[Korea University AI Seminar]
The AI Architecture Decoded - Orchestrating Technology and Hardware Innovation

Sunghee Yun

Co-founder & CTO - AI Technology & Biz Dev @ Erudio Bio, Inc.

About Speaker

- *Co-founder & CTO @ Erudio Bio, San Jose & Novato, CA, USA*
- Advisor & Evangelist @ CryptoLab, Inc., San Jose, CA, USA
- Chief Business Development Officer @ WeStory.ai, Cupertino, CA, USA
- Advisory Professor, Electrical Engineering and Computer Science @ DGIST, Korea
- Adjunct Professor, Electronic Engineering Department @ Sogang University, Korea
- Global Advisory Board Member @ Innovative Future Brain-Inspired Intelligence System Semiconductor of Sogang University, Korea
- *KFAS-Salzburg Global Leadership Initiative Fellow @ Salzburg Global Seminar*, Salzburg, Austria
- Technology Consultant @ Gerson Lehrman Gruop (GLG), NY, USA
- *Co-founder & CTO & Head of Global R&D & Chief Applied Scientist & Senior Fellow @ Gauss Labs, Inc., Palo Alto, CA, USA* *2020 – 2023*

- Senior Applied Scientist @ Mobile Shopping Team, Amazon.com, Inc., Vancouver, BC, Canada – 2020
- Principal Engineer @ Software R&D Center of DS Division, Samsung, Korea – 2017
- Principal Engineer @ Strategic Marketing & Sales Team, Samsung, Korea – 2016
- Principal Engineer @ DT Team of DRAM Development Lab, Samsung, Korea – 2015
- Senior Engineer @ CAE Team - Samsung, Korea – 2012
- MS & PhD - Electrical Engineering @ Stanford University, CA, USA – 2004
- Development Engineer @ Voyan, Santa Clara, CA, USA – 2001
- BS - Electrical Engineering @ Seoul National University, Seoul, Korea – 1998

Highlight of Career Journey

- BS in EE @ SNU, MS & PhD in EE @ Stanford University
  - *Convex Optimization - Theory, Algorithms & Software*
  - advised by *Prof. Stephen P. Boyd*
- Principal Engineer @ Samsung Semiconductor, Inc.
  - AI & Convex Optimization
  - collaboration with *DRAM/NAND Design/Manufacturing/Test Teams*
- Senior Applied Scientist @ Amazon.com, Inc.
  - e-Commerce AIs - time-series anomaly detection, deep reinforcement learning & recommender system
  - Jeff Bezos's project - increase sales by *$200M* via Amazon Mobile Shopping App
- Co-founder & CTO & Head of Global R&D & Chief Applied Scientist & Senior Fellow @ Gauss Labs, Inc.
- Co-founder & CTO - AI Technology & Business Development @ Erudio Bio, Inc.

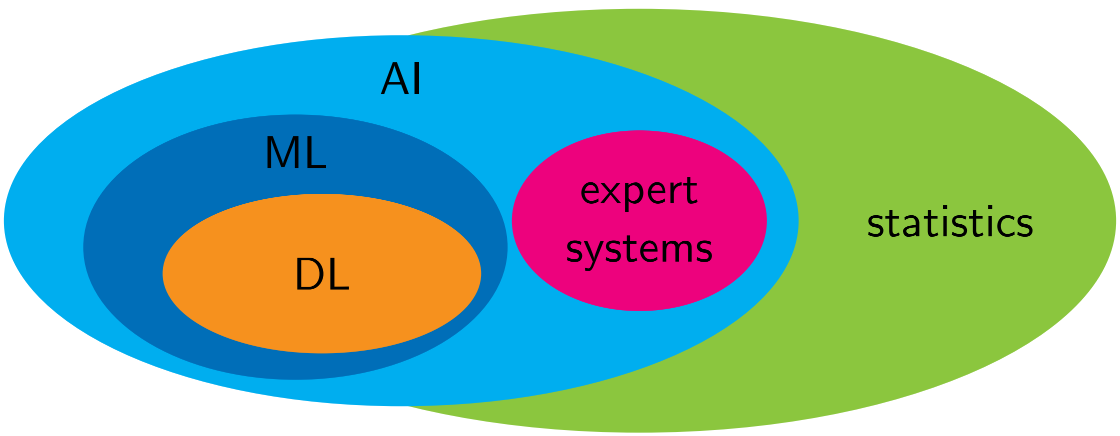
Today

Artificial Intelligence

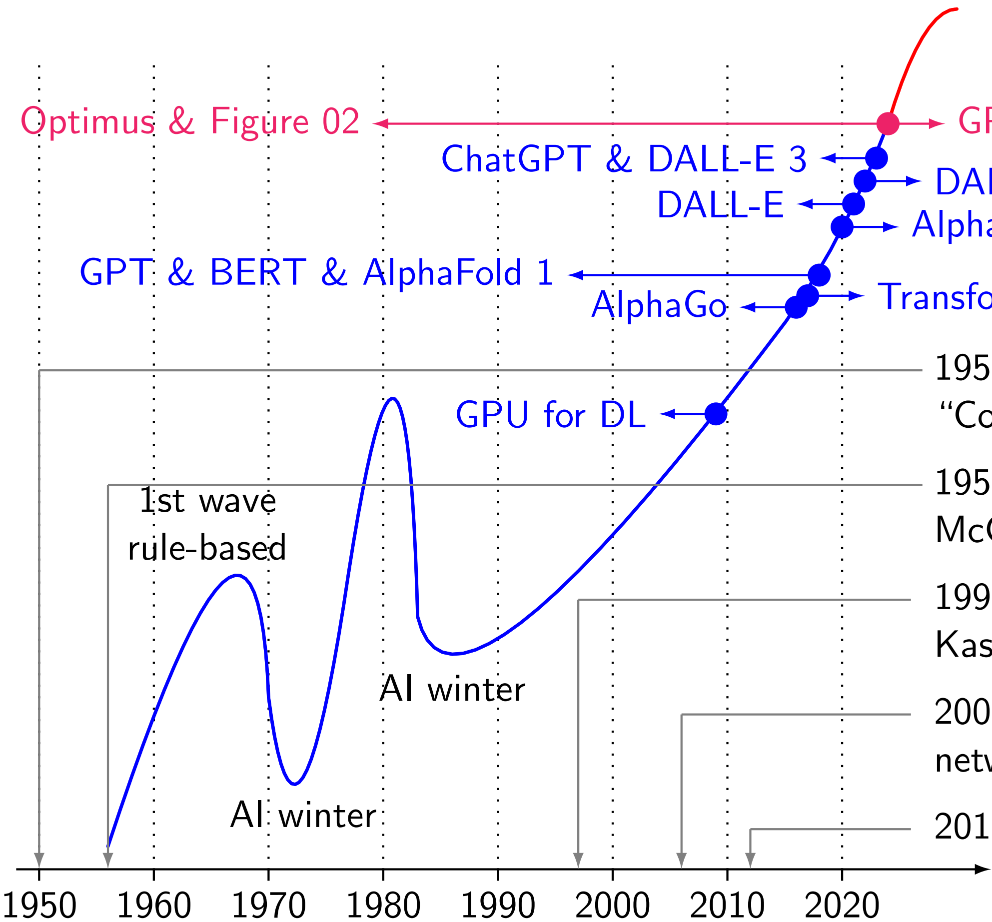
Definition and History

Definition of AI

- AI is
  - technology enabling machines to do tasks requiring human intelligence, such as learning, problem-solving, decision-making & language understanding
  - *not* one thing - encompass range of technologies, methodologies & applications
- relationship of AI, statistics, ML, DL, NN & expert system [HGH+22]



History of AI



1950 - Turing Test introduced by Alan Turing in "Computing Machinery & Intelligence"

1956 - "Artificial Intelligence" coined by John McCarthy

1997 - IBM's Deep Blue defeated Garry Kasparov, (then) the world chess champion

2006 - DL Learning gained traction deep belief networked introduced by Geoffrey Hinton

2012 - AlexNet won ImageNet Competition (ILSVRC) - widespread attention on DL

Significant AI Achievements - 2014 – 2024

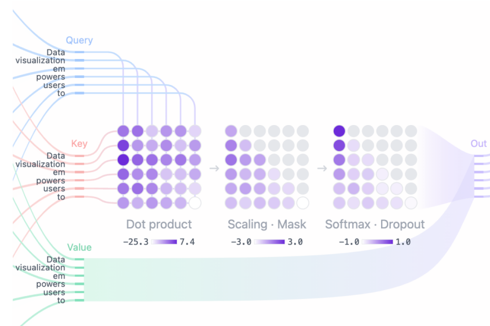
Deep learning revolution

- 2012 – 2015 - DL revolution[1]
  - CNNs demonstrated exceptional performance in image recognition, *e.g.*, *AlexNet's victory in ImageNet competition*
  - widespread adoption of DL learning in CV transforming industries
- 2016 - AlphaGo defeats human Go champion
  - DeepMind's AlphaGo defeated world champion in Go, extremely complex game *believed to be beyond AI's reach*
  - significant milestone in RL - AI's potential in solving complex & strategic problems

