Semiconductors: A Strategic U.S. Advantage in the Global Artificial Intelligence Technology Race

Falan Yinug, Director, Industry Statistics & Economic Policy, Semiconductor Industry Association

August 2018

EXECUTIVE SUMMARY

In recent years, artificial intelligence (AI) has emerged at the forefront of technological development and is poised to change how individuals, societies, and nations interact. AI has the power to fundamentally transform society for decades. Semiconductor technology is at the core of all present and future AI-related innovations. At present, the United States and its private sector remain at the forefront in AI-related technology development and commercial deployment. The United States maximizes its unique AI-related advantages, most critically, through its dominance in semiconductor technology; most of the critical silicon powering AI today is developed by the U.S. semiconductor industry. Yet, other countries such as China are investing heavily to close the gap. The U.S. government and industry must deepen their partnership to enable breakthrough innovations that can sustain and strengthen this essential leadership position in AI semiconductor technology. In particular, at the federal level, the government should pursue the following actions:

1) *Fund precompetitive basic research for AI hardware*
2) *Develop the AI workforce*
3) *Promote open data policies*
4) *Mitigate potential undesirable impacts of AI*
5) *Protect security and privacy of sensitive data*

AI is a Reality, and it will Transform Society

AI has emerged from the convergence of several key developments in the technology space over the past few years: the proliferation of the Internet of Things (IoT) and the creation of extremely large data sets, Big Data analytics, a marked increase in computing power due to the adoption of cloud computing, and new developments such as accelerated computing and heterogeneous computing in the datacenter. Advances in data collection and aggregation, algorithms, and processing power have paved the way for computer scientists to achieve significant breakthroughs in AI. In short, the convergence of massive amounts of data, enormous amounts of computing, and talented data scientists have enabled AI.

AI's emergence is significant, because it has the power to fundamentally transform society. AI has moved beyond the lab, with many machine learning systems already in commercial use for a wide variety of applications. While we may be most familiar with AI-enabled consumer home devices such as digital assistants, AI is poised to transform major sectors of our economy and society in ways not seen since the invention of electricity and computer technology. By some estimates, AI could add over $8 trillion in gross value added to the U.S. economy boosting annual growth from 2.6 percent to 4.6 percent by 2035.[1]

What is Artificial Intelligence?

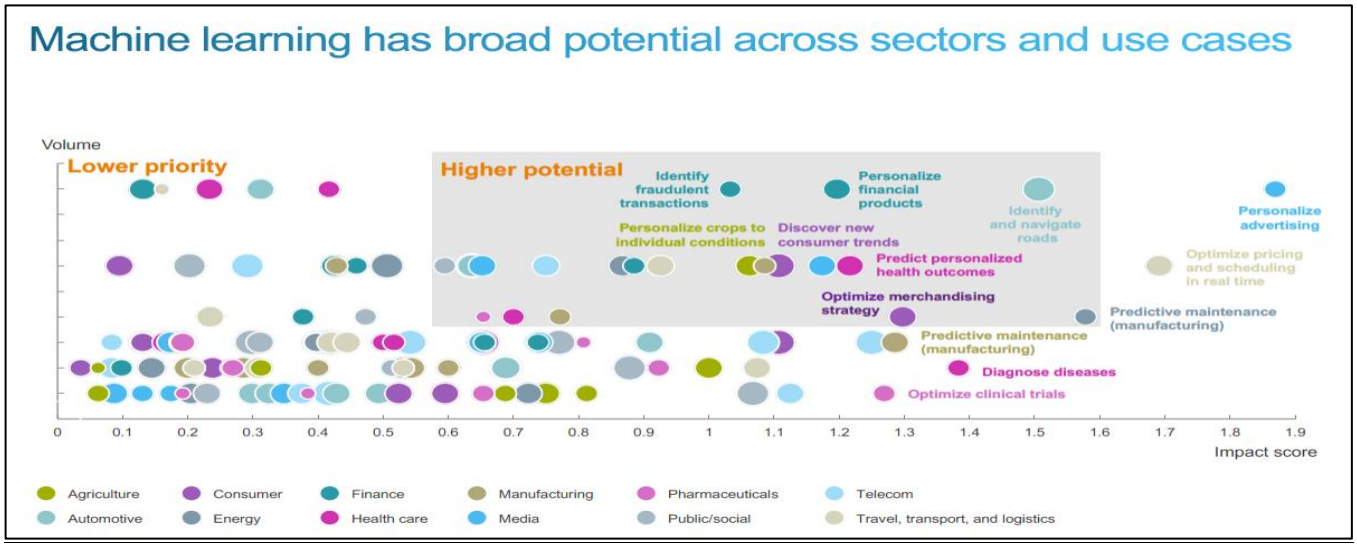
AI is often defined as the ability of machines to perform cognitive functions associated with the human mind. Machine learning (ML), a particular approach to AI, involves powerful computers digesting massive amounts of data to develop models, which can be used to infer the desired output when given similar parameters in other scenarios. This is vastly different from traditional programming, where humans instruct a predefined set of algorithms and the role of computers is merely to process and execute. In machine learning, computers take one step further, to "learn" from the given data and figure out the program themselves. Where computer systems once had to be programmed to execute rigidly defined tasks, they can now be given a generalized strategy for learning, enabling them to adapt to new data inputs without being explicitly reprogrammed.

