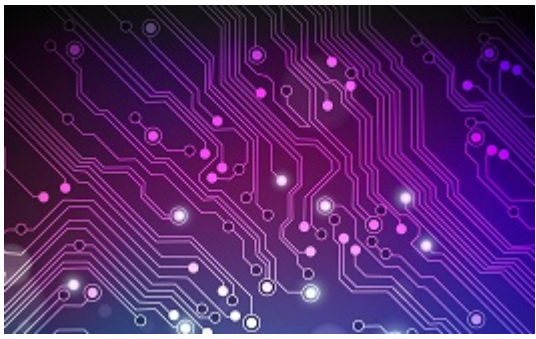
Tractica





RESEARCH REPORT

Deep Learning

Enterprise, Consumer, and Government Applications for Deep Learning Software, Hardware, and Services: Market Analysis and Forecasts for 112 Use Cases

Published 2Q 2017

JESSICA GROOPMAN
Principal Analyst

ADITYA KAUL
Research Director

SECTION 1

EXECUTIVE SUMMARY

1.1 INTRODUCTION

The sheer volume, variety, and velocity of data generated every day renders *learning* (from that data) the principal objective in order to justify investment of digitizing in the first place. Significant improvements in hardware speed and machine learning algorithms are opening up new capabilities and applications for artificial intelligence (AI), most notably in the area of deep learning. The fundamental business question of "what do we do with all of this data?" is colliding with our endless fascination for technological biomimicry of human intelligence.

Deep learning is a computing construct based loosely on the architecture of the human brain. A subset of machine learning, it is distinct in that it is composed of multi-layer ("deep") neural networks wherein each layer of the network corresponds to different levels of abstraction. When exposed to vast amounts of data, deep learning systems develop basic pattern recognition, enabling algorithms *to train themselves* to perform tasks and adapt to new data. What differentiates deep learning from other machine learning is its ability to infer outcomes without explicit instructions, instead drawing from patterns within the data. This is why deep learning has emerged as one of the most promising, if controversial, enabling technologies in the world of AI.

Part of the intrigue and difficulty of understanding deep learning is that it is application-agnostic. Deep learning can be applied, and is particularly well-suited to autonomously extracting nuance from any (structured or unstructured) data set large enough to identify statistically significant patterns. Many deep learning applications in development today center around incremental, if very practical, advancements in areas such as image recognition, text analysis, product recommendations, fraud-prevention, and content curation. But these are merely the tip of the proverbial AI iceberg.

Innovations within the deep learning space today are paving the way for more disruptive applications of tomorrow, such as driverless cars, personalized education, preventative healthcare, and many other capabilities previously reserved for science fiction. In future scenarios, technologists see no reason why deep learning would not be applied to software development itself, wherein deep learning systems iteratively code and design products and services themselves, perhaps beyond the realm of current human imagination or engineering capabilities.

Its potential is what has attracted wallet share, and even entire strategic pivots from many of the most powerful technology companies in the world; current leaders in the deep market include Google, Facebook, Microsoft, IBM, Amazon, Baidu, and others. The Chief Executive Officer (CEO) of Google, Sundar Pichai, recently described deep learning as "a core, transformative way by which we're rethinking how we're doing everything."

Tractica research estimates deep learning to be the largest technology category in terms of revenue. Our research into AI has unearthed well over 100 use case categories in which deep learning constitutes a majority. Deep learning is often applied in conjunction with other technologies, such as computer vision, natural language processing (NLP), sensors, and others. Based on our research and forecasting, Tractica believes the opportunity for deep learning spans a wide range of industries and geographies and is particularly disruptive in highly domain-specific markets with high-volume data needs and ontologies, and those with

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

growing applications for machine perception. Deep learning revenue is estimated to grow from $654.9 million in 2016 to $34.9 billion by 2025, which will represent 57% of the overall AI market in 2025.

What are the use cases for which current processes can be replaced with a deep learning-based approach because it is more efficient, comprehensive, cheaper, or smarter than the current model? Deep learning faces a number of fundamental hurdles to adoption, which will be outlined in this report. It also introduces new opportunities for improving accuracy, efficiency, costs, and streamlining workflows, and more empirical decision-making.

1.2 Market Drivers

Although many of the concepts underlying AI and technological biomimicry of human intelligence are well over 50 years old, deep learning's growth today is the result of a rather sudden convergence of three key trends: big, even colossal data generation; advancements in hardware speed; and improvements in algorithms. This has accelerated research efforts across academia, as well as startups and enterprises.

Enterprise interest in the space is also driven by a number of benefits and efficiencies that commercial application of the technology enables. Tractica's analysis also includes a detailed assessment of business model impacts in both cost and resource efficiency gains, as well as new incremental revenue generation. To illustrate these potential business model impacts, Tractica assesses the benefits of deep learning in its ability to drive efficiencies in the form of speed, accuracy, agility, and access in the following areas:

- Product development and improvement
- Process optimization and functional workflows
- Personalization and customer insight
- Sales optimization
- Innovation and longer-term strategy

Furthermore, deep learning offers a number of potential benefits to society. Tractica's research explores both the benefits and the risks of AI and deep learning in society, exploring areas such as energy conservation, safety, public health, and others.

Taken in sum, these drivers have turned the heads of almost every large technology company, enterprise adopters across industries, herds of investors, countless ".ai" startups, and even governments and policymakers.

1.3 Market Barriers

Despite the array of benefits deep learning could enable, the technology faces numerous barriers to achieving widespread or enterprise-scale adoption. AI, in general, suffers a significant gap between high expectations of intelligence and the reality of current limitations of software and computing. In fact, understanding the limitations of the artificially "intelligent" technology may be the most instructive step toward accurately grasping its utility. This research explores the following areas that threaten the adoption of deep learning and AI in commercial environments:

- Lack of accessibility and simplicity
- Trust issues, including employee and consumer sensitivities

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- Ethical risks and unintended consequences
- Job displacement and the transformation of work
- Business challenges to obtaining data
- Significant shortages in talent
- Challenges of scale
- Opacity in "explainability"

1.4 Technology Issues

This report provides a high-level examination of deep learning architecture and technological considerations. Within these structures, Tractica explores a number of technical questions enterprises face:

- Defining what deep learning is (and is not)
- Contextualizing deep learning within broader machine learning and AI
- Criteria for deep learning application
- Application in conjunction with other technologies
- Technical challenges of obtaining proper training data
- Training and supervising deep learning models
- Various hardware, firmware, and software configurations of deep learning

Deep learning is an enabling technology, for certain applications and often used in conjunction with other technologies and techniques. It is not a panacea for all information technology (IT) [in]efficiency. The technology is still in its infancy in both architecture and application, and is characterized by its open-source development. Although most machine and deep learning take place today as software applications in the cloud, advancements in firmware-level machine learning and maturity in edge processing are shifting the narratives around data storage requirements, architectural development, and market access. These developments are outlined in this report as well and explored in greater depth in Tractica's report on *Deep Learning Chipsets*.

1.5 Use Cases for Deep Learning

Tractica's research finds that use cases for deep learning span a wide range of more than 100 distinct use cases, touching at least 28 distinct industries. Some of the most common applications our research surfaced across multiple industries include:

- Static image recognition, classification, and tagging
- Object detection, including navigation
- Predictive maintenance
- Trend identification and prediction (e.g., weather, demand, fraud, etc.)
- Sensor data fusion

Within each of the use cases Tractica identified, countless sector-specific variations exist. As with most technological innovations, different industries will adopt pieces of this technology at varying paces. The qualitative portion of this report defines the top 52 use

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

cases for deep learning based on revenue, investment, and market activity. The quantitative forecast model accounts for all 112 use cases Tractica has identified in its ongoing coverage of the AI market.

1.6 MARKET FORECAST

Tractica forecasts that annual software revenue for deep learning applications will increase from $655 million worldwide in 2016 to $34.9 billion in 2025, representing a compound annual growth rate (CAGR) of 56%. Total annual revenue for deep learning software, services, and hardware will increase from $3.3 billion in 2016 to $160 billion in 2025, at a CAGR of 54%. According to our analysis of six distinct technology segments of AI, deep learning accounts for the largest segment, representing 48% of the overall AI revenue in 2016 growing to 57% of the revenue by 2025.

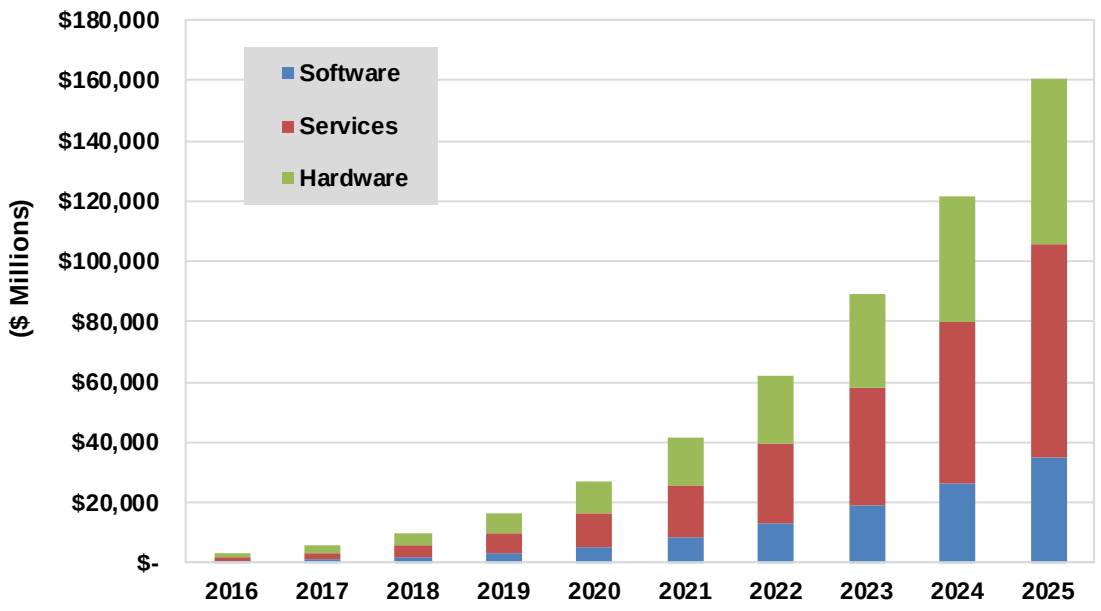
This report forecasts revenue across five world regions, 28 different industry verticals, and more than 100 use cases across enterprise, consumer, and defense markets. Some of these use cases include static image tagging, localization and mapping, predictive maintenance, and human emotion analysis, among many others.

