

AI Trends after DeepSeek and its Economic Implications

YYQ 20250428@Soeul



Outline

- DeepSeek
- AI Frontier
- AI Economics
- Strategic Competition?

AI Hype, justified?

TOP *of* **MIND**

GEN AI: TOO MUCH SPEND, TOO LITTLE BENEFIT?

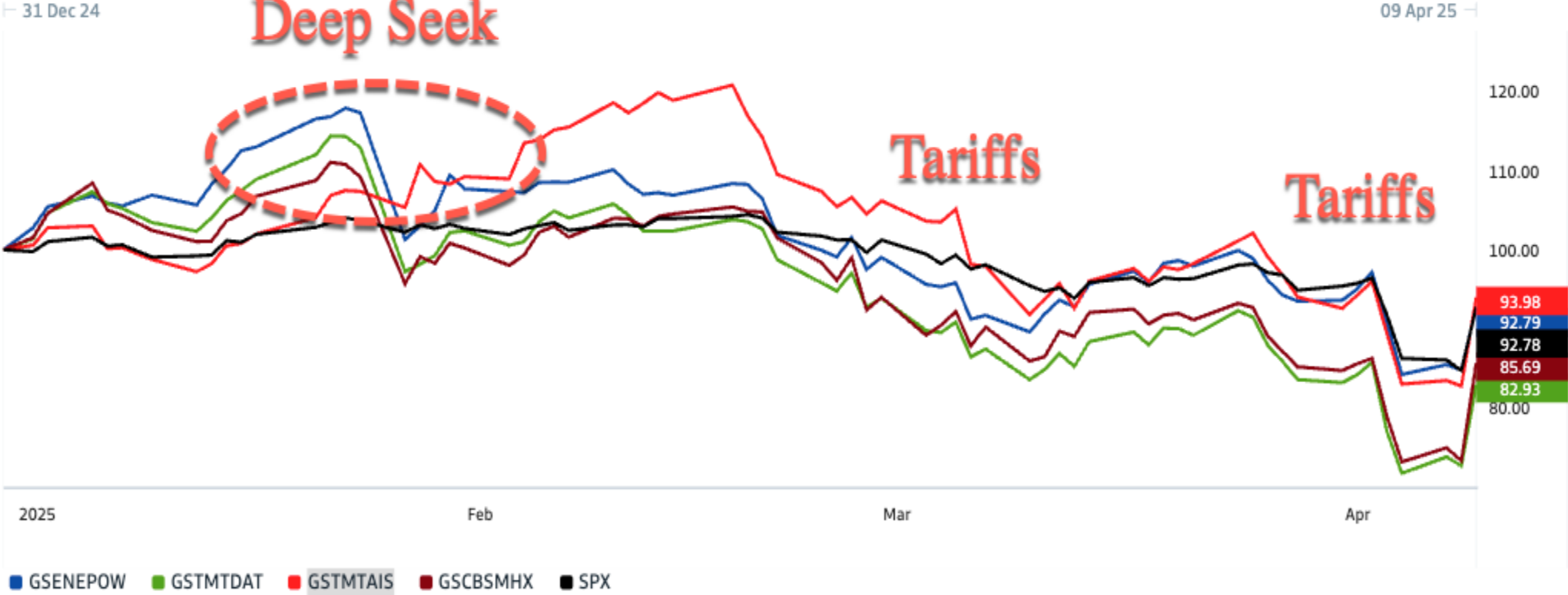
“

Given the focus and architecture of generative AI technology today... truly transformative changes won't happen quickly and few—if any—will likely occur within the next 10 years.

- Daron Acemoglu

Compute Peak?

How are Power and AI Custom Baskets performing YTD?



Scaling law

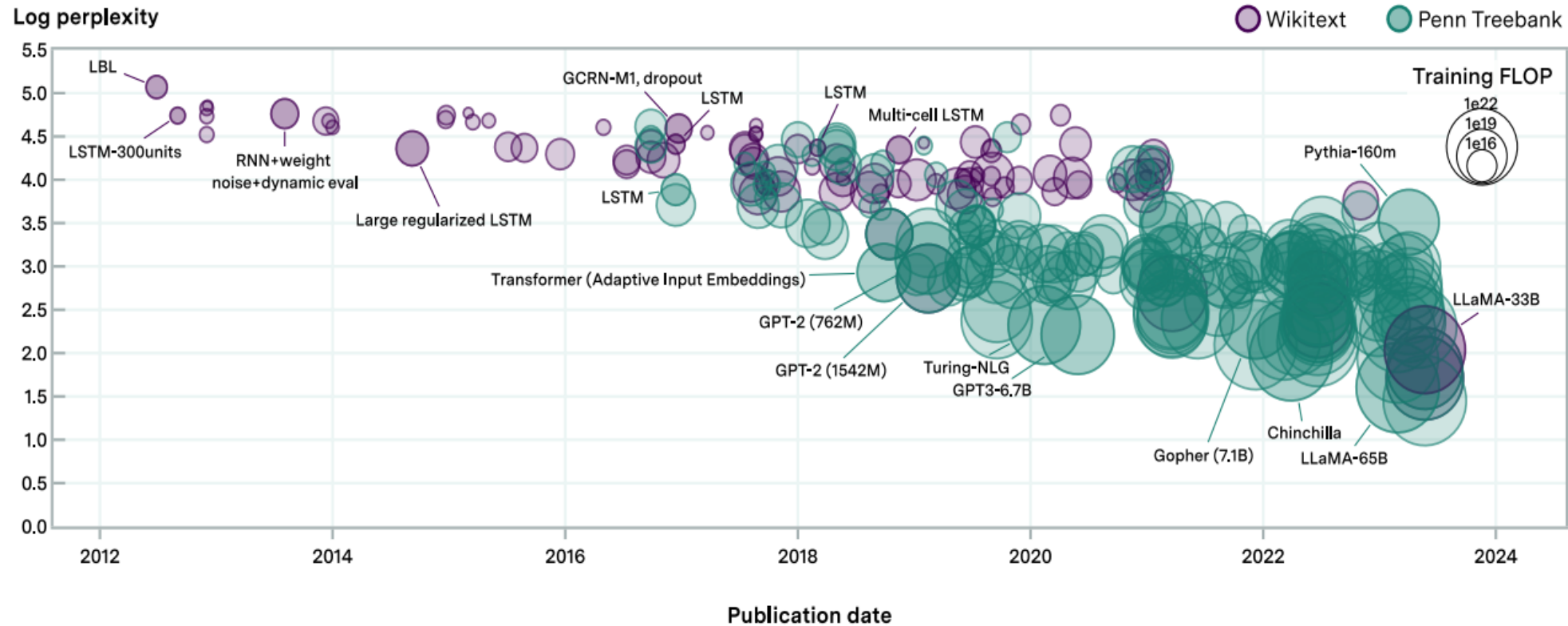
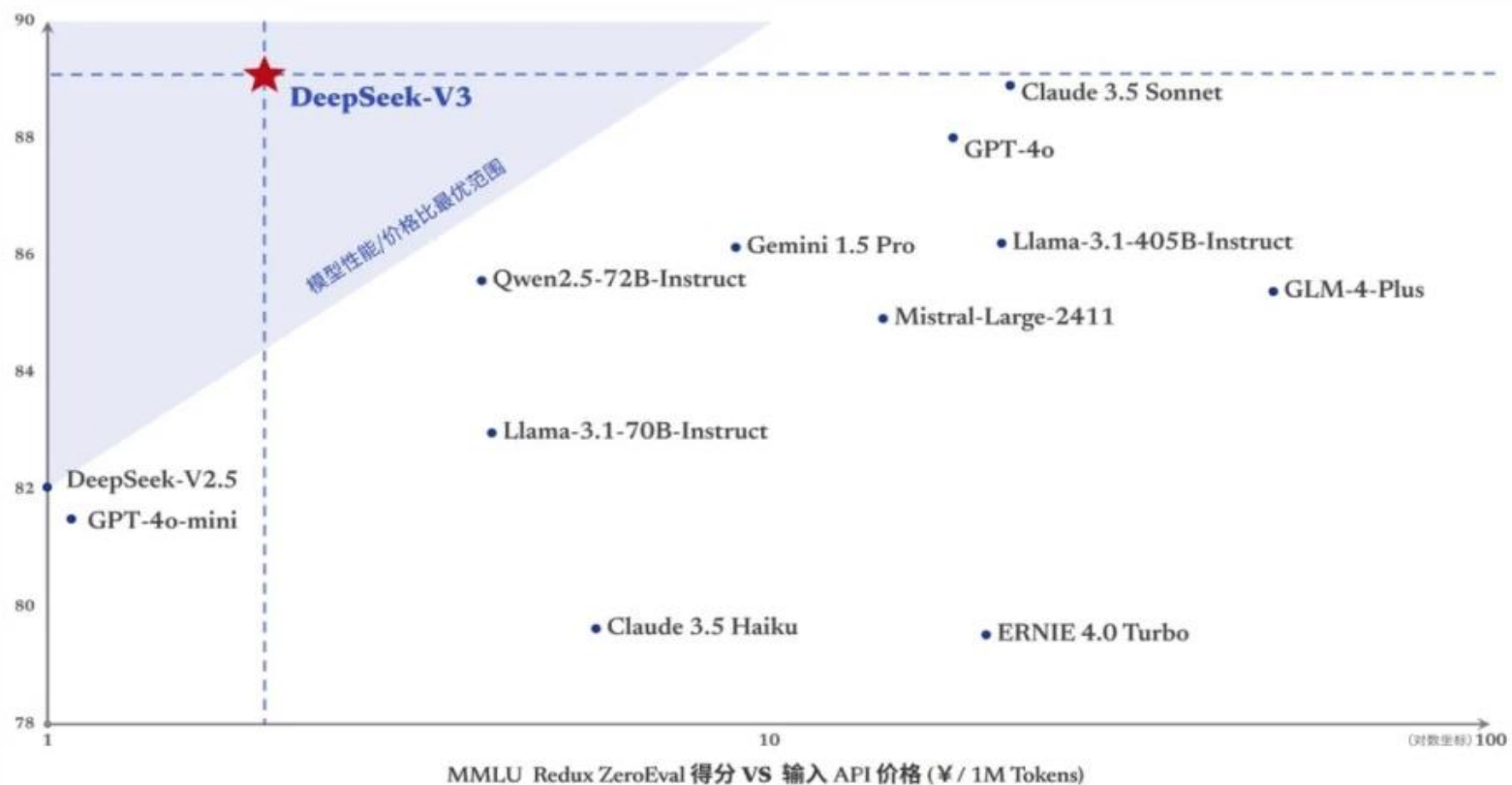


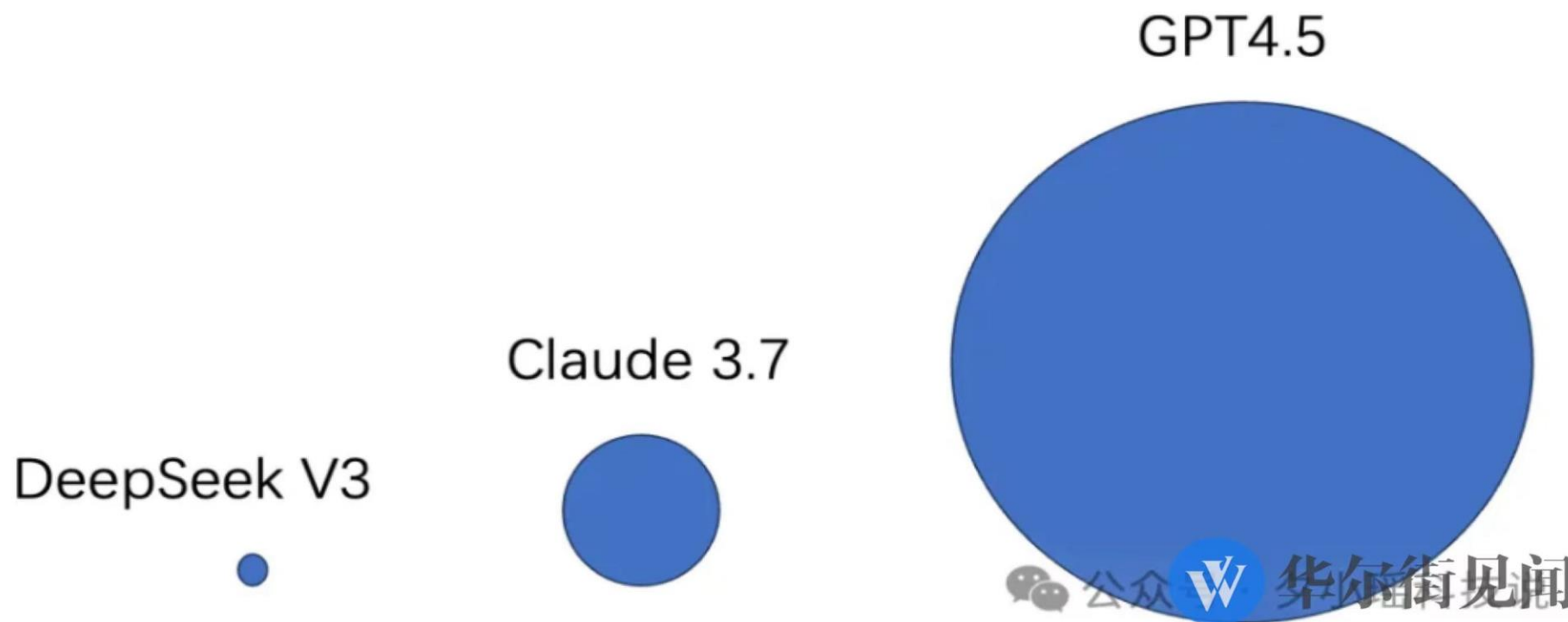
Figure 2: Log of perplexity of models used in our work, of over 231 language models analyzed in our work spanning over 8 orders of magnitude of compute, with each shape representing a model. The size of the shape is proportional to the compute used during training. Comparable perplexity evaluations are curated from the existing literature and from our own evaluations.

