ARTIFICIAL INTELLIGENCE – Software That Writes Itself

Jim McFadden, CFA **April 10, 2018**

203-912-7983

SUMMARY

True artificial intelligence (termed artificial general intelligence – see below), where smart machines perceive the world as we do, fully comprehend our language and instructions, and, perhaps most importantly, continuously learn from training and examples, is the holy grail of computer science. After numerous fits and starts in the sixty years since artificial intelligence (AI) was initially contemplated, this technological discipline seems now to be on an uninterruptible growth path. Much of the explosive growth experienced over the past few years, and projected for many years to come, can be attributed to three very propitious factors: a torrent of data in the form of text, images, mapping data, etc. (Big Data); a virtually unlimited capacity to store that data; and wide availability at fairly cheap prices of parallel processing computer power to analyze and manipulate the data (see Graphics Processing Unit section below). Phrased another way, the virtuous cycle of the increased interconnectedness of information technology resources and cloud computing architecture, easing the ability to organize and analyze big data, has provided a tremendous boost to the development of AI.

AI technologies have been responsible for significant advances in diverse industries such as automotive, health care, retail, manufacturing, and financial services. These and many other industries will be altered by this discipline over the near and intermediate terms, and AI's impact on these industries – and the investment implications of these actions -- will be profound.

The early adopters of AI have used the technology primarily for cost-cutting reasons: customer service, human resources, finance, and accounting. Increasingly, AI will be used to boost corporate revenues. Overall AI spending is expected to explode over the next few years: the market research firm International Data Corporation expects worldwide spending on AI hardware, software and services will jump to $58 billion by 2021, up from $12 billion in 2017. (Other industry analysts and consultants forecast similar increases.)

ARTIFICIAL INTELLIGENCE – THE BASICS

By definition, AI entails machines that can perform tasks that are characteristic of human intelligence such as object recognition, sound recognition, learning, and problem solving. Machine learning is simply a way of achieving AI and is considered to be perhaps the bedrock concept of AI. In machine learning, a designer creates an algorithm which optimizes automatically with limited or no additional human involvement. In other words, the algorithm "writes itself" by learning from the data. The algorithm, which has a "base architecture" built on linear regression models, searches for patterns in large amounts of data from diverse sources. But more than this, a machine learning algorithm is not limited to traditional linear relationships. Instead, a successful machine learning step-by-step mathematical

process entails the automated optimization, prediction and categorization of data, but does not attempt to deduce or infer causation of the data. For example, an algorithm might predict interest rates movements based on a variety of economic and social factors, but the algorithm does not deal with the cause of those economic or social factors. The following are the broad categories of machine learning algorithms:

- Supervised Learning. Under this form of AI, an algorithm analyzes a set of training data whereby one set of the data points may have the designation of "fraudulent" or incorrect; another set may be labeled "genuine" or correct. Under supervised learning, the mathematical process "learns" a general rule that allows prediction of labels for the remaining points in the data set. As an aside, NVDA GPU chips are considered to be particularly strong in the "training" aspect of AI. See page 4.

- Unsupervised Learning. Unlike the above case, data analyzed by the algorithm in unsupervised learning do not contain any "fraudulent" or "genuine" labels. Instead, the algorithm detects patterns in the data by itself by identifying groups of observations that depend on similar baseline attributes.

- Reinforcement Learning represents a blend of supervised and unsupervised learning. As in unsupervised learning, the algorithm analyzes raw, unlabeled data and makes a choice or decision for each observation. A human then provides a response to the algorithm on its choice, which in turn assists the program to "learn."

- "Deep Learning" algorithms, also termed artificial neural networks, operate in layers, similar to the interconnected neurons in the human brain. (Of course, in the human brain, a neuron can connect with practically any other neuron within a certain physical distance; whereas, artificial neural networks have discrete layers and directions of data propagation.) In deep learning, each neuron attempts to identify a data point. In turn, the neurons combine to work on a goal, such as determining what is actually in a photo or what is actually said in spoken words. Deep learning can be used in supervised, unsupervised or reinforcement learning.

  Figures 1 and 2 on the next pages are graphical representations of first, the relationship between AI, machine learning and deep learning; and second, the mechanism of a simple artificial neural network.

Figure 1: RELATIONSHIP BETWEEN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DEEP LEARNING