Tractica also forecasts hardware and services revenue driven by deep learning. The hardware forecasts are further segmented into separate forecasts of the demand for central processing units (CPUs); field-programmable gate arrays (FPGAs), application-specific integrated circuits (ASICs); graphics processing units (GPUs); networking products, and data storage devices.

Deep learning applications will also generate significant demand for professional services, such as installation, training, customization, application integration, and maintenance, which are included in Tractica's forecast. In addition to this market forecast, this Tractica report offers a comprehensive analysis and overview of the entire market opportunity and challenges for deep learning products and services.

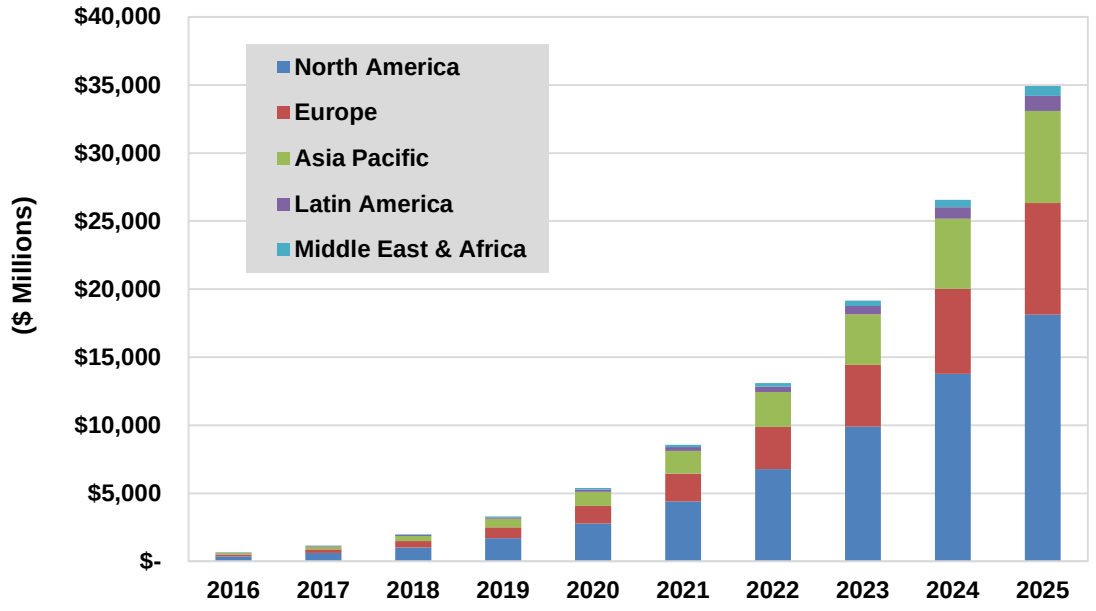
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Chart 1.1 Deep Learning Total Revenue by Segment, World Markets: 2016-2025



(Source: Tractica)

Chart 1.2 Deep Learning Software Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 2

MARKET ISSUES

2.1 MARKET DRIVERS

The primary driver behind the adoption of deep learning is an unprecedented amount of data caused by the digitization of society. Although many of the concepts underlying AI and technological biomimicry of human intelligence are well over 50 years old, deep learning's growth today is the result of a rather sudden convergence of three key trends, as well as a number of benefits and efficiencies that commercial application of the technology enables.

2.1.1 THE DIGITIZATION OF SOCIETY AND DATA AT SCALE

The digitization of society has been underway for decades, but as costs of computing have decreased, access to componentry, devices, and networked services has skyrocketed, for consumers and enterprises worldwide. This explosion of access is responsible for exponential increases in the production, utilization, sharing, and [potential] monetization of data. As more people, processes, institutions, and infrastructure come online, the result is more data across more applications, distributed more widely.

The amount of data that exists already today, and that we expect to be generated in the future is hard to fathom. As technologists have been saying for the last 5 years, it continues to be true that we have created more data over the previous 2 years than in all of human history combined. At the World Economic Forum in 2017, new data points illustrating our voracious data appetites show that *every minute* of every day:

- Americans alone use 18,264,840 megabytes of wireless data
- Dropbox users upload 833,333 files
- The Weather Channel receives 13,888,889 forecast requests
- YouTube visitors share 400 hours of new video
- Skype users connect for 23,300 hours

And this does not even scratch the surface of industrial applications, such as heavy machinery manufacturing, aerospace, agriculture, logistics, and transportation. In 1992, global internet traffic accounted for 100 gigabytes (GB) per day; In 2015, that number hit 15 billion GB per day. The digital universe is doubling in size every 12 months. By 2020, it is expected to reach some 44 zetabytes; some scientists estimate that number to be more bytes in the digital universe than there are stars in the physical universe.

Of course, digitization is not just a function of growing volume, but of velocity and variety. These characteristics of Big Data have a uniquely complementary relationship to deep learning: deep learning models require massive amounts of data to successfully train themselves; statistically significant sample sizes of vast ontologies and scenarios. As AI pioneer Andrew Ng famously puts it, "deep learning is like a rocket engine, and data is like rocket fuel. Without enough fuel, it is difficult to get off the ground, no matter how big the engine." But it is not just deep learning that needs data, the opposite is true as well: Big Data lakes and/or high velocity and variety data generation are currently only as useful/usable as their ability to be rapidly processed, analyzed, and acted on.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

2.1.1.1 Emerging Hardware and Interfaces Accelerate Digitization

The velocity of our consumption and demand is also, increasingly, enabled through a series of technologies, some of which are described below:

- **Processing Speed:** In the last 5 years, the tech industry has seen a few critical advancements in processing speed, or the time needed to process X amount of data. The more data that can processed in [near] real-time, the more sophisticated and varied the potential applications for learning from that data and the stickier applications are from an adoption standpoint. Refer to Section 3.4.2 for a deeper examination of advances in processing speed.

- **IoT and M2M:** For the last 30 years, telematics and sensor-enabled automation have been making their way into industrial sectors, but more recent advancements in network connectivity and cloud technologies are driving a rapid uptick in networked services via connected devices and infrastructure. From smart home to automotive, retail, and everything in between, the Internet of Things (IoT) is demanding programming intelligence like deep learning to support and sustain its business models.

- **Mobile Interfaces:** The super-computing phones that live in the pockets (and influence behavioral shifts) of some 5 billion users are another macro force influencing AI adoption. As the organizations must not only be, but must *learn from* where their customers are, mobile as a primary interface means it is important to the adoption, consumption, and monitoring of AI-driven services. Speaking of interface, emerging modes of technology interaction, such as voice or motion recognition, rely heavily on deep learning for optimization and an improved user experience (UX).

- **Adjacent AI Technologies:** AI is best understood as an umbrella category containing at least six subcategories with different capacities for different circumstances. Tractica identifies these categories as deep learning, machine learning, NLP, computer vision, machine reasoning, and strong AI. Among these, deep learning is already being combined with varying configurations of these, most often NLP and computer vision. This trend and the increase of different configurations will continue to define and redefine deep learning and its applicability across industries.

Each of these technologies generates massive variety in data itself; structured, unstructured, text, sensor data, audio, video, clickstreams, images, log files, etc. The variety of data flowing outward from our digitizing all aspects and elements of society must be processed. Broadly speaking, what the growing experimentation and adoption of deep learning, machine learning, and AI reflect is the inevitable shift in emphasis and maturity from data collection to analysis, and most importantly, toward decision-making.

2.1.2 Computation at Scale

Data at scale is scarcely possible, nor particularly insightful without computation at scale. Exponential advances in hardware speed have helped lift AI, especially deep learning, out of its final “winter,” and as such, it is helping deep learning emerge and move from academic experimentation to enterprise applications.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

2.1.2.1 Exponential Advancements in Hardware Speed

The most notable advancement in computation at scale is significant improvements in hardware speed, in GPUs, and beyond. For decades, and still today, most computers run using a CPU. CPUs run based on registers where instructions are executed, numbers manipulated, and memory addresses are maintained. CPUs process these instructions one at a time. While some computer models have several CPUs and multi-core processor chips are becoming more standard, the majority of computers today have a single CPU.

Parallel processing is a more robust form of processing in which instructions are divided among multiple processors in order accelerate the time it takes to run a program because there are more processing units working on the same code. While it was historically difficult to divide a program across separate units executing different portions without interfering with each other, a development in 2009 led to a breakthrough in parallel processing. A group of Stanford University researchers, including Andrew Ng at the time, discovered that a cluster of a certain kind of chip used in video gaming was also able to support the parallel processing of neural networks. The real breakthrough of this discovery was speed; the team found that a cluster of GPUs could accomplish the same task in a day that took weeks using traditional methods of neural net processing.

The inherent parallel design accounts for significantly faster processing of applications that need to process large blocks of data simultaneously. Researchers have found that 12 GPUs in a 3-machine cluster can rival the performance of the 1,000-node CPU cluster.

GPUs are defined as “a single chip processor with integrated transform, lighting, triangle setup/clipping, and rendering engines that is capable of processing a minimum of 10 million polygons per second simultaneously,” according to NVIDIA, the company that first developed the chips and invented the term. Developed initially to offload computationally-intense image processing tasks from CPUs, and thereby expand and improve gaming experiences, these chips allow augmented aesthetics without sacrificing overall system performance. For a deeper discussion on firmware and semiconductor innovations driving deep learning, reference Section 3.4.2.

2.1.2.2 Significant Improvements in Algorithms

The rise of deep learning in the last few years is not solely a function of increases in data and hardware speed, but can also be attributed to significant improvements in machine learning algorithms. To understand what came before and why developments since have introduced improvement, it is important to understand the idea of algorithm efficiency and iteration, and their impact on resources required to compute. Unlike previous “brute force” techniques, wherein all possible scenarios are systematically calculating all possible solutions, comparing results, and determining which solution is best based on those results, deep learning’s architecture “learns” through a sort of trial and error approach. While this method is logical, the model’s ability to learn through iteration quickly becomes untenable as the time and power required to calculate outcomes expand exponentially with every additional or alternative scenario.

In a deep learning construct, iteration occurs in the succession of approximations, each set of which is built based on the previous layer, enabling a far greater degree of abstraction, nuance, and accuracy. Furthermore, neural networks run probabilistic calculations to approximate the answer with a percentage degree of certainty.

