zenontech

Zen on Tech Newsletter

V17 - AI in Great Power Competition

2nd Oct 2023

Summary

AI in the Context of Great Power Competition
In this volume, we navigate the intricate web of artificial intelligence as it stands at the heart of global great power competition. Unraveling the vast potential and tangible realities of AI, we offer a comparative analysis of the U.S. and China's positions in this technological race. With AI poised to redefine economic, military, and societal landscapes, this assessment aims to shed light on the strategic decisions and repercussions in the larger context of geopolitical dynamics.

Promise of AI

- **Economic Growth**: Generative AI offers the potential for a 1.5% annual boost in U.S. labor productivity over a decade, and a 7% rise in global GDP.
- **Military Advancements**: AI's role in defense encompasses Lethal Autonomous Weapons Systems, Drone Swarms, logistics, electronic warfare, and even battlefield medical care. Moreover, AI training modules are forging a new era of military personnel.
- **Accelerated Scientific Breakthroughs**: AI's ability to process immense data and make predictions is set to revolutionize scientific research in material/molecule discovery, design and optimisation .

AI's Real-World Challenges

- **Diffusion and Application**: AI's true challenge lies in real-world applications, requiring both data interpretation and actionable insights in real-time. The first sectors to reap rewards will be ones that rely on information processing, without the need for complex real-world interactions
- **Civilian-Military Tech Exchange**: The private sector is the new epicenter of AI innovation, with the U.S. DoD adopting, rather than creating, these advancements. However, with DARPA's history of high risk research, a two-way transfer between the DoD and tech giants could spur unprecedented innovations in both military and civilian realms.
- **AI Chip Landscape**: While Nvidia is the current AI chip market leader, evolving software and system-level dynamics may provide opportunities for other contenders like Google.

The U.S. vs. China in AI

- **Current AI Landscape**: The U.S. leads in data and algorithms, but China's surge in hardware and talent indicates a growing rivalry. Despite China's academic contributions, its commercial execution remains behind the U.S.
- **Impact of Sanctions**: Following U.S. sanctions, China will have to defensively innovate to shield its AI sector. Major Chinese cloud players may be resilient to GPU scarcity though huge inventories, whereas smaller entities will struggle. Meanwhile in the U.S, easy hardware access for startups spurs innovation.
- **Algorithm and Compute Power**: Efficient algorithms might drive AI progress, but long-term leadership will require prowess in both software and hardware. China's AI chips mirror Nvidia's from five years ago, the challenge is not just in creating superior chips but building a holistic ecosystem.

The U.S.'s agile, entrepreneurial ecosystem contrasts China's state-backed might. While the U.S. seems primed for breakthroughs, China focuses on achieving self-sufficiency and replicating advancements in hardware. The tug-of-war between U.S. innovation and China's resource-heavy approach will dictate the AI supremacy battle's outcome.

Contents

1. Do It For The TAM

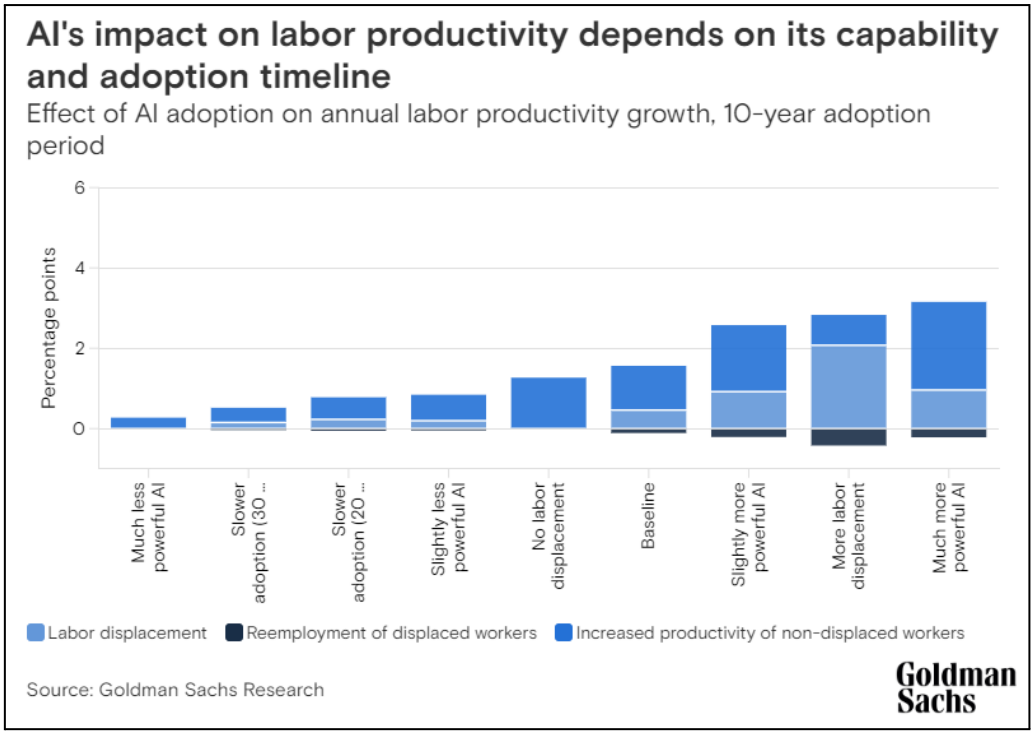
How to Prevent an AI Catastrophe | Foreign Affairs
Economic potential of generative AI | McKinsey
Why Generative AI Could Have a Huge Impact on Economic Growth and Productivity | AEI

The transformative potential of Generative Artificial Intelligence (AI) is set to usher in an economic paradigm shift. As sectors ranging from banking to retail anticipate profound shifts in productivity and revenue, it's imperative to grasp the total addressable market (TAM) of AI. In this first section we optimistically explore the opportunities for AI's financial prospects, anchored in its automation capabilities, labor productivity growth, and its imminent role in reshaping industries globally.

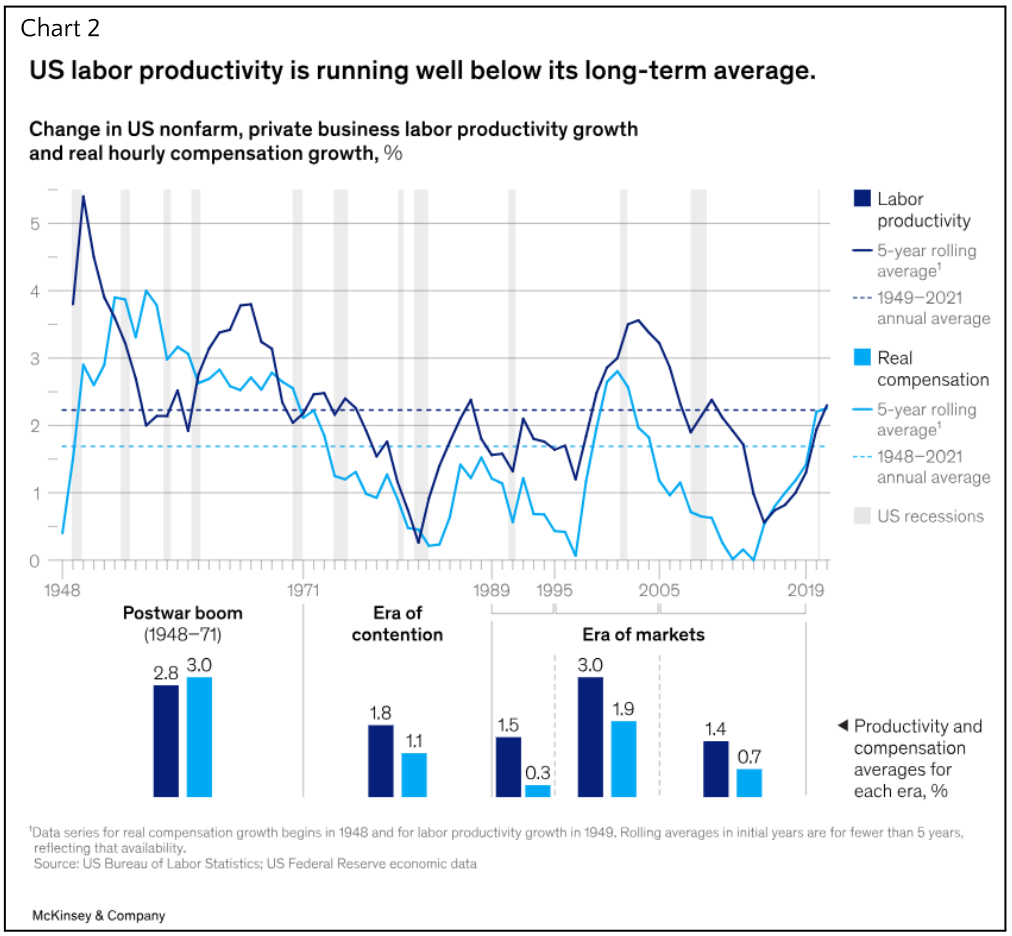
Numerous research firms are eagerly dissecting the transformative effects of Artificial Intelligence (AI). Amidst the cacophony, voices like Goldman Sachs have distinguished themselves. The numbers for generative AI are nothing short of astonishing (**Chart 1**):

- Generative AI could raise annual **US labor productivity growth by just under 1.5% per year over a 10-year period** following widespread business adoption.
- Generative AI could eventually **increase annual global GDP by 7%**, equal to an almost $7 trillion increase in annual global GDP over a 10-year period.
- Generative AI will be disruptive to jobs: "We find that roughly **two-thirds of current jobs are exposed to some degree of AI automation, and that generative AI could substitute up to one-fourth of current work.**"
- AI investment could approach 1% of US GDP by 2030 if it increases at the pace of software investment in the 1990s. (That said, US and global private investment in AI totaled $53 billion and $94 billion in 2021, a fivefold increase in real terms from five years prior.)

Chart 1



Generative AI Could Raise Global GDP by 7%



Rekindling US productivity growth for a new era | McKinsey

McKinsey produced a report on generative AI's potential impact, summarised in bullet points:

- Generative AI's potential economic contribution ranges from $2.6 trillion to $4.4 trillion annually.
- Four main sectors to benefit: customer ops, marketing and sales, software engineering, and R&D.
- Banking could see gains of $200 billion to $340 billion yearly; retail may benefit by $400 billion to $660 billion annually.
- Generative AI can automate tasks that occupy up to 60-70% of workers' current time.
- Half of today's work could be automated between 2030 and 2060 due to generative AI.
- Labor productivity might grow by 0.1% to 0.6% yearly until 2040.
- **Combined with all tech, productivity could rise by 0.2% to 3.3% annually**.

Different countries, different pace of adoption

Adoption is also likely to be faster in developed countries, where wages are higher and thus the economic feasibility of adopting automation occurs earlier. Even if the potential for technology to automate a particular work activity is high, the costs required to do so have to be compared with the cost of human wages. In countries such as China, India, and Mexico, where wage rates are lower, automation adoption is modelled to arrive more slowly than in higher-wage countries.