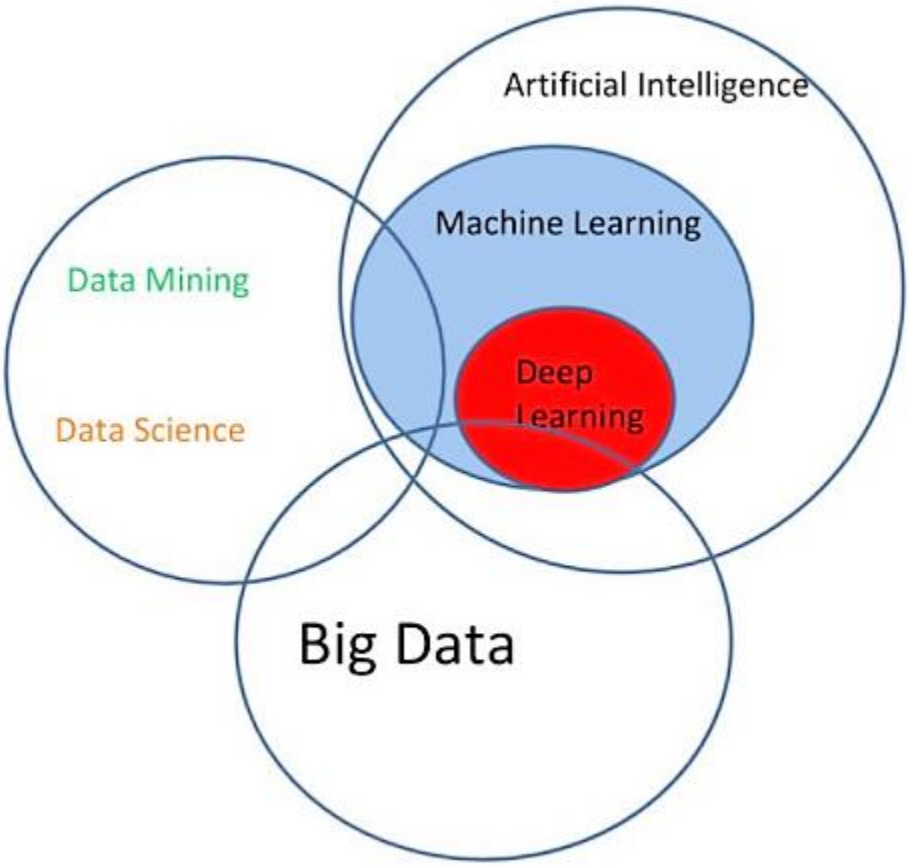
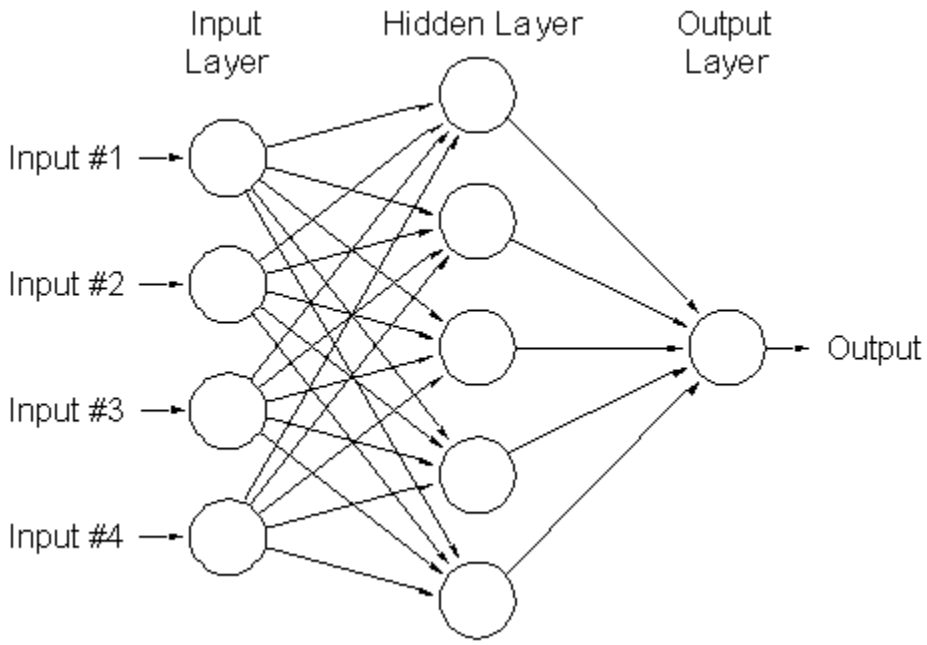


Figure 2: SIMPLIFIED REPRESENTATION OF AN ARTIFICIAL NEURAL NETWORK



In order to train a neutral network, a designer must choose the appropriate weights and preconceptions of the algorithm so that the output of the algorithm closely matches the known inputs (or training data). Of course, after the training is complete, the model has to complete a task or make a decision. This decision-making is termed inference, whereby the trained model deduces, or infers, an answer in real time based on its trained reasoning process.

Deep learning has resulted in breakthroughs in image recognition and natural language processing, where computers are able to understand and respond to spoken language (for example, Apple's Siri). Indeed, image recognition by some deep learning machines has proved to be better than humans. In addition, the concept of driverless cars is predicated on the concept of deep learning (see NVIDIA Automotive section).

Looked at from a broader perspective, deep learning models hinge on two primary factors, computing power and data, both of which are growing at exponential rates. NVDA has for instance increased the speed of its deep learning chip system by 500x over the past five years. As a result, growth in deep learning applications promises to be very rapid. Deep learning end-to-end AI platforms will be built in a way which bridges training, inference and many interconnected smart devices. For example, in the not-too-distant future, drones will independently maneuver through a warehouse; find, say, the appropriate pipe and pipe fittings

a customer has ordered; and deliver them to the loading dock. To extend the concept just a little further, Amazon's vision of a drone's delivering a product to the front door of a customer seems quite realistic.

- Artificial General Intelligence (AGI). AGI, that level of intelligence when a machine is capable of abstracting concepts from limited experience and transferring knowledge between domains, has of course not yet been achieved. This remains the ultimate goal of AI.

GRAPHICS PROCESSING UNIT

A graphics processing unit, or GPU, is a specialized electronic circuit designed to accelerate the creation of images for a display device and is very efficient in image processing and for computer graphics. The GPU's primary focus for many years was to improve the graphics on PCs. However, the highly-parallel structure of GPUs is ideal for the processing of algorithms as large blocks of data can be processed in parallel by the component. Indeed, NVDA found in the mid-2000s that its GPUs greatly accelerated deep-learning algorithms. As a consequence, the GPU has become a key building block for AI. See NVIDIA Summary section on page 10.