Based initially on research by Professor Geoffrey Hinton, of the University of Toronto, his team’s research showed that each layer of neural nets could be pre-trained, which would

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

accelerate subsequent supervised learning. As a result, today's neural networks are able to leverage training methods that pre-train individual layers of the network to recognize features at different layers, which makes it easier and more efficient to train networks with several layers. Minimizing the number of machine instructions required to iterate a layered solution has been essential to the growth of deep learning in the last few years, as hardware speed and data volumes are now available to support neural networks at far greater scale and speed.

2.1.3 Efficiency Gains and Accuracy Improvements

The most powerful aspect of deep learning is that when given a set of data, it can learn the features (and, in some instances, rules) of that data without being told what the features (or rules) are. Deep learning holds the greatest promise in its ability to deliver or at least streamline efficiency. Many point to AI and deep learning as equally profound in efficiency gains as the internet or mobile have been. This time, instead of replacing interface, AI replaces (or at least can significantly automate) decision-making and improves over time, with increasing autonomy, with more data.

Tractica's research finds application for deep learning in every function of the enterprise, including sales, marketing, customer service, supply chain, and operations automation and prediction, as well as wielding diverse data sets for legal, finance, human resources (HR), etc. Perhaps no function is more impacted than the product itself; deep learning's impact on product management, research, and development will transform design cycles, product roadmaps, and customer experiences.

Taken in broader context, deep learning helps drive a much greater goal of platform optimization in that, theoretically, it fuels ongoing improvement based on each phase of the data lifecycle wherein the network can inform itself. More users create more data, which feeds smarter algorithms, which improve product and user experience, which repeats the cycle. Put simply: data-driven product → more data → smarter algorithms → better insights → more profit → more investment → better product → more users → more data.

Figure 2.1 The Virtuous Cycle of Intelligent Algorithms

(Source: Tractica)

Efficiencies enabled through deep learning run the gamut, including but not limited to the following areas:

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- **Speed:** Deep learning allows organizations to handle various tasks at scale far more efficiently than possible in the past. Rapid processing, parsing, curation, analysis, and presentation of text, images, language, gaming, and time-series data have historically been very cost-prohibitive and inaccessible to most. With the advent of deep learning and adequate hardware speed and data for training, processing such information becomes a question of developing and refining a computing model, not of hiring droves of employees to tag content or worse, remaining in the dark with unstructured data. The more information one can digest at scale, the faster one can learn. Deep learning also reduces the need for feature engineering, a process that typically takes months of engineering work to achieve.

- **Agility:** Deep learning is often an architecture that can be adapted to new problems relatively easily. Using techniques like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory, applications around vision, time series, language, etc. can be shaped and re-shaped to meet changing circumstances and contexts.

- **Accuracy:** When layers are able to learn from each other and apply learnings to the next layers, such cumulative learning can drive significant improvements to accuracy. Not only do networks take less time to train, they get better and smarter over time as new data is introduced. Qualitatively, such learning enables abstraction and nuance detection at a much higher level than was ever possible for computers. From a quantitative perspective, models offer statistically-generated percentages of certainty (e.g., the model is 93.4% certain this image is a cat; or this piece of content will resolve the customer's issue).

- **Access:** While technology giants like Google and Microsoft are leading adoption of deep learning today, the openness of the research community and ubiquity of software as a service (SaaS) solutions are slowly increasing access for this technology to mid-size and smaller enterprises. Salesforce.com's acquisition of MetaMind is geared toward "baking in" certain machine and deep learning capabilities to each cloud of its customer relationship management (CRM) suite, for whom a large customer segment is small to medium-sized businesses (SMBs).

- **Process Optimization:** Deep learning can help drive intelligence at each level of an organization. From sales, marketing, support, and product to finance, HR, operations, logistics, legal, and beyond, the ability to augment tasks has use cases across the enterprise. As companies work to integrate these processes and their respective insights, this technology can help accelerate efforts toward the broader Big Data holy grail: learning from internal and external information as rapidly as possible to create as efficient and intelligent an IT/operational technology (OT) architecture as possible.

- **Personalization and Customer Insight:** Using deep learning technologies, organizations can target individuals at a hyper-personalized level, based on learnings not only from diverse data sets, but from the model's ability to detect patterns and automate decisions based on those patterns. Although organizations should exercise significant discretion around sensitive data, privacy concerns, and customer protections, the opportunity is to use insights to craft an experience tailored precisely (and even predictively) to what each user desires.

- **Product Improvement:** While not unique to deep learning, the ability to learn from product interactions and content begets greater insights from which to improve.

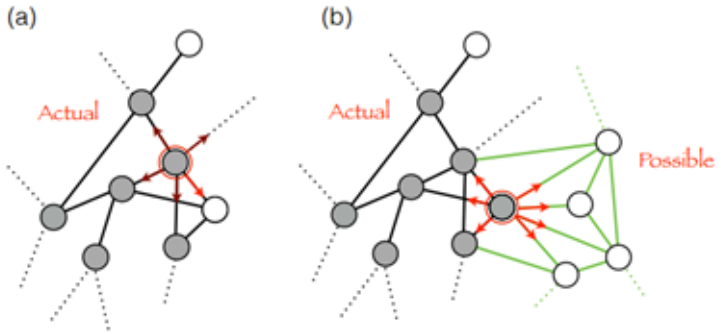
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Where deep learning enhances this is in the ability to improve the model for learning and the content and interactions that ensue. Companies have used deep learning to significantly enhance voice recognition and e-commerce recommendations. These improvements drive accuracy and a better customer experience (CX) for personalized end-user interactions, while also feeding research and development (R&D) for the broader product.

- **Sales Optimization:** By analyzing and modeling patterns across sales, customer, product, operations, market, and other data sets, neural nets can help drive predictive sales activities (i.e., predictive lead scoring, preemptive product failure alerts, predictive support, etc.).

Deep learning offers a wide range of potential benefits to enterprises, most of which can be applied to multiple business functions, as well as partners and customers. Yet, it also signals potential for an even deeper level of intelligence gathering over time. What if organizations could not only leverage their own device failure information, but that of all (similar) device failures? In certain industries like energy or oil and gas, equipment sequences are chained together so that if one device fails, the entire production may be threatened, costing companies millions of dollars in downtime and its cascading effects. Effectively handling this issue is not solely about enabling predictive failure for one device, but about collecting data of different kinds of failures across different devices. In such a scenario, deep learning could support a sort of shared learning platform in which multiple companies collaborate to share their device data from the field. The result of such a platform would mean that those companies not sharing their data are not able to learn from others' experiences; a significant disadvantage in such a scenario.

Figure 2.2 Massachusetts Institute of Technology's Mathematical Models Reveal Patterns for How Innovation Arises



(Source: Massachusetts Institute of Technology)

The Massachusetts Institute of Technology (MIT) recently published research that mathematically depicts this potential for innovation. As more data are shared, collective vocabulary emerges and expands, which develops a language. As new ways of expressing possibilities are enabled, new constructs for language (the foundation for transmissible human knowledge) and greater innovation are made possible.

Broader learning, across devices, industries, and contexts is one of the essential building blocks to deeper benefits gained in ubiquitous connectivity and the Internet of Everything. Consider the potential for aggregate (identity-protected) data analysis in healthcare, agriculture, science, academia, and beyond.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

2.1.4 Societal Benefits

Every new technology abounds with a kaleidoscope of hopes and fears surrounding its impact. Most of the time, humans are bound by a current myopic view of present structures and struggle to accurately extrapolate long-term implications, negative or positive. AI and deep learning, in particular, are unique in their (at least theoretical) semblance to human capacities. Whether or not this equivalency is false is less important than the bar it sets for humans to consider it successful. While our demands to achieve human-level discretion, accuracy, meaning, and autonomy leave room for a great number of trust and security issues (reference Section 2.2.2), it is worth noting that this technology may also engender certain societal benefits. Tractica's research identified the following areas as having the potential for a positive impact of deep learning:

- **Energy Efficiency and Conservation:** Deep learning can be used to detect areas of energy inefficiency, redundancy, and re-allocation of energy. Today's function of monitoring sensors and actuators could be tomorrow's models based on years of sensor, time-series, weather, and other data that learn and support a highly dynamic energy utilization schema. Google famously used DeepMind to achieve a 40X improvement in energy expended to cool its data centers. This equated to a 15% reduction in overhead for power usage effectiveness.

  Deep learning is also being applied to problems associated with conservation. For example, online data science platform Kaggle has partnered with the whale researchers at the National Oceanic and Atmospheric Administration (NOAA) to work on facial recognition applications to detect and monitor certain endangered species of whales. Leveraging deep learning to drive greater efficiency and conservation is a critical use case, if not for the rising temperatures of the planet and sustenance of life on earth, then for the desperate need to more efficiently scale computing structures worldwide.

- **Safety and Security:** Deep learning could help identify anomalies and signal preemptive action to prevent the failure of surgical devices, mission-critical machinery, power-generating technology, or any other technology that could impact end-user safety. Deep learning tools could also enhance areas pertaining to security, both digital and physical. Learning diverse details of current threats, integrating safety precautions into workflows, and using the model to identify new threat vectors are use cases targeted by startup Deep Instinct, among a growing number of companies.

- **Quality of Work:** Although many are understandably concerned about the prospect of mass job elimination by way of bots, another possibility may be simultaneously true. Could software automation improve the quality and experience of working? In some cases, work may become less manual and tedious and more mentally and creatively engaging; in other cases like customer service, which is less repetitive and more case-specific, it could mean fewer angry customers and more rapidly routed, information-equipped ones. Automation has been targeting repetitive jobs that, unfairly, have been associated with lower intelligence of the uneducated classes. Could machines help elevate such workforces to gain skills that are more transferrable, offering them more mobility than currently possible when focused on single, repetitive tasks? If AI is viewed as "lowering the cost of thinking," perhaps it will demand that society upgrades its collective structures to support greater intellectual cultivation for all, not just those who can afford it.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- **Education:** Education is an industry both ripe for disruption and greater scale. In a world building toward hyper-personalization and real-time demands, the existing education model of curricula at scale, didactic teaching, and standardized learning and testing methods grows more obsolete by the day. Deep learning is being applied to language and visual applications that help tailor learning methods, content, and interactions at the individual level. Already, companies are providing "intelligent" tutors for children, diagnostic training for service providers, and augmented reality (AR)-based training for mechanics and repair technicians. When coupled with mobile, this has the potential to increase engagement and success in current education programs at every level. In addition, it can open up or enhance new avenues (i.e., means to jobs, degrees, and mobility) for millions of people worldwide who lack access to quality education.