AI adoption is currently taking place across nearly all sectors, and especially in finance, healthcare, and manufacturing, but AI application is only limited by one's imagination (Figure 1). For example, potential major challenges or problem areas where AI can help include:

- **Transportation** – AI is being used to develop autonomous vehicles that make driving safer and more efficient.
- **Healthcare** – AI can help diagnose disease earlier, develop more personalized treatment, and accelerate cancer treatment.

- **Defense** – AI can be used to develop faster cybersecurity systems.
- **Agriculture** – AI can assist farmers better maximize crop yields, reduce environmental impacts, and meet growing demand.
- **Law Enforcement** – AI helps authorities combat crime by automating and accelerating the process of going through thousands of tips.

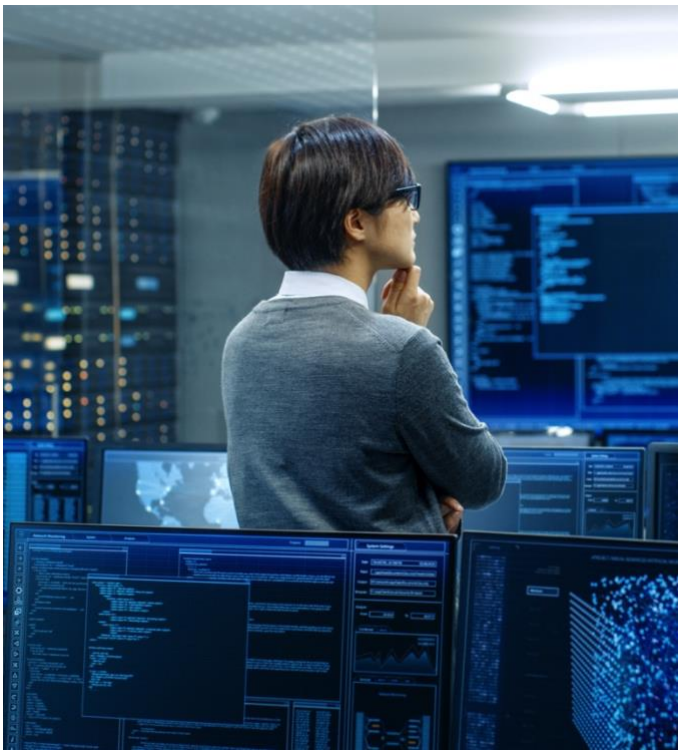
Figure 1: Potential Sectors and Use Cases for AI



Source: McKinsey Global Institute.[2]

Semiconductors are Central to Enabling the Growth of AI

Innovations in semiconductor technology enable AI. Without advances in semiconductor process technology and chip design, combined with huge amounts of data, and innovation in computing algorithms, AI could not have moved so rapidly from futuristic speculation to present-day reality. And the U.S. semiconductor industry's annual research and development (R&D) investments of $36 billion in 2017 – which also supplements investments in basic research – are a big part of what drives advances in AI applications.

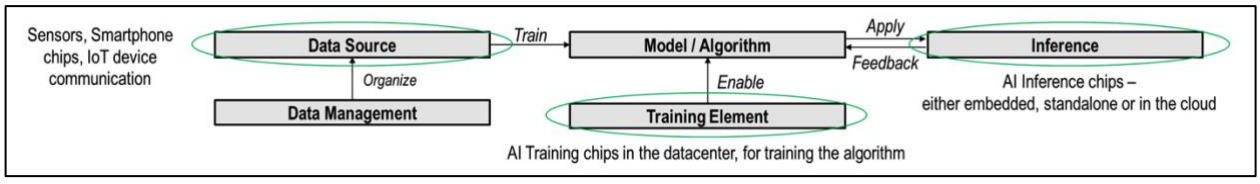


The advancement of semiconductor technology and innovation is at the core of machine learning (ML), a form of AI described above. Because ML requirements for the amount of data and computational power are much higher than for traditional programming, it also demands more semiconductor memory (to store the data) and more efficient semiconductor processors (to fulfil the need of computational power).