INDUSTRIES OR AREAS WITH THE MOST INTENSE FUTURE APPLICATIONS OF AI

As noted above, AI is likely to affect most industries and therefore investment opportunities in the positively- or negatively-affected sectors. In order to start the discussion, listed below is the research firm Tractica's top ten use cases for AI by the year 2025:

Algorithmic trading strategy performance improvement;

Static image recognition, classification and tagging;

Efficient, scalable processing of patient data;

Predictive maintenance;

Content distribution on social media;

Text query of images;

Automated geophysical feature detection;

Object identification, detection, classification, and tracking from geospatial images;

Object detection and classification – avoidance, navigation;

Contract analysis.

A few observations regarding this list:

- Giant internet retailers and social media companies which utilize Big Data to reach customers with product offerings aligned with their tastes and interests and to gain increasing advertising revenues because of their ability to target those customers seem extremely likely to benefit from increased AI availability and usage -- *the rich get richer*. The volume of data which internet giants already possess on individual consumers is astonishing. See the article below on the mountain of very personal information that Facebook possesses on each of its users

  https://www.cnbc.com/2018/03/27/facebook-knows-a-lot-about-me.html

  Using ever more sophisticated deep learning algorithms will only increase the already-sizable advantages of Google, Amazon, Facebook, Alibaba, Tencent, and Baidu. The companies' relationships with customers will be even tighter, and barriers to entry will be even more pronounced as information technology costs rise.

  Early adopters have realized eye-opening – and perhaps surprising – results. Fully 55% of AMZN's sales are driven by recommendations of its machine learning algorithms. In addition, Netflix has said publicly that it believes that its AI-driven personalization and recommendation engine generates $1 billion per year for the company, as NFLX believes many subscribers would abandon its service if the recommendation algorithm were not operating. (Note that NFLX's gross margin in 2017 was about $4 billion.)

- E&P companies and oil service companies will likely become large consumers of AI technology for the detailed analysis of seismic data (although probably not for deep sea projects). AI could also analyze production and sensor data like pump pressure, flow rate and temperature, which in turn could predict the timing of equipment failure. To this point, most of the AI spending in this sector has centered on an Exxon initiative to develop intelligent robots which detect oil seeps from rock on the ocean floor. Detecting these seeps is not only a positive for ecological reasons but can also point to the possibility of a large energy resource. Baker Hughes and Devon Energy have also begun to train a deep learning algorithm for use in future drilling.

- AI's ability to identify and track objects could translate into substantially higher military spending, particularly on missile defense systems. Another point: Russia's recent announcement that it is developing smart missiles that can evade current missile defense systems could hasten the introduction of AI to U.S. defense spending.

- A variety of health care costs could be reduced through improved diagnosis techniques afforded by AI and from vast improvements in the categorization of patient data. Nevertheless, according to a 2017 survey done by Healthcare IT News and HIMSS Analytics, only about 5% of

healthcare organization are currently using AI. Listed below are a few examples of how AI could save lives (and costs) in the medical field:

Seventeen million people die each year due to cardiovascular diseases. It is known, however, that markers of this disease can often be found in the eye. AI retinal imaging, combined with deep learning methods which have been used for years to help diagnose eye diseases, could potentially play a key role in slowing the rate of cardiovascular deaths.

AI technologies can reconstruct and enhance images obtained in conventional scans of the body. For example, using cinematic rendering (see Ray Tracing on page 15), one can now get a 3-D view of the left ventricle of a beating heart, allowing a determination of the amount of blood being injected into it. Similarly, researchers are training AI to identify tuberculosis on chest x-rays and to detect intracranial bleeding resulting from head trauma and stroke. To illustrate the scope of the opportunity, Massachusetts General Hospital has a data base of 10 billion images from years of patient data. The hospital is engaged in 40 AI projects with NVDA.

Mount Sinai Hospital in New York is an early adopter of deep learning to aid in disease diagnosis.

The UK's National Health Service has entered into a partnership with Google's DeepMind AI unit to improve the process of delivering care with digital solutions.

Many sectors in the health care space could benefit from better diagnoses and cost savings related to the widespread implementation of AI. In particular, insurers' costs should decline noticeably.

- AI will be a huge factor for pharma companies in the development of future drugs. Studies show that it costs more than $2 billion over a 10-year period to develop and test a drug candidate. AI can evaluate huge reams of public health data such as results of prior clinical trials and high-resolution medical images. In turn, the technology can find patterns in the data, cutting drug development costs.

  In addition, prior to gaining FDA approval, drugs have to pass rigorous testing procedure that must be documented and submitted to authorities. AI can generate reports, slicing months off the process.

- The race to build a fully and smooth-functioning self-driving car is one of the most heavily publicized applications of AI. The technology challenge is of course daunting: an algorithm must be able to analyze and interpret data from multiple cameras, radar and sensors to steer the vehicle around static and ever-changing dynamic objects. At the same time, the vehicle must determine in real time the particulars of the road and lanes it must navigate relative to a high-definition map, while factoring in changeable weather, road conditions and construction zones. Perhaps equally important, the algorithm must monitor the behavior and movements of the passengers in the vehicle's cabin. Building a precise virtual reality driving simulator is also of

paramount importance to factor in all possible scenarios, as the miles to be driven by any fleet of self-driving test cars is dwarfed by the actual number of miles driven manually – 10 trillion miles per year worldwide. In a nutshell, a self-driving computer platform must perform four tasks: collect data, train various automotive models, simulate multitudes of driving conditions, and of course ultimately drive the vehicle.