- **Social Input:** While the negative consequences of humans' increased interactions with bots (especially when compounded with decreased interactions with other humans) are unsettling and probably not fully comprehensible, there are glimmers of benefit in certain cases. Humans are extremely social creatures. Scientific research has repeatedly emphasized the negative impacts of loneliness, such as cognitive decline and increased risks of heart disease. For many, perhaps most especially the elderly and/or disabled, social interactions, intellectual stimulation, and general companionship can drastically improve quality (and length) of life.

- **Recognition at Scale:** Humans far outcompete computers when it comes to recognizing social cues, understanding cultural nuances, and exercising discretion, but there are other areas that are simply too complex and too vast for our brains to handle. Detecting trends based on patterns occurring across multiple (different types of) data sets and learning from millions of inputs in a formulaic way and retaining learnings without memory loss are prime examples of environments in which deep learning excels. It will augment humans' ability to better understand, predict, and take action in the world.

- **Medical Breakthroughs:** Although we are still in the early days of deep learning application in healthcare, developments underway hold incredible promise in areas like medical image analysis, disease detection, and treatment recommendation. Even today, researchers at the University of Ohio are using deep neural networks (DNNs) and machine learning to transform the efficacy of the hearing aid by enhancing sound segregation. The team identified and used 85 unique attributes to train the model to distinguish speech from noise. Early results show a boost in speech comprehension in noisy environments from 10% to 90%. Longer term, the ability to combine, analyze, and learn from disparate (often unstructured) data sets could accelerate drug development, forecast patient illnesses, and even develop highly personalized virtual assistants for patients and doctors.

2.1.5 Enterprise Giants Progressing Rapidly

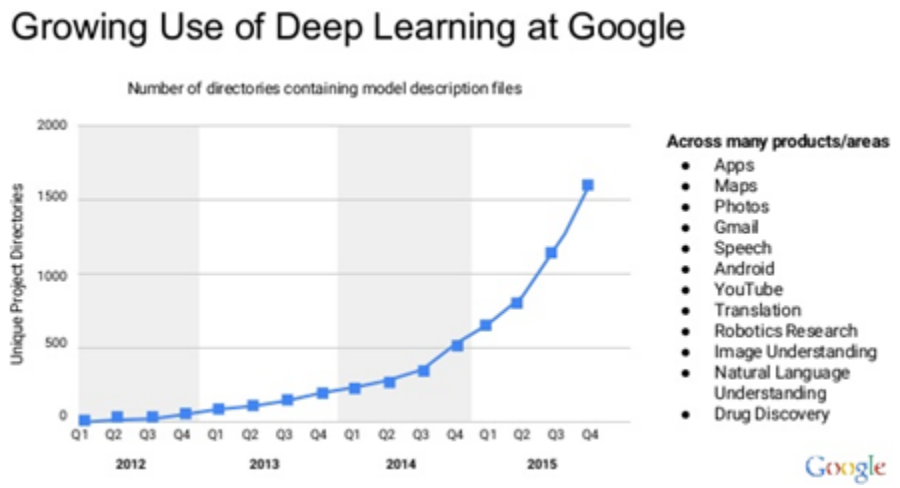
The broader AI space is one whose momentum is driven heavily by the technology giants of the world. Although Google, Facebook, Microsoft, Amazon, Baidu, Tencent, IBM, and NVIDIA are working with startups and poaching talent right out of academia (and from each other), the reality is the sheer data volumes these companies have, rendering them trailblazers in the enterprise market for deep learning. Another reason for these companies' dominance in deep learning is due to the scale of their high-performance computing (HPC) resources and the resources they can devote to training DNN algorithms. Given that DNN performance is directly proportional to the number of layers in the network, extremely robust

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

hardware is required to process and train these algorithms at scale. This hardware has historically been out of reach for smaller organizations.

Companies like Google, Baidu, Tencent, and Facebook today have thousands of AI apps in production and running at scale, compared to just 5 years ago, and they are developing, recruiting, open-sourcing, and deploying rapidly. Google alone has been using deep learning in search for years now, never mind in merchant services, advertising, online services, enterprise solutions, Google cloud services, and the Android platform. The company has scaled Google Brain from about three employees in 2011 to hundreds and growing today.

Figure 2.3 Growing Use of Deep Learning at Google



(Source: Google)

Facebook recently announced that it was training 2 trillion examples every day and over 1.2 million models are trained each month. Facebook has been working on image recognition for years now and invests in a range of open-source application programming interface (API) projects supporting diverse use cases such as disease diagnostics in healthcare and computer vision enabling remote internet infrastructure development. It also has an entire dedicated wing for AI research (Facebook AI Research, or FAIR), and is making significant contributions (e.g., frameworks, code, best practices, etc.) to the broader AI community.

Amazon is a pioneer in product recommendation and e-commerce machine learning. But the company is betting on AI much more strategically with its launch of the voice-controlled Amazon Echo, alongside its machine learning component to Amazon Web Services (AWS). Both have made significant impacts in the smart home and cloud markets, respectively.

It is fair to say that IBM has been dabbling in AI since its origins in the 1950s. One of the most primitive, but important examples of machine learning, a bot that could play checkers, was developed by IBMer Arthur Samuel. This laid the groundwork for IBM's famous DeepBlue chess-playing program, which beat Gary Kasparov, the world's chess grandmaster in 1997. The company's more recent development of Watson AI has more than 22 APIs available to developers and is increasingly a central platform for its strategy across all products.

Baidu just announced it had established China's first national engineering lab on AI and deep learning. Notably, the laboratory will be less of a physical structure and more of a digital network of researchers across the country, focusing specifically on human-computer

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

interaction, biometric identification, intellectual property rights, and computer vision.

Many of these companies have come out expressly stating that their strategies are no longer "mobile-first," but "AI-first" instead. Such a shift is required, as Google CEO Sundar Pichai put it, in order to support "a world where computing becomes universally available—be it at home, at work, in the car, or on the go—and interacting with all of these surfaces becomes much more natural, intuitive, and intelligent."

Learn more about each of these organizations' contributions and activities to the deep learning space in Section 6, Industry Participants.

2.1.6 THE DEEP LEARNING INDUSTRY IS COALESCING

While the advent of deep learning may feel sudden, the reality is that various forms (research, applications, frameworks) of AI have been around for decades. The hallmark of the latest wave of excitement and investment is marked by an emerging landscape of technology providers. Data is the essential fuel for developing deep learning applications, so companies investing in the technology tend to fall into three overall camps:

- **The Data Superrich:** These companies possess their own (massive) data sets and are using it to build their own neural networks. Such giants are few in number, but the sheer amounts of clean, structured data they have do give them a significant advantage over all others developing machine learning initiatives.

- **(Big) Data Service Providers:** These companies provide machine or deep learning services to other companies, which help them increase cleanliness, structure, and use of their own data. They require access to others' data sets to provide such services, and improve their own services.

- **Application Innovators:** These companies focus on very specific applications and must gain access to others' data to develop their models. Such companies are not service-related, but rather require data for product development. (e.g., technology for self-driving cars, stock market predictions, medical diagnostics, etc.).

2.1.7 COST DRIVERS AND BUSINESS MODEL IMPACTS

The whole point of deep learning is to turn data into actionable information. This potential lends itself to an infinite number of business model impacts, both cost savings and new revenue generation, by enabling fundamentally new context for business process workflows, service design, and decision making.

Although each use case and industry's financial, technological, and regulatory requisites and opportunities vary, cost reductions associated with deep learning tend to fall into the following categories:

Fewer People | Faster Decision-Making | Less Risk | Less Waste | Reduced Costs

- Reduced latency through automation of services (e.g., transfer of energy, data, documentation, money, etc.)
- Reduction in time and costs associated with data cleansing, processing, analysis, and empirically-based decision-making
- Reduction in labor required for reconciling errors, issues, claims; identifying and diagnosing problems; as well as tagging, scanning, and processing data manually (e.g.,

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

support, legal, healthcare, hospitality services)

- Reduction in logistical labor required for scanning goods, transport, oversight, etc. (e.g., truck drivers)
- Reduction in legal costs and fees associated with processing, investigating, and reconciling errors, issues, claims, fraud, and unmet compliance
- Reduction in costs associated with regulatory penalties
- Reduction in IT/OT security costs due to more robust and preemptive threat intelligence
- Reduced costs associated with infrastructure malfunction, waste, and inefficiency
- Reduced costs associated with managing data centers
- Reduction in costs associated with market intelligence gathering and R&D due to greater, faster, and more real-time learning about customers
- Longer-term potential for deep learning to power software writing itself, reducing costs associated with expensive human labor (e.g., developers, software engineers, etc.)

Opportunities for new revenue generation associated with deep learning tend to fall into the following categories:

Increased Conversion | Greater Precision | Increased Visibility | Automated Services

- Increased conversion (advertising, sales, etc.) through more accurate personalization and targeting
- Increased pipeline via brokerage services uniting buyers and sellers based on data services
- Increased algorithmic improvement and training via high-volume crowd-sourced platforms (e.g., users submitting billions of photos, videos, content, etc.)
- Monetization of services made possible through data collected across disparate participants or sources (e.g., sensor data fusion, diagnostics, forecasting)
- Enablement of new paid services that could differentiate “freemium” from paid subscription models
- New products/services made possible through deep learning (e.g., recognition services, translation services, security services, gaming or simulation services, etc.)
- New service categories made possible through technological capabilities enabled by deep learning (e.g., autonomous driving services)
- Increased utilization of unstructured, disparate, or altogether “dark” data could unlock previously unforeseen opportunities for unique product or service offerings
- Longer-term potential for deep learning to power software writing itself, identifying new products and services

Tractica’s research finds that net new revenue associated with deep learning emerges when the technology underlies or streamlines other emerging business technologies, namely IoT/machine-to-machine (M2M) applications, computer vision, NLP and generation, drones, and other enablers of “computing perception.” For adopters of the technology, new business models and net new revenue streams made possible by deep learning are somewhat less predictable than its cost reductions. In all likelihood, this will enable business models yet to be conceived of, perhaps even inventing new business models on their own.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

2.2 MARKET BARRIERS

Despite the array of benefits deep learning *could* enable, the technology faces numerous significant barriers to achieving widespread or enterprise-scale adoption. AI, in general, suffers a significant gap between high expectations of intelligence and the reality of current limitations of software and computing. The complexity of the technology and nascence of the market are exacerbated by a significant shortage of talent and a wide range of social, ethical, legal, and regulatory questions and concerns.