Both McKinsey and Goldman Sachs present a compelling picture of its transformative potential, especially the **productivity boost to developed economies (Chart 2)**. While predicting a precise 10-year trajectory remains elusive, their projections suggest an impact to the tune of trillions on the global economy, serving as a valuable base case. These figures could swing higher or lower depending on a myriad of factors. Beyond numbers, the geopolitical implications of AI's ascendancy are profound and multifaceted. We explore the ripple effects of these dynamics further in this newsletter.

A New Era of Warfare

The way wars are fought has been continually evolving throughout history. From cold steel and gunpowder to tanks and jet fighters, each technological advancement has brought its own set of strategic implications. However, nothing seems to be reshaping modern warfare quite like the integration of Artificial Intelligence (AI). Based on recent analyses, AI's involvement in war goes beyond strategy and tactics—it influences the very psychology of conflict.

AI in Action: The Many Facets

From Intelligence, Surveillance, and Reconnaissance (ISR) to Cyber Warfare, AI's influence is profound. The recent conflict in Ukraine, for instance, highlighted how drones, backed by AI-driven ISR, can provide real-time actionable insights. Lethal Autonomous Weapons Systems (LAWS) and Drone Swarms are small examples of how AI can change the dynamics of physical combat.

Furthermore, the role of AI in cyber warfare can't be understated. Both defensively and offensively, AI algorithms can counter cyber threats or exploit vulnerabilities at unprecedented speeds. In logistics, war-gaming, electronic warfare, and even medical care on the battlefield, This isn't limited to battles alone. Training modules augmented by AI are creating a new generation of soldiers, further emphasizing the symbiotic relationship between man and machine in future combat scenarios.

Redefining the Metrics of Military Power

Will Roper, Former Assistant Secretary of the Air Force for Acquisition, Technology and Logistics, commented on the Pentagon's hardware-centric mindset. Traditional metrics, like the number of ships or planes, still dominate budgetary discussions. However, as Roper pointed out, the real measure in today's world lies in the digital prowess of these assets. From sensors and algorithms to intelligent munitions, the "digital capabilities" are what determines a force's real strength. The call is clear: the Pentagon needs to pivot from traditional assets to "decision superiority."

> "Ships, airplanes, tanks and ground troops still matter, but what matters more is their **digital capabilities: sensors to detect enemy forces, algorithms to process data, networks to transmit information, command-and-control to make decisions, and intelligent munitions to strike targets.** Roper said the DoD should be focusing on "decision superiority... but its not the way the budget is built" **Four Battlegrounds, Paul Scharre**

The AI Influence: Beyond Cognition

While humans, with their emotions, can be unpredictable and, at times, inefficient, AI systems function without fear, exhaustion, or desperation. The emotionless nature of AI means they would fight with equal tenacity, whether they're on the offense or defense. This could revolutionize siege warfare or extended battles, where human emotions previously played a considerable role. Kenneth Payne, an International relations Politics researcher at King's College London, observed that AI will fight with "mindless dedication" on both fronts suggests a paradigm shift in warfare psychology.

> "AI systems will not just change the cognitive dimensions of war - the processes of synthesizing information, making decisions and executing tasks among distributed units. AI systems will also change the psychology of war. War is a human endeavour, and it brings with it flaws and limitations not just of human cognitions but of human emotion. The quality of human fighting units depends on intangible factors such as morale, esprit de corps and unit cohesion. Human soldiers suffer fear, anger, exhaustion fatigue, and desperation. AI systems will experience none of these emotional burgens War with AI systems will remain violent and chaotic, but AI systems will likely fight differently as a result." **Four Battlegrounds, Paul Scharre**

Past Wars, Future Implications

Reflecting upon the Gulf/Iraq Wars provides a lens to envisage AI's potential role. The technological dominance displayed by the coalition forces, chiefly the U.S., in these conflicts demonstrated the shift from traditional to technologically advanced warfare. Drawing parallels, just as precision-guided munitions and advanced communication systems were game-changers then, AI promises a similar revolution now. The essence lies in the cognitive advantage and unwavering psychological stance that AI brings, which might redefine military metrics in the future.

The rapid ascent of China as a technological powerhouse underscores a new era where the U.S. and its allies face a near-peer adversary that is not only catching up but, in some domains, outpacing them. Much like Japan's meteoric rise in the early 20th century, China's trajectory in the realms of AI, quantum computing, and advanced missile systems has redrawn the strategic landscape. **The challenges are twofold: not only does the West have to successfully integrate and adapt to AI warfare, but it must also anticipate and counter similar or even superior capabilities from China**. It's a dynamic reminiscent of the intense naval competition between the U.S. and Japan during WWII, but with an added layer of complexity. **While AI has the potential to revolutionize military strategy and tactics, its ubiquity and accessibility mean that dominance is no longer guaranteed by mere adoption.** The key will lie in mastery, constant innovation, and the agility to deploy AI effectively in a rapidly evolving battle environment. The stakes are higher, and the margin for error, thinner than ever before.

Warfare, like all other human endeavors, evolves. Today, as we stand at the cusp of a digital revolution in warfare, driven by AI, understanding its multifaceted role becomes crucial. For geopolitical analysts, investors, and policy analysts, this isn't just a technological shift—it's a comprehensive transformation of strategy, tactics, psychology, and even the very essence of military power.

A Catalyst for Innovation

AI's potential as a catalyst for innovations across various scientific domains stems from its ability to process vast amounts of data, recognize patterns, and predict outcomes more efficiently than traditional methods. AI acts as a force multiplier for scientific endeavours, enhancing our understanding, accelerating discoveries, and offering innovative solutions across various domains. As computational power continues to grow and algorithms become more sophisticated, the synergy between AI and scientific research is expected to yield transformative breakthroughs. Here's a breakdown of how AI could drive advancements in areas like material science, batteries, and other scientific domains:

- **Material Science**:
  - **Discovery and Design**: Using AI to analyse patterns in vast datasets can lead to the discovery of new materials with specific desired properties. This could revolutionise fields such as electronics, aerospace, and medical devices.
  - **Simulations**: AI can enhance molecular simulations, helping scientists understand material behaviours under various conditions.
  - **Optimization**: AI can determine the best conditions for manufacturing and processing materials, minimising defects and enhancing properties.
- **Batteries**
  - **Design Optimization**: AI can predict how different materials will perform as electrodes, leading to better energy storage and faster charging batteries.
  - **Lifetime Prediction**: AI models can forecast battery lifespan based on usage patterns and other conditions, aiding in the design of more durable batteries.
  - **Safety Improvements**: AI can detect early signs of battery failures, such as overheating or swelling, leading to safer battery designs and usage protocols.
- **Pharmaceuticals and Drug Discovery**
  - **Drug Repurposing**: AI can identify existing drugs that could be effective for new conditions, significantly reducing the time and cost of drug development.
  - **Molecular Design**: AI can suggest molecular structures that could have desired therapeutic effects, streamlining drug discovery.
- **Agriculture**
  - **Crop Optimization**: AI can predict which genetic traits in plants will lead to desired outcomes, such as drought resistance or higher yield.
  - **Pest and Disease Prediction**: AI can identify patterns that lead to pest infestations or disease outbreaks, enabling preventative measures.
- **Chemistry**
  - **Reaction Predictions**: AI can anticipate how molecules will react under specific conditions, leading to more efficient chemical processes.
  - **Catalyst Discovery**: AI can suggest new catalysts that will speed up reactions, reducing energy consumption and waste.
- **Climate Modelling**
  - **Predictive Analysis**: AI can process vast datasets from global sensors to create more accurate climate models, leading to better predictions of climate change and its effects.
  - **Cross-Domain Collaborations**: An overarching benefit of AI is its ability to integrate insights across multiple domains. For instance, a breakthrough in material science might lead to better battery technology, which could then impact renewable energy storage solutions.

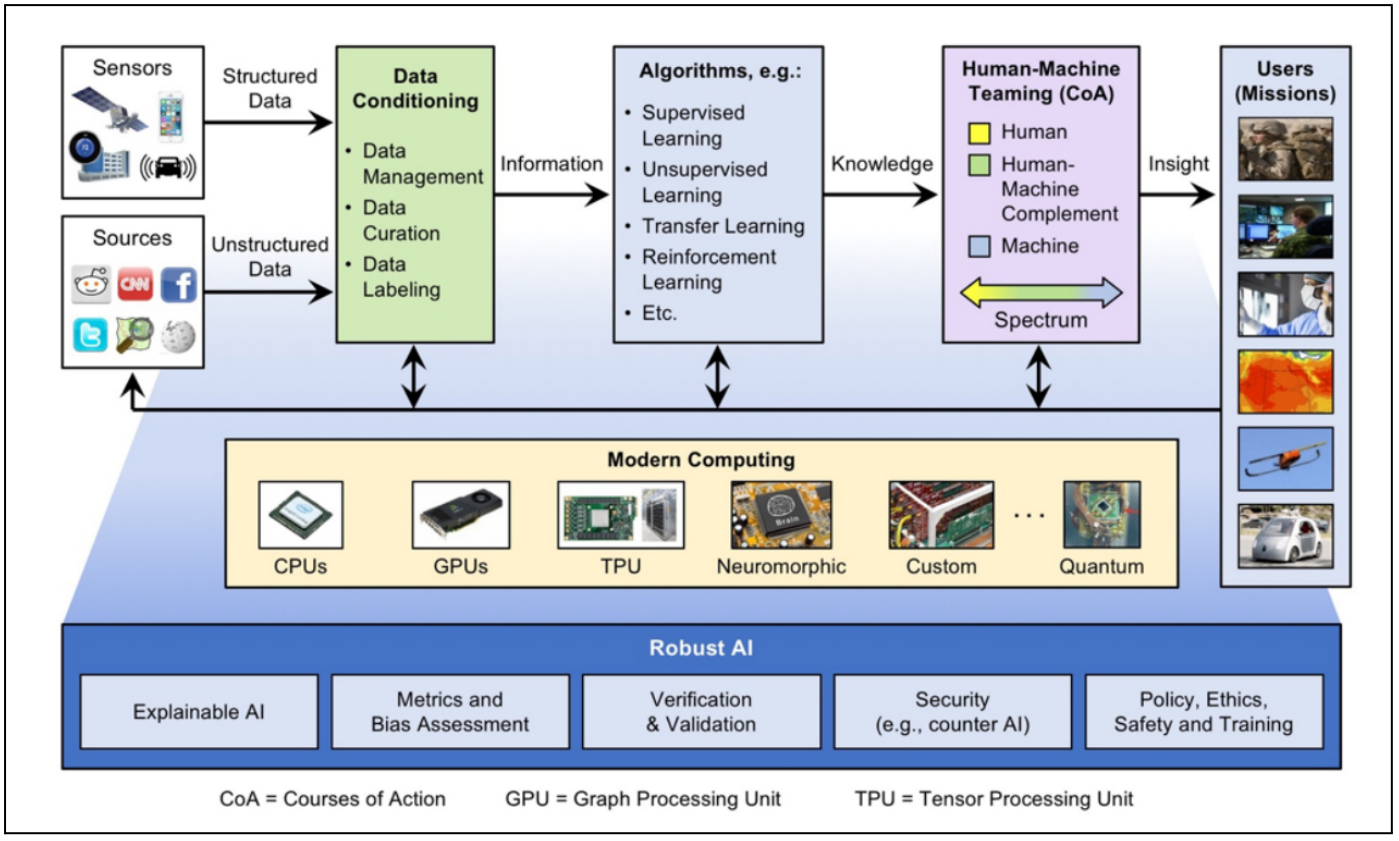
2. Managing Expectations

Generative AI exists because of the transformer

While the preceding section extolled the vast potential of AI, it's crucial to temper our enthusiasm with a dose of reality. It's easy to be enamored by the dazzling prospects of artificial intelligence, but like any nascent technology, it comes with its own set of challenges and limitations. In this segment, we'll introduce a counter-narrative, shedding light on instances where AI might not match up to its often lofty expectations and widespread apprehensions. It's essential to recognize that while AI is transformative, it's not a panacea; there are situations where its application might be premature or unsuitable, and numerous hurdles that might take considerable time and effort to surmount.