AI should a very positive effect on the trucking industry. Initially, deep learning should extend the range of truckers by mapping out more efficient paths and times of travel. Longer term, self-driving trucks would also solve one of the limiting factors in that industry: finding and keeping qualified drivers.

- AI applications will gradually find their way onto the manufacturing factory floor, but probably at a slower pace than in other industries. One example: GE built a $200 million "brilliant factory" in India which manufactures diverse products from jet engine parts to locomotive components. According to GE, AI has improved equipment effectiveness by 18%. GE now has seven brilliant factories. At those facilities, unplanned downtime has been cut 10-20%.

  Another example: in late 2016, FANUC, the Japanese industrial robotics giant, announced plans to build the factory of the future based on a NVIDIA AI platform. FANUC will use NVIDIA chips as the artificial engine for its industrial robots, creating "thinking robots" that learn skills on their own and communicate with fellow robots to boost productivity.

  According to CFO Magazine, for the average Fortune 100 company, a shortening of the supply chain by just one day would free up $50-100 million in annual cash flow. So, if AI could streamline a company's supply chain – which seems quite likely for a large company -- that would create immediate value.

  A key consideration regarding the potential impact of AI in manufacturing: If the technology allows manufacturing to be more flexible and were to become widely deployed, causing customization to become cheap enough, that would cause a tremendous shift in numerous markets.

- Because of AI, cyberattacks will be more damaging and cyberattacks will be more effective. Bots, or software applications that run automated tasks over the Internet, have for many years probed web servers searching for vulnerabilities. In the future, AI-powered bots will be developed to learn, adapt and evade cybersecurity. To combat these advances, cybersecurity companies like Palo Alto Networks and Check Point Software could realize substantial new sources of revenues.

- AI will dramatically reduce the costs of legal review and compliance in the financial industry. For example, JP Morgan is employing AI-powered software in order to review and approve commercial loan agreements, eliminating about 360,000 hours annually of work that had been

performed by lawyers and loan officers – with fewer mistakes. In the banking industry as a whole, compliance and regulation costs total about $270 billion per year, or about 10% of the operating costs of a typical financial service company. A huge portion of these costs will be eliminated.

- Enterprise software vendors like SAP, Oracle and Salesforce.com are embedding AI into their offerings.

NVIDIA

SUMMARY

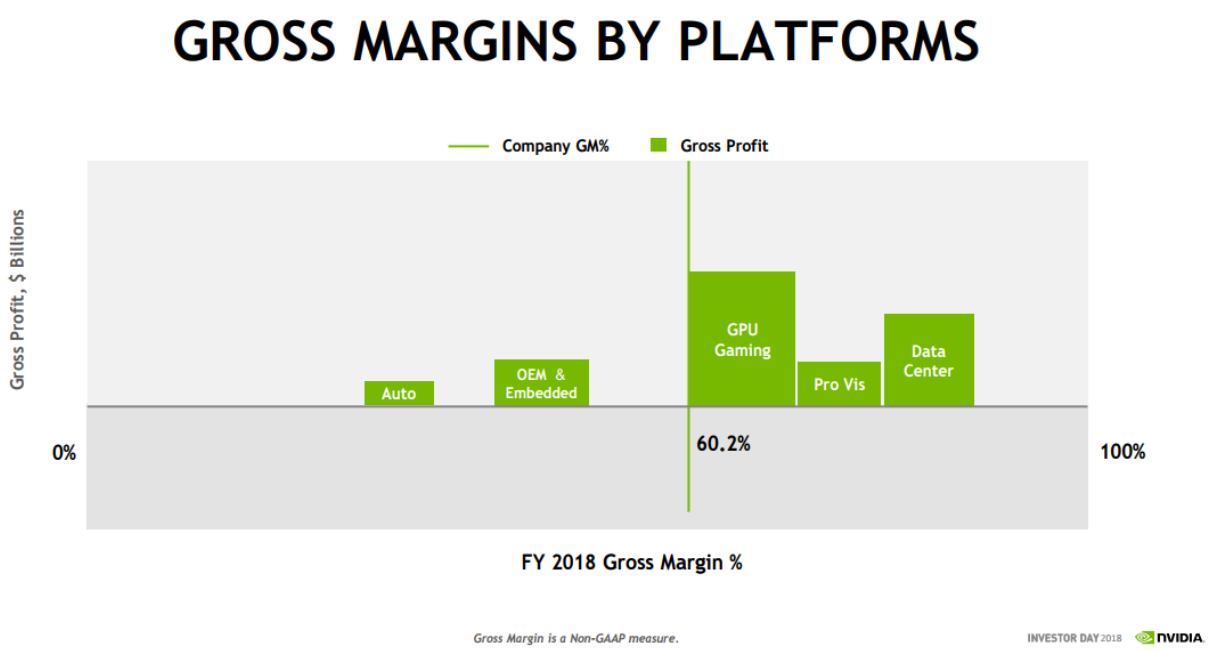
Nearly 20 years ago, NVDA introduced tremendous advances to the graphics processing unit (GPU). In turn, the GPU graphics accelerator became the backbone of more complex and detailed video games and has played a key in the growth of virtual reality technology. Equally important, the GPU has proven to be very efficient in processing large blocks of data simultaneously given its parallel processing structure, making it an ideal solver of machine learning and deep learning artificial intelligence algorithms.