2.2.1 BARRIERS TO UNDERSTANDING

One of the most basic but inconvenient issues from which deep learning suffers is complexity. It is not only difficult to understand conceptually, but it is challenging to capture the breadth of its applications. Even those intimate with the technology struggle to explain its architecture beyond its standard building blocks. Below are a few of the major areas of confusion around deep learning:

- **Ontology and Naming Conventions:** While deep learning is a subset of AI, its ontology is hardly well defined. *A learning algorithm by any other name…* Even experts in AI struggle to consistently define what it means with any standardization in nomenclature. Those grasping to understanding this technology will encounter a range of new vocabulary: neural networks, expert systems, cognitive computing, machine learning, machine reasoning, CNNs, knowledge discovery in databases (KDD), heuristics, etc. Deep learning's apparent analogous relationship with the human brain, in all its enigmatic, poorly understood glory, only compounds this struggle for everyday business leaders, practitioners, and end users.

- **Amorphous and Broad:** While deep learning is not the tool for every job, its applications are vast; Tractica has identified some 112 commercial use cases for this technology. It can be used for applications that include landing on an asteroid, identifying cancer, rendering art and music, powering natural language generation in social robots, and beyond. So many manifestations of deep learning are part of its impressive potential, but they also make it difficult for adopters and end users to understand what it is and is not, how it is different from other segments of AI, and when it would be the right solution to their specific problem(s).

- **Hype versus Reality:** Since the dawn of AI in the 1950s, the technology has been prone to hyperbole and unrealistic expectations. Predictions then that any feature of intelligence could be mathematically modeled and, therefore, simulated in a machine have fallen short, and caused numerous "winters" in which interest and investment in the space were all but abandoned. Such predictions also spawned global fascination with (and demand for) science fiction media of all manner. Even today, as data and computing at scale reignite the decades-old dream, the reality is the vast majority of deep learning in production today targets a very specific part of a very specific problem; what we call narrow AI. By contrast, more generalized or strong AI where a machine exhibits behavior and reasoning that is at least as skillful, introspective, aware, and flexible as humans does not yet exist in the wild, never mind at scale. Inflated expectations inspired by science fiction (and perhaps, hubris), coupled with the current state of narrow AI make AI especially easy to oversell.

- **Limitations of Neuroscience:** Since 1956, the scientific understanding of neurons has expanded significantly, but a comprehensive understanding of the brain remains elusive. Neurons differ from one another structurally, functionally, and genetically. There are currently thought to be about 1,000 different types, but new neurons are

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

being discovered every day. As for issues of higher cognition (i.e., how we perceive, how we remember, how we act), we have very little idea how neurons are storing information, how they make inferences, what the rules for higher thinking are, and what the algorithms are. Thus, while deep learning may be inspired by the brain, the reality is we are not yet in an era in which we can confidently or comprehensively use the brain as an accurate guide to the construction of intelligent systems. As Michael Jordan, the Pehong Chen Distinguished Professor at the University of California, Berkeley, likes to point out, deep learning neural nets are only "cartoon models" of the brain.

- **Explainability:** It is perhaps ironic, in its (albeit primitive) likeness to the human brain, that deep learning struggles from an introspection problem. More than just about any other technology, explaining *how and why* a neural network produced a specific outcome is extremely difficult, even impossible in some cases. The super high dimensionality and vast amount of parameters incorporated into hundreds or thousands of layers make explainability (sometimes called "interpretability") extremely opaque. It also makes such models difficult to fix, tune, and de-bug. Companies face many pressures around accountability, compliance, anti-discrimination, and other consumer protections, for which explainability could be a legal imperative. This is yet another way in which deep learning's complexity hinders trust and adoption. Reference Section 2.2.7 for a deeper technological explanation on this issue.

2.2.2 Social Conventions and Trust Issues

In addition to the complexity surrounding deep learning as a technology, it also has a long way to go before it acts as seamlessly and intuitively as a human being. Although the drive to create machines in our image informs this technology, the reality is that we will disrupt ourselves and social hierarchies.

2.2.2.1 Human Biases and Trust in Uncharted Territory

The most important ingredient in technological adoption is trust. All too often, deep learning and AI suffer from underperforming in the context of overblown expectations. This is in no small part due to aspiring to have AI (and particularly deep learning) resemble ourselves.

Humans hold themselves in high intellectual esteem, and intelligence is heralded as a cardinal virtue in modern societies. People are often accorded higher social status based on their intelligence. Educational institutions rank students the world over according to their ability to perform on cognitive tests. Intelligent people often are financially better rewarded as well. This is an article of faith in many cultures around the world that education is a stepping stone toward a better life. This makes AI controversial because it challenges some peoples' sense of self-worth and upsets the social order.

"The whole point is to make artificially intelligent systems less artificial, otherwise people won't trust them," explains Steve Ardire, AI investor and advisor. "This is why emotional intelligence becomes the future of AI; people don't change behavior on information, they change is based on emotion. Also, people think differently than machines; people think cause and effect; machines think in correlation. This is a fundamental barrier to reckoning our relationship to AI."

Below is a list of example areas in which social conventions lag behind, are unclear, or contort the perceived value of AI-driven interactions:

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- **Facial Recognition**: Are consumers comfortable with businesses using facial recognition to personally identify them? Who is responsible for privacy protections? If a crime takes place, will all individuals (identified using facial recognition) in the area be subject to inquiry?
- **Always-On Voice Recognition:** For many voice-enabled devices, defaults are set to always-on, so that the software is constantly listening for a "wake word" (e.g., "Siri" or "Alexa"). If a voice-enabled assistant processes all language it hears, where, how, and when will these data be used? A 2016 murder case was the first time that speech data from the Amazon Echo was summoned for evidence; Amazon declined to provide this or comply with the search warrant, citing customer privacy.
- **Virtual Assistants:** When computers and devices start speaking to end users within the context (and sometimes dialect) of their own language, interactions are often unpredictable and dance along human vulnerabilities and emotional sensitivities. Japanese company Gatebox recently opened pre-orders for a 3D anime character virtual assistant, marketed as a digital companion, designed not only to aid in smart home functionality, but social isolation and loneliness.
- **Hyperpersonalization:** The ability to learn from multiple data sets, some unstructured, is one of the key strengths of deep learning. As a result, many advertisers and service providers envision finally reaching the proverbial holy grail of customer experience: right person, right product, right time, right place, right content, etc.

"It will change us culturally," said the leader of an AI program at a large technology supplier who wished to remain anonymous. "Certainly it will force people at every level to think differently. If we are going to engage super human powers, we're going to have to understand these solutions are going to make decisions that we simply cannot. So much of what we do is subjected to our expectations, and our mental models and biases."

AI has been controversial since its inception. It remains controversial to this day and it will be controversial in the future. The more successful AI becomes, the more controversy it will generate. This has three consequences:

- AI technology will attract controversy regardless of the technology's actual faults or merits. The more the technology threatens the social order, the more controversial it will be perceived.
- Companies that successfully use AI will be reticent to publicize results. They may intentionally or unintentionally sandbag their competitors and make marketing difficult for their vendors.
- As technologies like deep learning become more successful, they will stop being called AI and be re-labeled with one or more socially acceptable terms.

2.2.2.2 Challenges to Adoption and Actionability

One of the greatest challenges to adopting deep learning is getting end users, both employees and customers, to actually embrace it. This is especially important, not only to justify investment in the deployment, but to enable the model to grow smarter over time. The more interactions, the more data, the smarter the algorithms, the better the product, and repeat.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

The following areas limit adoption and value creation in deep learning applications:

- **Employee Resistance:** Even with the fastest processors, best data, and successful training of the model, certain deep learning applications can be thwarted by employee resistance. Some worry that such new tools merely preface their own job displacement. Others are uncomfortable with explaining the technology's role in unsolicited outreach or interactions with customers. Others worry they could instigate an unwanted media crisis. Others may disagree with the recommended action or outcome a model provides.

  Many of the companies Tractica interviewed cited this as a critical hurdle requiring education, training, and assurance that such tools are meant to enhance, not replace employees. One large financial institution is implementing deep learning to identify new upsell opportunities. In confidence, the leader interviewed shared that employees resisted using the model's recommendations because they lacked justification for why they suddenly had new information about customers and were using the information to recommend them new products.

- **Customer Sensitivity:** Consumers have never been more armed with information, nor have they been more influential to their peers, via digital and social media channels. Not only are consumers more discerning about brand interactions, they are (generally) more tech savvy, and intolerant of bad user interfaces and experiences. Such sensitivity is compounded by the ever-growing persona of "big brother" with which many brands and technology companies are negotiating. Some retailers, for example, are hesitant about deploying facial recognition technology in their store environments due to fears of consumer backlash and a public relations (PR) crisis.

  "Even within the constraints of client requirements for maintaining proprietary database repositories, we have still been able to use deep learning to extract highly targeted consumer insights from different organizational data lakes," explains the head of AI for a global advertising company who requested anonymity. "We do this by aggregating common entities and piecing them together to the point where we're able to give a statistical level of confidence of 70-80% that this is the same individual. Therefore, we are able to model what individuals are doing digitally from the time they wake up to the time they go to bed."

  Despite, or perhaps because of, the fact that most consumers do not realize the tremendous amount of data companies have about them, organizations risk backlash if they do not wield such insights responsibly and appropriately.

- **Ongoing Management and Overhead Required:** Implementing deep learning algorithms is no small feat. It typically requires expensive hardware, sometimes customized for specific applications, and lots of training time to learn. Moreover, these are not "set and forget" initiatives, but rather require ongoing resources and overhead to manage. Management is also not just a matter of monitoring basic performance, but of reassessing outcomes, assumptions, and impacts.

  "We're still in the early days and a phase where machine learning can't yet fully reason about environments; with the exception of a few game-related agents, we fake reasoning by adding handmade rules," says Luca Regazio, Director of Engineering at Panasonic's Silicon Valley Innovation Lab. "The task is one of both staying apprised of the open source community and its advancements and

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

movement in this direction, while also working to better understand how systems are making decisions given new inputs. This means that we have to constantly engage in a virtuous cycle in which we're looking at current sets of data and evaluating recognition, inferences, and how systems are taking in environmental context."

