Machine, Data and Learning

ML Introduction

Machine Learning

- Scientific study of algorithms and statistical models that computer systems use
  - To perform a specific task effectively without using explicit instructions
  - Rely on patterns and inference instead.
- Involves
  - Building a **mathematical model** based on sample data, known as "training data“ to make predictions or decisions
  - No explicit programming done to perform the task

Machine Learning

- Term coined around 1960
- Why learn ? Why not just hire enough programmers and code in rules ?
  - Lots of patterns for an activity/event
  - Events can be dynamic
  - **Data** is increasing exponentially
  - **Data** is also in various formats [Text, Audio, Video]
  - Higher quality **data** due to cheaper storage
- Can be broadly classified into three categories
  - Unsupervised, Supervised and Reinforcement learning

Unsupervised Learning

- Takes a set of data that contains only inputs and finds structure in data E.g., Grouping or Clustering of data points

- **Marketing:** Finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records.
- **Biology:** Classification of plants and animals given their features.
- **Earthquake studies:** Clustering observed earthquake epicenters to identify dangerous zones.
- **World Wide Web:** Clustering weblog data to discover groups of similar access patterns.

Supervised Learning

- Builds mathematical model using data set that has both inputs and desired outputs E.g., Classification and Regression tasks

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

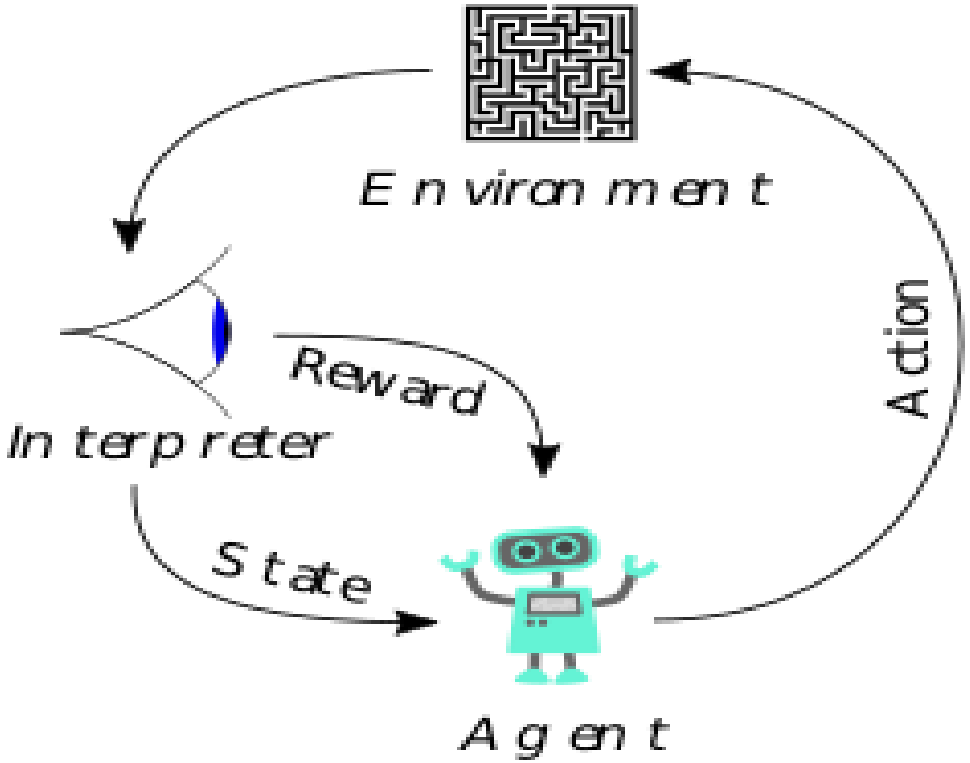
Figure A: CLASSIFICATION

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Figure B: REGRESSION

Reinforcement Learning

- Concerned with how software agents should take actions in an environment to maximize cumulative reward E.g. Autonomous vehicles, Computer games



Some Applications

- Search engines
- Information retrieval
- Recommendation systems
- Credit card fraud detection
- Disease diagnosis
- Election prediction
- Image processing
- Speech translation
- …

AlphaGo

- First computer Go program to defeat a 9-dan professional player
- Uses Monte Carlo Tree search algorithm based on knowledge learned by a deep learning method
- Beat World No. 1 ranked player in 2017
  - Retired after this match
- https://deepmind.google/research/breakthroughs/alphago/
- https://www.youtube.com/watch?v=WXuK6gekU1Y

- AlphaGo Zero – Version without human data and stronger than AlphaGo [defeated 100-0]

AlphaZero & MuZero

- AlphaZero, a generalized version of AlphaGo Zero Took 4 hours to learn Chess and defeat reigning world computer chess champion 28 to 0 in 100 matches
- https://www.youtube.com/watch?time_continue=7&v=tXlM99xPQC8

- MuZero: Master games without knowing rules
- Uses approach similar to AlphaZero, developed in 2019
- Trained via self-play and play against AlphaZero with no access to rules, opening books or endgame tables
- Viewed as significant advancement over AlphaZero

AlphaFold: solution to a 50 year old grand challenge in biology

- https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology
- Figuring out what shapes proteins fold into is known as the “protein folding problem” - grand challenge in biology for the past 50 years
- Focus of intensive scientific research for many years, using a variety of experimental techniques such as nuclear magnetic resonance and X-ray crystallography.

AlphaFold

- Number of ways a protein could theoretically fold before settling into its final 3D structure is astronomical.
- Cyrus Levinthal estimated 10^300 possible conformations for a typical protein.
- Estimated would take longer than the age of universe to enumerate all possible configurations. Yet in nature, proteins fold spontaneously, some within milliseconds - referred to as Levinthal's paradox.

Nobel Prize in Chemistry 2024 !!!

- The Nobel Prize in Chemistry 2024 is about proteins, life's ingenious chemical tools.
- David Baker has succeeded with the almost impossible feat of building entirely new kinds of proteins.
- Demis Hassabis and John Jumper have developed an **AI model** to solve a 50-year-old problem: predicting proteins' complex structures. These discoveries hold enormous potential.

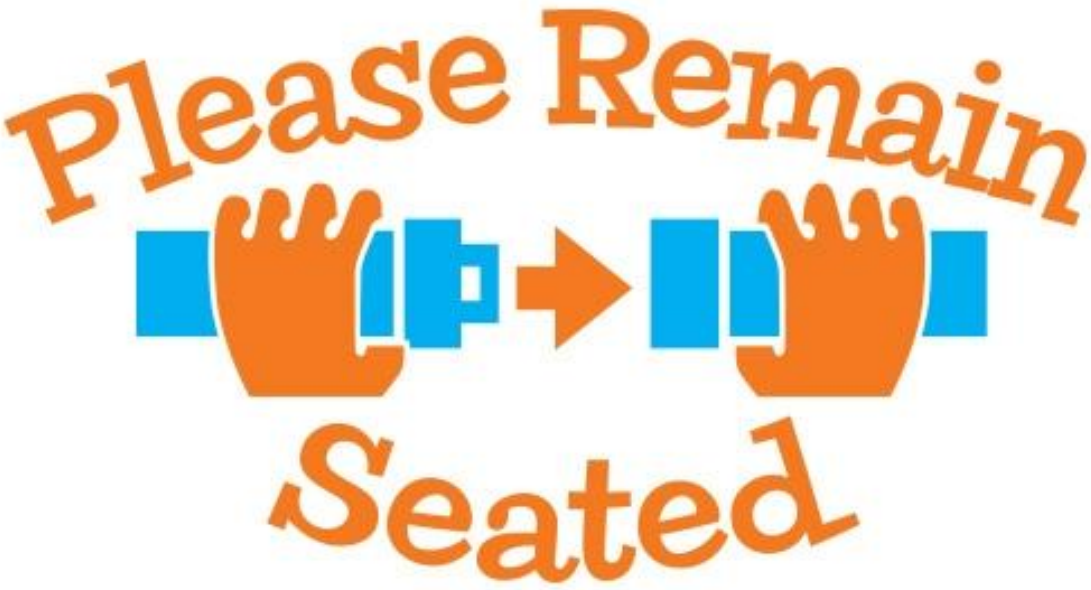
Nobel Prize in Physics 2024 !!!

- They used physics to find patterns in information
- John J Hopfield and Geoffrey Hinton
- “for foundational discoveries and inventions that enable machine learning with artificial neural networks”

Turing Award in 2024 !!!

- Nobel prize of computing
- Awarded to Andrew Barto and Richard Sutton
- “for developing the foundational concepts and algorithms for **Reinforcement Learning (RL)**, a key area in Artificial Intelligence, particularly for training large models like LLMs”

Admin



Admin

- ……
- Any others ?
- Surprise quizzes…

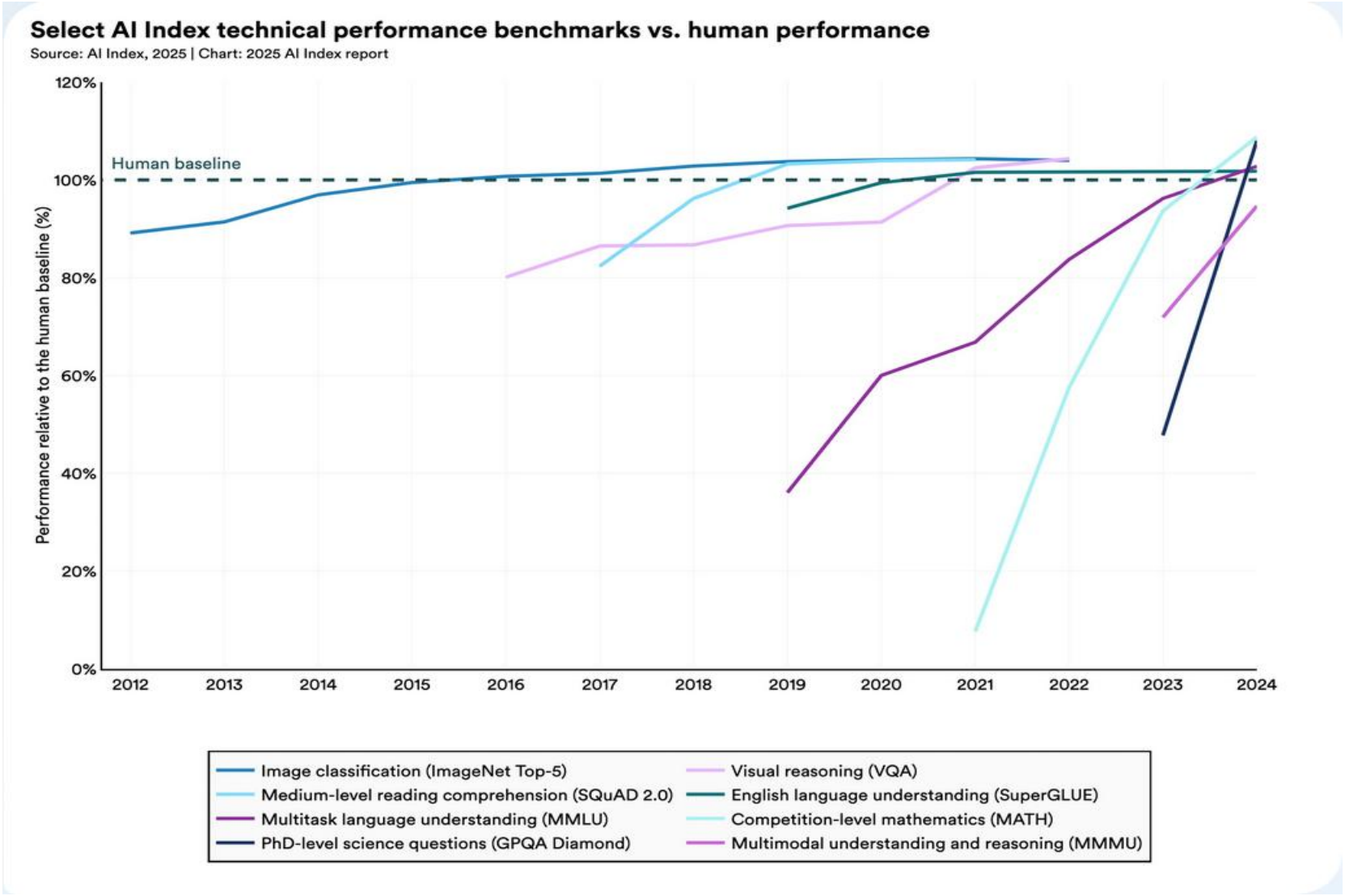
AI Index 2025

- Source: Stanford Institute for Human-Centered Artificial Intelligence (HAI)
- Report: 2025 AI Index Report
- https://hai.stanford.edu/ai-index/2025-ai-index-report
- Slides follow

AI Performance on Demanding Benchmarks

- AI achieved large gains on harder benchmarks
- Strong progress in coding and video generation
- AI agents outperform humans in time-constrained tasks