DeepSeek

- DeepSeek大模型成为全球开源模型领军者，DeepSeek应用登顶iOS应用下载榜第一。
- 市场对于中国大模型的认知从“等待海外开源模型”转变为“中国大模型引领开源”



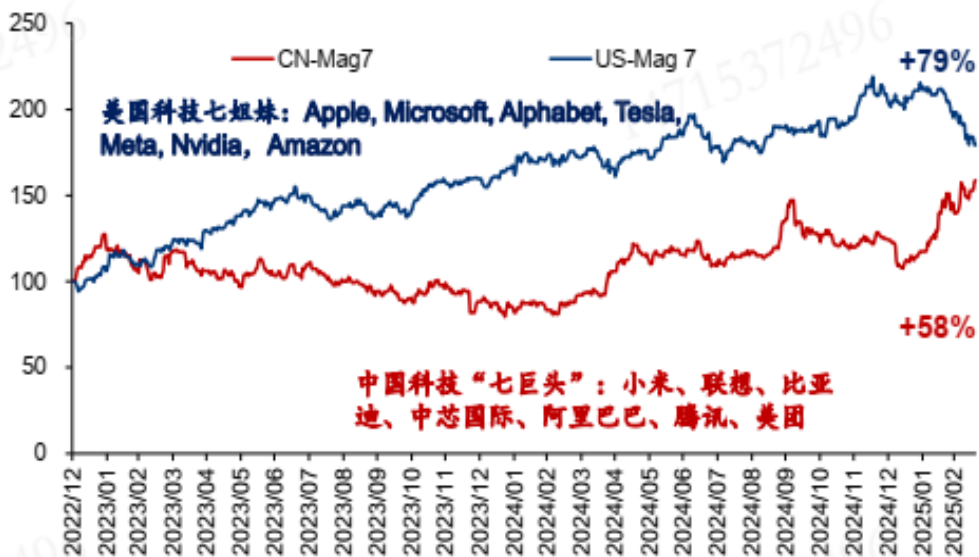
Cost: DeepSeek V3-0324



“US Mag 7” vs “China Mag 7”

图表17: 中国科技“七巨头” vs 美国科技“七姐妹”

中国科技“七巨头” vs 美国科技“七姐妹” (起点为2022/12/31, 截至2025/3/18)

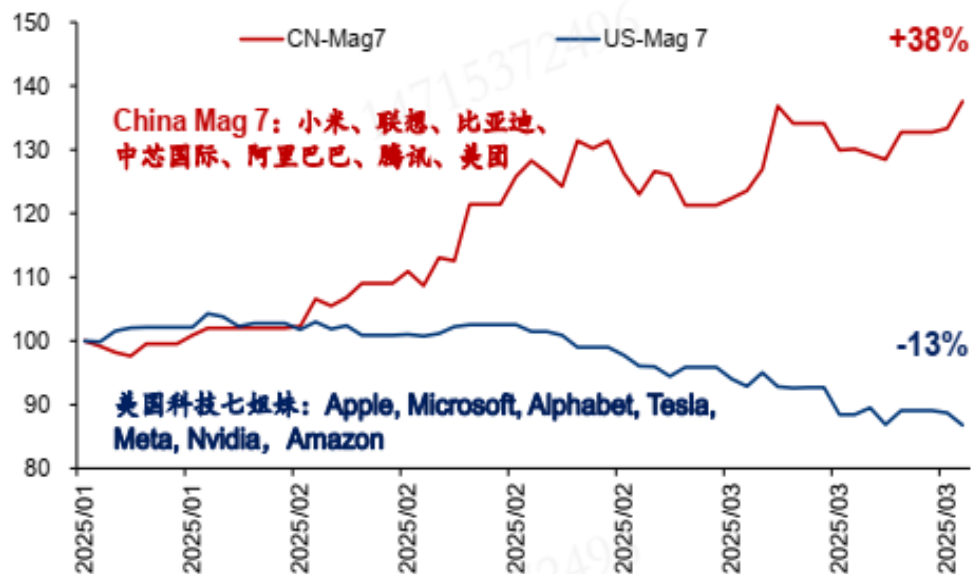


注: 股价起点为 2022/12/31, 截至 2025/3/18; 起点按 100 计算, 股价计算方法为市值加权平均

资料来源: Factset, 华泰研究

图表18: 中国科技“七巨头” vs 美国科技“七姐妹”

中国科技“七巨头” vs 美国科技“七姐妹” (起点为2025/1/20, 截至2025/3/18)



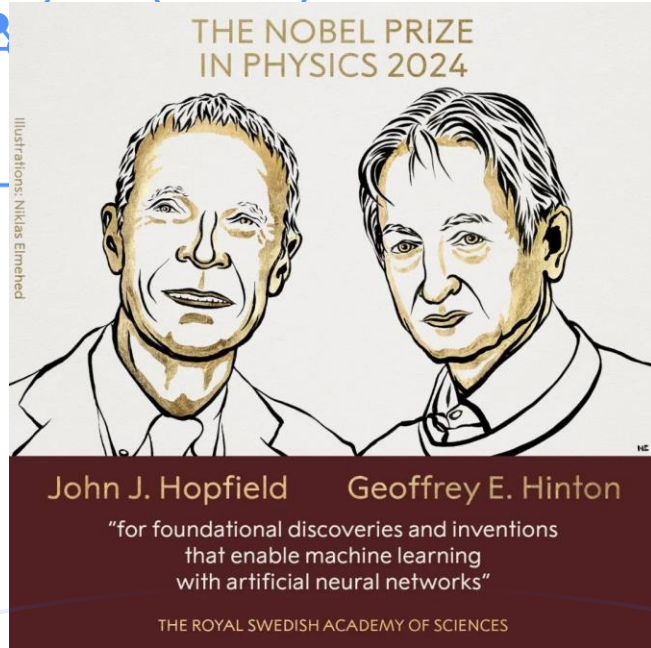
注: 股价起点为 2025/01/20, 截至 2025/3/18; 起点按 100 计算, 股价计算方法为市值加权平均

资料来源: Factset, 华泰研究

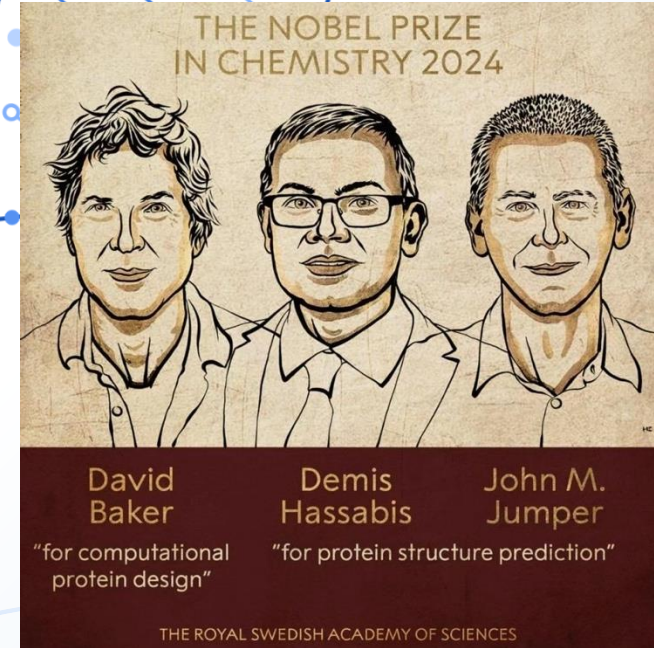
AI Frontier

AI& Science

科学和AI的 双螺旋引擎



Science for AI



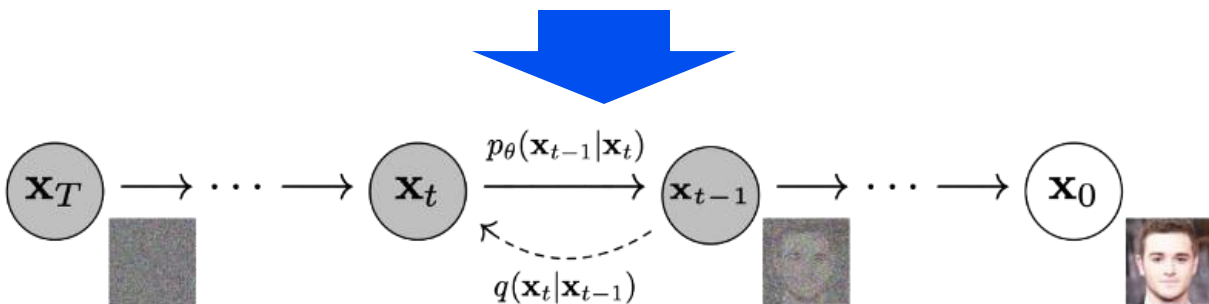
AI for Science

Science for AI: diffusion model

热力学的扩散过程启发了AI的扩散模型，引发了AIGC的热潮



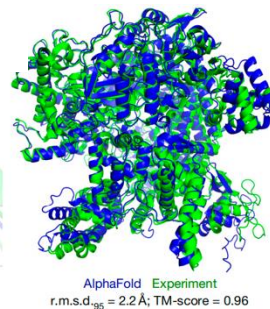
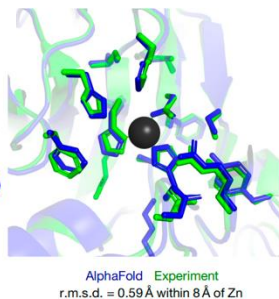
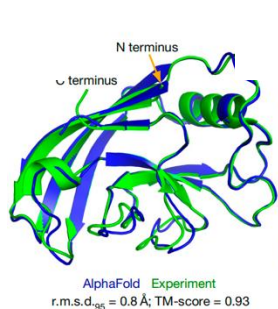
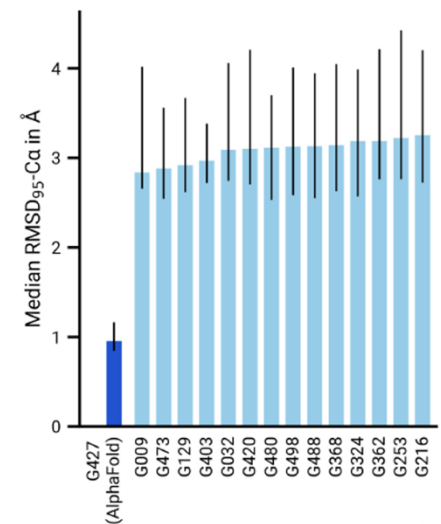
扩散模型 (Diffusion Model)



AI for Science: AlphaFold

AlphaFold 2

解决了困扰人类**半个世纪**的蛋白质结构预测难题。
原本通过湿实验需要数月甚至数年才能得到的结构，现在利用AlphaFold可以在**几分钟**之内得到，且达到实验精度

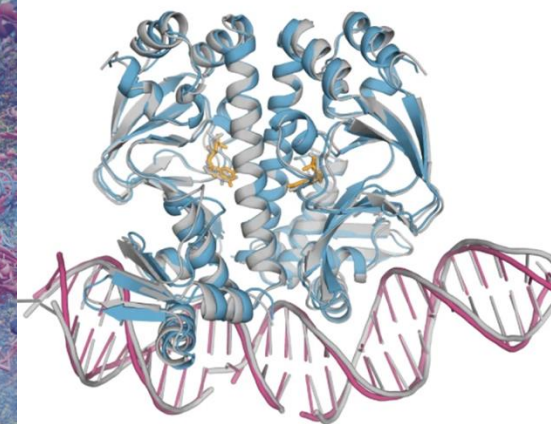


RMSD 小于 1 Å 表明预测达到了实验精度

Jumper, J. et al. Nature, 2021.

AlphaFold 3

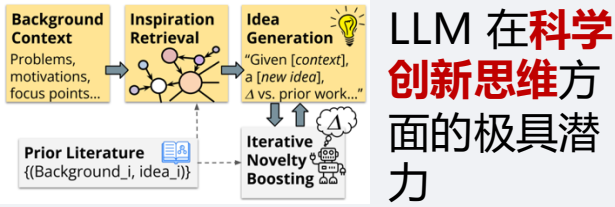
基于扩散模型架构，将蛋白质结构预测任务扩展到各类生物分子的结构和他们的**相互作用**上



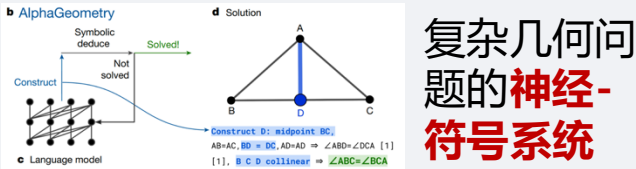
12

Abramson, J. et al. Nature, 2024.

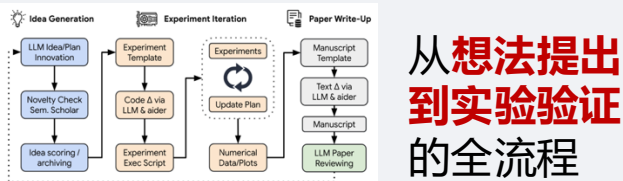
基于LLM的科学研究



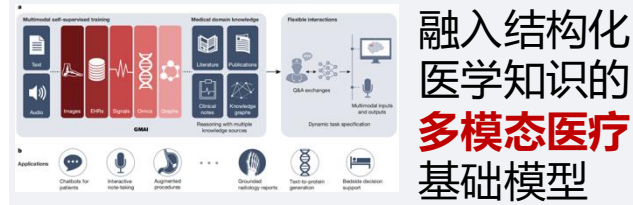
融入先验的AI模型



全流程AI科学家

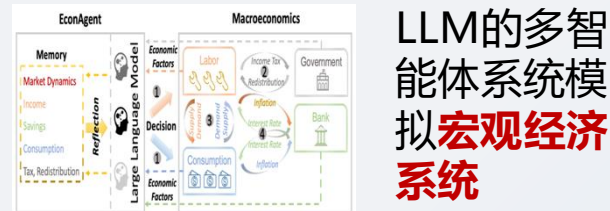


垂直领域科学大模型

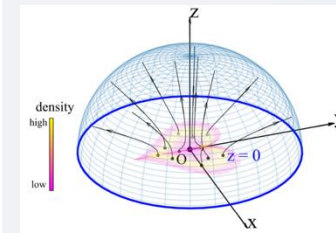


AI for Science

复杂多智能体建模



物理世界第一性原理



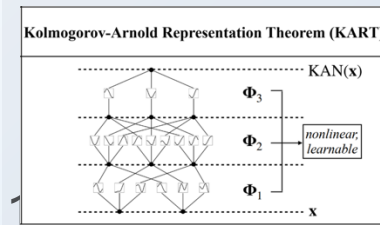
电磁启发的“泊松流”生成模型PFGM速度比扩散模型**提升10-20倍**

深度
融合

Science for AI

双向
赋能

科学启发的AI新架构



相比于传统的MLP架构, KAN更具备**可解释性**

Agentic AI

What Are AI Agents?

AI agents are poised to accelerate the adoption of digital applications and create an epochal shift in human-computer interaction.