2.2.3 Lack of Experienced Talent

Deep learning's unique blend of machine learning, data science, and software engineering demands a very specific skill set that very few people, particularly existing employees, possess. The lack of experienced talent in the deep learning space is an expensive and difficult problem for the market. This gives large technology enterprises (with deep pockets) another advantage over smaller startups and innovators. Most of the hands-on experience within the field is concentrated around a handful of professors and their graduate students.

Although deep learning R&D is not exclusively North American, most of the deep learning expertise centers around three Canadian professors: Geoffrey Hinton, Yann LeCun, and Yoshua Bengio. Many who have the most experience in creating successful deep learning systems passed through the University of Toronto's deep learning group. LeCun now leads Facebook's AI research lab. Geoffrey Hinton works at Google and Yoshua Bengio recently joined Microsoft, and was previously working with IBM's Watson group.

Google, Facebook, and other big tech firms are paying engineers proficient in deep learning over $400,000 a year. Those with the most experience can earn incomes over seven figures. Companies seeking to compete directly with the technology giants, such as in the automotive sector, are going to have to pay top dollar to attract this talent.

Last year at the prestigious Neural Information Processing Systems (NIPS) academic conference for AI research, in an area reserved for graduate students' poster presentations, many of the world's largest technology companies had paid to set up recruitment tables. Interestingly, about half of the enterprises present at the event were not technology giants, but hedge funds and financial firms. Some companies in sectors less exciting than high tech (industrial manufacturing, agriculture, etc.) are creating university collaboration programs or simply acquiring smaller startups solely to obtain best-in-class talent.

A limited pool of talent also inhibits diversity in developing the technology. Joi Ito, Director of MIT's Media Lab, recently expressed one of his greatest concerns about the development of AI and deep learning:

> *This may upset some of my students at MIT, but one of my concerns is that it's been a predominately male gang of kids, mostly white, who are building the core computer science around AI, and they're more comfortable talking to computers than to human beings. A lot of them feel that if they could just make that science-fiction, generalized AI, we wouldn't have to worry about all the messy stuff like politics and society. They think machines will just figure it all out for us.*

Beyond demographic diversity, implementing deep learning applications for enterprise requires extensive enterprise expertise. For graduate students accepting new roles in deep learning, such business experience can be very limited. Corollary to pure business expertise are myriad adjacent areas of consideration, areas that must be incorporated into the code and algorithms as accurately as possible: legal, regulatory, cultural, sociopolitical, and ethical considerations and biases.

Organizations must consider the talent pool for AI beyond traditional computer science degrees, as algorithms become increasingly intertwined and influential with the experiences

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

we have in daily life, with advanced fluency in other disciplines, such as law, ethics, international policy, sociology, history, etc. This is a talent development tool already underway in some universities. Harvard, for example, has brought on renowned computer scientist and ethicist, Cynthia Dwork, to help lead a new research practice in algorithmic fairness, including computer scientists, law professors, and scientists.

Finally, some technology giants like Facebook and IBM have taken to developing internal and external platforms and tools to familiarize less tech-savvy employees and practitioners to contribute to and learn from and about machine and deep learning. "Like in the early days of mobile, where demand far outpaced talent, we see the same (even greater) need to democratize access to deep learning by allowing folks with adjacent skills to more easily leverage programming, analytics, and data skills," says Anthony Stevens, Offering Manager for IBM Watson's Deep Learning solution.

2.2.4 Job Displacement, Human Augmentation, and Transformation of Work

One of the most contentious aspects of AI, and particularly deep learning (considering its role in more general AI applications) is job displacement. In December 2016, the White House released a report stating that AI could cost millions of jobs. In particular, the report found that workers earning less than $20 per hour and without a high school diploma would be at the highest risk of displacement by automation.

Automation via AI generally threatens a wide range of job markets, including:

- **Low Training, High Tedium:** Functions performing highly tedious tasks, or functions that require (sometimes prohibitively) large numbers of individuals to perform a simple task at scale
  - **Example**: tagging content; reviewing paperwork
- **Low Training, High Sensory Perception:** Functions performing logistical tasks that previously relied on human perception
  - **Example**: Inventory movement in warehouses; driving and navigation
- **High Training, Low Bandwidth:** Functions with high training, but [human] limitations for data input, analysis, non-bias
  - **Example**: Economic advisors, inventory managers, real estate agents
- **Beyond Human Capacity:** Capabilities beyond, or superior to, human capacity or ability
  - **Example**: Big Data analytics, empirically-based prediction and decision-making

Although these categories account for millions of jobs worldwide, the reality is that displacement will not happen overnight and we are still early in AI adoption, not to mention trusting systems more than humans. The threat of "blue-collar" job displacement could exacerbate the economic divide between socioeconomic classes and in competitive versus uncompetitive geographic locations, further bifurcating the haves and have nots in a time of widespread distrust and uncertainty. Although most research in prospective job displacement finds that jobs involving high degrees of analytical thinking, creativity, or interpersonal communication and empathy are most secure, longer term, "white-collar" jobs are not necessarily immune either. Some point out that software developers themselves may eventually be replaced by code-writing machines.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Of course, technological phenomena replacing entire job categories are nothing new. All technological revolutions render certain job categories obsolete, but they also create new roles. In 2016, the World Economic Forum famously pointed out that many high-demand jobs of today, such as social media managers, Big Data architects, mobile operating system (OS) engineers, UX designers, etc., simply did not exist 10 years ago. Similarly, AI will undoubtedly drive numerous new types of jobs. Humans are not just encoding AI systems, they must *continuously* program, train, repair, and optimize them based on their relationship and evolution within each application context. There will be countless, if yet-to-be-defined, roles involved in evaluating and updating laws, protections, regulations, and ethical frameworks for an increasingly automated, even autonomous world.

The inevitable limitations of computers, robotics, and machine learning-designed componentry also signal that the impact on jobs will not be one of replacement, as much as transformation. Rarely, say AI experts, will a job be fully eliminated, but rather a percentage changed. Teams will not necessarily be replaced, but forced to reorient. Doctors, for example, may spend less time working to retain new research and diagnostic evidence in their heads, and more time interacting directly with patients. Customer service agents will spend less time pouring through past records in search of specific case resolutions, and more time speaking more productively with customers.

The White House report suggests that the best ways to enhance services will be to pair humans and machines in situations where the human compensates for the computer's weakness and the computer compensates for the human's. The report offers a recent comparative study in radiology to illustrate:

> *When given images of lymph node cells, and asked to determine whether or not the cells contained cancer, an AI-based approach had a 7.5 percent error rate, where a human pathologist had a 3.5 percent error rate; a combined approach, using both AI and human input, lowered the error rate to 0.5 percent, representing an 85 percent reduction in error.*

Meanwhile, businesses will continue to prioritize productivity and more efficient revenue generation, as AI promises leaner operations and potential new business models. (See Section 2.1.7 for discussion on business models.) The potential advantages of the technology are not just for the enterprise. Such "augmentation" of job roles could lead to greater productivity, meaning workers need fewer hours to produce the same amount, potentially leading to greater leisure time and higher quality of life.

Ultimately, some jobs will be eliminated by AI, others enhanced, and still others created. Impacts will be fragmented, happening more rapidly in certain sectors than others. What is critical is that enterprises, governments, and academic institutions work to ease the difficulty of those displaced through training, re-training, entrepreneurial guidance, skills development, and access to new opportunities.

2.2.5 Business Challenges Obtaining High Integrity/Accuracy Data for Training

One of the greatest challenges of implementing deep learning applications is obtaining quality data for training. Deep learning, in particular, is very data hungry; with each layer of the neural network, there are multiple nodes, and each represents coefficients that require statistically significant amounts of data to learn. Moreover, the demand for data is not just one of volume, but of integrity and usability. Exacerbating this challenge is the fact that exponential growth of data itself does not equate to advancement of deep learning; more likely, it merely drives up costs associated with obtaining, cleaning, and storing data in the first place. Today, most enterprise data is "dark data," which is data that organizations

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

collect, process, and store, but never actually use. International Data Corporation (IDC) estimates that 90% of unstructured data is dark data.

Broadly speaking, the challenge inherent to obtaining the volume and integrity of data required to train models is that of proving value. It is difficult to foresee full value of a system without a full build. As organizations struggle to understand deep learning, experimenting with the technology *without clearly defining the problem* is risky. Technology vendors warn against setting expectations too high at the outset, creating a sort of chicken and egg problem. Without starting with a premise, you will not know how valuable the data will be until you build out a solution.

See Section 3.3.2 for a deeper assessment of technical challenges associated with obtaining high-integrity data for training.

2.2.6 Ethical Risks and Unintended Consequences

With all technologies come a host of unintended risks and consequences. AI and deep learning generate unprecedented questions and considerations for enterprises, governments, and consumers alike.

2.2.6.1 Algorithmic Accountability and Risks of Offloading Discretion and Freewill to Machines

As the world adopts deep learning, and machine learning more broadly, algorithms will increasingly inform or control processes and outcomes. This presents unprecedented and unfathomable impacts on social, economic, political, and security dynamics. Algorithms are conceived of by humans; at the mercy of biases in programming, code, data, and labeling. They are also deaf to cultural nuance, human discretion, and common sense notions of fairness. Increasingly, however, algorithms are being used to determine the content we consume online, the advertisements we see, the universities to which applicants are accepted, the loans for which we qualify, and the “segments” in which we fall as evermore targeted consumers.

“Algorithms are never entirely clear of bias,” says Josh Sutton, head of AI for Publicis Sapient, “because they are based on data we feed it. Transparency into training data is key so that we can have a sense of what biases are built in and how to account for them.”

Below is a list of critical ethical areas enterprises and governments must address when applying deep learning and AI:

- **“Playing God”:** Certain questions or situations simply transcend machine capacity for decision-making. The classic, but quintessential conundrum ethicists pose is: how would an autonomous car prioritize (in real time) who or what to sacrifice if it faced a situation in which it had to choose between avoiding a collision with an old lady, a young child, a dog, a group of people, running off a cliff, or potentially killing the passenger. Who decides and how is it reflected, accounted for, or audited in the code?

- **Risk Assessment:** As algorithms become increasingly responsible for determining levels of risk by evaluating diverse data streams, individuals could fall subject to discrimination, disenfranchisement, privacy violation, or other compromise given inadvertent weighting of some data points over others.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- **Inherent (Encoded) Ethical Judgements:** The inability for code to discern between socioeconomic, political, or other cultural dynamics runs the risk of algorithms making judgements or appropriations that could harm or disenfranchise vulnerable populations. If a refugee is unfairly categorized under one political regime as a criminal or terrorist, such labeling could impact their rights, entitlements, and access in another geopolitical climate. Predictive policing, wherein law enforcement rely on predictive models for crime detection and police force allocation, risks intrinsic bias in the code if algorithms rely on historical data alone.

- **Offensiveness:** Machine learning risks developing biases that socially-aware humans simply would not. Microsoft recently had to shut down its online chatbot, Tay, when in the course of 24 hours, it began spouting outrageous hate speech (based, in some unknown part, on user interactions and public data). Guarding against such offensive interactions or abuses of civil or even human rights is particularly difficult when there are no current tools to account for, not to mention automating programming for ever-evolving linguistic, ethnic, and sociological oppression dynamics.

- **Liability:** Questions as big as the autonomous car conundrum also apply to countless other areas, where even if lives are not at stake, financial livelihoods could be. In areas of insurance, healthcare, regulation, supply chain, etc., questions around liability and who (i.e., consumer, service provider, algorithm, manufacturer, advertiser, etc.) are responsible for harmful outcomes, remain unclear, and, in many cases, are totally unprecedented.

- **Accountability of Autonomous Devices:** The question of liability takes on an entirely new set of considerations when autonomous devices (e.g., cars, robots, etc.) themselves are actors. As devices increasingly make decisions, communicate, and even transact on behalf of humans and organizations, who is responsible? Governments around the world are just beginning to ask questions around the "personhood," rights, protections, or exceptions for robots and other autonomous devices. Again, these questions are almost entirely uncharted in the course of human history.

- **Privacy and Surveillance:** Deep learning offers companies new capabilities for pattern detection, involving or aggregating personally identifiable information, which could intentionally or inadvertently be used in ways that threaten privacy. Even with differential privacy, which aims to maximize accuracy of queries from statistical databases while minimizing chances of identifying individual records, aggregate data from different sources could be juxtaposed to infer behavior, location, and lifestyle, and inform precision targeting of individuals. Social robots and autonomous devices that require identity authentication, but monetize through advertising or insurance services present other privacy concerns, particularly when they involve speech, facial, or emotion recognition.

- **Wealth Distribution:** Another unknown of deep learning and AI is the impact on wealth distribution. Automation technologies are slowly eating away at both low- and high-skilled labor. Meanwhile, in the last few decades, the average income of the top 1% has risen by approximately $1.2 million per year, roughly 250%. As more wealth reaches fewer hands, reductions in demand, investment, and consumption resonate outward. How to sustain demand and wealth distribution and preserve social mobility in this context remains a hotly debated issue in the field.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- **In Warfare:** The use of AI, and particularly deep learning, computer vision, autonomous weapons, and robotic militants, in warfare, presents a host of ethical questions around the ethics of war and international humanitarian law. The Geneva Convention of 1949 states that human conduct in war must satisfy three criteria: military necessity; discrimination between combatants and non-combatants; and proportionality between the value of the military objective and the potential for collateral damage. These are complex subjective judgements that are beyond the capability of current deep learning or AI models to make.

Ultimately, these risks illustrate an intense need for enterprises to deeply examine and forecast, to the best of their abilities, the consequences of using algorithms to surpass humans and manual operational structures. In the example of credit scoring, deep learning might reveal unknown knowledge about a relationship between data that a person might not think to look for, such as first name and probability of paying back a loan. Theoretically, such a system would be immune to human bias, treating everyone the same regardless of their age, race, creed, color, sex, national origin, religion, sexual orientation, gender identity, disability, and marital status, unless the data was biased when collected. On the other hand, the system might identify a statistical relationship between one of the variables that might result in a violation of the law. Statistics is subject to its own mistakes, such as false positives and false negatives, so the loan application might be accepted or denied for reasons that no one could explain.

It is essential that the industry (enterprise, government, and academics) prioritize ethical safeguards, research, and frameworks so as to not automate and encode human bias at internet speed.

2.2.7 Opacity in Explainability

The array of risks and experiential outcomes above are not just unknown from a societal or economic perspective, but suffer from technical opacity as well. The ability to explain algorithmic decisions is largely absent today. In the words of famed Android co-creator Andy Rubin, “It’s just like your brain. You can’t cut your head off and see what you’re thinking.”

In supervised learning scenarios, this presents conflicts around “data overfitting” in which models are overly tuned to account for each variable and historical data point in order to explain how a system works. The risk that arises is that such models fail to generalize or adapt to new situations because samplings of past data do not necessarily reflect the reality of the trend. To illustrate, consider a model built to automatically trade financial assets that included data from the recession of 2008, which also detected that the market crashed at the same time a chain of new retail stores opened. Such a spurious correlation could then become part of the financial model for financial asset trading, and the next time a similar retail chain opened, it would trigger the automated sell-off of assets, which has massive cascading effects on the broader financial landscape. Not only is this “correlation over causation” fallacy problematic, it could go altogether overlooked. Incorrect predictions could arise, or worse, be exploited by a malicious adversary.

In unsupervised learning scenarios, particularly involving neural networks, algorithmic explainability is even more opaque. The ability to see “inside” such networks not only to understand *why* an outcome was produced, not to mention which layers and nodes carried the greatest weight in the decision-making, remains poorly understood. Not only is this problematic from the enterprise perspective in terms of low accountability, regulatory compliance, and erroneousness in the model, it is problematic for external parties with an interest in knowing whether such organizations are overtly or inadvertently behaving nefariously or irresponsibly.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

"We often forget the most important thing we have to do is build trust with humans based on the output of systems," explains Andy Hickl, Chief Product Officer for the Saffron Technology Group at Intel. "To do that, we have to be able to provide enough provenance or rationale behind decisions."

Ultimately, society will reject this technology if credible safeguards are not in place. This issue is not lost on the frenetic AI market. Recently, the Association for Computing Machinery (ACM) released a statement on algorithmic transparency and accountability, including seven core principles developers and organizations can apply. They include:

- **Awareness** (of biases and potential harm)
- **Access and Redress** (mechanisms for those adversely affected)
- **Accountability** (institutional responsibility for algorithmic decision-making)
- **Explanation** (of algorithmic procedures and decisions made)
- **Data Provenance** (procedures, biases, governance related to data gathering and training processes)
- **Auditability** (recordability of data, algorithms, models, and decisions)
- **Validation and Testing** (initial and ongoing documentation, assessment, publication of performance and potential harm)

Most organizations are just beginning to wrap their minds around AI and deep learning, so these principles are well-timed. The sooner companies apply best practices with ethical considerations in mind, the less crisis aversion, clean-up, or legal repercussion they will encounter.

Meanwhile, significant efforts are coalescing in the academic field. One notable example comes from the University of Washington's Marco Tulio Ribeiro, a researcher specializing in interpretability. Ribeiro's Local Interpretable Model-Agnostic Explanation (LIME) framework is a method of "perturbing" each part of the input to assess its weight of contribution to the output. Explanations are designed to be local, i.e., they should pertain precisely to the instance predicted; interpretable, insofar as a human can understand them; and model-agnostic, where the method can apply to any model without having to penetrate it.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 3

INDUSTRY AND USE CASE OVERVIEW

3.1 INDUSTRY OVERVIEW

Deep learning has a wide range of applications, impacting workflows and decision-making across just about every industry. It is often used in cases where feature extraction with classical machine learning is too difficult. Also, deep learning is often a "tool" in the software intelligence toolkit, and, depending on the needs of the use case, may be implemented in combination with other programming techniques, such as computer vision and NLP. Below are summaries of the top industries and use cases in terms of the amount of revenue generated, as measured by Tractica's AI forecast model for 2016 through 2025.

3.2 ADVERTISING

The advertising industry is highly competitive and technically adept. It consists of a few large, well-established companies like Google, Facebook, Yahoo, digital advertising networks, and advertising agencies, as well as traditional print, radio, and TV advertising. There are tens of billions of trades daily across all digital advertising exchanges, thousands of times more than the number of daily trades executed by NASDAQ and the NYSE combined. Because of this volume of data, the ad services business has always been highly automated; deep learning systems are now being applied to the mix. This industry is increasingly being run by algorithms, as buyer preferences and the ads they are exposed to are optimized. The more ads are optimized, the more "value" can be generated from a piece of screen real estate.

3.2.1 STATIC IMAGE RECOGNITION, CLASSIFICATION, AND TAGGING

Image recognition, classification, and tagging is one of the most popular and widely applied use cases for AI and deep learning today. Many algorithms are being developed to calculate the optimal ad to show to the current user at the present time. Typically, these algorithms focus on variations of the ad, optimizing among different properties, such as background color, image size, or set of images. Companies like Google, Facebook, and Yahoo are actively using image recognition and classification algorithms to improve advertising by tagging and classifying images, or suggesting improvements.

A company called Ditto Labs uses deep learning to identify company brand logos in photos posted on social media. The software then evaluates the environments and related sentiments in which the brands appear, and then offers companies the ability to target advertising campaigns accordingly and compare brand performance against competitors.

3.2.2 AD INSERTIONS INTO IMAGES AND VIDEO

Companies are increasingly using deep learning to detect patterns and infer opportunities for ad insertion into images and video consumed by customers and prospects. Using image recognition, classification, and tagging helps companies automate what ads to place where, when, and for whom, and to drive intended actions.

For example, computer vision specialist GumGum uses AI to embed ads or links into photos where it finds relevance and helps brands target and expand their advertising. It has used the technique to post ads about an upcoming TV series on targeted photos that featured the star of the show. Facebook uses AI to look for text on an image used for advertising, and labels it as "high text," "medium text," or "low text," helping advertisers achieve a higher

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

success rate with ads that have low text.

3.2.3 Advertising-Related Use Cases

The use cases for deep learning in the advertising industry are as follows:

- Ad insertions into images and video
- Human emotion analysis
- Interactive window displays
- Performance reporting and analytics of ad campaigns
- Static image recognition, classification, and tagging
- Targeted advertising using multi-domain customer data (social, web, context)
- Video content analysis
- Voice/speech recognition

Reference Section 6.3.1 for Tractica's 2016 to 2025 forecast of deep learning in advertising.

3.3 Aerospace

Deep learning is increasingly being used in aerospace to augment navigation, detection, and computer perception, as well as safeguard against threats to aircraft and related componentry.