Indeed, semiconductors are critical in all three areas of a typical AI process flow: 1) data generation or data source – through smartphones, automotive, and multiple "Internet of Things" devices; 2) training the AI/deep learning algorithms – using graphics processors (GPUs), microprocessors, or other heavy performance-centric processors; and 3) AI inference in real world use cases on the premises or in the cloud (Figure 2).

More recently, semiconductor firms and an array of new entrants to semiconductor design have begun developing specialized semiconductor hardware – often in the form of field programmable gate arrays (FPGAs), and Application Specific Integrated Circuits (ASICs) when economically viable – that is uniquely tuned for ML workloads. Particularly for inferencing applications, these specialized chips can dramatically speed up calculations and make more computation possible at the edge, rather than in the cloud. Specialized accelerators, including GPUs tuned for lower precision calculations, FPGAs, and ASICs, are increasingly being used for training applications, as well.

Figure 2: Where Semiconductors are Needed in AI Process Flow

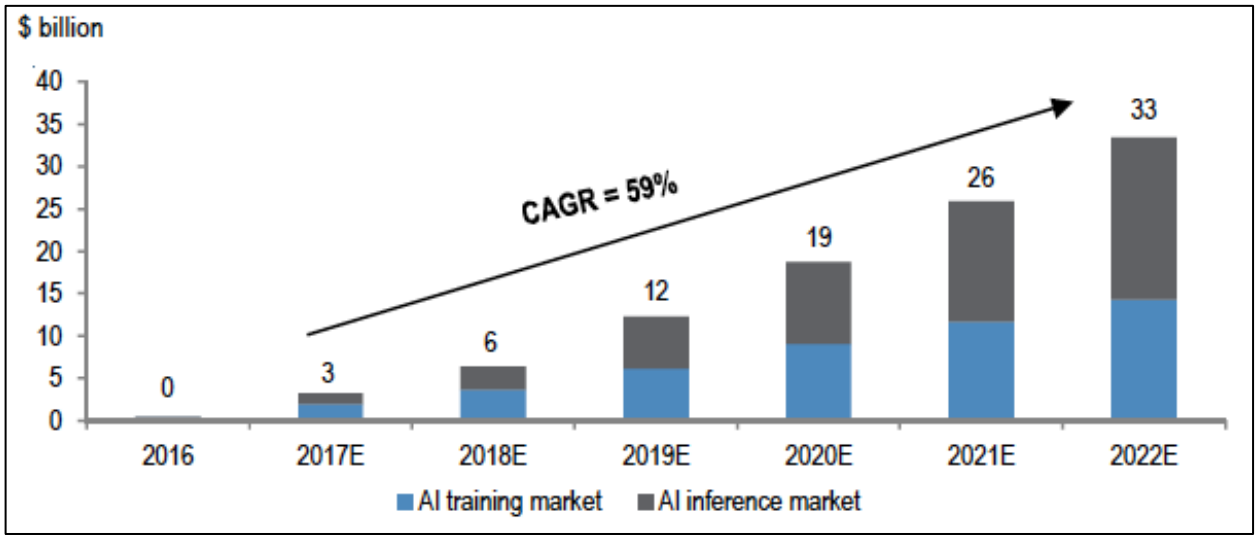


Source: J.P. Morgan.[3]

The U.S. Industry is Well-Positioned to Lead in AI, but Intense Competition Exists

The U.S. semiconductor industry is well-positioned to take advantage of AI market growth. According to some experts, the total AI semiconductor market is forecast to grow at a 59 percent compound annual growth rate (CAGR) to reach $33 billion by 2022 (Figure 3).[4] This growth will be driven by increasing use cases for AI applications, increases in AI training needs across Internet vendors and enterprises, and rising adoption of inference use cases in smartphones, cars, healthcare, agriculture, and manufacturing.

Figure 3: Semiconductor Market Revenue for AI



Source: J.P. Morgan estimates. Note: global semiconductor market to grow 5-6% CAGR from 2018-2022.[5]

The U.S. semiconductor industry currently leads in AI investment, putting the United States in a strong competitive position. In 2016, North American firms (of which U.S. firms are the leader) collectively invested $15-23 billion in AI, while Asian and European firms invested $8-12 billion and $3-4 billion, respectively.[6] Besides understanding and investing in AI, the U.S. semiconductor industry already maintains leadership in key subproduct segments that are central to AI development. Namely, U.S. semiconductor firms are currently among the global leaders in meeting increased processor demand and providing more advanced memory solutions.