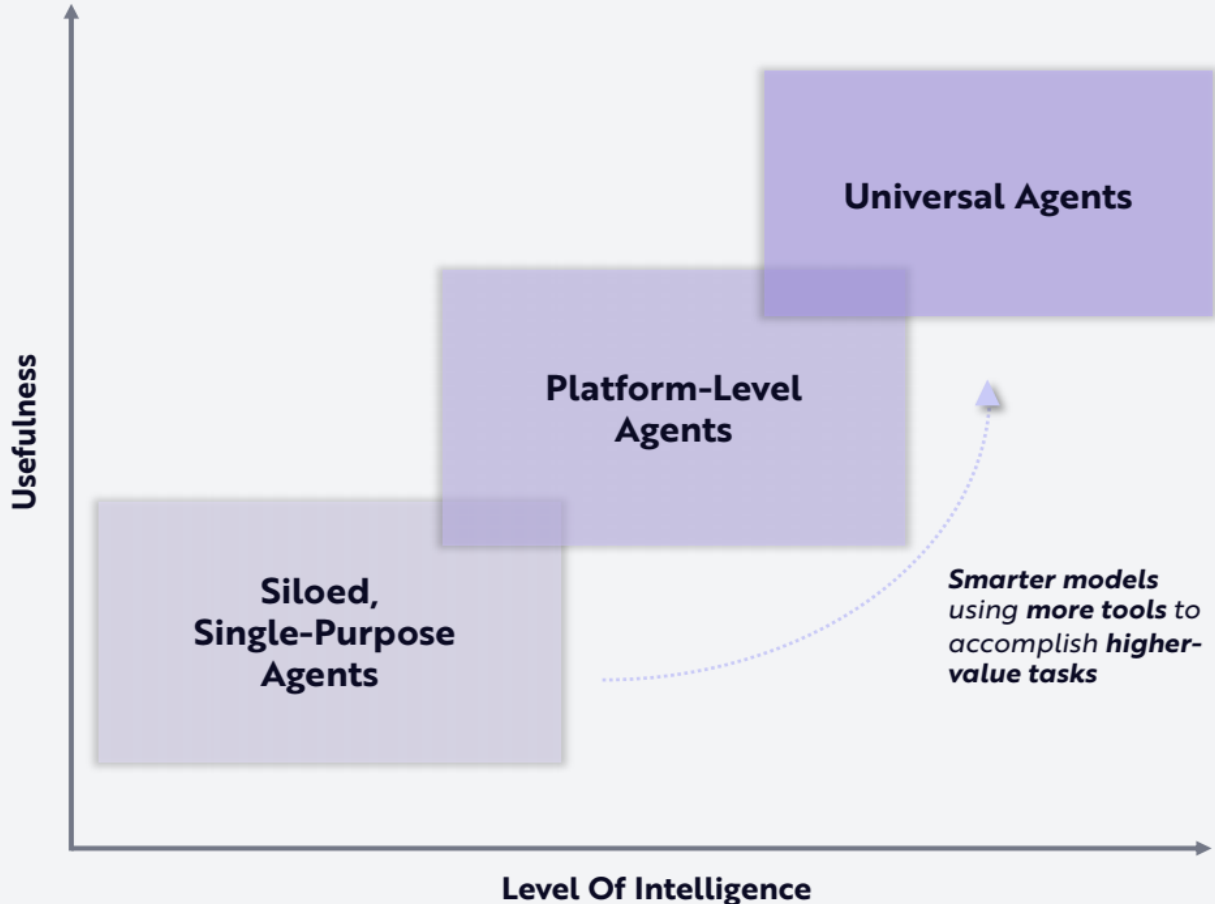
AI Agents:

Understand intent through natural language

Plan using reasoning and appropriate context

Take action using tools to accomplish the intent

Improve through iteration and continuous learning



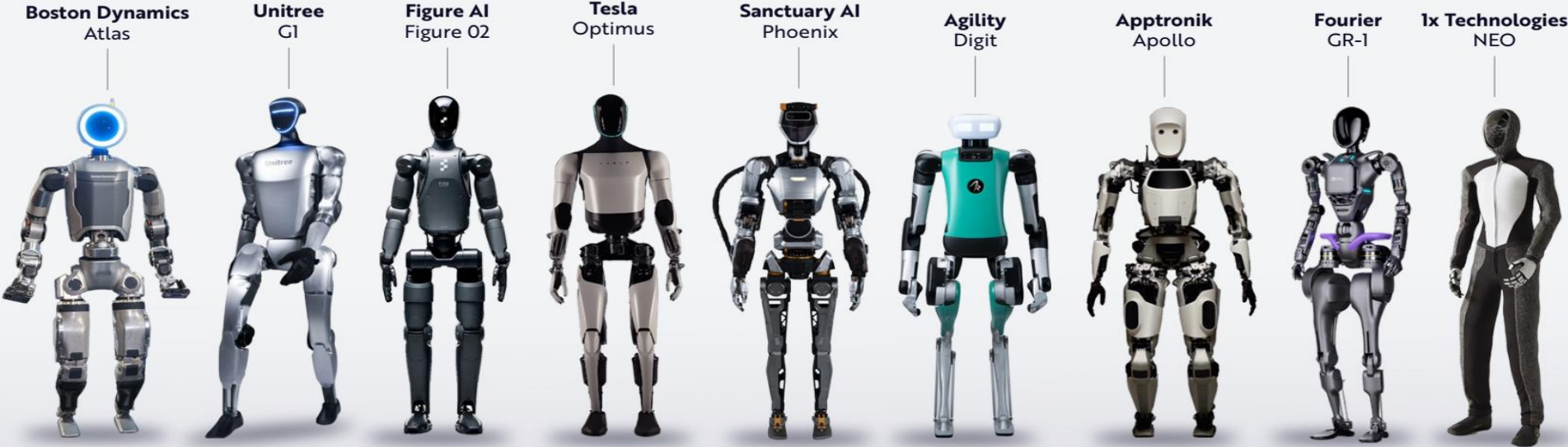
Embodied AI

ARK's Big Ideas 2025: Robotics



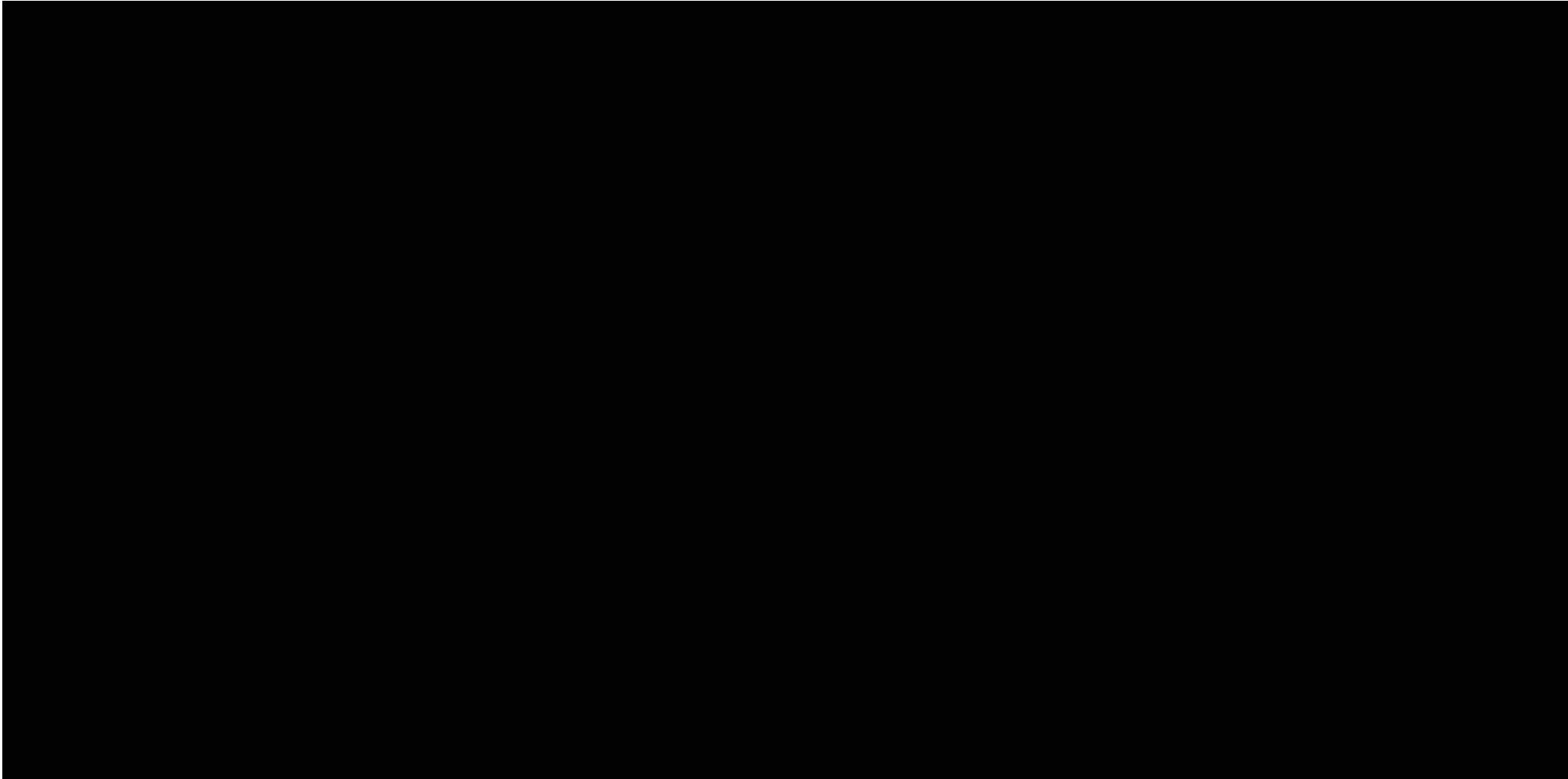
Humanoid Robots Are Debuting Around The World

Why the human form factor? Key is that a humanoid robot is generalizable. While a wrench can tighten nuts better than a human hand can, it is not a generalizable tool. The human hand is generalizable, particularly in an environment built by and designed for humans.

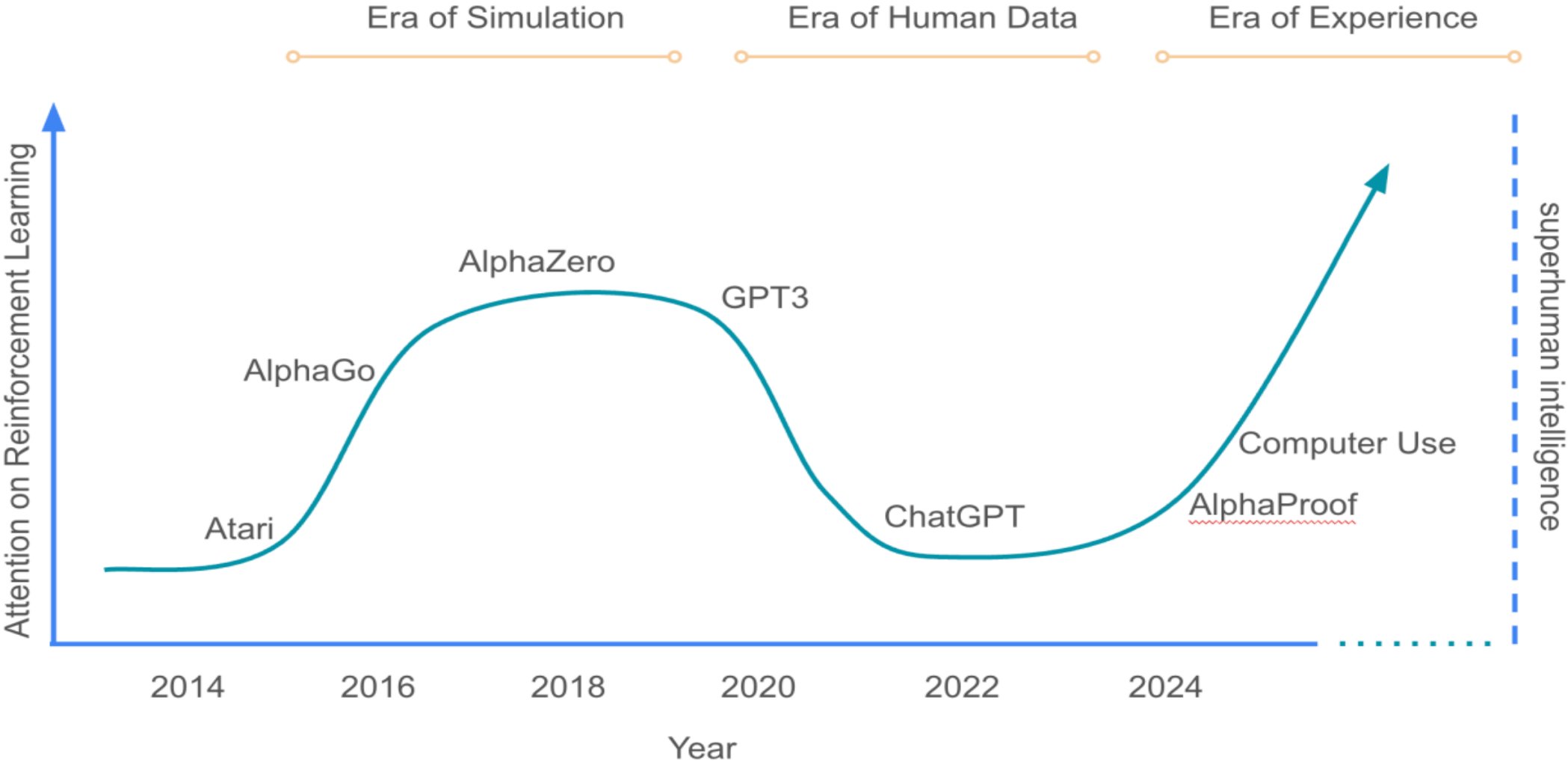


Source: ARK Investment Management LLC, 2025. For informational purposes only and should not be considered investment advice or a recommendation to buy, sell, or hold any particular security.