As a result, NVDA plays a central role in the growth – in many cases, the creation – of artificial intelligence technologies and autonomous driving vehicles, businesses with seemingly uncapped growth potential. At the same time, NVDA's core gaming business continues to expand rapidly, as interest in gaming broadens and deepens across all age classes (see below) and a new precise graphics technology promises to usher in cinematic-quality video games.

GPUs are less power-efficient than specialty chips, but, as noted above, can handle with dexterity complex graphics like virtual reality, augmented reality and gaming. Power efficiency is a critical factor for cell phones and data centers. (GPUs are, however, making rapid inroads in data centers. See page 12.) NVDA has very little presence in mobile GPUs; Qualcomm dominates that space, making mobile GPUs for Samsung, HTC and LG, among others.

NVDA has four major business segments – Gaming, Datacenter, Professional Visualization, and Automotive. GPU Gaming, Datacenter and Professional Visualization are extremely high-margin businesses; each had a FY 2018 (January) gross margin well above NVDA's consolidated figure of 60%. Automotive, which will be in investment mode for the next couple years, posted a solid gross margin, but below the corporate average. On the following page, please see Figure 3.

Figure 3: FY2018 GROSS MARGINS OF NVDA'S KEY SEGMENTS



NVDA's Datacenter, Gaming and Automotive segments are discussed in detail below.

DATACENTER

	Revenue	Year-Over-Over Revenue Growth	Gross Margin
FY 2018 (Jan.)	$1,932 million	133%	Greater than 60.2%
FY 2017	830 million	145%	N/A
FY 2016	339 million		N/A

NVDA's datacenter revenues over the past few years have tracked the spectacular growth in AI applications, projects and research initiatives. The company's GPU-powered accelerated computing platform which is designed to speed both the training and inferencing of neural networks is available in servers made by every major computer manufacturer and from every cloud services provider, including Google Cloud, Microsoft Azure and Amazon Web Services. Of special note, sales to hyperscale

computing (Big Data and cloud computing) customers increased 155% in FY 2018, even faster than the overall Datacenter segment pace.

A recent Deloitte Global study illustrates the exponential growth in datacenter AI projects and the demand for machine learning chips, primarily GPUs. The consulting firm projects that about 800,000 machine learning chip systems will be sold to datacenters in calendar 2018, up from 100,000 to 200,000 in 2016. To put this in perspective, a top-of-the-line NVDA GeForce 1080 Ti GPU graphics card used in gaming lists for about $700 (and sells for about $1,000 on Amazon.com and newegg.com). Until fairly recently, some of these devices were used in datacenters (see just below). A NVDA Tesla V100 BPU graphics card which is designed for datacenter use sells for around $9,000. Multiplying any of these figures by 800,000 units implies a huge potential market opportunity for NVDA.

(NVDA does have a high-class problem: too much demand for its products. Some data centers have purchased strings of GeForce gaming GPUs or $3,000 NVDA Titan V products that are designed for AI research to utilize in their servers as opposed to the V100 GPU described above. In late 2017, NVDA changed the terms of its licensing agreements to deter such substitutions. These product substitutions by data centers also probably served to understate NVDA's true Datacenter revenue in FY 2018 and before.)

A key positive for NVDA's products: gains in speed and efficiency far surpassing the confines of "Moore's Law" (which states the number of transistors per device, and therefore device speed, doubles approximately every two years). The speed of NVDA's top-end GPU-based server product (now named the Volta 100) has increased by 25x since 2013. Similarly, NVDA's standalone supercomputer DGX-2 can train machine learning models 500x faster than it could five years ago. (A DGX-2 can replace 300 conventional servers.)

To this point, training neural networks has been the specialty of NVDA's chip systems; however, inferencing represents a huge additional growth avenue for the company. Indeed, 2,000 customers have adopted NVDA's inference platform in FY 2018. Inferencing is an extremely complex problem – significant improvements in speech recognition must be accomplished – but it represents a substantial opportunity. According to NVDA, its total addressable market in training and inferencing combined for hyperscale computing customers could be $20 billion. The company has a very small inferencing market share, but it is rapidly accelerating. Cloud computing datacenters downloaded NVDA's inferencing optimization tool 30,000 times in FY 2018 versus only 70 times in FY 2017. Similar downloads of NVDA's training optimization tool reached 840,000 in FY 2018 versus 280,000 in FY 2017.

Accelerating the performance of high performance computers (HPCs), also known as supercomputers, represents another growth path for NVDA – and an affirmation of the computing power of the company's GPUs. A total of 102 of the world's 500 fastest supercomputers have been accelerated; eight-six of the 102 were accelerated with NVDA's system. Over time, every supercomputer will have to be accelerated to handle growing deep learning requirements; NVDA is very likely to capture the lion's share of this high-margin business.

Perhaps more importantly, datacenters across the world will similarly have to be accelerated to handle deep learning needs – perhaps with the new DGX-2 computer. According to NVDA, the company holds a market share of 90% or more in datacenter accelerations.