However, as with any market with significant growth potential, the AI semiconductor market will see fierce global competition. Many governments are prioritizing and supporting AI development. For example, in China, one municipal government has committed to invest $5 billion in AI, while the government recently announced a $2 billion investment towards an AI development park. South Korea ($1 billion) and Canada ($125 million) have also committed significant funding to develop AI.[7]

China, in particular, views leadership in AI as a national priority. In July 2017, the Chinese government released its New Generation Artificial Intelligence Development Plan. The plan sets the ambitious goal of building China into the "world's primary AI innovation center" by 2030.[8] However, as it currently stands, China has catching up to do to be capable of domestically producing those types of high-powered, high-functioning semiconductors required for AI applications. The majority of these types of chips such as graphics processing units (GPUs), microprocessors, and field-programmable gate arrays (FPGAs) are produced by U.S.-headquartered semiconductor firms, giving the U.S. industry an invaluable, though not insurmountable, early lead in the race for future AI technology leadership.

Future Semiconductor Innovations Necessary to Sustain Rapid U.S. AI Development

In order to maintain U.S. leadership in AI-enabled semiconductors, the United Sates must maintain its position as the dominant force shaping critical technology trends. Today, we are seeing the beginning of a revolution in semiconductor advancements that will help propel AI innovations for key applications like autonomous vehicles. The semiconductor industry and its ecosystem have found innovative ways to apply existing semiconductor technology to AI applications, but moving forward, entirely new materials, architectures, and designs are necessary to achieve significant breakthroughs in the performance of AI applications. On its own, the industry invested $36 billion in R&D spending in 2017 to achieve its technology development goals, and this figure is only expected to increase.

Both traditional semiconductor industry players, and non-traditional players, such as Google and Microsoft, are investing heavily in hardware innovation to accelerate AI training and inference. Industry investments are largely focused on the design and fabrication of new forms of hardware accelerators well-tuned to AI applications.

While GPUs continue to serve many AI applications well, established firms and startups alike are investing in revolutionary chip designs tuned to the dominant AI algorithmic paradigm: deep neural networks, a branch of machine learning. The basic computations that underlie these algorithms are simple matrix or tensor multiplications – where many individual multiplications can be performed in parallel and summed at the end. By designing traditional semiconductor hardware that carries out these computations naturally – by setting parameters of an FPGA to the values of the matrix or tensor, for instance – firms have found they can dramatically speed up calculations and increase power efficiency.

At the far edge, firms are marking investments in chips based on entirely new computing paradigms, including cognitive/neuromorphic computing and quantum computing. Currently emerging devices are beginning to enter the compute ecosystem as advanced accelerators, with capabilities well beyond the FPGAs and ASICs already flooding into the compute landscape for targeted applications.

In cognitive and neuromorphic computing, companies across the industry are experimenting with hardware naturally suited to the less precise calculations needed for machine learning – e.g., memristors and phase change memory. Some early devices are already finding limited applications in the market as powerful AI inferencing chips.

Government Policy Will Play a Crucial Role in Cementing U.S. Leadership in AI Semiconductors

Unique policy challenges will need to be addressed in order for the U.S. semiconductor industry to achieve and maintain global leadership in this dynamic and exciting market. At the federal level, there are several specific areas where the government can play an appropriate role:

1) *Fund precompetitive basic research in AI hardware* – the federal government needs to increase research funding through agencies such as the NSF, DARPA, and NIST for programs that promote basic research in AI. Such funding is particularly critical in the wake of other countries' prioritization of research funding for AI. In addition, continued specific government funding for the development of faster supercomputers, such as exascale, is essential for AI research.

2) *Develop the AI workforce* – the federal government should support policies that help prepare workers to create and work alongside AI technologies. This starts with supporting STEM education in elementary and secondary schools, including research funded by NSF to evaluate effective teaching techniques and practices, but it also includes federal funding for the training of graduate-level scientists and engineers who are the future AI innovators. A secondary focus in this area should be effectively addressing displaced workers with retraining and other support. AI will transform society, and such a change will clearly redefine work as we know it. The government should anticipate such changes and begin now to explore policy solutions.

3) *Promote open data policies* – the federal government should provide access to some of the most comprehensive data sets it possesses. Data is the fuel that drives AI, and to realize AI's fully potential, large government datasets must be available to researchers.

4) *Mitigate potential undesirable impacts of AI* – the federal government should partner with industry to establish voluntary guidelines for the ethical use of AI and to help prevent it from being used in ways that could harm the public. This should include efforts to identify which AI applications should require further explanation and investment in R&D on needed explainability.