One of its most transformative components: the transformer architecture. Introduced in the paper "Attention is All You Need" in 2017, transformers have reshaped the landscape of deep learning, particularly in the realm of Natural Language Processing (NLP). At its core, the transformer leverages an attention mechanism that enables it to focus on different parts of the input data, akin to how humans pay selective attention to specific details when comprehending information. This allows for more sophisticated handling and understanding of contextual relationships in data.

Contrary to some portrayals, the transformer is not "sci-fi" but a highly intricate and efficient statistical model. While its mathematical underpinnings might appear abstract, the principle is straightforward: it assesses and weighs the importance of different pieces of information and determines how they interrelate. **However, it's essential to note that while transformers exhibit an impressive capacity for pattern recognition and prediction, they don't "understand" content in the way humans do.** Instead, they excel at identifying and exploiting patterns in vast datasets, making them powerful tools in the AI toolkit but not without their own limitations and nuances.



https://arxiv.org/pdf/2210.04055.pdf

Preaching to the Converted

AI will change American elections, but not in the obvious way

In an era of heightened political discourse, the potential for AI-driven disinformation has stirred concern among experts and laypeople alike. The constant exposure to conflicting narratives on various platforms can leave the average American fatigued and, perhaps surprisingly, resistant to persuasion.

The Immune Response to Misinformation

Academic studies suggest that the American public, accustomed to an incessant influx of political claims and counterclaims, has evolved a sort of immunity against persuasion. While the rise of AI undoubtedly amplifies the dissemination of untrustworthy information, leading to heightened mistrust, cynicism, and entrenched views, it doesn't necessarily change minds in any substantial way.

As Democratic Congressman Jake Auchincloss aptly remarked, the primary goal of disinformation campaigns, particularly those spearheaded by foreign adversaries, seems less about misleading citizens into distrust of specific individuals or institutions. Instead, it's about sowing the seeds of universal mistrust: **"*Our adversaries abroad, and the worst actors here at home, are at the cutting edge of using disinformation—less to make citizens not trust a particular person or institution, but to make them not trust anything*."**

Research Breaks the Misconception

Yet, it appears that the vast majority of social media users aren't as gullible or susceptible to such tactics as one might fear. A significant study by Matthew Gentzkow and Hunt Allcott of Stanford University, examining the **Russian disinformation campaign, deduced its impact on vote shares to be almost negligible. A mere 0.01% shift, to be precise.** Another study published in Brendan Nyhan, in Nature Communications in 2023 echoed this sentiment, revealing that exposure to **tweets by Russian bots had negligible effects on individual political attitudes or polarization**.

Interestingly, **genuine news on platforms like Facebook doesn't fare much better in terms of influence**. In an experiment conducted during the 2020 presidential election run-up, Meta allowed researchers to modify the type of news some users were exposed to. Even when the news challenged users' pre-existing beliefs or when it was presented in a non-algorithmic, chronological order, there seemed to be **no measurable shift in political perspectives**.

Preaching to the Converted

One notable finding is that **fake news predominantly reaches those who are already deeply entrenched in their political beliefs. As Ms. Chang suggests, the issue lies more with an insatiable demand from hyper-partisans than with the supply of fake news itself. Floating voters, on the other hand, seem relatively untouched.**

But if persuasion is the ultimate goal, even the best in the business face an uphill battle. The multi-billion dollar expenditures on political advertising during election campaigns are a testament to the challenges of swaying voter opinions. Dubbed the "minimal-effects hypothesis" by political scientists, the negligible impact of political advertising underscores the complexities of persuasion. In the words of Brendan Nyhan, "Persuasion is very difficult." His work seeks to debunk myths around the influence of misinformation. The evolving skepticism isn't limited to just textual content; people are increasingly questioning the authenticity of other media formats, including recordings and film.

In conclusion, while the digital age, compounded by AI-driven content, presents a plethora of challenges, the public's inherent skepticism and discernment offer a silver lining. The battle may not be against fake news per se, but in fostering critical thinking and discernment in a media-saturated world.

The I/O Challenge

Broadly defined, AI is a field of computer science dedicated to creating systems capable of performing tasks that normally require human intelligence. AI's potential as a catalyst for innovations across various scientific domains stems from its ability to **process vast amounts of data, recognize patterns, and predict outcomes more efficiently than traditional methods**. However, despite the impressive strides made in the realm of AI, it is essential to recognize its limitations, especially when venturing from digital domains into the physical realities of our world. One significant hurdle, which we'll term the "I/O problem," stands out: the **challenge of seamlessly integrating AI's input/output mechanisms with real-time, physical environments**.

AI's Core: Data Processing and Inferences
At its core, AI operates on data. Machine learning models, a subset of AI, are trained on massive datasets, "learning" from them to make predictions or decisions without being specifically programmed to perform the task. For example, a machine learning model trained on images of cats can recognize and categorize new cat images it has never seen before. Similarly, natural language processing models can generate human-like text by training on vast amounts of textual data.

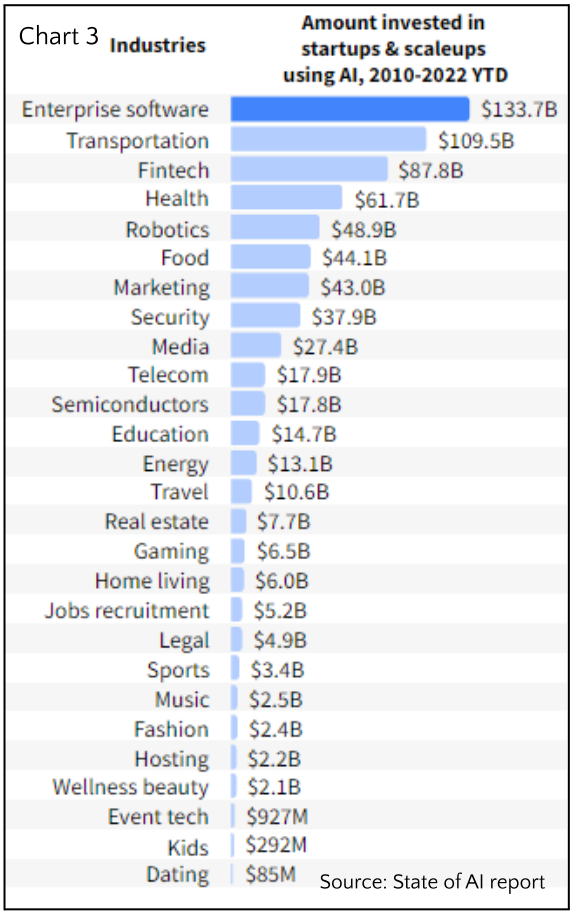
Venturing Into Physical Reality: The I/O Challenge
The real complexity emerges when AI attempts to interact with the physical world, such as driving a car, controlling a drone, or performing a surgical procedure. This interaction demands not just data processing but also the accurate collection of real-time data and the ability to act on that data through various actuators.

- **Input Challenge**: The quality and timeliness of data become paramount. Sensors must collect data in real-time, and any delay or inaccuracy can have serious implications. For instance, an autonomous vehicle's sensors must detect obstacles instantaneously and with pinpoint accuracy to ensure safety.
- **Output Challenge**: Once the data is processed, AI must then relay commands to actuators, devices that convert the AI's digital output into physical actions. This step too can face lag, especially if the actuator is complex or the command intricate. In robotics, for instance, executing a command to grasp an object involves intricate motor movements, and any delay or misjudgment can result in failure.

The I/O problem has broader implications:
- **Safety**: In cases like autonomous driving or medical surgeries, inaccurate data or actuator lag can be life-threatening.
- **Costs**: High-quality sensors and actuators are expensive. Integrating them can escalate the costs of AI-driven systems, making them less accessible.
- **Complexity**: A holistic AI system, integrated with sensors and actuators, requires multidisciplinary expertise, from software engineering to robotics, increasing the complexity of projects.
- **Maintenance**: Physical components, like sensors and actuators, are prone to wear and tear and require regular maintenance, adding to the long-term costs and challenges of AI systems.

Low-Hanging Fruit

The rise of Artificial Intelligence promises transformative economic benefits, but like any paradigm shift, it brings along its unique set of challenges. Among these is the previously mentioned I/O problem, which has profound implications for the trajectory of industries, investments, and employment in the coming years.



The Low-Hanging Fruits: Information-Based Systems

In the world of AI, the first to reap rewards are often the systems that rely primarily on **information processing, without the need for complex real-world interactions (Chart 3)**. Examples include chatbots, recommendation systems, or financial forecasting tools. Here, the I/O is streamlined: data is readily available in digital form, and the output is typically a digital action, like sending a message or recommending a song. The economic advantages are clear:

- **Rapid Deployment**: With minimal hardware requirements, these AI systems can be swiftly integrated into existing digital infrastructures.
- **Cost-Effective**: Absence of sensors or actuators makes these applications more financially accessible.
- **Scalability**: Information-based systems can be scaled rapidly, providing immediate value to a broad user base.

Navigating the Complex Physical Landscape: Self-Driving Cars: Consider an AI system that boasts a 96% accuracy rate. In the world of email filtering or music recommendations, a 4% error margin is acceptable. However, apply that to self-driving cars, and the narrative changes. A 96% accuracy rate implies that 4 out of every 100 decisions could be wrong – a figure too risky for public roadways. The economic repercussions:

- **Higher R&D Costs**: Achieving the near-perfect accuracy required for such applications demands significant investment in research and development.
- **Regulatory Hurdles**: Safety concerns lead to stricter regulations, which can delay product launches and increase compliance costs.
- **Insurance and Liability**: The complexity of integrating AI into the real world raises questions about liability, leading to potential changes in insurance structures.