GAMING

	Revenue	Year-Over-Over Revenue Growth	Gross Margin
FY 2018 (Jan.)	$5,513 million	36%	Greater than 60.2%
FY 2017	4,060 million	44%	N/A
FY 2016	2,818 million		N/A

NVDA's GeForce gaming platform (which is based on NVDA 1080 Ti and 1070 Ti GPUs) had 100 million users in FY 2018, up from 90 million the year before. Further sharp increases in users seems assured over the foreseeable future. Note the following points:

- The number of worldwide gamers has tripled over the last ten years to more than 2 billion. Perhaps more important, the average age (and presumably the income status) of "gamers for life" is likewise growing rapidly. Last year, nearly 50% of this group was 35+ years old versus less than 25% in 2003. Even more positive, given our technology- and computer-driven culture, it is not much a stretch to say nearly everyone born from now on in non-third world countries is likely to be a gamer.

- NVDA's gaming GPUs that were introduced two years ago – NVDA's Pascal GTX 10-series graphics cards -- still notably outpace the newest release by AMD, NVDA's only real competitor in the manufacture of desktop discrete GPUs. The following review excerpt issued in March 2018 by pcgamesN.com is representative: "Unfortunately right now, the AMD RX Vega cards suffer by comparison to their smaller, more efficient, and more powerful competition." The AMD Vega cards were released in late 2017.

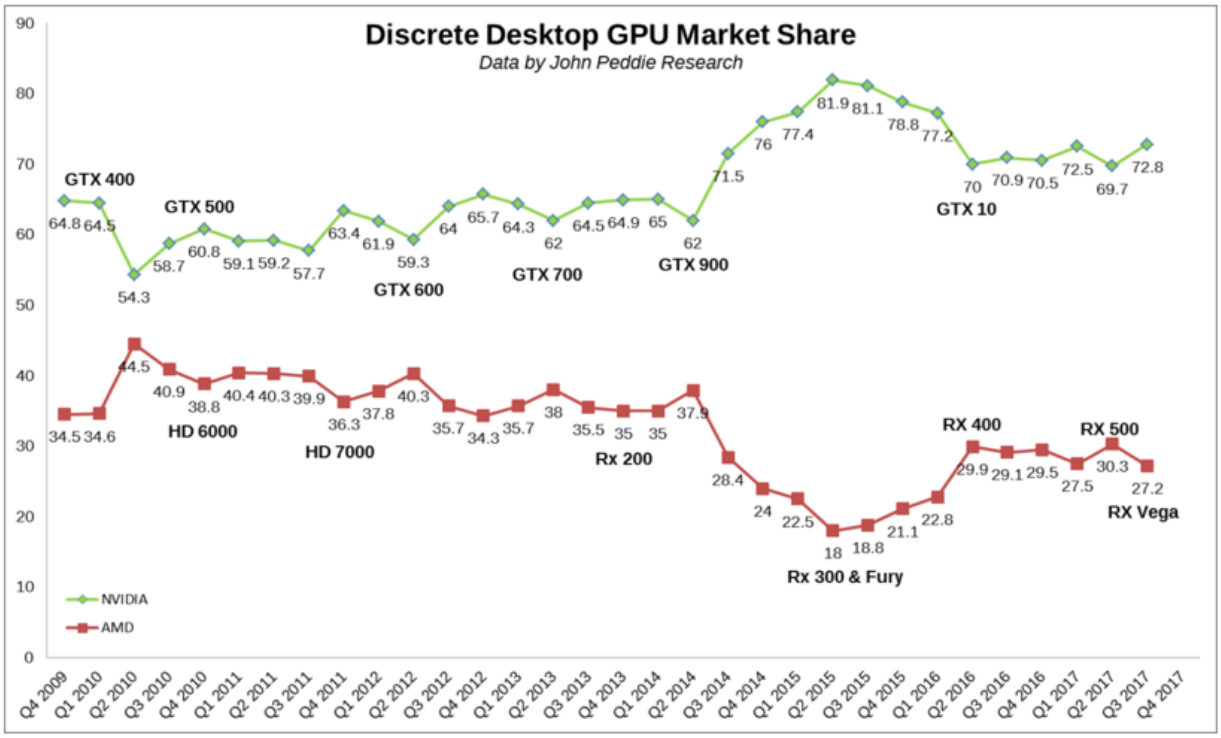
- NVDA released two professional level GPUs in 2017: the Tesla V100 for AI applications and for data centers and the Titan V for high-performance computing needs. A new Volta gaming GPU (the 11-series) is expected to be released sometime in 2018. Given the advanced technical parameters of the V100 and the Titan V, NVDA's new graphics seems likely to extend its technical superiority vis-a-vis AMD.

- According to NVDA, just 30% of gamers which use NVDA graphics cards use cards with the current Pascal GPU architecture. Some 70% of NVDA gamers use older, slower NVDA graphics cards, implying substantial upgrade demand potential.

- Cryptocurrency miners became large buyers of GPUs starting about a year ago, as the chips are well suited to the miners' massive computational requirements. In an effort to limit sales to miners in favor of gamers, NVDA in early 2018 began encouraging retailers to curtail sales of gaming chips to miners. So far, AMD has made no similar provisions. Clearly, sales to crypto miners boosted NVDA's and AMD's sales in calendar 2017. One well-regarded industry source (Peddie Research) estimates that miners bought $776 million of graphics cards in 2017, primarily

from AMD. (This $776 million estimate seems overstated as AMD has said its "blockchain" revenues represented a mid-single digit percentage of its 2017 revenues, equivalent to about $200-$300 million.) AMD's sales to cryptocurrency miners has played a key role in boosting its discrete video card market share to 33.7% in 4Q 2017, up from 29.5% in 3Q 2017. Figure 4 below shows the NVDA/AMD discrete GPU market share split since late 2009.