5) *Protect security and privacy of sensitive data* – the federal government should partner with industry to facilitate the development of effective security and privacy mechanisms from their theoretical foundations to their effective implementations and lifecycle management. This includes funding research into novel approaches to threat modeling in an ever-more-connected ecosystem, security-aware software execution models to counter both known and unexpected threats, security-aware hardware and software development design/co-design tools to build security in our systems at their foundations, and security and privacy foundations, including cryptography.

---

[1] Accenture, press release, September 28, 2016.
[2] Figure extracted from McKinsey Global Institute, "What's Now and Next in Analytics, AI, and Automation?" page 6.
[3] Figure extracted from JP Morgan, "AI and Semiconductors," page 9.
[4] JP Morgan, "AI and Semiconductors," February 7, 2018.
[5] Figure extracted from JP Morgan, "AI and Semiconductors," page 4.
[6] McKinsey Global Institute, "Artificial Intelligence: The Next Digital Frontier?" November 2017.
[7] New York Times, "As China Marched Forward" and Reuters, "Beijing to Build $2 billion AI Research Park."
[8] New America, "From Riding the Wave."

BIBLIOGRAPHY

Accenture, Press Release: "Artificial Intelligence Poised to Double Annual Economic Growth Rate in 12 Developed Economies and Boost Labor Productivity by up to 40 Percent by 2035, accounting to New Research by Accenture," September 28, 2016. Accessed at: https://newsroom.accenture.com/news/artificial-intelligence-poised-to-double-annual-economic-growth-rate-in-12-developed-economies-and-boost-labor-productivity-by-up-to-40-percent-by-2035-according-to-new-research-by-accenture.htm.

Bernstein, Report: "Artificial Intelligence: Hope or Hype for the Semiconductor Industry?" released September 5, 2018.

Intel Corporation, written submission to the U.S. House of Representatives, Committee on Oversight and government reform, Subcommittee on Information Technology, Hearing "Game Changers: Artificial Intelligence Part I", February 14, 2018. Accessed at: https://oversight.house.gov/wp-content/uploads/2018/02/Khosrowshahi-Intel-Statement-AI-2-14.pdf.

J.P. Morgan, Investor Report: "AI and Semiconductors: Exponential Growth from AI Adoption in the Cloud and at the Edge," released February 7, 2018.

McKinsey & Company, "Artificial Intelligence: The Time to Act is Now," January 2018. Accessed at: https://www.mckinsey.com/industries/advanced-electronics/our-insights/artificial-intelligence-the-time-to-act-is-now.

McKinsey Global Institute, presentation: "Artificial Intelligence: The Next Digital Frontier?" November 2017.

McKinsey Global Institute, "What's Now and Next in Analytics, AI, and Automation," May 2017. Accessed at: https://www.mckinsey.com/global-themes/digital-disruption/whats-now-and-next-in-analytics-ai-and-automation.

New America, "From Riding a Wave to Full Steam Ahead," February 28, 2018. Accessed at: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/riding-wave-full-steam-ahead/.

New York Times, "As China Marches Forward on A.I., the White House Is Silent," February 12, 2018. Accessed at: https://www.nytimes.com/2018/02/12/technology/china-trump-artificial-intelligence.html.

New York Times, "Is China Outsmarting America in A.I.?" May 27, 2017. Accessed at: https://www.nytimes.com/2017/05/27/technology/china-us-ai-artificial-intelligence.html.

NVIDIA Corporation, written submission to the U.S. House of Representatives, Committee on Oversight and government reform, Subcommittee on Information Technology, Hearing "Game Changers: Artificial Intelligence Part I", February 14, 2018. Accessed at: https://oversight.house.gov/wp-content/uploads/2018/02/Buck-NVIDIA-Statement-AI-2-14.pdf.

Reuters, "Beijing to Build $2 Billion AI Research Park: Xinhua," January 3, 2018. Accessed at: https://www.reuters.com/article/us-china-artificial-intelligence/beijing-to-build-2-billion-ai-research-park-xinhua-idUSKBN1ES0B8.

Semiconductor Industry Association and Semiconductor Research Corporation, "Semiconductor Research Opportunities: An Industry Vision and Guide," March 2017. Accessed at: https://www.semiconductors.org/clientuploads/Research_Technology/SIA%20SRC%20Vision%20Report%203.30.17.pdf.

World Economic Forum, "Artificial Intelligence is going to Completely Change Your life," November 10, 2017. Accessed at: https://www.weforum.org/agenda/2017/11/artificial-intelligence-is-going-to-completely-change-your-life?utm_content=buffer9a9a8&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer.