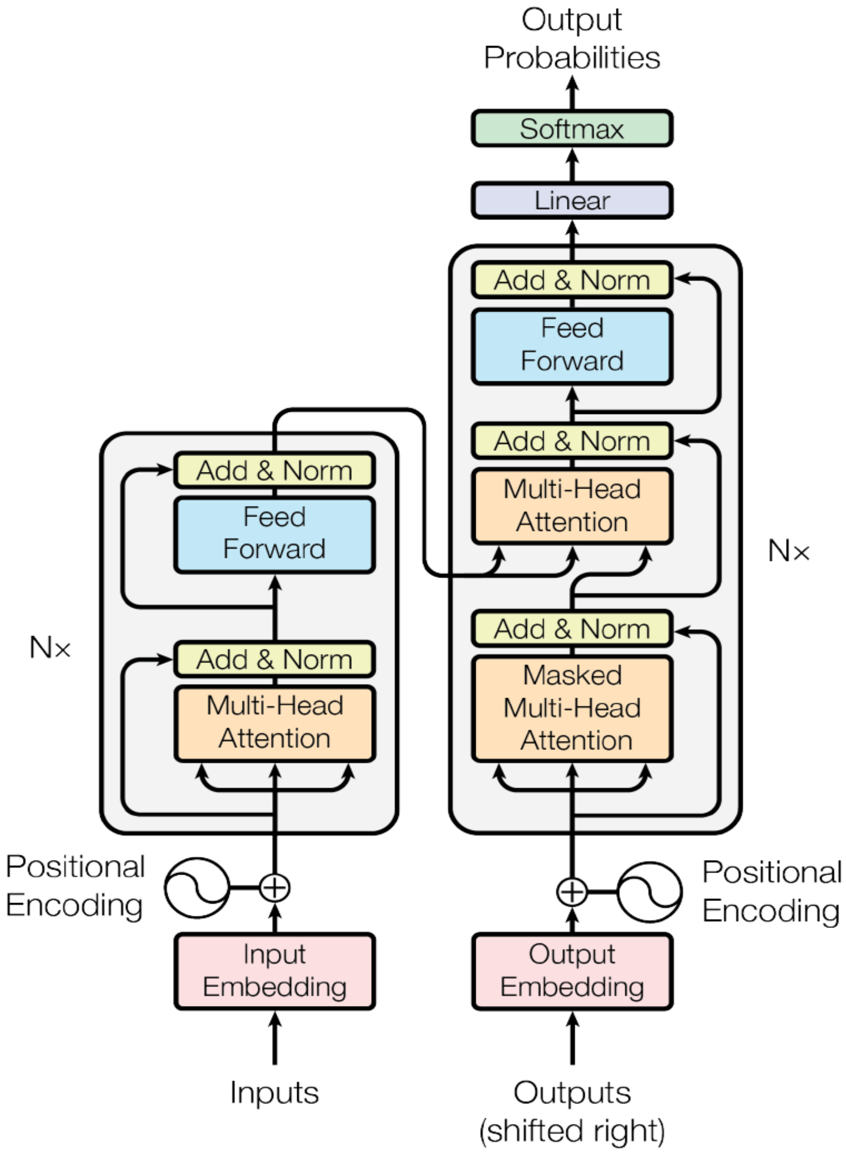
Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu → co:here

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com → NEAR INCORPORATED

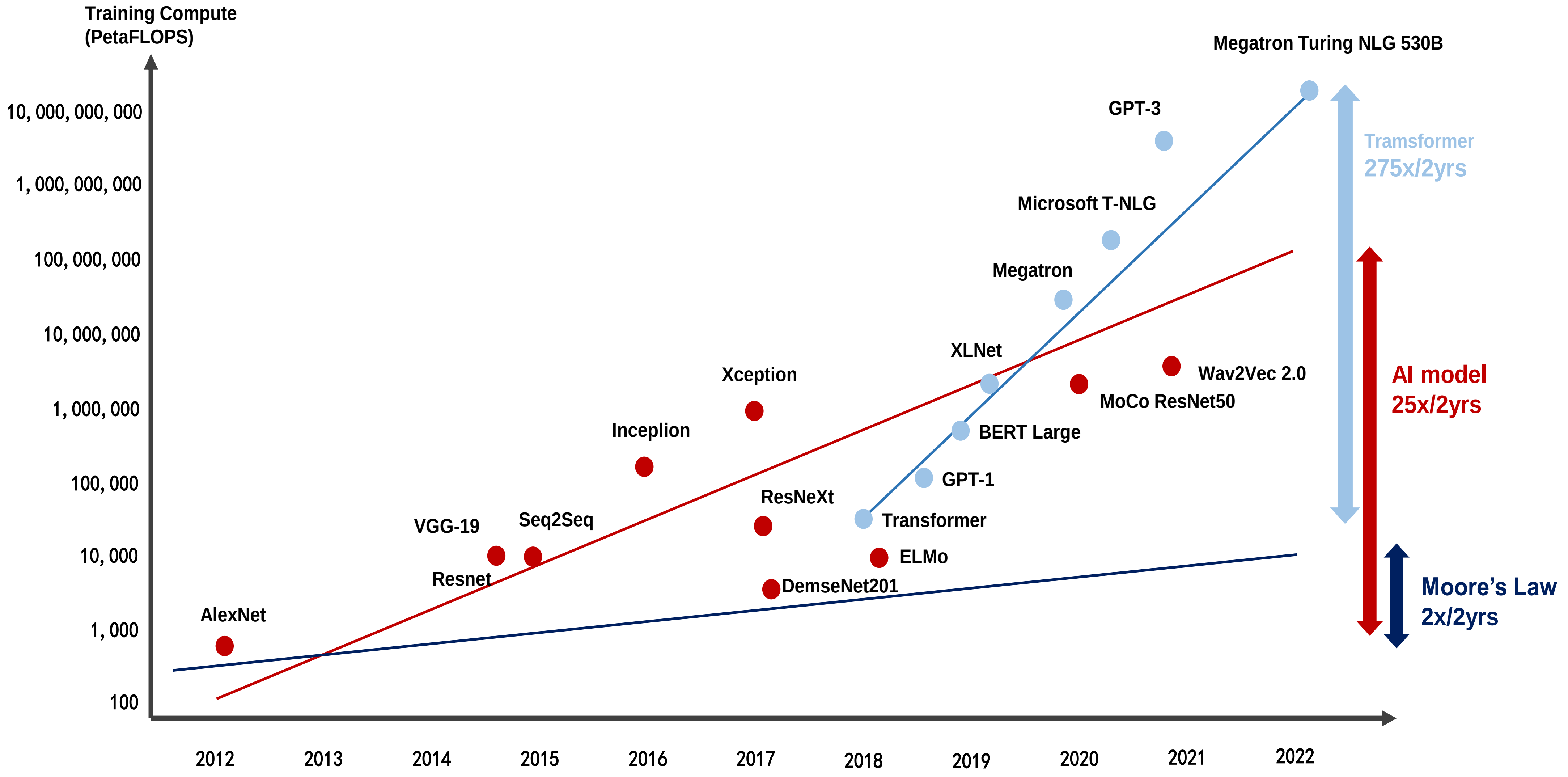


Vaswani, et al (2017). "Attention is all you need." *NeurIPS*.
(被引用次数：118159)

AI 2.0：基础模型+ 生成式AI

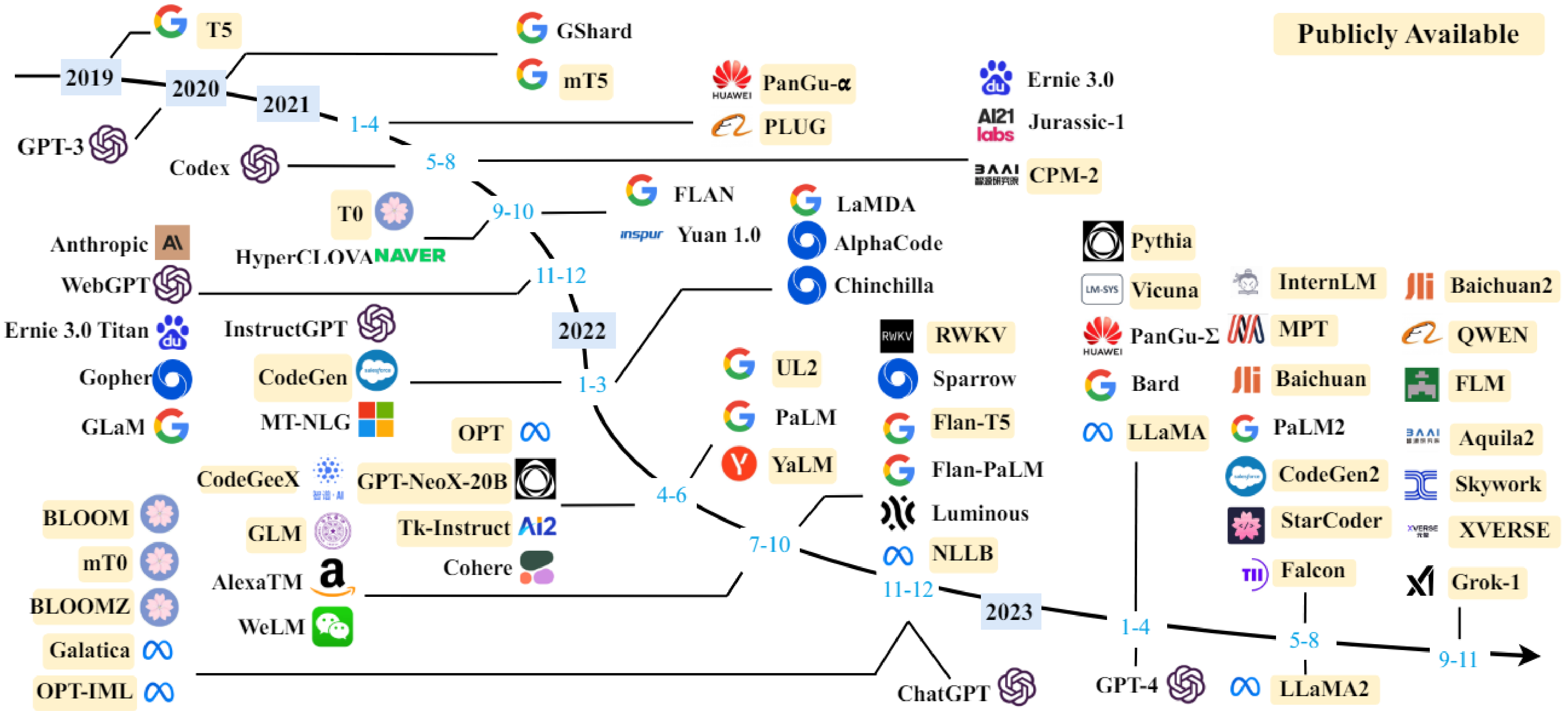
1. **基础模型：Transformer作为基础架构，ChatGPT/GPT4［技术膨胀期］**
2. **生成式人工智能：LLM/Diffusion Models突飞猛进。［技术膨胀期］**
3. 以数据为中心的人工智能：从模型为中心到数据为中心。[技术萌芽期]
4. 因果人工智能：因果关系与人工智能的结合。[技术萌芽期]
5. 复合型人工智能：“连接主义”与“符号主义”相结合。[技术萌芽期]
6. NeuroAI：神经科学与人工智能交叉研究。[技术萌芽期]

算力驱动：远超摩尔定律



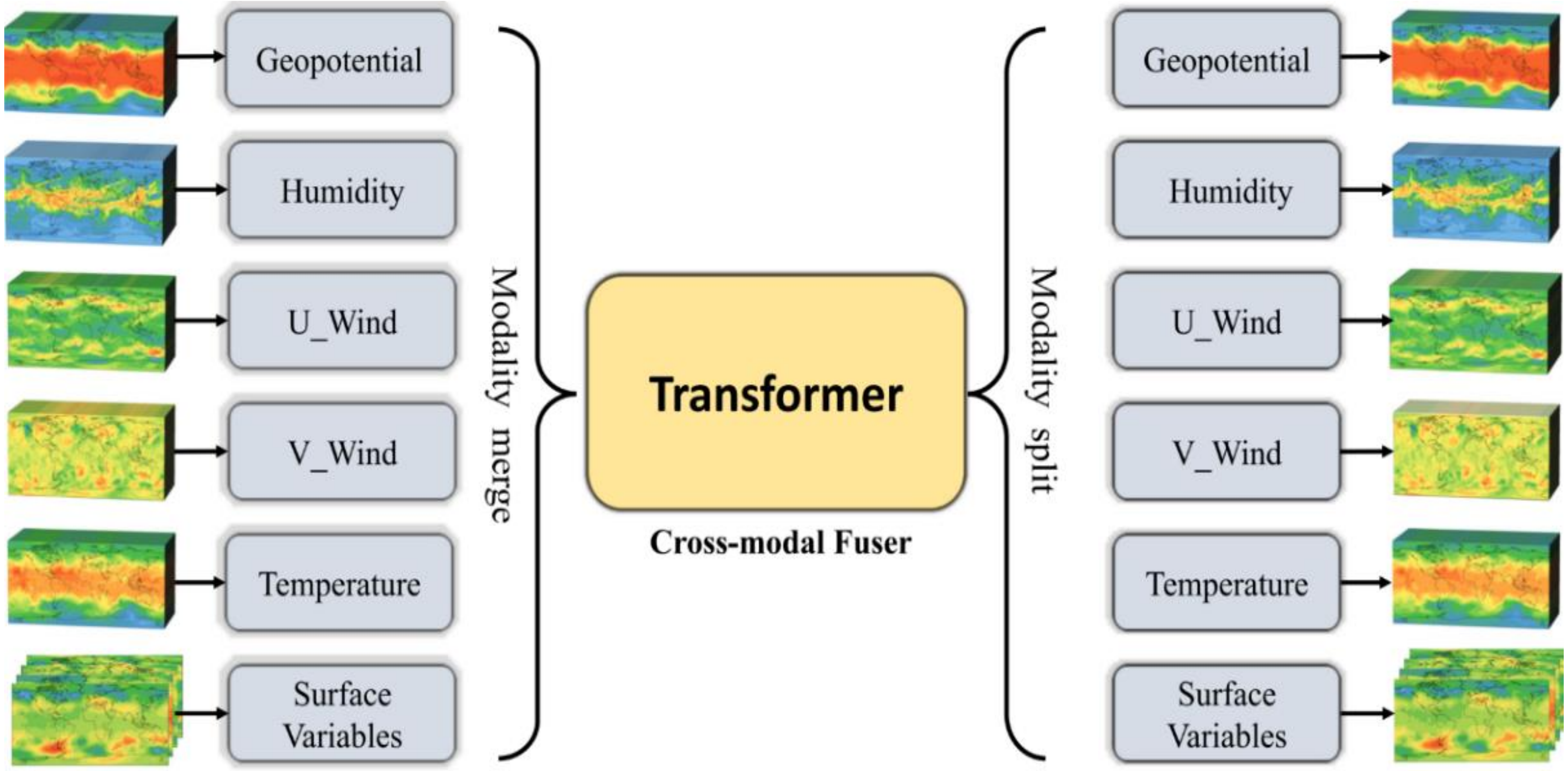
Source: NVIDA, 2023

LLMs



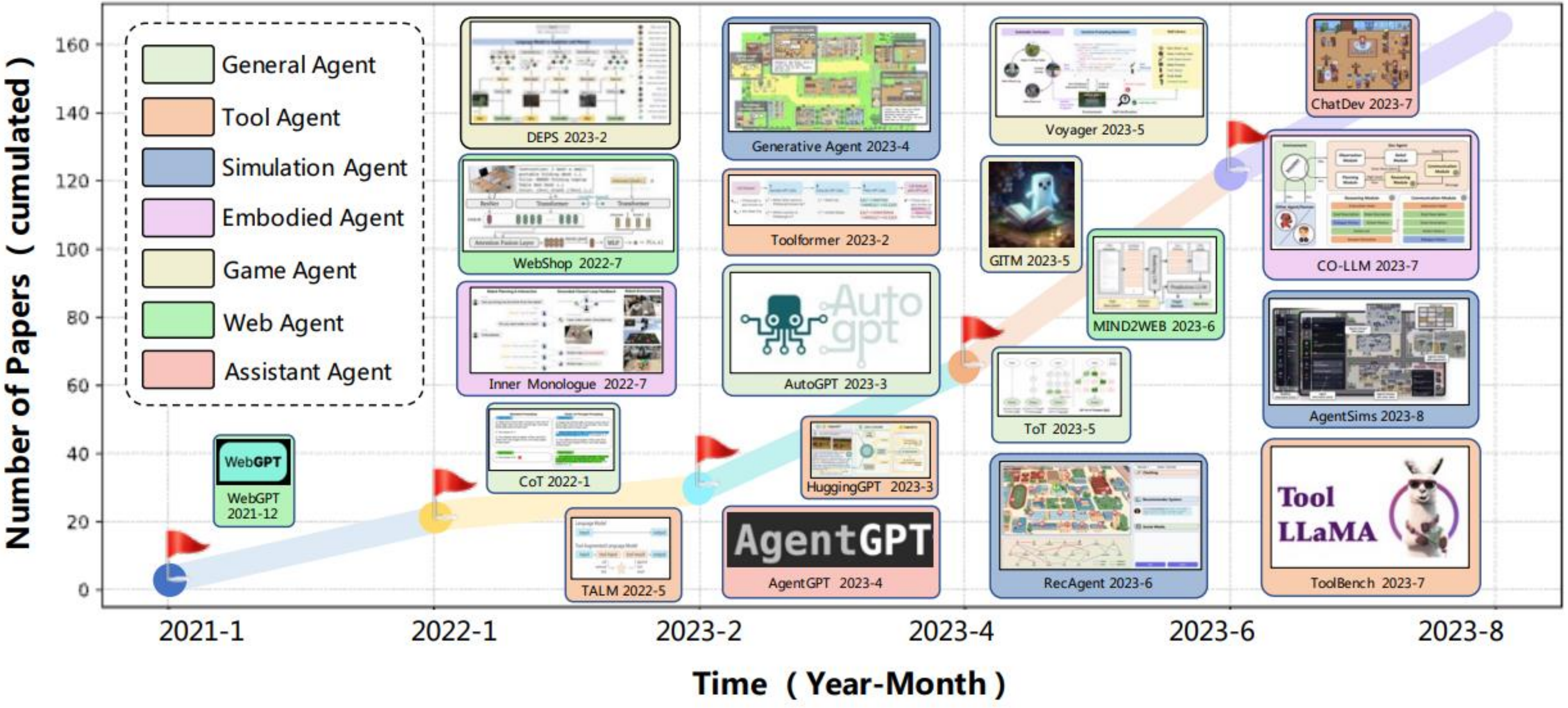
Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

AI4Science: FengWu



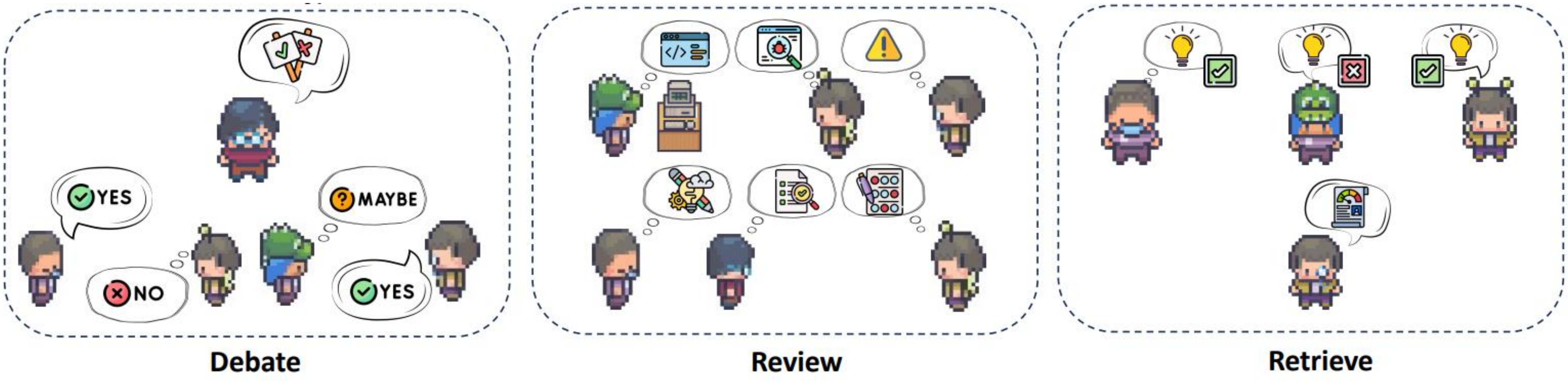
Chen et al. (2023). FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead.

AI Agents



General Agent
Tool Agent
Simulation Agent
Embodied Agent
Game Agent
Web Agent
Assistant Agent
Number of Papers (cumulated)
160
140
120
100
80
60
40
20
0
WebGPT 2021-12
DEPS 2023-2
WebShop 2022-7
Inner Monologue 2022-7
CoT 2022-1
TALM 2022-5
Generative Agent 2023-4
Toolformer 2023-2
AutoGPT 2023-3
HuggingGPT 2023-3
AgentGPT 2023-4
Voyager 2023-5
GITM 2023-5
MIND2WEB 2023-6
ToT 2023-5
RecAgent 2023-6
ChatDev 2023-7
CO-LLM 2023-7
AgentSims 2023-8
ToolBench 2023-7
Tool LLaMA
2021-1
2022-1
2023-2
2023-4
2023-6
2023-8
Time (Year-Month)

AI Agents: Corex



An intuitive illustration of Corex, employs LLMs as agents to collaboratively solve a problem. The strategies encompass the Debate, Review, and Retrieve modes, leveraging both the reasoning process and code synthesis. This framework facilitates interactions between models that foster a collaborative environment for the derivation of a well-reasoned answer.

Sun et al. (2023). Corex: Pushing the Boundaries of Complex Reasoning through Multi-Model Collaboration.

Safety and Alignment

The A.I. Dilemma: Growth versus Existential Risk
Charles I. Jones
NBER Working Paper No. 31837
November 2023
JEL No. J17,O40

ABSTRACT

Advances in artificial intelligence (A.I.) are a double-edged sword. On the one hand, they may increase economic growth as A.I. augments our ability to innovate. On the other hand, many experts worry that these advances entail existential risk: creating a superintelligence misaligned with human values could lead to catastrophic outcomes, even possibly human extinction. This paper considers the optimal use of A.I. technology in the presence of these opportunities and risks. Under what conditions should we continue the rapid progress of A.I. and under what conditions should we stop?

02

AI for Economics

AI for Economics

- 预测 Prediction/Forecasting
- 计算 Economic Computation
- 度量 New Data and New Measurement
- 非结构化数据分析 Quantification of Unstructured Data
- 模拟 Economic Simulation
- 假设 Hypothesis Generation
- 因果 Causal Inference

AI for Economic Forecasting

Macroeconomic Forecasting using AI/ML

I. 计量模型
   1. 自回归模型
   2. 因子模型
II. 机器学习模型
   1. 正则化回归模型（LASSO/Ridge）
   2. 核回归方法
   3. 随机森林
   4. XGBoost、GBDT
III. 交叉耦合模型
IV. 衍生组合模型

Yanqing Yang, Xingcheng Xu, Jinfeng Ge, Yan Xu. “Machine Learning for Economic Forecasting: An Application to China’s GDP Growth.” arXiv preprint arXiv:2407.03595, 2024.