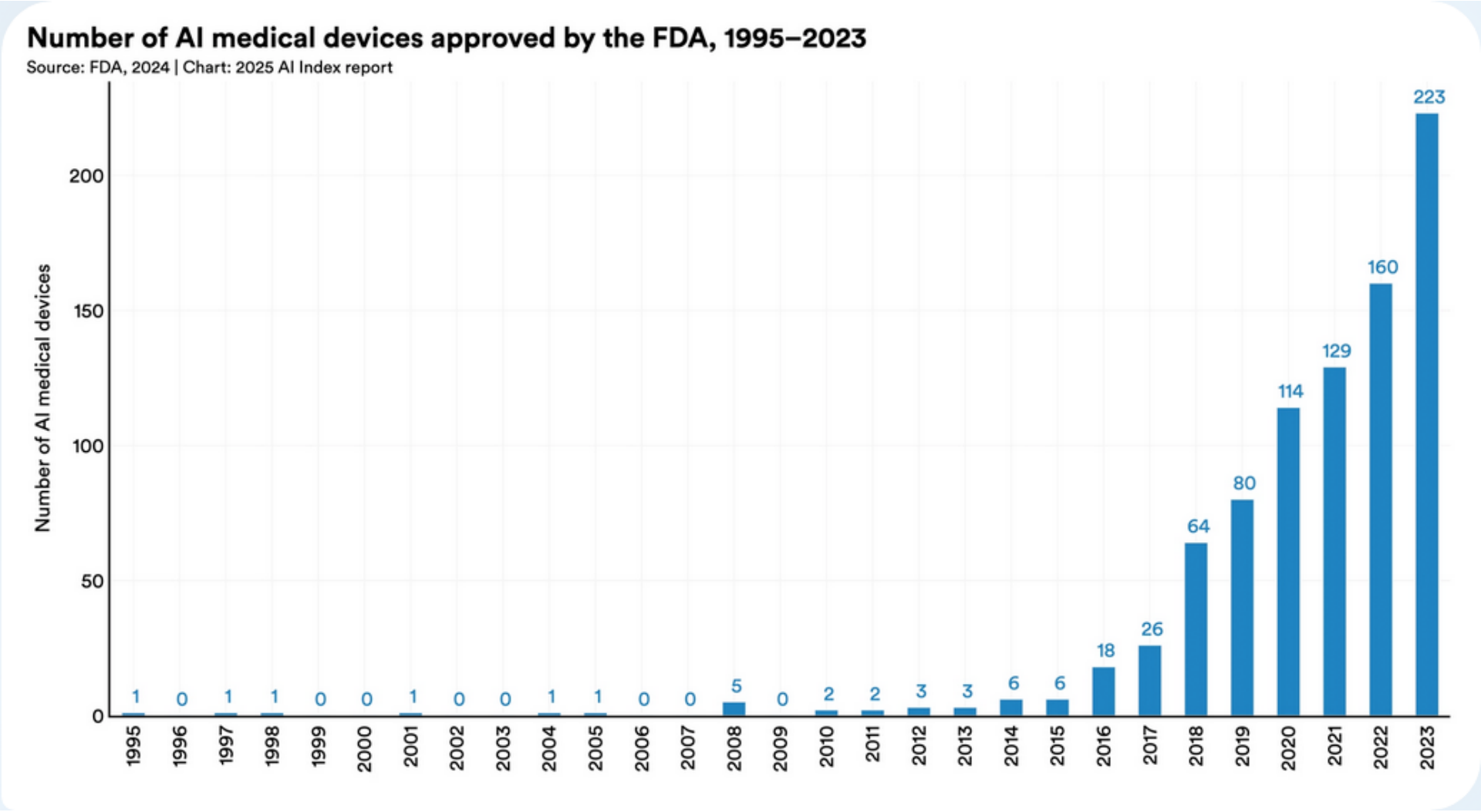
AI Performance on Demanding Benchmarks — Data View



AI Embedded in Everyday Life

- AI adoption expanded across healthcare and transport
- FDA approvals of AI medical devices surged
- Autonomous vehicles reached commercial scale

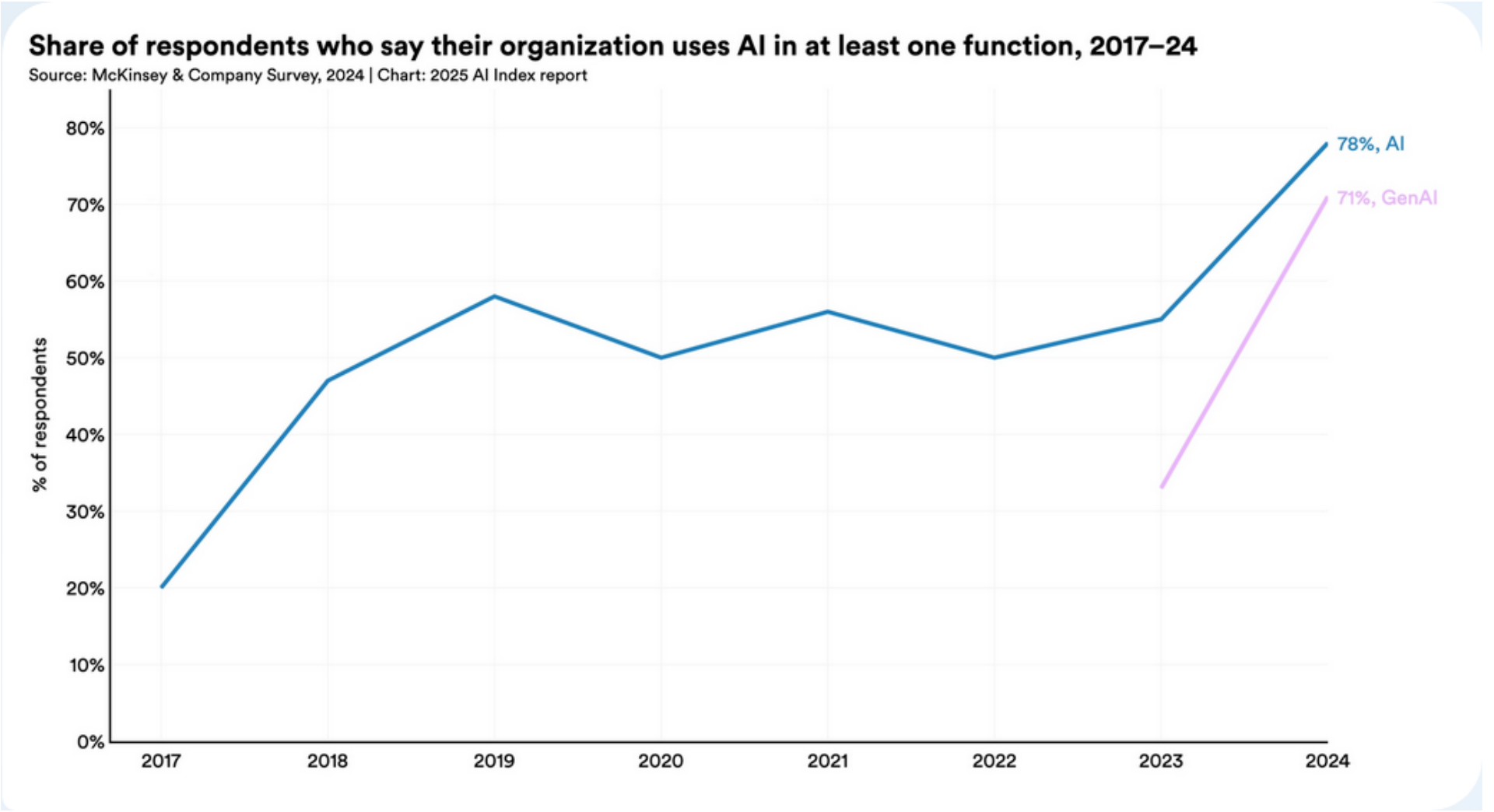
AI Embedded in Everyday Life — Data



Business Investment in AI Hit New Highs

- U.S. private AI investment reached $109.1B
- Generative AI funding remained strong
- 78% of organizations report AI use