Embodied AI



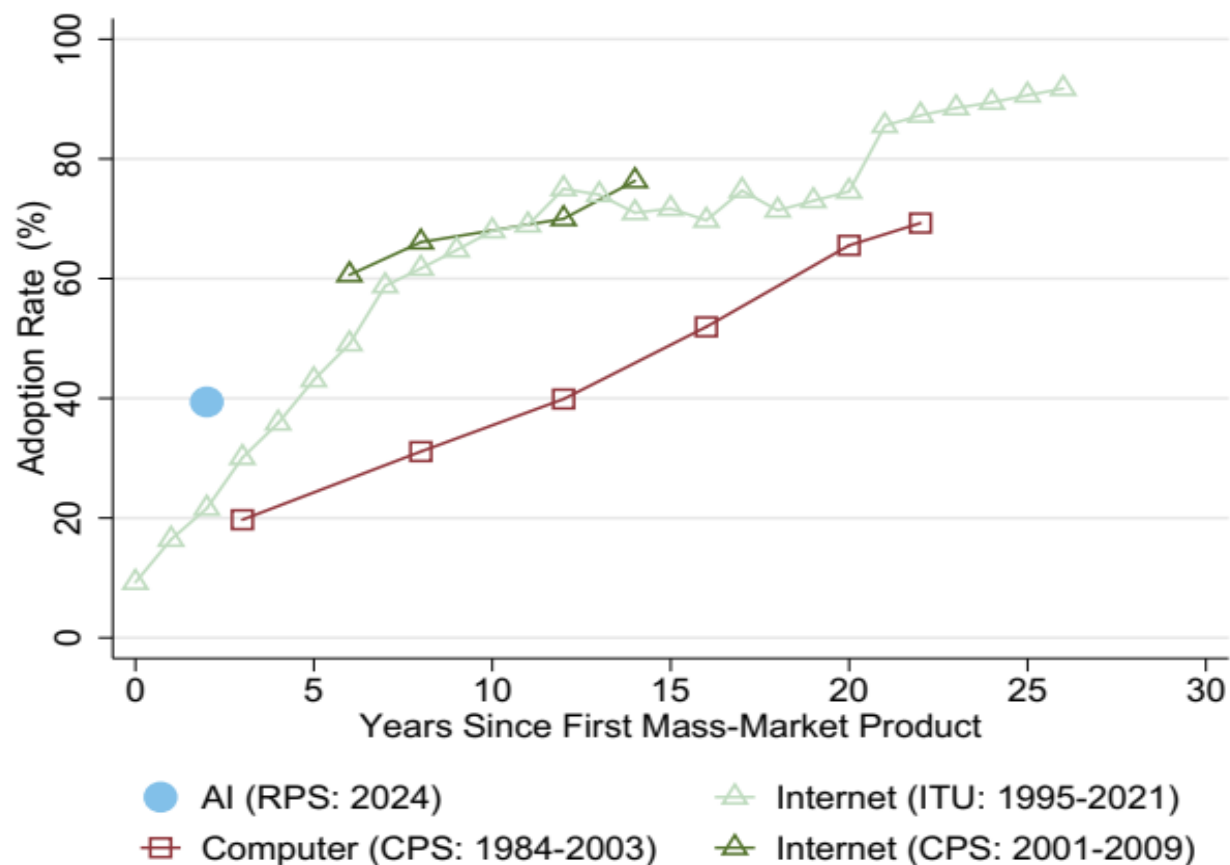
Era of Experience



AI Econ

Adoption

Figure 4: The Trajectory of Computer, Internet, and AI Adoption

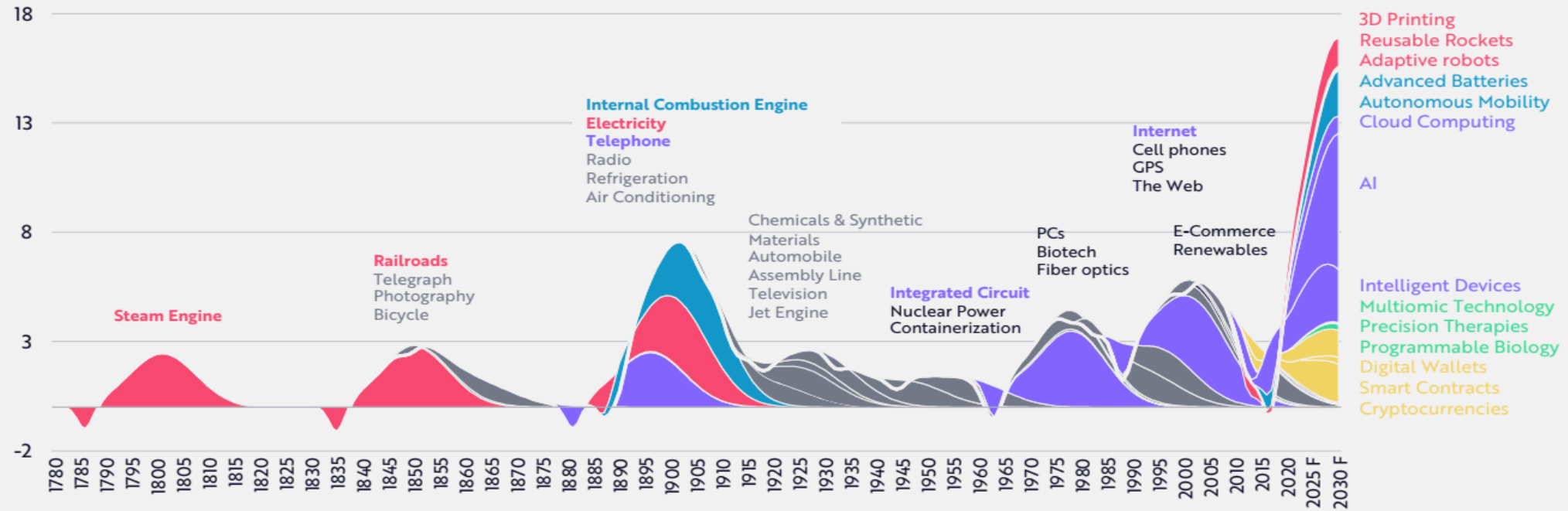


GPT

Converging Technologies Are Generating A Historic Technological Wave

Estimated Economic Impact of General Purpose Technologies

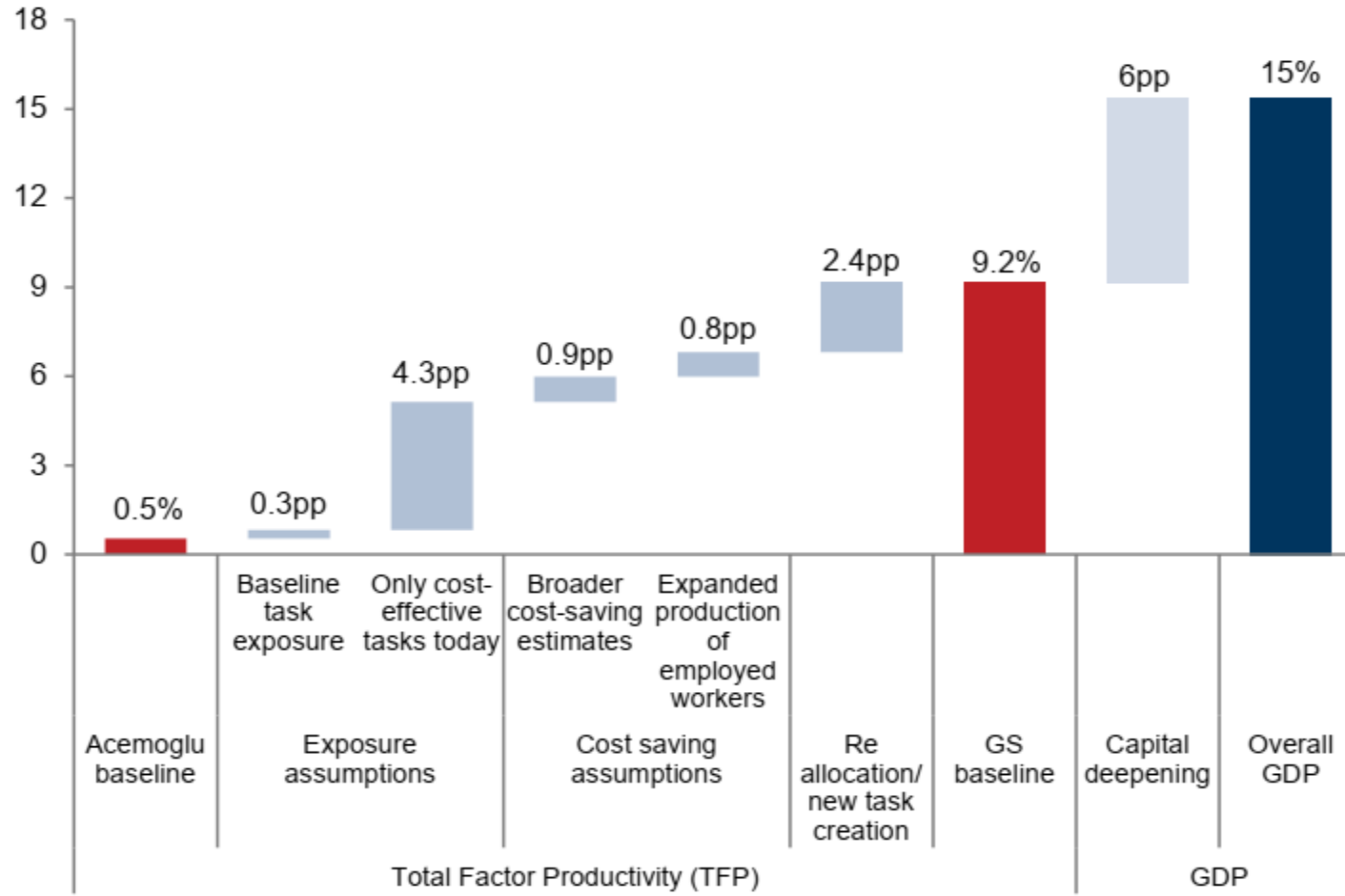
(Annual Percentage Point Additions to Real GDP Growth And Consumer Surplus)



Sources: ARK Investment Management LLC, 2024. This ARK analysis is based on a range of underlying sources, including Bekar et al. 2017, which may be provided upon request. The chart uses GPT 4 prompting to survey a comprehensive list of general purpose technologies using the identification framework detailed therein. Where available, academic literature is also used to assess attributable economic impact. A GPT-4 scoring rubric assesses technology-by-technology impacts. The impact measured directly is matched against the scoring to tune all scores to produce technology-by-technology estimates of economic impact (even when direct measures of economic impact are unattainable). Consistent with General Purpose Technology theory, these technologies are assumed to go through a period of investment in which economic impact is negative before productivity advances begin to realize into economic data. All technologies are assumed to have the same diffusion and realization cycle. If recent technologies are assumed to diffuse more quickly, the current wave would appear steeper. Forecasts are inherently limited and cannot be relied upon. For informational purposes only and should not be considered investment advice or a recommendation to buy, sell, or hold any particular security. Past performance is not indicative of future results.