表 2 模型列表

#	模型	核心方法	模型分类
1	AR	自回归	计量模型（基准）
2	FM-AR-SE	因子模型	计量模型
3	XGB-GBTREE	XGBoost	机器学习模型
4	XGB-GBLINEAR	XGBoost	机器学习模型
5	GBDT-AE	梯度提升树	机器学习模型
6	GBDT-HUBER	梯度提升树	机器学习模型
7	GBDT-SE	梯度提升树	机器学习模型
8	RF-AE	随机森林	机器学习模型
9	RF-SE	随机森林	机器学习模型
10	FM-XGB-GBLINEAR	因子模型+XGBoost	交叉耦合模型
11	FM-XGB-GBTREE	因子模型+XGBoost	交叉耦合模型
12	FM-GBDT-AE	因子模型+梯度提升树	交叉耦合模型
13	FM-GBDT-HUBER	因子模型+梯度提升树	交叉耦合模型
14	FM-GBDT-SE	因子模型+梯度提升树	交叉耦合模型
15	FM-RF-AE	因子模型+随机森林	交叉耦合模型
16	FM-RF-SE	因子模型+随机森林	交叉耦合模型
17	FM-KRR-POLY	因子模型+核岭回归	交叉耦合模型
18	FM-KRR-RBF	因子模型+核岭回归	交叉耦合模型
19	FM-LASSO	因子模型+Lasso 回归	交叉耦合模型
20	机器学习模型	中位值	衍生组合模型
21	机器学习模型	平均值	衍生组合模型
22	交叉耦合模型	中位值	衍生组合模型
23	交叉耦合模型	平均值	衍生组合模型
24	机器学习相关所有模型（机器学习+耦合）	中位值	衍生组合模型
25	机器学习相关所有模型（机器学习+耦合）	平均值	衍生组合模型

Macroeconomic Forecasting using AI/ML

真实值基本落在预测值的最大与最小值之间。尤其是经济平稳阶段，预测的中位数值趋近于真实值。在经济波动较大时期，真实值与中位值的差距相对于经济平稳时期有所增大，且模型预测范围比经济平稳时期更宽，真实值更偏向预测的极端值附近。

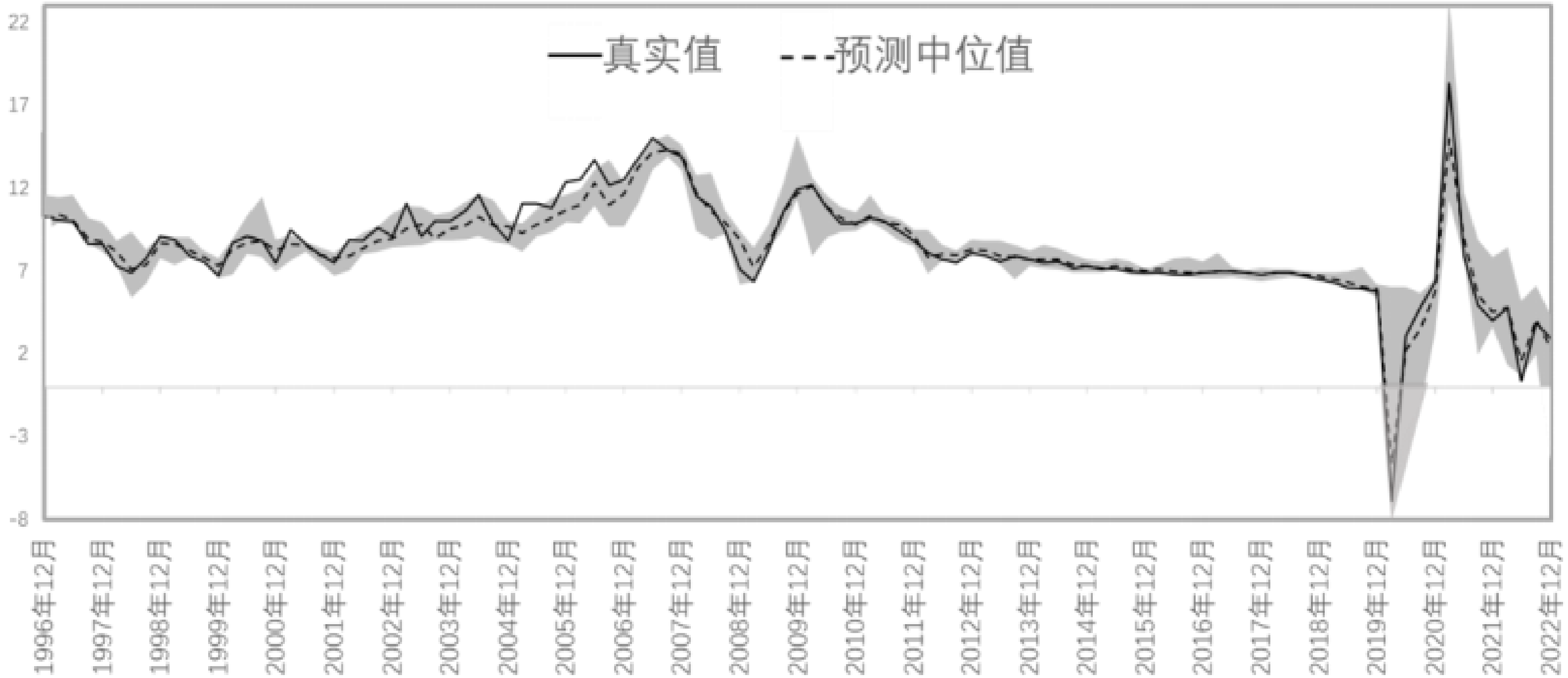


图 2　全国季度 GDP 增速、机器学习模型预测范围和中位值

注：真实值为实线，所有机器学习相关模型（包括机器学习模型和交叉耦合模型）的预测范围为灰色区域，中位值为虚线。

预测结果比较

机器学习模型预测精确度优于“朗润专家预测”和计量模型预测。（2005Q3-2015Q4期间）

具有非线性特征的机器学习模型能够有效的提高预测精确性。这在梁方等（2021）计量模型结果的比较中也得到印证：非线性计量模型的误差小于线性计量模型的误差。

表 3　单个模型预测结果比较（2005Q3-2015Q4）

本文单个模型 RMSE			梁方等(2021) RMSE		
机器学习模型	RF-AE	0.5845	朗润专家预测	第一名	0.6028
	XGB-GBTREE	0.5869		第二名	0.6685
	RF-SE	0.5869		第三名	0.6807
			计量月度模型	MIDAS	0.8107
				ECM-MIDAS	0.8151
交叉耦合模型	FM-RF-SE	0.7146		FM-VAR	0.8771
	FM-GBDT-AE	0.7262	计量非线性	MS-VAR	0.8430
	FM-GBDT-HUBER	0.7434		TVTP	0.8561
				MS	0.8648
			计量线性模型	ARMIA(1,1,1)	0.8705
本文计量基准模型	AR	1.1436		ARMIA(1,1,2)	0.8940
	FM-AR-SE	1.1932		ARMIA(2,1,1)	0.9116

表 4　模型组合预测结果比较（2005Q3-2015Q4）

本文组合模型 RMSE			梁方等(2021) 组合模型 RMSE		
机器学习模型	平均值	0.5702	朗润专家预测	平均值	0.6528
	中位值	0.5197		中位值	>0.6528
交叉耦合模型	平均值	0.7599			
	中位值	0.7745	计量模型	平均值	>0.8701
机器学习相关所有模型	平均值	0.6464		中位值	0.8701
	中位值	0.6282			

预测结果比较

在经济较为平稳的2014-2019年，机器学习相关模型的预测表现优于“一财专家预测”。然而，在2020年1季度至2022年4季度新冠疫情时期，“一财专家预测”表现优于机器学习相关模型。由于新冠疫情这类事件的不可预料性，及其经济冲击性质的独特性，在预测中加入其他实时信息尤为重要。且信息的使用中需善于运用经济学知识，理解经济波动背后的核心推动因素。这时，专家预测相对于机器学习模型具有较大优势。

表 5 衍生组合模型预测结果比较（2014-2022 年）

衍生组合模型		2014Q1-2022Q4		2014Q1-2019Q4		2020Q1-2022Q4	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
机器学习模型	平均值	1.19	0.65	0.22	0.18	2.04	1.60
	中位值	1.29	0.57	0.19	0.15	2.21	1.43
交叉耦合模型	平均值	1.11	0.63	0.15	0.11	1.91	1.65
	中位值	0.89	0.49	0.13	0.09	1.53	1.29
机器学习相关所有模型	平均值	1.05	0.56	0.15	0.12	1.81	1.45
	中位值	0.90	0.48	0.14	0.11	1.55	1.21
一财专家预测	平均值	0.33	0.26	0.17	0.15	0.51	0.45
	中位值	0.30	0.23	0.17	0.15	0.48	0.39

注：一财专家预测缺少 2015 年 2 季度、2015-2022 年 4 季度数据，所以在比较中，相应的时期也在其他模型中被剔除。

经济变量重要性与可解释性分析

全局角度：Shapley值反映了某个变量对全样本最终预测值的边际影响。通过使用Shapley值，可以从全局角度评估了各变量的重要性，进而解释模型预测和性能的驱动因素。

对中国GDP增长预测值最为重要的五个变量是：工业增加值、社会消费品零售总额、财新服务业PMI、钢材产量和PMI。

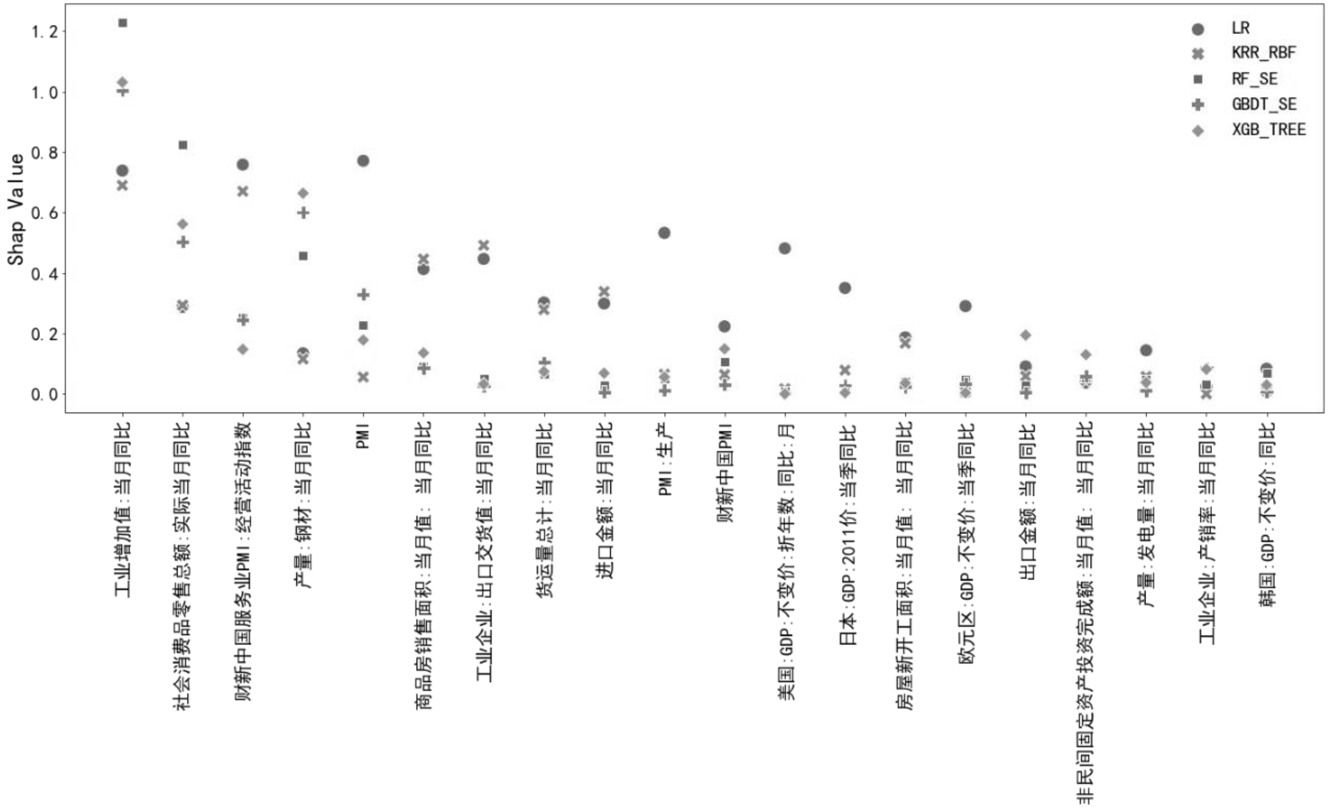


图 6 各变量的 Shapley 值占比（2006-2022 年）

经济变量函数形式

每张图描述了局部Shapley值（纵轴）与观察到的变量输入值（横轴），展现了不同模型学习的各经济变量对预测的函数形式。

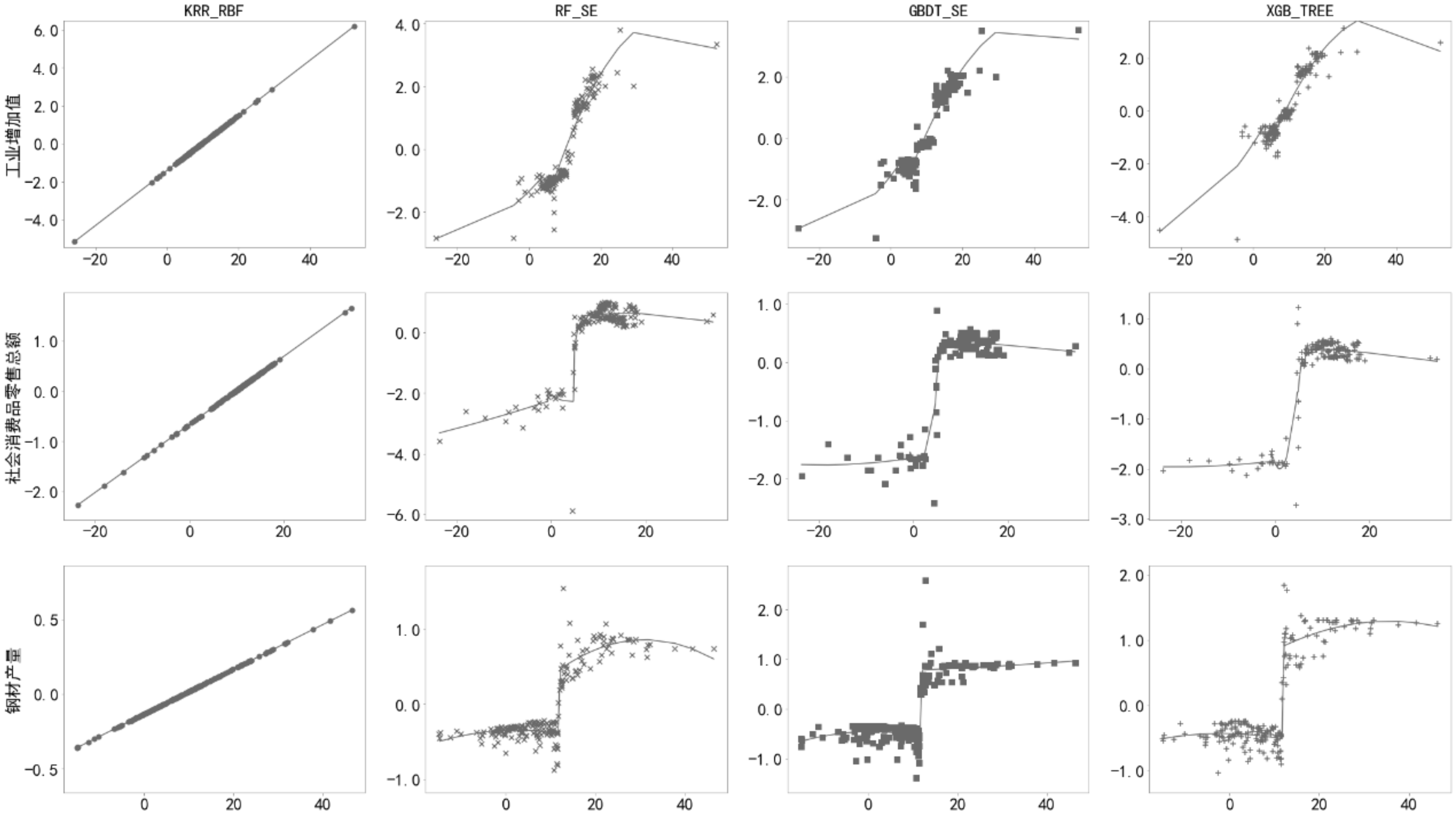


图 9 各模型的函数形式比较（2006-2022 年）

Machine Learning and Financial Crises Prediction

Forecasting the financial crisis of 2008

Visiting the LSE and being shown how terrible the situation was and had been, the Queen asked: "Why did nobody notice it?"

Machine Learning and Financial Crises Prediction

Answering the Queen: Machine Learning and Financial Crises
Jeremy Fouliard, Michael Howell, Hélène Rey, and Vania Stavrakeva
NBER Working Paper No. 28302
December 2020, Revised October 2023
JEL No. G01,G15

ABSTRACT

Financial crises cause economic, social and political havoc. Macroprudential policies are gaining traction but are still severely under-researched compared to monetary and fiscal policy. We use the general framework of sequential predictions, also called online machine learning, to forecast crises out-of-sample. Our methodology is based on model aggregation and is "meta-statistical", since we can incorporate any predictive model of crises in our analysis and test its ability to add information, without making any assumption on the data generating process. We predict systemic financial crises twelve quarters ahead out-of-sample with high signal-to-noise ratio. Our approach guarantees that picking certain time dependent sets of weights will be asymptotically similar for out-of-sample forecasts to the best ex post combination of models; it also guarantees that we outperform any individual forecasting model asymptotically. We analyse which models provide the most information for our predictions at each point in time and for each country, allowing us to gain some insights into economic mechanisms underlying the building of risk in economies.

Machine Learning and Financial Crises Prediction

Online learning:
NOT big data but model AGGREGATION

This framework is very suitable for crisis prediction in real time:

- **Multivariate** : Which variables cause a financial crisis?
- **Time-varying weights** : Causes of financial crises may be different over time.
- **Statistically robust** : overfitting is a problem in the literature.

- **Not "black-box"** : assess the role each model plays to predict the pre-crisis.
- **Theoretically grounded** : asymptotic properties of our aggregation rules ensure convergence.
- **More general than Bayesian Model Averaging**
- This framework has been used to predict French electricity load (EDF); the tracking of climate models; the network traffic demand.

Machine Learning and Financial Crises Prediction

Sequential predictions

Online learning is performed in a sequence of consecutive rounds where at time instance t the forecaster:

1. Receives a question.
2. Uses expert advice $\{f_{j,t} \in \mathcal{D} : j \in \mathcal{E}\}$
3. Predicts $\hat{y}_t \in \mathcal{Y}$
4. Receives true answer $y_t \in \mathcal{Y}$
5. Suffers a loss $\ell(\hat{y}_t, y_t)$.

To combine experts' advice, the forecaster chooses a sequential aggregation rule $\mathcal{S}$ which consists in setting a time-varying weight vector $(p_{1,t}, ..., p_{N,t}) \in \mathcal{P}$:

$$\hat{y}_t = \sum_{j=0}^{N} p_{j,t} f_{j,t}$$

The forecaster and each expert incur a cumulative loss defined by :

$$L_T(\mathcal{S}) = \sum_{t=1}^{T} \ell\Big(\sum_{j=0}^{N} p_{j,t} f_{j,t}\Big) = \sum_{t=1}^{T} (\hat{y}_t - y_t)^2$$

Machine Learning and Financial Crises Prediction

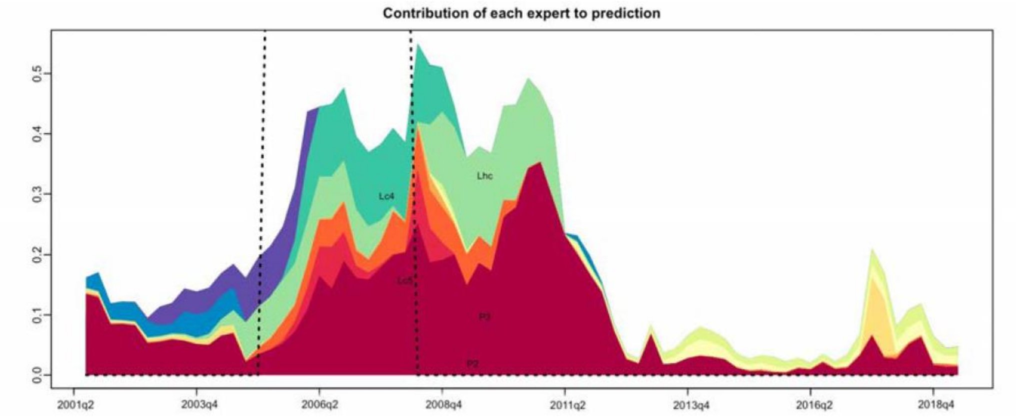


Figure 13: France: Experts contribution to forecast. OGD aggregation rule

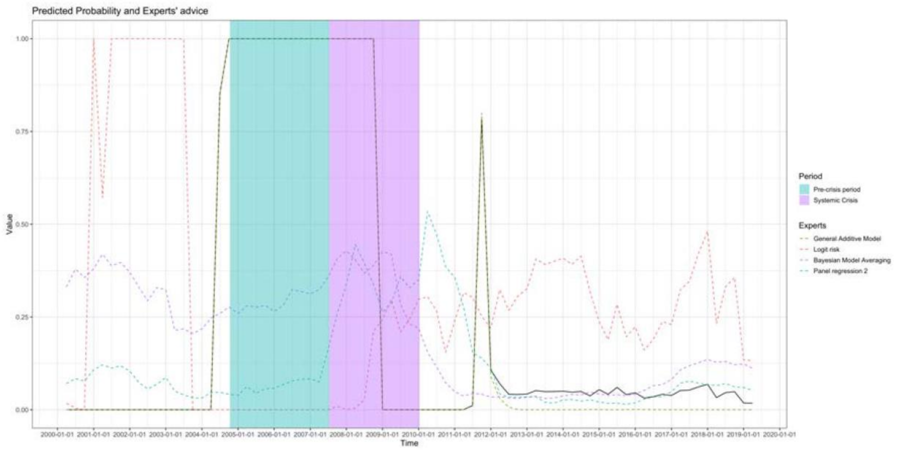


Figure 21: UK: A Subset of Experts' Predicted Probability of Crises versus EWA aggregation.

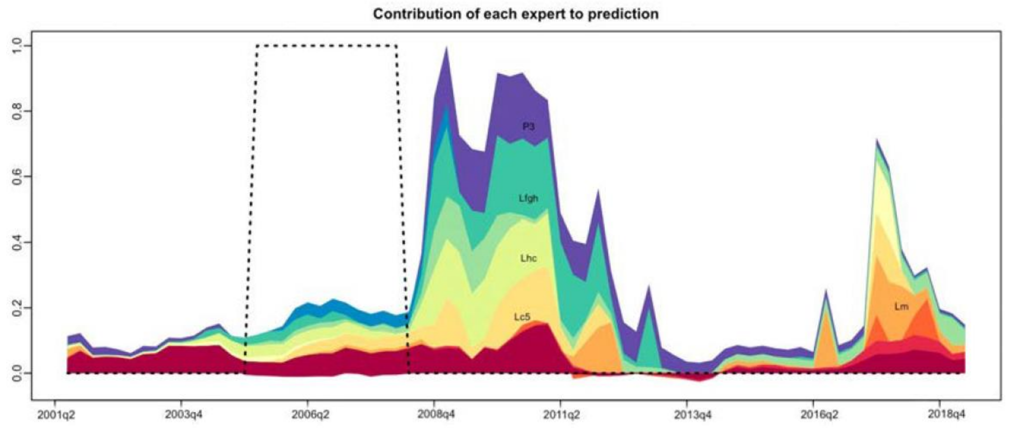


Figure 14: France: Experts contribution to forecast. Ridge aggregation rule

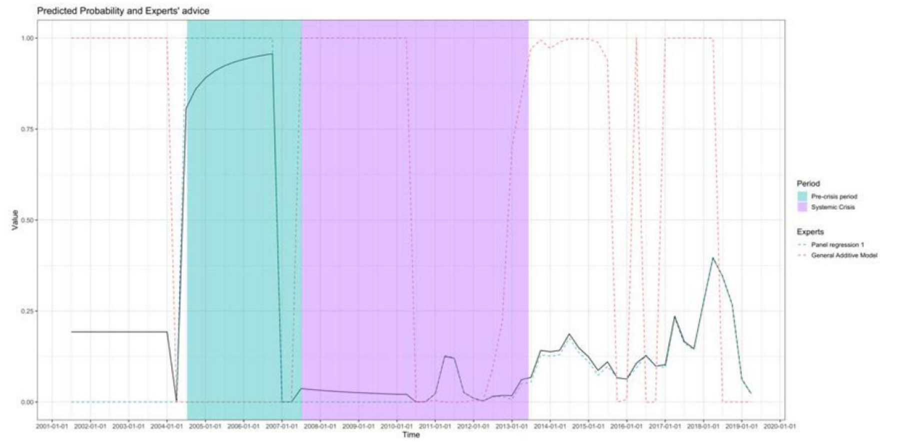


Figure 24: Germany: A Subset of Experts' Predicted Probability of Crises versus EWA aggregation.

Fouliard, Jeremy, Michael Howell, and Hélène Rey. Answering the queen: Machine learning and financial crises. No. w28302. National Bureau of Economic Research, 2021.

Financial Machine Learning

Financial Machine Learning
Bryan T. Kelly and Dacheng Xiu
July 2023
JEL No. C33,C4,C45,C55,C58,G1,G10,G11,G12,G17

ABSTRACT

We survey the nascent literature on machine learning in the study of financial markets. We highlight the best examples of what this line of research has to offer and recommend promising directions for future research. This survey is designed for both financial economists interested in grasping machine learning tools, as well as for statisticians and machine learners seeking interesting financial contexts where advanced methods may be deployed.

AI for Computation: Quantitative Trade Model

AI for Global Trade and Industrial Policies

- Many governments around the world have resorted to trade and industrial policies in recent years

- We study optimal trade and industrial policies in a quantitative trade model

Zi Wang, Xingcheng Xu, Yanqing Yang, Xiaodong Zhu. "Optimal Trade and Industrial Policies in the Global Economy: A Deep Learning Framework." arXiv preprint arXiv:2407.17731, 2024.

AI for Global Trade and Industrial Policies

- Optimal trade and industrial policies in multi-country-multi-sector quantitative trade models
  - Combination of policies: e.g. Trump's tariffs + Biden's industrial policies
  - Non-cooperative games among inter-connected countries
- Challenges
  - High-dimensional equilibrium system: $3NJ + N = 931$ for $N = 7$ and $J = 44$
  - High-dimensional policy space: $(N-1)J + J = 308$ for $N = 7$ and $J = 44$
  - Multiple players: iterative optimization
- Existing literature
  - Simplifying assumptions $\Rightarrow$ analytical characterizations: e.g. Lashkaripour and Lugovskyy (2023)
  - Restrictions on policy space or model: e.g. Ju et al. (2023); Ossa (2014)

AI for Global Trade and Industrial Policies

- Computing optimal tariffs: Ossa (2014); Judd and Su (2012)
    - This paper is the first attempt to use gradient-based learning algorithm and machine-learning implementation to solve for optimal policies in multi-country-multi-sector GE models

- Jointly consider optimal trade and industrial policies: Bartelme et al. (2021); Lashkaripour and Lugovskyy (2023); Ju et al. (2023)
    - This paper computes the fully optimal policies without simplification assumptions or restrictions on policy space

Global Trade and Industrial Policies

- N countries with labor endowments $\{L_i\}_{i=1}^N$. J sectors.

- Labor: immobile across countries but perfectly mobile across sectors.

- Preference:

$$U_i = \sum_{j=1}^{J} \alpha_i^j \log \left[\left(\int_0^1 \left[C_i^j(\omega) \right]^{\frac{\sigma_j - 1}{\sigma_j}} d\omega \right)^{\frac{\sigma_j}{\sigma_j - 1}} \right]$$

- Perfect competition: the unit cost of variety ω of intermediate j in country i is $c_i^j(\omega) = \frac{1}{z_i^j(\omega)} c_i^j$ where

$$c_i^j = \underbrace{\frac{1}{\left(L_i^j\right)^{\psi_j}}}_{\text{Sectoral Scale Economy}} w_i^{\beta_i^j} \left[\prod_{s=1}^{J} (P_i^s)^{\gamma_i^{sj}} \right]^{1-\beta_i^j}, \quad \sum_{s=1}^{J} \gamma_i^{sj} = 1,$$

where P_i^s is the price index of good s in country i and L_i^j is the labor allocated to sector j of country i

- Productivity $z_i^j(\omega)$ is drawn from:

$$Pr\left[z_i^j(\omega) \le z \right] = \exp\left\{ -T_i^j z^{-\theta_j} \right\}, \quad z > 0, \quad \theta_j > \max\{\sigma_j - 1, 1\}$$

General Problem

- N players (countries): action space $a \equiv (a_i, a_{-i}) \in \mathcal{A}$

- Given a_{-i}, country i solves for the best-response function $\mathbf{b}_i^*(.)$

$$\mathbf{b}_i^*(a_{-i}) \equiv \max_{(a_i;X)} W_i(a_i, a_{-i}; X), \quad \text{s.t. } G_i(X, a_i, a_{-i}) = 0, \tag{1}$$

where $G_i(.)$ is the equilibrium system, X refers to equilibrium outcomes, and $W_i(.)$ is the reward function

Gradient-based Learning

- Best-response dynamics

$$a_i^{t+1} = \eta_t \mathbf{b}_i^* \left(a_{-i}^t\right) + (1 - \eta_t) a_i^t, \quad \forall i = 1, 2, \ldots, N, \quad \eta_t \in (0, 1]. \tag{2}$$

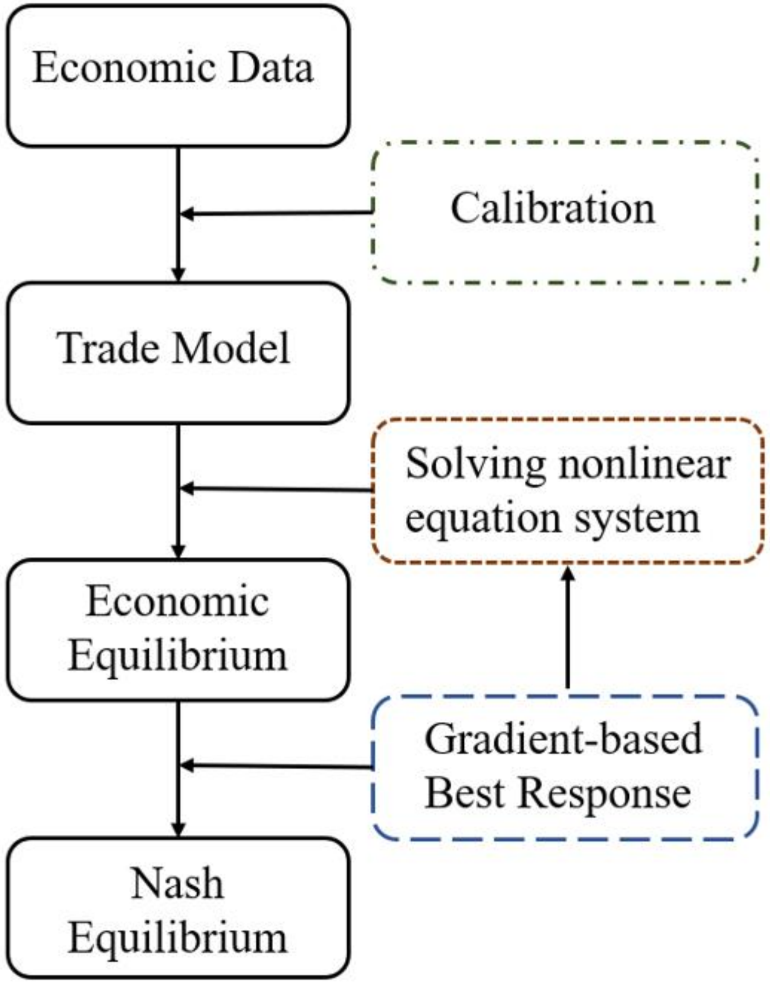
- Let $O_i(a_i) \equiv -W_i(a_i, a_{-i}^t; X)$ s.t. $G_i(X, a_i, a_{-i}) = 0$. Then the general update rule is

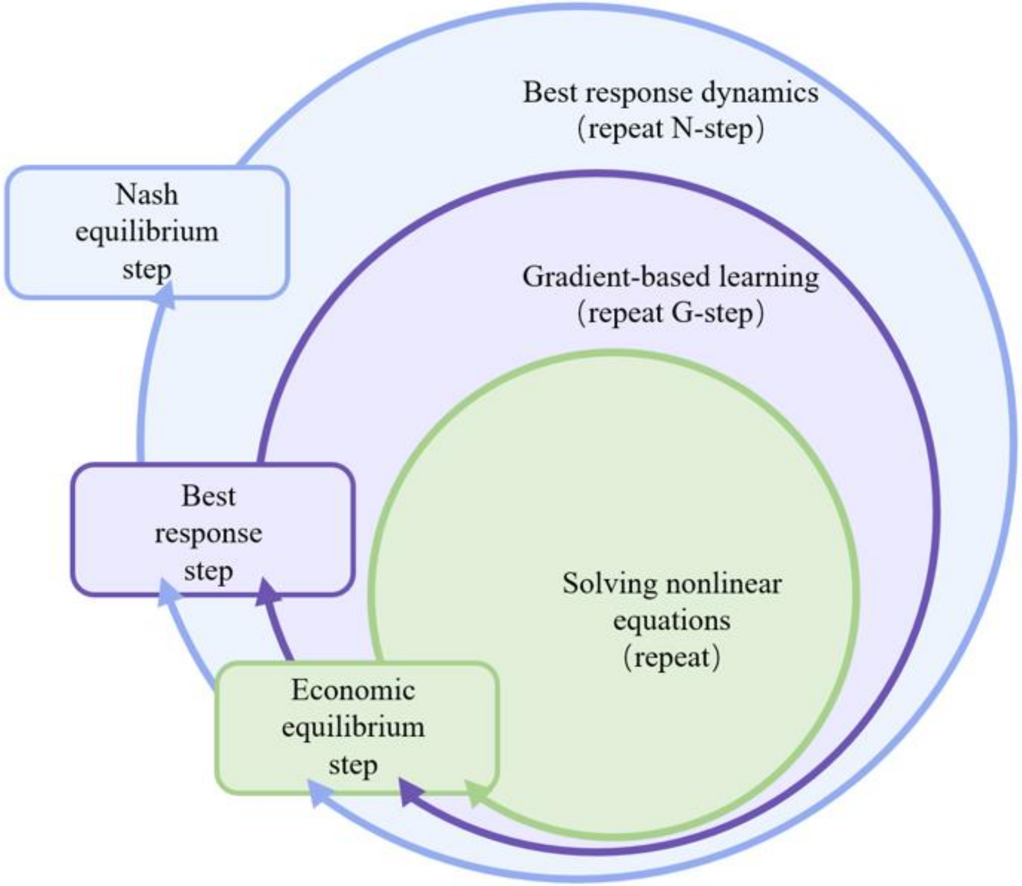
$$a_i^{t+1} = a_i^t - \gamma\omega(\nabla O_i|_{a_i = a_i^t}), \tag{3}$$

where a_i^t is the value at timestep t, γ is the learning rate (step size), $\nabla O_i|_{a_i = a_i^T}$ is the gradient of the objective function with respect to a_i at timestep t, and $\omega(.)$ is a functional of the gradient

- Countries play following **random shuffle playing sequence** for each round
  - Presuming cyclic order of policy functions may offer some players a positional advantage

Global Trade and Industrial Policies





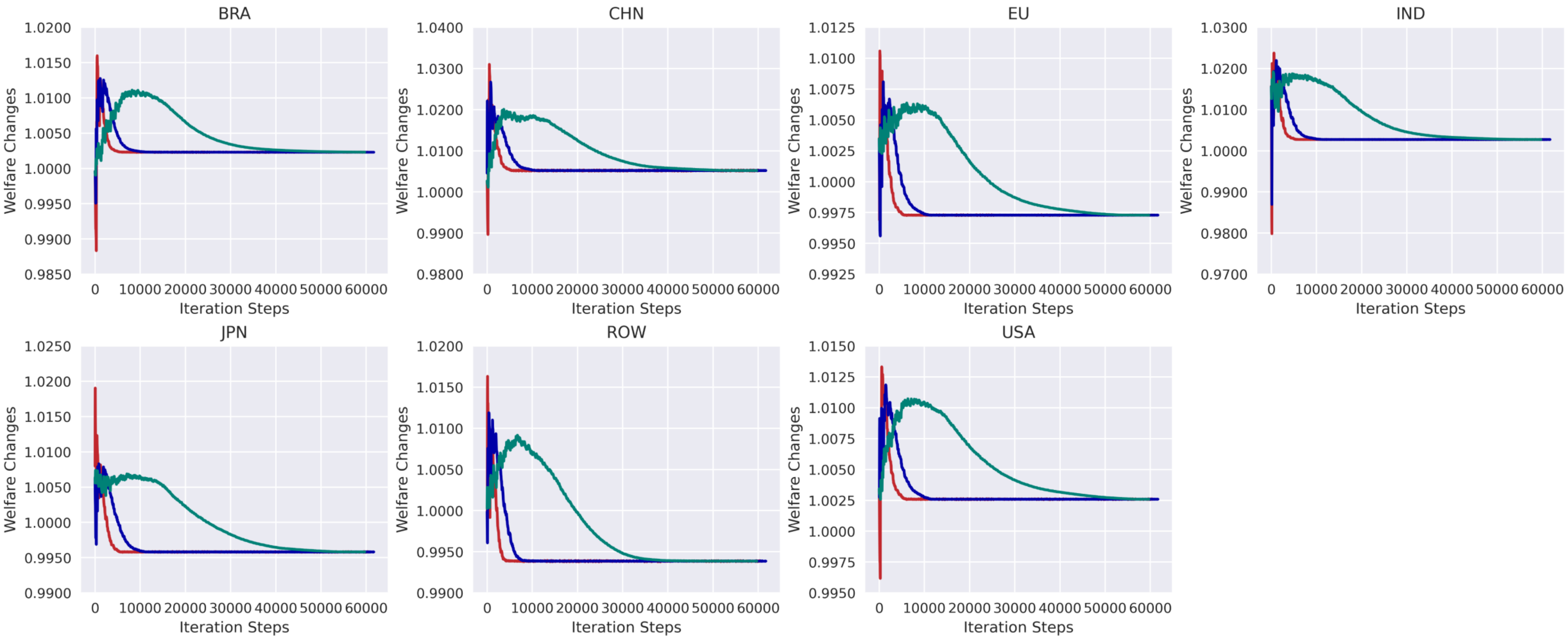
Global Trade and Industrial Policies

Table 2: Welfare Changes at Nash Equilibrium with Scale Economies

	China and US (Δ%)			World (Δ%)				China (Δ%)		
	Subsidy	Tariff	Dual	Subsidy	Subsidy-Uni	Tariff	Dual	Subsidy	Tariff	Dual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
United States	1.28	-0.07	1.23	1.31	1.31	-1.01	0.26	-0.09	0.08	-0.03
China	3.31	-0.33	3.17	2.42	1.80	-2.56	0.52	3.71	2.22	5.78
European Union	-0.51	0.02	-0.49	1.03	0.53	-1.80	-0.27	-0.17	-0.11	-0.28
Japan	-0.61	0.03	-0.57	1.14	0.30	-2.21	-0.42	-0.36	-0.42	-0.67
India	-0.50	0.01	-0.51	2.56	0.60	-1.84	0.27	-0.01	1.20	-0.47
Brazil	-0.05	0.00	-0.05	1.97	1.71	-2.08	0.23	0.17	-0.56	-0.44
Rest of the World	-0.44	0.05	-0.41	1.47	1.76	-2.33	-0.62	-0.17	-0.93	-1.19

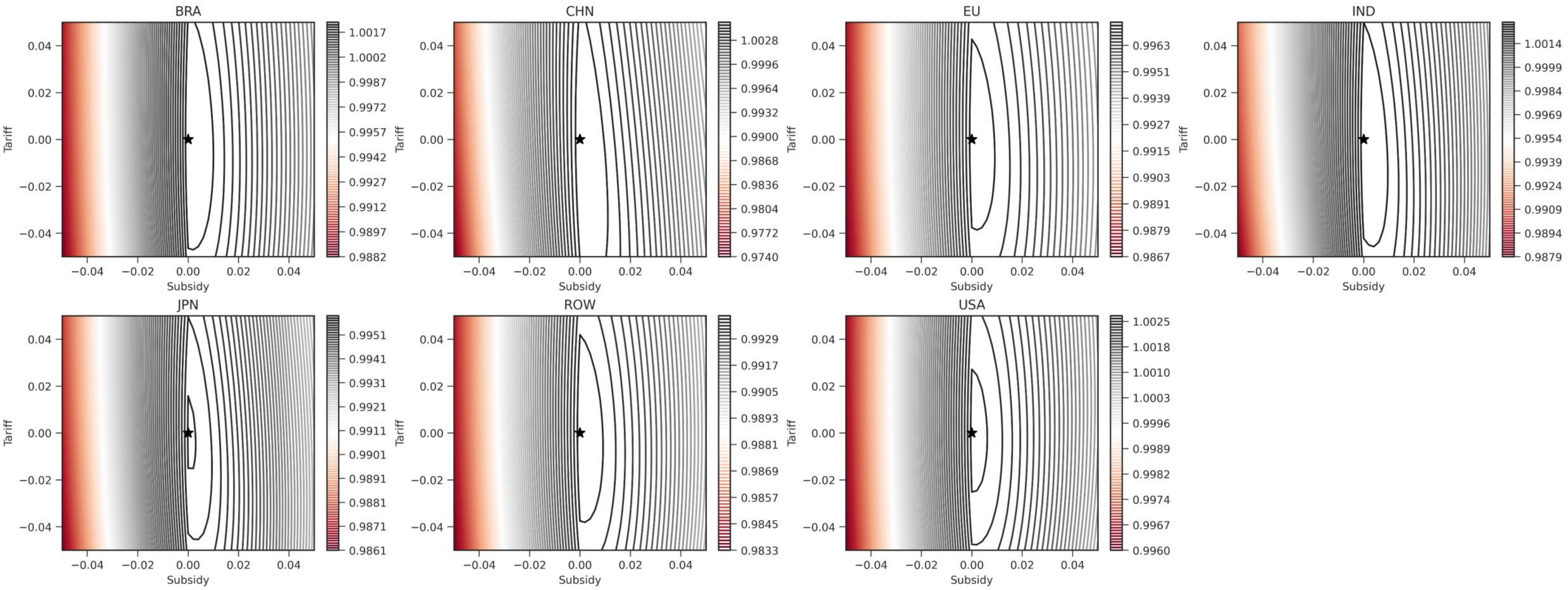
Note: "Subsidy" refers cases where players can adjust their industry subsidies only. "Subsidy-uni" refers to the cases where each player can only choose a uniform subsidy rate for all manufacturing sectors (sector 6-22 in Table A.1). "Tariff" refers to cases where players can modify their import tariffs solely. "Dual" refers to cases where players have the flexibility to adjust both their industry subsidies and import tariffs. "China and US" refers to cases where only China and the US are allowed to adjust their policies, whereas "World" refers to cases where all economies can adjust their policies.

Global Trade and Industrial Policies



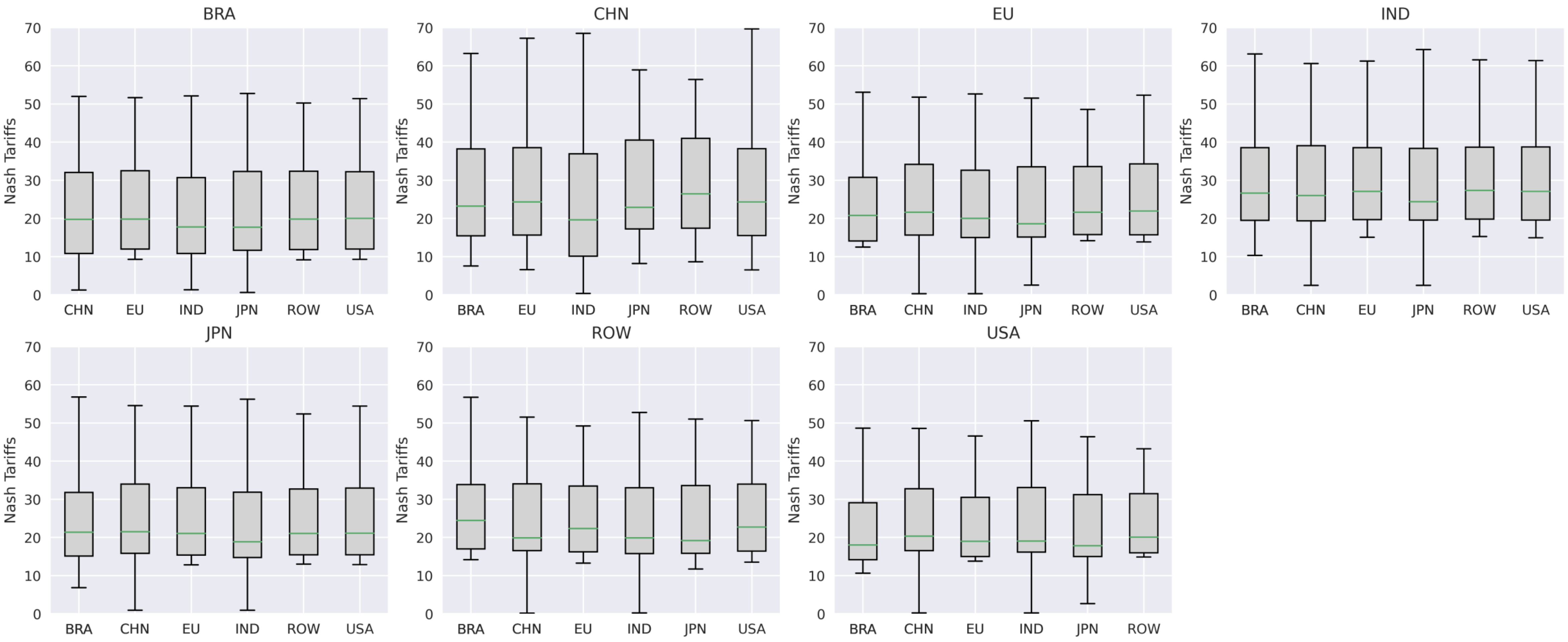
Iteration Curve for Nash Equilibrium of Global Dual Policy Competition with Scale Economies

Global Trade and Industrial Policies



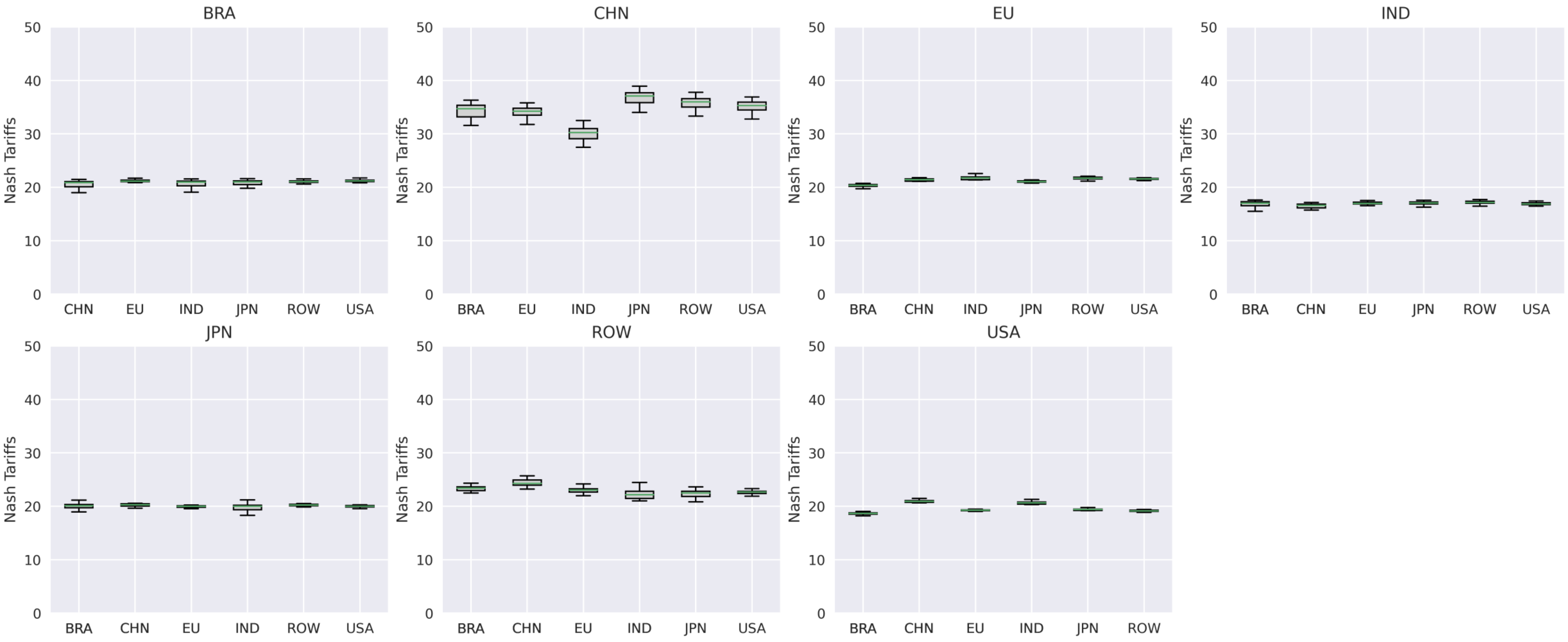
Landscape Near Nash Equilibrium for Global Dual Policy Competition with Scale Economies

Global Trade and Industrial Policies



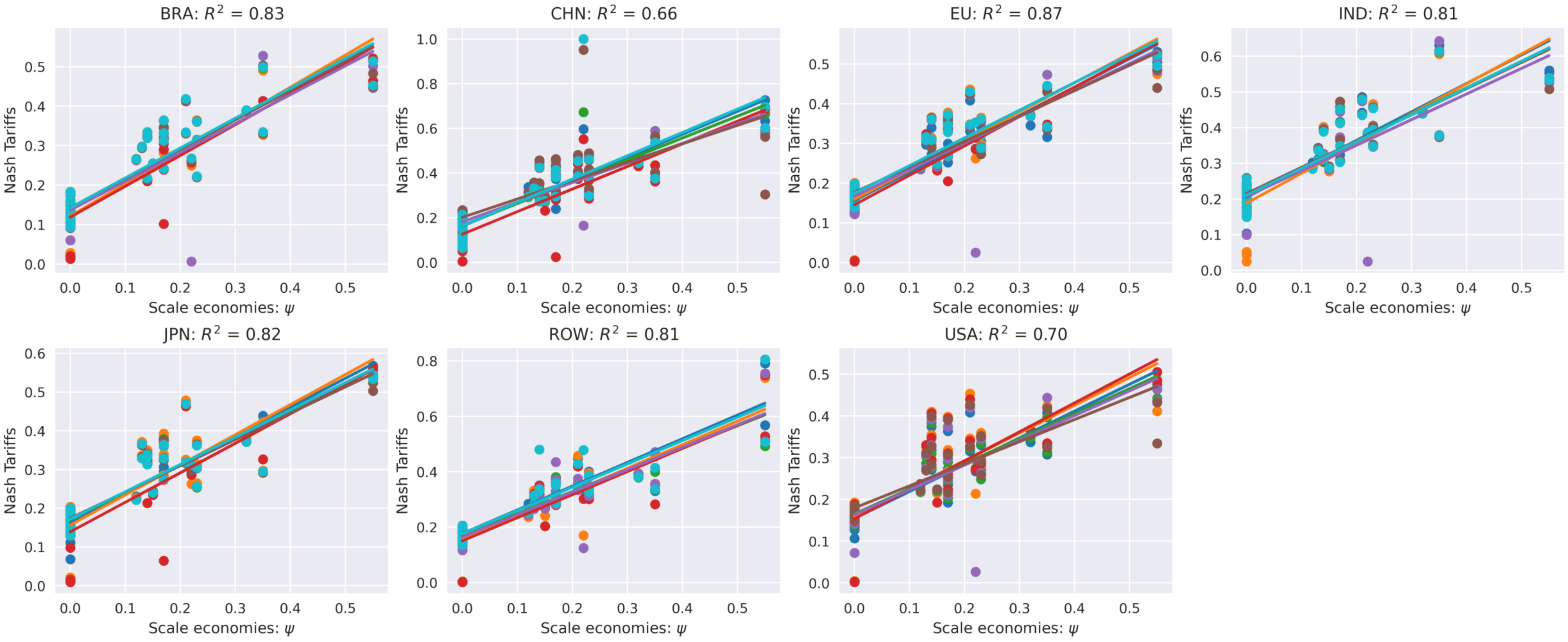
Nash Tariffs for Global Dual Policy Competition with Scale Economies

Global Trade and Industrial Policies



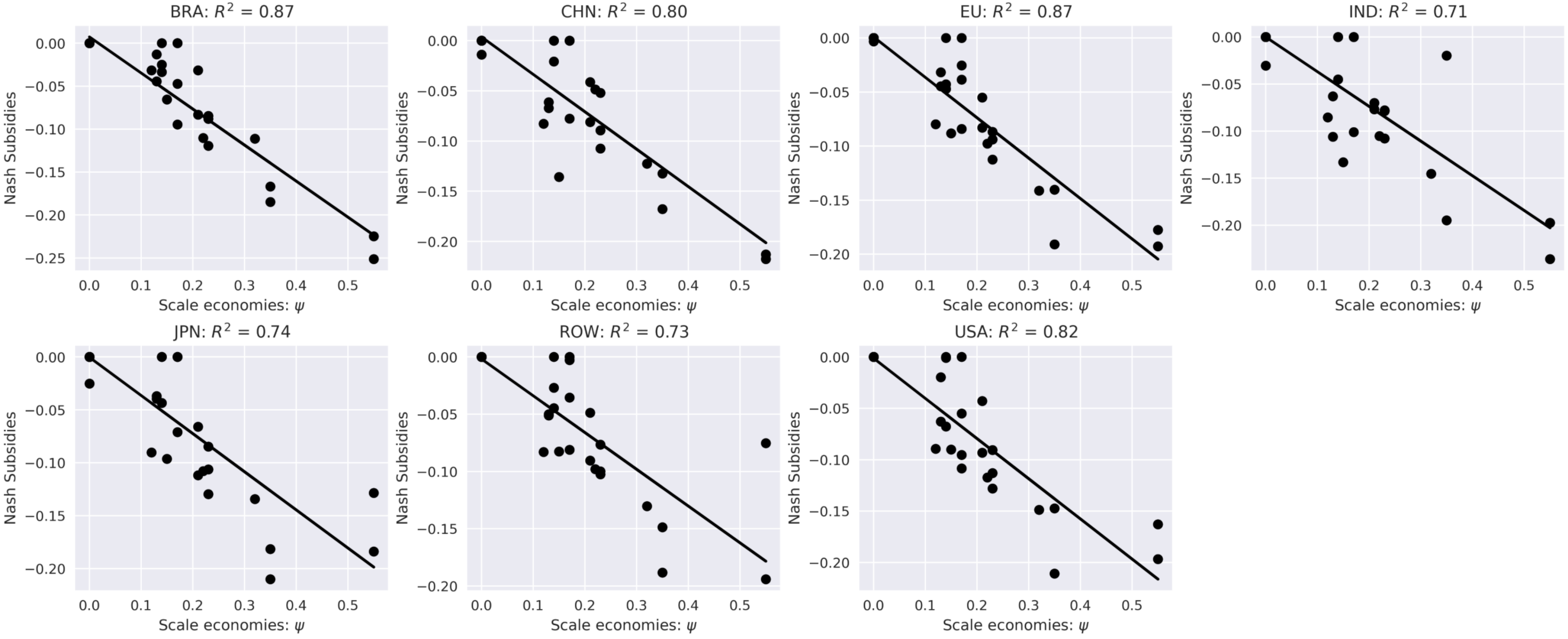
Nash Tariffs for Global Dual Policy Competition without Scale Economies

Global Trade and Industrial Policies



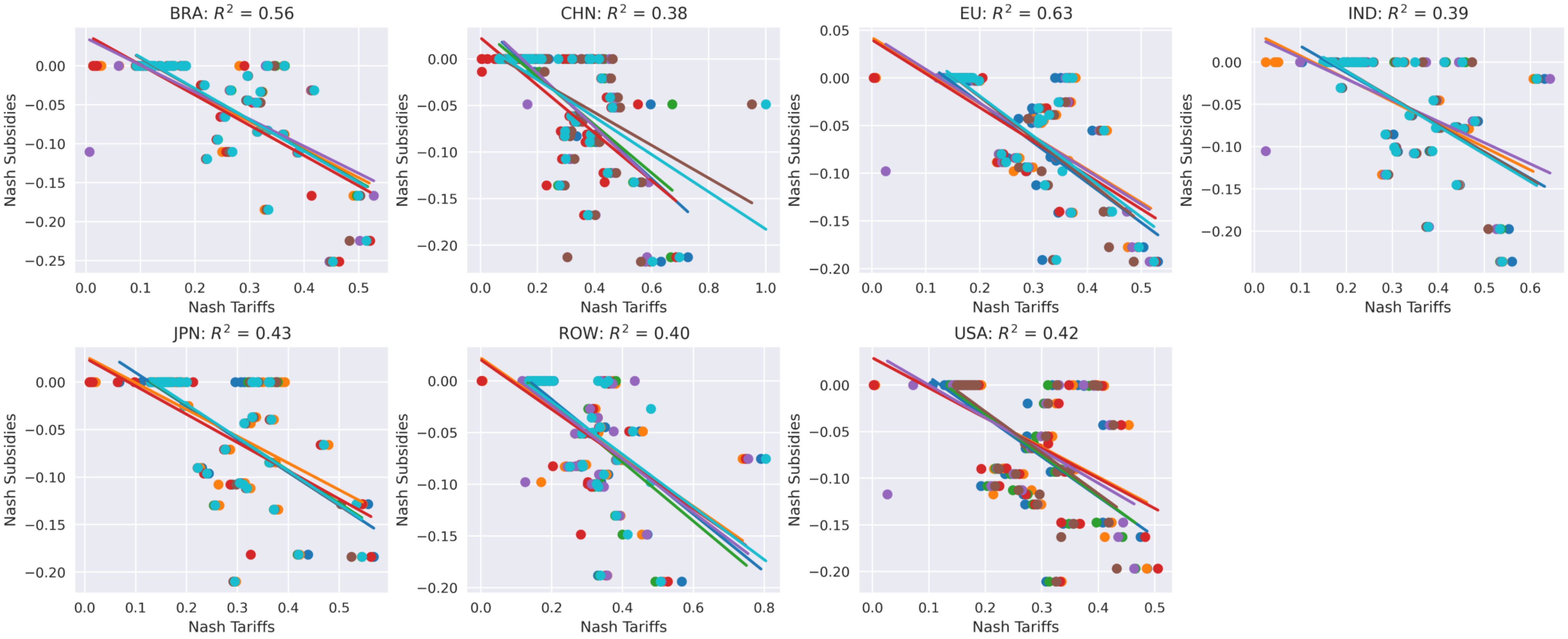
Nash Tariffs for Global Dual Policy Competition with Scale Economies

Global Trade and Industrial Policies



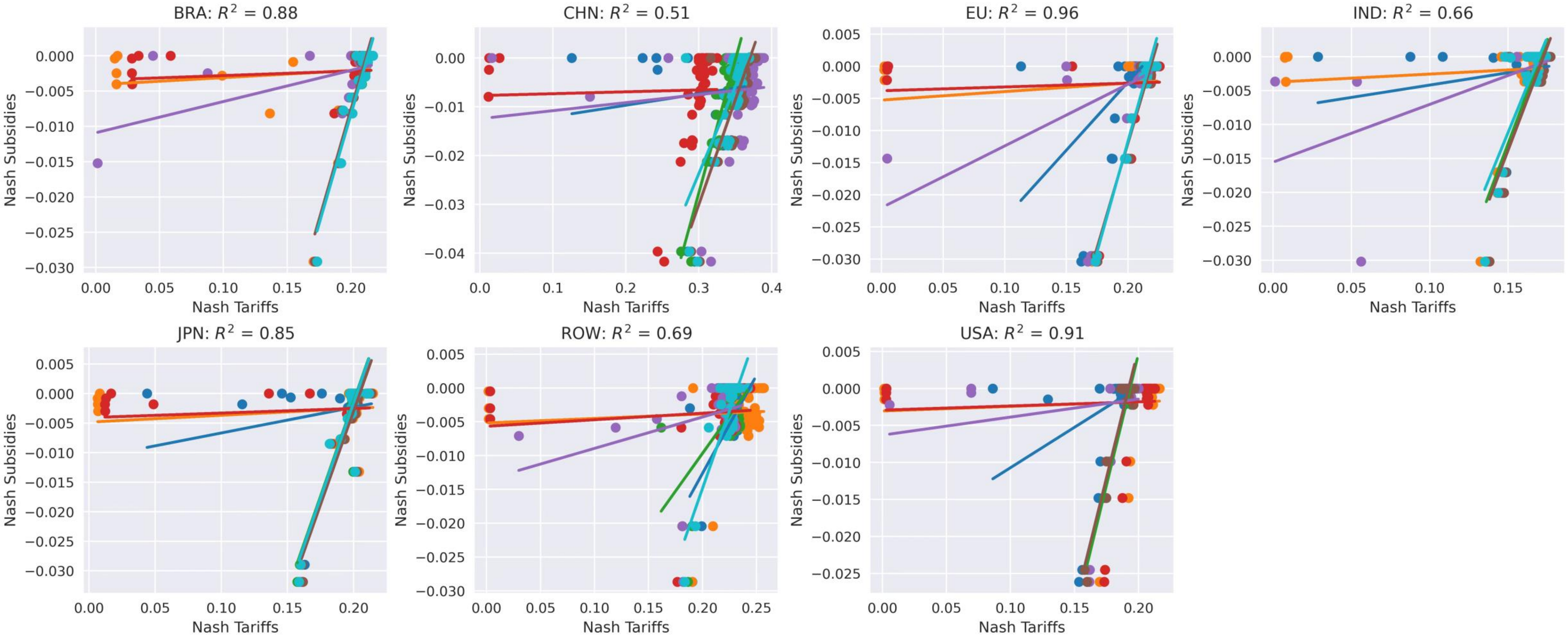
Nash Subsidies for Global Dual Policy Competition with Scale Economies

Global Trade and Industrial Policies



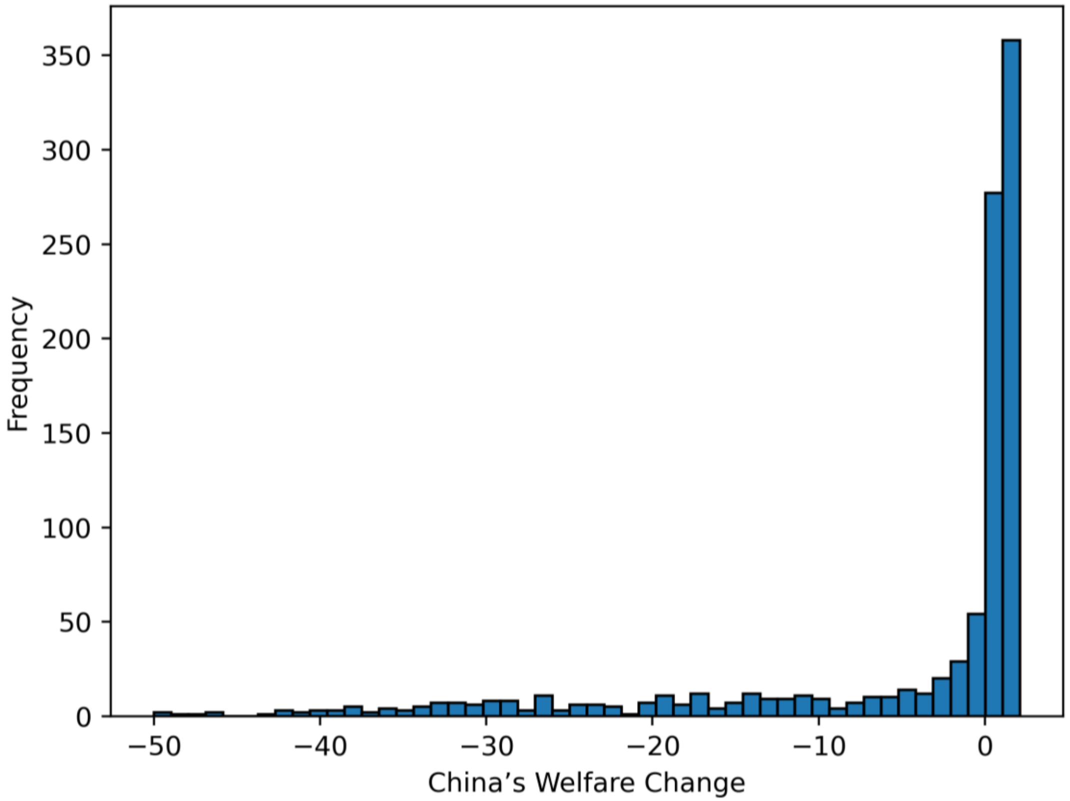
Complementary Between Nash Tariffs and Subsidies for Global Dual Policy Competition with Scale Economies

Global Trade and Industrial Policies



Substitutes Between Nash Tariffs and Subsidies for Global Dual Policy Competition without Scale Economies

Global Trade and Industrial Policies



China's Welfare Change (%) under Randomly Drawn Industrial Subsidies.

Mean: -4.61%, Median: 0.66%, Max: 2.11%.

(Note: Each e^j_{CHN} is drawn uniformly from $\left[0.1e^{j*}_{CHN}, 1.9e^{j*}_{CHN}\right]$. We draw $\left(e^j_{CHN}\right)^{22}_{j=1}$ for 1000 times.)

Global Trade and Industrial Policies

- This is the first attempt to use gradient-based learning algorithm and machine-learning implementation to solve for optimal policies in multi-country-multi-sector GE models with scale economies.
- Efficiently solve high-dimensional optimal policies under high-dimensional (nonlinear) equilibrium system.
- Nash tariffs lead to significant welfare losses, whereas Nash subsidies, if properly implemented, lead to considerable welfare gains for all countries.

Heterogenous Agent New Keynesian Model (HANK)

- Idiosyncratic and aggregate shocks, incomplete markets and borrowing constraints, see Krusell and Smith (1988).
  - The economy has ℓ heterogeneous agents that differ in capital and productivity that have a joint distribution $\{k_t^i, y_t^i\}_{i=1}^{\ell}$ for $i = 1, ..., \ell$.
  - The economy has also aggregate productivity z_t.
  - Thus, the economy has $2\ell + 1$ state variables $\left(\{k_t^i, y_t^i\}_{i=1}^{\ell}, z_t\right)$, for example, if $\ell = 1,000$ the state space has 2001 state variables.
- HANK is heterogeneous-agent new Keynesian model that includes monetary policy, for example, Taylor rule for the nominal interest rate.

Heterogenous Agent New Keynesian Model (HANK)

Krusell-Smith (1998) model with savings through capital

- Heterogeneous agents $i = 1, ..., \ell$ solve

$$\max_{\{c_t^i, k_{t+1}^i\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t^i\right) \right]$$

$$\text{s.t. } c_t^i + k_{t+1}^i = R_t k_t^i + W_t y_t^i,$$

where c_t^i, $k_t^i \geq 0$ are consumption and capital; idiosyncratic labor productivity follows $\ln y_{t+1}^i = \rho_y \ln y_t^i + \sigma_y \epsilon_t^i$ with $\epsilon_t^i \sim \mathcal{N}(0,1)$; and $\beta \in (0,1)$; $\rho_y \in (-1,1)$ and $\sigma_y \geq 0$; and $\left(k_0^i, y_0^i\right)$ is given.

- The Cobb-Douglas production implies the interest rate R_t and wage W_t

$$R_t = 1 - d + z_t \alpha k_t^{\alpha-1} h_t^{1-\alpha} \text{ and } W_t = z_t (1-\alpha) k_t^{\alpha} h_t^{-\alpha},$$

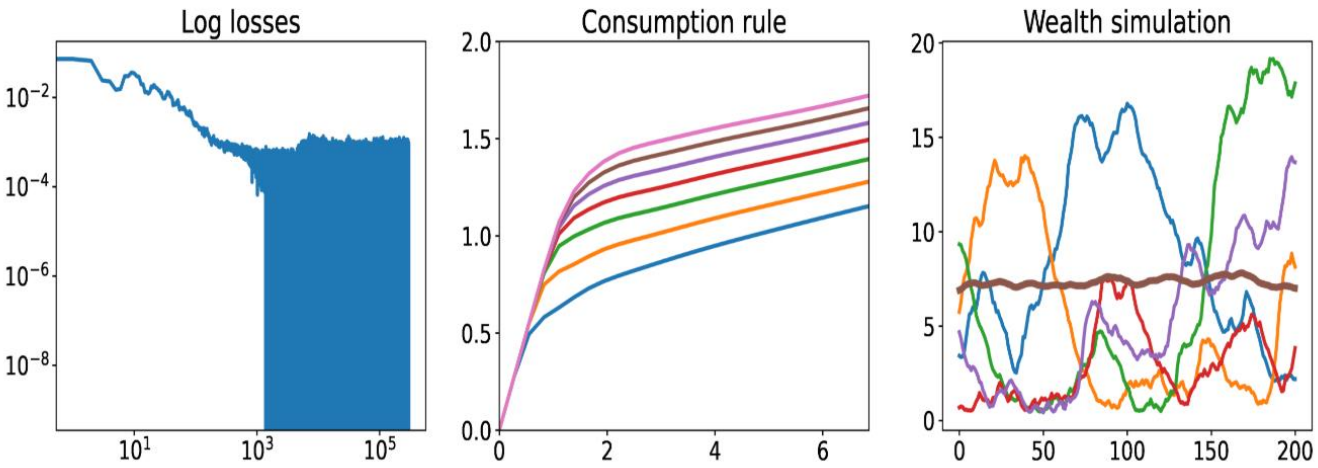
where $d \in (0,1]$ is the depreciation rate, $k_t = \sum_{i=1}^{\ell} k_t^i$ is aggregate capital, $h_t = \sum_{i=1}^{\ell} y_t^i$ is aggregate efficiency labor, and aggregate shock follows $\ln z_{t+1} = \rho_z \ln z_t + \sigma_z \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0,1)$, $\rho_z \in (-1,1)$ and $\sigma_z \geq 0$.

Heterogenous Agent New Keynesian Model (HANK)

Algorithm 1: Deep learning for the model with capital
Step 0: (Initialization).
Construct initial state of the economy $\left(\left\{k_0^i, y_0^i\right\}_{i=1}^{\ell}, z_0\right)$ and parameterize three decision functions by a neural network with three outputs $$\left\{\frac{c_t^i}{w_t^i}\right\} = \sigma\left(\zeta_0 + \varphi\left(k_t^i, y_t^i, \left\{k_t^i, y_t^i\right\}_{i=1}^{\ell}, z_t; \theta\right)\right),$$ $$\mu_t^i = \exp\left(\zeta_0 + \varphi\left(k_t^i, y_t^i, \left\{k_t^i, y_t^i\right\}_{i=1}^{\ell}, z_t; \theta\right)\right),$$ where $w_t^i \equiv R_t k_t^i + W_t y_t^i n_t^i$ is wealth; μ_t^i is Lagrange multiplier associated with the borrowing constraint; $\varphi(\cdot)$ is a neural network; $\sigma(z) = \frac{1}{1+e^{-z}}$ is a sigmoid (logistic) function; ζ_0 is a constant; θ is a vector of coefficients.

Algorithm 1: Deep learning for the model with capital
Step 1: (Evaluation of decision functions).
Given state $\left(k_t^i, y_t^i, \left\{k_t^i, y_t^i\right\}_{i=1}^{\ell}, z_t\right) \equiv s_t^i$, compute μ_t^i, $\frac{c_t^i}{w_t^i}$ from the neural networks, find the prices R_t and W_t; and find k_{t+1}^i from the budget
constraint for all agents $i = 1, \ldots, \ell$.
Step 2: (Construction of Euler residuals).
Draw two random sets of individual productivity shocks $\Sigma_1 = \left(\epsilon_1^1, \ldots, \epsilon_1^{\ell}\right)$, $\Sigma_2 = \left(\epsilon_2^1, \ldots, \epsilon_2^{\ell}\right)$ and two aggregate shocks $\epsilon_{1,}$, ϵ_2, and construct Euler residuals $$\Xi(\theta) = \left\{\left[\Psi^{FB}\left(1 - \frac{c_t^i}{w_t^i}, 1 - \mu_t^i\right)\right]^2 + \varpi_\mu \left[\frac{\beta\left[\left(c_{t+1}^i\right)^{-\gamma} R_{t+1} \middle\vert \Sigma'_{t+1}, \epsilon'_{t+1}\right]}{\left(c_t^i\right)^{-\gamma}} - \mu_t^i\right]\left[\frac{\beta\left[\left(c_{t+1}^i\right)^{-\gamma} R_{t+1} \middle\vert \Sigma''_{t+1}, \epsilon''_{t+1}\right]}{\left(c_t^i\right)^{-\gamma}} - \mu_t^i\right]\right\},$$ where ϖ_n, ϖ_μ are given weights and $\Psi^{FB}(a, b) = a + b - \sqrt{a^2 + b^2}$ is a Fischer-Burmeister function.

Algorithm 1: Deep learning for the model with capital
Step 3: (Training).
Train the neural network coefficients θ to minimize the residual function $\Xi(\theta)$ by using a stochastic gradient descent method $\theta \quad \theta - \lambda\nabla_\theta \Xi(\theta)$ with $\nabla_\theta \Xi(\theta) \approx \frac{1}{N}\sum_{n=1}^{N} \nabla_\theta \xi\left(\omega_n; \theta\right)$, where $n = 1, \ldots, N$ denotes batches.
Step 4: (Simulation).
Move to $t+1$ by using endogenous and exogenous variables of Step 3 under $\Sigma_1 = \left(\epsilon_1^1, \ldots, \epsilon_1^{\ell}\right)$ and ϵ_1 as a next-period state $\left(\left\{k_{t+1}^i, y_t^i\right\}_{i=1}^{\ell}, z_{t+1}\right)$.



DeepHAM

In iteration k, given $V^{(k)}(s)$, optimize policy $\mathcal{C}^{(k)}(s)$ on simulated paths.

In N-agent competitive equilm problem, when solving agent i's problem, fix other agents' policy from last "play". Iterate the following:

1. At "play" $\ell+1$, last play's policy $\mathcal{C}^{(k,\ell)}(s)$ is known.
2. For agent $i=1$, solve for her optimal policy $\mathcal{C}^{(k,\ell+1)}(s)$:

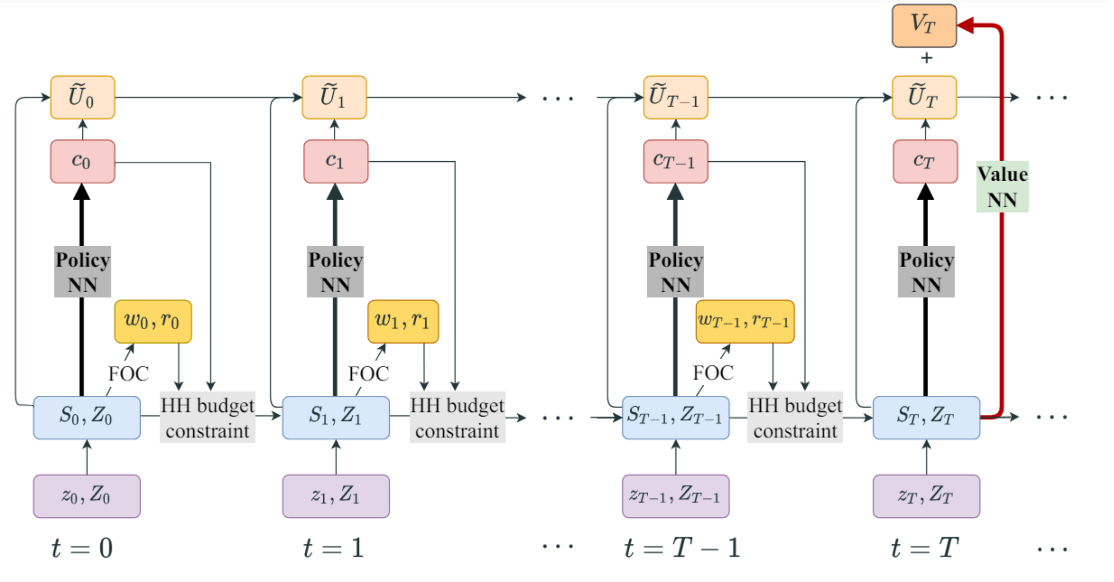
$$\max_{\mathcal{C}^{(k,\ell+1)}(s)} \mathbb{E}_{\mu(\mathcal{C}^{(k-1)}),\mathcal{E}}\left(\sum_{t=0}^{T}\beta^t u\left(c_{i,t}\right)+\beta^T V^{(k)}(s_{i,T})\right)$$

   subject to others all following $\mathcal{C}^{(k,\ell)}(s)$ in the first T periods.

