State of AI 2026,
02 February 2026

Introduction

This report is intended to summarize the advancements in AI and its media perception in the last 12 months (January 2025 - January 2026): the progress in research, specific AI news in the media and emerging trends (catalysts) that will impact the near future. Our understanding of the AI advancements and the trends allows us to make trend predictions we believe will happen in the next year (last 9 slides). We will review accordingly in one year's time.

2026
Table of Contents

Technologies
Bring the future to humanity

Author

Andrea Isoni
Chief AI Officer

Physicist with a PhD from Imperial College, Former Founders Factory (FF), Chief AI Officer at AI Technologies. Experienced in leading edge AI algorithms and how to translate them into business applications (including pioneering Human Machine Interface systems). Advisor to Waed Saudi Aramco and AI writer ('Machine Learning for the Web' published by Packt in English, Chinese and Korean and more than 10,000 copies). Committee member of ISO and IEEE for AI standards and International AI speaker (London, New York, Dubai, Saudi Arabia). AI Newsletter 'Thoughts about AI by a Human' (3.1K+ subscribers).

TECHNOLOGIES
Bring the future to humanity

Who read the State of AI (shortlist)

Who read the State of AI (audience)

The State of AI Report is designed for a diverse audience of strategic decision-makers, technologists, and investors who require a comprehensive yet accessible overview of the current AI landscape. Its primary readers include C-suite executives, policy makers, and researchers who rely on the report to make sense of the complex web of current ("contingent") events, breakthrough research, and market dynamics without getting lost in the weeds. The report analyzes the forces at play—both the long-standing "secular" trends driving fundamental change and the emerging "catalysts" sparking immediate disruption—to help leaders anticipate how AI will shape the near future of society, industries, and consumer adoption. Each scenario presented is governed by a specific logic framework: Secular Trend + Catalyst → Contingent Event + Prediction. Crucially, the analysis is not intended to be an exhaustive encyclopedia or an overly technical manual; rather, it provides just enough detail to ensure readers understand the mechanics and implications of each scenario discussed. Ultimately, the report's predictions (whether right or not) serve as a guide, giving readers a ballpark understanding of how AI is likely to impact life and business in the year ahead. For more details on this methodology, see "How to Read the Report" on page 7.

Disclaimer: the content of this report is for education/information only, not financial, legal, or tax advice.

2026
Notes on the Report

*

This year's data is considered until 30 January 2026: so from January 2025 to January 2026 included.

*

AI Leaderboard used: https://huggingface.co/spaces/ArtificialAnalysis, https://lmarena.ai/?leaderboard, https://opendfm.github.io/MULTI-Benchmark/#leaderboard, https://vlaleaderboard.com/ https://artificialanalysis.ai/leaderboards/models, https://www.wolfram.com/llm-benchmarking-project

*

Definitions (only very technical ones. It is expected the reader knows what GPU means etc.):
-pre-training: a process where a model is trained on a large, general dataset to learn general features.
-mid-training: additional training with synthetic data, optimizing data mixes, domain-specific data and long-context training stages.
-post-training: set of processes and techniques applied to a model after it has been initially trained on a dataset. It focuses on refining and optimizing the model to improve performance, efficiency and meet specific practical requirements.
-multimodal: AI systems that simultaneously process and integrate various data types.
-SOTA: state of the art, it refers to a model achieving best scores on benchmarks.

*

The BONUS prediction is half a joke, half true. We believe many more will grasp some coding terms and that may naturally spill out into common conversations.

AI
Technologies

How to read the Report

> ## 'secular' trends + 'catalyst' trends → 'contingent events' + trend prediction

* 'Events' means contingent events or reasons that are happening or just happened this year (left box). They are 'signals' that the prediction (detailed on the right box) may likely happen soon.

* 'Catalyst Trends' means trends and events developing in the last 5 years or less. Each trend has immediate effects, it may not last and it can accelerate or slow down depending on political or other reasons. These trends impact some predictions (indicated with numbers at the bottom of each trend prediction page).
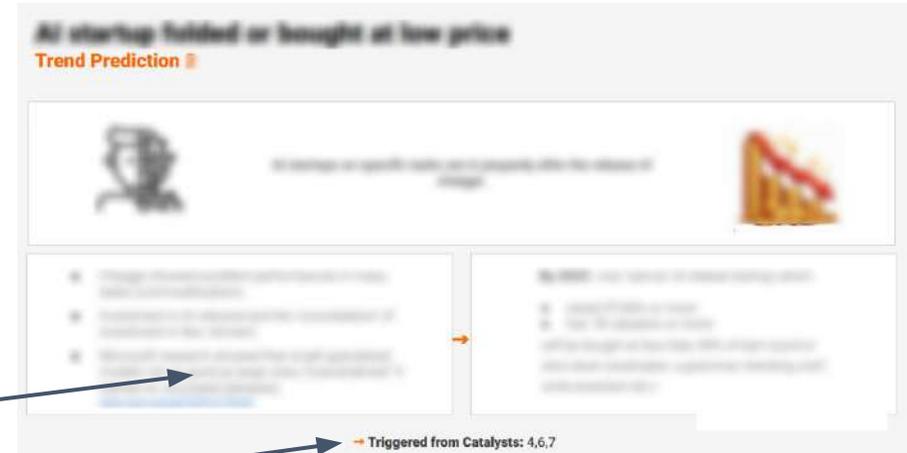
* 'Secular Trends' means trends that are developing in the last decade or more. Each trend is slow to take effect but durable over time. These trends are always impacting predictions (indicated with numbers at the bottom of each trend prediction page).

Pages' guide

- Secular trends from page 17 to page 21.
- Catalyst trends from page 22 to page 32
- Trend predictions, from page 33 to page 53. We consider and give predictions for the following sectors: hardware progress, industry adoption, consumer adoption, cybersecurity, agents and robotics. Each page describes what is actually happening now, references the catalyst trends acting on the sector and then our educated guess on what is supposed to happen in the next year.

A **TECHNOLOGIES**
Bring the future to humanity

2025
Catalyst Trends / Checks

Download 2025 report here

01
Governance AI: EU AI regulations are coming in 2024

01
Happening. Regulations are coming but industry reaction is slower than expected.

02
Synthetic VS human data generation

02
Happening. Going forward there is a hybridization of data (synthetic/human). Renamed to 'Human, synthetic and hybrid data generation'.

03
Cloud and data centers growth

03
Happening. Due to AI adoption or foreseen adoption, new data centers were planned. Sovereign trend also helped.

04
Energy supply and consumption

04
Partially. Overall, energy demand is growing but only in certain regions and the grid is the main issue. Renamed as 'Energy supply and demand imbalances'.

05
Crypto Ecosystem Development (and AI)

05
Happening. As the crypto ecosystem is developing, some blockchain projects merge with AI.

06
Supply chain localization, sovereign cloud and workforce

06
Happening. US and other countries tend to build factories within their territory or near their borders. Also the sovereign trend incentivizes localization.

A TECHNOLOGIES
Bring the future to humanity

2025

Catalyst Trends / Checks

Download 2025 report [here](#)

07	**07**
VC investments in tech	Happening. There is a particular concentration in mega rounds.
08	**08**
Labor shortages	Happening. However, the gravity and the industries involved may be different than expected.
09	**09**
Public spending & GDP/debt ratios	Happening. Debt worldwide keeps rising as does public spending. Renamed to 'Public spending & GDP/debt ratios growth' .
10	**10**
Virtualization of living	Happening. Especially younger generations spend more time in virtual environments. Now with GenAI tools, we also automate personal processes: renamed to 'Virtualization and automation of living'.

2025 Predictions

Download 2025 report here

Outcomes

01

The best 'overall' pre-trained model will be by a BigTech or Lab (>$100M in funding) with 40%+ synthetic data, non-transformer architecture, >$40M training cost, and scores at most 15% better than smaller models. Within 6 months, the result will be replicated by open source models.

01

Almost. The best overall is (family) Gemini 3/GPT 5.2. Best score ~92%, from Big Tech and proprietary first, had definitely a final run $40M+ and used 40% synthetic data (speculative). Many 'cheaper models' within 15% (80%+ score). They are not (speculative) from a mixture of (hybrid) architecture. Although many competitive hybrid models emerged as expected (Nemotron 3, Zamba, Minimax 2 etc.) these were not top models architecture. To be 100% right, by May 26 an open source model should reach Gemini 3 performance (very likely).

02

A top performing 'post-trained' model will rank top 2 on specific benchmark hard to reproduce with a 'mixture/hybrid' of post-training techniques. It should use 'proprietary data' and by either a 'not' foundation AI model lab/BigTech or open data by a community in blockchain..

02

Almost. DeepSeekMath-V2, a V2 specific 'post trained' model (base is DeepSeek-V3.2-Exp-SFT) by a top lab, DeepSeek, reached 'gold-level scores on IMO 2025 and CMO 2024 and a near-perfect 118/120 on Putnam 2024' using different post training techniques (first 'fine-tuned on reasoning data related to mathematics and code' then GRPO on verified data). Wrong on data would be proprietary.

03

A data labelling firm will enter in the race with a 'data proprietary' model ranking top 10 in the benchmarks.

03

Yes. Amazon has a data labeling service (SageMaker Ground Truth), Nova 2.0 Premium is Top 11 on Artificial Analysis and Top 9 in Vellum.ai. Note: Models as 'families' (i.e. OpenAI Gpt 5.2 high/medium, 5.1 etc.) occupy the #1 ranking spot.

04

A GLP-1 variant which AI software/model contributed to optimise/discover/design will start FDA approval process.

04

Half Right. MindWalk AI found a 'novel dual-pathway regimen' to both keep benefits of GLP-1 drug and enhance longevity. ImmunoPrecise Antibodies AI designed a GLP-1 receptor agonist achieving 'in vitro' results similar to commercial GLP-1 drugs. It's unknown if they started the FDA approval process.

2025

Predictions

Download 2025 report [here](#)

Outcomes

05

On the top 5 models' leaderboard for video or images there will be a new Lab/company coming from an emergent economy country (excl. China). ✔

05

Almost. Yes. Kandinsky 5.0 (SberAI, Russia) was released in November 2025 and was ranked first in open weights leaderboard, 11 overall ([LMArena](#)).

06

A text to image model not diffusion based will achieve SOTA Top 10 on the leaderboard. ✔

06

Yes. Flux 2, Rectified Flow Transformer from BlackForest (Germany) was released on December 25 reaching top 4 (LMArena). Also Qwen Image 2512 (Alibaba), Hunyan Image 3 (Tencent), Seedream 3 (Bytedance) all use a Multimodal Diffusion Transformer (MMDiT) architecture instead of a traditional Diffusion.

07

A mobile video game built with genAI tools will be in the news. ✔

07

Yes. Aliens vs Zombies: Invasion reached the news claiming '4x less time' in development due to AI tools. Clair Obscur: Expedition 33 reached the news as it was stripped of the Indie Game prize due to the use of GenAI.

08

Nvidia will remain the go-to chips for large clusters and big orders. ✔

08

Yes. Plenty of new orders for all around the world including with governments (UAE, Vietnam etc.).

09

AMD will have 10%+ of the market for single or small GPU configurations. ✘✔

09

Half Right. Especially on PC graphics, Nvidia had about 94% at the end of December 24 (AMD 6%). During the year AMD recovered to about 8% until Q3 25. We have no data for Q4 25 to assess the prediction and even Oracle in October 25 announced a deal of 50k AMD GPUs. I give myself half right mark on this in fairness (I assume the year will close at least at 8% market share).

2025

Predictions

Download 2025 report here

Outcomes

10

A new chip player (not Intel, Nvidia, AMD) will announce orders worth >$500M. ✓

10

Yes. Cerebras announced a deal with Perplexity targeting the $100B search market (VentureBeat). Groq secured a $1.5B commitment from Saudi Arabia. Blaize had orders for about $400M.

11

A safety AI startup will be acquired by a large tech player for more than $100M. ✓

11

Yes. SentinelOne's $250 million acquisition of Prompt Security. Cato Networks acquired Aim Security for an estimated $300−350M.

12

There will be a serious remote vehicle attack using AI which will raise awareness on cars vulnerability. ✗

12

No. Simply did not happen. I still think it is possible in the near future.

13

AI regulations only related to automotive overall will increase in western countries and China. ✗

13

No. On the contrary, Trump may also de-regulate the autonomous vehicles industry (AutoNews).

14

At least one new product for watermarking combining blockchain, AI and cryptography will be presented but it will fail to make significant traction. ✓

14

Yes. Nethermind's zkWM, IBM TabWak most notable products. No signal of traction.

15

ChatGPT and competitors combined are used at least once by 40% of the US workforce. ✓

15

Yes. According to a Stanford study, USA workforce adoption was 43.2% as of March/April 25 45.9%, June/July 25 and 35.9% December 25. Considering the prediction was to use AI at least once, it was right.

A
TECHNOLOGIES
Bring the future to humanity

2025
Predictions

Download 2025 report here

Outcomes

16

Top 3 non first world countries with large, outsourced workforces (like India etc.) will have a combined 25%+ of traffic of chatGPT and competitors. ✔

16

Almost. In August 25, India, Brazil and Indonesia had 17.3% traffic (Visual Capitalist); in December 25, India, Brazil and Mexico had 25.9% (FirstPageSage).

17

At least 50 agents' accounts (>100K followers) will be built on blockchain. ✘

17

No. On X there are few (aixbt, terminal of truths, Moby, etc.) I cannot claim we have reached 50 agents.

18

Adoption of blockchain based agents will be higher than OpenAI GPTs. ✔

18

Yes. OpenAI and Anthropic traction on their 'apps platform' remained low. On blockchain agents platforms, like Virtuals.io
, agents, even if few days old, reached 24h trading activity of $20–100K.

19

KPO will grow faster than predicted (Precedence research: $110.69B in 2024 expected a 13.7% growth). ✘

19

No. Precedence research kept the expected CAGR at 13.7%.

20

We keep seeing 'small/narrow' AI solutions the most adopted in specific functions while large ('sophisticated') models adoption in processes will be slow especially in regulated industries. ✔

20

Almost. According to Databricks 77% of users choose models with 13B parameters or fewer. According to McKinsey industries like Financial Institutions, Insurance, Engineering and Construction, Transportation and Retail. Surprisingly, Healthcare (regulated industry) had a certain degree of adoption as well as certain support functions in Insurance.

AI TECHNOLOGIES
Bring the future to humanity

2025 Predictions

Download 2025 report here

Outcomes

21
US admin workforce will not lose more than 1% due to AI.

✅

21
Yes. According to the [Federal Reserve Bank](#), there was no observed loss of workforce for Office and Administrative Support Occupations.

22
Average AI roll out will still be around 8–12 months in large companies.

❓

22
As of January 26, I could not find reliable sources that confirm or deny the prediction.

23
US will establish a nation-wide AI education program to reskill the workforce.

✅

23
Yes. April 25, Trump announced a youth AI education policy and program ([White House](#)).

24
VC interest in the early stage will grow 15%+ compared to previous year (rounds <$15M).

✅

24
Yes. Per [Crunchbase](#), early stage funding is up 36% compared to 2024.

25
A startup with data labelling capabilities reaches $30B+ valuation.

✅

25
Almost. Scale AI is valued at $29B. Surge AI has a valuation between $15–30B.

26
New data centers construction due to AI will cool off but grid stability will emerge as the bottleneck.

✅

26
Partially. Spain blackouts in April possibly triggered many media to talk about the problems ([The Conversation](#)). Some data centers have been paused (also due to environmental concerns), while some were announced.

27
2% of news X accounts (>100K followers) will be agents

❓

27
As of January 26, I could not find reliable sources that confirm or deny the prediction.

Technologies
Bring the future to humanity

Best AI advancements

01

'DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning'--> It is so important that it was included even last year. Open weights model reaching SOTA with less resources. Seminal for Group Relative Policy Optimization (GRPO, removed separate critic model need) and RL with Verifiable Rewards (RLVR).

02

'Less is More: Recursive Reasoning with Tiny Networks' --> Tiny Recursive Model (TRM) 7M parameters, that performs even better than LLMs at solving logic puzzles (e.g. reasoning like the ARC-AGI benchmark).

03

'Block Diffusion: Interpolating Between Autoregressive and Diffusion Language Models'--> First successful approach to combine Diffusion and Autoregressive (e.g. Transformer) models.

04

'Training Agents Inside of Scalable World Models'-->Introduced Dreamer 4, training a video world model with just 1 GPU A100 for 17 days, the first to reach diamonds in Minecraft using offline data only. As a comparison OpenAI's VPT used 720 GPUs and ~2900 days in 2022. The approach may work for a robot.

05

'Qwen3 Technical Report'--> open model replacing Llama as the default model to build on. A family of models (different size), it combined Gated DeltaNet + Attention, replacing the standard attention, used Ultra-Sparse MoE (but also has dense versions of models to ease fine tuning).

A Technologies
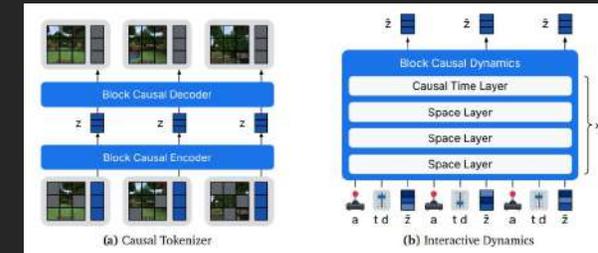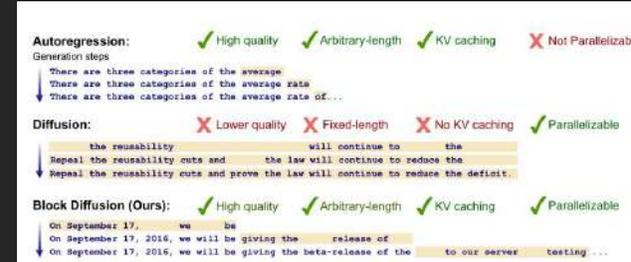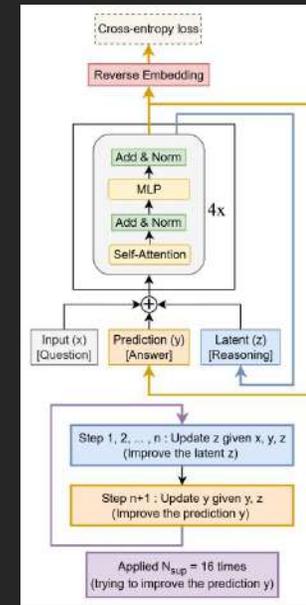
Best AI advancements

06

'SAM 3D: 3Dfy Anything in Images' --> Open weights model able to reconstruct 3D objects from a single image even if cluttered. It also showed that there is a 'scaling law' as LLMs: the more data the better results.

07

'Hybrid Quantum Transformer for Language Generation' --> The first hybrid quantum-classical Transformer for language generation requiring only 10 qubits and approximately 80 quantum gates—can effectively replace over 10% of the classical parameters in a 150M-scale model without any decrease in performance.

08

'LeJEPA: Provable and Scalable Self-Supervised Learning Without the Heuristics'--> First truly multimodal (image, text, or video) embedding representations that are scalable across dataset and architectures.
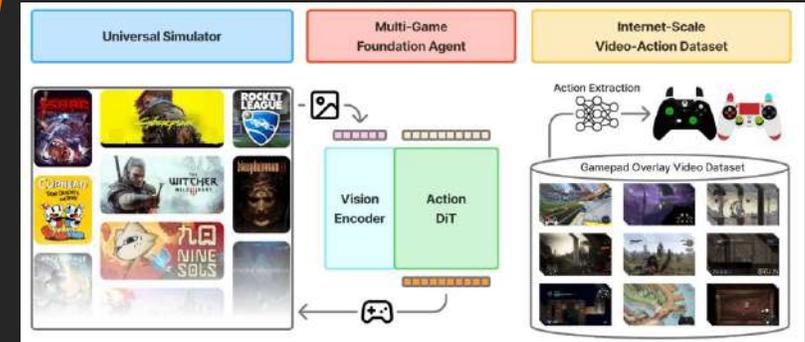
09

'NitroGen: An Open Foundation Model for Generalist Gaming Agents'-->Based on GROOT N1.5, it plays 1000s of games (racing, RPGs etc.). Achieved ~52% higher success than trained models in unseen games.
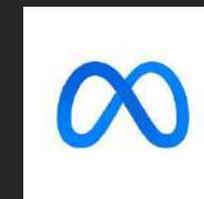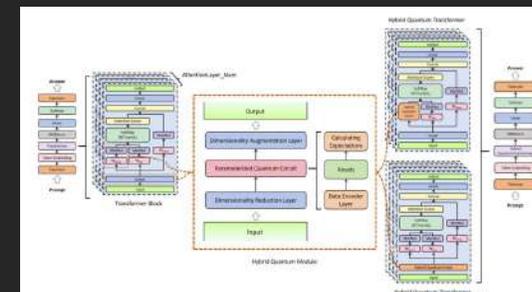
10

'The Art of Scaling RL Compute for LLMs'--> It unifies the different RLHF, RLAIF, and Reasoning into a predictable law and it creates a standardized, open-source recipe (ScaleRL) which can be used to predict RL training hours into expected performance of the final model.

11

'It's All Connected: A Journey ....'--> it proved that Transformers, RNNs, and SSMs (like Mamba) are all actually the same architecture (Associative Memory Machines). It shows why a model forgets.

Human Machine Hybridization

Humans are progressively integrating with technology, hardware and software.

- Mobile phones, search engines (2000s)

- Smartwatch (2010s), iPods

- Smart Home (2015s)

- **Now (2020s), (chatgpt) AI assistants, VR headset, good proportion of GenZ watches Youtube at 1.5x/2x**
- **Assisted wearables (2030s) Nike Amplify**

- Future.. wearable pins, Brain Computer Interface

TECHNOLOGIES
Bring the future to humanity
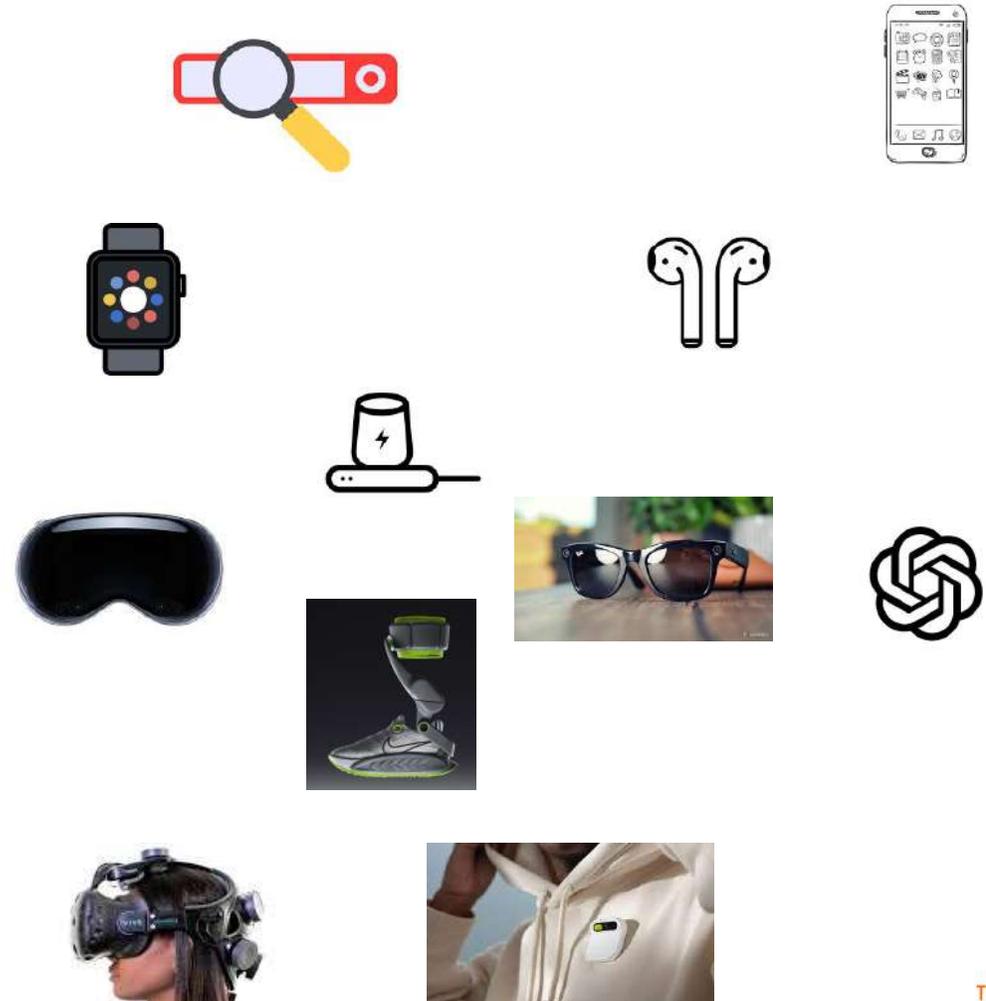
Secular Trends 2
Aging

Most countries have an increasingly older population with birth rates less than 2. Age dependency ratio (bottom right) is rising in Western countries but stable or decreasing in China and low Income countries.
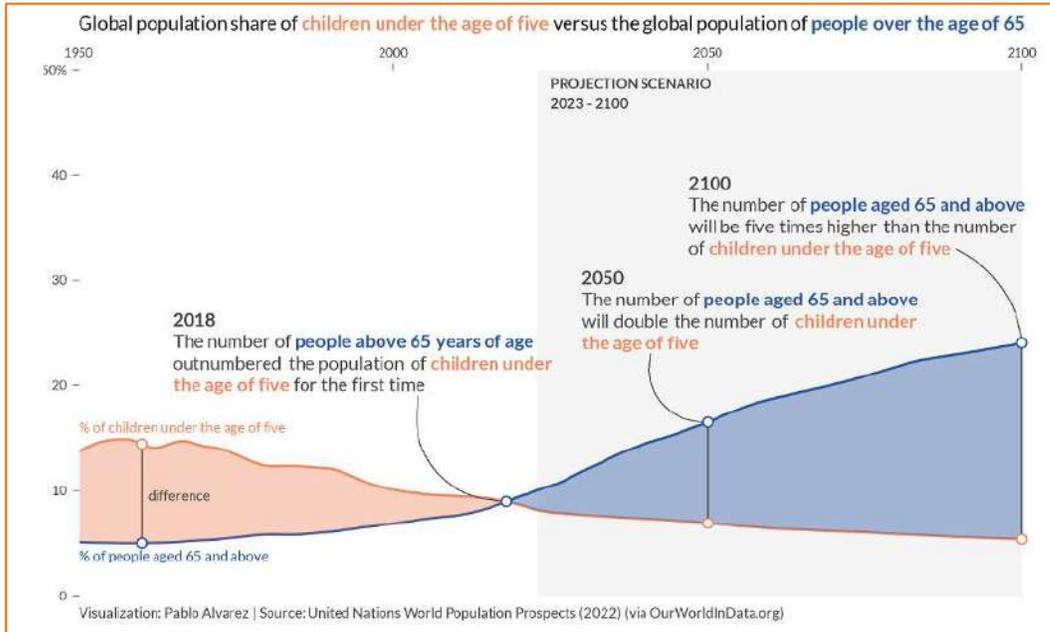
Obesity and Health

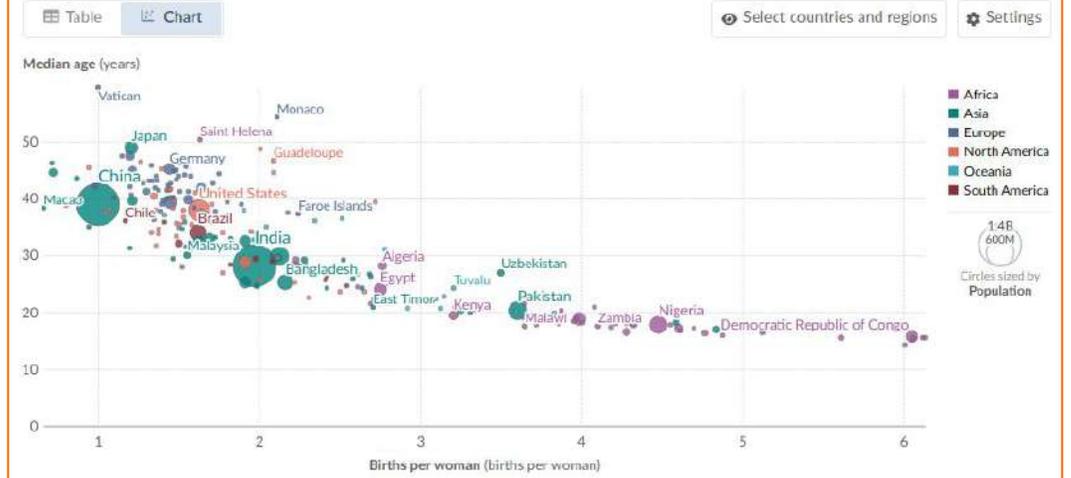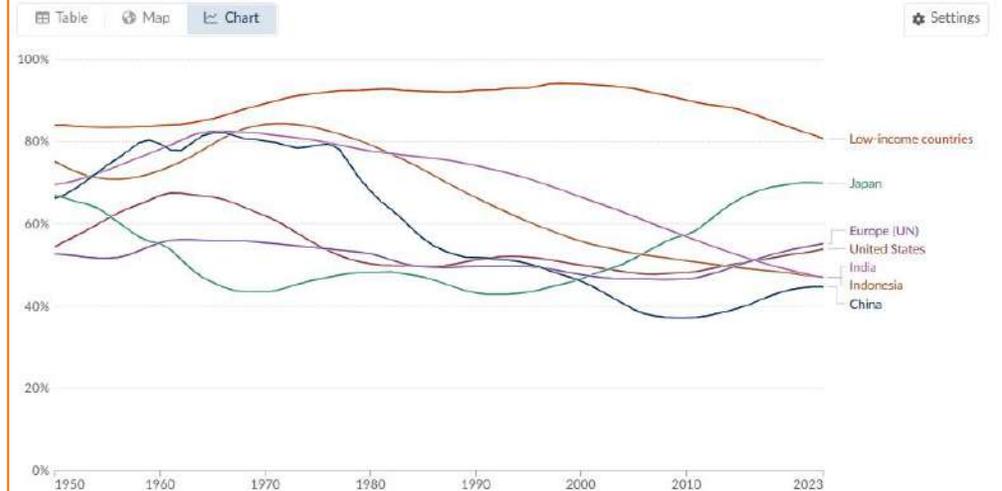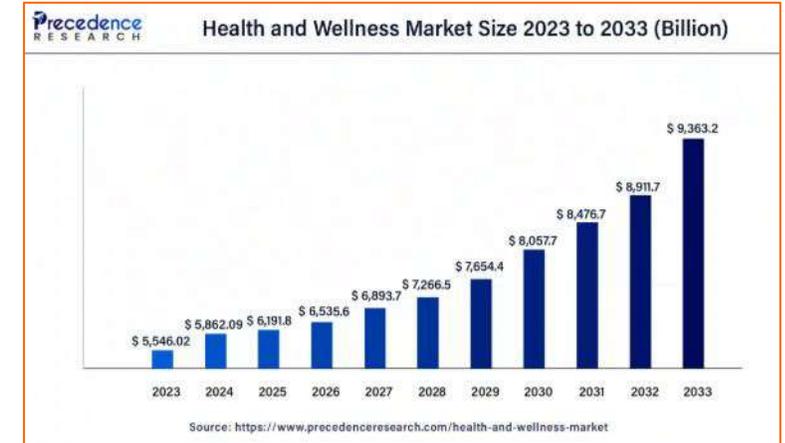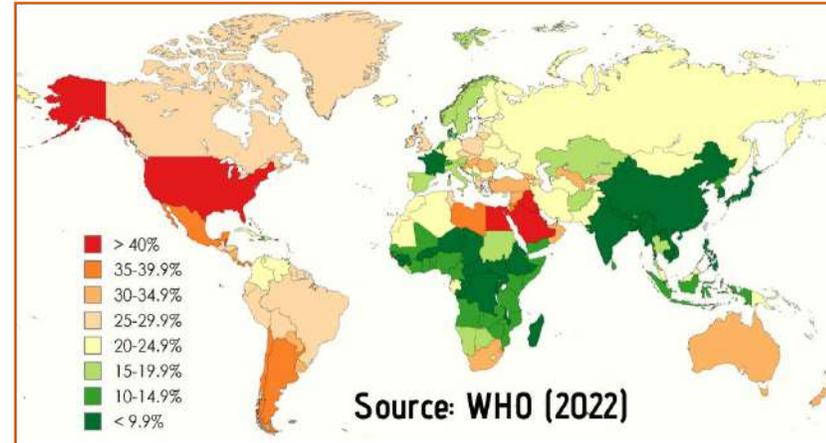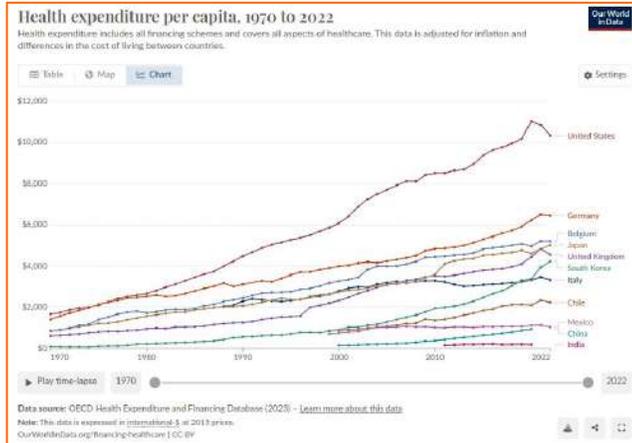Worldwide population spent on healthcare is increasing due to aging and obesity:

- From 1990 to 2022, global obesity rates more than quadrupled in girls (1.7% to 6.9%) and boys (2.1% to 9.3%), with increases seen in almost all countries. In adults, obesity rates more than doubled among women (8.8% to 18.5%) and nearly tripled in men (4.8% to 14.0%) between 1990 and 2022. (NCD Risk Factor Collaboration, NCD-RisC in collaboration with the World Health Organization, WHO).

- 79% of overweight adults and obesity will live in Low- and Middle-Income Countries (LMICs) by 2035. It is projected that the number of adults living with obesity will rise to 1.53 billion in 2035 (worldobesity.org).

- Qatar, Egypt and USA have obesity rates> 40%, China >8% and Japan has the lowest rate of obesity among the OECD member countries at 3.2% (WHO, 2022).

- Health expenditure per capita over the last decades always increased (OECD).

- Global health and wellness market size keeps growing and it accounted for USD 5,862.09 billion in 2024 and is expected to be worth around USD 9,363.2 billion by 2033, at a CAGR of 5.34% from 2024 to 2033. (Precedence Research).

Asian growing economies

Percentage Share of Global GDP: 2030
Using 5-year CAGR for projections

Asia (58%)
Europe (15%)
North America (13%)
South America (7%)
Rest of World (7%)

GDP in Purchasing Power Parity terms with added estimates for the size of the Informal economy and adjustments for out-of-date GDP base year

Map 2: Projected GDP in PPPs in 2050

Lower GDP in PPPs — Higher GDP in PPPs

Source: PwC analysis

2035 projection of top 20 economies by nominal GDP (US$B)

US 44,578.5
Mainland China 38,773.2
India 11,016.8
Japan 8,007.3
Germany 7,381.1
UK 5,340.3
France 4,854.8
Indonesia 3,969.8
Canada 3,507.6
Italy 3,419.5
South Korea 3,154.1
Brazil 3,906.2
Australia 3,043.7
Spain 2,634.1
Poland 1,938.1
Netherlands
Russia 3,570.7
Mexico 2,756.6
Türkiye 2,472.6
Saudi Arabia

Data collected Sept. 1, 2024.
Source: S&P Global Market Intelligence.

defense budget is increasing while GDP/debt ratio is stable, marginally improving:

- Asian share of global GDP is projected to reach around 58% by 2030, up from 55% currently. EU share is expected at 15% and USA 13% (worldeconomics.com).

- S&P Global projects 4.2% expansion in 2025 and 4.1% in 2026, slightly lower than earlier forecasts of 4.4% for both years (S&P).

- By 2035, emerging markets will contribute about 65% of global economic growth. By 2050, emerging markets' old-age dependency ratio is expected to reach 35%, still well below the 50% expected for high-income countries (S&P).

- China's current share of world GDP at PPPs stands at almost 18%, and it is projected to increase to 20% by 2050 while India is projected to rise steadily to over 15% by 2050 (PWC).

- Six of the seven largest economies in the world are projected to be emerging economies in 2050 including China, Japan, India and Indonesia (PWC).

TECHNOLOGIES
Bring the future to humanity

World polarization
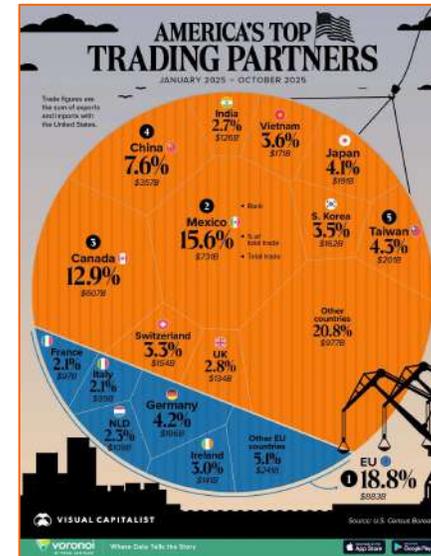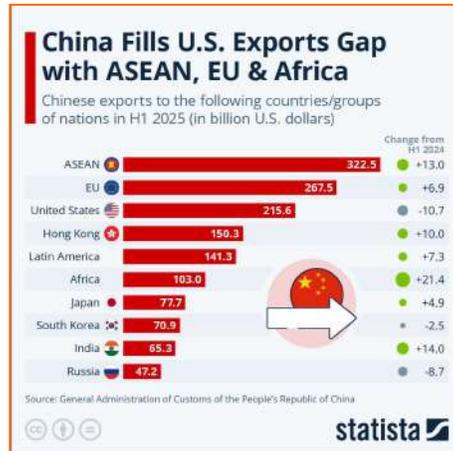
With China competing with US on all fronts, the world is polarized:

- China has been bolstering its investment in the BRICS countries as a counterweight to the G7 (various sources).

- The European Union is navigating a complex position between the US and China, debating whether to pursue equidistance or alignment. Equally, generally Asian and MENA countries and specifically Southeast Asian Nations (ASEAN) countries have consistently emphasized their desire to avoid taking sides in the US-China competition: both countries are ASEAN's largest trading partners (various sources). A recent EU-India trade agreement goes in that direction of balancing USA and China relationships.

- Taiwan faced significant incidents in 2024: 650% increase in cyber attacks on the telecommunications sector, 57% increase on defense supply chain targets and 70% increase on transportation targets. Record number of military aircraft breached Taiwan's ADIZ (various sources).

- Technology has become a key battleground, with both countries implementing restrictions and investing heavily in critical sectors like AI, quantum computing, and semiconductors. Famous examples: the US Chip Act, China (CCP) Action Plan (supply Global South with their AI solution, cautioning Western chip purchases), and US AI Action Plan (deregulation, incentivize energy and compute, faster AI adoption in defense).

- The new Trump administration is expected to continue the tension with China (tariffs, trade wars, chips and other restrictions).

governance AI: EU (and others) AI regulations in 2025

AI Risk Categories

Unacceptable Risk — social scoring, biometric id in public, emotion recognition

High Risk — health, credit scoring, education exams

AI systems with specific Transparency obligations

Limited Risk — customer service chatbot etc.

Minimal Risk — not affected, GDPR etc. in place
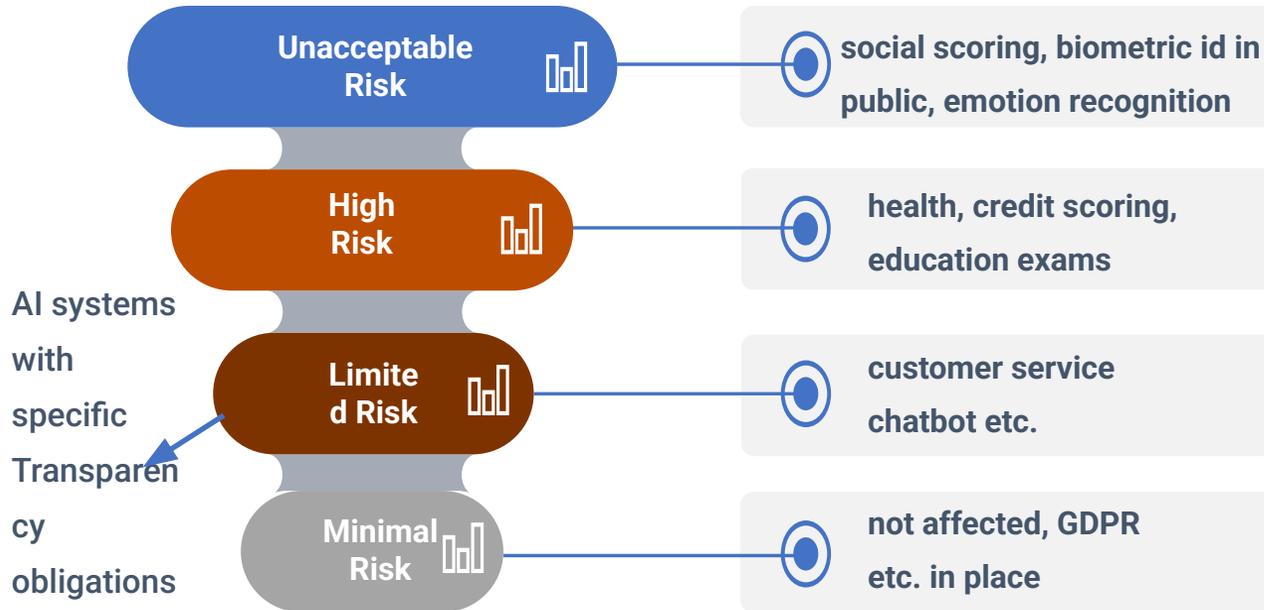
Limited Risk

- **Limited risk regulations (general purposes):**
- **apply from August 25**
- technical documentation available for clients of the AI system
- 'summary' of content and data used in training
- policy in place to respect copyright law

High Risk

- **High risk regulations:**
- **apply from August 26**
- all requirements of limited risks
- AI model evaluation records
- AI model assessment and mitigation of risks
- management and response policy of incidents
- cybersecurity measures and adversarial testing
- **High risk regulations (Annex II): from August 2027 (toys, health devices, cars etc.)**

-Most regulations follow or draw inspiration from the EU AI Act: the focus is on high-risk or 'systemic risk' uses of AI (e.g., healthcare diagnostics, education etc.) to ensure accountability and prevent harm on life. Using AI in this category requires developing and maintaining model evaluation records, incident response and cybersecurity measures.

-The definition of 'risk' may be different in different jurisdictions. Ex: in California is large-scale AI systems ($500M+ in revenue) or pose significant risks (e.g., mass casualties or material damages $500M,+ EU depends on sectors), EU is 'harm to life principle'.

-USA: while there are AI (some strict) laws in some US states, the trend is a 'unified federal AI framework' (in progress).

-China vs EU: China has 'vertical' regulations focused on specific technologies (AI) while EU has a 'horizontal' framework applicable to all industries using AI. Chinese approach requires 'traceability' of output back to source while EU approach requires disclosing data information.

-Light touch regulations: UK, Australia, Israel, Japan, India, Singapore, Russia, KSA and UAE.

A TECHNOLOGIES
Bring the future to humanity

Human, synthetic and hybrid data generation

Human and synthetic data are still increasing but the difference is now blurring: 'hybrid' data is emerging.
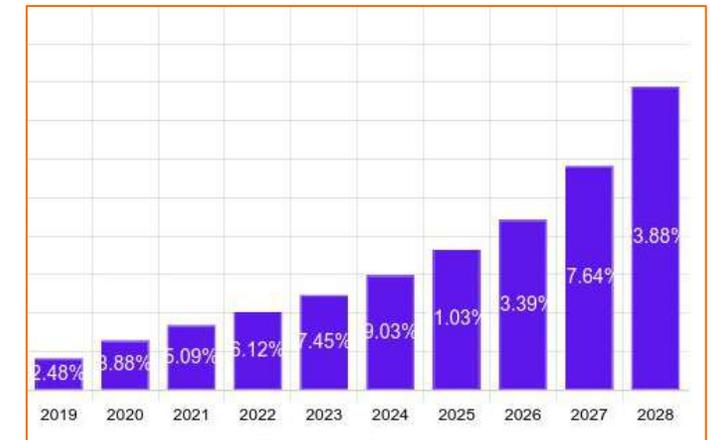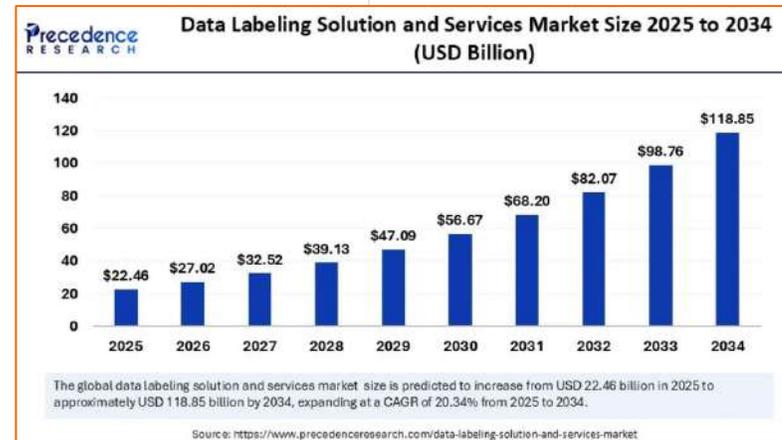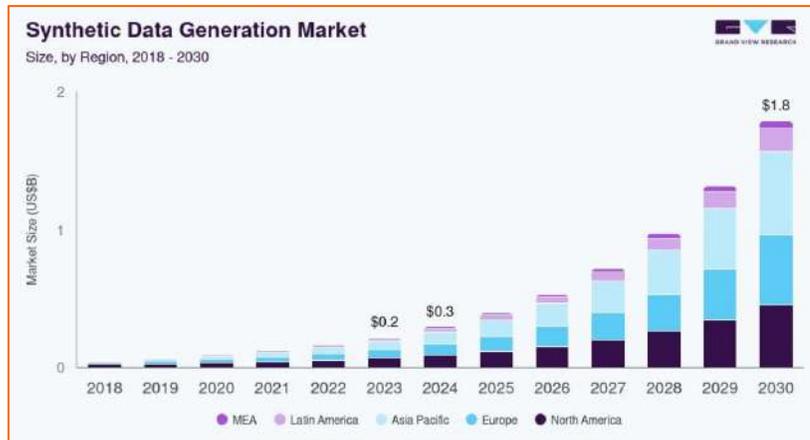
- By 2025, the global volume of data created (includes data that is newly generated, captured, copied, or consumed) is projected to rise to 181 zettabytes. Videos account for over half of internet data traffic. Other estimates suggest it could reach 394 zettabytes by 2028 (digitalsilk.com).

- The data labeling market—critical for "hybrid" datasets where humans refine AI outputs. Global data labeling market was valued at $22.46B in 2025 and is projected to grow at a CAGR of 20.3% to 2034. The market is growing due to the increasing demand for AI and ML across industries. Most of the labeling is outsourced. Scale AI deal with OpenAI confirmed the trend. (Precedence research).

- Various estimates project synthetic data market size will grow at a CAGR of 32-35% between 2025-2030. Tabular and Visual data are expected to grow more than other types.

Cloud and data centers growth

Cloud and data centers are growing and becoming more 'sophisticated' (AI, automation, data platforms etc.) with specialized services on top of the infrastructure. Sovereign cloud pushes local demand but the world still relies on USA.
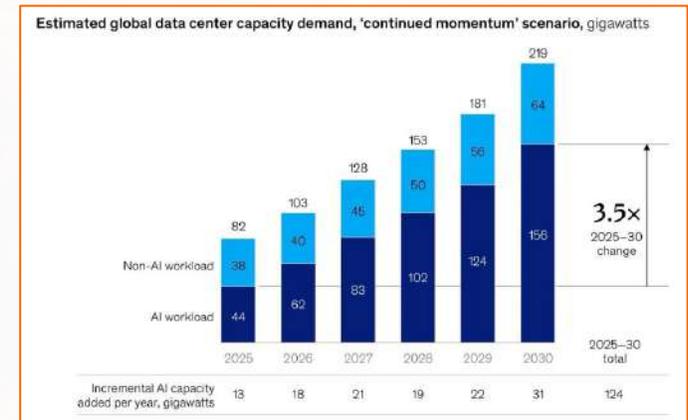
- Global cloud computing market reached $943B in 2025 with BFSI (Banking, Financial Services, Insurance) 23% and government 7%. Software (SaaS) still the largest (53%) but infrastructure as a service (IaaS) will grow faster. North America is 40% (Grandview Research). The global data center market is expected to reach $1.1T by 2035, growing at a CAGR of 11% from 2025 to 2035, faster in Asia Pacific (Precedence Research).

- USA has ~4K data centers, Germany, UK (~500 each) and China (487) (Statista). Ranking countries by TWh : USA 180TWh, China 102 TWh, EU+UK 60-100 TWh and Japan 19TWh (EIA data, 2024). BigTech (Google, AWS, Meta, Microsoft etc.) invested more than $300B in data center expansions and related in 2025. Northern Virginia's capacity is 3,046 MW (Q125), larger than the next 5 markets (Cushman&Wakefeld).

- World dependency: 90% of LatAM traffic goes through USA (Iacnic), ~50% of world hosting is USA (including AWS and Azure) with high concentration of world bandwidth (~20%) in Virginia (Mordor, Cardinalnews). ~54% of total Hyperscale data centers are in USA (Synergy Research).

- Demand for AI-ready data center capacity to rise at an average rate of 33% per year between 2023 and 2030. AI workloads (GW) will grow 3.5X and Non-AI 1.7X by 2030. In data centers, Arm CPUs will have 70% market share vs 30% x86 Intel (mckinsey, semiengineering).

Energy supply and demand imbalances
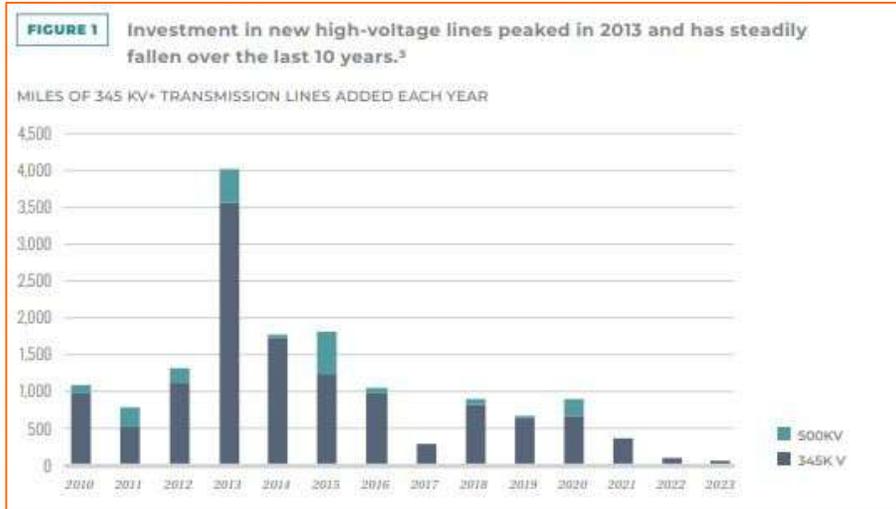
AECG
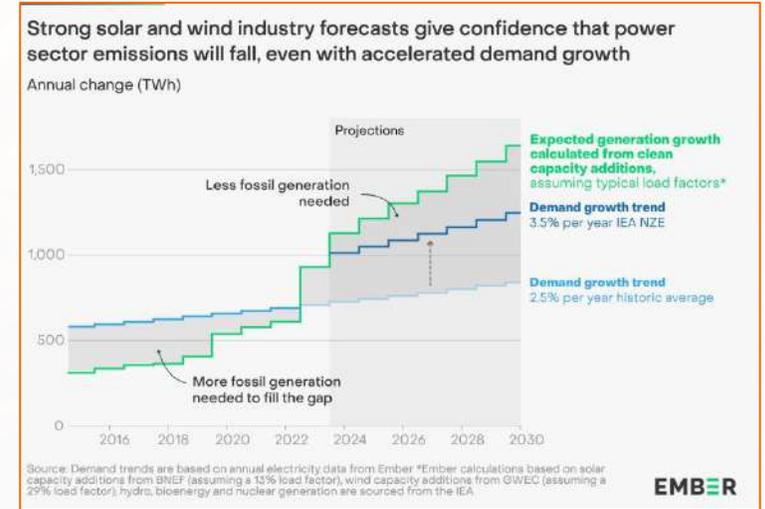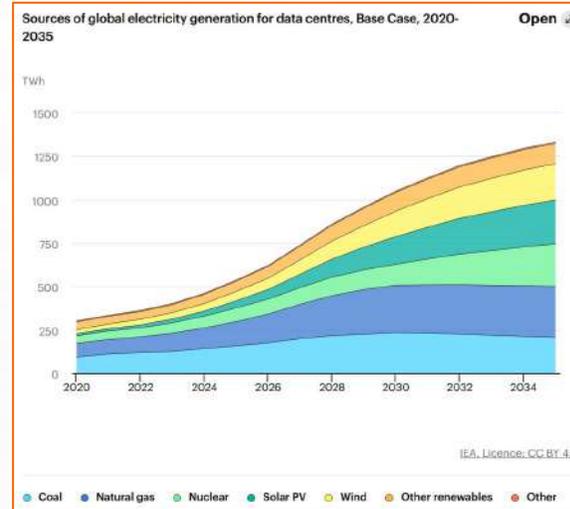
As overall energy supply increases, the grid to manage the distribution becomes the bottleneck:

- World supply of Oil and LNG is still projected to exceed the demand until 2030. The USA is still largest producer (~10%) but expected to peak in 2026 (EIA reports). Coal is set to decrease in the West, potentially increase in Asia (China consumes 60% of world coal). Demand is driven mostly by data centers/AI, manufacturing.

- Still long term 40-60% energy supply comes from fossil (by 2050) rest renewable/low emissions (EIA). Data centers will absorb ~10% of total energy by 2030 (various sources). ~40%-70% will be fossil fuels (coal, gas). By 2035 additional demand will be met by renewables (EIA).

- EU in 2025 for the first time Solar+Wind (30%) produced more energy than fossil (29%).However, the demand is flat as the growth in data centers is offset by deindustrialization (cars, metals factories etc.) (Ember, Eurelectric). UK supply is 49% renewable but demand is flat.

- Worldwide the main problem is the grid (PMJ, Eurelectric, Ofgem, GridStrategies, US Energy, WEF, AECG): especially long range high voltage lines are underinvested and inadequate as solar/wind supply increases. Main issues are permission and legal issues (ex. Not In My Backyard movements,NIMBY, etc.) , scarcity of LPT (Large Power Transformer), lack of funds and incentives.

Crypto Ecosystem Development (and AI)

Top 10 Countries by Crypto Ownership (GWI Data 2026)
Source: GWI (via InvestingInTheWeb) | Leading adoption across diverse economies

Top 10 Countries by Cryptocurrency Users (2025)
Source: Statista | India leads with over 100m users globally
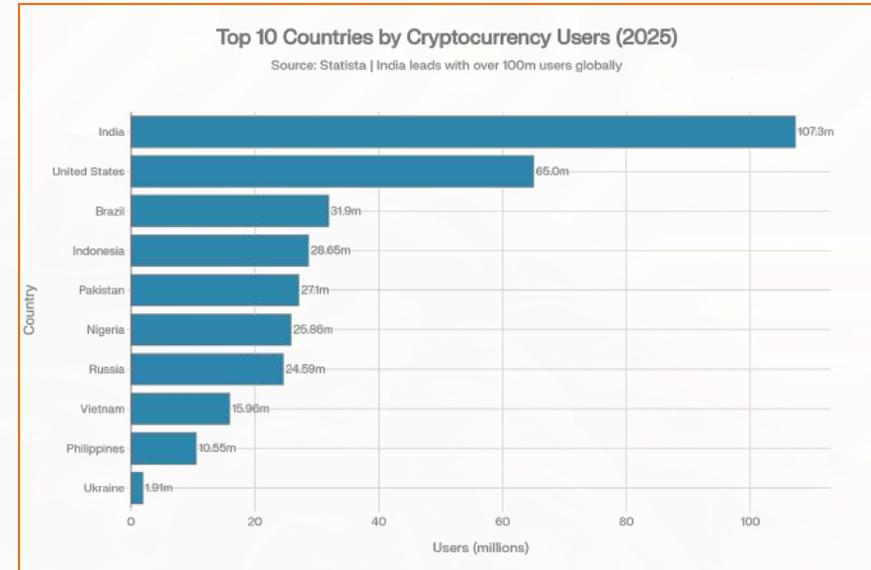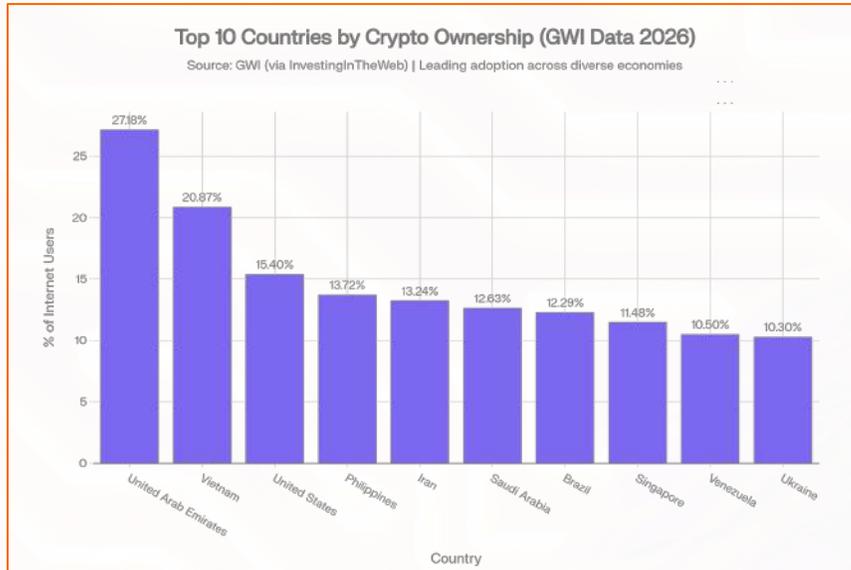
As AI, crypto and its applications are increasing:

- Bitcoin 'stabilized' below $100K and keeps bringing new media level of attention to the whole crypto ecosystem. Adoption PPP and GDP adjusted is higher in developing countries (India, Nigeria, Indonesia etc.). Stablecoins processed $9T in 2025 , 87% YoY, 5x Paypal.

- As of 2024, global cryptocurrency ownership rates are an average of 6.8%, with over 560M cryptocurrency users worldwide. 61% are male and 34% aged between 25-34 (triple-a.io). Crypto is now "property" in the UK (Property Act 2025) and "regulated innovation" in the US (GENIUS Act 2025).

- In 2024 ETFs on Ethereum (July) and Bitcoin (January). New ETF approvals in 2025: Litecoin, XRP, Solana etc. pushing crypto demand even higher.

- New cryptos are emerging in the AI space to solve various 'decentralized' issues, share of computing power (ex. Render), NFTs to authenticate data, train AI models (ex. Bittensor, ASI), build and share 'agents', share LLM inference (petals, gentorrent) revenue from its usage (ex. Virtuals, Griffain).

TECHNOLOGIES
Bring the future to humanity

Supply chain localization, sovereign cloud and workforce

Countries trying to be less vulnerable are localizing supply chains and implemented 'industrial policies'.
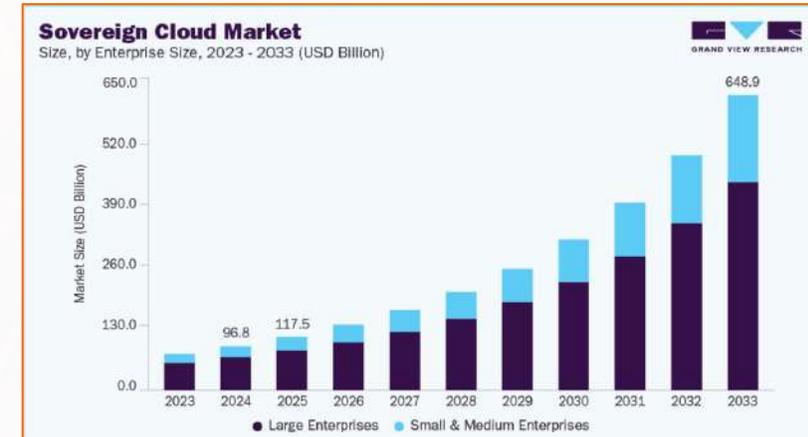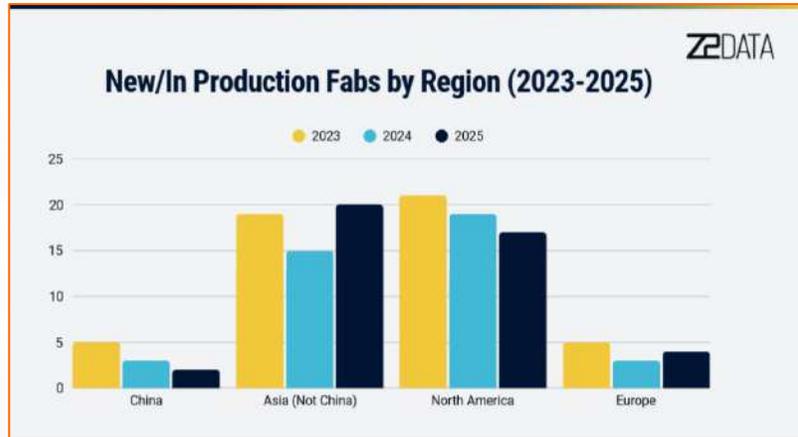
Given the political worldwide instability, supply chains are becoming localized or shifting to 'nearshoring' countries:

- US and other countries' bans on export of chips and related technology: US, Japan, KSA, UAE, EU and others have plans to expand local production capabilities (fabs). Inshore fabs increase resilience and local employment but raise costs to end customers (mckinsey).

- Re/Nearshoring:82% of companies reporting improved resilience, 77% reported cost reduction (pwc). Supply chain is considered a top priority for 68% of companies and nearshoring is a top 3 strategy for tariff mitigation (reuters).

- Sovereign cloud: AI/data regulation and geopolitics favors data processing within a country. The sovereign cloud market is growing at ~23% CAGR (grandviewresearch). Due to costs and lack of expertise, 'hybrid sovereign' systems (US tech with local company) is a solution (bcg, convotis)

- According to bls.gov, 2025 US workforce is employed 14% in professional services, 14.8% healthcare and social, 14.8% public sector, 8% manufacturing. Employment growth is expected in healthcare, possibly flat in other sectors.

VC investments in tech

Again, tech investments have remained concentrated in mega rounds and in USA. We are starting to see an increase of early stage deals and the push on robotics.

Global Venture Dollar Volume Through Q4 2025

Figure 1 Global venture capital deal count, Q1 2020 - Q3 2025
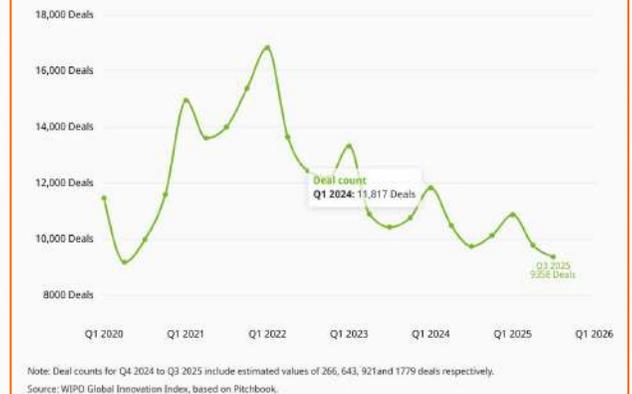
Venture capital investments in tech startups grows:

- Global VC investments in 2025, depending on sources, increased by 30-47% compared to 2024 in terms of total deals, valued at ~$425-469B (cbinsights, crunchbase). Vast majority of global VC money (~40-80%) is still spent in the US deals (different stages, sectors).

- Per Crunchbase, AI funding was ~50% ($202B) of the total funds. $118B (58% of $202B) went to mega rounds ($500M+) similar to 2024. $80B went to foundation labs. As in 2024, it confirms 'concentration' of rounds into few shortlisted winners (in $B): OpenAI (40), Scale AI (14.3), Anthropic (13), Project Prometheus (6.2), xAI (5.3), Databricks (4), Anduril (2.5), Anysphere (2.3) then Polymarket, ReflectionAI, SSI, Thinking Machines, Binance and Mistral (2 each). Unsurprisingly, # of deals is decreasing (~24K deals compared to 30-35K in 2024).

- Robotics reached $40B in 2025, a 74% YoY, the sector with the largest increase (cbinsights).

- Seed stage deals have increased 36% compared to 2024 (Crunchbase).

TECHNOLOGIES
Bring the future to humanity
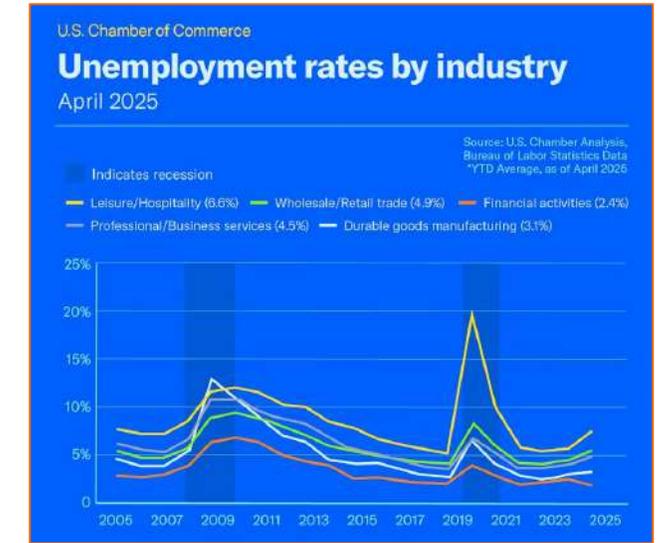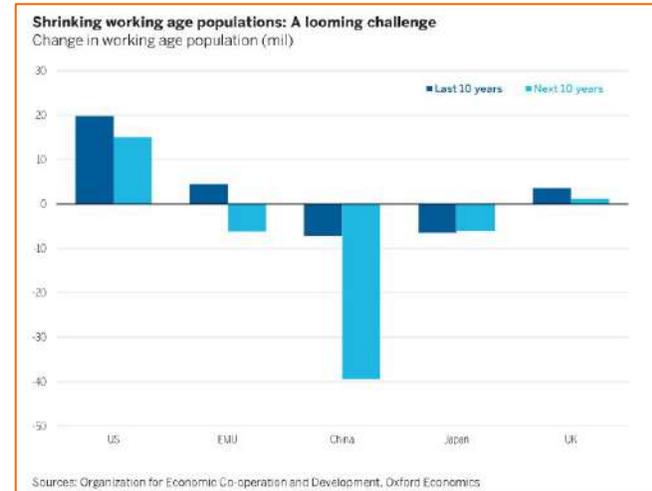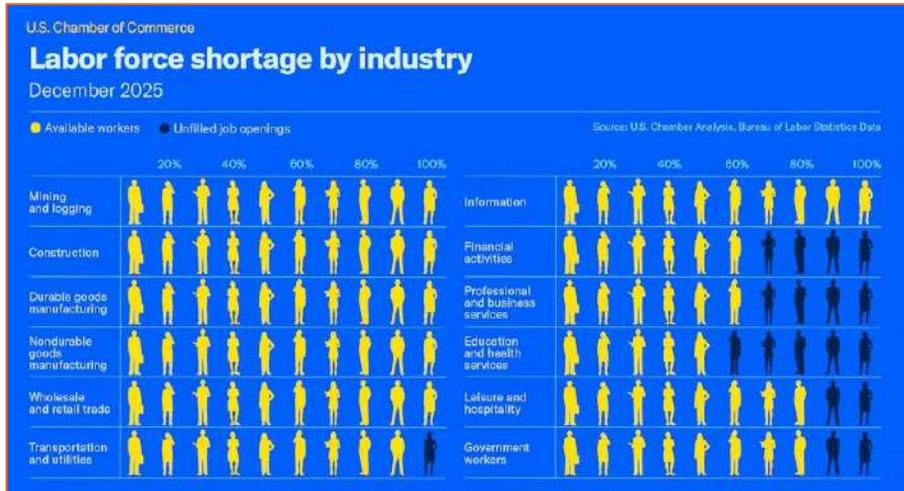
Labor shortages

Labor shortages are expected especially in certain sectors (healthcare, manufacturing, hospitality etc.)

- Large baby boomer generation now retiring and being followed by the smaller Gen X and the elongated profile of Gen Y and Gen Z (i.e., Gen Y and Gen Z are large but births were spread out over a long period), it may induce relative labor scarcity to be one of the enduring themes of the next decade. (Wellington Management).

- First time in years, as of December 2025, there were approximately 0.9 to 0.95 job openings for every unemployed person (bls).

- In the USA, the healthcare sector is expected to face acute worker shortages, particularly in nursing and specialized medical roles. Mercer projects a deficit of over 100,000 healthcare workers in the U.S. by 2028 (https://www.shrm.org).

- In 2025 ~4.2% roles per US manufacturer are unfilled (amtec). By 2030, the US manufacturing industry could have 2.1M unfilled jobs, potentially costing the economy $1T (https://www.shrm.org). 79% of manufacturing execs say labor shortage is greatest challenge (caddi research).

- By 2030, more than 85 million jobs could go unfilled globally due to skills shortages and aging population (kornferry.com).

Public spending & GDP/debt ratios growth

Defense budget is increasing while GDP/debt ratio is stable, marginally improving:

- Global military expenditure reached an all-time high of $2.74T in 2024, marking a 7.4% increase from 2023 (sipri.org).

- The U.S. remains the world's largest defense spender, with a budget of $997B in 2024, accounting for over 37% of global military spending. (sipri.org) NATO allies committed to a new target of investing 3.5% to 5% of GDP on defense by 2035.

- China's defense spending has increased by over 50% in the last decade, from $132 billion in 2014 to $234 billion in 2024 (aei.org). While China officially reports ~$247B, the real figure is $314-471B when accounting for purchasing power parity (PPP) and off-budget R&D (sipri.org, csis.org).

- Public debt keeps increasing as % of GDP in China and US, stable in the EU. (IMF, see image above). Inflation is expected 2-3% in Western countries (various sources). GDP is expected to grow in Asia, China marginally in EU and Japan (IMF).

TECHNOLOGIES
Bring the future to humanity

Virtualization and automation of living

Working from home

Working from home in the US has stabilized at about 25 percent of days.
(US full days worked from home, percent)

SOURCE: Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021. "Why Working from Home Will Stick." National Bureau of Economic Research Working Paper 28731, Cambridge, MA

Gaming Statistics By the Numbers

77% of Millennials — Play video games — 81% of Gen Z

Average Hours Playing Video Games per Week
Gen Z gamers — 7 hours and 20 minutes
Millennials — 6 hours 50 minutes

Gaming brings 90% of U.S. gamers joy

55% of gamers identify as male
45% identify as Female

57% of U.S. gamers play on smartphones

Cloudwards

THE PUBLIC'S USE OF ONLINE DATING SITES OR APPS

SSRS Opinion Panel Omnibus Poll. "The Public and Online Dating," January 5 – 7 and 19 – 21, 2024. Base: Nationally representative sample of 2,011 adults aged 18 and older.
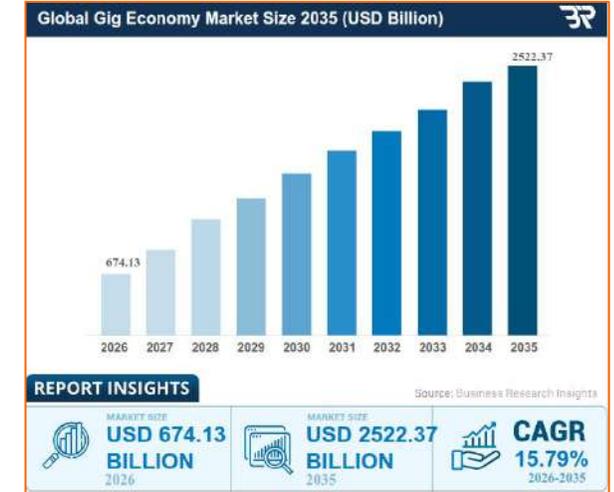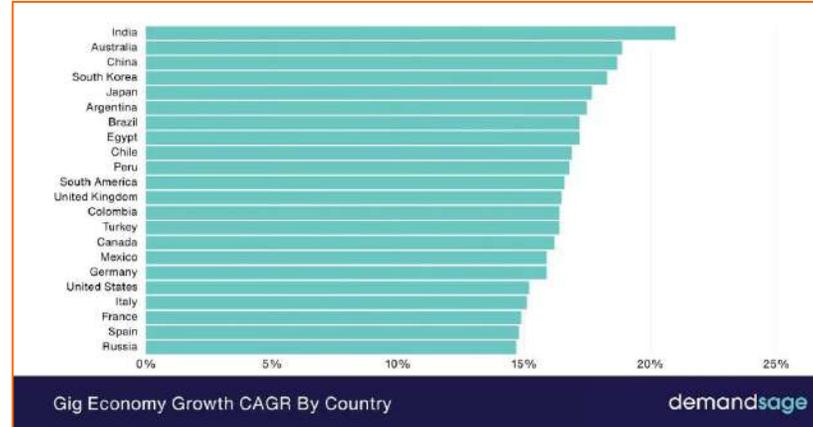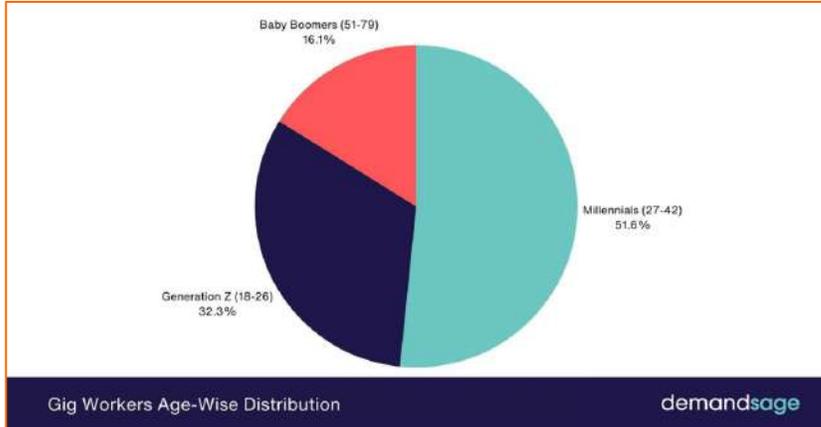
New generations increasingly use virtual environments compared to physically going out and automate tasks:

- Millennials (Born 1981-1996) and Generation Z (Born 1997-2012), ~50% view online experiences as meaningful substitutes for in-person interactions. 40% of Gen Z socializes more in video games than offline (zdnet).

- By 2025, the gaming industry is expected to generate ~$211B in revenue, with mobile gaming contributing $116B (EY). The esports audience reaching 640.8M in 2025 (fortunly). 77% of Millennials and 81% of Gen Z consumers play video games (cloudwards.net).

- Top-tier e-sports players (League of Legends, Dota 2) can earn up to $2 million/year including sponsorship and streaming and the average player makes $500,000/year. In comparison, a top football player (Premier League, Liga etc.) earns $2-4 million or more (tribuna.com).

- 37% of US adults used online dating at some point. A majority of adults (61%) believe relationships that begin on dating sites are as successful as those that begin in person. 37% of current users are aged 18-29, 38% are aged 30-49; 19% are aged 50-64 and only 6% are aged 65+ (ssrs).

- 22% of the US workforce will work remotely by 2025 and 14% of Americans work from home all of the time, according to Pew Research Center. 41% of employees work remotely on a hybrid basis (usatoday.com). US employees work remotely 25% of the time (IMF).

- Smartphone and internet overuse may decrease cognitive abilities (https://bit.ly/3PXY9ev) and similar for AI usage (https://bit.ly/3CFyi7Z).

TECHNOLOGIES
Bring the future to humanity

Freelancer (gig) Economy

Gig Workers Age-Wise Distribution — demandsage

Gig Economy Growth CAGR By Country — demandsage

Global Gig Economy Market Size 2035 (USD Billion)

An increasing portion of the workforce has flexible or project based employment:

- The gig economy has evolved from a niche sector into a massive economic engine, growing approximately 3x faster than the traditional workforce(fortune).

- The global gig economy is valued between $450-674B in 2026 with a CAGR of 16-17%. Labor sourced strictly through digital platforms (like Upwork, Uber, DoorDash) alone generated over $556B in 2024.

- By 2025, the total global gig workforce reached ~400-500M people. The online gig workers are ~12% of the total online workforce (worldbank).

- USA is the largest market with ~76M freelancers, ~36% of the national workforce (carry). Surprisingly, the largest employer of US gig workers is actually the government/public sector, 14% (statista).

- While the total market value is growing, the % of population in the gig workforce is stable since 2014: 48.92% vs 47.97% in 2023 (worldbank). India is the fastest-growing gig economy with a 21% CAGR, Argentina with 17.5% and Brazil with 14.9% are growing at a rate higher than the average.

- A permanent global inflation, economic uncertainty, affordable access to (AI) tools to provide services, 'knowledge' needs (Knowledge Process Outsourcing, KPO: gigs in healthcare, education, legal consulting, and AI training etc.) are the main reasons of market growth.

AI capital, energy and economy (Events)

Trend Prediction 1
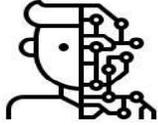
AI investments keep growing and hardware/software improvements
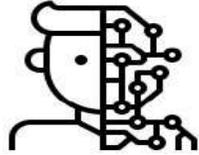should make energy demand under control.

- New Nvidia Rubin GPU requires +2x power (up to 2.3KW) but it can generate 10x more AI output (tokens) for the same cost and crucially **does not require (energy-intensive) chillers cooling systems to run** (ambient to cool the water to 45°C, even in hot climates).
- Training models (<150B parameters) is affordable: LLama3 70B ~$1-2M, 7B $50-500 (localaimaster). Fine tuning starts from $500.
- Large and reasoning models are more energy draining: assuming an essay of 1K words, a reasoning model will consume (Wh) up 10-30X a non reasoning model ([University of Rhode Island's AI lab](#)). This is consistent with current service costs: non reasoning price ($/1M tokens) is significantly less than reasoning (Ex: as of Jan 2026, GPT4.1 non reasoning $12 /M tokens vs GPT5.2 $14/M tokens).
- **Performance of small-medium AI models (<150B parameters) are now '[sufficient enough](#)'** for many tasks and keep improving. Also ['local' (e.g. small) AI models are now 5.3x more efficient than cloud based for most tasks](#).
- Beyond China and USA, countries are planning ambitious data centers (ex: KSA 480MW, UAE 200 MW, Japan 3.1GW etc). Due to the grid energy issues, most new data centers follow the Bring Your Own Energy (BYOE) strategy (build their own power plant).
- Accounting **GPU depreciation is 5-6 years while operative lifespan is 2-3 years** (various).
- As already mentioned in [2025 report](#), space data centers are the new frontier for energy efficiency: Lumen Orbit, Loft Orbital, Axiom have all shown it is possible and large players (SpaceX, Google etc.) are coming.
- New 'types' of AI labs raised significant funds ($1B+): SSI (aiming superintelligence beyond current AI methods), Thinking Machines (Custom enterprise infrastructure).

➜Triggered from Catalysts: 1,3,4,6,7,9

Technologies

AI capital, energy and economy (Predictions)

Trend Prediction 1

AI investments keep growing and hardware/software improvements should make energy demand under control.

By 2027

- Due to the lowering barrier to training small-medium models, early stage VC funds will be about 30% and focusing on AI applications (no foundation, infrastructure etc.).
- AI fund concentration in mega rounds will continue but it will slow down or flat compared to 2025.
- Robotics will reach $50B+ in funding.
- Due to the improvement in cooling, NIMBY issues and demand pressure, retrofitting dated data centers will be as common as new built data centers.
- Data centers' energy demand will be under control due to the BYOE and efficiency improvements. However the grid will still be under pressure.
- As an orbital object is under the law of the country where the owner is incorporated, a country will propose a law designed for 'space data centers' to incentivize and regulate the matter.
- There will be at least one neocloud company or GPU data centers with impairment issues due to GPU depreciation and chip manufacturer will come to rescue ('circular' deal).

➡️**Triggered from Catalysts:**   1,3,4,6,7,9

Technologies

AI Hardware 1 (Events)

Trend Prediction 2
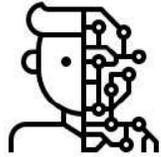
Hardware performance are still dominated by Nvidia, competitors are coming and a robotics market is emerging with Nvidia as a top player.

Overview

- A striking contradiction: **GPU market is valued at $136B and the whole semiconductor industry $627B ([precedenceresearch](), [precedenceresearch]()) while Nvidia and AMD have 2025 revenues of $130B and $34B and market caps ~$4T and ~$420B**. Some analysts may think companies are overvalued (various) others may think the whole industry instead is undervalued (various).
- Nvidia chips were overbooked (including 'circular' deals with Microsoft, AWS, Oracle etc. and [smuggled chips into China]()). AMD got significant orders (Oracle 50K GPUs, OpenAI deal, KSA $1B deal etc.) and Google started selling its TPUs (Anthropic, Meta etc.). As per 2025 predictions review (pg 8), AMD has about ~8% market share on discrete GPUs ([carboncredits]()) and 10% on data centers ([mordor]()).
- **Small Models (fit in laptop hardware) are becoming 'sufficiently powerful for most tasks'** (various, [including Nvidia and Georgia Tech]()).
- Neoclouds vs Local: constant improvement of GPUs and infrastructure is pushing companies to hire instead of buying (Coreweave, Nebius etc.) or security and the 'small AI performance' push companies to 'local' AI training/inference hardware (HP, Asus, Apple etc.)

Nextgen GPUs

- Nvidia: Rubin R200 is expected for mid 2026, should be up to 5x faster in inference 3.5x faster in training, compared to current Blackwell B200-300 (configuration of 72 GPUs, Vera Rubin NVL72). The energy efficiency should be 10x for inference and 4x in training, especially in 'sparse AI models' (MoE or similar).
- **Nvidia vs AMD and Intel hardware: both products (Rubin vs MI450) in mid 2026 with similar performances** (slightly better for Nvidia): memory bandwidth (~20TB/s), TGP (total graphic power) 2.3-2.5KW, compute ~40-50 PFLOPs. AMD beats Nvidia on memory (~432 vs ~250-380GB). Intel is seriously lagging: best chip Gaudi 3, performances on par with Hopper H100 (2 generations ago).

➡️**Triggered from Catalysts:**   3,4,5,6,7

Technologies

AI Hardware 2 (Events)

Trend Prediction 2
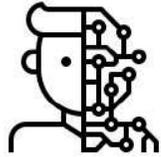
Hardware performance are still dominated by Nvidia, competitors are coming and a robotics market is emerging with Nvidia as a top player.

Nextgen GPUs (continued)
- Nvidia software vs others: the **usability of Nvidia CUDA is ahead of AMD ROCm and others**. AMD is fast improving and there are CUDA 'open source' (MUSA or [ZLUDA](#)), to work on other chips. Also NVIDIA announced NVQLink to combine GPUs and quantum processors.
- Others vs Nvidia: xAI and Meta in development, AWS Tranium v3-4 are cost effective especially in inference and compatible with Nvidia racks (NVLink), Google TPU (v7) on par with Blackwell and more efficient possibly better in 'specific architectures' (e.g. Transformer).
- Chinese: top chips (e.g. Huawei) are on par with Hopper (2 generations away), other cheap options (Moore Threads , Biren and Cambricon) all raising revenue due to Nvidia absence. The strategy now is to invest in software platform and optical/photonics connection.
- A startup, Ricursive raised $35M to reduce the chip design cycle from 2-3 years to few weeks (wsj).

Nextgen Inference only
- Nvidia CPX: as AMD is catching up, Nvidia announced by the end of 2026 a chip with disaggregated phases of inference (prefill, decoding). With the Groq acquisition, **Nvidia will be ahead at least in long context processing**. Others (Microsoft Maia 200 etc.) are in progress.
Other approaches raised >$100M: model specialist (Etched only Transformer models), cost-efficient designs (d-Matrix, Rebellions, PositronAI), racks architecture (Lightmatter), fine tuning/inference (Sambanova, Untheter, FuriosaAI), new chip design (Tenstorrent, Graphcore, Cerebras).
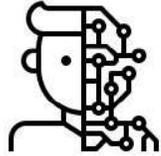
Nextgen robotics GPUs:
- Nvidia vs others: **Jetson platform is the industry standard in performance while other cheaper options exist** (Qualcomm RB5, AMD Kria/FPGA and Raspberry+Halio). Again Nvidia provides a software (Isaac) which simplifies training/usability compared to competitors.

➜**Triggered from Catalysts:**   3,4,5,6,7

AI Hardware (Predictions)

Trend Prediction 2

Hardware performance are still dominated by Nvidia, competitors are coming and a robotics market is emerging with Nvidia as a top player.

By 2027:

As the process started last year (see [2025 report](#)), roughly the same overview is valid on Nvidia and competitors. However:

- Nvidia will continue to dominate but the market share (data centers, clusters etc.) will be eroded below 85%.
- Nvidia will start a major lawsuit to limit CUDA portability into other chips.
- AMD will have 10%+ market on discrete/laptop GPUs.
- IPO of a new chip maker >$1B.
- First $1B order for non AMD/Nvidia chip player.
- As local models are now performant enough, discrete GPU market will grow more than 6-6.5% CAGR currently expected (marketresearchfuture, dataintelo).
- As the next AI chip market growth will be due to the expected new robotics, there will be at least one player that announces a platform competitive in performance with Nvidia Jetson.

➔ **Triggered from Catalysts:** 3,4,5,6,7

Technologies

AI adoption in governance, defense, cybersecurity 1 (Events)
Trend Prediction 3
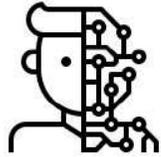
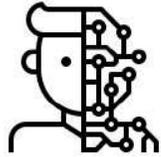**Beyond deepfakes: GenAI's first malware, defense adopt AI wingmen drones.**

- The **USA is the world's most internet dependent country as 1h outage will cost $300M**, then China $240M, UK $93M, Japan $78M, Spain $64M, Netherlands $45M, Germany $42M, India $42M, Switzerland $37M and South Korea $35M. Good news: numbers are down compared to 2024 (e.g. 1h in USA cost $450M) ([merchantmachine](#)).
- Attacks are increasing in 2025: Cloudflare records +30%DDoS attacks in H2 compared to the full year 2024 (also 1Bpps+ attacks are surging); Ransomware attacks +32% and manufacturing was hit the most +61% (socradar) but ~60% did not pay (compared 50% in 23); Phishing +31% in H1 25 compared to H1 24 (mail phishing via AI now counts 80% of the total) (sqmagazine).
- China and USA have the largest amounts of vulnerable ("open") devices (called amplifiers, which are the main source of DDoS weapons) ~1.5-1.4M, 3rd is Italy with ~800K (the highest concentration of weapons/person), 4th South Korea ~619K. (a10networks).
- Bots (good or malicious) are ~51% of web traffic (Imperva), ~20%+ of social media (various).
- CAVs (connected and autonomous vehicles) have increasingly [more vulnerabilities](#) and their supply chains discovered vulnerabilities in 2024, yet to materialize into incidents ([vicone](#)).
- Cybersecurity market growth is 9-14% CAGR compared to last year 7-11% (researchanmarket,fortunebusinessinsights, mordor).
- AI incidents (per AI Incident Database) are increasing: 276 in 2024 compared to 317+ in 2025 ([time](#)).
- New Attacks by GenAI: Chinese group via Claude Code performed espionage with 10-20% human supervision ([anthropic](#)). **Google reported 5 new AI malware**: FRUITSHELL, PROMPTFLUX PROMPTLOCK, PROMPTSTEAL and QUIETVAULT able to connect to server, execute malicious scripts, collect sensitive data and exfiltrate credentials while re-generate code/itself to avoid detection.

➜ **Triggered from Catalysts:** 1,2,6,7,9,10

Technologies

AI adoption in governance, defense, cybersecurity 2 (Events)
Trend Prediction 3

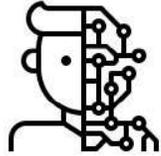Beyond deepfakes: GenAI's first malware, defense adopt AI wingmen drones.

- Deepfakes: Video accounts for 46%, image 32% and audio 22% (zerothreat). However, **vishing (voice impersonation) surged 449%** (sqmagazine). A survey reports 27% of people struggle most with identifying deep fake text/images compared to video (mimecast).
- **AI can hack blockchains**: tests showed that the latest LLMs (Claude 4.5, GPT 5 etc.) can find smart contract vulnerabilities.
- AI Safety scores of major players: Future of Life Institute assessed AI safety major players and the best is Anthropic with C+, OpenAI C, Google C-. F for Chinese players (Deepseek and Zhipu).
- Even with BigTech and labs attempting to stop malicious usage and improve safety, there is **a proliferation of malicious AI tools** (based on open source). Examples: WormGPT 4, FraudGPT, OnlyFake, Morris II (phishing, malware, forgery, identity fraud, exfiltrate data etc.).
- As we use more AI, AI itself is vulnerable: +540% in prompt injection incidents, fastest-growing attack especially via RAG or Third-Party Inputs like EchoLeak (sonnylabs, lasso). Also research found that as little as 250 documents in training can allow exfiltration of data.
- Safety products are evolving to tackle the threats and secure AI: as per last year's report Virtue AI, Protect AI, Lakera, HiddenLayer, PromptArmor etc. act as firewalls between user (or tool) , the model and the supply chain /pipeline related to the model.
- Due to AI new regulations (EU AI Act and others), the AI governance market is valued $308-620M in 2025 with a CAGR of 36-51%. Gartner predicts it will drive $1B in compliance by 2030 (Forrester $15B). (grandviewresearch, nextmsc). Most US AI players are abiding by the EU AI Act, Meta (Llama) refused, xAI partially. **Mistral and Black Forest are compliant, Chinese models mostly not**. EU Commission proposed to delay the EU AI Act August 26 deadline of 6-12 months (Digital Omnibus).
- AI in defense industry alone is valued at ~$11B with a CAGR 13% (grandviewresearch, gminsights).
- US Military embraces AI: autonomous/wingmen drones are in testing by Anduril, Saronic, General Atomic, AeroVironment etc.

➜ **Triggered from Catalysts:** 1,2,6,7,9,10

Technologies

AI adoption in governance, defense, cybersecurity (Predictions)
Trend Prediction 3

Beyond deepfakes: GenAI's first malware, defense adopt AI wingmen drones.

By 2027 :
- There will be a serious drone/autonomous vehicle attack using AI which will raise awareness on vulnerabilities. (similar from last year).
- AI regulations only related to drone/autonomous overall will increase in western countries and China. (similar from last year).
- A drone or robotics defense startup will be bought (or IPO) for at least $40B+.
- A small LLM model <5GB in RAM space or ~<4B parameters will be used in cyber attacks.
- A western country will put a ban on Chinese models usage at least in certain industries.
- A major cyber attack reported in the news will involve Italy or South Korea.

➜

Technologies

Industry AI adoption (Events)
Trend Prediction 4
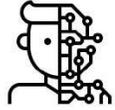
As hallucination and complexity are still practical hurdles, industry adoption is now progressing beyond pilots especially in 'narrow' scope cases.

- Organizations using GenAI in one or more functions rose again to 88% in 2025 compared to 78% in 2024. More companies are experimenting but no more than 10% are scaling AI solutions. Industries that deploy at scale the most are Tech, Media, Healthcare, Energy and Insurance. The most common usages are IT, knowledge management and marketing/sales or operational efficiency (mckinsey, cisco).
- Research in retail and recruiting show valid ROI for AI solutions in business. GenAI spent was $37B in 2025. **However, 27% of the AI commercial spending is from individual users (coding assistant etc.) even using their personal credit card** (menlo VC).
- Other studies, however, show AI pilots have a high failure rate (88-95%) with studies identifying the issue as 'GenAI lacks memory and adaptability.' However, they believe **AI can succeed in 'narrow scope' workflow** (document automation for contracts and forms, voice AI for call summarization and routing, code generation for repetitive engineering tasks) (idc, mit nanda).
- 60% of AI projects will be abandoned due to lack of "AI-ready" data by 2026 and 40% of agentic projects by 2027 due to complexity. Only 11% of AI projects reach production (Gartner). Significant ROI using agentic AI solutions requires significant business restructuring (as the agents touching different business functions at once) and it is benefitting ambitious/audacious companies (mckinsey, mit nanda).
- **Externally provided AI solutions have up to 2x chances to succeed compared to building internally**. (mit nanda, uipath,menlo VC). The AI solutions will not replace employees but rather outsourcing agencies, BPOs (and KPOs) (mit nanda).
- 20-30% of companies in ETR survey declared they will limit hiring due to leveraging AI.
- France announced discontinuing contracts with US video calls providers (Zoom, Microsoft) in favor of local company Visio.
- AI will displace 92M jobs but create 170M by 2030. There will be a massive skill mismatch (wef).
- Data labeling market is growing at 29% CAGR (coherent insights) and even Uber is converting Indian drivers into data labelers.
- The American Arbitration Association launched an AI powered arbitrator for construction disputes.
- Hallucination is still a 'structural' issue plaguing GenAI solutions (various research).

➜**Triggered from Catalysts:**   1,2,3,6,8,9,10,11

Technologies

Industry AI adoption (Predictions)

Trend Prediction 4
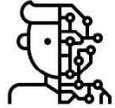
As hallucination and complexity are still practical hurdles, industry adoption is now progressing beyond pilots especially in 'narrow' scope cases.

By 2027:
- GenAI spent will reach $50B+.
- As data labeling market grows this will offset potential reduction in Knowledge Process Outsourcing (KPO) and we will not see a reduction in CAGR growth (Precedence Research set 13.7%).
- Success rate of AI pilots in production (scaling) will be <15% for 'complex/multi agents' but above 30% for 'narrow scope' projects.
- Most of AI commercial adoption and spend will still be in AI tools and small-medium ('specific') models.
- Due to AI adoption and efficiencies (i.e. no recession driven), hiring will be reduced or frozen but the admin workforce reduction will be no more than 3% as counted by the [Federal Reserve Bank](#) compared to January 2026.
- EU politics will incentivize the usage of 'local AI champions' over the rest i.e. Mistral , Black Forest etc.
- New AI powered arbitrators will be launched beyond construction.
- Since significant gains of AI/agentic solutions require restructuring, words like 'restructuring workforce, processes' will become common on public companies' earnings calls.

- Additional (as it is not measurable): There will be layoffs and some labeled 'due to AI' will likely be AI redundancy washing as the companies doing them may be in financial distress.

➔Triggered from Catalysts: 1,2,3,6,8,9,10,11

Technologies

Consumer AI adoption of LLMs/GenAI tools (Events)
Trend Prediction 5

As the market is shared by BigTech, AI 'local' products with are emerging beyond large vendors and platforms.

- In 2025, 26.3% of US adults used AI solutions at least monthly. UAE has the highest number 59% ([microsoft](#)).
- December 2025 monthly visits (similarweb):
  - OpenAI (5.5B): USA 16.6%, India 9.1%, Japan 4.7%, Brazil 4.65% and UK 3.7%;
  - Gemini (1.7B): USA 12.6%, India 7.4%, Japan 7.4%, Brazil 5.3% and Vietnam 4.8%.
  - Grok (271M): USA: 22.5%, India 9%, Brazil 4%, Korea 3.5%, Vietnam 3.5%;
  - Claude (172M): USA 30%, India 6.3%, UK 4.3%, Indonesia 3.9%, Japan 3.6%;
  - Deepseek (328M) China 38%, Russia 11%, USA 5.5%, HK 3.5% Indonesia 3.2%,
  - Qwen (32M): Russia 29%, China 17%, USA 4.6%, India 3.8%, Indonesia 3.6%.
- Top 5 open models in terms of tokens processed: Deepseek 42%, Qwen 16%, Llama 11.7%, Mistral 8.6% and OpenAI 4.9% (openrouter).
- The smarter you are, [the smarter you still are when using AI](#) (compared to less smart employees).
- Referrals from ChatGPT (1.1B visit/month) convert better than any other marketing channel: direct, paid search etc (vertu, similarweb).
- Claude and Grok Code are consistently the most used tools for programming (openrouter).
- **Roleplay (if you were a doctor, a teacher etc.) and programming are the most common use cases for using LLMs** (openrouter).
- AI revenue is $37B in 2025, 3x YoY with $19B in products and $18B in AI infrastructure. Also BigTech now counts as 50% of total VC spending in AI (Menlo VC, dealroom). **Chinese models have ~30%+ market share, more in low income countries** (openrouter, microsoft).
- Individuals or labs are experimenting with 'agentic' 'local' solutions, supervising their own LLM(s): crewAI, mem0, relay.app.
- AI Companion market: already at $32B market growing at 30% CAGR. **Users engage with AI companions for an average of 1.5 to 2.7 hours daily, often exceeding time spent on TikTok or Instagram**. Audience is young: 60%+ under 30 and retention rate is 48% higher than traditional social media apps (nikolaroza, mktclarity). Notable startups in number of users: [Character.ai](#) (20M+), Replika (30M+), [Candy.ai](#) (1M+), [Chai.ai](#) (10M+) and Pengu (15M+).

➜ **Triggered from Catalysts:** 2,3,6,8,10

Consumer AI adoption of LLMs/GenAI tools (Predictions)
Trend Prediction 5
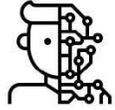
As the market is shared by BigTech, AI 'local' products with are emerging beyond large vendors and platforms.

By 2027:
- Chinese models will continue to dominate market share in low and middle income countries.
- Outsourcing countries (Brazil, Vietnam, Indonesia, India etc.) will have at least 15% of traffic of GenAI services.
- A major ecommerce platform (Shopify,Ebay, Amazon etc.) will announce 1%+ of all purchases will come from GenAI referrals.
- An AI companion startup in NSFW category will exceed 50M users.
- There will be an agentic platform low/no code (or framework) that gets traction on at least one industry/niche (**not** coding). It will not be built by a major player (e.g. OpenAI etc.), get media attention and installable locally, not cloud based. At least 1M users (installation).

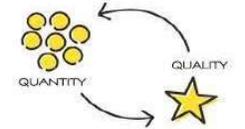➡ **Triggered from Catalysts:**   2,3,6,8,10

Technologies

Models' Advancements: Text/Multimodal (Events)
Trend Prediction 6
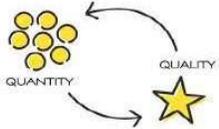
Advancement in AI models is rapid and beyond current Transformer architecture and post training RL. Models are now performing even at small-medium sizes.

- Models are released as a 'family': within a family, most Western models are proprietary, most Chinese models are open. Examples: Qwen 3 family, all open but Qwen 3 Max; Kimi 2 family all open; Gemini 3 all proprietary.
- **Top models still proprietary and (mostly) USA made, top open models are mostly China made.**
- China grows the % of AI publications and citations, USA is flat, India is growing (including github AI projects) ([HAI](#), [study1](#))
- Korea for the first time entered the top 10 rankings for the open weights models with K-Exaone in January 2026 .
- **In terms of pricing, best open models are comparable to similar (in performance) proprietary models but they are slower.** Kimi K2 Thinking cost per 1M tokens is the same as Gemini 3 Flash ($1.1) and their overall Intelligence scores are 40 and 41 respectively. However Kimi is significantly slower (Output Tokens/s): 83 compared to 223 of Gemini 3 Flash. Deepseek 3.2 V2 is even cheaper ($0.3 per 1M tokens) scores slightly better in intelligence (41) but is very slow (29 tokens/s). If open weights are necessary GLM 4.7 is the best compromise: 42 intelligence, $0.9 per 1M tokens and 137 tokens/s speed. (All data from [artificialanalysis.ai](#), January 2026). ➜
- Beyond Transformers: Significant research in diffusion models and hybrid architectures (e.g. Nvidia Nemotron 3 etc. See 'Best AI advancement papers' pg12). Research on 'continual learning' (i.e. continuous update of the weights) has attention in 2025 but no definitive breakthrough yet. From Deepseek, Reinforcement Learning with Verifiable Rewards (RLVR) becomes the standard in post training. Inference time scaling still behind top models (i.e. the model elaborates draft answers then assesses and selects one).
- Top performance models now have a 1M+ tokens context window but transformer still suffers from 'attention collapse' (forgets previous information). **Also transformer based [LLMs show 'structural' hallucination](#) in tasks beyond a certain complexity.**
- Episodes of benchmaxxing: top models have trained directly on testing data artificially boosting their scores (Llama 4 etc.).
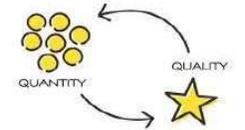
➜ **Triggered from Catalysts:** 2,3,4,5,6,7,10

Models' Advancements: Text/Multimodal (Predictions)

Trend Prediction 6
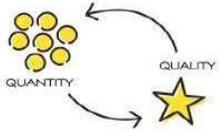
Advancement in AI models is rapid and beyond current Transformer architecture and post training RL. Models are now performing even at small-medium sizes.

By 2027 :

- Best overall (i.e. not only open weights) 'family of models' will still be from a BigTech or Lab (>$1B) with
    - hybrid architecture (i.e. not only transformer based) for reasoning/thinking performance
    - diffusion based for 'fast/small' model size
    - proprietary at least at thinking level (best of the family)
    - no more than 10% better than the closest open weights model
    - at least 20 points better than the best 2025 model in one LLM overall benchmark leaderboard released in 2025 (GDPval, Humanity's Last Exam etc.)
    - a new algorithm/approach to overcome 'attention collapse' (i.e. the model forgets after a period of time previous info).
- A Chinese lab or BigTech will have the best models <150B parameters in overall benchmarks.
- We will have a model of around 8B parameters or less with scores 20%+ on Humanity's Last Exam Benchmark or similar (i.e. more than double current best intelligence performance of a 8B model).
- There will be a benchmark for reasoning on noisy/messy data/context.
- At least one lab not from China, USA will hit the top 10 in one LLM benchmark leaderboard for the first time.
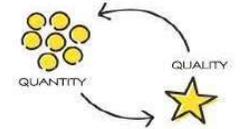
➜ **Triggered from Catalysts:** 2,3,4,5,6,7,10

Technologies

Models' Advancements: Agentic, Coding, platforms (Events)
Trend Prediction 7
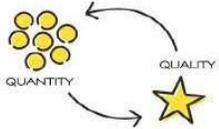
AI and applications market is maturing. Memory and agents integration are emerging as key elements to reach mass adoption.

- The length of tasks (measured by how long they take human professionals) that generalist frontier models can complete with 50% reliability is doubling every ~ 7 months (metr).
- Many coding LLMs are open, the best are still proprietary and from USA (leaderboards). Coding startups are coming (Cursor Composer 1 etc.), stackoverflow traffic is falling.
- The best small model that scores better than 60% (MMLU-Pro) is shrinking: now Qwen3 VL 4B (~ 8GB in memory).
- **Many best open models do not perform well with tools (like search, database query, compiler etc.).**
- **Models are becoming 'brands' as each is 'different' : Kimi K2 is renown for creative writing, Claude Code for coding, Nova for customer service and instructions.**
- There is an increased interest in research (Magma, MemRec, CompassMem etc.) and startups (Zep, Letta, Lettria Mem0 etc.) how agents should share memory and organize data in nodes (context graph) both for learning and performing long tasks. At the same time a user should be able to delete from memory via a biologically inspired 'technology of forgetting'.
- Protocols to connect models, agents to systems have been proposed: MCP(tools), A2A(cross platforms), ACP (local network).
- Performing models now fit in a laptop or less: **Nvidia and Georgia Tech found small models (<8B parameters) more suitable for ('narrow') agentic applications.**
- There is a 'convergence' of traditional RPA with Agents (Ex. agents can automatically fix broken RPA scripts): while agents may suffer from hallucination and safety issues, **there is a trend of 'governance as code' to limit unwanted agent actions**. Triggering agents by simple approved RPA bots may solve issues and be more efficient (forrester, octoparse, gleecus).
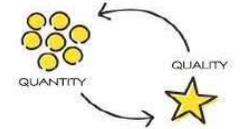
➜ **Triggered from Catalysts:**   1,2,3,4,5,6,7,10

Technologies

Models' Advancements: Agentic, Coding, platforms (Predictions)

Trend Prediction 7
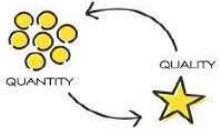
AI and applications market is maturing. Memory and agents integration are emerging as key elements to reach mass adoption.

By 2027 :

- Anthropic Claude Code will still be the best coding LLM on major benchmarks and with more traction among coders.
- A multi-agent system will complete tasks for longer than 10h (80% success rate).
- While different agent protocols will be used, MCP will be the most used and adopted.
- A memory startup founded no later than 2022 will reach $150M+ valuation.
- Given the need of regulations embedded in agentic (governance as a code) and the evolution of RPA into agentic, UiPath will at least be +30% YoY compared to January 2026.

➜

➜ **Triggered from Catalysts:** 1,2,3,4,5,6,7,10
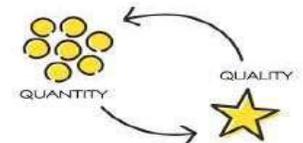
Technologies

Models' Advancements: Audio/Images/Video (Events)
Trend Prediction 8
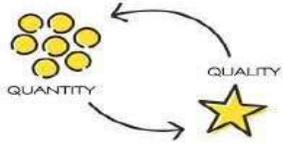
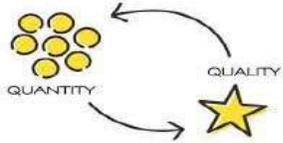AI models improved but none is a dominant player. A new image to 3D sector is emerging.

- **Many labs from around the world are joining the race to develop top performing models**. Beyond China and USA, notable companies in 2025 were Black Forest (Germany), SberAI (Russia), Bria and Lightricks (Israel), Moonvalley (Canada), Leonardo (Australia), Animon (Japan), Recraft (UK).
- Image generation: quality overall is similar across the top products (Nano Banana, GPT Image, Flux, Seedream, Qwen etc.). What makes the difference is the UX, features of each product (slide creation, text insert, editing, time, pricing etc.). No clear lead.
- Video generation (text2vid and img2vid): among the top products (Kling, Veo, Sora, Runway etc.) all have strengths and weaknesses from consistency, physical realism, length of generation etc. No clear lead.
- Audio/music generation: this is the only sector with little interest by big players (both USA and China). Notable best models are from Suno, Producer.ai (both USA), Elevenlabs (UK) Maps AI (China). Note: text2speech and speech2text are dominated by the usual BigTech/labs (both Chinese and American).
- 2D to 3D generations: those who read this report, may remember we started to talk about 3D generation from images in 2023 (State of AI 2023 page 3 Neural Radiance Fields NeRF). Few products currently exist (tripo, alpha 3D) and getting better. The research is improving significantly (see pages 15-16, AI best advancements).
- Coca Cola used GenAI for its commercials 2024 and 2025 (team of 5 AI specialists and creatives, a mixture of the products: KlingAI, Veo 3, Luma AI, and Sora, etc.). Mixed reviews. **AI 'slop/brainrot' takes Youtube**: low quality AI video generated accounts for 20-30% of Youtube shorts (kapwing). Spain has the most (20M) AI slop channel subscribers, followed by Korea and USA.

➔ Triggered from Catalysts: 2,3,4,6,7,10,11

Technologies

Models' Advancements: Audio/Images/Video (Predictions)
Trend Prediction 8

AI models improved but none is a dominant player. A new image to 3D sector is emerging.
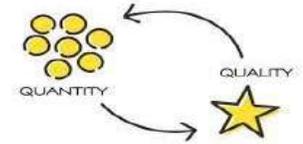
By 2027 :
- On the top 10 models' leaderboards for video or images there will be companies from at least 7 different countries during the year.
- A new leaderboard (and relative benchmarks) for 2D to 3D models will emerge.

→ **Triggered from Catalysts:**   2,3,4,6,7,10,11

AI adoption in science, robotics (Events)
Trend Prediction 9

AI applications and models in science, robotics and related are exploding. New breakthroughs and adoption are expected.

- Games: different approaches Runway Gen-4.5, Deepmind Dreamer 4 (world model), Google Genie 3, Nvidia Cosmos/NitroGen,can play games without human input and without large GPU clusters. Vision-Language-Action (VLA) models are improving fast (various).
- Robotics AI is maturing: spinning out from gaming (a mix of RL, imitation learning, world models) we have foundation pre-trained models: Nvidia GROOT 1.6, ByteDance's GR-3 and robot-controlling models Gemini Robotics 1.5, Molmo-Act 'scalable' to different robots (hardware): Unitree, Boston Dynamics, UBTech etc. However, **robot tasks are limited to 3-5 steps, no complex chains (various)**.
- Beyond factories: The market for robotics is expected to reach $200B+, Amazon deployed 1M+ robots and you can already see robots serving food in restaurants. 70% of the market is still on industrial robots (China, Japan and Korea the largest markets. India is the fastest growing 7% YoY). Consumer robots are now in pre-order (ex: x1).
- Self Driving and Drones: American Waymo, Tesla, Zoox expand but Chinese competitors are growing faster. Drones delivery market is expected to grow at 35% CAGR both USA and China have adoption growth. Drone startups are USA dominated followed by India, while China has corporates (JD, Meituan etc.). **World drone manufacturing is China dominated (various sources, 70-80%).**
- Surge of science models: beyond US/China known labs, **various countries are developing 'vertical' (sector e.g. agriculture, chemistry etc.) foundation models**. Examples: India with healthcare, material discovery, agriculture etc.; China model Sinong for agriculture.
- Top models keep making breakthroughs: GPT-5.2 solved Erdos Math problems; OpenAI, Gemini and open-weights DeepSeekMath-V2 all reached gold medal level in math competitions. Deepmind model found new Navier Stokes solutions.
- Automation of research, labs, healthcare: AI Scientist-v2 (Sakana AI) autonomously formulated hypotheses, ran experiments and wrote a paper, GPT-5 (Red Queen Bio) optimized (79× efficiency gain) a gene-editing protocol, NVIDIA's Apollo physics model reported 10× faster design iterations for materials, FDA announced an AI framework to shorten the drug approval process, an AI platform (Doctronic) approved to assist in renewing prescriptions, OpenAI introduced ChatGPT Health, a Robot in a chemistry lab can speed up experiments 10x.
- US Genesis Mission: the USA to build an integrated AI platform to harness the world's largest collection of scientific datasets to train AI.

➔**Triggered from Catalysts:** 1,2,3,4,5,6,7,8

AI adoption in science,robotics (Predictions)

Trend Prediction 9

AI applications and models in science, robotics and related are exploding. New breakthroughs and adoption are expected.

By 2027 :
- A model from a lab/company not from China or USA will make a scientific breakthrough or reach gold level in a math competition.
- There will be at least a significant advancement using AI models in an unsolved scientific problem with a monetary prize (Millenium, Xprize etc.).
- Robot hardware will still be manufactured in China (60%+).
- The best AI robotics foundation model will score > 6 hours on CALVIN (chained tasks) benchmark and it will be Chinese.
- A large Pharma company will acquire an AI powered science discovery startup for $100M+.

➜

Technologies

BONUS: We will talk Pynglish → From 2024
Trend Prediction *Bonus* (semi-serious one)

Non coders will start picking up some code language.

ENG

- Developers are adopting copilot and alike AI software development tools quickly.
  https://survey.stackoverflow.co/2023/#ai

- Interns and others software related professionals (project managers etc.) learn to become proficient in coding faster using AI tools. https://bit.ly/498yVSn

- Demand for developers overall will not shrink (apart from front end).

→

By 2027 :

- 10% of non-coders professionals will at least try GenAI tools to make their own software scripts.
- Non-coders will pick up developer jargon in common life.
- An authoritative English dictionary (like Oxford English Dictionary) will add a software word as a new recognized word.

➜**Triggered from Catalysts:** 6,7,8

A Technologies

CONTACT US

andrea@aitechnologies.co

Andrea Isoni, Chief AI Officer, PhD

AI Technologies

TECHNOLOGIES
Bring the future to humanity