The Differential Impact on Industries

The I/O challenge means that AI's economic boon won't be uniform across industries:

- **Information Technology & Services**: Stand to gain the most in the short term, given the ease of integrating information-based AI systems.
- **Manufacturing & Robotics**: While there's vast potential, the tangible integration of AI demands substantial investments in sensors, actuators, and safety mechanisms.
- **Healthcare**: Applications like robotic surgeries require absolute precision, leading to a slower adoption rate due to the high stakes involved.

For investors, understanding the I/O challenge offers a lens to evaluate the risk-reward ratio. Information-based AI startups: Might offer quicker returns given their ease of integration and scalability. Real-world AI applications: Represent a longer-term bet, with potentially higher returns, but accompanied by increased risks and longer maturation periods.

Transferring AI Into Military Power

Historically, the U.S. Department of Defense (DoD) was at the forefront of technological innovation, with groundbreaking advancements like GPS, microprocessors, and the early internet. This dominance in the 1960s and '70s solidified the military's role as a pacesetter in the realm of Research & Development (R&D). However, recent shifts in R&D expenditure trends suggest a significant transition. **The lion's share of innovation is now emanating from the private sector, rendering the DoD more of a consumer than an instigator of cutting-edge technologies.**

The implications of this transformation are twofold. On the one hand, the **DoD stands to benefit from the burgeoning technological dynamism of the private sector, potentially allowing them to capitalize on innovations without the accompanying R&D expenditure**. Yet, on the other hand, the DoD finds itself in uncharted waters. Without the leverage to steer the direction of technological innovation, the DoD is faced with the Herculean task of adapting and integrating these external advancements rapidly. The pressing concern is that any failure in this **'spin-in' mechanism** might see the U.S. military side-lined from pivotal technological shifts, thus undermining its global supremacy.

AI, with its promise of ushering in an era of unprecedented economic growth, comes tethered to its unique challenges. For instance, while AI startups operating in the information realm present quicker investment returns due to their scalability, real-world AI applications, though potentially more rewarding, are fraught with higher risks and prolonged maturation periods. The trend indicates a potential divergence in leadership roles in AI implementation. **Tech giants are poised to spearhead innovations in enterprise software. Simultaneously, the DoD seems primed to champion many real-world implementations of these AI systems, potentially revolutionizing warfare**. The vision of retrofitting Tesla's autonomous driving tech into military jets, submarines, and drones paints a picture of a formidable, digitally-enhanced military force. Yet, these are not mere speculations; the DoD's foray toward spinning-in the AI ecosystem, is a testament to this trajectory.

However, this vision, as grand as it sounds, has its roadblocks. The DoD's notorious slow pace of tech implementation, rooted in its industrial-era metrics and mindset, poses a significant challenge. In an era where the yardstick of military might is shifting from sheer numbers to digital prowess, **the DoD's budgetary and operational focus requires a recalibration. For the DoD to truly harness the potential of AI and other innovations, a cultural and strategic overhaul is imperative.**

The symbiotic potential between the Department of Defense (DoD) and corporate giants, particularly after a so-called "Cold War II," presents an exciting future canvas for technological advancements. Historically, as with the development of the internet and GPS, military innovations have often been groundbreaking and well-resourced. If the Department of Defense were to pass down these advancements to the commercial realm, we could see a revolutionary leap in civilian tech, with Tesla's vehicles or Boeing's aircrafts benefiting from defense-grade AI and navigation systems. This becomes all the more compelling when considering the **DoD's appetite for tackling long-term investments and complex challenges like the I/O problem, areas where the private sector often exercises caution due to associated risks**. Through this partnership, the DoD provides the rigorous research and development foundation, while companies like Tesla and Boeing provide scalability and widespread application.

3. AI Hardware

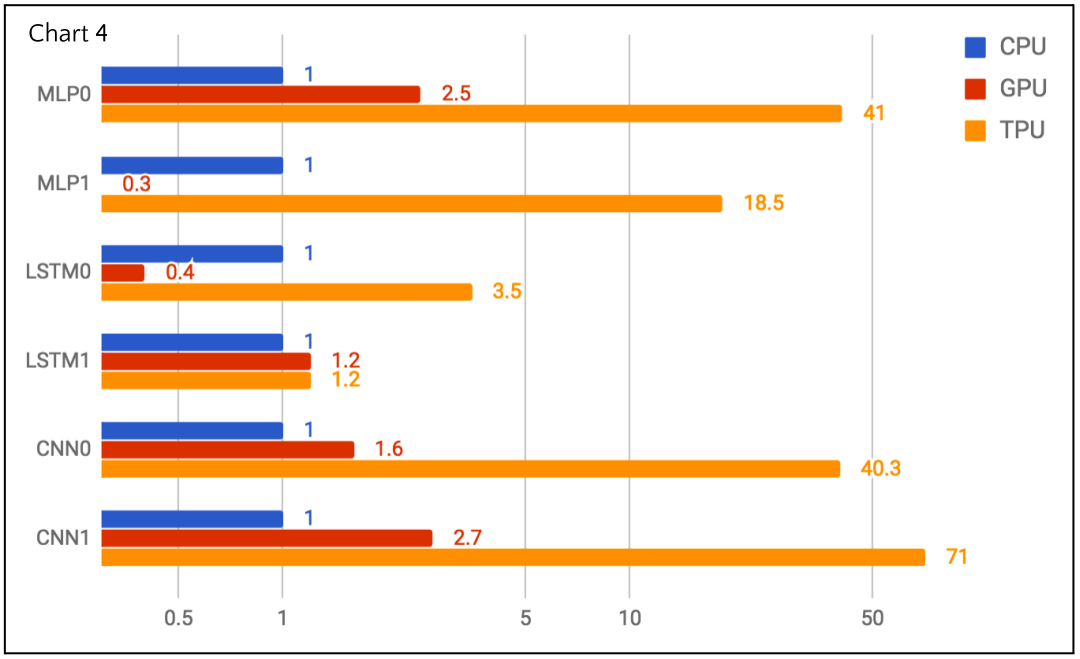
AI and ML Accelerator Survey and Trends
Reimagining Our Infrastructure for the AI Age | Meta
Recent breakthroughs in AI owe a significant debt to advancements in computing hardware, such as Nvidia's GPUs, which have adeptly catered to the needs of computation-intensive machine learning models, particularly Deep Neural Networks (DNNs). The intrinsic design of Nvidia GPU architecture is especially apt for running dense matrix operations, making it powerful for the rapidly advancing transformer model architectures. To better understand the landscape of AI hardware, it's pivotal to discern the different terms and their implications. Here's a brief overview:

AI Chip: This is the fundamental processing unit optimised for AI operations. Here are some types:

- **ASIC (Application-Specific Integrated Circuit)**: Custom-designed for particular tasks, Google's TPU is a prime example. It's tailored to execute specific AI workloads with high efficiency (**Chart 4**).
- **GPU (Graphics Processing Unit)**: Initially designed for rendering graphics, GPUs like those from Nvidia have evolved as versatile chips that can also efficiently handle a variety of tasks, including AI computations.
- **CPU (Central Processing Unit)**: The primary general-purpose processing unit of a computer, CPUs can handle AI tasks but are often outperformed by specialised hardware in efficiency and speed.
- **FPGA (Field-Programmable Gate Array)**: A reconfigurable chip, FPGAs can be customized post-manufacture, offering flexibility and making them useful for specific AI tasks and prototyping.

AI Card: An augmentative hardware component, it typically comprises one or multiple AI chips coupled with supporting hardware. While NVIDIA's Tesla or A100 GPUs were crafted as graphics cards, their robust computational faculties make them adept for AI acceleration.

AI System: The holistic ensemble, which is a computer or server that's fine-tuned for AI tasks. **It harmonises CPUs, memory, storage, and networking, often with specialised software, to streamline AI workloads. NVIDIA's DGX systems are an example; these servers, constructed around an array of NVIDIA GPUs, are dialled in for deep learning.**



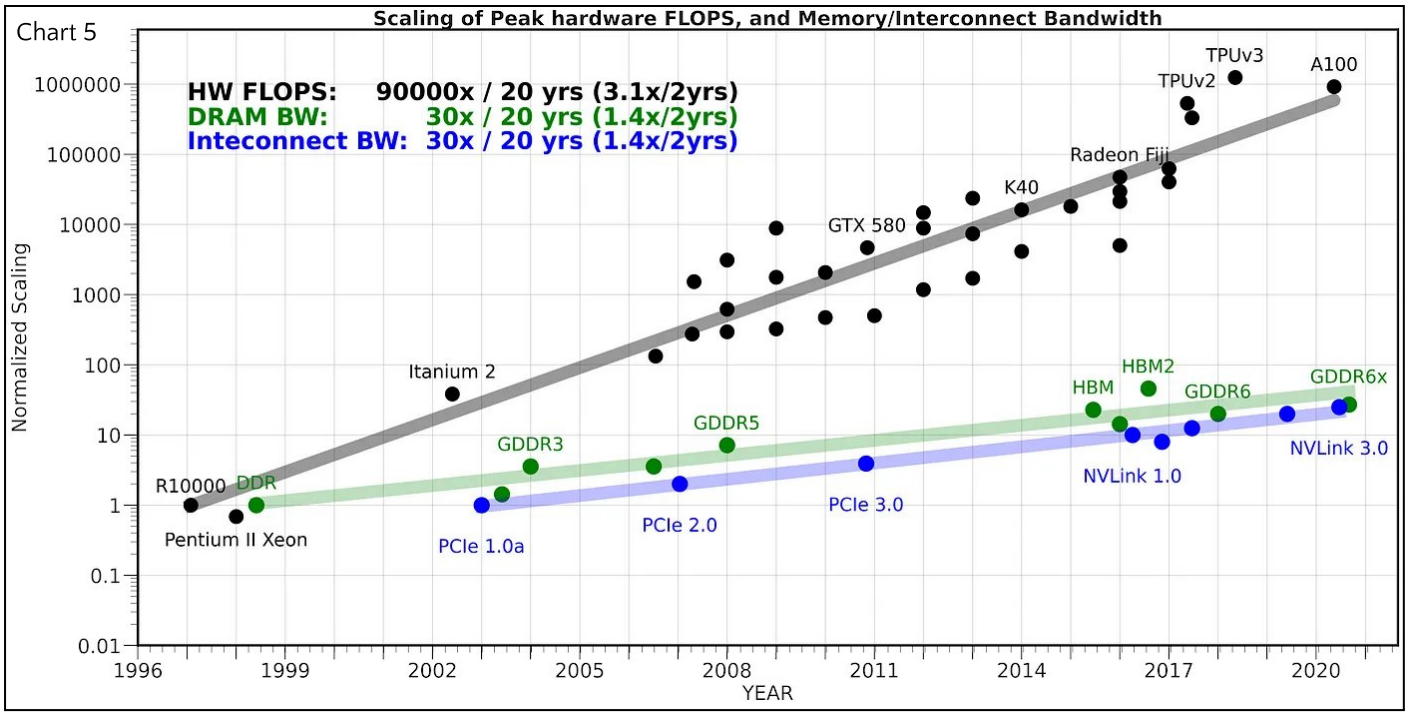
An in-depth look at Google's first Tensor Processing Unit (TPU)

Memory Wall

Machine learning model training has traditionally been dominated by computational demands, particularly matrix multiplications. Yet, with the exponential growth of AI models since 2018, the memory requirements have surged due to expansive model parameters. Current GPUs, while powerful in computation, struggle with the "Memory Wall" challenge (**Chart 5**). Models from companies like Baidu and Meta demand terabytes of memory, leading to more time spent waiting for data access than on computation.

Memory exists in a hierarchy from fast and costly, close to the processor, to the slower and cheaper. Though some chips utilise vast SRAM pools for model weights, capacity remains inadequate for sprawling models. Nvidia, as an example, can't incorporate adequate on-chip memory without facing prohibitive costs, a situation Moore's Law doesn't alleviate much. Even prospective solutions like 3D SRAM offer just short-lived solace.

This "memory wall problem" underscores the widening gap between computational power and memory bandwidth and capacity. Distributing training over multiple accelerators isn't a silver bullet either due to communication bottlenecks. **Addressing this looming challenge calls for innovations in training algorithms, efficient deployment strategies, and a reimagining of AI accelerator designs that harmonise computational and memory capabilities. Basically, system level innovation, not just more grunty chips.** The escalating computational demands and ballooning model sizes mean that the memory challenge needs to be dealt with, or it will become the primary bottleneck in AI model training.



AI and Memory Wall by Amir Gholami

Nvidia's Hype Machina

Nvidia's reign in the AI market is largely attributed to its symbiotic hardware-software ecosystem. Their software has been meticulously optimised for their GPU architecture, making competition difficult. With software frameworks like PyTorch historically tailor-made for Nvidia GPUs, there was a cyclical dependency that bound developers to Nvidia's platform.

However, cracks in this dominance are emerging. **In 2022, 20,350 academic papers cited using Nvidia GPUs, but this number is forecasted to drop to 17,669 in 2023**. While these numbers still dwarf competitors, they hint at a waning monopoly. However, amidst the decrease of Nvidia's usage in academic papers their sales have gone completely gangbusters with revenue of $13.51 billion vs. $6.70 billion y/y.

Nvidia's attempt to retain its monopoly by bundling software may be eroding. With PyTorch 2.0's introduction (an open source tool managed by Meta) and its shift towards a hardware-agnostic model, the AI landscape is evolving. While Nvidia's A100 GPUs still benefit with a performance boost from their software, PyTorch 2.0 promises gains for other hardware solutions too, levelling the playing field. Furthermore, tech giants like **Meta investing in PyTorch 2.0 seek to decentralise AI hardware reliance, thereby subtly challenging Nvidia's dominance**.

In summary, Nvidia, through strategic hardware-software integrations, solidified its AI market lead. But, the evolving software ecosystem, especially with PyTorch 2.0's enhancements, and the looming memory wall challenge could diversify the AI hardware landscape, potentially diluting Nvidia's hegemony.

In the AI environment, the hardware architecture—most notably the chip microarchitecture and overall system design—is a key determinant in the development and scalability of AI-driven applications. It has a profound impact on operational costs and gross margins, underlining the critical nature of optimizing AI infrastructure for practical deployment. Google moved early on this and developed specialized hardware, Tensor Processing Units (TPUs), to facilitate the efficient scaling of its AI initiatives.

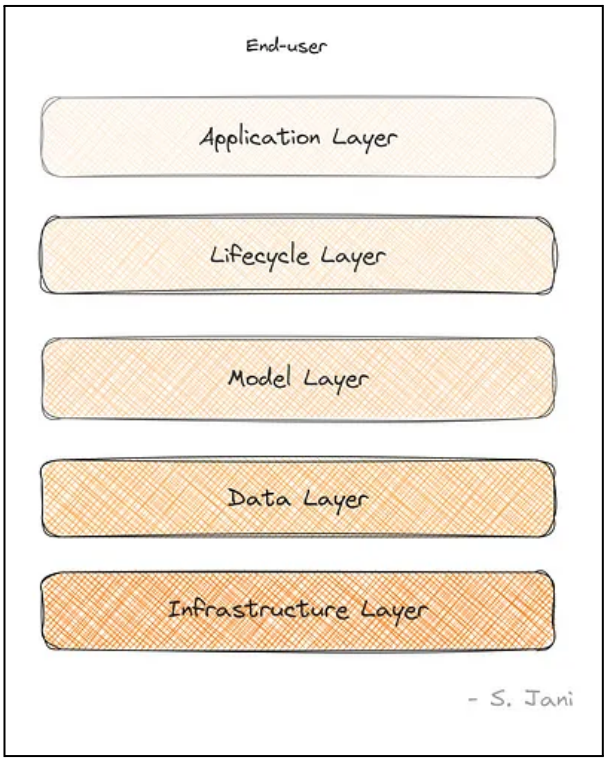
Embarking on its quest for AI-specific infrastructure back in 2006, Google made a significant move in 2016 by launching its first TPUs specifically designed for handling AI tasks. Google's continuous roll-out of TPU iterations—ranging from the original TPU to the latest TPUv5—serves as a testament to their ongoing investment in AI infrastructure.

A unique advantage for Google lies in their all-encompassing strategy that spans from chip microarchitecture to system design and even to "deployment slicing." Google's TPUs are engineered to excel at AI-specific tasks, complemented by their proprietary networking stack, known as ICI, which enhances the scalability and efficiency of their systems. In comparison, Nvidia's solutions, while system-aware, face limitations particularly concerning scalability and networking capabilities. Google has further refined its infrastructure by employing custom optical switches and distinctive networking topologies, thereby minimizing networking costs while enhancing overall performance. **Google has a near-unmatched ability to deploy AI at scale reliably with low cost and high performance.**

Where Will Value Accrue?

[Exploring the Microcosmic Layers of the Generative AI Stack]

Algorithms are universal, so the edge often lies in proprietary data and advanced hardware. Nonetheless, the bridge between these technical facets and real-world utility is human expertise, and the institutions that channel these assets into tangible applications play a decisive role in AI leadership.



Organizations that are more likely to prefer Nvidia's ownership solution over Google's rental option for AI infrastructure are typically large enterprises with substantial AI workloads, research institutions, AI model development companies, and businesses with **stringent security and compliance needs**. These entities often have specific hardware requirements, long-term investment perspectives, and a desire for greater control and customization of their AI infrastructure. Nvidia's ownership model aligns with organizations planning for extended AI infrastructure investments. Additionally, existing Nvidia ecosystem users may opt to continue with Nvidia's hardware to maintain compatibility and leverage previous investments.

Organizations favoring Google's rental model for AI infrastructure often include smaller businesses, startups, and those seeking cost-effective and scalable solutions (most non-AI businesses). **Google's cloud platform offers accessibility, flexibility, and a suite of managed AI services that simplify AI model development and deployment**. This is particularly advantageous for organizations lacking extensive AI expertise, or ones that have no requirement to invest in the expertise. Google's global data center network ensures low-latency access to resources worldwide. Moreover, Google's focus on the model layer of the AI ecosystem, including machine learning APIs, AI platform solutions, and pre-trained models, streamlines the AI development process. Additionally, its integration with other Google services and initiatives to support startups with AI solutions further make Google's rental model appealing to a diverse range of organizations seeking accessible and efficient AI infrastructure solutions. Similar to the way software is bought today "as a service," we expect the same for the LLMs Model-as-a-Service, MaaS. **This 'turnkey' solution optimizes for companies seeking a fast time to value with the assurance of well-performing models. Google will be able to offer access to AI development that is lower cost and less labour and capital intensive.**

Google's recent decision to incorporate [Nvidia's H100 GPUs into their new supercomputer], alongside their established use of TPUs, signals a strategy to diversify their supply chain. GPUs have a mature ecosystem with extensive software tools and a large developer community. While Google continues to use TPUs in their core infrastructure, the embrace of H100 GPUs suggests their openness to other AI accelerators, optimizing for both versatility and performance.

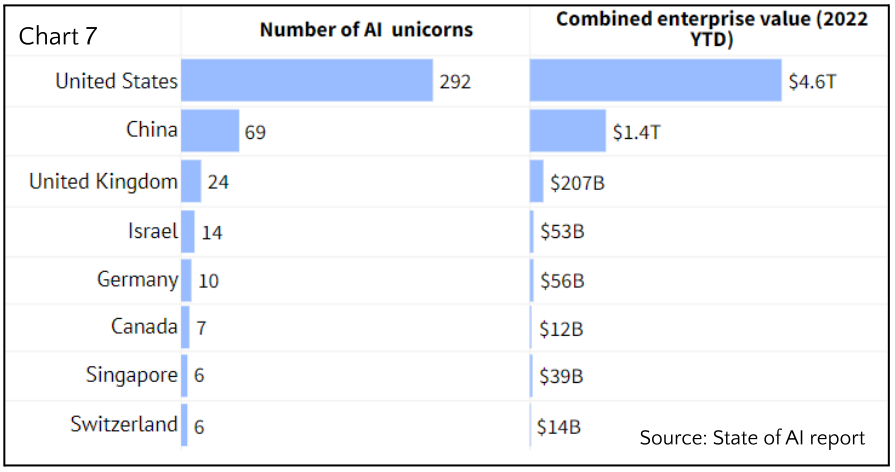
The generative AI stack presents diverse layers, each with its own competitive landscape. While the infrastructure layers have been quick to capture value, the evolution of the higher layers is yet to unfold. As technology progresses, these layers will transform, consolidate, or even introduce new layers. However, the fundamental principles of building durable businesses with attractive unit economics poised for efficient growth remain constant.

4. AI Matchup: Contested Leadership

[Artificial Intelligence and Great Power Competition, With Paul Scharre | Council on Foreign Relations]
[Four Battlegrounds: Power in the Age of Artificial Intelligence | Center for a New American Security (en-US)]

In evaluating global AI leadership, there is a quadrant of vital elements: Data, Algorithms, Hardware, and Talent. Data drives AI, and as more devices gather it, AI systems sharpen their accuracy. Algorithms transform this data into meaningful insights, while high-powered hardware ensures processing. Yet, these systems are only as potent as the human experts behind them, underscoring the competition for top AI talent. Algorithms are universal and easily replicated, so the edge often lies in proprietary data and advanced hardware. Nonetheless, the bridge between these technical facets and real-world utility is human expertise, and the institutions that channel these assets into tangible applications play a decisive role in AI leadership.

The current AI landscape presents a nuanced picture. While the U.S. publishes around **30% more papers than China**, the latter is rapidly narrowing this academic chasm (**Chart 10**). Remarkably, when considering publications in the CNKI (China National Knowledge Infrastructure, a leading Chinese academic database), **China's output is a whopping five times that of the U.S** (**Chart 9**).

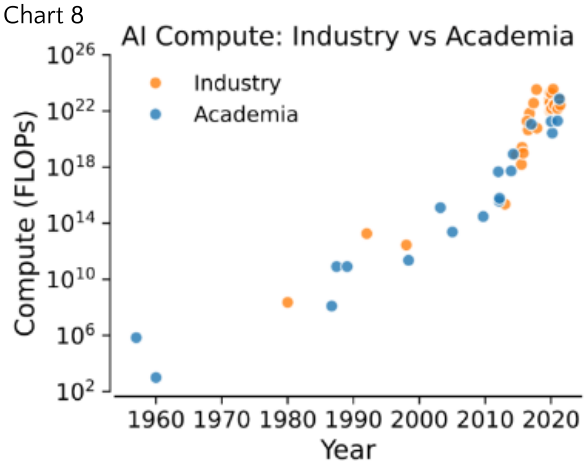


However, there's a significant shift in the AI sector's dynamics. Before 2010, academic institutions spearheaded 80-100% of large-scale AI outcomes. By 2020, this plummeted to below 10%, with startups and private enterprises wielding commercially-driven agendas stepping into prominence (**Chart 8**). Analysing the startup ecosystem, the U.S. has 292 AI "unicorns" (startups valued over $1 billion) with a cumulative worth of $4.6 trillion. In contrast, China boasts 69 such firms, valued at $1.4 trillion (**Chart 7**). **Thus, while China's academic contribution is vast, its commercial translation lags behind the U.S. ecosystem.**

Expert consensus suggests China's Large Language Models (LLMs) trail by roughly **two years**. A stark distinction between the U.S. and China is evident in China's LLM approval mechanism; overseen by the Cyber Security Administration, it mandates model-specific evaluations. This translates to formidable barriers for public-facing, consumer-centric solutions, like chatbots. Conversely, business-focused applications, such as corporate productivity tools, may encounter a more lenient regulatory environment. Further compounding the issue, there's a discernible hesitance among Chinese consumers towards SaaS subscription models. However, this doesn't preclude the potential for military innovations within China.

The breadth and depth of tech available to the U.S. DoD courtesy of its private sector necessitates integration for defense applicability. The People's Liberation Army (PLA) may begin with a weaker commercial foundation. Yet, the malleability of China's state-centric economic approach could funnel resources to projects that are identified as important. The U.S. holds the upper hand in algorithms, and notably, hardware. But, **China's relentless pursuits in hardware, talent development and the transferability of algorithms make the U.S. advantages marginal at this point**. However if dramatic changes were executed on the U.S. hardware control regulations, this gap would become far more pronounced. This begs that question: Can China underpin its AI trajectory by ensuring access to top-tier hardware?

Chart 8



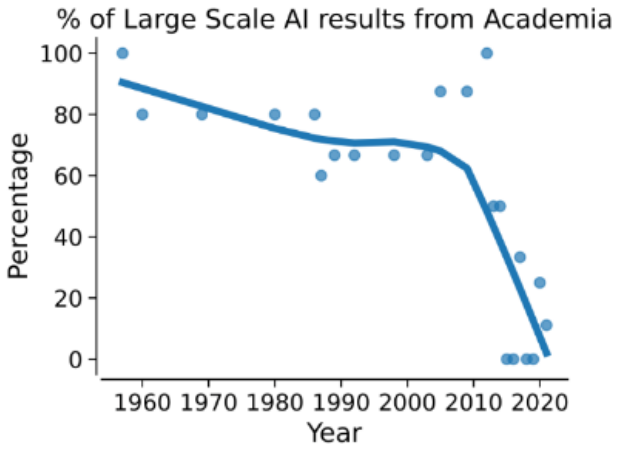


Chart 9

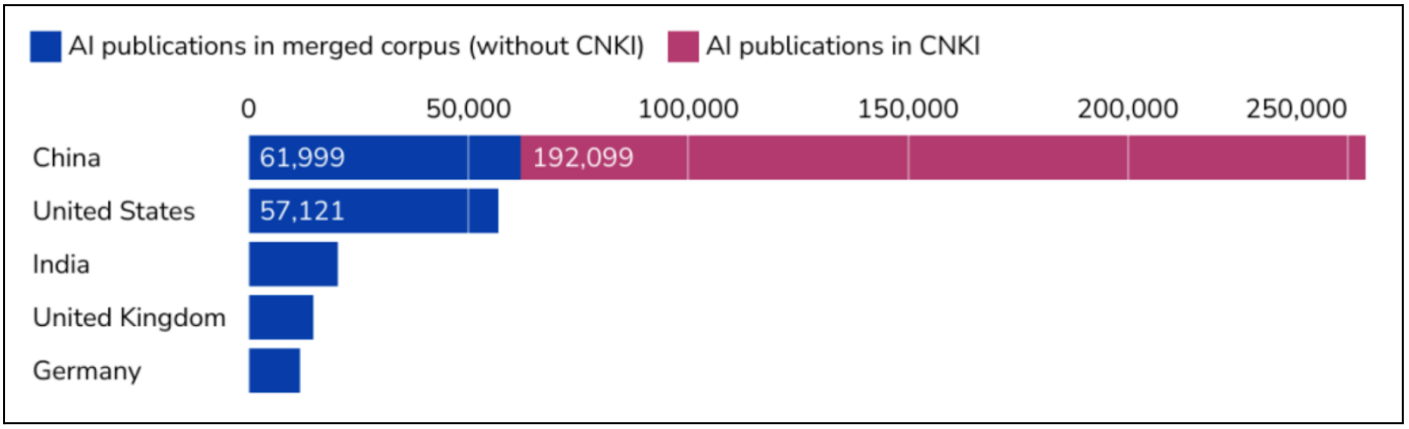
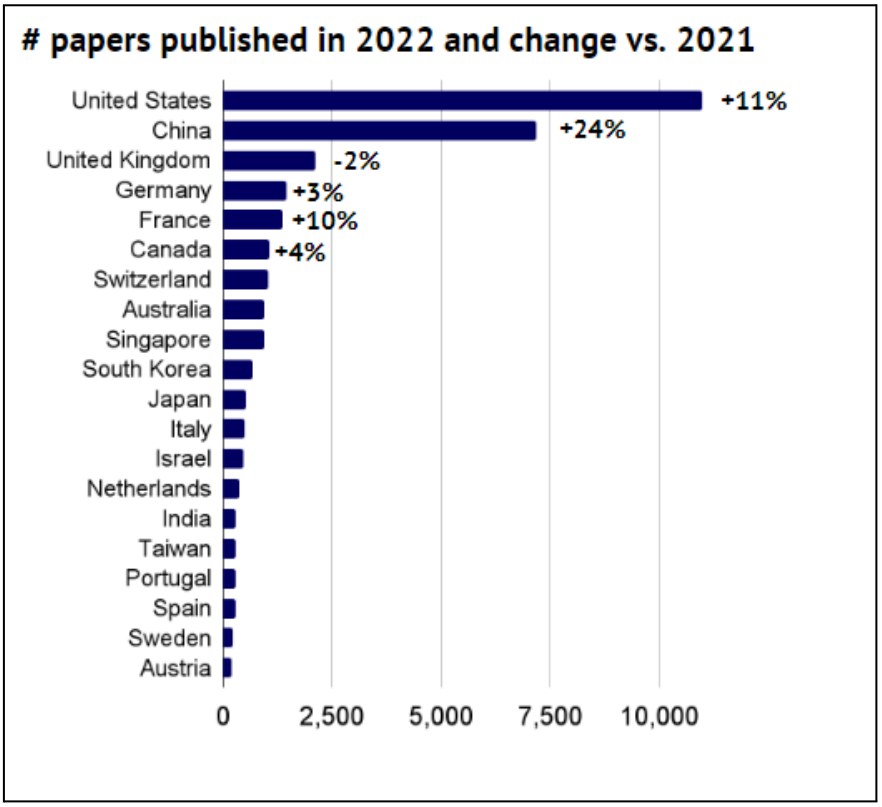


Chart 10



[State of AI Report Compute Index]

5. Can China Protect their AI Industry?

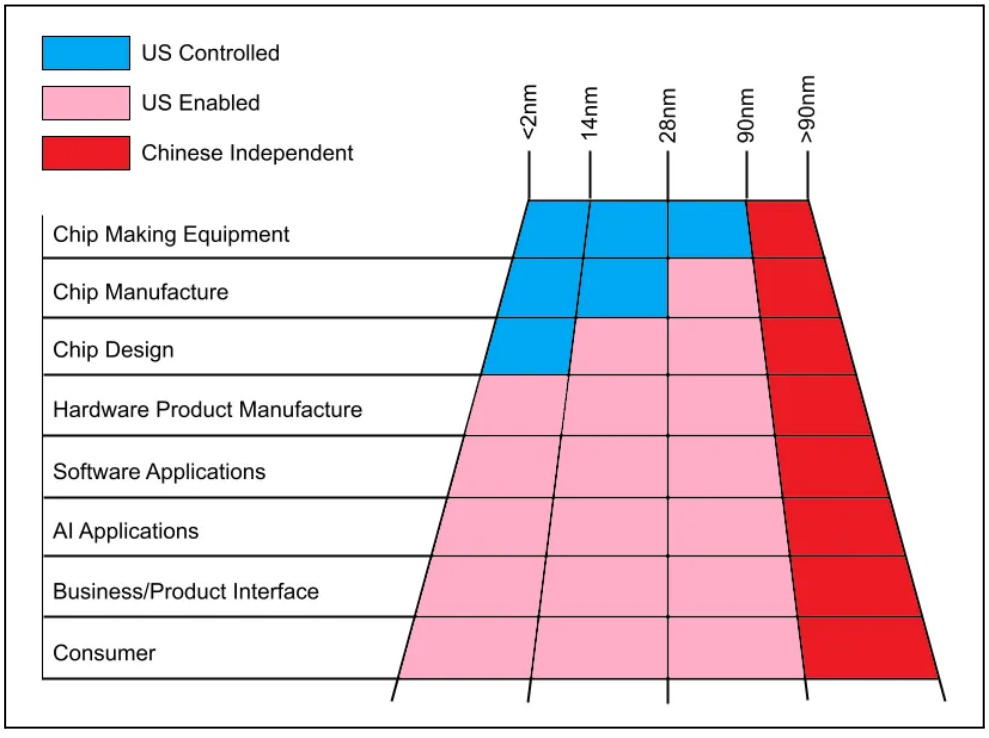
Huawei Got 7nm Chips, Your Move America

Fundamental to global power rivalry hinges on whether major powers can secure crucial resources and technology to reinforce their security infrastructures. The ramifications of U.S. technology constraints on China's defense, economy, and technological progress warrant deeper examination.

Beijing's prioritization of tech investments over broader stimuli underscores its significance in the eyes of Chinese leadership (Zen on Tech, 22-Sept). The procurement of state-of-the-art semiconductors and related chip-making machinery is pivotal for China's AI development, though it may not immediately jeopardize its military or economic capacity. While these constraints may shape future warfare capabilities, the more pressing concern could be their effect on economic productivity growth - a linchpin for China's economic vitality. **Given the transformative potential of AI in bolstering productivity and military power, this could be decisive for China's long-term trajectory.** In the absence of foreign cutting-edge chips, a looming question is whether China's AI evolution can remain competitive, ensuring significant productivity and military enhancements even when reliant on dated chip technology.

Under tech restrictions, **Chinese companies will need to innovate and effectively mix AI chips, run efficient AI programs on massive yet potentially less efficient supercomputers, and develop satisfactory chip-making equipment.** It wouldn't be unprecedented for an industry to thrive outside of the U.S. due to a focus on efficiency. American engineering ethos often leans towards maximising resources rather than optimising them. Consequently, the current restrictions on AI chips and chip manufacturing technology may not pose an existential long term problem for Beijing. Yet, their relative success in resolving these technical issues may make Beijing a more critical threat to US leadership. If China fails in overcoming the technical challenges imposed by U.S. tech restrictions Beijing could become more aggressive and desperate. However, the more successful China is, the higher the likelihood of the U.S. ramping up its technology restrictions, potentially giving rise to a self-reinforcing feedback loop.

"I definitely think it's true that the interventions that have been taken so far on the export controls, are likely to have a seismic impact on China's ability to train the next generation of frontier models. Fundamentally all training depends on these chips, Nvidia GPUs and each generation is way more powerful than the previous." Mustafa Suleyman, founder of DeepMind



Mix and Match the Chips

U.S. sanctions have prompted Chinese tech firms to hasten research for advanced artificial intelligence (AI) capabilities without depending on American semiconductors. An analysis by the Wall Street Journal indicates that companies, including Huawei, Baidu, and Alibaba, are exploring methods to maximize the performance of their existing chips without relying solely on the newest models. Additionally, they are experimenting with combining various chip types to diversify their hardware dependence.

Strategic Resource Allocation in Chinese Tech Giants

After February this year, companies such as ByteDance, which already have significant cloud computing stakes, increased their orders with NVIDIA. ByteDance, for example, placed orders worth over $1 billion on GPUs. Meanwhile, firms that had already stocked up on A100s, like Alibaba and Baidu, are now limiting the use of these advanced foreign chips, reserving them for more intensive computational tasks. Baidu has even paused using its A100s for departments like its self-driving unit to allocate them for the development of its ChatGPT equivalent, Ernie Bot.

NVidia's Crafty Response to U.S. Restrictions

The US banned leading edge NVidia chips from being sold to China in October 2022. However, driven by the allure of short-term profits, NVidia responded to the BIS constraints strategically. The tech giant developed AI chips tailored for the Chinese market, the A800 and the H800, ensuring it remained just beneath the performance benchmarks set by the U.S. Commerce Department. This crafty move allowed the A800/H800 to substitute the A100/H100 in data centers, bypassing the newly imposed restrictions with comparable performance for AI training workloads.

Diverse Chip Combinations as a Solution

Chinese companies are now merging three or four older-generation chips to emulate the capability of Nvidia's top processors. This methodology, however, can be expensive. This scenario has motivated some enterprises to fast-track the creation of techniques for training expansive AI models across an assortment of chips. Giants like Alibaba, Baidu, and Huawei have been documented trying various combinations of A100s, older Nvidia chips, and Huawei Ascends to optimize performance while managing costs. China's internet giants are dominant forces in the cloud realm, with sufficient processing capabilities and a vast reserve of chips/GPUs built up over time. **In China, the scarcity of GPUs might benefit these major cloud players on a relative basis while posing challenges for smaller firms who lack the scale and inventory to access hardware. This is in contrast to the U.S., where startups have easy access to the necessary hardware, fostering innovation and the development of new models.**

Source: NVidia

	V100	A100	A800	H100	H800
Release date	Jun 2017	May 2020	Nov 2022	Sept 2022	Mar 2023
ASP	10,000	20,000	15,000	35,000	25,000
Interconnect restrictions	N/A	N/A	⅔ of A100	N/A	⅔ of H100
AI training speed (benchmark A100)	0.3x	1.0x	0.7x	9.0x	6.0x
AI training cost efficiency (benchmark A100)	0.7x	1.0x	0.9x	5.1x	4.8x

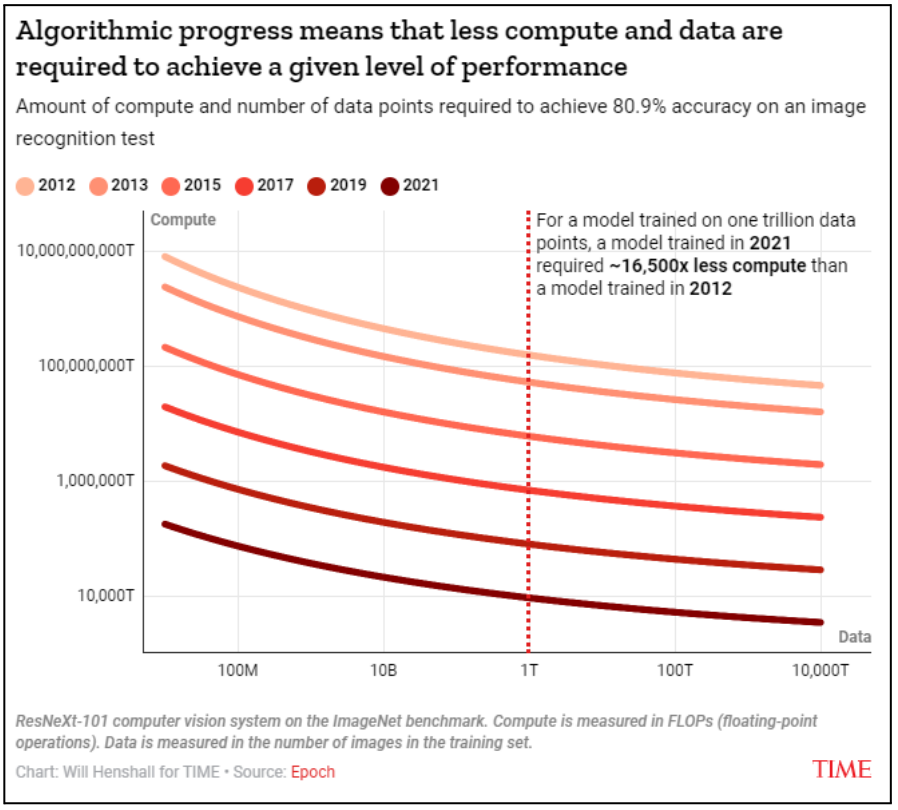
Efficiency or Power: AI Algorithms

Innovation isn't merely about sheer power—it's about using resources effectively. While the U.S. often emphasizes maximizing resources, other nations have thrived by focusing on efficiency. History provides ample evidence of this dynamic.

The Japanese Case Study: Vehicles and Appliances
During the latter half of the 20th century, the Japanese auto industry disrupted the global market with a unique proposition: highly efficient, reliable, and affordable vehicles. Instead of the gas-guzzling behemoths popular in the U.S., brands like Toyota and Honda introduced models that were fuel-efficient and required fewer repairs. Similarly, in the realm of home appliances, Japanese companies like Daikin took the lead in air conditioner innovation. While American manufacturers focused on raw cooling power, Japanese models were designed with energy efficiency and compactness in mind. This strategic emphasis on efficiency propelled Japanese brands to the forefront of the global market, with their products being synonymous with reliability and value.

The Algorithmic Era and AI's Evolution
Algorithms are sets of rules driving operations, allowing AI systems to utilize computational resources to analyze data. While processing more data is one method, there's a trend of refining these algorithms to achieve results with fewer resources. One study showed that "every nine months, the introduction of better algorithms contributes the equivalent of a doubling of computation budgets" (**Chart 11**). When compared with Moore's Law, where computer power doubles every two years, it's evident that **efficient algorithms will contribute more to the rate of change of AI than raw compute power. However, multiplied out over the long term the world leaders will develop both hardware and software.**

Chart 11



China's Path in AI: Algorithms
Understanding this is important when examining the trajectory of Chinese firms in the AI sector. While they may face challenges in developing competitive hardware, the nature of algorithms—which can be replicated—offers a pathway for development.

However, in the realm of hardware, sheer processing capacity isn't the sole determinant of progress. The "memory wall problem" highlights the gap between computational speed and memory accessibility. Distributing tasks across multiple accelerators presents its own set of challenges, including communication limitations. Addressing these challenges might require a combination of refined training algorithms, more strategic deployment methods, and a rethinking of AI accelerator designs to balance computational and memory needs. **For continued growth, China needs to invest in algorithm development and consider new designs for AI accelerators, rooted in their own chip innovations. Software alone won't be enough.**

Design AI Chips Domestically

"[The Nvidia processor ban] will have no impact on NIO's operations in the short term, and in terms of AI computing power, NIO's current resources are sufficient for self-driving algorithm training. There are local companies in China with similar chips, and NIO will evaluate their solutions." NIO's Q2 earnings call

"Huawei is committed to building a solid computing power base in China – and a second option for the world," Meng Wanzhou, Huawei CFO