3. All agents adopt the new policy $\mathcal{C}^{(k,\ell+1)}(s)$ in "play" $\ell+1$.

Optimization solved on Monte Carlo simulation with N agents on a large number of sample paths in a computational graph.

$$\max_{\Theta^C} \mathbb{E}_{\mu(\mathcal{C}^{(k-1)}),\mathcal{E}}\left(\widetilde{U}_{i,T}+\beta^T V_{\mathsf{NN}}(s_{i,T};\Theta^V)\right)$$



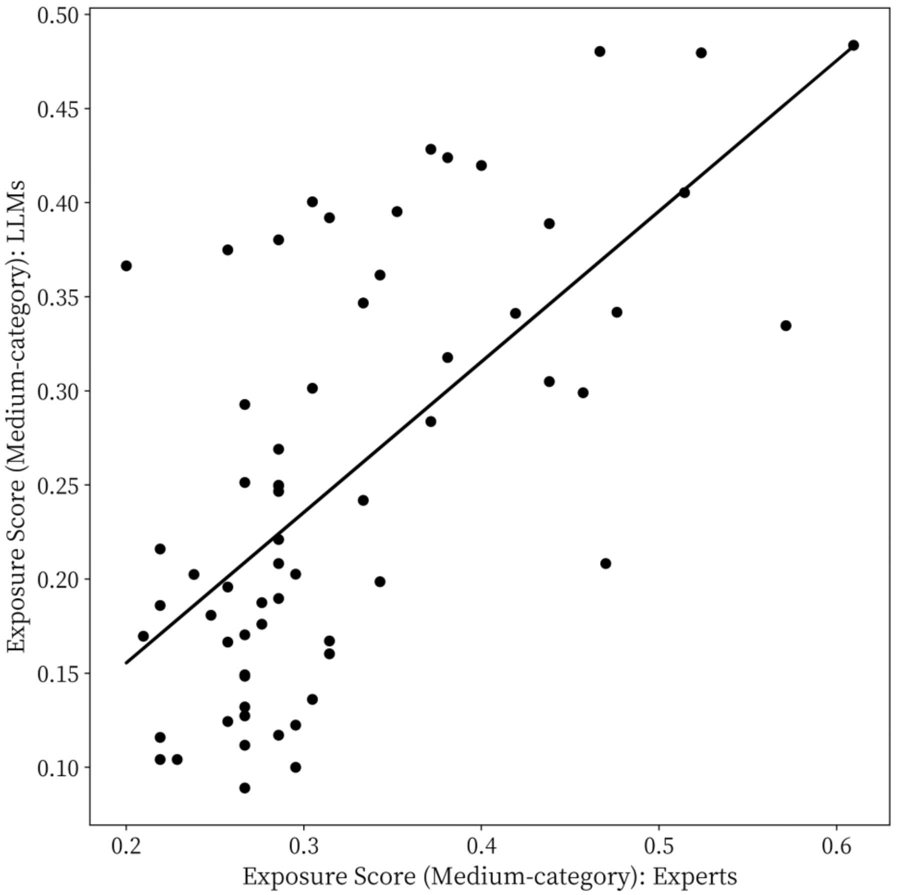
Budget constraint $a_{i,t+1}=(r_t+1-\delta)a_{i,t}+w_t\bar{\ell}y_{i,t}-c_{i,t}$. $s_t=(a_{i,t},y_{i,t},Z_t,\mathbf{\Gamma}_t)$.
Cumulative utility $\widetilde{U}_{i,t}=\sum_{\tau=0}^{t}\beta^\tau u\left(c_{i,\tau}\right)$

Han, Jiequn, Yucheng Yang and Weinan E. "DeepHAM: A global solution method for heterogeneous agent models with aggregate shocks." arXiv preprint arXiv:2112.14377 (2021).

AI for Text Analysis: Labor Market

LLMs at Work in China’s Labor Market

中国



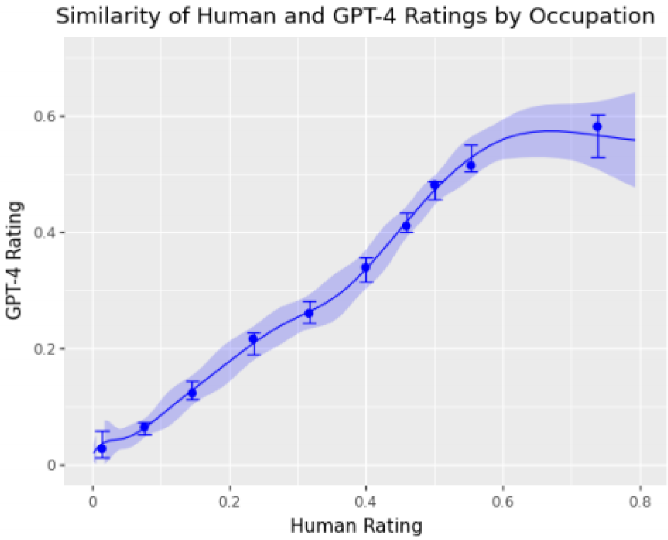
美国

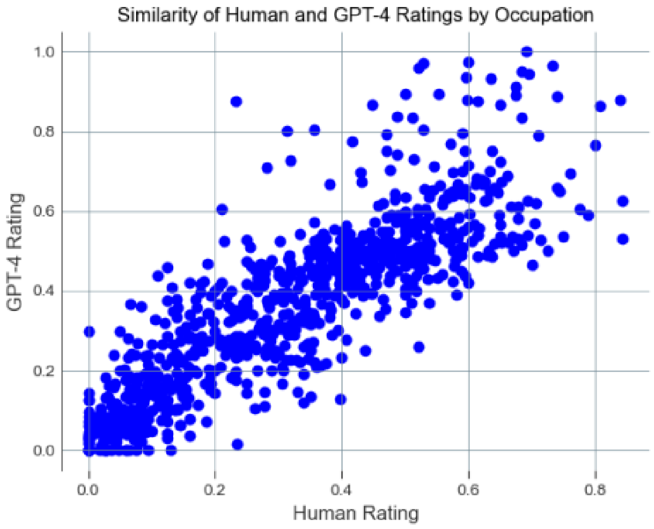
GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou[1], Sam Manning[1,2], Pamela Mishkin*[1], and Daniel Rock[3]

[1]OpenAI
[2]OpenResearch
[3]University of Pennsylvania

August 22, 2023





Summary Statistics of LLMs Exposure

Fine Categories Occupation Level Exposure			
	GLM	InternLM	GPT-4
count	1606	1606	1606
mean	0.44	0.18	0.24
std	0.26	0.18	0.21

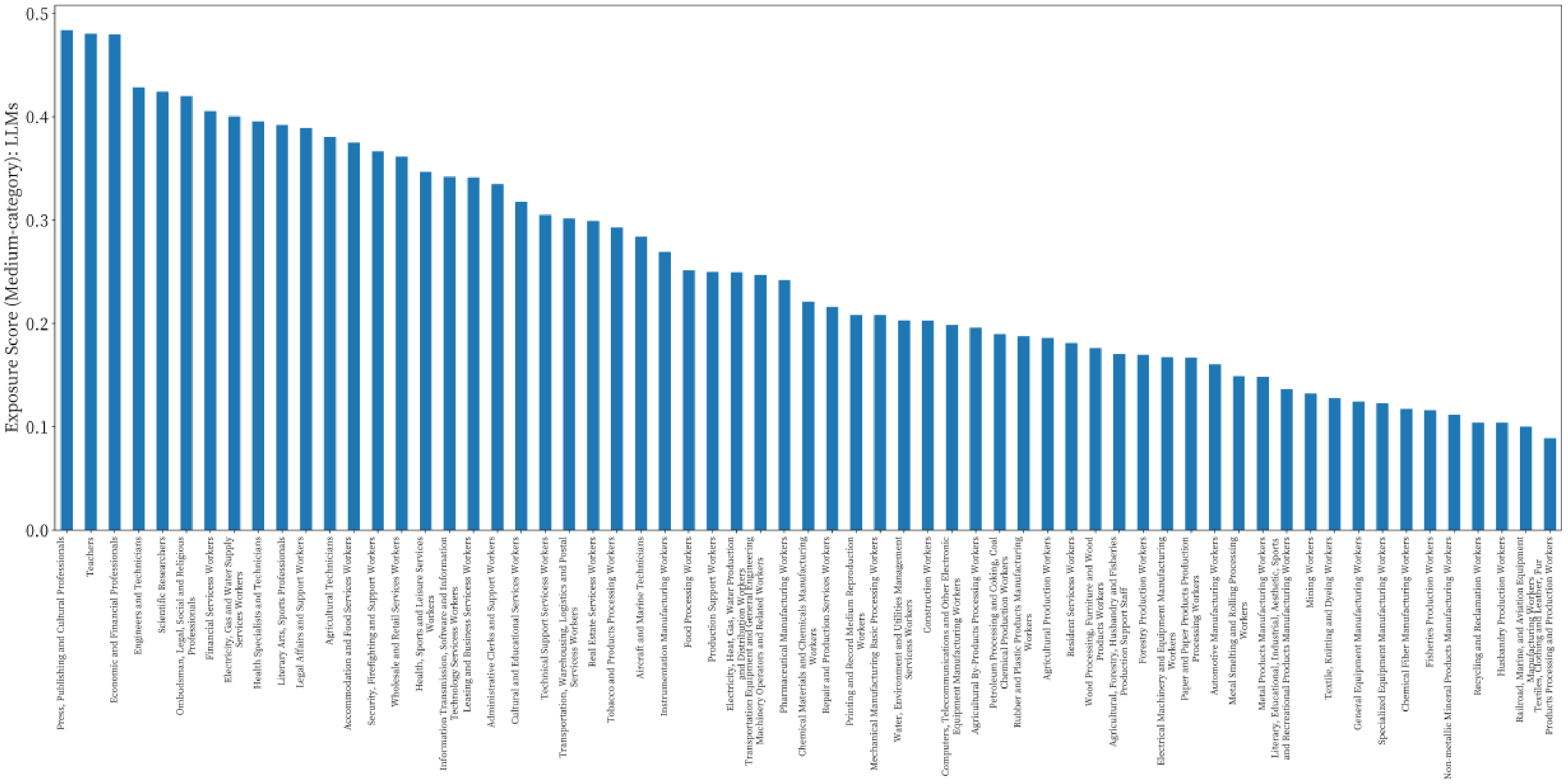
Fine Categories Occupation Level Exposure Corr.			
	GLM	InternLM	GPT-4
GLM	1.0***	0.284***	0.1915***
InternLM	0.284***	1.0***	0.2887***
GPT-4	0.1915***	0.2887***	1.0***

Medium Categories Occupation Level Exposure			
	GLM	InternLM	GPT-4
count	63	63	63
mean	0.40	0.14	0.22
std	0.15	0.10	0.18

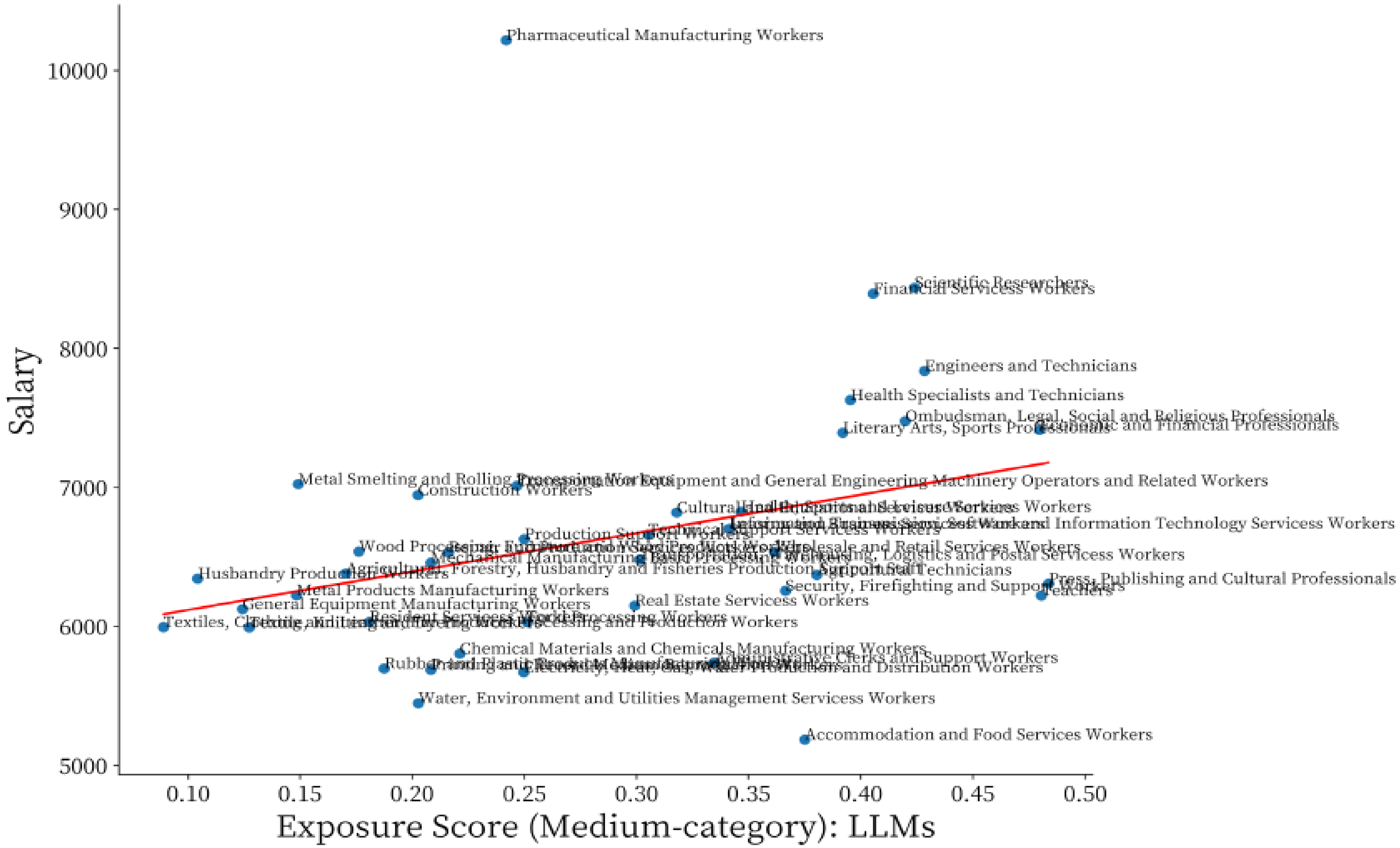
Medium Categories Occupation Level Exposure Corr.			
	GLM	InternLM	GPT-4
GLM	1.0***	0.5938***	0.306*
InternLM	0.5938***	1.0***	0.4807***
GPT-4	0.306*	0.4807***	1.0***

Qin Chen, Jinfeng Ge, Huaqing Xie, Xingcheng Xu*, Yanqing Yang. “Large Language Models at Work in China’s Labor Market.” arXiv preprint arXiv:2308.08776, 2023.

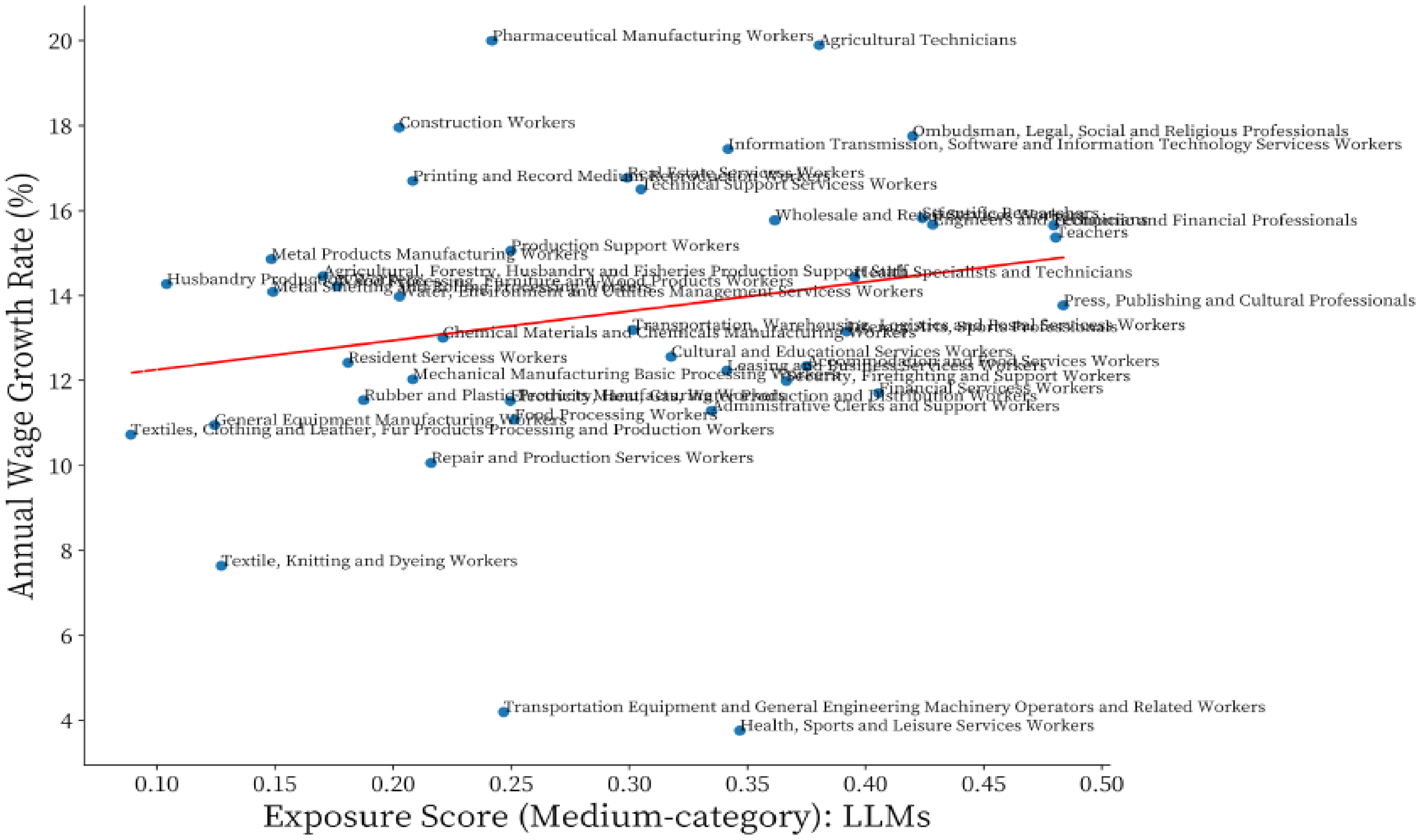
Exposure Score on Medium-category Level: LLMs



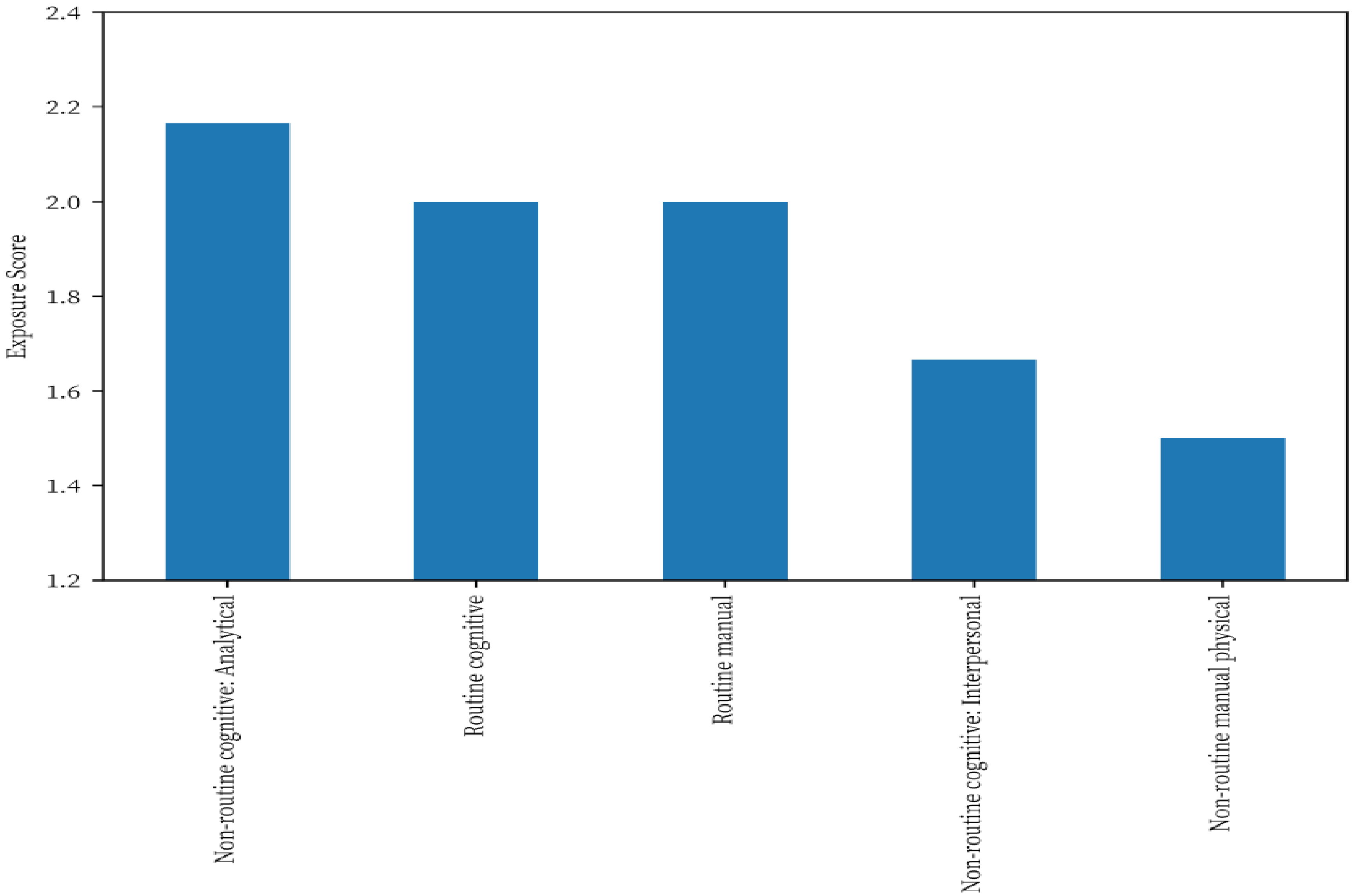
Salary and Exposure Score (Medium-category): LLMs



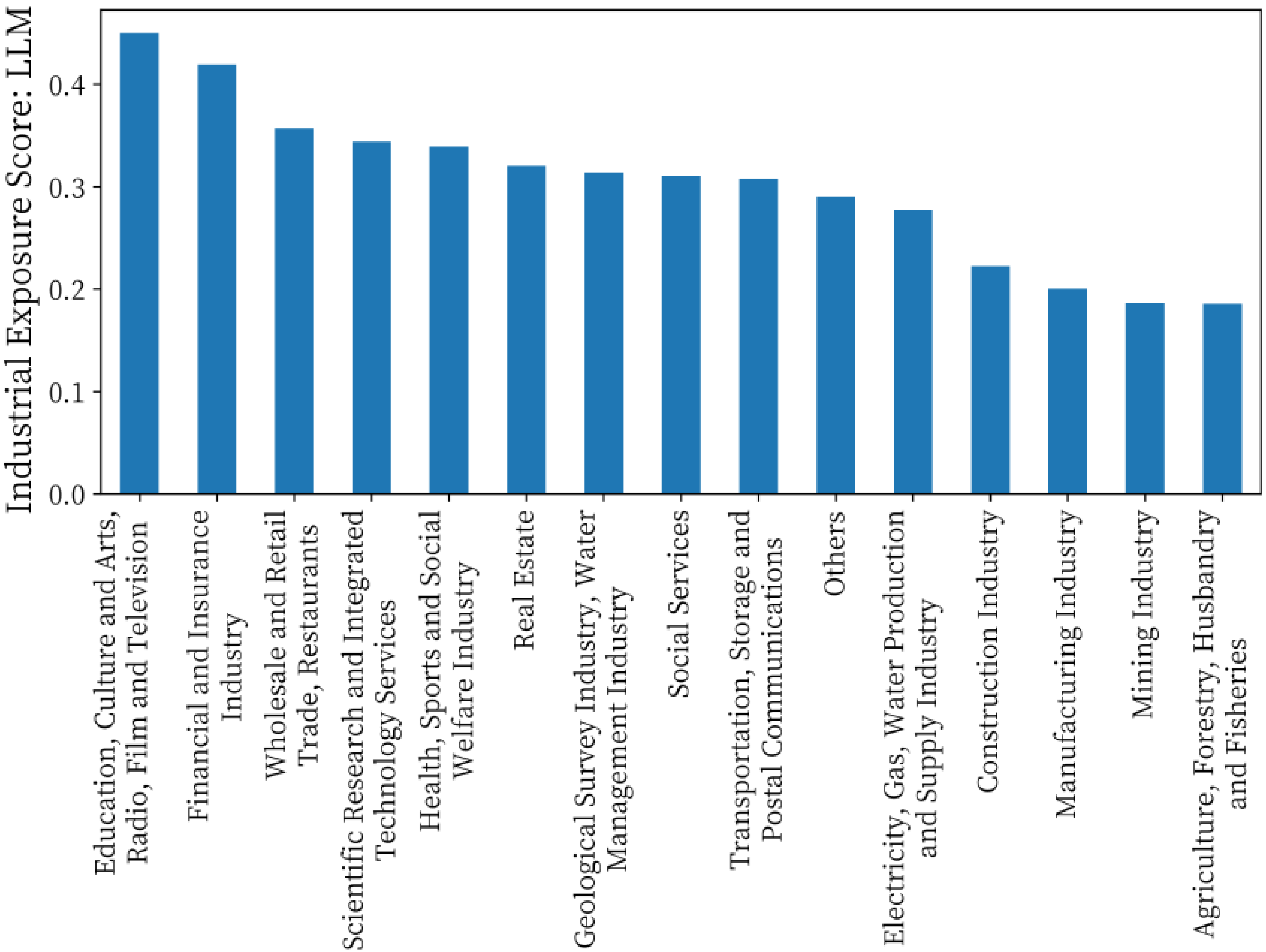
Annual Wage Growth Rate and Exposure Score



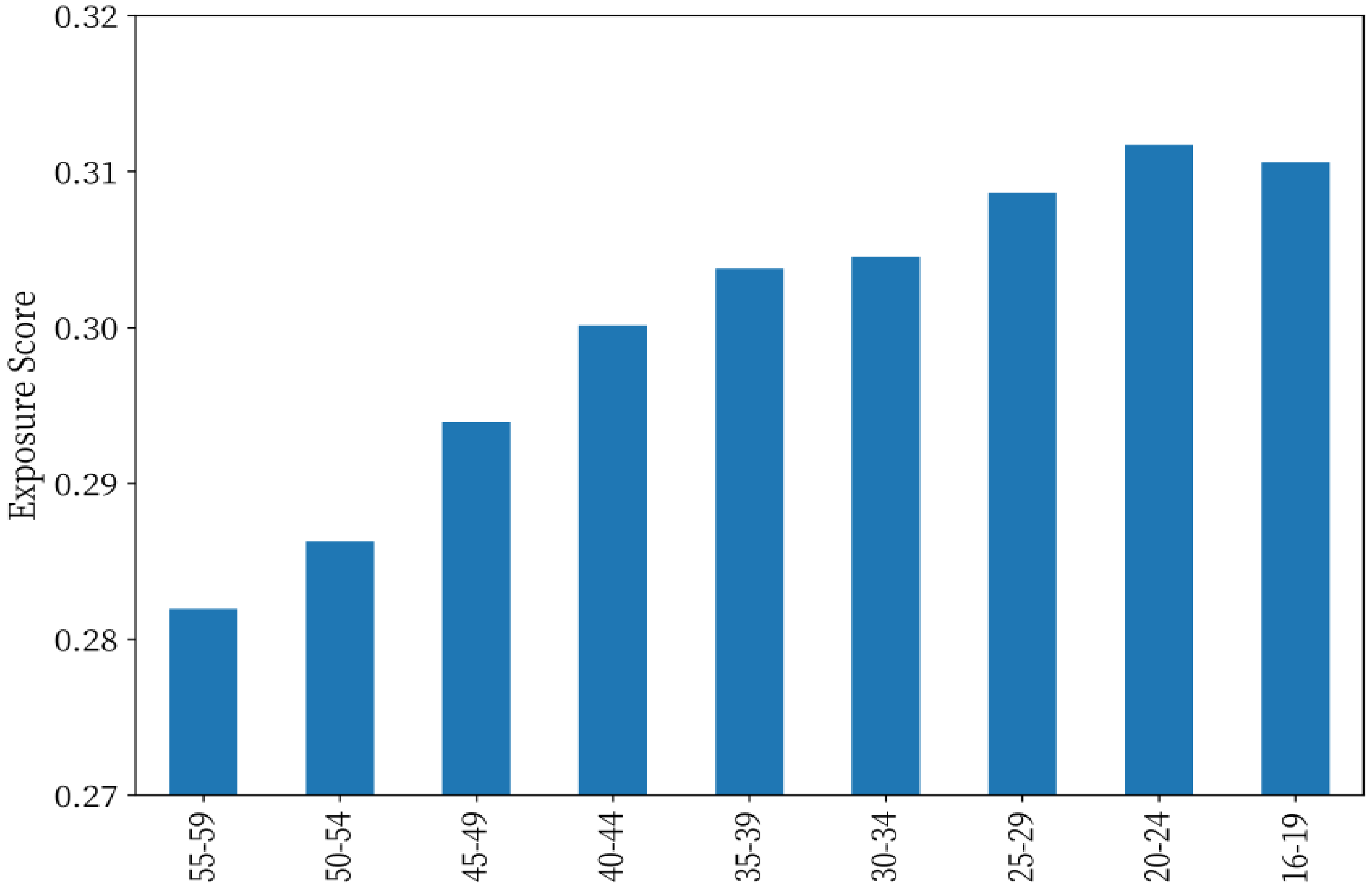
Occupational skills and LLMs exposure



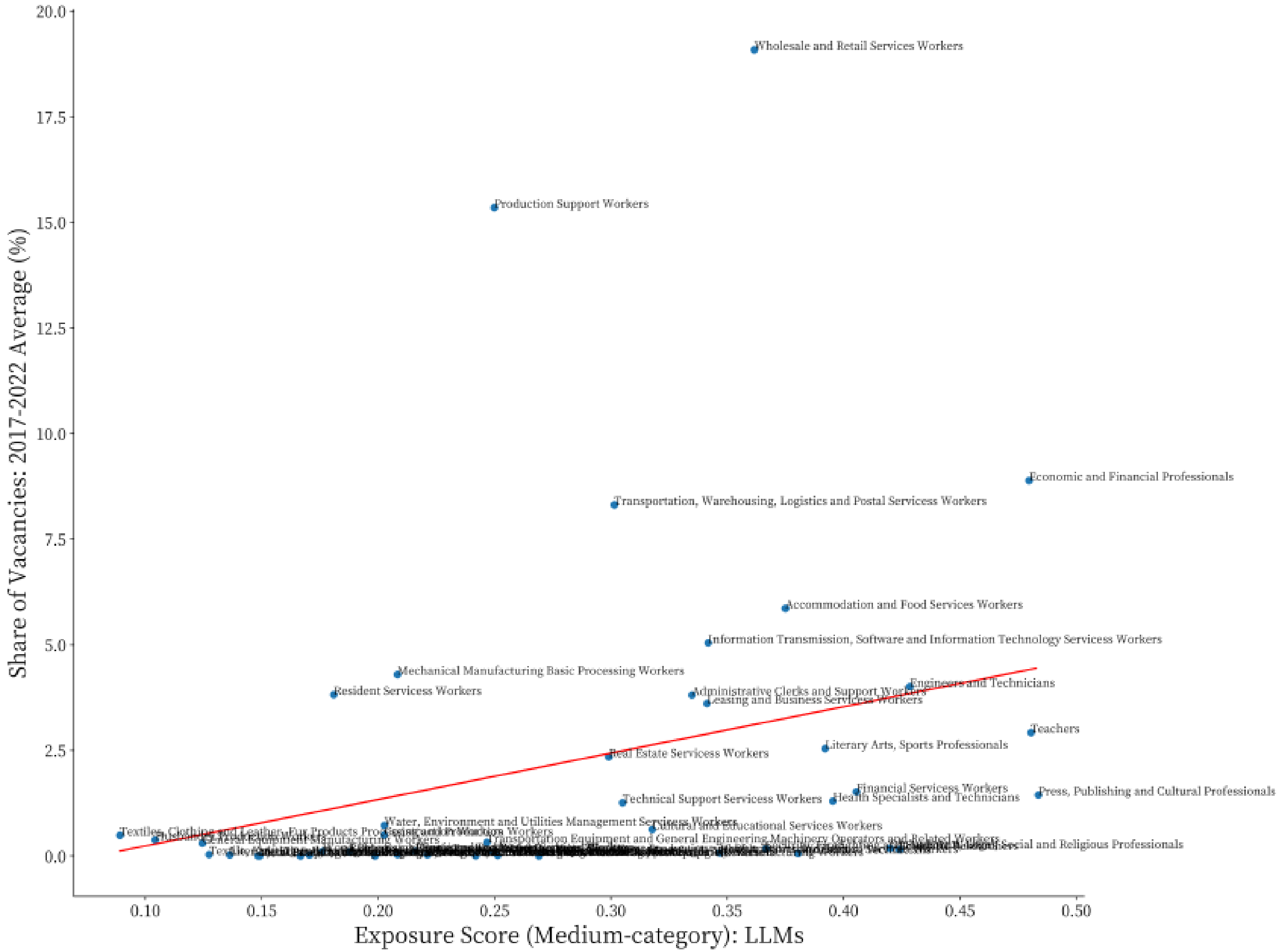
Industrial Exposure Score: LLMs



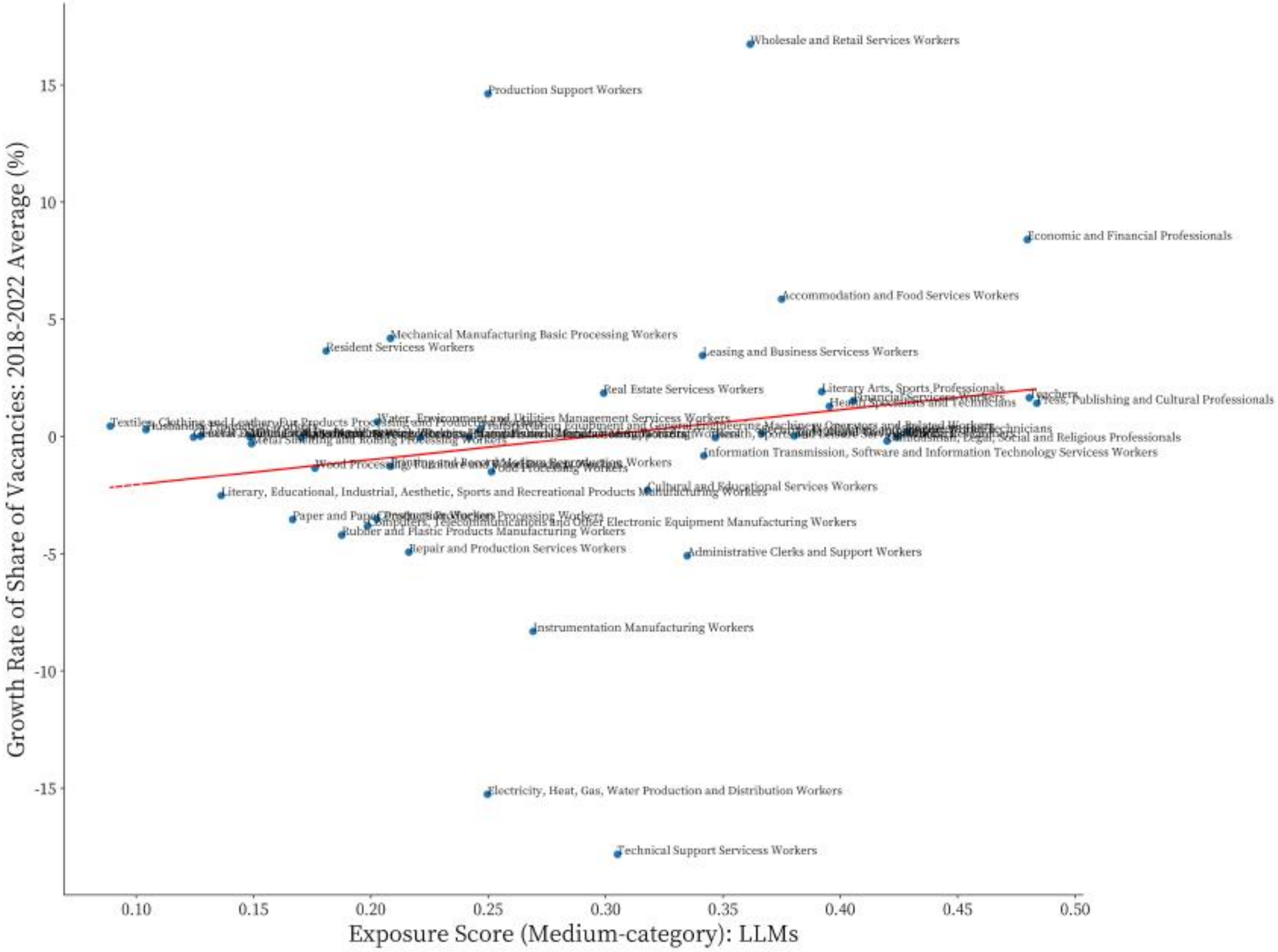
Demographic Exposure Score: LLMs



Share of Vacancies and Exposure Score: LLMs



Growth Rate of Share of Vacancies and Exposure Score



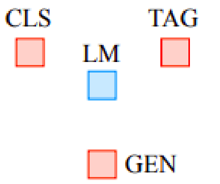
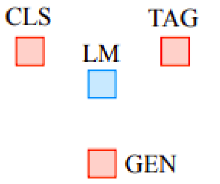
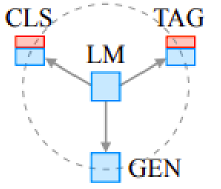
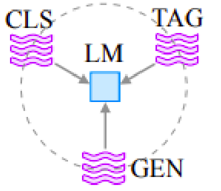
NLP, LLM and Text Analysis

- NLP I
  - ➢ TFIDF, Topic model
- NLP II
  - ➢ Word2Vec
- NLP III
  - ➢ BERT
- NLP IV
  - ➢ GPT, LLMs

✓ Representation/Embedding

✓ Probability

✓ Output Text/Analysis

Paradigm	**Engineering**	**Task Relation**
a. Fully Supervised Learning (Non-Neural Network)	Features (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	

LLM and Text Analysis

AER: Insights 2021, 3(3): 303–320
https://doi.org/10.1257/aeri.20190499

Measuring Technological Innovation over the Long Run†

By Bryan Kelly, Dimitris Papanikolaou, Amit Seru, and Matt Taddy*

We use textual analysis of high-dimensional data from patent documents to create new indicators of technological innovation. We identify important patents based on textual similarity of a given patent to previous and subsequent work: these patents are distinct from previous work but related to subsequent innovations. Our importance indicators correlate with existing measures of patent quality but also provide complementary information. We identify breakthrough innovations as the most important patents—those in the right tail of our measure—and construct time series indices of technological change at the aggregate and sectoral levels. Our technology indices capture the evolution of technological waves over a long time span (1840 to the present) and cover innovation by private and public firms as well as nonprofit organizations and the US government. Advances in electricity and transportation drive the index in the 1880s, chemicals and electricity in the 1920s and 1930s, and computers and communication in the post-1980s. (*JEL* C43, N71, N72, O31, O33, O34)

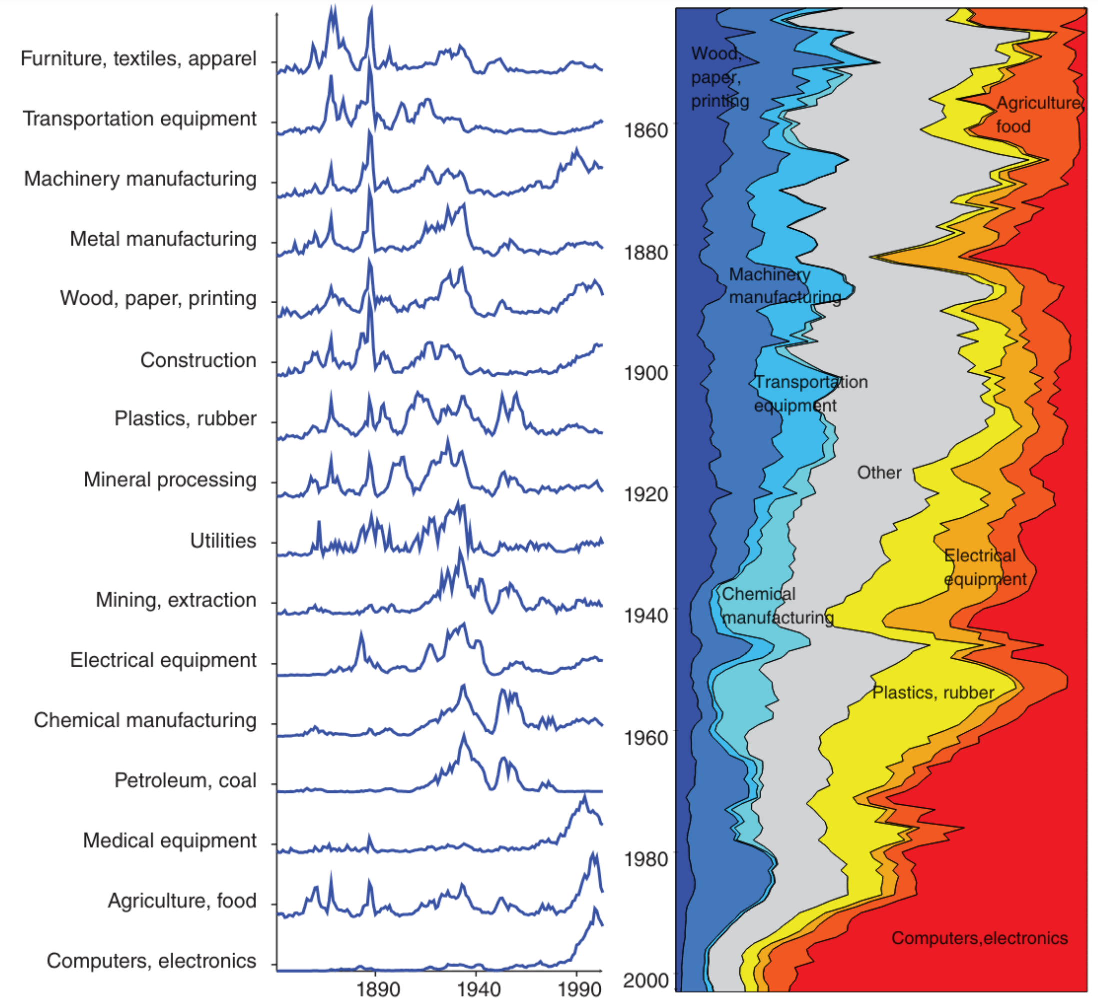


Figure 5. Breakthrough Innovation across Industries

LLM and Text Analysis

The Diffusion of New Technologies

Aakash Kalyani, Nicholas Bloom, Marcela Carvalho, Tarek Hassan,

Josh Lerner, and Ahmed Tahoun[1]

October 30, 2023

Abstract: We identify phrases associated with novel technologies using textual analysis of patents, job postings, and earnings calls, enabling us to identify four stylized facts on the diffusion of jobs relating to new technologies. First, the development of new technologies is geographically highly concentrated, more so even than overall patenting: 56% of the economically most impactful technologies come from just two U.S. locations, Silicon Valley and the Northeast Corridor. Second, as the technologies mature and the number of related jobs grows, hiring spreads geographically. But this process is very slow, taking around 50 years to disperse fully. Third, while initial hiring in new technologies is highly skill biased, over time the mean skill level in new positions declines, drawing in an increasing number of lower-skilled workers. Finally, the geographic spread of hiring is slowest for higher-skilled positions, with the locations where new technologies were pioneered remaining the focus for the technology's high-skill jobs for decades.

Keywords: Employment, Geography, Innovation, R&D

JEL Classification: O31, O32

LLM and Text Analysis

Recovering Overlooked Information in Categorical Variables with LLMs: An Application to Labor Market Mismatch
Yi Chen, Hanming Fang, Yi Zhao, and Zibo Zhao
NBER Working Paper No. 32327
April 2024
JEL No. C55,J16,J24,J31

ABSTRACT

Categorical variables have no intrinsic ordering, and researchers often adopt a fixed-effect (FE) approach in empirical analysis. However, this approach has two significant limitations: it overlooks textual labels associated with the categorical variables; and it produces unstable results when there are only limited observations in a category. In this paper, we propose a novel method that utilizes recent advances in large language models (LLMs) to recover overlooked information in categorical variables. We apply this method to investigate labor market mismatch. Specifically, we task LLMs with simulating the role of a human resources specialist to assess the suitability of an applicant with specific characteristics for a given job. Our main findings can be summarized in three parts. First, using comprehensive administrative data from an online job posting platform, we show that our new match quality measure is positively correlated with several traditional measures in the literature, and at the same time, we highlight the LLM's capability to provide additional information conditional on the traditional measures. Second, we demonstrate the broad applicability of the new method with a survey data containing significantly less information than the administrative data, which makes it impossible to compute most of the traditional match quality measures. Our LLM measure successfully replicates most of the salient patterns observed in a hard-to-access administrative dataset using easily accessible survey data. Third, we investigate the gender gap in match quality and explore whether there exists gender stereotypes in the hiring process. We simulate an audit study, examining whether revealing gender information to LLMs influences their assessment. We show that when gender information is disclosed to the GPT, the model deems females better suited for traditionally female-dominated roles.

AI for New Measurement:
Unstructured Data (Other than Text)

AI and Creativity

- Ever since industrial revolution, there's debate about whether machines "replace" or "augment" human labor
- Many existing literature studies the impact of automation on job creation/destruction
  - For example, a lot of papers study the impact of "robots", which perform repetitive tasks for humans, on jobs and corporate finance.
- But one observation is that AI is affecting a lot of "creative" jobs, which are not "repetitive" in nature. why and how?
  - Chess/GO
  - Art design/movies
- AI change how creative jobs (industries) are organized; and it also change creativity of humans

Example: Self-Play + RL has changed the industry of Go and Chess

"It made me question human creativity. When I saw AlphaGo's moves, I wondered whether the Go moves I ha[d] known were the right ones. Its style was different, and it was such an unusual experience that it took time for me to adjust. AlphaGo made me realize that I must study Go more." (1)

- Sedol Lee, a former world Go champion

Go AI and Human Learning

PNAS

RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Check for updates

Superhuman artificial intelligence can improve human decision-making by increasing novelty

Minkyu Shin[a,1], Jin Kim[b,c,1], Bas van Opheusden[d,1], and Thomas L. Griffiths[d,e]

Edited by Michael Gazzaniga, University of California Santa Barbara College of Letters and Science, Santa Barbara, CA; received August 31, 2022; accepted December 19, 2022

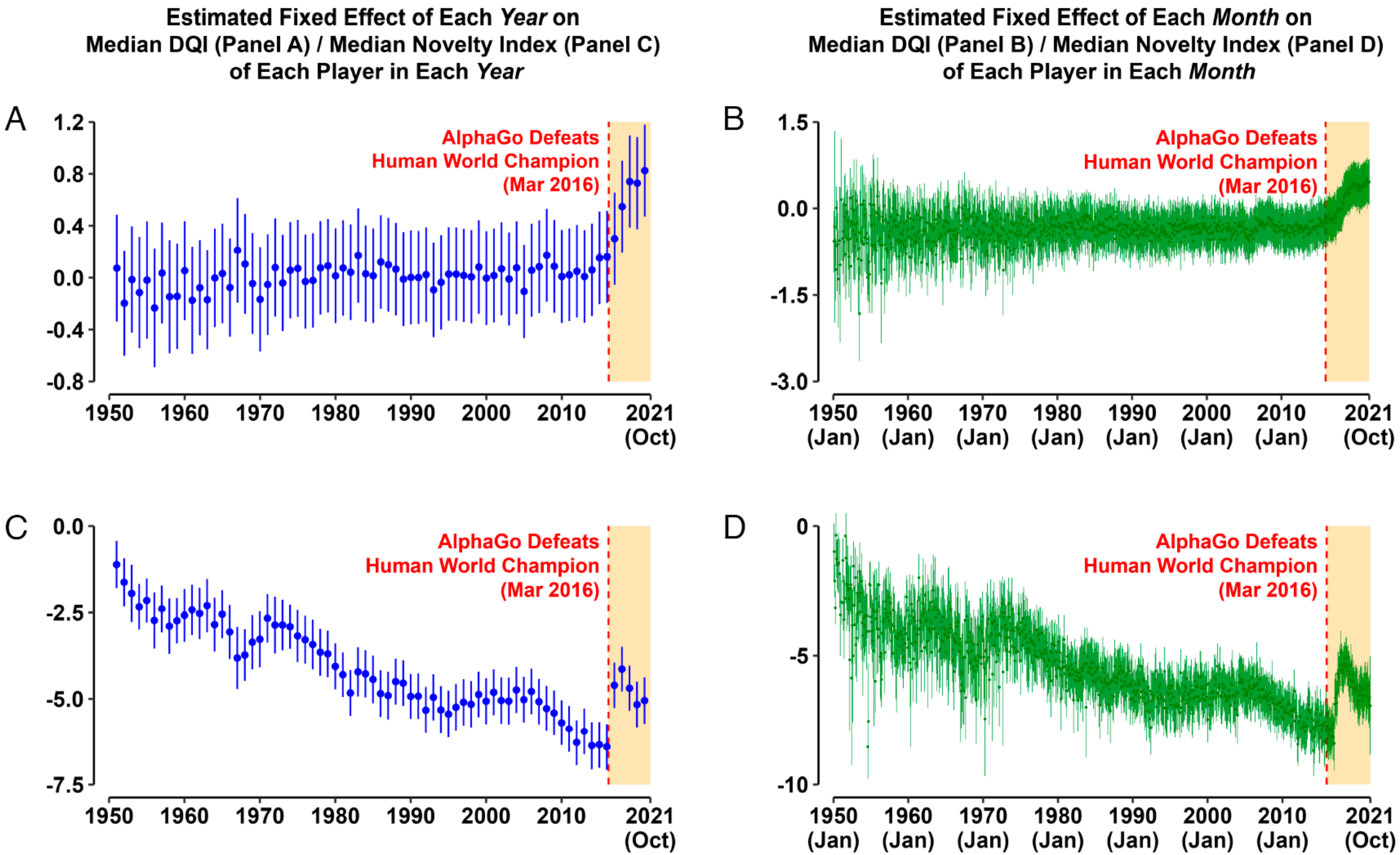
How will superhuman artificial intelligence (AI) affect human decision-making? And what will be the mechanisms behind this effect? We address these questions in a domain where AI already exceeds human performance, analyzing more than 5.8 million move decisions made by professional Go players over the past 71 y (1950 to 2021). To address the first question, we use a superhuman AI program to estimate the quality of human decisions across time, generating 58 billion counterfactual game patterns and comparing the win rates of actual human decisions with those of counterfactual AI decisions. We find that humans began to make significantly better decisions following the advent of superhuman AI. We then examine human players' strategies across time and find that novel decisions (i.e., previously unobserved moves) occurred more frequently and became associated with higher decision quality after the advent of superhuman AI. Our findings suggest that the development of superhuman AI programs may have prompted human players to break away from traditional strategies and induced them to explore novel moves, which in turn may have improved their decision-making.

judgment and decision-making | artificial intelligence | novelty | cognitive performance | innovation

Significance

Although advances in artificial intelligence (AI) created superhuman AI systems, little is understood about how such AI systems will affect human decision-making. We examine historical changes in decision-making by professional Go players over the recent seven decades, focusing on changes after the advent of superhuman AI (e.g., AlphaGo). We find that superhuman AI may have improved human

Go AI and Human Learning



AI for Simulation Modeling

AI for Simulation Modeling: Tax Policy

SCIENCE ADVANCES | RESEARCH ARTICLE

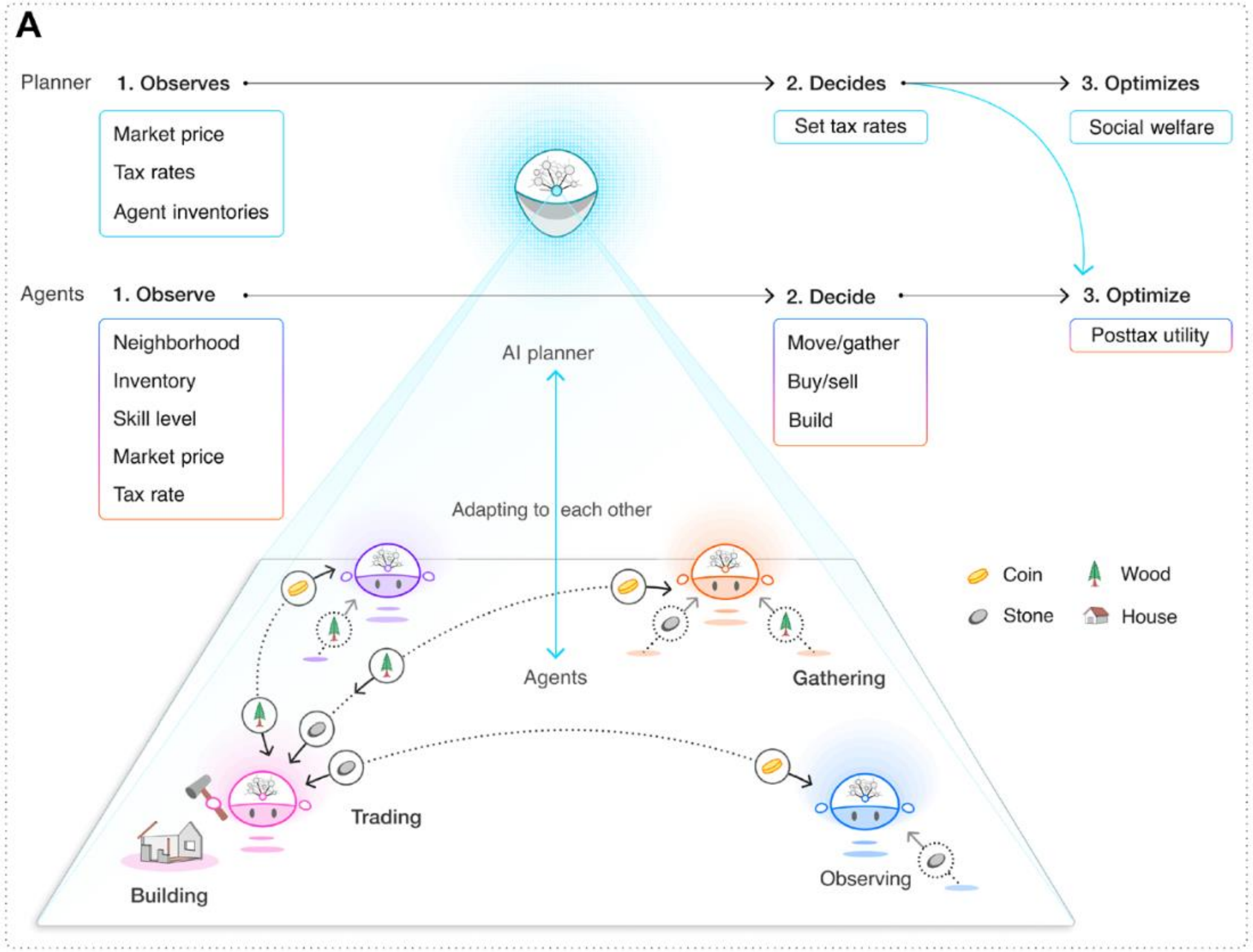
COMPUTER SCIENCE

The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning

Stephan Zheng[1]*†, Alexander Trott[1]†, Sunil Srinivasa[1], David C. Parkes[2], Richard Socher[3]

Artificial intelligence (AI) and reinforcement learning (RL) have improved many areas but are not yet widely adopted in economic policy design, mechanism design, or economics at large. The AI Economist is a two-level, deep RL framework for policy design in which agents and a social planner coadapt. In particular, the AI Economist uses structured curriculum learning to stabilize the challenging two-level, coadaptive learning problem. We validate this framework in the domain of taxation. In one-step economies, the AI Economist recovers the optimal tax policy of economic theory. In spatiotemporal economies, the AI Economist substantially improves both utilitarian social welfare and the trade-off between equality and productivity over baselines. It does so despite emergent tax-gaming strategies while accounting for emergent labor specialization, agent interactions, and behavioral change. These results demonstrate that two-level, deep RL complements economic theory and unlocks an AI-based approach to designing and understanding economic policy.

AI for Simulation Modeling: Tax Policy



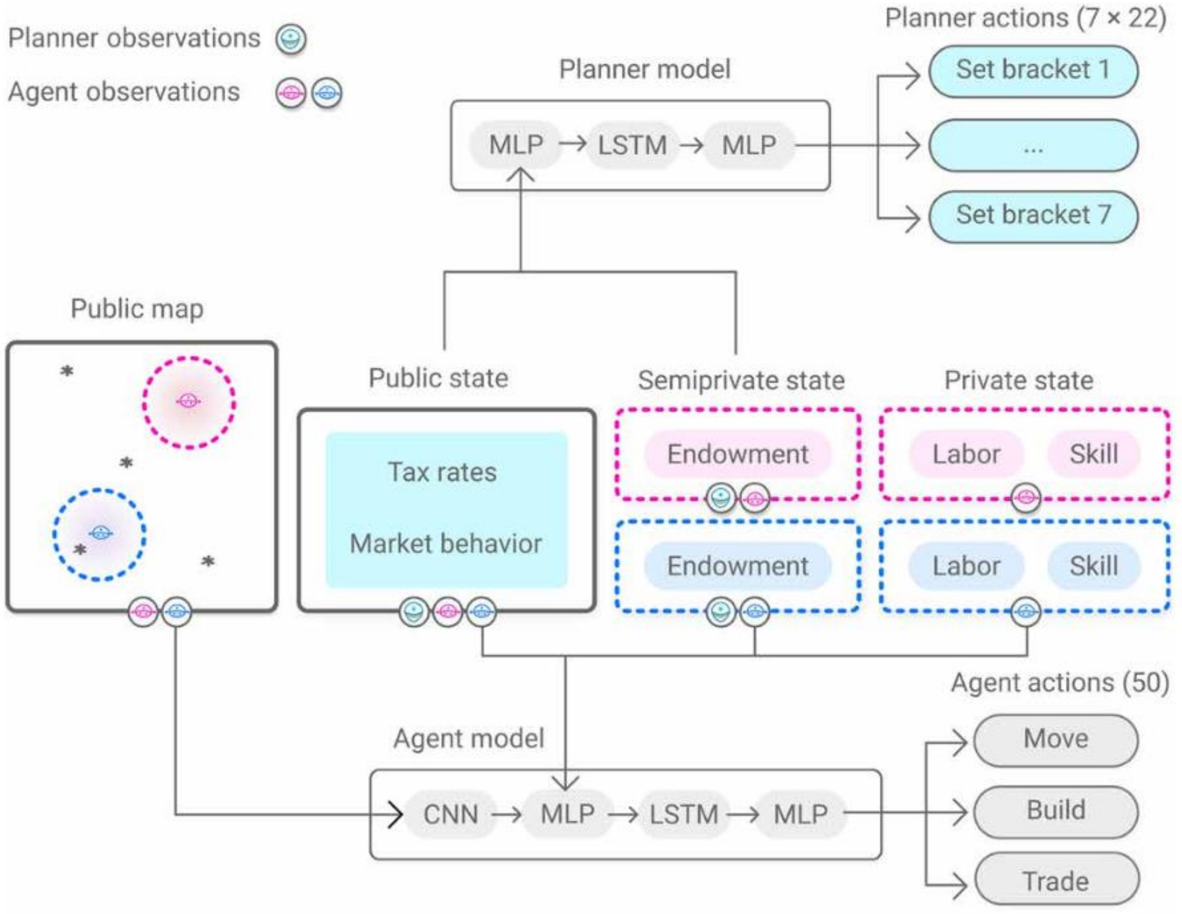


Fig. 9. Observation and action spaces for economic agents and the social planner.

AI for Simulation Modeling: Tax Policy

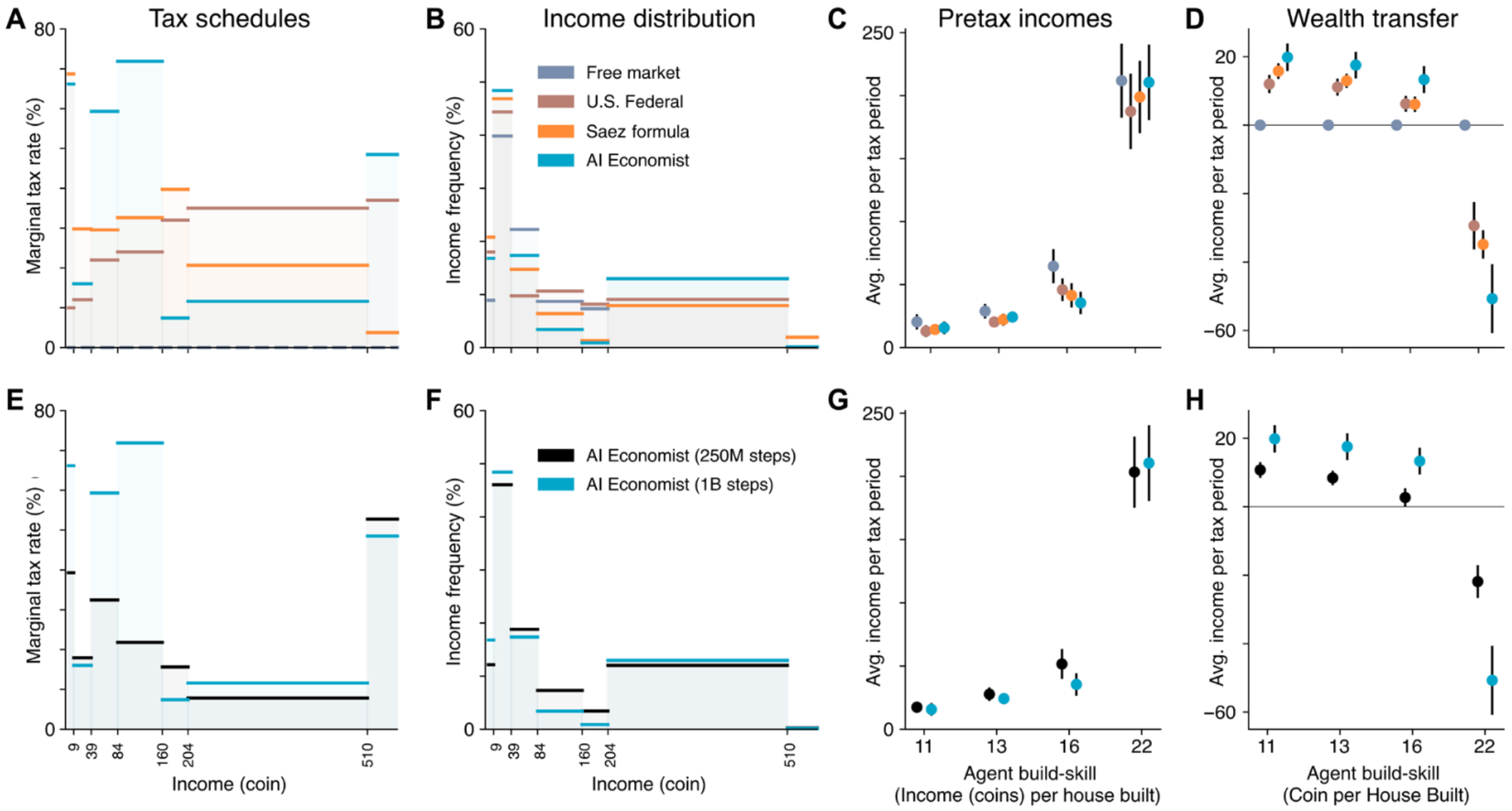


Fig. 8. Comparison of tax policies in the four-agent Open-Quadrant Gather-Trade-Build environment. (**A**) Average marginal tax rates within each tax bracket. (**B**) Frequency with which agent incomes fall within each bracket. (**C**) Average pretax income of each agent (sorted by build-skill) under each of the tax models. (**D**) Average wealth transfer resulting from taxation and redistribution. (**E** to **H**) Same as (A) to (D), comparing the AI Economist from early during training (250 million training samples) versus at the end of training (1 billion training samples). Dots denote averages and error bars denote standard deviation across episodes.

AI for Simulation Modeling: Algorithmic Behaviors

American Economic Review 2020, 110(10): 3267–3297
https://doi.org/10.1257/aer.20190623

Artificial Intelligence, Algorithmic Pricing, and Collusion

By Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello*

Increasingly, algorithms are supplanting human decision-makers in pricing goods and services. To analyze the possible consequences, we study experimentally the behavior of algorithms powered by Artificial Intelligence (Q-learning) in a workhorse oligopoly model of repeated price competition. We find that the algorithms consistently learn to charge supracompetitive prices, without communicating with one another. The high prices are sustained by collusive strategies with a finite phase of punishment followed by a gradual return to cooperation. This finding is robust to asymmetries in cost or demand, changes in the number of players, and various forms of uncertainty. (*JEL* D21, D43, D83, L12, L13)

AI for Simulation Modeling: Algorithmic Behaviors

We find, first of all, that the algorithms consistently learn to charge supracompetitive prices, obtaining a sizable extra-profit compared to the static Nash equilibrium. To quantify this extra-profit, we use the following normalized measure:

$$\Delta \equiv \frac{\bar{\pi} - \pi^N}{\pi^M - \pi^N}, \tag{9}$$

where $\bar{\pi}$ is the average per-firm profit upon convergence, π^N is the profit in the Bertrand-Nash static equilibrium, and π^M is the profit under full collusion (monopoly). Thus, $\Delta = 0$ corresponds to the competitive outcome and $\Delta = 1$ to the perfectly collusive outcome. Taking π^M as a reference point makes sense when δ is

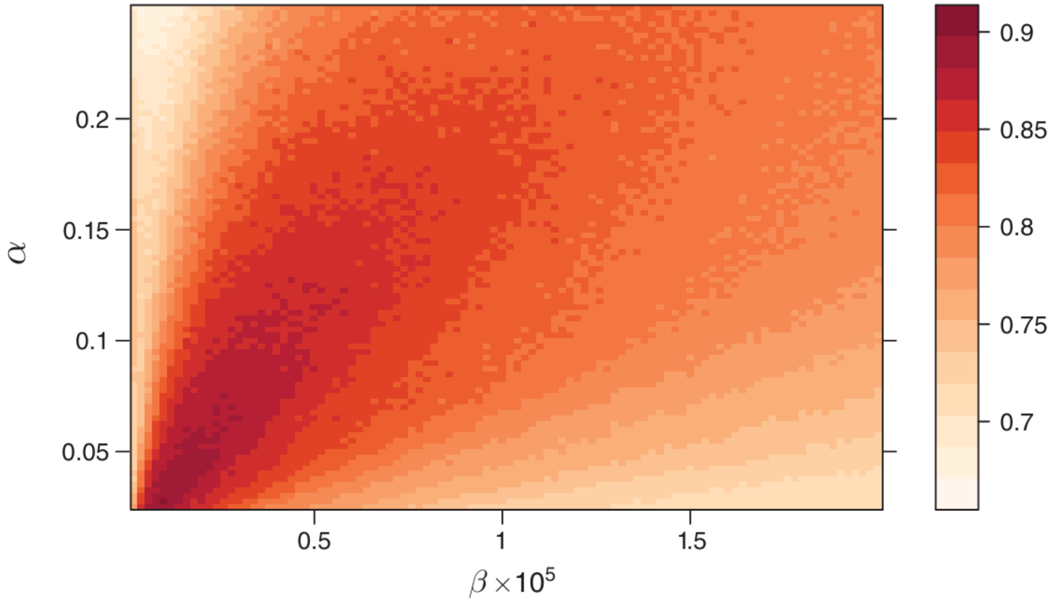


FIGURE 1. AVERAGE PROFIT GAIN Δ FOR A GRID OF VALUES OF α AND β

LLM-based Agents

Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

Meredith Ringel Morris
Google DeepMind
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu

Figure 1: Generative agents are believable simulacra of human behavior for interactive applications. In this work, we demonstrate generative agents by populating a sandbox environment, reminiscent of The Sims, with twenty-five agents. Users can observe and intervene as agents plan their days, share news, form relationships, and coordinate group activities.

AI for Hypothesis Generation

AI for Behavioral Science/Economics

Machine Learning as a Tool for Hypothesis Generation
Jens Ludwig and Sendhil Mullainathan
NBER Working Paper No. 31017
March 2023
JEL No. B4,C01

ABSTRACT

While hypothesis testing is a highly formalized activity, hypothesis generation remains largely informal. We propose a systematic procedure to generate novel hypotheses about human behavior, which uses the capacity of machine learning algorithms to notice patterns people might not. We illustrate the procedure with a concrete application: judge decisions about who to jail. We begin with a striking fact: The defendant's face alone matters greatly for the judge's jailing decision. In fact, an algorithm given only the pixels in the defendant's mugshot accounts for up to half of the predictable variation. We develop a procedure that allows human subjects to interact with this black-box algorithm to produce hypotheses about what in the face influences judge decisions. The procedure generates hypotheses that are both interpretable and novel: They are not explained by demographics (e.g. race) or existing psychology research; nor are they already known (even if tacitly) to people or even experts. Though these results are specific, our procedure is general. It provides a way to produce novel, interpretable hypotheses from any high-dimensional dataset (e.g. cell phones, satellites, online behavior, news headlines, corporate filings, and high-frequency time series). A central tenet of our paper is that hypothesis generation is in and of itself a valuable activity, and hope this encourages future work in this largely "pre-scientific" stage of science.

AI for Behavioral Science/Economics

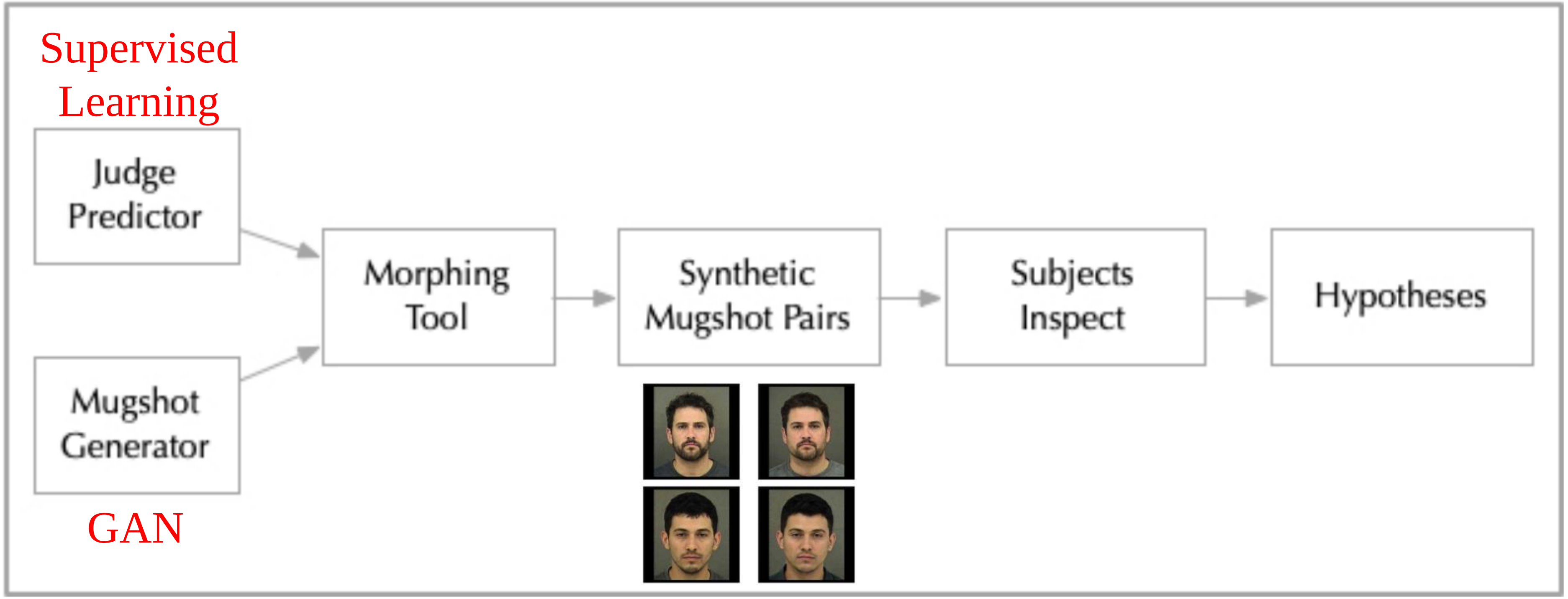


Figure IV: Hypothesis generation pipeline

Notes: The above diagram illustrates all the algorithmic components in our procedure by presenting a full pipeline for algorithmic interpretation.

Source: Ludwig, Jens, and Sendhil Mullainathan. "Machine learning as a tool for hypothesis generation." The Quarterly Journal of Economics 139.2 (2024): 751-827.

AI for Behavioral Science/Economics

Image number 1.

Questions for image 1

1.

a. Please move the slider to describe how well the face matches each description, from 1 (low) to 9 (high).

Attractiveness: unattractive or unnapealing looks (low) or very attractive (high)

Competence: incompetent appearance (low) or qualified and competent (high)

Dominance: weak or timid (low) or strong and assertive (high)

Trustworthiness: dishonesty (low), or dependable and reliable (high)

Well-groomed Unkempt appearance (low) or well-groomed (high)

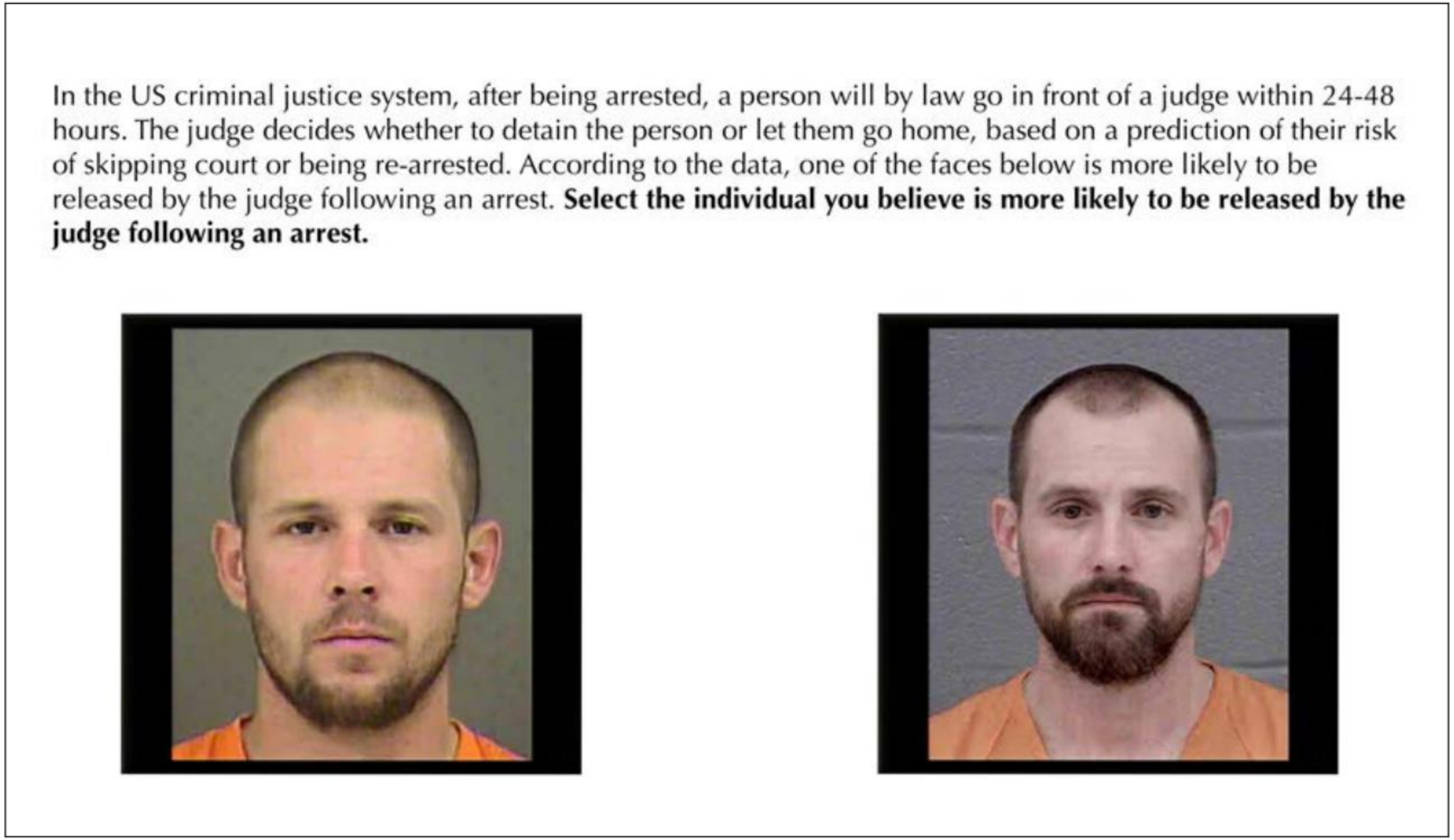
Full faced: has gaunt or lean features (low), or chubby, wide set face with broad features (high)

b. Please select the response that you feel best answers the following questions.

What race does this individual appear to be?

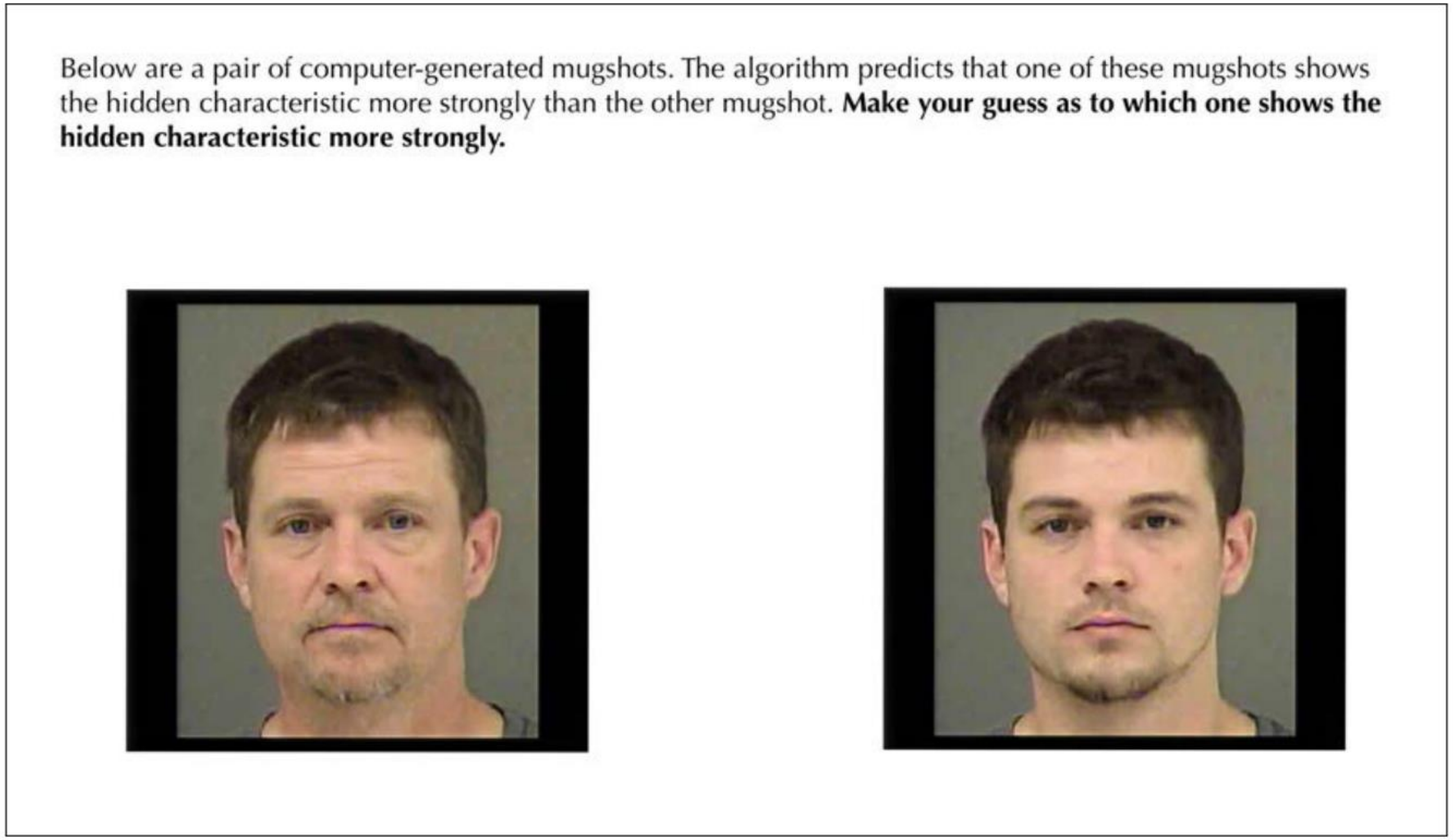
- Asian
- Black
- Caucasian / white
- Hispanic
- Indian

AI for Behavioral Science/Economics



(a) The screen presented to workers when selecting an image.

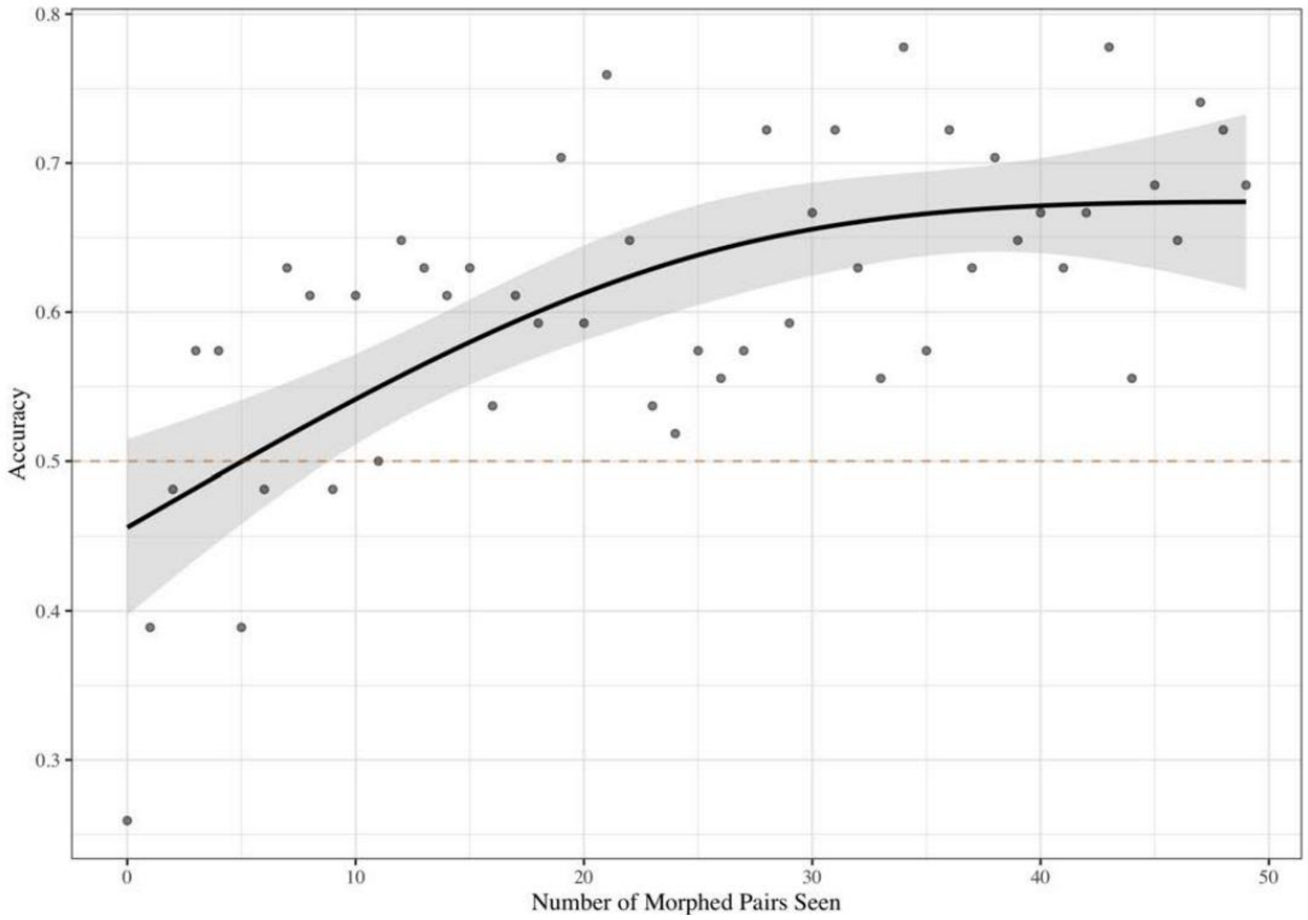
AI for Behavioral Science/Economics



(a) The screen presented to workers when selecting an image.

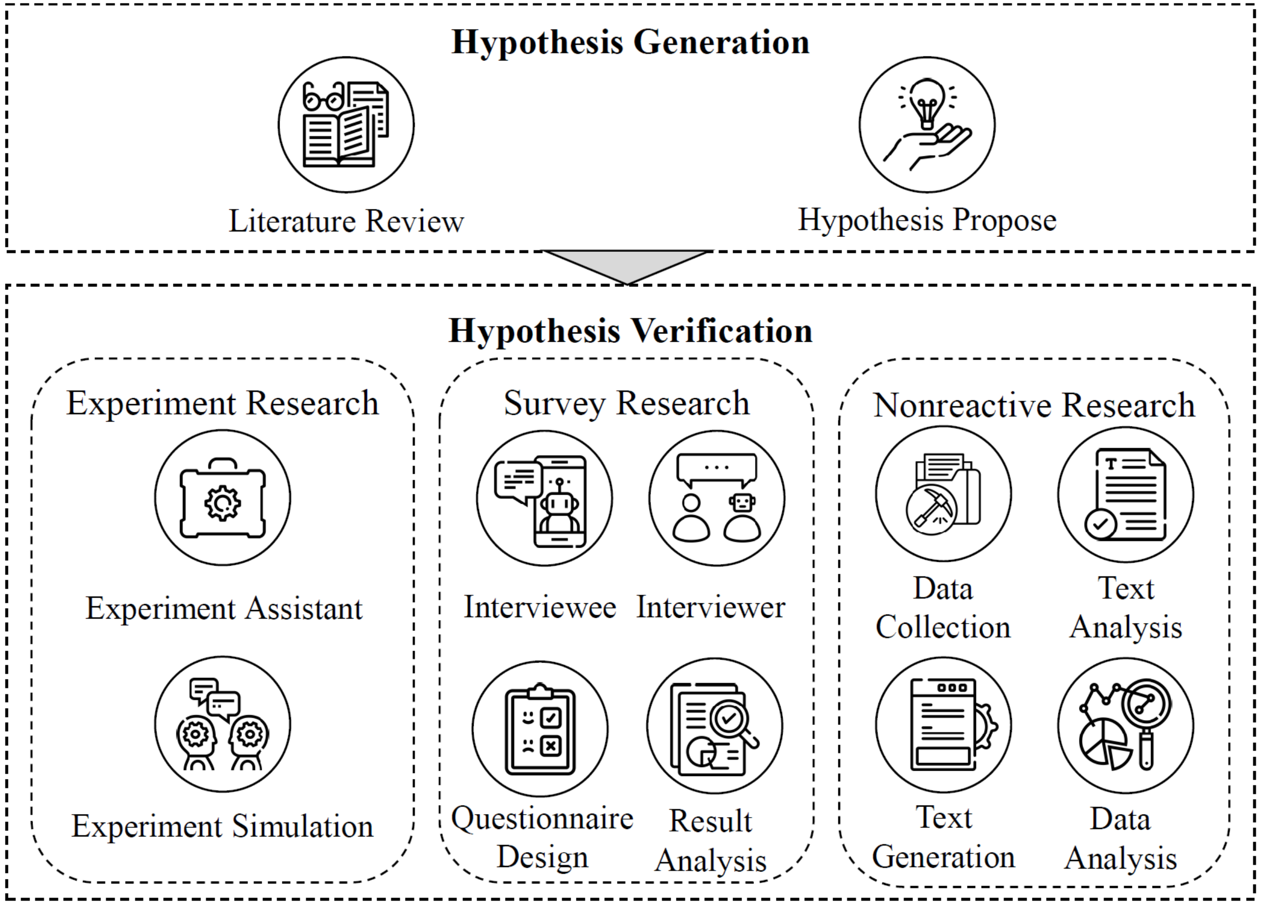
Subjects were shown age-risk-morphed image pairs and asked to make a guess about the image that exhibited that hidden characteristic more strongly. After completing this guessing exercise on 50 image pairs, subjects were asked to write down the facial features that they believed were related to the algorithm's predictions.

AI for Behavioral Science/Economics



Notes: The figure above shows subject accuracy rates in guessing which morphed image pair has a higher detention risk, and how that changes as the subjects see more images. Each subject was shown 50 image pairs matched on race, skin tone, age and gender; in our analysis, we treat the data from the first 10 images each subject sees as learning examples and carry out our analyses using the last 40 image-pair results from each subject.

AI for Social Science: Research Stages



Research Stages	Traditional Methods	Large language models
Hypothesis Generation		
Speed	Low	**High**
Validity	**High**	Low
Novelty	Low	**High**
Hypothesis Verification		
Experiment Research		
Cost	High	**Low**
Speed	Low	**High**
Reproducibility	Low	**High**
Scalability	Low	**High**
Fidelity	**Entire**	Not Sure
Survey Research		
Cost	High	**Low**
Engagement	Low	**Entire**
Interaction	Fixed	**Natural**
Bias	**Low**	Not Sure
Nonreactive Research		
Generality	Single-purpose	**Multiple-purpose**
Accessibility	Low	**High**
Numerical analysis	**Accurate**	Not Sure

Source: Xu et al. "AI for social science and social science of AI: A survey." Information Processing & Management 61.3 (2024): 103665.

AI for Economic Research: Micro Tasks

Journal of Economic Literature 2023, 61(4), 1281–1317
https://doi.org/10.1257/jel.20231736

Generative AI for Economic Research: Use Cases and Implications for Economists[†]

Anton Korinek[*]

Generative artificial intelligence (AI) has the potential to revolutionize research. I analyze how large language models (LLMs) such as ChatGPT can assist economists by describing dozens of use cases in six areas: ideation and feedback, writing, background research, data analysis, coding, and mathematical derivations. I provide general instructions and demonstrate specific examples of how to take advantage of each of these, classifying the LLM capabilities from experimental to highly useful. I argue that economists can reap significant productivity gains by taking advantage of generative AI to automate micro-tasks. Moreover, these gains will grow as the performance of AI systems continues to improve. I also speculate on the longer-term implications of AI-powered cognitive automation for economic research. The online resources associated with this paper explain how to get started and will provide regular updates on the latest capabilities of generative AI in economics. (*JEL* A11, C45, D83, I23, O33)

Category	Task	Usefulness
Ideation and Feedback	Brainstorming	●
	Feedback	◐
	Providing counterarguments	◐
Writing	Synthesizing text	●
	Editing text	●
	Evaluating text	●
	Generating catchy titles & headlines	●
	Generating tweets to promote a paper	●
Background Research	Summarizing Text	●
	Literature Research	○
	Formatting References	●
	Explaining Concepts	◐
Coding	Writing code	◐
	Explaining code	◐
	Translating code	●
	Debugging code	◐
Data Analysis	Creating figures	◐
	Extracting data from text	●
	Reformatting data	●
	Classifying and scoring text	◐
	Extracting sentiment	◐
	Simulating human subjects	◐
Math	Setting up models	◐
	Deriving equations	○
	Explaining models	◐

AI for Causal Inference

AI for Causal Inference

Estimation and Inference of Heterogeneous Treatment Effects using Random Forests*

Stefan Wager
Department of Statistics
Stanford University
swager@stanford.edu

Susan Athey
Graduate School of Business
Stanford University
athey@stanford.edu

March 6, 2017

Abstract

Many scientific and engineering challenges—ranging from personalized medicine to customized marketing recommendations—require an understanding of treatment effect heterogeneity. In this paper, we develop a non-parametric *causal forest* for estimating heterogeneous treatment effects that extends Breiman's widely used random forest algorithm. In the potential outcomes framework with unconfoundedness, we show that causal forests are pointwise consistent for the true treatment effect, and have an asymptotically Gaussian and centered sampling distribution. We also discuss a practical method for constructing asymptotic confidence intervals for the true treatment effect that are centered at the causal forest estimates. Our theoretical results rely on a generic Gaussian theory for a large family of random forest algorithms. To our knowledge, this is the first set of results that allows any type of random forest, including classification and regression forests, to be used for provably valid statistical inference. In experiments, we find causal forests to be substantially more powerful than classical methods based on nearest-neighbor matching, especially in the presence of irrelevant covariates.

AEA Papers and Proceedings 2019, 109: 65–70
https://doi.org/10.1257/pandp.20191069

APPLIED MACHINE LEARNING

Ensemble Methods for Causal Effects in Panel Data Settings†

By Susan Athey, Mohsen Bayati, Guido Imbens, and Zhaonan Qu*

Machine Learning and Causal Inference for Policy Evaluation

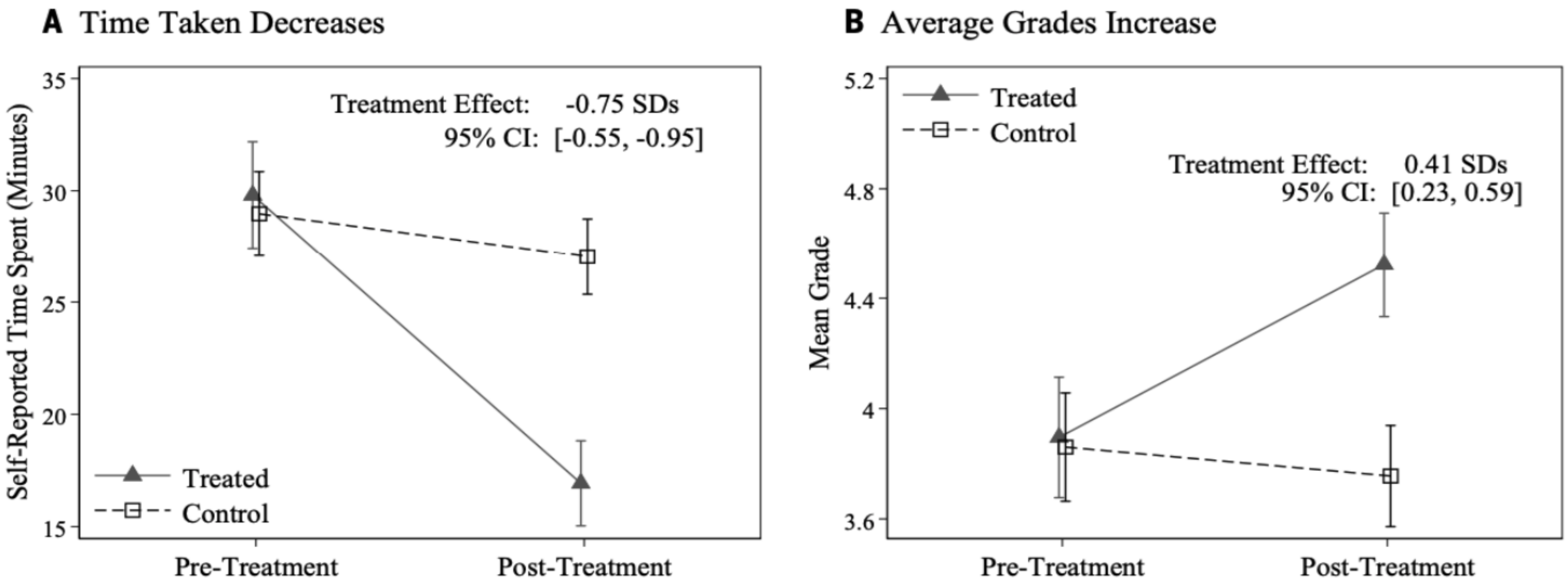
Susan Athey
Stanford Graduate School of Business
655 Knight Way
Stanford, CA 94305
1-650-725-1813
athey@stanford.edu

More: New Problems on AI and Human Interaction

ChatGPT and Task Performance: Experiment Evidence

ChatGPT drives productivity in (repetitive, boring?) writing

- A new MIT study supports popular wisdom: ChatGPT helps with writing. Specifically, for "mid-level professional writing" the study showed that, compared to a control group, workers using ChatGPT took 40% less time to complete their task, and the output quality was measured to be 18% better.



ChatGPT and Economic Rationality

RESEARCH ARTICLE | ECONOMIC SCIENCES

The emergence of economic rationality of GPT

Yiting Chen, Tracy Xiao Liu, You Shan, and Songfa Zhong Authors Info & Affiliations

Edited by Jose Scheinkman, Columbia University, New York, NY; received September 22, 2023; accepted November 13, 2023

December 12, 2023 | 120 (51) e2316205120 | https://doi.org/10.1073/pnas.2316205120

Abstract

As large language models (LLMs) like GPT become increasingly prevalent, it is essential that we assess their capabilities beyond language processing. This paper examines the economic rationality of GPT by instructing it to make budgetary decisions in four domains: risk, time, social, and food preferences. We measure economic rationality by assessing the consistency of GPT's decisions with utility maximization in classic revealed preference theory. We find that GPT's decisions are largely rational in each domain and demonstrate higher rationality score than those of human subjects in a parallel experiment and in the literature. Moreover, the estimated preference parameters of GPT are slightly different from human subjects and exhibit a lower degree of heterogeneity. We also find that the rationality scores are robust to the degree of randomness and demographic settings such as age and gender but are sensitive to contexts based on the language frames of the choice situations. These results suggest the potential of LLMs to make good decisions and the need to further understand their capabilities, limitations, and underlying mechanisms.

ChatGPT and AI Creativity

scientific reports

Check for updates

OPEN

The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks

Kent F. Hubert[1,2], Kim N. Awa[1,2] & Darya L. Zabelina[1]

The emergence of publicly accessible artificial intelligence (AI) large language models such as ChatGPT has given rise to global conversations on the implications of AI capabilities. Emergent research on AI has challenged the assumption that creative potential is a uniquely human trait thus, there seems to be a disconnect between human perception versus what AI is objectively capable of creating. Here, we aimed to assess the creative potential of humans in comparison to AI. In the present study, human participants (N = 151) and GPT-4 provided responses for the Alternative Uses Task, Consequences Task, and Divergent Associations Task. We found that AI was robustly more creative along each divergent thinking measurement in comparison to the human counterparts. Specifically, when controlling for fluency of responses, AI was more original and elaborate. The present findings suggest that the current state of AI language models demonstrate higher creative potential than human respondents.

ChatGPT and AI Behavior

PNAS

RESEARCH ARTICLE | ECONOMIC SCIENCES

OPEN ACCESS

Check for updates

A Turing test of whether AI chatbots are behaviorally similar to humans

Qiaozhu Mei[a,1], Yutong Xie[a], Walter Yuan[b], and Matthew O. Jackson[c,d,1]

Contributed by Matthew O. Jackson; received August 12, 2023; accepted January 4, 2024; reviewed by Ming Hsu, Juanjuan Meng, and Arno Riedl

We administer a Turing test to AI chatbots. We examine how chatbots behave in a suite of classic behavioral games that are designed to elicit characteristics such as trust, fairness, risk-aversion, cooperation, etc., as well as how they respond to a traditional Big-5 psychological survey that measures personality traits. ChatGPT-4 exhibits behavioral and personality traits that are statistically indistinguishable from a random human from tens of thousands of human subjects from more than 50 countries. Chatbots also modify their behavior based on previous experience and contexts "as if" they were learning from the interactions and change their behavior in response to different framings of the same strategic situation. Their behaviors are often distinct from average and modal human behaviors, in which case they tend to behave on the more altruistic and cooperative end of the distribution. We estimate that they act as if they are maximizing an average of their own and partner's payoffs.

Significance

As AI interacts with humans on an increasing array of tasks, it is important to understand how it behaves. Since much of AI programming is proprietary, developing methods of assessing AI by observing its behaviors is essential. We develop a Turing test to assess the behavioral

Thanks! 谢谢