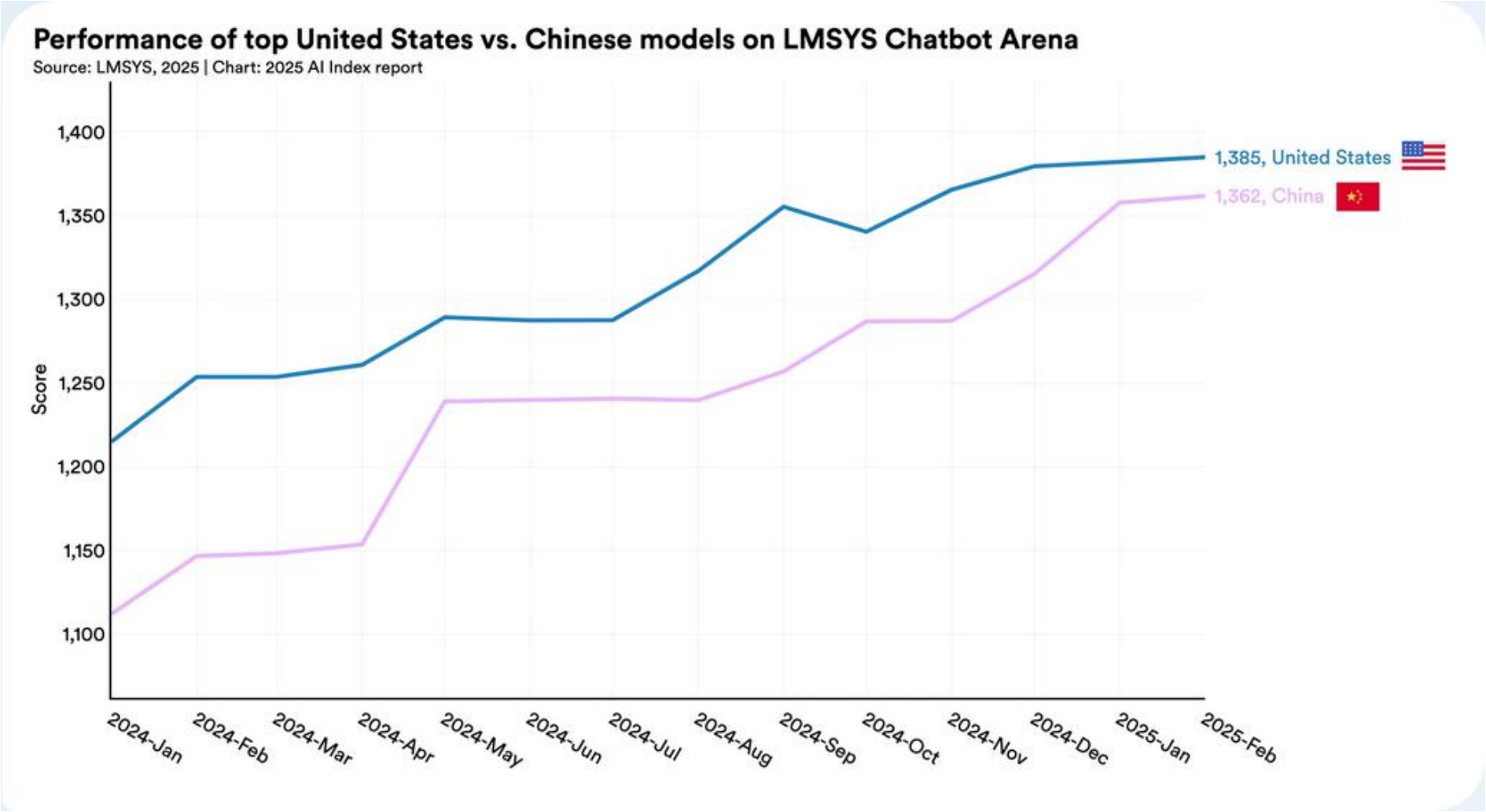
Business Investment in AI Hit New Highs — Data View



U.S. Leads, China Is Closing the Gap

- U.S. produced most notable AI models
- China narrowed performance gaps
- AI development increasingly global

U.S. Leads, China Is Closing the Gap — Data View



Responsible AI Progress Is Uneven

- AI incidents increased faster than safeguards
- New safety benchmarks emerged
- Governance frameworks expanded

Responsible AI Progress Is Uneven — Data View

Reported safety and responsible AI benchmarks for popular foundation models

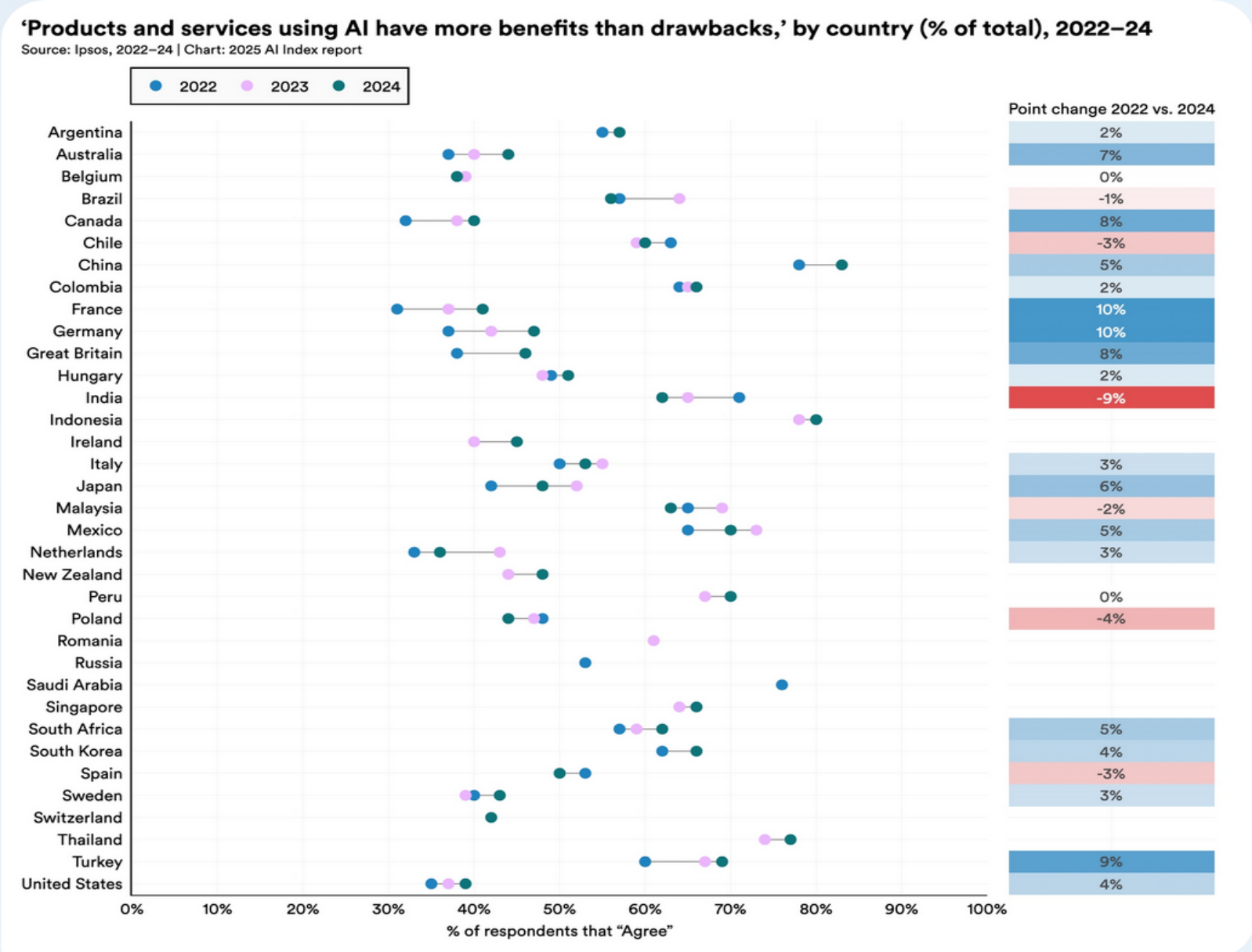
Source: AI Index, 2025 | Table: 2025 AI Index report

Responsible AI benchmark	o1	GPT-4.5	DeepSeek-R1	Gemini 2.5	Grok-2	Claude 3.7 Sonnet	Llama 3.3
BBQ	✓	✓				✓	
HarmBench							
Cybench						✓	
SimpleQA			✓	✓			
Toxic WildChat	✓	✓				✓	
StrongREJECT	✓	✓					
WMDP benchmark	✓	✓					
MakeMePay	✓	✓					
MakeMeSay	✓	✓					

Global AI Optimism Is Rising

- Optimism strongest in Asia
- Sentiment improving globally
- Large regional differences remain

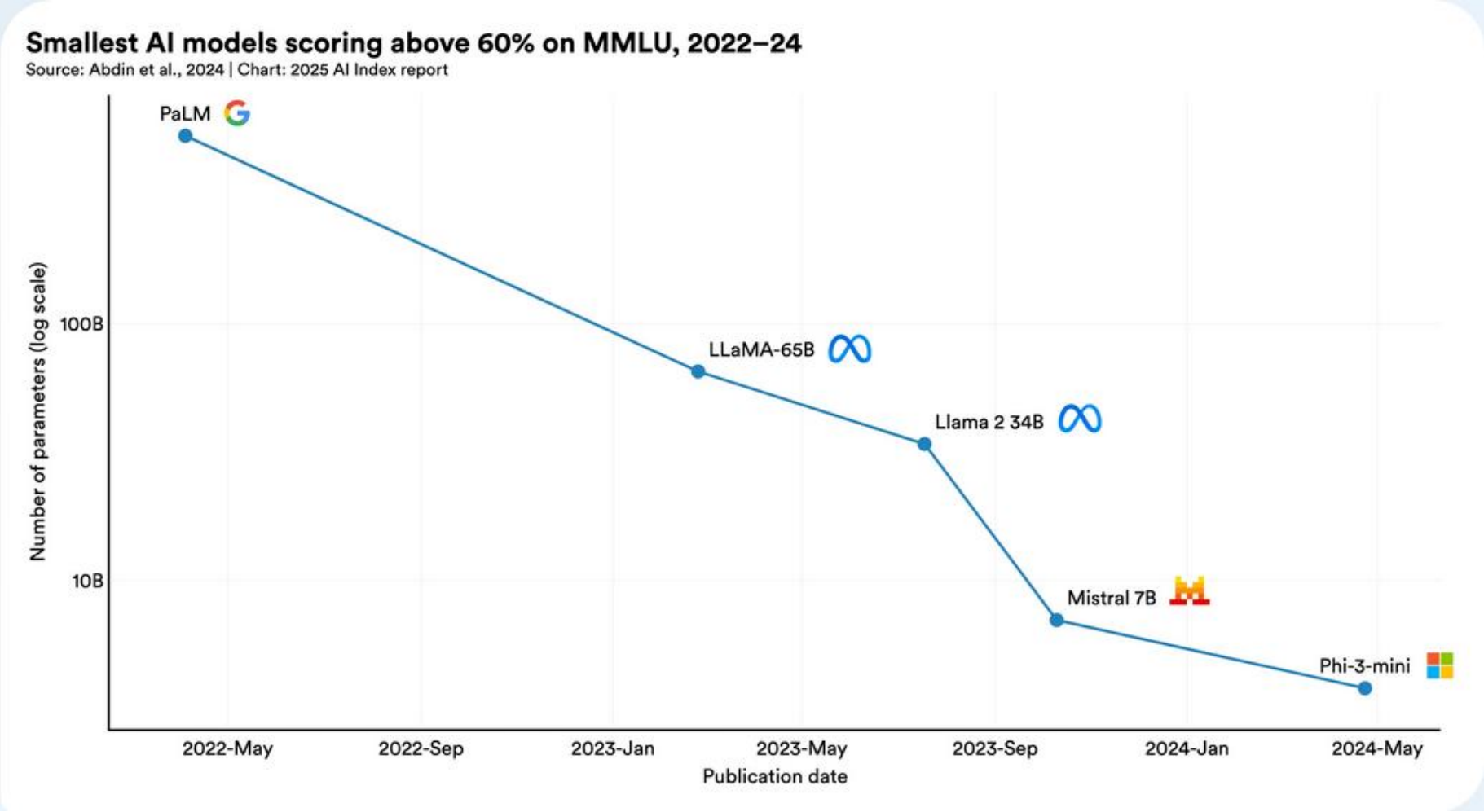
Global AI Optimism Is Rising — Data View



AI Became Cheaper and More Accessible

- Inference costs dropped up to 280x
- Energy efficiency improved
- Open models closed performance gaps

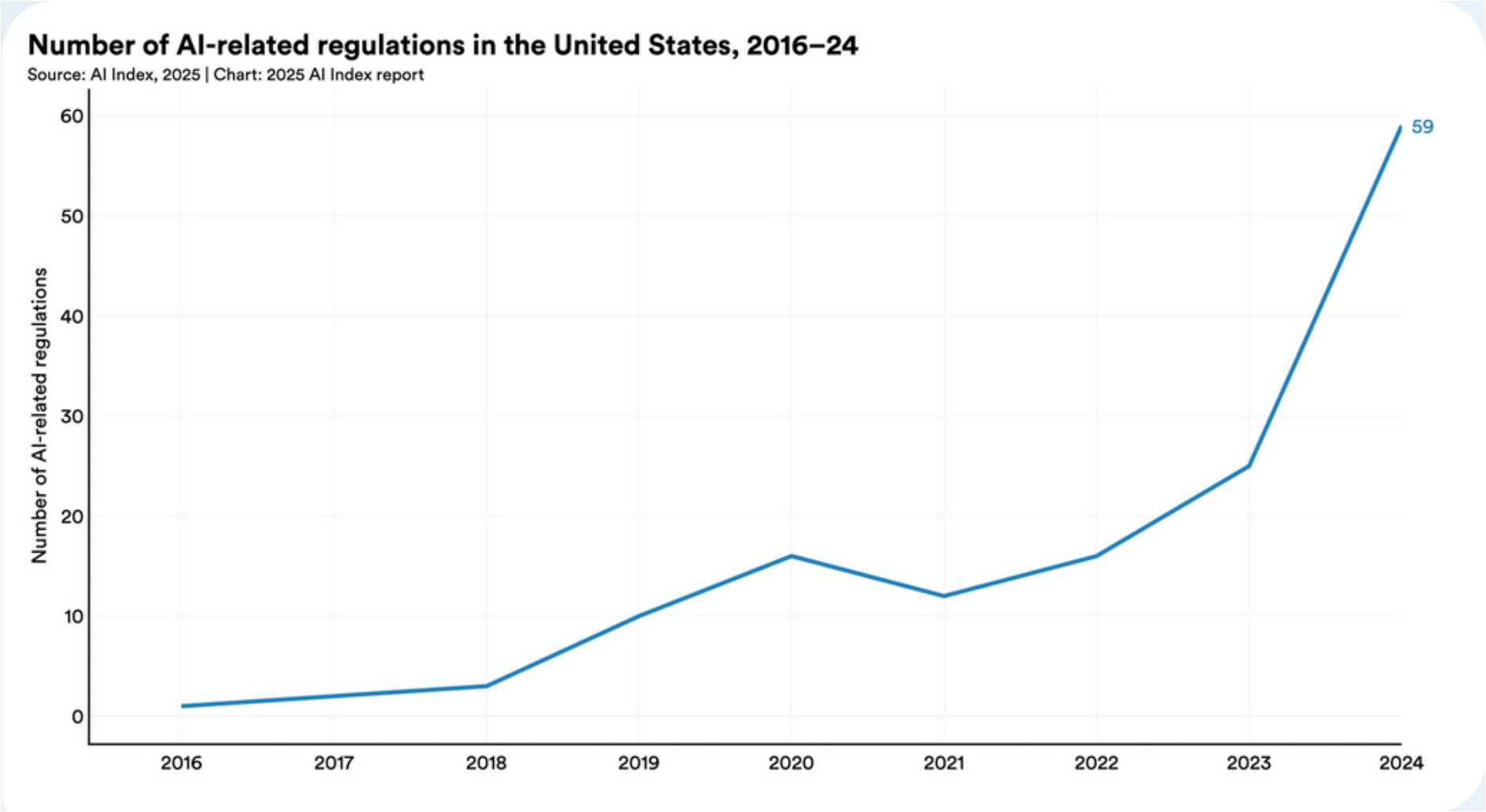
AI Became Cheaper and More Accessible — Data View



Governments Expanded AI Policy

- 59 U.S. federal AI regulations in 2024
- Global AI legislation mentions rose
- Major public investment in AI infrastructure

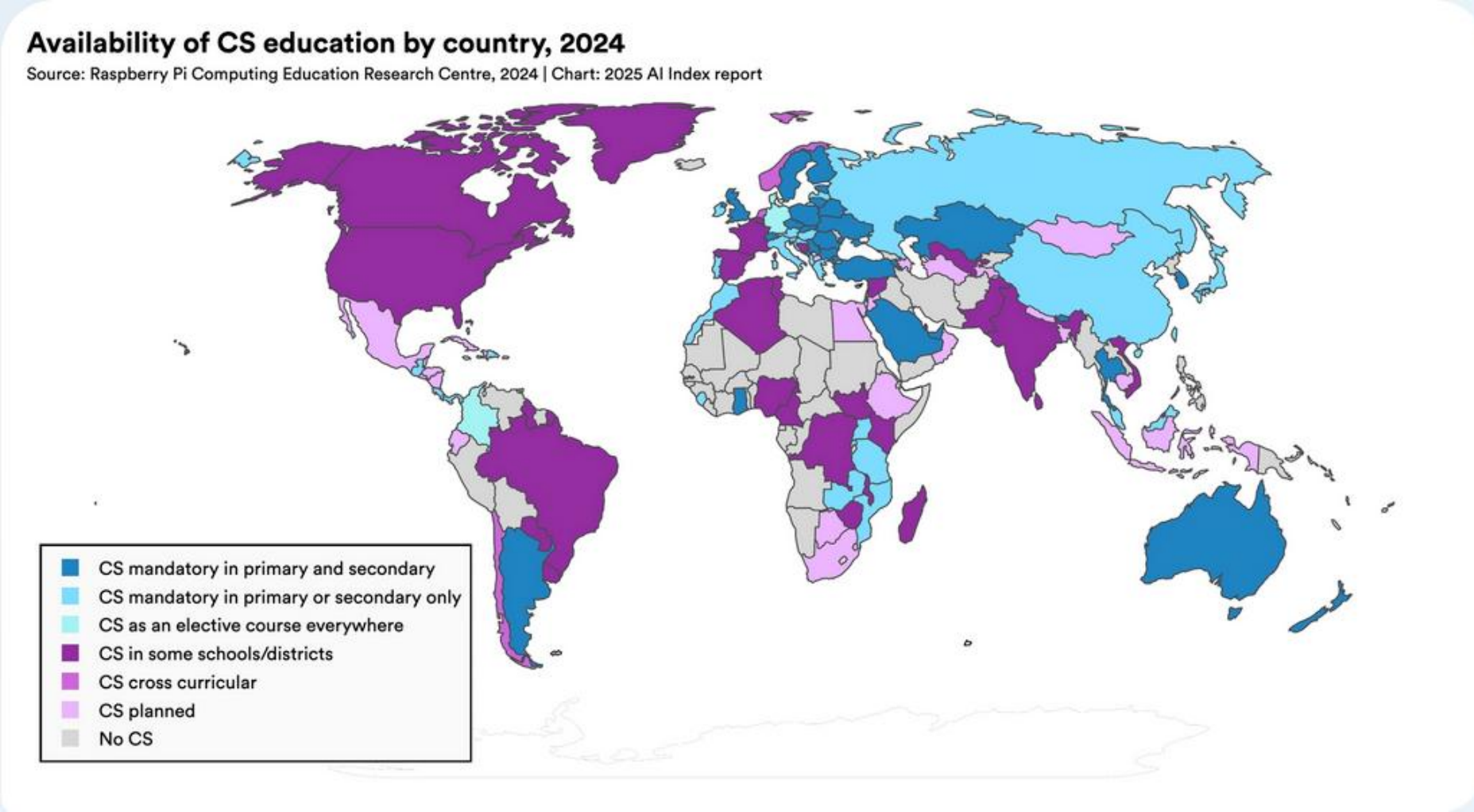
Governments Expanded AI Policy — Data View



AI & CS Education Expanded

- Two-thirds of countries offer AI/CS education
- CS graduates increased in the U.S.
- Infrastructure gaps persist

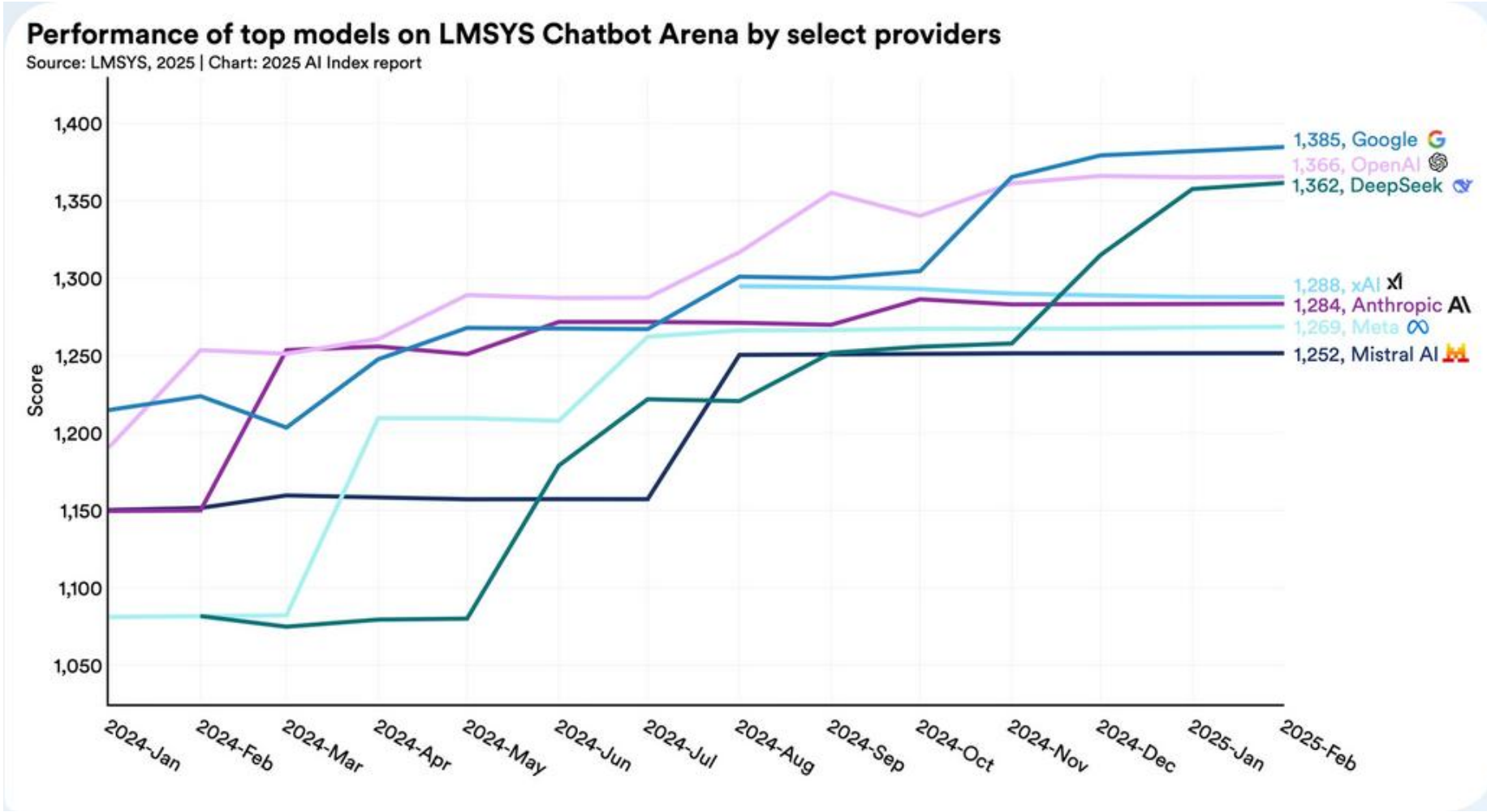
AI & CS Education Expanded — Data View



Industry Dominates AI Development

- ~90% of notable models built by industry
- Performance gaps narrowed
- Compute scale continues to rise - training compute doubles every five months, datasets every eight, and power use annually.

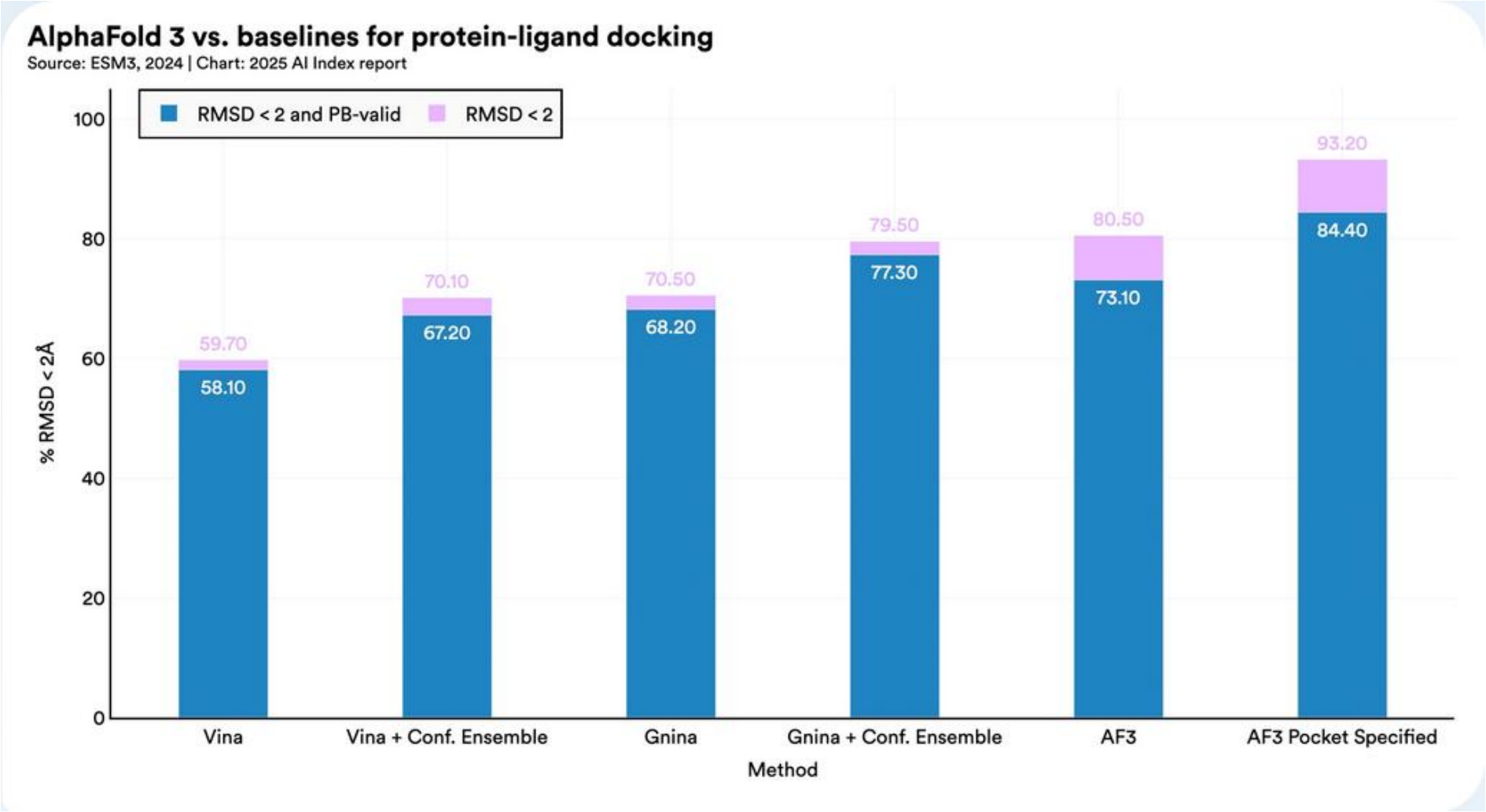
Industry Dominates AI Development — Data View



AI Earned Top Scientific Honors

- Nobel Prizes recognized AI breakthroughs
- Turing Awards highlighted reinforcement learning
- AI reshaping scientific research

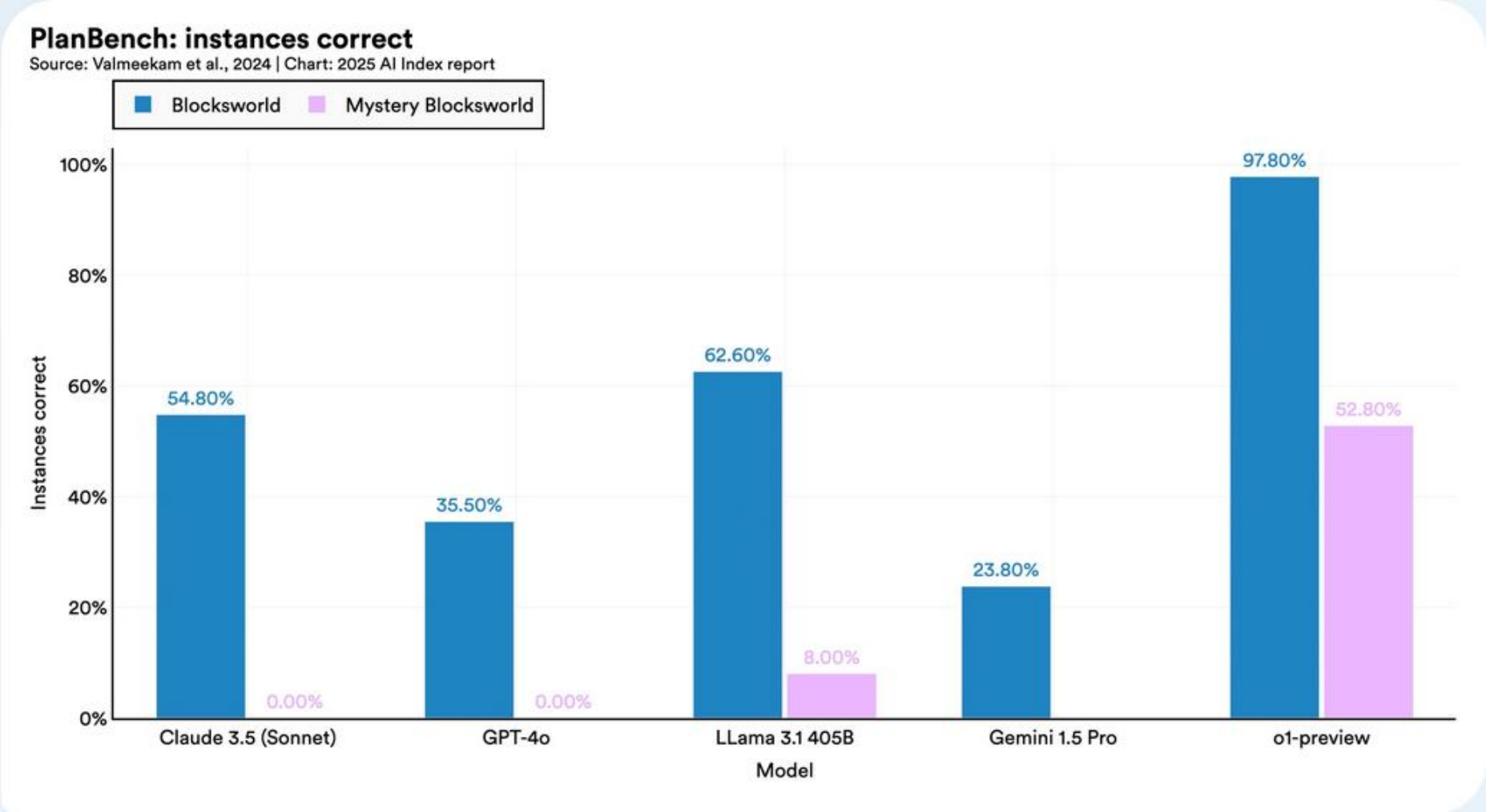
AI Earned Top Scientific Honors — Data View



Complex Reasoning Remains a Challenge

- Strong structured task performance
- Weaknesses in deep reasoning
- Limits high-stakes deployment

Complex Reasoning Remains a Challenge — Data View



Overall

- Lot of developments expected going forward…