[1] DL: deep learning, CNN: convolutional neural network, CV: computer vision, RL: reinforcement learning

Transformer changes everything

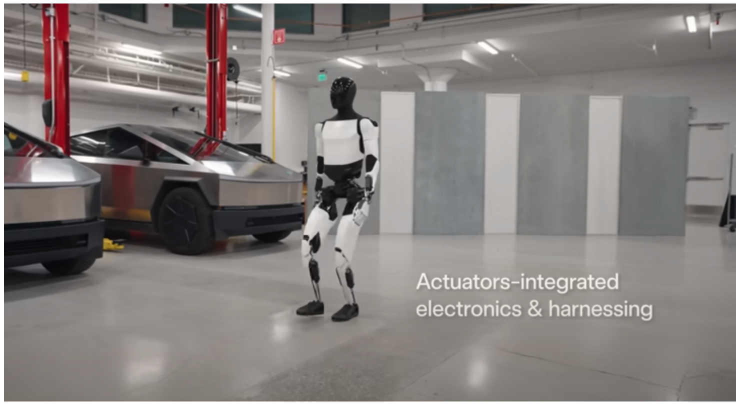
- 2017 – 2018 - Transformers & NLP breakthroughs[2]
  - *Transformer (e.g., BERT & GPT) revolutionized NLP*
  - major advancements in, *e.g.*, machine translation & chatbots
- 2020 - AI in healthcare – AlphaFold & beyond
  - DeepMind's *AlphaFold solves 50-year-old protein folding problem* predicting 3D protein structures with remarkable accuracy
  - accelerates drug discovery and personalized medicine - offering new insights into diseases and potential treatments



[2] NLP: natural language processing, GPT: generative pre-trained transformer

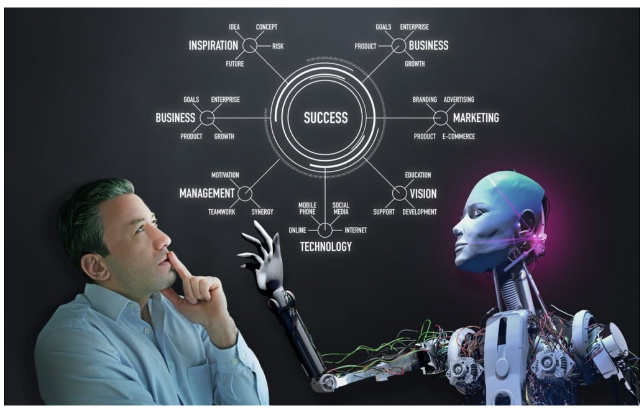
Lots of breakthroughs in AI technology and applications in 2024

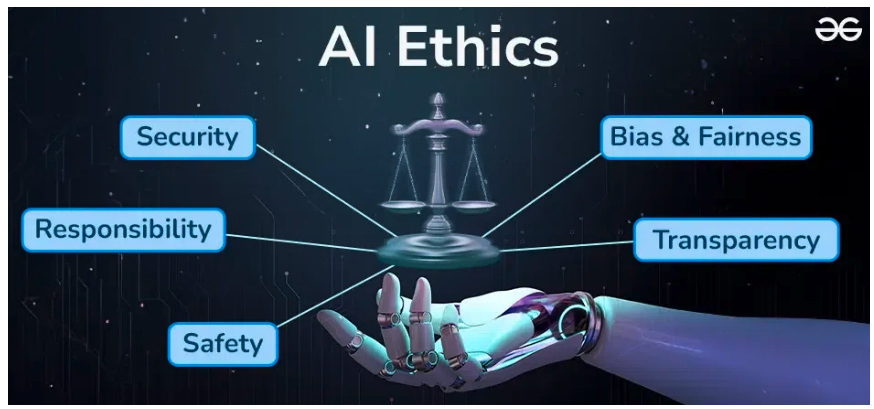
- proliferation of advanced AI models
  - GPT-4o, Claude Sonnet, Llama 3, Sora
  - *transforming industries* such as content creation, customer service, education, *etc.*
- breakthroughs in specialized AI applications
  - Figure 02, Optimus, AlphaFold 3
  - driving unprecedented advancements in automation, drug discovery, scientific understanding - *profoundly affecting healthcare, manufacturing, scientific research*



Transformative impact of AI - reshaping industries, work & society

- accelerating human-AI collaboration
  - not only reshaping industries but *altering how humans interact with technology*
  - AI's role as collaborator and augmentor redefines productivity, creativity, the way we address global challenges, $e.g.$, *sustainability & healthcare*
- AI-driven automation *transforms workforce dynamics* - creating new opportunities while challenging traditional job roles
- *ethical AI considerations* becoming central not only to business strategy, but to society as a whole - *influencing regulations, corporate responsibility & public trust*

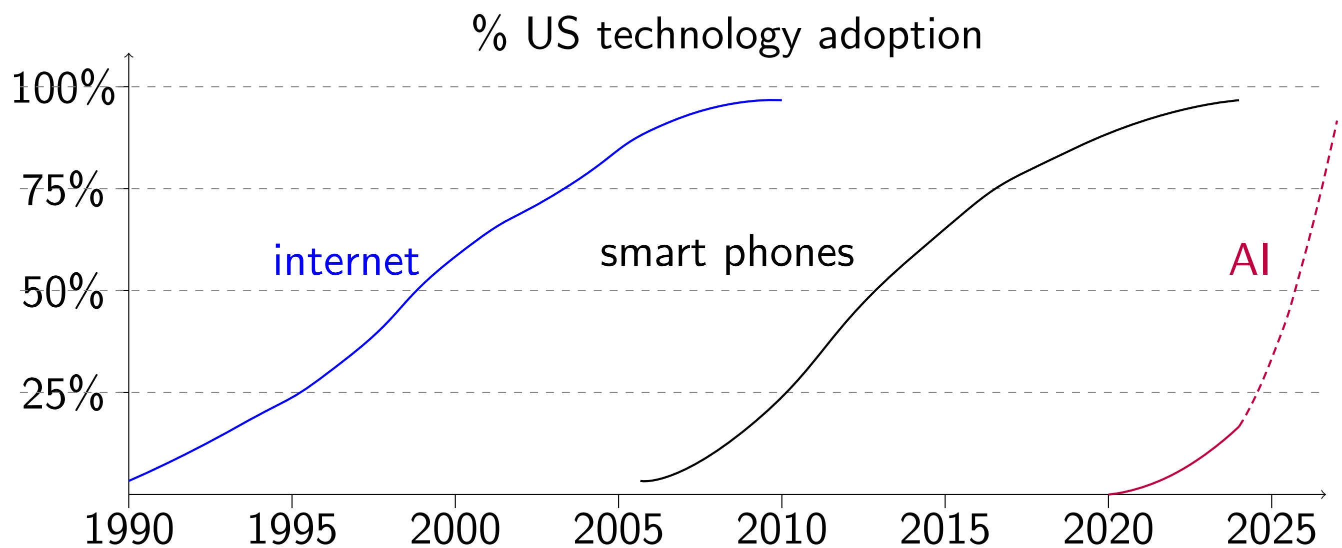




Recent Advances in AI

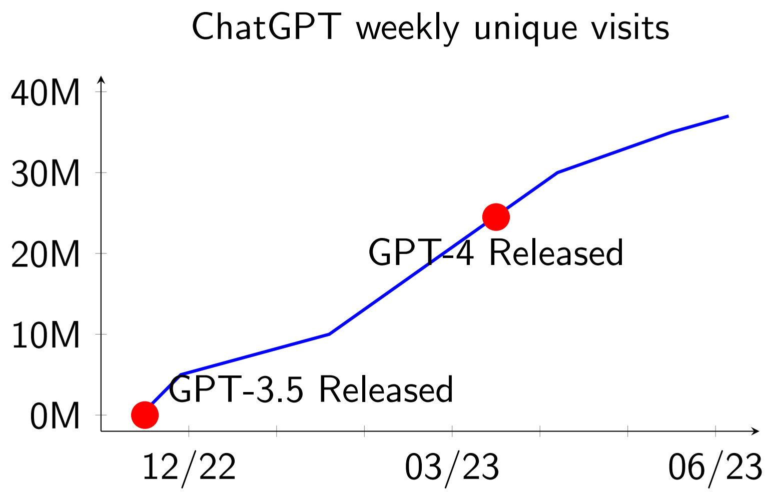
Where are we in AI today?

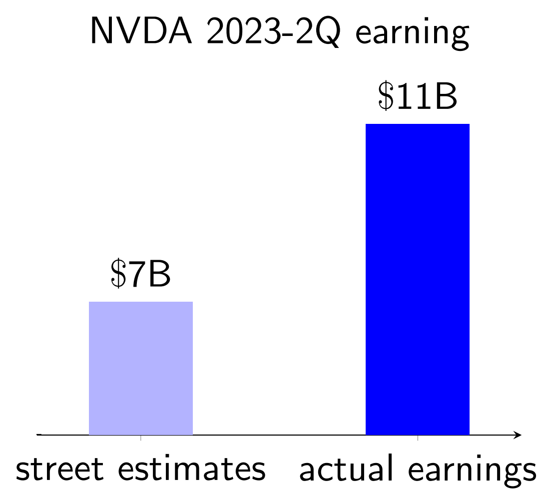
- sunrise phase - currently experiencing dawn of AI era with significant advancements and increasing adoption across various industries
- early adoption - in early stages of AI lifecycle with widespread adoption and innovation across sectors marking significant shift in technology's role in society



Explosion of AI ecosystems - ChatGPT & NVIDIA

- took only *5 months for ChatGPT users to reach 35M*
- NVDIA 2023 Q2 earning exceeds market expectation by big margin - $7B vs $13.5B
  - surprisingly, *101% year-to-year growth*
  - even more surprisingly *gross margin was 71.2%* - up from 43.5% in previous year[3]

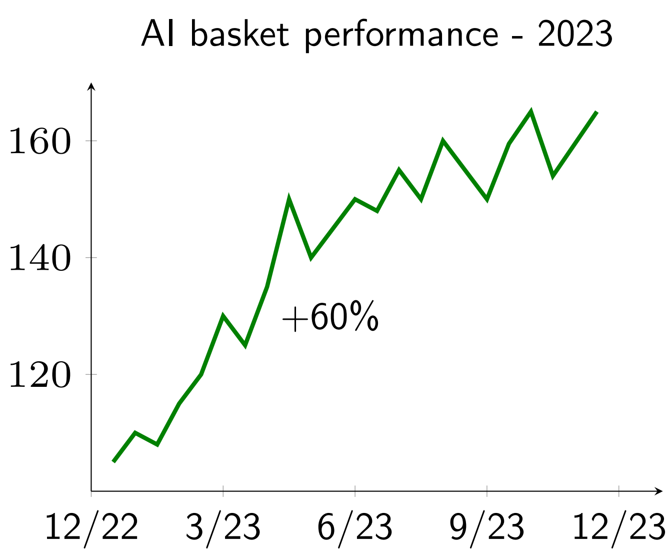


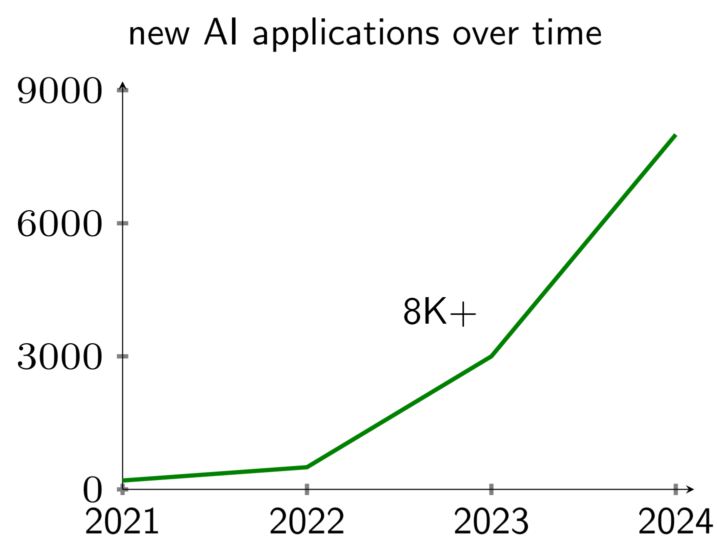


[3] source - Bloomberg

Explosion of AI ecosystems - AI stock market

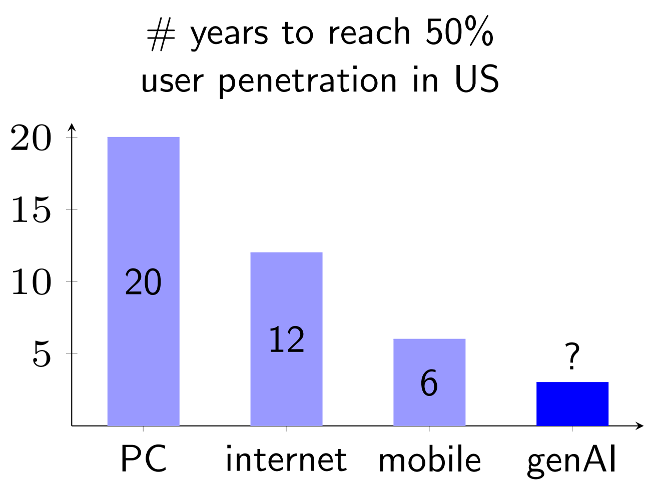
- *AI investment surge in 2023 - portfolio performance soars by 60%*
  - AI-focused stocks significantly outpaced traditional market indices
- *over 8,000 new AI applications* developed in last 3 years
  - applications span from healthcare and finance to manufacturing and entertainment

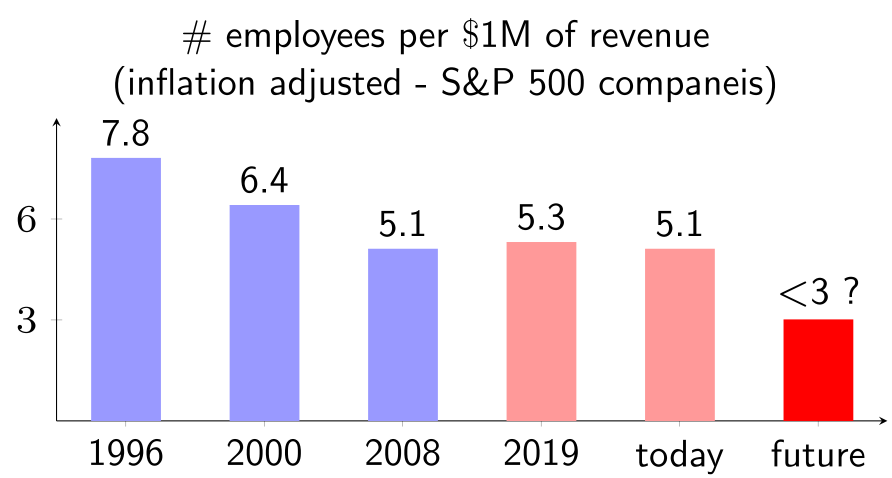




AI's transformative impact - adoption speed & economic potential

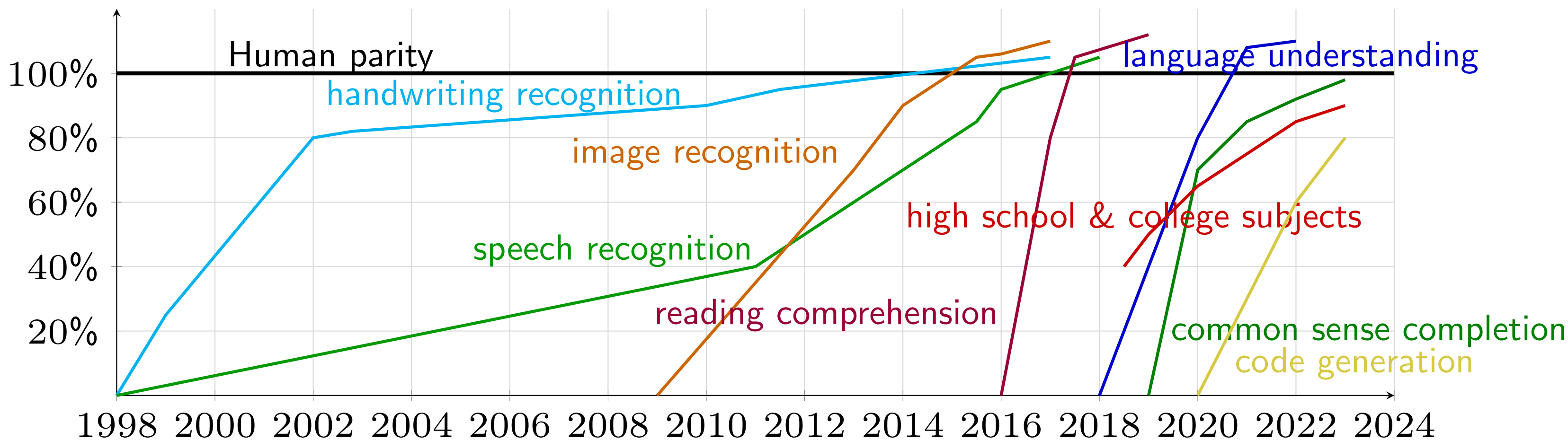
- adoption - has been twice as fast with platform shifts suggesting
  - increasing demand and readiness for new technology improved user experience & accessibility
- AI's potential to drive economy for years to come
  - 35% improvement in productivity driven by introduction of PCs and internet
  - greater gains expected with AI proliferation





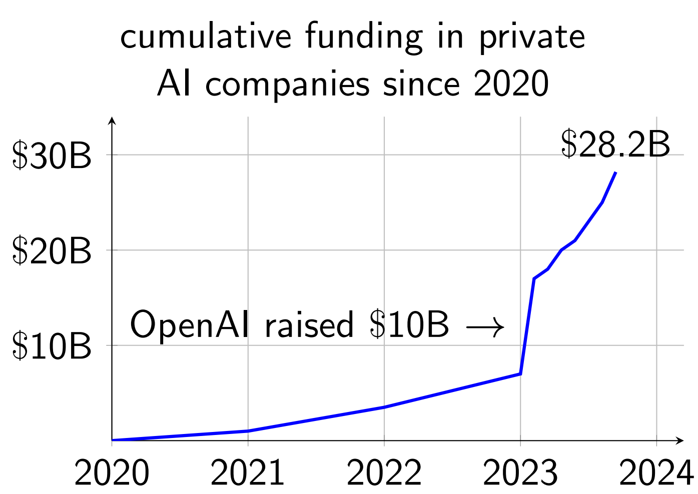
AI getting more & more faster

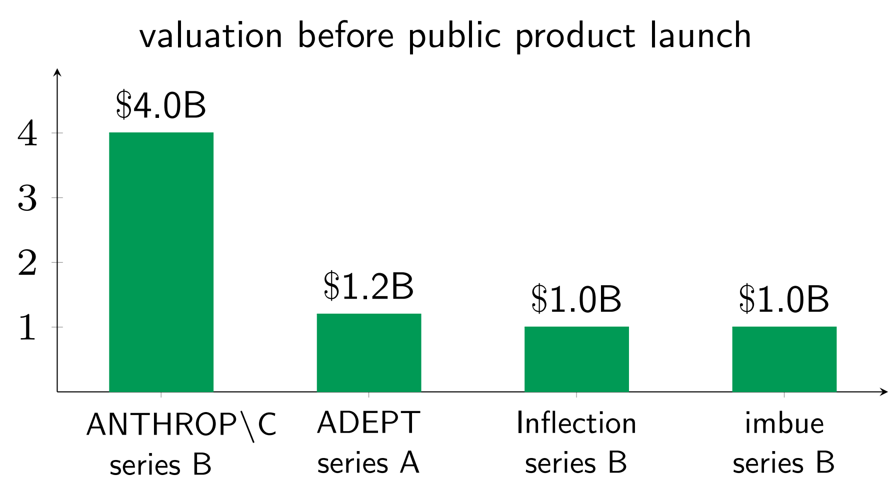
- steep upward slopes of AI capabilities highlight accelerating pace of AI development
  - period of exponential growth with AI potentially mastering new skills and surpassing human capabilities at ever-increasing rate
- closing gap to human parity - some capabilities approaching or arguably reached human parity, while others having still way to go
  - achieving truly human-like capabilities in broad range remains a challenge



Massive investment in AI

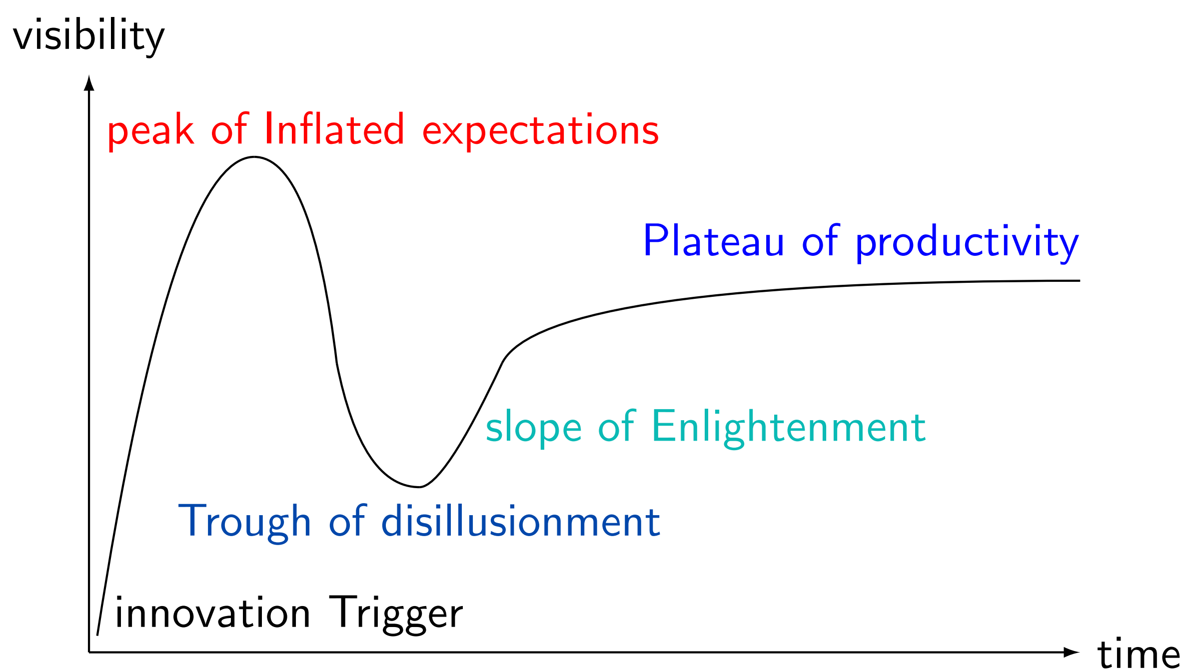
- *explosive growth* - cumulative funding skyrocketed reaching staggering $28.2B
- OpenAI - significant fundraising (=$10B) fueled rapid growth
- *valuation surge* - substantial valuations even before public products for stella companies
- *fierce competition for capital* among AI startups driving innovation & accelerating development
- massive investment indicates *strong belief in & optimistic outlook for potential of AI* to revolutionize industries & drive economic growth





Is AI hype?

Technology hype cycle



- innovation trigger - technology breakthrough kicks things off
- peak of inflated expectations - early publicity induces many successes followed by even more
- trough of disillusionment - expectations wane as technology producers shake out or fail
- slope of enlightenment - benefit enterprise, technology better understood, more enterprises fund pilots

Fiber vs cloud infrastructure

- fiber infrastructure - 1990s
  - Telco Co's raised \$1.6T of equity & \$600B of debt
  - bandwidth costs decreased 90% within 4 years
  - companies - Covage, NothStart, Telligent, Electric Lightwave, 360 networks, Nextlink, Broadwind, UUNET, NFS Communications, Global Crossing, Level 3 Communications
  - became *public good*

- cloud infrastructure - 2010s
  - entirely new computing paradigm
  - mostly public companeis with data centers
  - *big 4 hyperscalers generate* \$150B + annual revenue

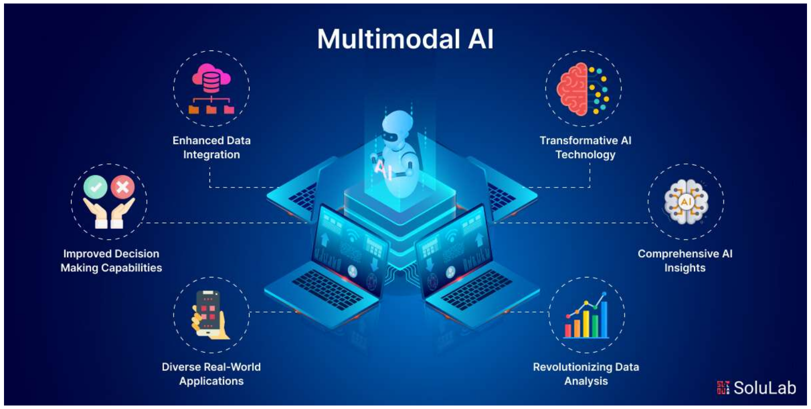
Yes & No

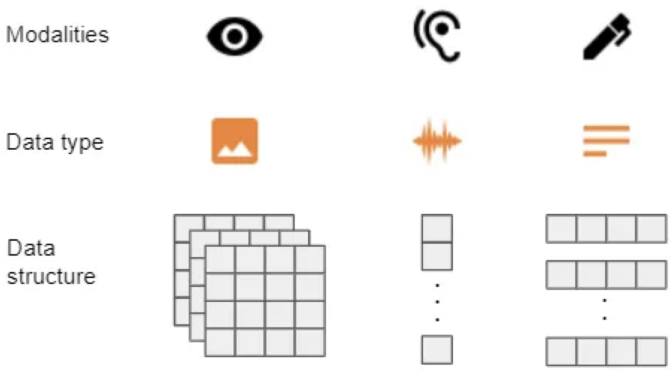
characteristics of hype cycles	speaker's views
value accrual misaligned with investment	• OpenAI still operating at a loss; business model *still* not clear • gradual value creation across broad range of industries and technologies (*e.g.*, CV, LLMs, RL) unlike fiber optic bubble in 1990s
overestimating timeline & capabilities of technology	• self-driving cars delayed for over 15 years, with limited hope for achieving level 5 autonomy • AI, however, has proven useful within a shorter 5-year span, with enterprises eagerly adopting
lack of widespread utility due to technology maturity	• AI already providing significant utility across various domains • vs quantum computing remains promising in theory but lacks widespread practical utility

AI Agents

Multimodal learning

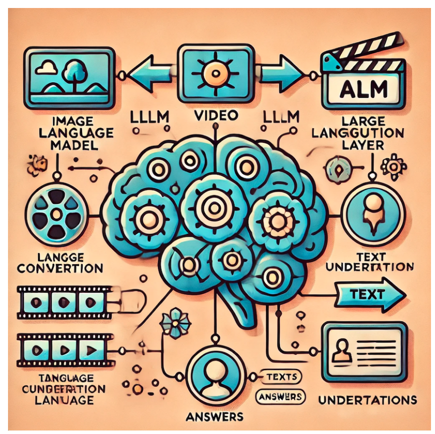
- understand information from multiple modalities, *e.g.*, text, images, audio, video
- representation learning methods
  - combine multiple representations or learn multimodal representations simultaneously
- applications
  - images from text prompt, videos with narration, musics with lyrics
- collaboration among different modalities
  - understand image world (open system) using language (closed system)

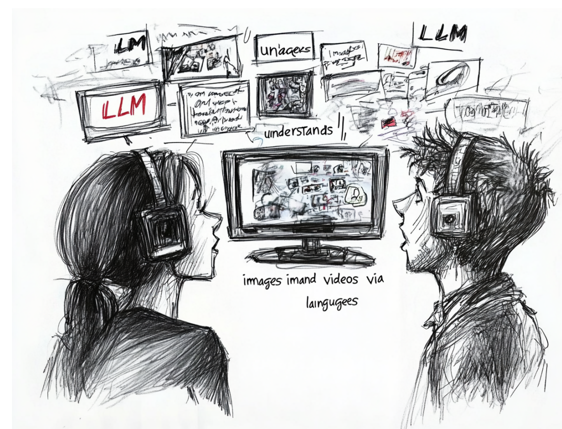




Implications of success of LLMs

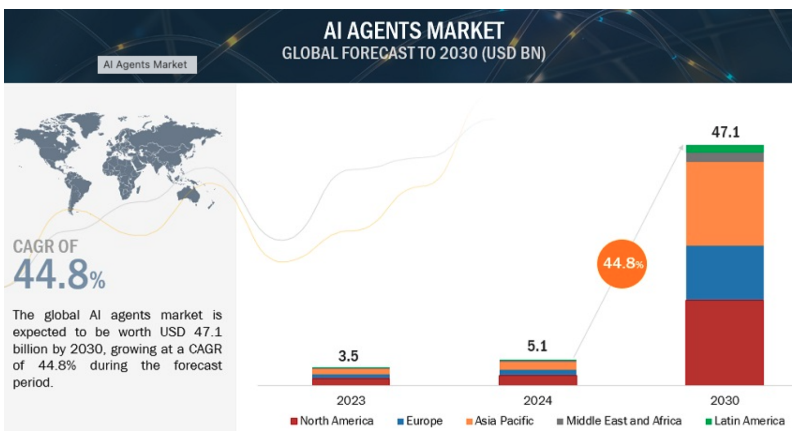
- many researchers change gears towards LLM
  - from computer vision (CV), speach, music, video, even reinforcement learning
- *LLM is not only about NLP . . .* humans have . . .
  - evolved to optimize natural language structures for eons
  - handed down knowledge using *this natural languages* for thousands of years
  - internal structure (or equivalently, representation) of natural languages optimized via *thousands of generation by evolution*
- LLM *connects non-linguistic world (open system) via natural languages (closed system)*

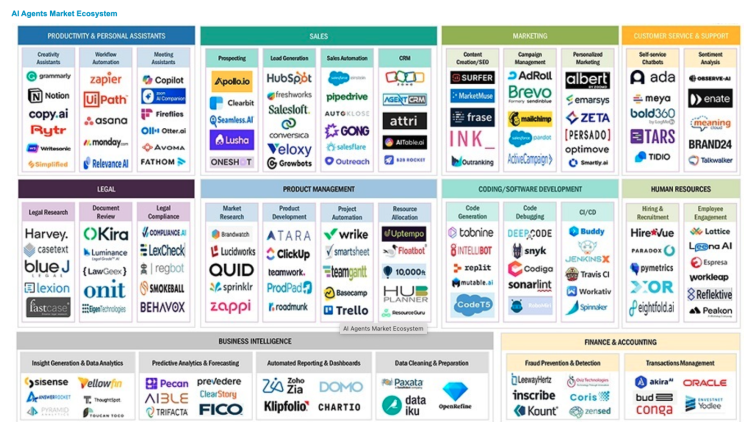




Multimodal AI (mmAI) - definition & history

- mmAI - systems processing & integrating data from multiple sources & modalities, to generate unified response / decision
- 1990s – 2000s - early systems - initial research combining basic text & image data
- 2010s - CNNs & RNNs enabling more sophisticated handling of multimodality
- 2020s - modern multimodal models - Transformer-based architectures handling complex multi-source data at highly advanced level
- mmAI *mimics human cognitive ability* to interpret and integrate information from various sources, leading to holistic decision-making



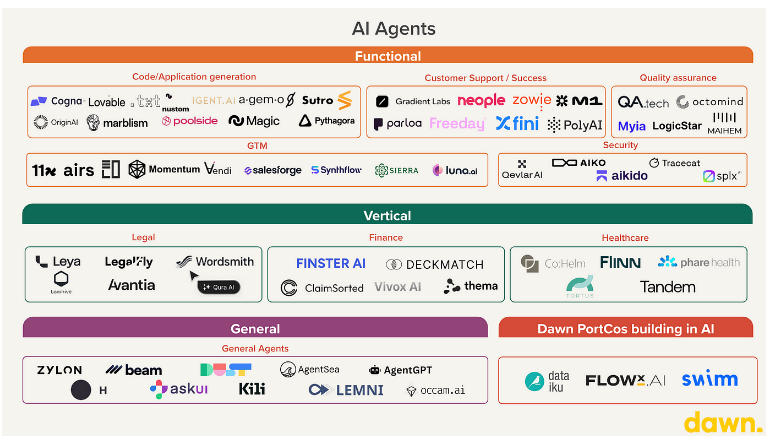


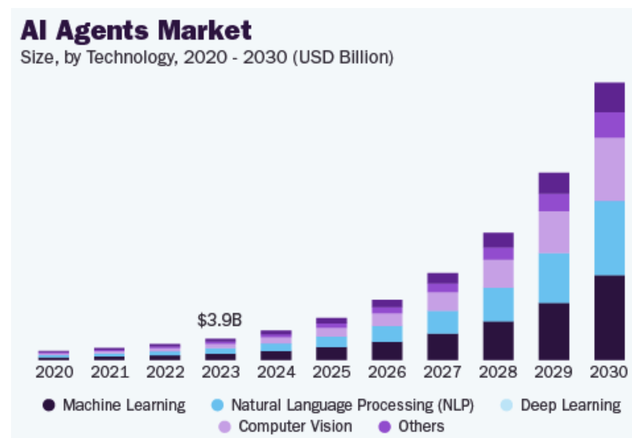
mmAI Technology

- core components
  - data preprocessing - images, text, audio & video
  - architectures - unified Transformer-based ($e.g.$, ViT) & cross-attention mechanisms / hybrid architectures ($e.g.$, CNNs + LLMs)
  - integration layers - fusion methods for combining data representations from different modalities
- technical challenges
  - data alignment - accurate alignment of multimodal data
  - computational demand - high-resource requirements for training and inferencing
  - diverse data quality - manage variations in data quality across modalities
- advancements
  - multimodal embeddings - shared feature spaces interaction between modalities
  - self-supervised learning - leverage unlabeled data to learn representations across modalities

AI agents powered by multimodal LLMs

- foundation
  - integrate multimodal AI capabilities for enhanced interaction & decision-making
- components
  - perceive environment through multiple modalities (visual, audio, text), process using LLM technology, generate contextual responses & take actions
- capabilities
  - understand complex environments, reason across modalities, engage in natural interactions, adapt behavior based on context & feedback





AI agents - Present & Future

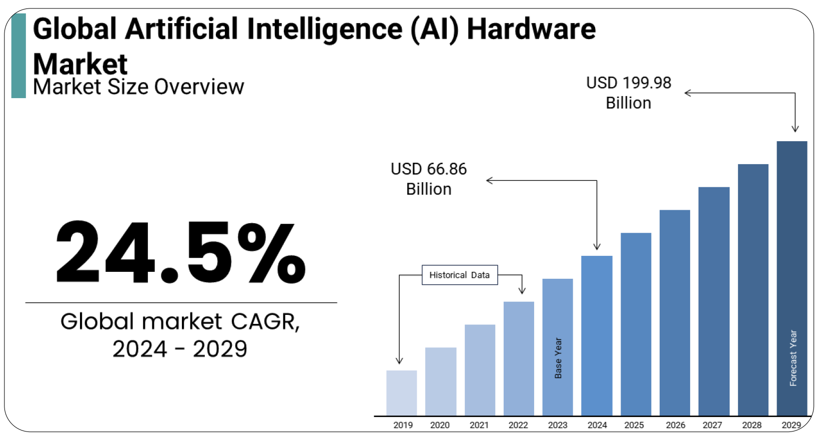
- emerging applications
  - scientific research - agents analyzing & running experiments & generating hypotheses
  - creative collaboration - AI partners in design & art combining multiple mediums
  - environmental monitoring - processing satellite sensor data for climate analysis
  - healthcare - enhanced diagnostic combining imaging, *e.g.*, MRI, with patient history
  - customer experience - virtual assistants understanding spoken language & visual cues
  - autonomous vehicles - integration of visual, radar & audio data
- future
  - ubiquitous AI agents - seamless integration into everyday devices
  - highly tailored personalized experience - in education, entertainment & healthcare

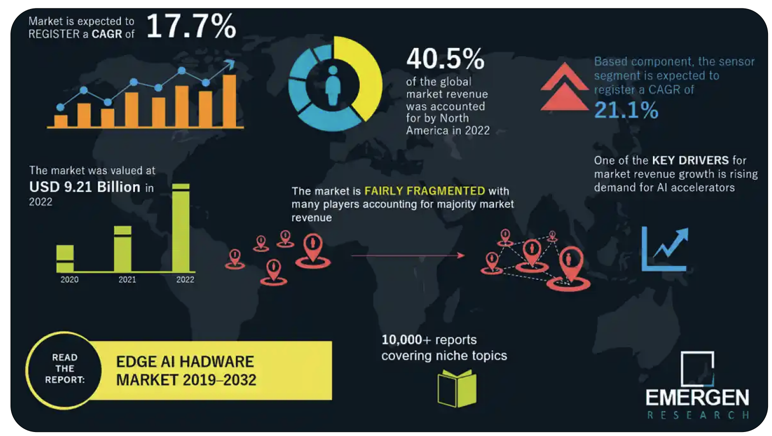
AI Hardware

AI Hardware Industry

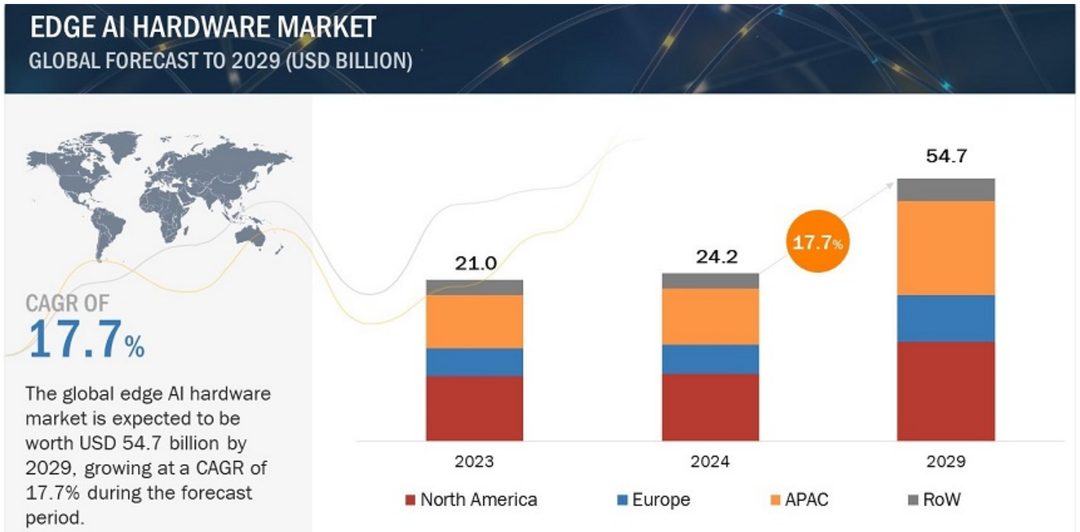
Landscape of AI hardware industry

- global AI hardware market valued at $66.96B in 2024, projected to grow significantly
- major companies - Nvidia, Intel, AMD, Qualcomm, and IBM w/ Nvidia holding substantial market share



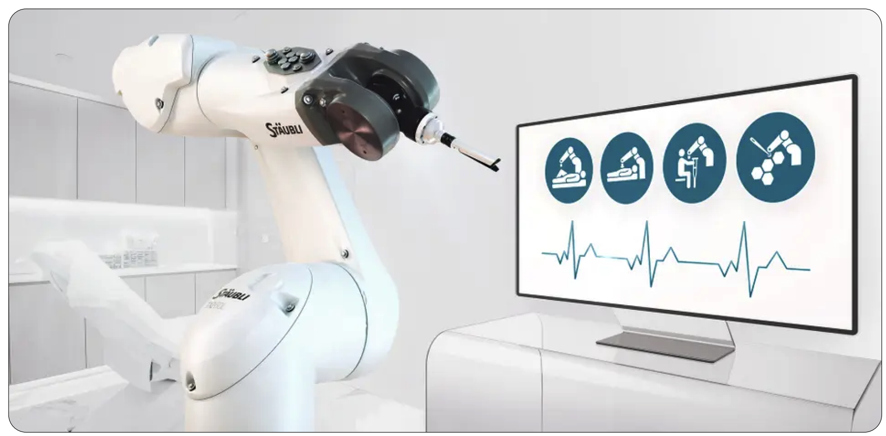


- North America leading market - high R&D investments & key industry players
- Asia Pacific rapidly expanding - strong semiconductor industries in South Korea, China & Japan
- demand for advanced processors such as GPUs, TPUs & AI accelerators rising due to complexity of AI algorithms & high computational power



Predictions for future of AI hardware market

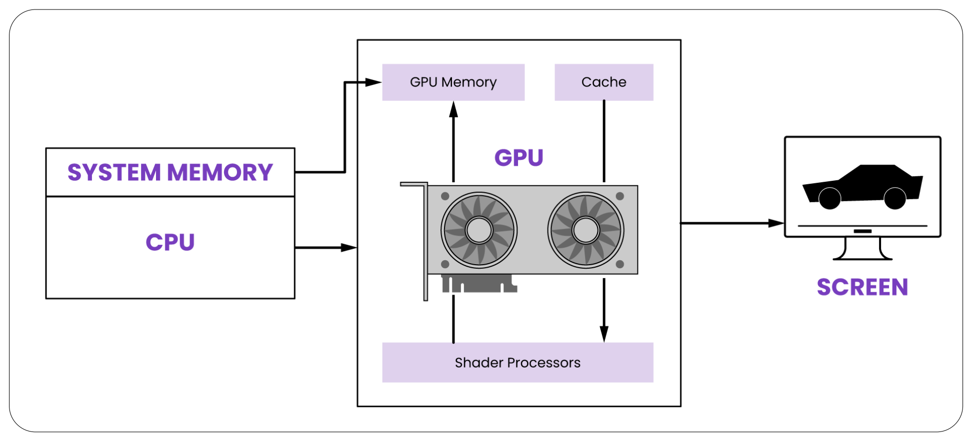
- AI hardware market expected to reach $382B by 2032 - significant growth in data center AI chips
- integration of AI w/ 5G & increased use of AI in edge computing anticipated to drive future demand
- AI hardware becoming crucial in sectors such as autonomous vehicles, robotics & medical devices
- need to address challenges such as heat and power management along with technical complexities



GPUs and AI Accelerators

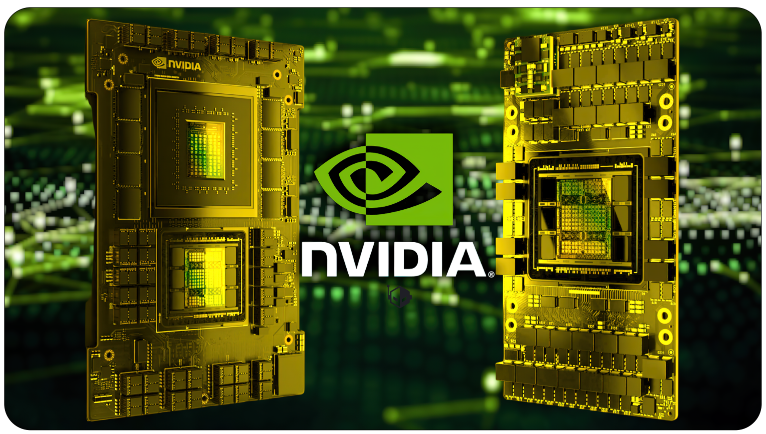
Technical challenges of GPUs & AI accelerators

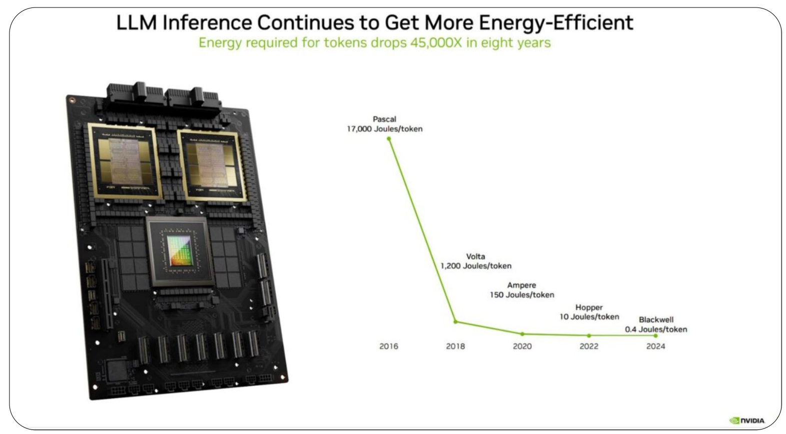
- facing challenges in scaling to handle increasingly large AI models and datasets - traditional architectures struggling w/ massive parallel processing demands of modern AI applications
- AI applications require extensive memory bandwidth often leading to bottlenecks - efficient memory management is crucial
- AI accelerators consume significant power - high operational costs and environmental concerns for both cloud-based & edge AI applications



Potential solutions for overcoming challenges

- development of AI-specific architectures such as tensor cores and custom ASICs to improve efficiency and performance - novel architectures like FPGAs for specific AI tasks, *e.g.*, for RAG & vectorDB
- implementing software optimizations to enhance hardware usability and performance - use of compilers and frameworks that maximize efficiency of existing hardware
- encouraging market competition to drive innovation and reduce monopolistic control - exploring alternative hardware solutions and improving energy efficiency standards





Big tech's in-house chip development

- shift towards in-house AI hardware - major tech companies increasingly developing their own AI chips - move to enhance AI capabilities and reduce dependence
- collaboration with specialized partners - partnering with specialized firms for manufacturing and technology blending in-house expertise with external innovation

	Microsoft	Google	Amazon	Meta
Chip	Maia 100	TPU v5e	Inferentia2	MTIA v1
Launch Date	November, 2023	August, 2023	Early 2023	2025
IP	ARM	ARM	ARM	RISC-V
Process Technology	TSMC 5nm	TSMC 5nm	TSMC 7nm	TSMC 7nm
Transistor Count	105 billion	-	-	-
INT8	-	393 TOPS	-	102.4 TOPS
FP16	-	-	-	51.2 TFLOPS
BF16	-	197 TFLOPS	-	-
Memory	-	-	-	LPDDR5
TDP	-	-	-	25W
Packaging Technology	CoWoS	CoWoS	CoWoS-S	2D
Collaborating Partners	Global Unichip Corp.	Broadcom	Alchip Technologies	Andes Technology
Application	Training/Inference	Inference	Inference	Training/Inference
LLM	GPT-3.5, GPT-4	BERT, PaLM, LaMDA	Titan FM	Llama, Llama2

AMD - Nvidia's new competitor

- key points
  - AMD launched new AI accelerator chip, *Instinct MI300X*, on Dec 6, 2023
  - CDNA 3 architecture, mix of 5nm and 6nm IPs, delivering 153B transistors
  - *outperforms Nvidia's H100 TensorRT-LLM* by 1.6X higher memory bandwidth and 1.3X FP16 TFLOPS
  - up to 40% faster vs Nvidia's Llama-2 70B model in 8x8 server configurations
- market impact
  - significant challenge to Nvidia's dominance in AI accelerator market
  - performance gains over Nvidia's offerings could drive *customer adoption and market share for AMD*
- future prediction
  - *AMD stocks soared* since launch indicating investor confidence in their competitiveness
  - Lisa Su, AMD's CEO, categorized Instinct MI300X as "next big thing" in tech industry
  - potential risks include need to *manage ROCm vs CUDA software ecosystem* & ensure rapid customer adoption and production coverage

AI Accelerator Startups

AI accelerator startups

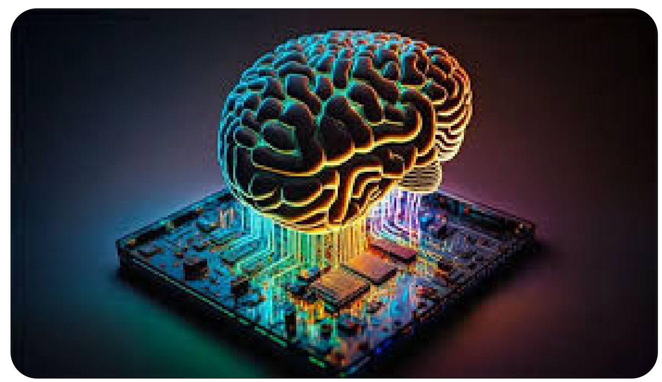
- innovative architectures - startups like Groq, SambaNova & Graphcore leading with *novel architectures designed to accelerate AI workloads*
  - *Groq* - tensor streaming processor (TSP) offering ultra-low latency & high throughput, high-performance AI inference chips enhancing speed & efficiency
  - *SambaNova* - reconfigurable dataflow architecture optimizing for various AI workloads
  - *Graphcore* - intelligence processing unit (IPU) tailored for graph-based computation excelling in sparse data processing
  - *Cerebras Systems* - develop wafer scale engine (WSE), largest chip built for AI workloads, unmatched computational power revolutionizing AI hardware capabilities
  - *Hailo* - specialize for edge devices optimizing AI processes for real-time applications, raised $120M emphasizing potential to disrupt traditional AI chip markets

Technological competitiveness

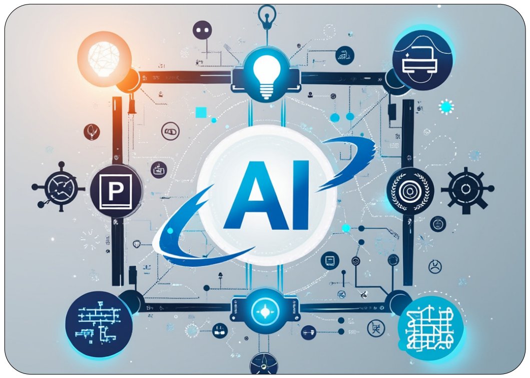
- energy efficiency
  - energy-efficient designs crucial for scalability in data centers and edge devices
  - startups developing solutions significantly reducing power consumption without compromising performance
- customization & flexibility
  - AI accelerators from startups often offer greater customization options for specific AI tasks compared to traditional GPUs
  - flexibility in hardware allows for tailored solutions that can outperform general-purpose accelerators in certain applications
- software integration
  - robust software ecosystems critical - startups investing in developing software stacks that optimize performance for their hardware
  - compatibility with existing AI frameworks is competitive advantage, *e.g.*, `TensorFlow` & `PyTorch`

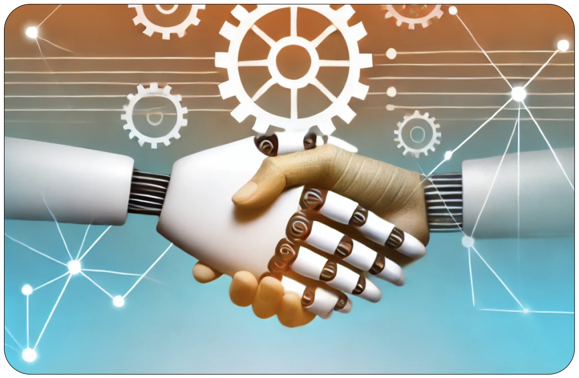
Industry and market influence

- disruption of traditional players
  - challenging dominance of established players like NVIDIA & Intel
  - unique architectures providing specialized solutions traditional GPUs and CPUs cannot efficiently handle
- driving down costs
  - offering competitive alternatives pushing down cost of AI computation
  - could lead to democratization of AI w/ more companies affording high-performance AI capabilities

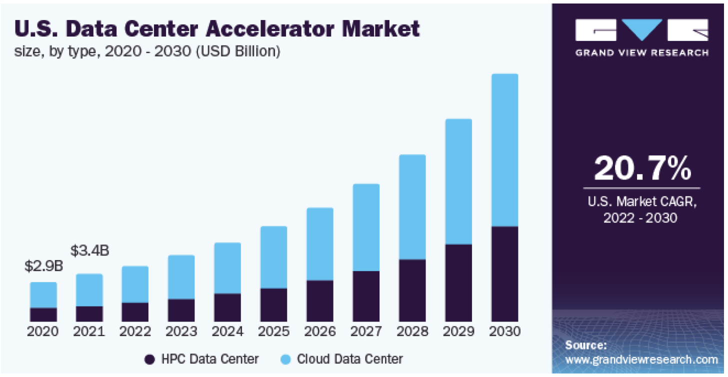


- accelerating AI innovation
  - contributing to rapid innovation providing hardware that can handle emerging AI models & workloads
  - adaptability and specialization enable advancements in AI research & faster development cycles
- strategic partnerships & acquisitions
  - big techs increasingly forming strategic partnerships or acquiring startups to stay competitive
  - collaborations can speed up integration of advanced AI hardware into mainstream products





- market growth & opportunities
  - AI accelerator market expected to grow significantly driven by demand in data centers, edge computing & autonomous systems
  - startups well-positioned to capture significant share of growing market particularly in niche applications
- future outlook
  - dependency on Asia for fabrication might lead to strategic shifts in global tech policies and investments in local manufacturing
  - increasing demand for efficient AI processing on edge devices and in data center.



Global Semiconductor Industry

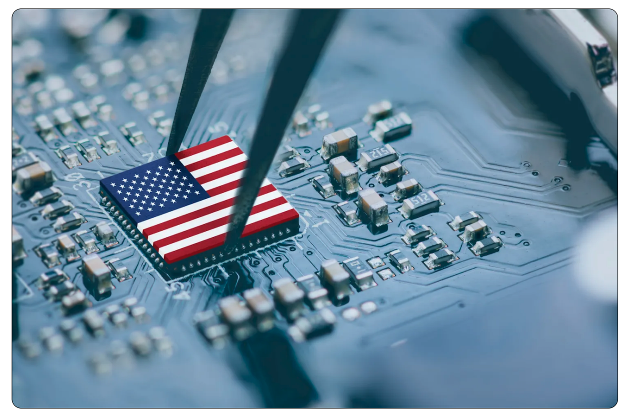
Hard-to-predict AI hardware markets

- US
  - birthplace for modern semiconductor chips driving PC market, internet, multi-media, mobile phones, and AI . . .
    - Intel, Texas Instrument (TI), Global Foundry
  - traditionally strong with design houses - NVIDIA, AMD, Broadcom, Apple, . . .
  - threatened experiencing global chip shortage & vulnerable supply chain via COVID
  - national security concerns & economic competitiveness
- China
  - strong fast followers - SMIC[4], Huawei, Hua Hong Semiconductor (foundry)
- South Korea
  - best memory chip makers - Samsung, SK hynix
  - struggling with LSI and foundry business

[4]SMIC - Semiconductor Manufacturing International Corporation

Reshoring semiconductor manufacturing industry

- trade & semiconductor WAR between US & China
  - export controls on advanced chips and equipment
- CHIPS & Science Act (Aug, 2022)
  - $52B in subsidies for domestic production, 25% investment tax credit for chip plants
  - (coerce) world-best semiconductor manufacturers build factories in US with support
    - GlobalFoundries - $1.5B @ Feb-2024
    - Intel - $8.5B @ Apr-2024 - Ohio - two fabs expandable to $100B
    - Samsung - $6.4B @ Apr-2024 - Talor, Texas
    - TSMC - $6.6B @ Apr-2024 - Phoenix, Arizona - two foundry fabs (3nm & 4nm)



Turmoils in global semiconductor business

- global context
  - EU Chips Act - €43B to boost European chip production
  - Japan & South Korea - significant investments in domestic capacity
- industry dynamics
  - Intel's foundry ambitions - targeting 50% global market share by 2030
  - TSMC expanding global footprint (US, Japan, possibly Germany)
- future outlook
  - projected shift in global semiconductor manufacturing landscape
  - increased geographical diversification of chip production

Export controls on US chip technology to China

- goal - limit China's access to advanced semiconductor tech to maintain US strategic advantage
- impacts on
  - China - advanced chips and equipment not allowed, domestic innovation increased
  - US - short-term - US lose market share and revenue in China
  - US - long-term - potential decline in US global competitiveness
- Chinese response - circumvent controls and adapt supply chains
- conclusion
  - US-China chip rivalry transforms global supply chains with deep implications for *security & industry*
  - US success hinges on better coordination and policy analysis

- reference - Balancing the Ledger - Center for Strategic & International Studies (CSIS)

China strikes back on US sanction

- Huawei's launch of Mate 60 Pro smartphone
  - these domestically produced chips represent major breakthrough against US sanctions
  - its success with *advanced 7nm Kirin 9000S chip* demonstrates significant progress in China's self-reliance in high-tech manufacturing - narrowing the technological gap with global leaders
- Huawei case highlights potential failure of US sanctions potentially leading to more aggressive US measures
  - US export controls on China's semiconductor industry are effective in the short term but insufficient to halt China's progress especially in legacy chip manufacturing
  - to maintain technological edge, US must balance further restrictions with supporting its semiconductor industry to avoid overreliance on export controls

Chinese semiconductor companies

- Chinese major semiconductor companies
  - SMIC - China's largest chip foundry, advancing 7nm technology
  - HiSilicon - Huawei's chip design arm, crucial for the Kirin processors
  - YMTC - leader in 3D NAND memory chip production
  - Huahong Group, CXMT, SMEE, GigaDevice, UnilC Semiconductors, ASMC, *etc.*
- *SMIC shows significant progress in producing 7nm chips* & YMTC leads memory chip manufacturer - both face challenges from US export controls
- industry faces internal challenges, *e.g.*, corruption & misallocation of resources
- but remains crucial to China's goal of technological self-reliance

Selected References & Sources

Selected references & sources

- Chris Miller “Chip War: The Fight for the World’s Most Critical Technology” (2022)
- Daniel Kahneman “Thinking, Fast and Slow” (2011)
- M. Shanahan “Talking About Large Language Models” (2022)
- A.Y. Halevry, P. Norvig, and F. Pereira “Unreasonable Effectiveness of Data” (2009)
- A. Vaswani, et al. “Attention is all you need” @ NeurIPS (2017)
- S. Yin, et. al. “A Survey on Multimodal LLMs” (2023)
- I.J. Goodfellow, ..., Y. Bengio “Generative adversarial networks (GAN)” (2014)
- T. Kuiken “Artificial Intelligence in the Biological Sciences: Uses, Safety, Security, and Oversight” (2023)
- Stanford Venture Investment Groups
- CEOs & CTOs @ startup companies in Silicon Valley
- VCs on Sand Hill Road - Palo Alto, Menlo Park, Woodside in California, USA

References

References

[HGH+22] Sue Ellen Haupt, David John Gagne, William W. Hsieh, Vladimir Krasnopolsky, Amy McGovern, Caren Marzban, William Moninger, Valliappa Lakshmanan, Philippe Tissot, and John K. Williams. The history and practice of AI in the environmental sciences. *Bulletin of the American Meteorological Society*, 103(5):E1351 – E1370, 2022.

[Mil22] Chris Miller. *Chip war: fight for the world's most critical technology*. New York: Scribner, 2022.

[MLZ22] Louis-Philippe Morency, Paul Pu Liang, and Amir Zadeh. Tutorial on multimodal machine learning. In Miguel Ballesteros, Yulia Tsvetkov, and Cecilia O. Alm, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, pages 33–38, Seattle, United States, July 2022. Association for Computational Linguistics.

[YFZ+24] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on multimodal large language models, 2024.

Thank You