Figure 4: **Discrete Desktop GPU Market Share**



Source: Peddie Research

With the need to purchase GPUs quickly, the miners were generally willing to buy any available product. To illustrate this, Peddie Research estimates that mid-range GPUs comprised 51.7% of shipments in the fourth quarter of 2017, up from 41.5% in the fourth quarter of 2016. A testament to the appeal of mid-range graphics cards: GPUs like AMD's RX Vega 64 graphics card, which originally costs $500, sells for over $1000 on Amazon. Similarly, NVDA's flagship GTX 1080 Ti transacts at around $1000 on Amazon, about $300 above its list price. (With the sharp recent declines in cryptocurrencies, graphics card prices have retreated somewhat over the past week or two.)

(Another reason that NVDA and AMD have struggled to keep up with demand has been a global DRAM shortage. In 2017, DRAM manufacturers increased production of DDR4 memory products used in cell phones at the expense of GDDR5 and HBM2 video memory products used in graphics cards and for high-performance computing.)

- NVDA's gaming revenue from the sale of GPUs totaled about $4.6 billion in FY 2018, up 21% from around $3.8 billion in FY 2017. As noted in the summary table on page 13, the company's total revenue in its gaming unit amounted to $5.513 billion and $4.06 billion in FY 2018 and FY 2017, respectively, implying that non-GPU gaming revenue was about $900 million in FY 2018, up dramatically from $260 million in FY 2017. This incremental and rapidly-growing non-GPU gaming revenue pushed total gaming revenue to a 36% increase year-over-year in FY 2018. NVDA's non-GPU gaming businesses include the sale of cloud-based devices (computer on a chip) and services for home entertainment and gaming; cloud-based gaming; and platforms for specialized console gaming devices.

- Esports, or competitive video gaming, will be a key driver of future growth. Esports began in South Korea, and it has grown exponentially over the last few years. The NewZoo Esports website says that 86% of Esports gamers play on NVDA's GeForce platform.

  A few Esports anecdotes: According to Rolling Stone in December 2017, "more than 360 million unique viewers watched this year's *League of Legends* Mid-Season Invitational." At a single Esports event in Poland, 175,000 people **attended** the event live. In March 2018, the Luxor Hotel in Las Vegas opened a 30,000-square foot permanent Esports arena that will hold future tournaments. Prize money in worldwide tournaments reached more than $100 million in 2017.

RAY TRACING

In March 2018, NVDA announced the creation of a ray-tracing technology, which brings real-time cinematic quality rendering (a photo-realistic image) to movie makers, content creators and video game developers and is considered the biggest breakthrough in computer graphics in many years. Until now, lighting and shadows has been introduced to content and games through programmable shading; in ray tracing, real-time computations replicate the way light and shadows truly behave in real life. In practical terms, about one billion frames of film are painstakingly rendered each year to ensure its cinematic-quality production. The rendering time for each frame sometimes takes hours. NVDA's new technology can eliminate these hours.

In addition, new video games will be more realistic, showing nooks and crevices that previously could not be displayed. A further benefit to NVDA: each year new video games seem to require a faster and more powerful GPU. With ray tracing, this need will be further accelerated.

AUTOMOTIVE

	Revenue	Year-Over-Over Revenue Growth	Gross Margin
FY 2018 (Jan.)	$558 million	15%	Considerably less than 60.2%
FY 2017	487 million	52%	N/A
FY 2016	320 million		N/A

NVDA has been at the forefront of autonomous driving, providing AI platforms powered by its GPU chips. Indeed, according to an August 2017 piece in Nikkei Asian Review, NVDA offers what are considered the best chips to serve as AI engines in self-driving cars. The company has partnered with 370 auto manufacturers, suppliers, mapping companies, and research organizations that are developing the algorithms, including Uber, Tesla, Audi, Mercedes-Benz, and Bosch.

While NVDA has entered into so many partnership, the revenue build has so far been relatively slow; its Automotive unit, which until recently reflected primarily "infotainment" revenues as opposed to the self-driving initiative, recorded sales of $558 million in FY2018, up (a pedestrian, for NVDA) 15%. The total addressable self-driving market promises to be huge; some industry estimates peg it to reach $40 billion or more in twenty years. By 2035, 100 million self-driving cars plus 10 million robo-taxis may be sold on an annualized basis. The first sale of a self-driving car is probably two to three years away.

Setbacks will doubtless happen as the technology is refined; witness the fatal Uber accident in Arizona in March 2018. In light of the accident, NVDA has temporarily suspended its self-driving car tests. It is important to note that Uber did not utilize NVDA's DRIVE PX autonomous driving platform (which collects data through real-time sensors, provides AI training of various car makes and models, provides simulations, and driving technology, among other functions). Instead, NVDA sold the company GPU chips which Uber employed in its own sensor and driving system.

Growing worldwide interest and advances in autonomous driving could have a tremendous knock-on effect on NVDA's AI business. For example, a successful self-driving vehicle will likely have multiple deep learning networks that the car maker utilizes. Each network will have to be backed up by a supercomputer (perhaps a NVDA DGX-2 which has a list price of $399,000) that holds millions and millions of images. To go a step further, each company working on autonomous driving will need to purchase a number of supercomputers.

VIABLE COMPETITIVE THREATS TO NVDA'S DOMINANCE IN AI AND THE GAMING SPACE? POSSIBLE BUT NOT LIKELY FOR SOME TIME.