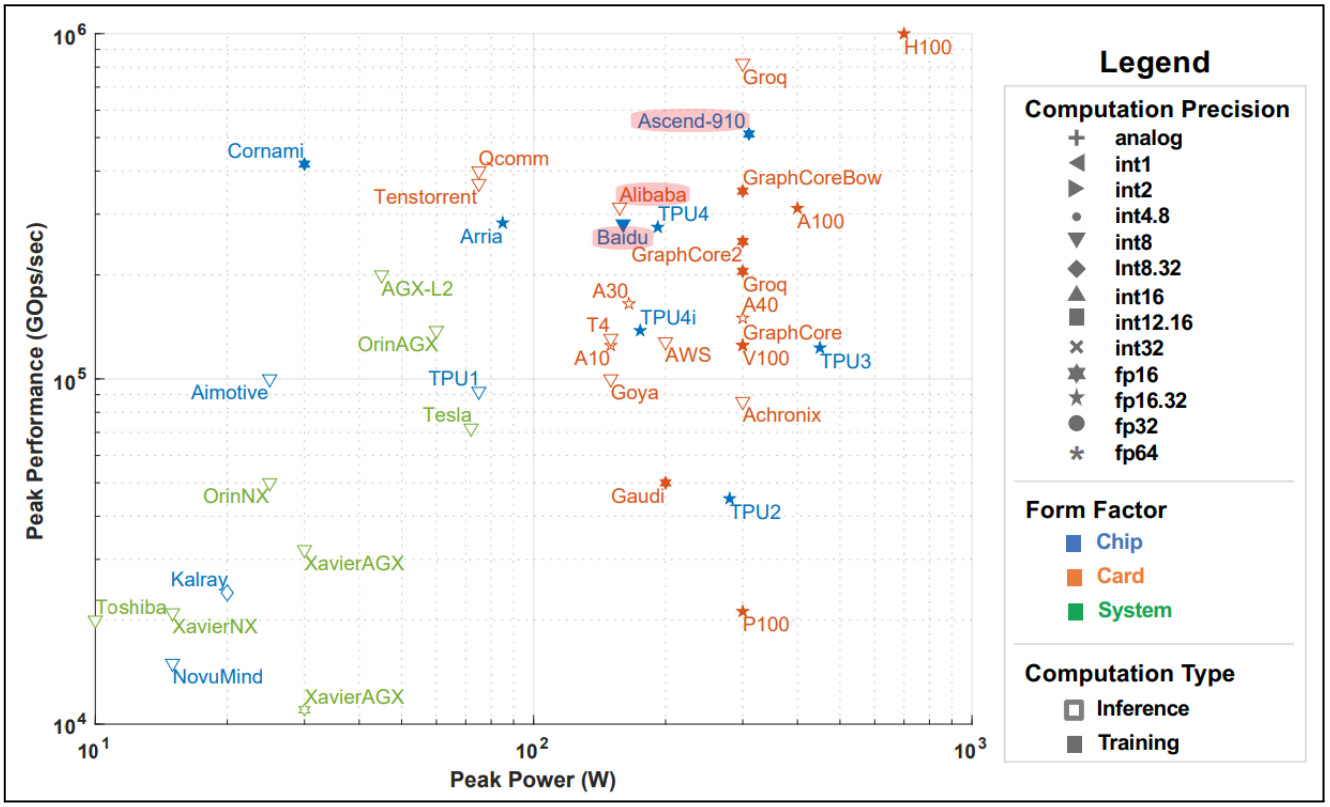
China's Response: Ramping Up Local Production

Despite the international restrictions on sub-14nm AI chips, China's semiconductor industry is showing resilience and innovation. A notable highlight is Baidu AI Cloud's unveiling of the "Kunlun" 2nd generation 7-nm general GPU. Boasting a performance three times superior to its predecessor, this chip offers enhanced cost efficiency compared to foreign counterparts. **Kunlun's 2nd gen chip has also successfully replaced foreign-made chips in quality inspections, slashing costs by up to 65%**. Furthermore, the R&D for the 3rd generation is underway, with mass production targeted for 2024, aiming to cater to the domestic high-end demand.

Comparative Analysis with Nvidia

While it's heartening for Chinese semiconductor manufacturers, it's essential to maintain a comparative perspective. **Top Chinese AI chips are broadly analogous to Nvidia chips that were state-of-the-art five years ago.** Certain benchmarks have demonstrated Huawei's Ascend 910 matching the performance of Nvidia's V100 in AI workloads. Yet, Nvidia's H100, which China currently cannot access, is a powerhouse, outperforming the older V100 by over 13-fold (**Chart 12**).

Chart 12



AI and ML Accelerator Survey and Trends, note the computation precision legend

Risks and Challenges

China's recent maneuver to manufacture 7nm phones, with Huawei's collaboration with SMIC, is a testament to its capabilities. However, producing AI accelerators on the same architecture could ratchet up international tensions and invite further sanctions.

Furthermore, there's a broader ecosystem to consider. Merely producing top-tier chips isn't enough. To make significant inroads in AI and high-performance computing (HPC), China needs an integrated approach encompassing CPUs, GPUs, and specialized accelerators. **While producing competitive GPUs is challenging yet feasible, building CPUs that can rival long-established market leaders remains a daunting task.**

China's Disadvantage Offset: Scale and Centralization

The centralized and large-scale approach China adopts in many of its technological endeavors could be its trump card: "*On August 30, Alibaba Cloud announced the official launch of its Zhangbei Super Intelligent Computing Center, with a total construction scale of 12 EFLOPS (12 quintillion floating-point operations per second) AI computing power, which will surpass Google's 9 EFLOPS and Tesla's 1.8 EFLOPS. The new construction is now the largest intelligent computing center in the world able to provide powerful intelligent computing services for AI large model training, automated driving, spatial geography and other AI exploration applications.*" However Google is already planning a supercomputer that is double the size of Alibaba's and announced in May a 26 EFLOP supercomputer made with NVidias H100's or Cerebras' (US) planned project of 36 EFLOPS in the UAE. **Chinese companies still lack the scale of compute of their American counterparts (Chart 13). While Chinese tech giants like Alibaba have the resources to maintain a somewhat competitive scale, budding AI startups in China will struggle to find their footing amidst these dynamics.**

Chart 13

Concentration of control

Market share of cloud computing providers

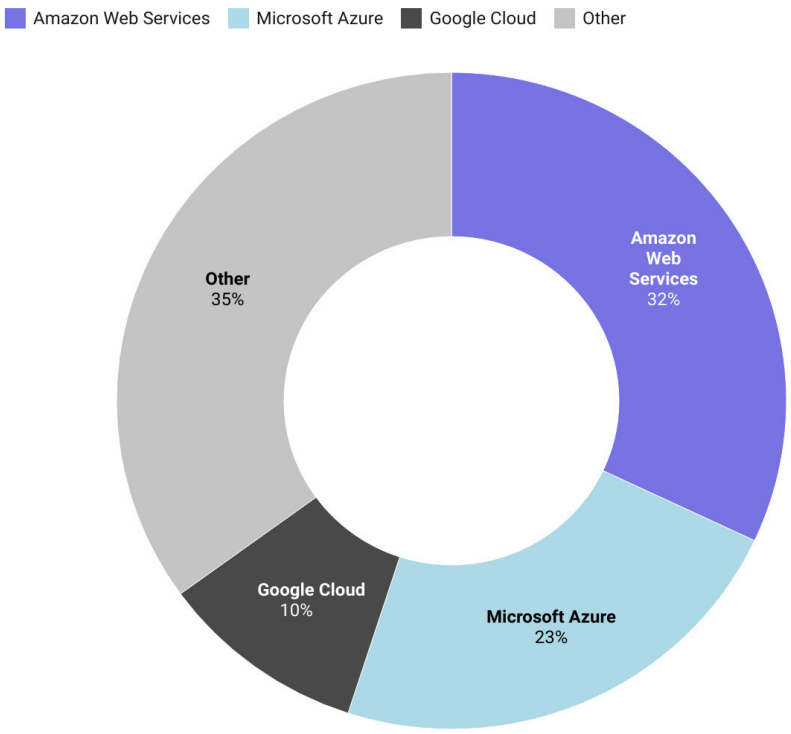
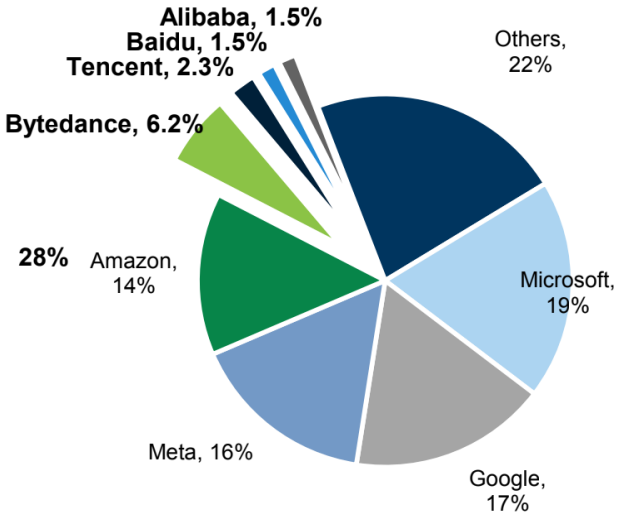


Chart: Exponential View • Source: CRN, AI Now Institute • Created with Datawrapper

Global AI server procurement share (2021)



Source: Goldman Sachs

6. AI in Great Power Competition

In assessing the trajectories of the U.S. and China within the realm of artificial intelligence (AI) and technology, several compelling narratives emerge. At present, given the framework of restrictions, the U.S. holds a slight edge in the AI competition. This advantage stems from the nation's hardware advantages and efficient allocation of resources, fostering a dynamic ecosystem conducive to technological innovation and entrepreneurship.

China's strength lies in its scale, centralization, and commitment to promoting indigenous technology. Such a strategy allows for the rapid deployment of resources on projects of national significance. The big players in China, the 'hyperscalers', benefit from this approach, however they have failed to amass the vast computational power of their American counterparts. Furthermore, Beijing's strategic approach might leave smaller entities—like nascent AI startups—scrambling for a foothold.

Drawing lessons from history, the 1970s—despite its economic challenges—saw the U.S. birth tech titans like Microsoft, Apple, and Intel (1968). These entities later provided the West with an unforeseen strategic advantage against the Soviet Union during the Cold War. Two factors, however, make this competition uniquely intense. Firstly, China's economy is large, having surpassed the U.S. in terms of purchasing power parity and inching closer on a current dollar basis. This economic might affords China a clout the Soviets never enjoyed. Secondly, China's unique blend of state-driven centralization and private sector creates an environment capable of innovation. This duality poses a significant challenge, as China's private sector is dynamic enough to potentially rival the innovative spirit of the West.

The U.S.' strategic deterrence in the ongoing AI and tech competition with China hinges on preventing China's dominance in key tech sectors and reducing reliance on it for essential commodities, while concurrently imposing systemic costs on China's tech development. Through sanctions against companies like Huawei, the U.S. hampers individual entities but also signals to the broader Chinese tech landscape its readiness to thwart significant technological leadership bids. Each time the U.S. increases sanctions on China, Beijing responds with more state support. This approach drains China's resources, forcing them into defensive technological investments. Conversely, while the U.S. struggles toward stability in tech supply chains, it has intensified efforts in cutting-edge technologies and global partnerships to consolidate its technological leadership. This duality seeks to balance immediate security with sustained innovation.

In sum, while the U.S. benefits from an efficient, entrepreneurial ecosystem, China's massive scale and fusion of state and private sectors present a formidable counter. The U.S. is better situated to innovate and breach the new frontiers of compute and AI while Chinese resources are funneled into catching up. However, when new approaches or applications emerge, Beijing has the resources and coordination to replicate foreign advances domestically. The key question remains: will the U.S's agile innovation or China's centralized resource-driven approach create the decisive advantage? The unfolding of this competition promises to shape the global AI landscape and, by extension, the strategic geopolitics of the 21st century.

However, this trajectory and these relative strengths are contingent upon China's continued access to US chip making equipment from 90nm down to 14nm as is currently the case. If that were to change, and the U.S. was able to convince its allies and force its private sector to expand the restrictions up to 28nm, the implications for the Chinese AI industry would be dire.