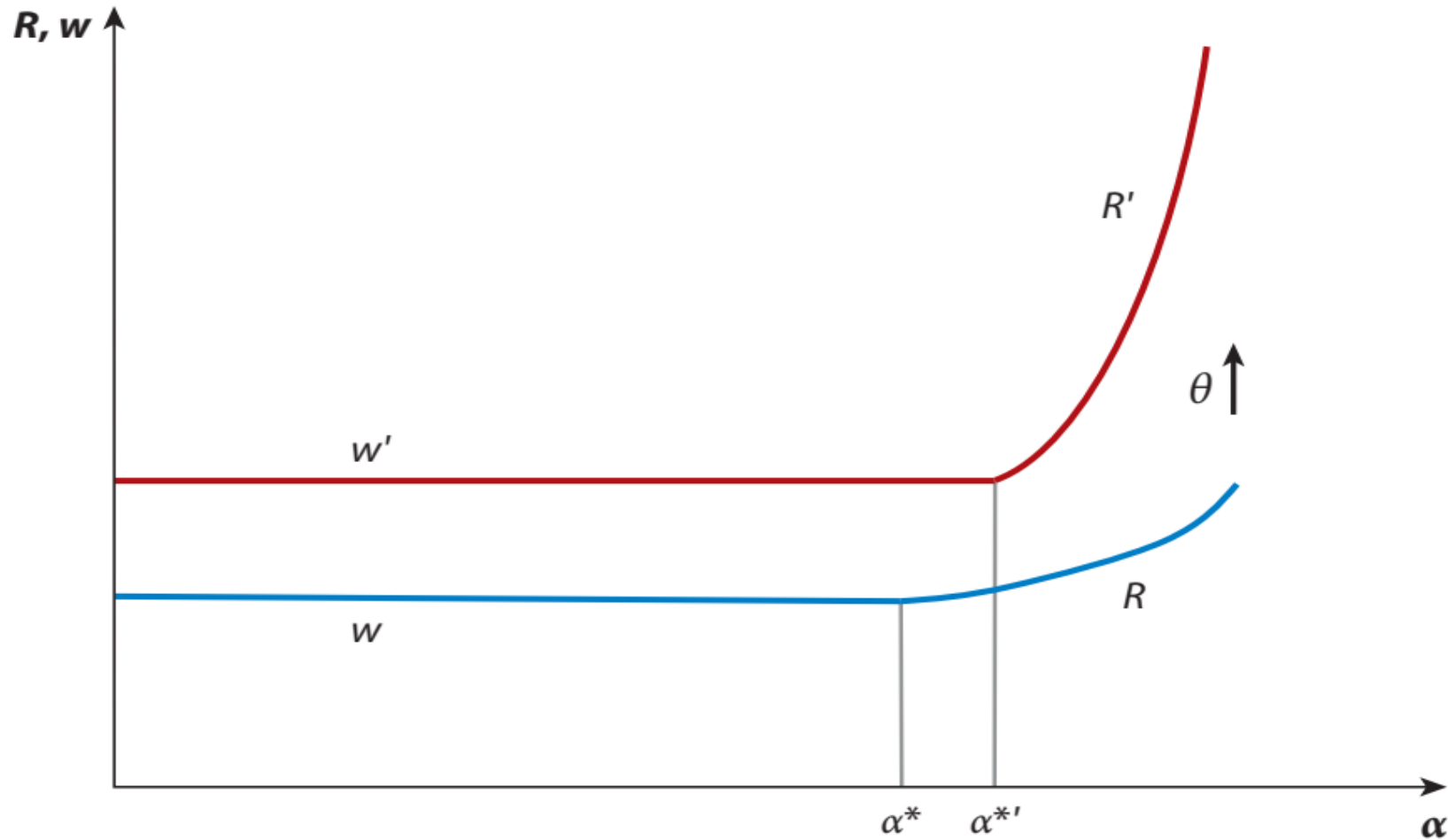
TFP and growth

Reconciling estimates of AI impact on GDP: Acemoglu (2024) vs. GS (2023), %

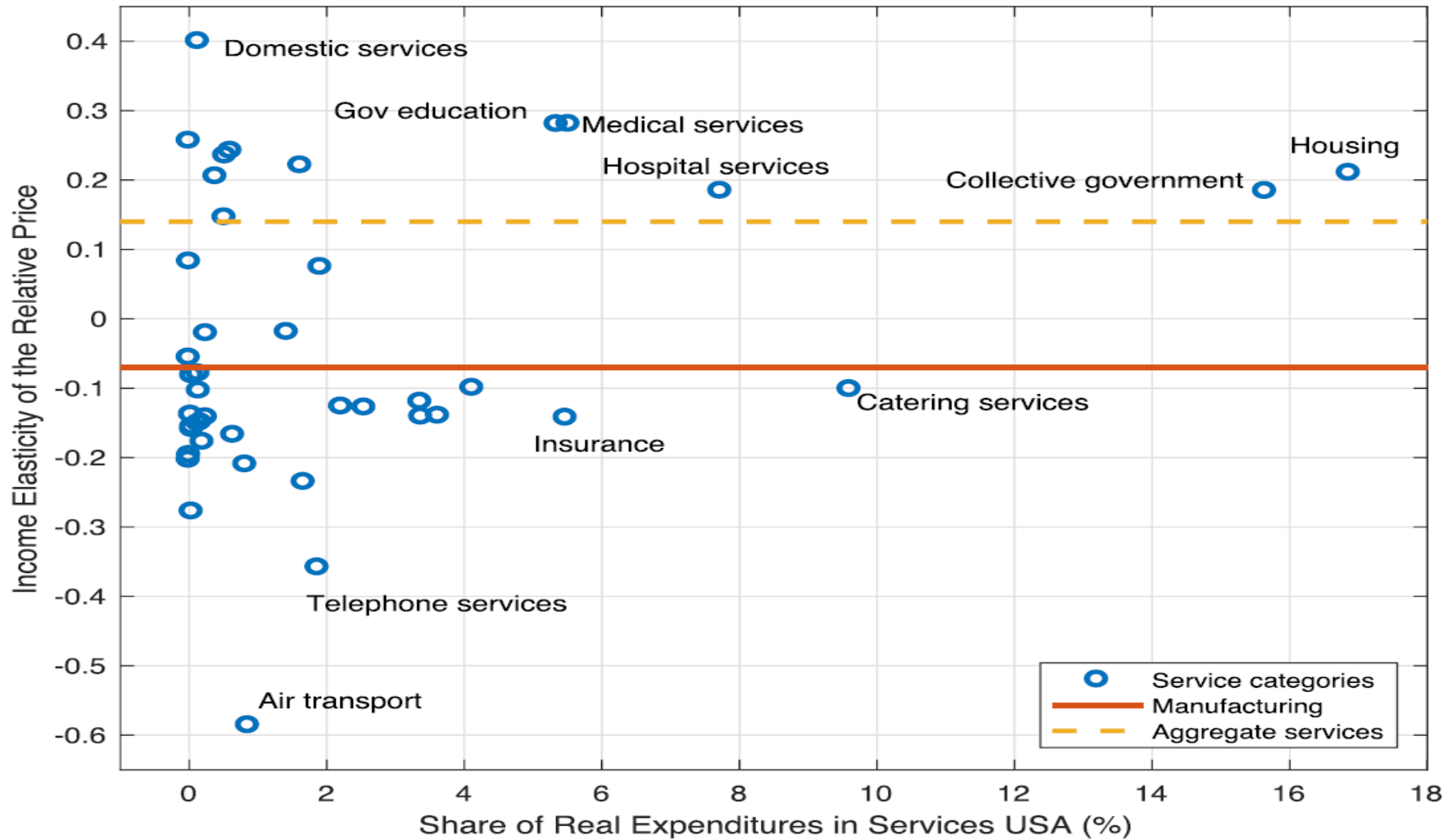


Source: Goldman Sachs GIR.

Productivity : Span of Control



Productivity : Baumol's Disease



Labor market

Exposure/GPT VS our work

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

Tyna Eloundou¹, Sam Manning^{1,2}, Pamela Mishkin*¹, and Daniel Rock³

¹OpenAI

²OpenResearch

³University of Pennsylvania

March 20, 2023

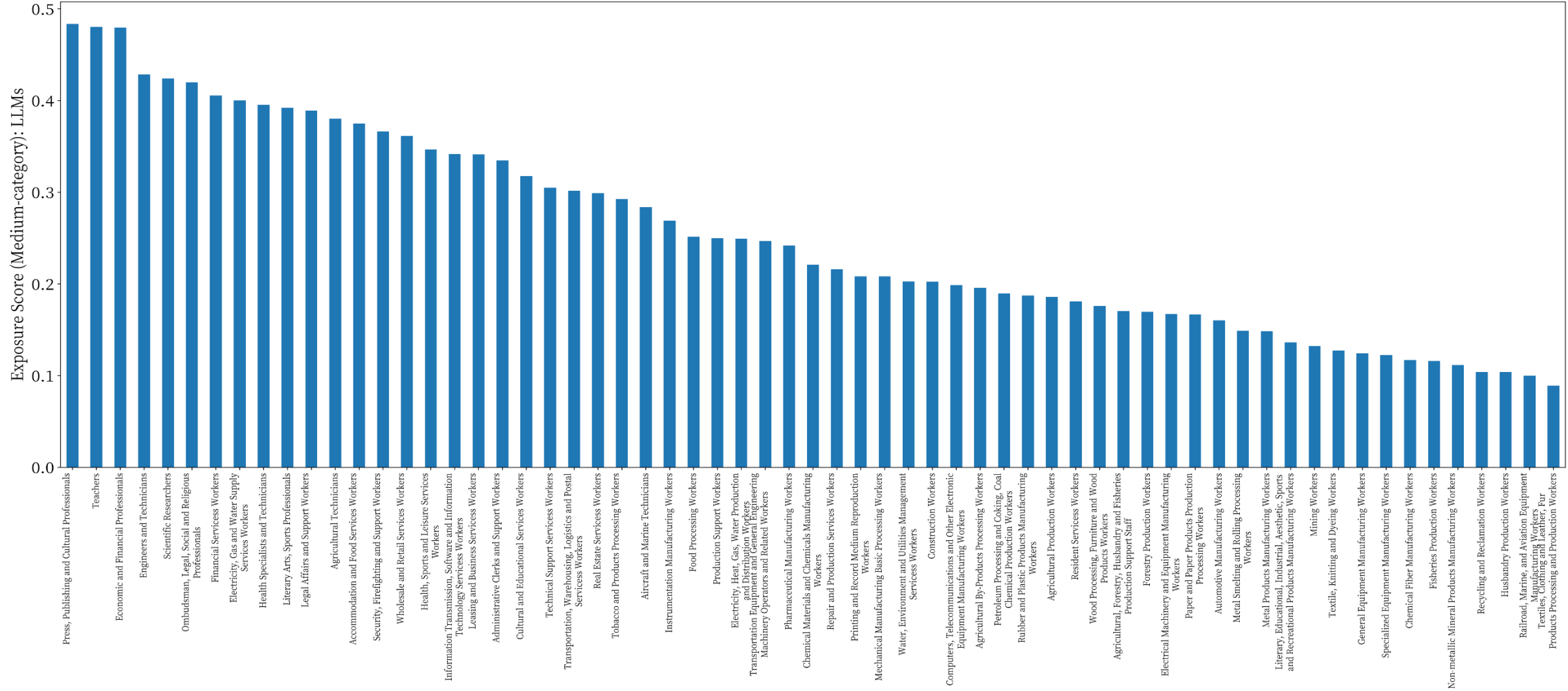
Abstract

We investigate the potential implications of Generative Pre-trained Transformer (GPT) models and related technologies on the U.S. labor market. Using a new rubric, we assess occupations based on their correspondence with GPT capabilities, incorporating both human expertise and classifications from GPT-4. Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted. The influence spans all wage levels, with higher-income jobs potentially facing greater exposure. Notably, the impact is not limited to industries with higher recent productivity growth. We conclude that Generative Pre-trained Transformers exhibit characteristics of general-purpose technologies (GPTs), suggesting that as these models could have notable economic, social, and policy implications.

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

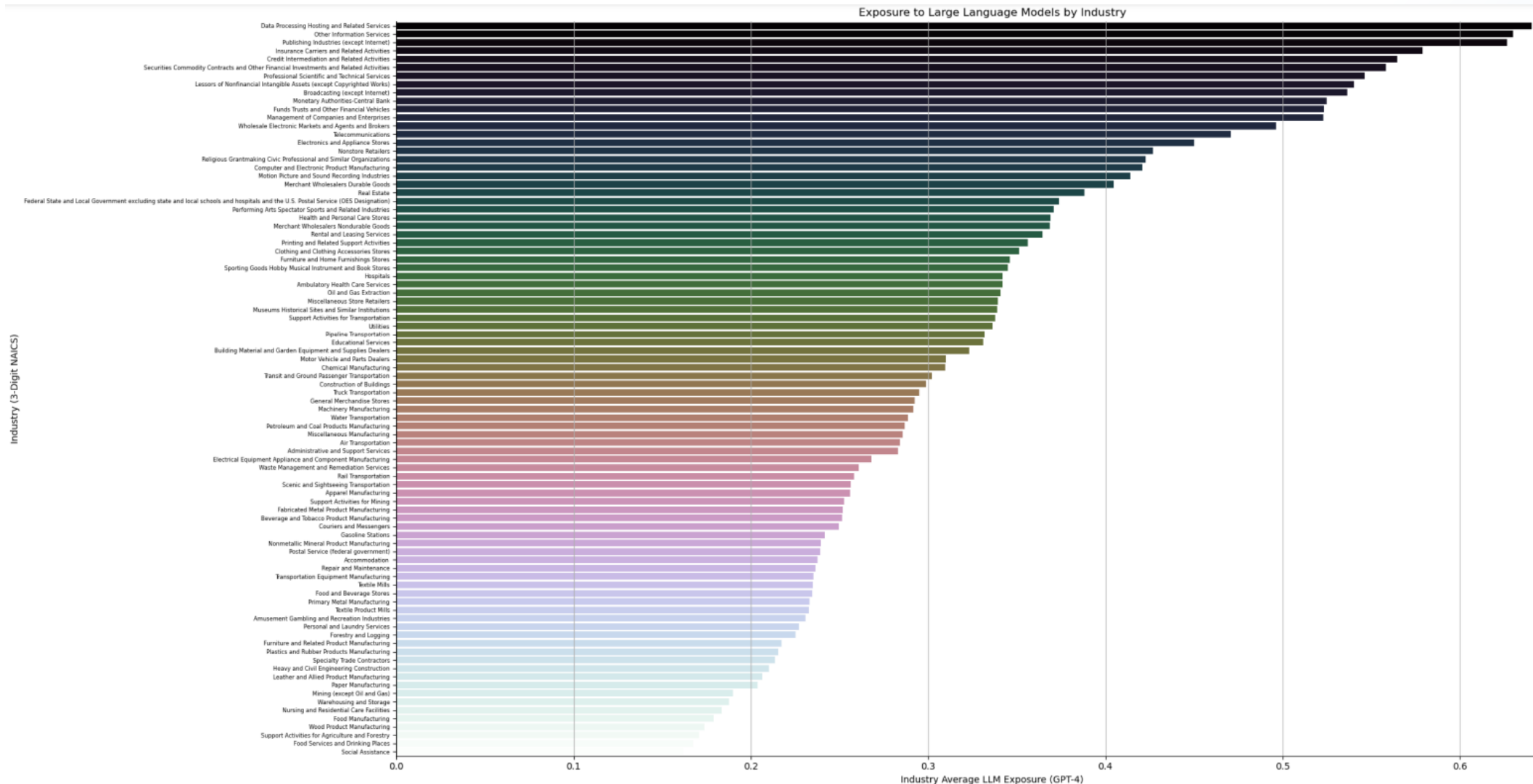
Exposure/CH

中国最高的Exposure Score职业为“新闻、出版和文化专业人员”、“教师”、“经济和金融专业人员”；最低的Exposure Score职业为“纺织品、服装和皮革毛皮产品加工和生产人员”、“制造生产人员”等。



Exposure/US

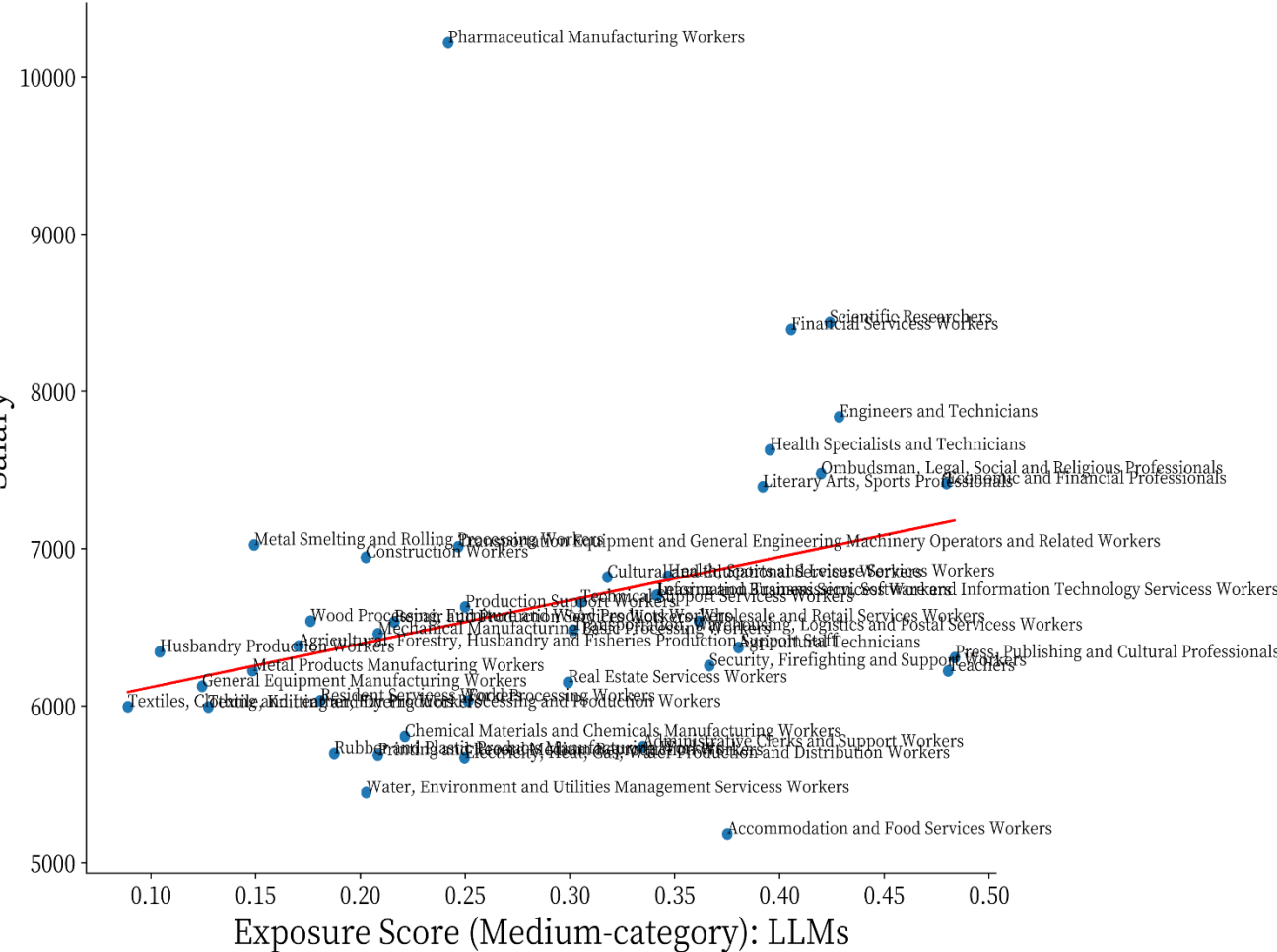
美国Exposure Score最高的行业为“Data Processing Hosting and Related Services”，“Other Information Services”，“Publishing Industries”和“Insurance Carriers”等行业；Exposure Score最低的行业为“Social Assistance”，“Food Services and Drinking places”，“Support Activities for Agriculture and Forestry”等。



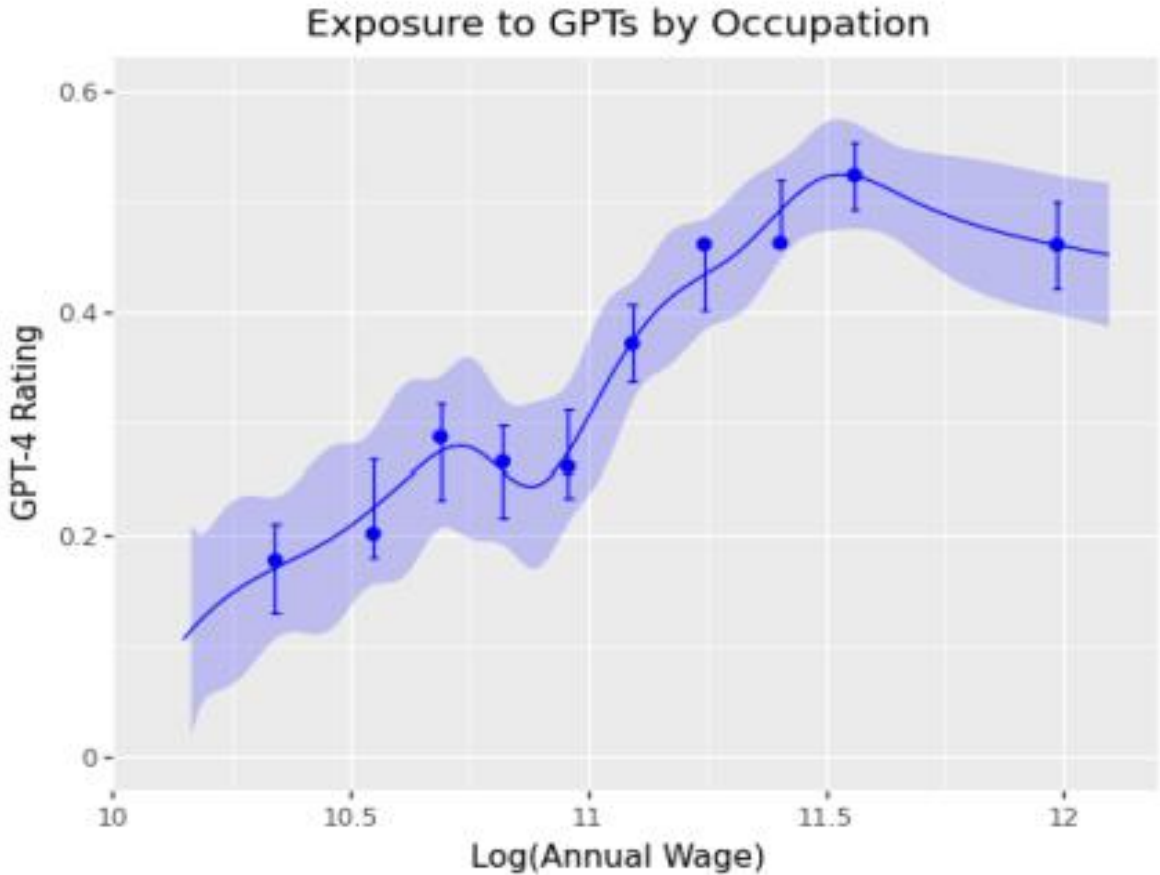
Salary and Exposure

中美数据都显示，Exposure Score和工资水平呈现正相关关系。

中国

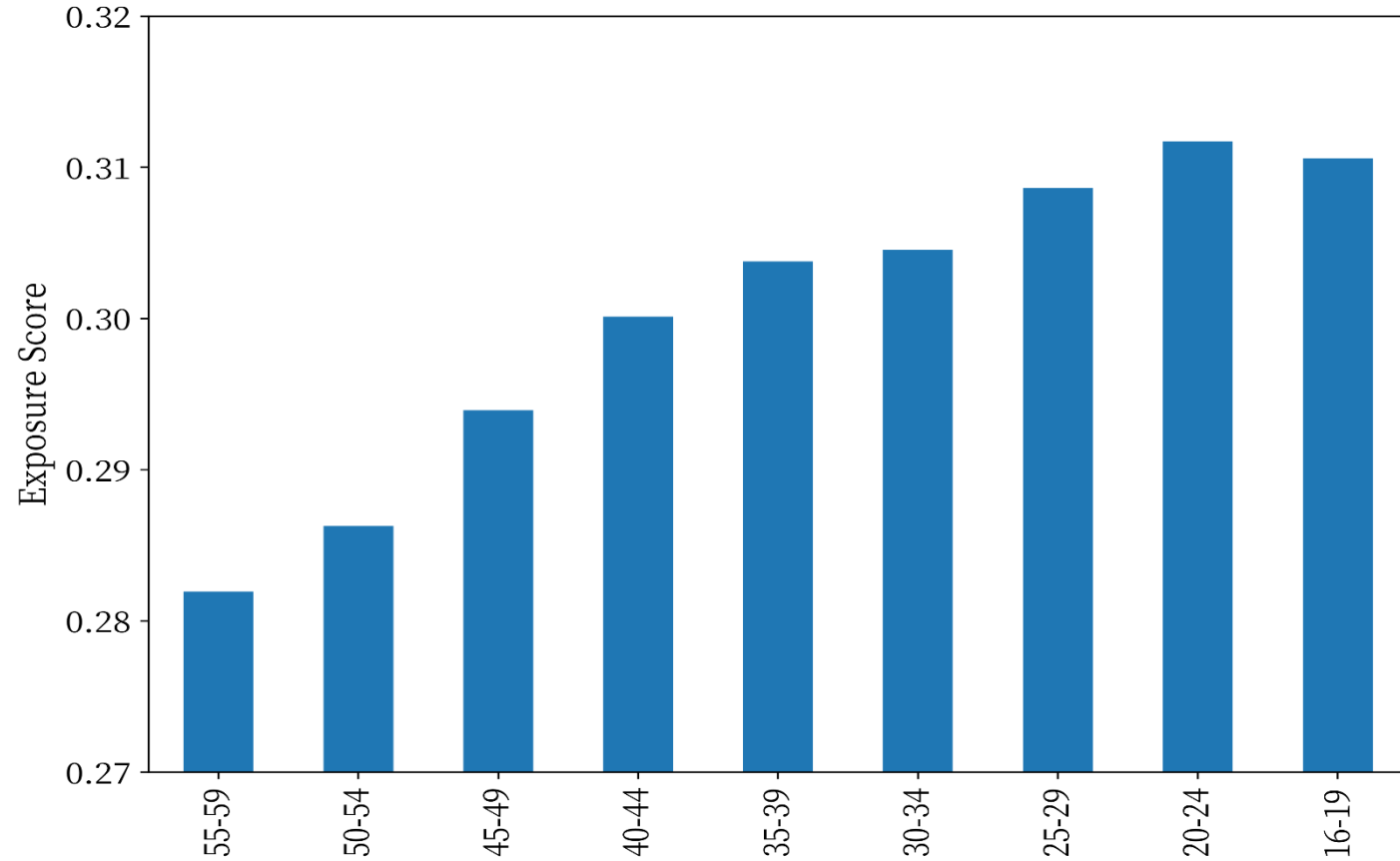


美国



Demographic Exposure/CH

中国的数据显示，20-24岁的年轻人更容易收到LLMs的影响。



AI for Econ

Global growth forecasting

Table 6: Prediction of the quarterly GDP growth with multi-indicator data results for linear regression and LSTM.

Dataset	Period	Light	Seq Length	Train: Test	Dims	Linear Regression			LSTM Best Valid Model			LSTM Final Model		
						MAE	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE
GDP with Other Variables	95-19	x	8	592: 108	21	0.6626	0.7761	0.8810	0.7845	0.9352	0.9670	0.6623	0.7122	0.8439
			10	558: 104	21	0.9189	1.8856	1.3732	0.8484	1.1802	1.0864	0.7097	0.8586	0.9266
			12	530: 100	21	0.9225	1.6033	1.2662	0.7122	0.8166	0.9037	0.7659	0.9043	0.9510
		x	8	202: 56	21	3.6646	34.8454	5.9030	0.6869	0.8769	0.9364	0.6958	0.8727	0.9342
			10	170: 56	21	2.5218	15.1833	3.8966	0.7555	1.0397	1.0196	0.7910	0.9904	0.9952
			12	143: 52	21	1.2476	2.8113	1.6767	1.0640	1.7409	1.3194	0.9097	1.2224	1.1056
		sum\mean\std	8	202: 56	24	5.0465	105.0620	10.2500	0.7447	1.0500	1.0247	0.7443	1.2163	1.1029
			10	170: 56	24	1.6424	4.0354	2.0088	0.7717	0.9947	0.9974	0.8862	1.2894	1.1355
			12	143: 52	24	1.1717	2.1186	1.4555	0.8111	1.1020	1.0497	0.7885	1.1107	1.0539
	13-19	mean	8	202: 56	22	3.1707	37.9830	6.1630	0.5667	0.7400	0.8602	0.7776	1.1846	1.0884
			10	170: 56	22	1.8926	7.0782	2.6605	0.6640	0.7797	0.8830	0.7645	0.9739	0.9869
			12	143: 52	22	1.2036	2.5812	1.6066	0.8069	1.0748	1.0367	0.8759	1.1421	1.0687
	every month mean	8	202: 56	24	3.4789	35.8779	5.9898	0.7356	0.9659	0.9828	0.6097	0.7380	0.8591	
		10	170: 56	24	1.8478	6.1222	2.4743	0.7755	1.1304	1.0632	0.6910	0.8486	0.9212	
		12	143: 52	24	1.0143	1.7284	1.3147	0.7108	1.0255	1.0127	0.9846	1.3968	1.1818	

Global growth forecasting

Table 7: Prediction of the quarterly GDP growth with multi-indicator data results for Time-LLM and PatchTST.

Dataset	Period	Light	Seq Length	Train: Test	Dims	Time-LLM best Valid model			Time-LLM Final Model			PatchTST best Valid model			PatchTST Final Model		
						MAE	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE
GDP with Other Variables	95-19	×	8	592: 108	21	0.5842	0.6392	0.7995	0.5348	0.5681	0.7537	0.6052	0.7117	0.8436	0.6854	0.8860	0.9413
			10	558: 104	21	0.5258	0.4999	0.7071	0.5392	0.6115	0.7820	0.4884	0.5130	0.7163	0.4928	0.4425	0.6652
			12	530: 100	21	0.4724	0.4376	0.6615	0.5421	0.5605	0.7487	0.5005	0.4671	0.6834	0.5386	0.5957	0.7718
	13-19	×	8	202: 56	21	0.4988	0.5435	0.7372	0.4992	0.4931	0.7022	0.5939	0.6322	0.7951	0.5125	0.5102	0.7143
			10	170: 56	21	0.4932	0.5202	0.7213	0.5305	0.5865	0.7658	0.9690	1.6141	1.2705	0.4497	0.4353	0.6597
			12	143: 52	21	0.5736	0.5998	0.7745	0.5992	0.6919	0.8318	0.5285	0.5526	0.7434	0.6126	0.7758	0.8808
		sum\mean\std	8	202: 56	24	0.5469	0.6442	0.8026	0.5545	0.5648	0.7515	0.5186	0.5373	0.7330	0.5265	0.5301	0.7281
			10	170: 56	24	0.5612	0.6347	0.7967	0.5651	0.5349	0.7314	0.5064	0.4725	0.6874	0.4697	0.4474	0.6689
			12	143: 52	24	0.5434	0.5756	0.7587	0.5749	0.5836	0.7639	0.5964	0.6630	0.8142	0.5264	0.5974	0.7729
		mean	8	202: 56	22	0.5544	0.5670	0.7530	0.5527	0.6226	0.7891	0.6648	0.7893	0.8884	0.5149	0.4606	0.6787
			10	170: 56	22	0.5569	0.6488	0.8055	0.5914	0.6663	0.8163	0.5345	0.5690	0.7543	0.4831	0.5147	0.7174
			12	143: 52	22	0.6126	0.7070	0.8408	0.6450	0.7164	0.8464	0.5962	0.6543	0.8089	0.5830	0.5950	0.7713
every month mean	8	202: 56	24	0.6115	0.7286	0.8536	0.5841	0.6204	0.7877	0.5996	0.6683	0.8175	0.5737	0.5634	0.7506		
	10	170: 56	24	0.6017	0.6898	0.8305	0.5902	0.6717	0.8196	0.5696	0.5047	0.7104	0.4942	0.4035	0.6352		
	12	143: 52	24	0.6826	0.8536	0.9239	0.6176	0.6912	0.8314	0.6162	0.6780	0.8234	0.7491	0.9456	0.9724		

Global trade war

- N countries with labor endowments $\{L_i\}_{i=1}^N$. J sectors.
- Labor: immobile across countries but perfectly mobile across sectors.
- Preference:

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

- 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

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

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:

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

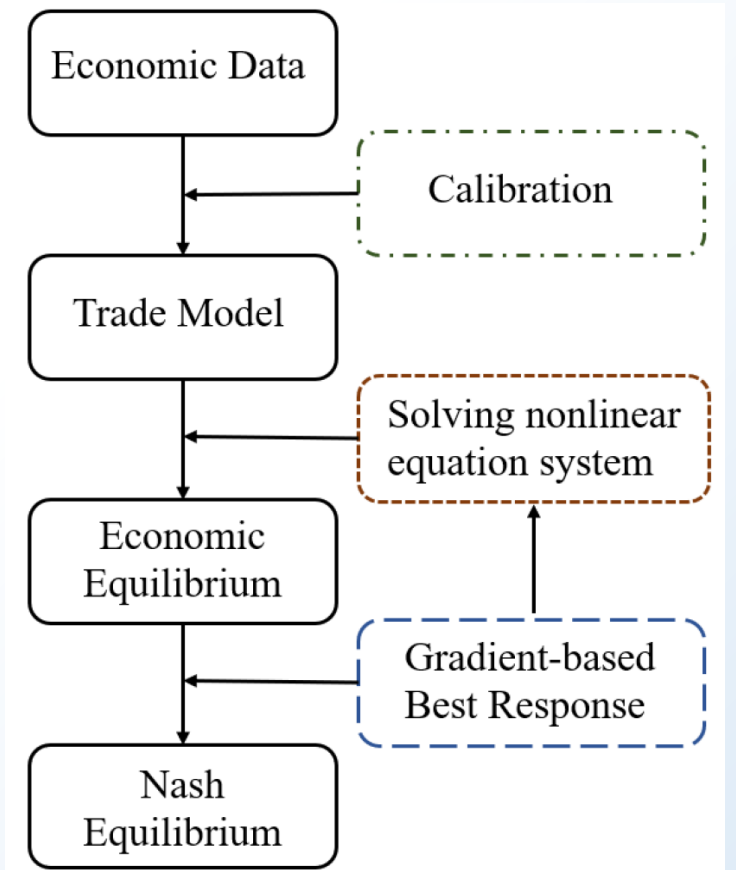


Table: Computation statistics of Solving for Global Dual Competition

Information	Value
ML Framework	PyTorch
Device	CPU
Num. epochs	~ 20
Iterations per player/epoch	~ 50
Num. players	7
Playing sequence	Random Shuffle
Optimizer	ADAM
Anneal learning rate	False
Learning rate	$10^{-4} \sim 5 \times 10^{-3}$
Is clipping grad	{True, False}
Max grad norm	10.0
Max Computation Time	15h

Nash Equilibrium

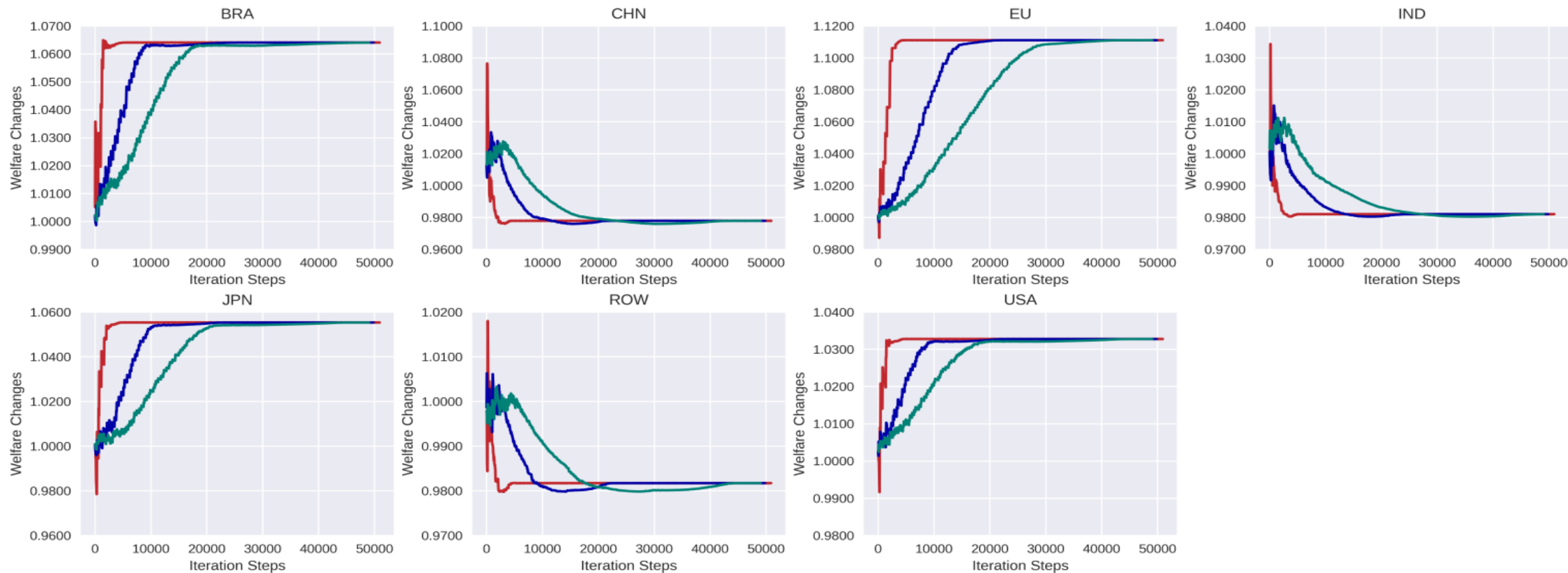
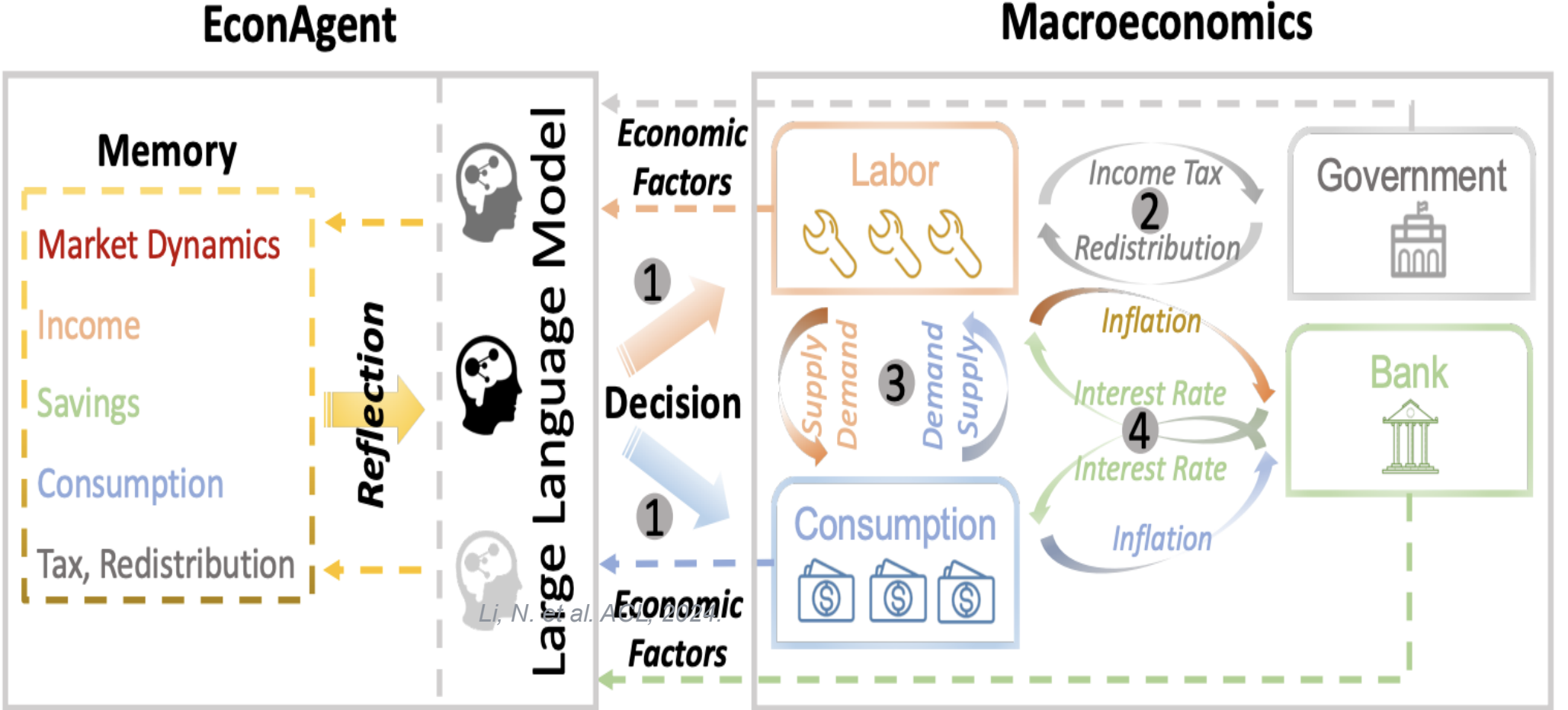


Figure: Iteration Curve for Nash Equilibrium of Global Dual Competition

Trade war impact

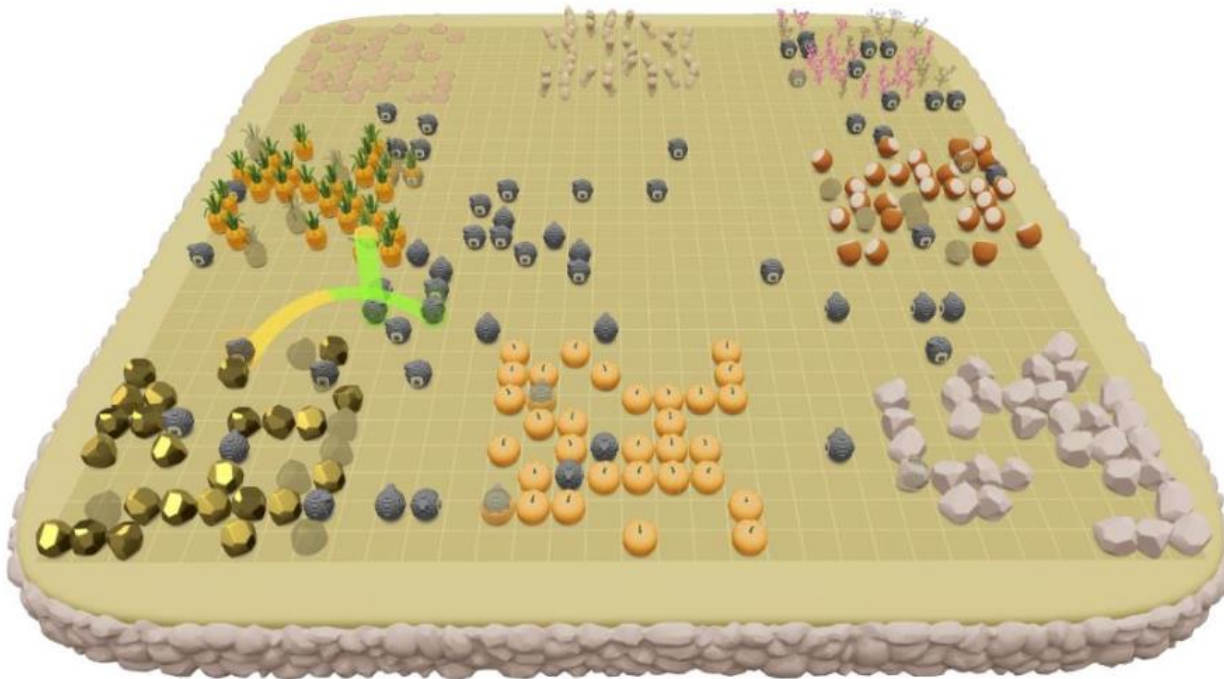
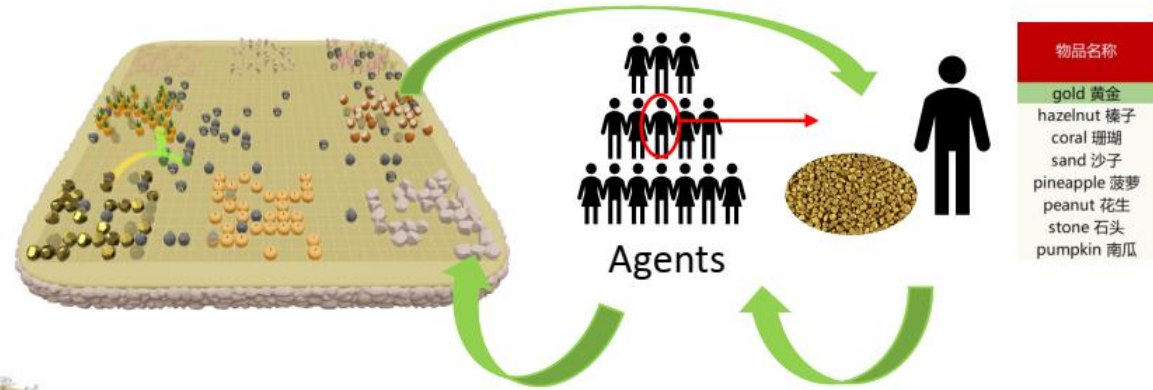
	美国对其关税增加幅度	对美国出口	实际GDP (不含关税收入)	实际GDP (含关税收入)
	%	变化(%)	变化(%)	变化(%)
巴西	10	-17.8	-0.26	-0.28
中国	145	-74.6	-1.20	-0.91
欧盟	10	-17.2	-0.23	-0.23
印度	10	-36.7	-0.71	-0.74
日本	10	-11.8	-0.17	-0.17
其他	10	-16.2	-0.50	-0.52
美国	n.a.	n.a.	-0.46	0.05

Agent in Econ



Agent in Econ

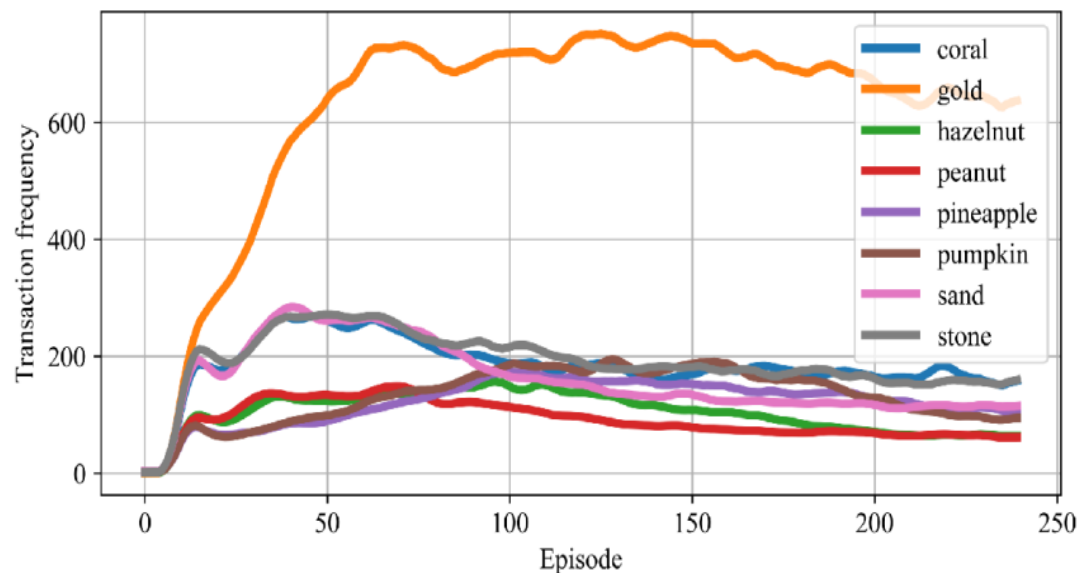
- 1 设定具有给定物理属性的物品
- 2 构建支撑多智能体交易的环境
- 3 设置规则与互动机制



Agent in Econ

物理属性对货币涌现的影响

耐用性、稀缺性、易分割性对货币形成是必要的，具备这三个属性的商品作为货币涌现；

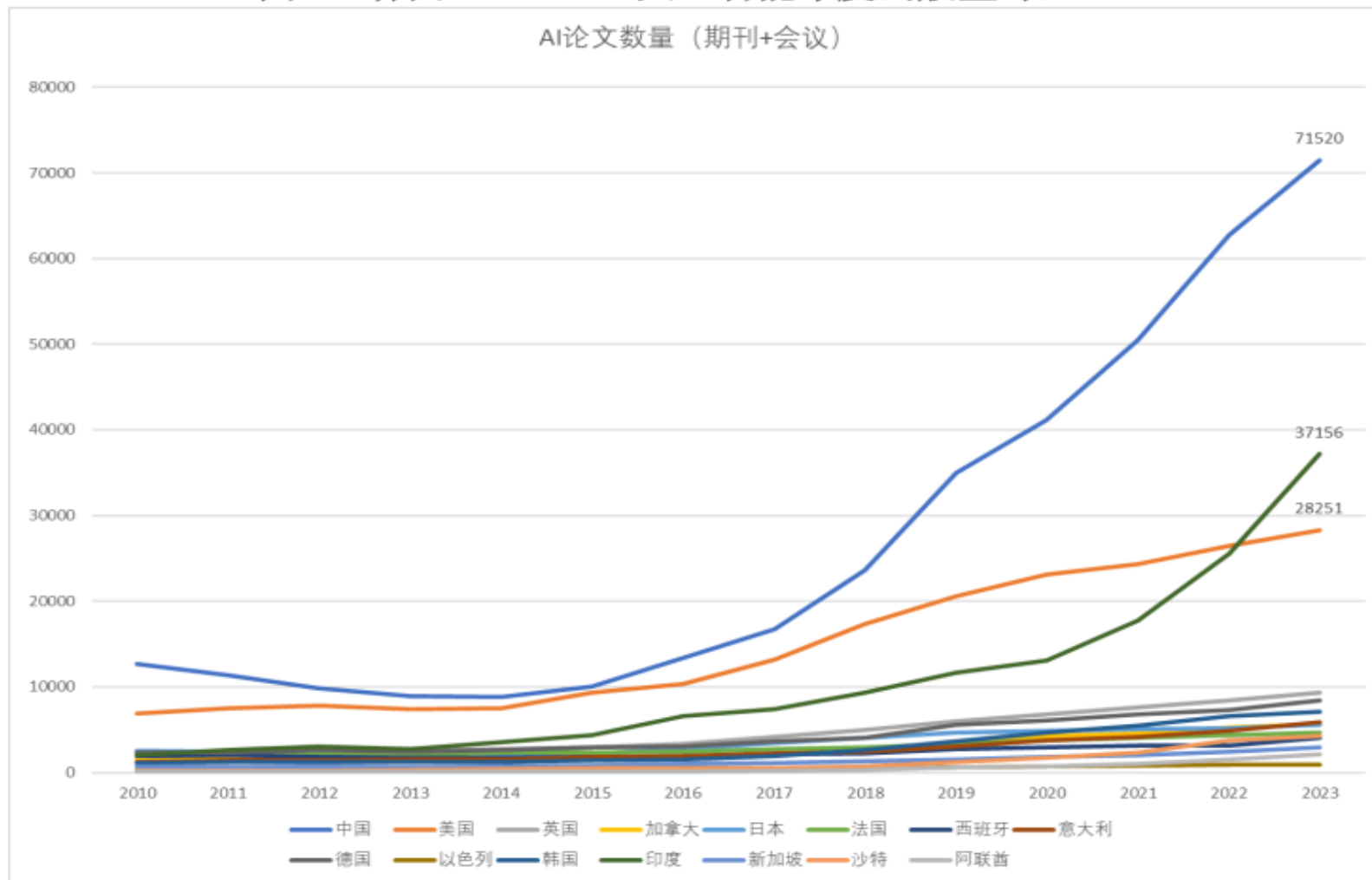


纵坐标为物品交易次数，横坐标为时间步。

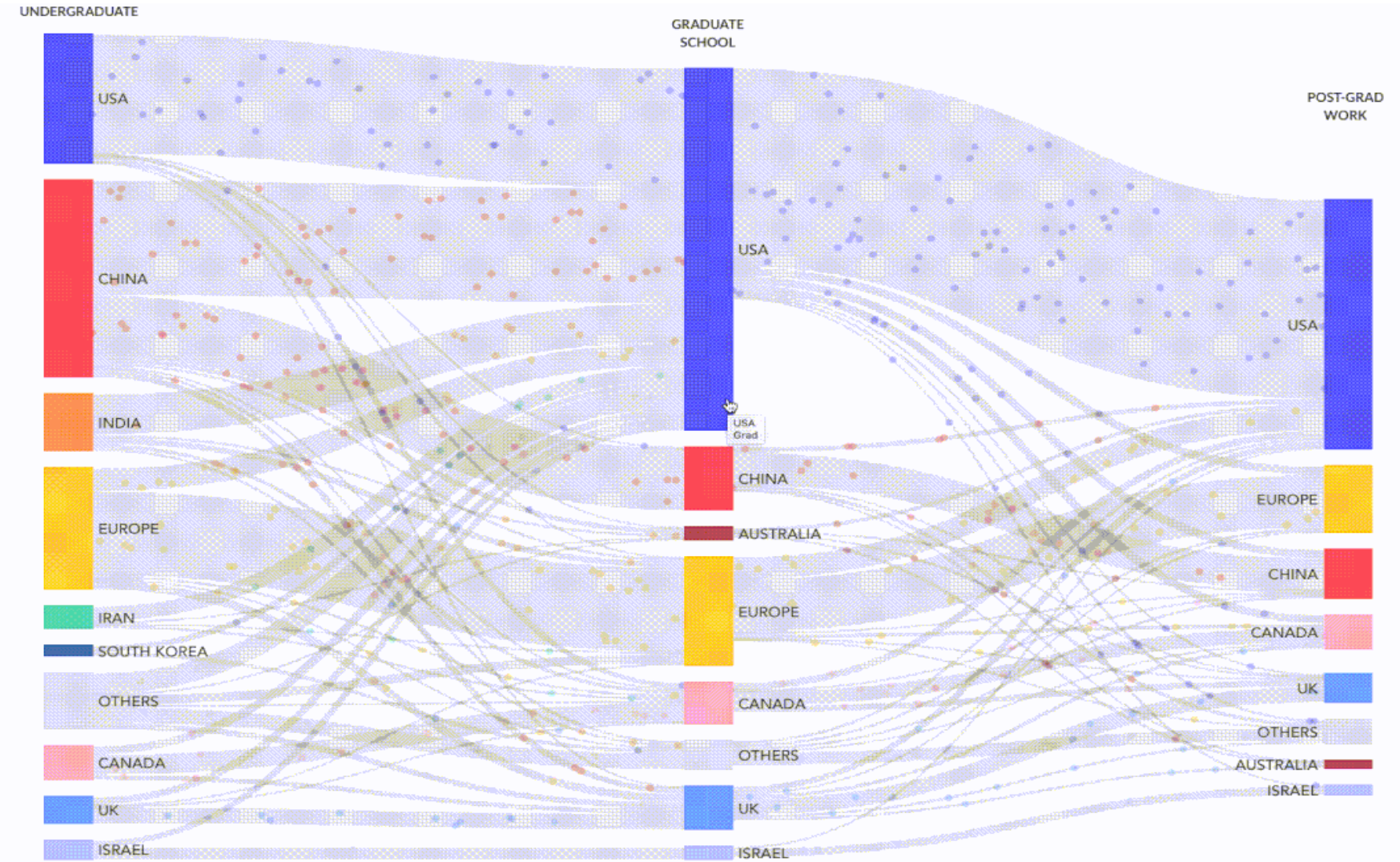
Strategic Competition

AI : global publication

图 1.1 各国 2010-2023 人工智能年度出版量对比



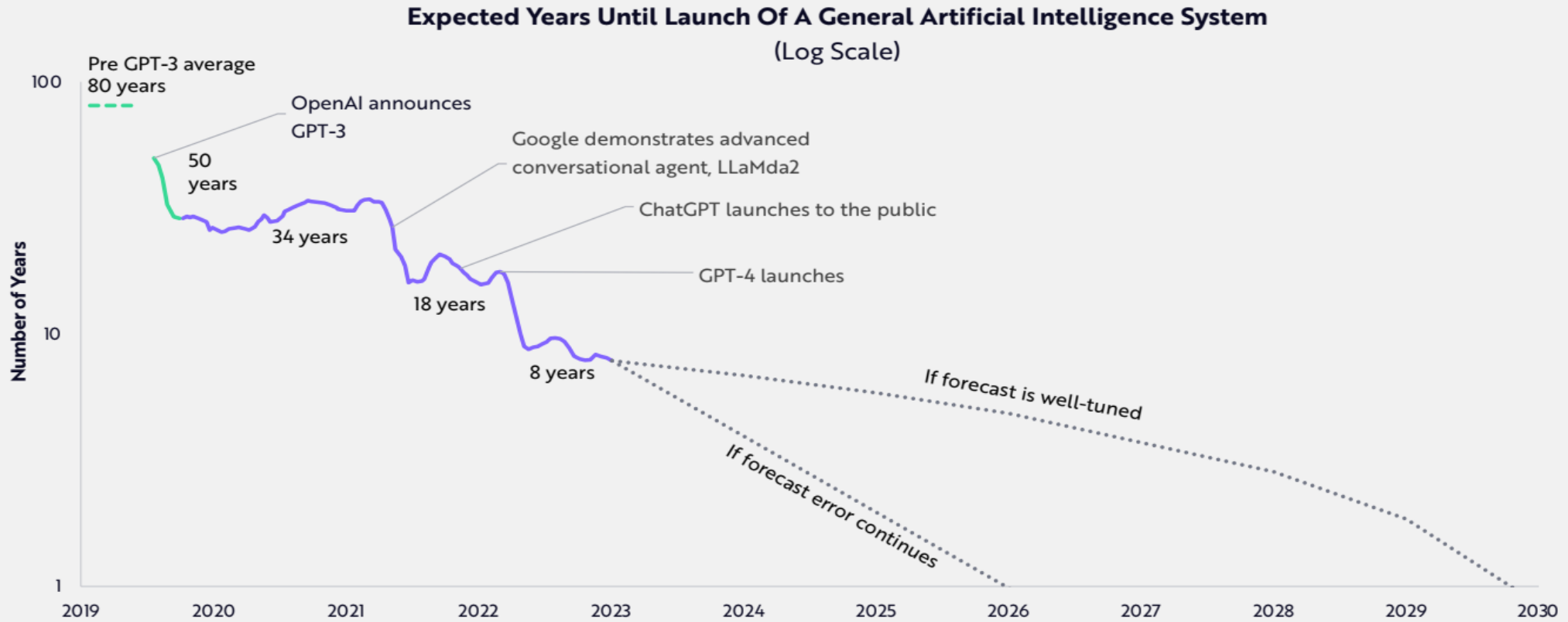
AI : global talent



Future

AGI

AI Is Accelerating Faster Than Forecasters Anticipated



Sources: ARK Investment Management LLC, 2024, based on data from Metaculus, including benchmark details, as of January 3, 2024. Benchmark broadly requires the successful passage of an adversarial two-hour Turing test, broad success on a Q&A knowledge and logic benchmark, and the successful interpretation of and execution complex model car assembly instruction, all within a single system. Green lines are derived estimates for time to general purpose AI (strongly formulated) based upon forecasts for a weaker benchmark. Forecasts are inherently limited and cannot be relied upon. For informational purposes only and should not be considered investment advice or a recommendation to buy, sell, or hold any particular security. Past performance is not indicative of future results.

Mechanical Interpretation

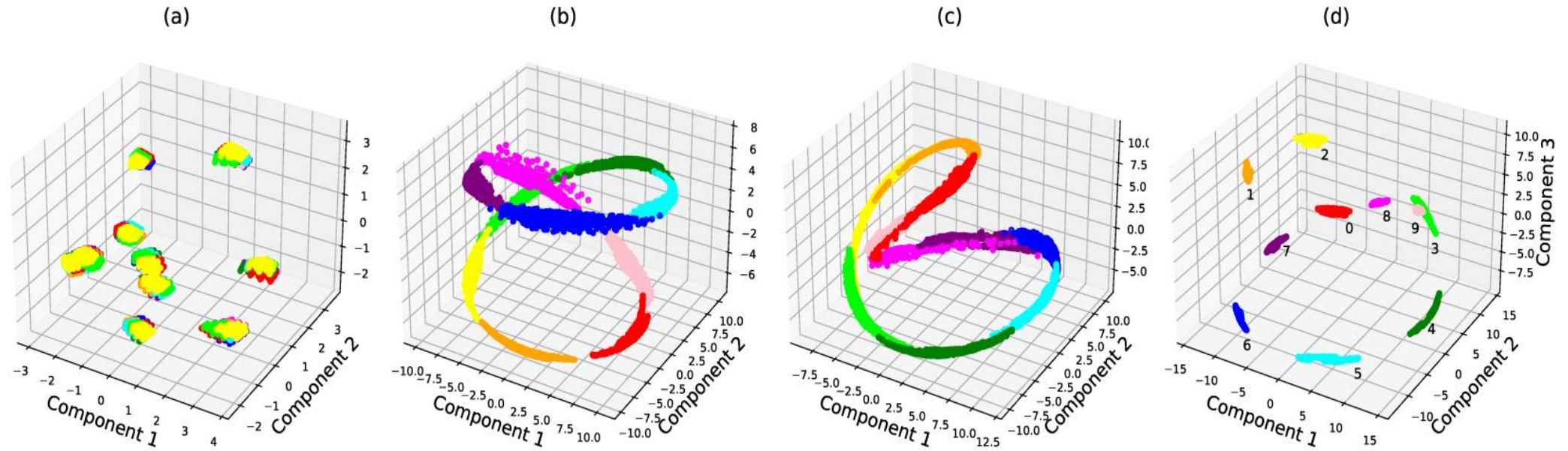


Figure 4: 3D representation structure of the first three principal components in the addition operation. Figure 4(a) to Figure 4(d) represent the initial model, model with 14%, 51%, and 100% test accuracy, respectively.

The
Guardian



=Web3天空之城=