China

In order to decrease its reliance on foreign-made chips for military and other applications, China in July 2017 articulated a goal to become the U.S.'s equal in AI technology within three years and to become the world leader by 2030. A key foundation of this plan is the country's commitment to develop semiconductor technology which can efficiently run artificial neural networks. In particular, the Ministry of Science and Technology's stated it wants to create a chip benchmark that delivers performance and energy efficiency 20x better than NVDA's three-year old Tesla M40 chip. The M40 is used for AI machine learning applications in data centers but is not NVDA's top-of-the line product.

However, China is considered multiple generations behind NVDA's technology. Furthermore and perhaps surprisingly given the Ministry's statements, few of China's chip companies are directly attacking NVDA's core market for servers. Instead, they are focusing on cell phone technologies. Also, the Chinese technology stalwarts, Alibaba, Baidu, and Tencent have announced AI agreements with NVDA.

Google's Tensor Processing Unit (TPU)

GOOG has been using an internally developed custom processor, termed a TPU, to accelerate its Deep Learning applications. The TPU is believed to be used primarily for inference; whereas, NVDA's V100 GPU has been primarily used for training (so far). GOOG does not plan to sell the TPU but to make access available on an hourly basis in the Google Cloud. (Obviously, this would differentiate the Google Cloud Platform from Amazon Web Services and Microsoft Azure.)

GOOG has said that the TPU is faster than others' CPUs and GPUs and has explicitly noted performance improvements over NVDA's technology (although NVDA does not agree with this interpretation). Note that a TPU is an ASIC built solely for AI. GPUs can not only facilitate AI, but also function in data center hardware or, say, in the transcoding of video from a format used in broadcast television to one used for streaming.

A few implications of GOOG's TPU technology for NVDA:

- The challenge that NVDA faces seems confined to AI inference work performed in the cloud. In addition, NVDA has a far-reaching position in the Deep Learning ecosystem of hardware, software and leading AI scientists and researchers.

- NVDA could choose to develop a new, presumably powerful GPU built specifically for inferencing to counter the competitive threat. Also, NVDA is almost certain to improve the efficiencies of its products. The company has achieved efficiency gains of 25x over the past five years in V100 data center GPU (see page 12).

- A longer-term concern is that new chip start-up companies like Nervana, Knupath and Movidius could change the set up of neural networks such that new chips are used which perform only one task very well.

Intel

In November 2017, Intel hired a high-level executive from AMD to spearhead its GPU development efforts. In February 2018, INTC announced a GPU prototype which is only about as powerful as one of NVDA's lowest-end graphics cards. In particular, the prototype contains 1.5 billion transistors, nearly 90% fewer than in NVDA's current high-end GPU for gamers. It is of course possible that INTC could build a more powerful GPU product in the future, but the company does not seem to pose a near or intermediate term threat in the GPU high-end graphics or AI markets.

NVIDIA VALUATION

The attached spreadsheet shows detailed income statement, cash flow and balance sheet projections for NVDA. Some of the key assumptions (bolded in red in the spreadsheet) are as follows:

- Driven by both GPU sales revenue and non-GPU gaming revenue (see page 15), NVDA's Gaming segment is projected to show sales growth of 32.5% in FY 2019 and 25% in FY 2020. While rapid, these increases are slower than the 36% and 44% figures recorded in FY 2018 and FY 2017, respectively. Keep in mind that an updated (and likely impressive) series of graphics gaming cards should be announced this year. NVDA's new ray tracing technology should contribute to growth as well. An offset: the potential reduction in sales to cryptocurrency miners in FY 2019.

- Unrelenting AI demand growth, coupled with acceleration opportunities in both data centers and supercomputer sites and new products (the DGX-2 supercomputer), should allow Datacenter revenues to increase 85% in FY 2019 and a further 40% in FY 2020. While red hot, such a growth pace would be markedly slower than that realized in FY 2018 and FY 2017.

- NVDA's self-driving initiative is not expected to have a large effect on results in FY 2019 and FY 2020, but could provide a huge boost in subsequent years.

- NVDA's overall gross margin is projected to remain constant at 60% over the next two years.

Factoring all this in, NVDA's EPS should reach about $6.45 in FY 2019 and $8.25 in FY 2020 – above consensus forecasts, particularly in FY 2020. Similarly, revenues should increase 38% and 27% in FY 2019 and FY 2020, respectively.

If NVDA were to achieve forecasted FY 2019 results, NVDA currently trades at a 35x multiple of FY 2019 earnings, 28.6 times EV-to-FY 2019E EBITDA and 10 times EV-to-FY 2019E revenue.

In the Comp's tab in the attached spreadsheet, NVDA's key valuation parameters are compared to other high-growth tech companies – NFLX, BABA and CRM. AMD is also displayed for completeness. NVDA's "current year" valuation parameters are more or less in line with these peer companies but are still on the high side of the Comp's averages.

On the basis of estimated FY 2020 earnings (essentially calendar 2019), NVDA is more reasonably priced at about a 27x P/E. Revenue and EBITDA multiples based on FY 2020 projections remain high.

In my opinion, the likelihood of above-consensus earnings over the next few years, especially in FY 2020, coupled with a fairly friendly FY 2020 P/E multiple, suggests that NVDA is a good long candidate, particularly after the recent 10% sell-off. Note that NVDA plays a central role in three fast-growing technologies – AI generally, driverless cars and the relentlessly-growing video gaming business. Indeed, NVDA may be the best horizontal play in AI, as its GPU chips seem likely to play a role in almost every AI initiative and the chips already have a huge (and hard to replace) ecosystem built around them.