3.3.1 Object Detection/Identification

For aircraft, drones, and satellites alike, deep learning and computer vision are becoming critical tools to enable or optimize sight and detection. Due to the high variability of data, as well as the criticality of aircraft being able to reliably and accurately detect objects, many existing techniques fell short. With the introduction of computer vision and deep learning, new commercial services are emerging that allow organizations to detect, measure, monitor objects, improve resolution, thermodynamics, and accurate assessment of contents, as well as model and predict patterns. This technology is being applied to support a variety of use cases, including image detection, segmentation, and classification, as well as to support navigation, search, change monitoring, research, and beyond.

Improvements in object detection and identification through aircraft offer many benefits in efficiency, accuracy, and costs in certain cases. Collecting information through aerial imaging may be cheaper than a full networked sensor and connectivity implementation, for example. Deep learning is particularly helpful given it requires low or no feature engineering. That said, some basic challenges remain when it comes to weather, viewpoint, lighting, and atmospheric unpredictability.

With the help of deep learning technology, Neurala, a Boston-based company specializing in AI, is tackling the problem of drone collisions. The company trained its software by feeding it video images of potential collisions from Microsoft Flight Simulator. Neurala's software notifies drone users and operators whenever it recognizes similar, real-time images from a single camera mounted on the drone. The technology is also able to be programmed to identify, find, and track specific types of objects.

Satellite imaging companies are in the process of launching dozens of new satellites in the next year or so, which will be able to provide a refresh rate of 24 hours for the entire planet. Apart from providing a way for humans to track the planet on a daily basis, this also means that image processing will have to be automated, in order to take advantage of this quick refresh rate and trove of imagery data. Satellite images are being mined for real estate

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

development, conservation efforts estimating deforestation, and forecasting growth by analyzing construction sites.

DigitalGlobe is a provider of high-resolution Earth imagery and analytics that processes 4 million square kilometers of satellite imagery every day. The company has been using deep learning, computer vision, and machine learning to more efficiently and accurately identify imagery, objects, and activities. Objects can be fixed, such as infrastructure, buildings, and bridges, or can be moveable, such as helicopters and airplanes. Activities may be events like wildfires or flooding. In certain cases, such as the 2015 earthquake in Nepal, the company had tens of thousands of volunteers pitch in to crowdsource damage assessment over a million tiles of imagery.

Orbital Insight uses computer vision and deep learning to take millions of geospatial images and provide insights based on these images. Providing, for example, the relative count and distribution of cars parked in a retailer's parking lot offers retailers insights into traffic patterns and inputs to forecast traffic over time. It provides similar services, via satellite, to measure crude oil stored in containers and assess oil supply in real time.

3.3.2 Predictive Maintenance

Using deep learning, often in conjunction with classic machine learning techniques, to drive predictive maintenance is another growing use of the technology in aerospace, among many other industries. Here, machine learning algorithms are used to identify failure patterns and detect anomalies, often triggering automated maintenance actions, such as service upgrades, scheduling service engineers, or managing spare parts in inventory chains.

Where deep learning is particularly useful is in its ability to automatically extract features from raw data that are most suitable. This has historically has been a manual, non-scalable, bias-prone process (requiring significant physical and mechanical expertise) of constructing the right features from the data set for detection, as well as derivative features for solving tasks. Thus, deep learning helps reduce efforts, delays, and costs associated with extracting good features, not to mention helping indirectly with the costs associated with broader downtime in the event issues are not preemptively identified and addressed.

Airbus is working with EasyJet to provide predictive maintenance capabilities for its fleet of more than 200 aircraft. Airbus is using EasyJet fleet data in conjunction with data from other carriers to improve prognostic tools and predict when parts need to be replaced, ultimately helping carriers like EasyJet improve fleet performance and reduce maintenance costs.

3.3.3 Sensor Data Fusion

Sensor data fusion is the process of combining data from multiple sensors in order to improve situational or environmental awareness. This is a critical underpinning for helping various types of heavy, mission-critical machinery adjust and adapt to their environments in real time. Sensor data fusion using traditional methods use fixed or hard-wired algorithms to combine data from multiple sensors, and then provide a real-time assessment of the environment, to make adjustments that go beyond object avoidance and navigation.

AI-based sensor fusion exploits statistical interdependencies between disparate data sources, using Bayesian networks and probabilistic graphical models. Deep learning, in particular, is being used to merge samples from diverse sensor types (e.g., accelerometer, gyroscope, magnetometer, barometer, satellite receiver, CPIS, etc.) and account for high dynamism.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Numerous research efforts characterize development in this space, including notable work from ONERA's French Aerospace Lab, which is using deep learning to perform optical and laser sensor data fusion, assess remote sensing images, introduce multi-kernel CNNs for fast aggregation and prediction of scene labeling and segmentation for urban areas.

3.3.4 Aerospace-Related Use Cases

The use cases for deep learning in the aerospace industry are as follows:

- Localization and mapping (commercial aircraft, consumer drones)
- Machine/vehicular object detection/identification/avoidance (commercial aircraft, consumer drones)
- Predictive maintenance (commercial aircraft, consumer drones, satellites)
- Sensor data fusion in machinery (commercial aircraft, consumer drones, satellites)
- Vehicle network and data security (commercial aircraft, consumer drones, satellites)
- Weather forecasting

Reference Section 6.3.2 for Tractica's 2016 to 2025 forecast of deep learning in aerospace.

3.4 Agriculture

The impact of emerging technologies on agriculture has triggered profound, but critical competitive shifts. As populations increase, climatological uncertainty grows, and increasing costs pressure every aspect of the supply chain, making increases in productivity even more essential. Deep learning and AI are cropping up across a variety of processes in agricultural contexts.

3.4.1 Object Detection/Identification

Uses of deep learning for object detection and identification in agriculture range from the detection of defects in poultry eggs, to harvesting tomatoes, to self-driving farming equipment, and image recognition for recognizing and killing weeds. The use of deep learning AI technology is projected to provide large cost savings, as well as reductions in pesticide and fertilizer use.

Startup Prospera uses deep learning to "see" threats that farmers and drones cannot. It uses a device equipped with computer vision and proximal red, green, blue (RGB) cameras to assess water and nutrients, detect pestilence and disease, and monitor current yields. It uses deep learning to process all of this information and predict output, recommend nutrient optimization, conserve resources, and analyze plant development approaches.

3.4.2 Satellite Imagery for Geo-Analytics

Like in aerospace, defense, and other industries, deep learning and computer vision techniques to better capture satellite imagery are being applied to agriculture use cases as well. Farmers and agricultural suppliers have traditionally relied on periodic (monthly, end of season, or less) releases of forecast data; new commercial methods offer updates to that information once every day or two with county-level accuracy. Long-term, this technology has profound implications for our ability to better understand natural resources, how resources move, and how humans and industry are impacting the planet.

Spaceknow and Orbital Insight are two companies that are using satellite imagery data and applying AI techniques to provide analytics around economic or environmental indicators. Another startup, Descartes Labs, is using 4 petabytes of satellite imaging data to assess

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

crop health from space. The company uses spectral information (not visible to the human eye) to measure chlorophyll levels and inform models for crop yield.

3.4.3 Sensor Data Analysis (Internet of Things)

Pervasive sensor application and related networked services, often termed the "IoT," have been transforming the agriculture sector for the last decade. As sensors are now being used to monitor everything from soil, rain, air quality, plants, livestock, and fertilizers to tractors, forklifts, containers, and beyond, agricultural suppliers' visibility into their operations is changing the way goods are produced. Deep learning and machine learning are enhancing sensor technology by analyzing and producing insights.

AI has been used for irrigation scheduling by using rainfall and drip irrigation sensors, a phenotype measuring system for greenhouse climate control, or for predicting the fermentation process of cattle, which can be used to determine their nutritional feed.

Crop health monitoring, both in outdoor fields and within greenhouses, is another area where AI tools gather data from multiple sensors like temperature sensors, soil sensors, pressure sensors, light sensors, water sensors, and wind sensors. This data is then combined and analyzed to predict the health of crops, identify pests, which could damage yields, and provide suggestions for improving crop yield.

The Weather Company recently announced Deep Thunder, a hyper-local weather forecaster that harnesses diverse data inputs, including historical weather reports, to predict and model future conditions.

3.4.4 Sensor Data Fusion

Deep learning is already being used to make sense of a wide variety of sensor data, including temperature, soil, and humidity. These analyses, combined with aerial images from satellites and drones, build detailed models regarding a wide range of environments: machinery performance, soil viability, weather models, etc. Sensor data fusion in agriculture is about extracting data from multiple data sources to facilitate optimal positioning or function of autonomous vehicles, processes, measurement, damage, or other parameters.

3.4.5 Weather Forecasting

Another emerging use case for deep learning is that of weather monitoring and analysis. Methods for tracking and predicting weather have evolved alongside technology since the beginning of agriculture, and today's use of deep learning helps farms and organizations with more accurate forecasting, and applies reinforcement learning on past predictions and actual outcomes. By comparing predictions with accuracies, the model is able to learn and improve simulation capabilities, as well as forecast much further into the future.

AI can be used to perform weather pattern detection, such as cyclonic activity or other extreme weather events. The U.S. National Energy Research Computing Center (NERSC) has used CNNs to classify threatening climate events like cyclones. This work was performed on a CPU-only Cray XC30 supercomputer, where both the training and inference were run on the same platform, although there was some effort involved in adapting the CNN algorithm to the climate data. The main goal for NERSC was to have a model learn the characteristics of a cyclone and classify it, an area where human decision-making variance is an issue. With the algorithm having between 80% and 90% accuracy in identifying extreme weather events, this is only the start and shows that AI techniques can be used for classification and identification of more complex weather systems and events.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.4.6 AGRICULTURE-RELATED USE CASES

The use cases for deep learning in the agriculture industry are as follows:

- Food safety
- Machine/vehicular object detection/identification/avoidance
- Satellite imagery for geo-analytics
- Sensor data analytics
- Sensor data fusion in machinery
- Localization and mapping
- Weather forecasting
- Weed identification

Reference Section 6.3.3 for Tractica's 2016 to 2025 forecast of deep learning in agriculture.

3.5 AUTOMOTIVE

The automotive industry is undergoing one of the greatest transformations in its history, in no small part thanks to advancements in AI and deep learning. The advent of deep learning is supercharging the progress in autonomous cars. The complexity of thousands of vehicles on a particular stretch of road at once is precisely suited for the processing capabilities of neural networks. Autonomous driving is about managing highly uncommon situations that lead to accidents, so deep learning can aggregate data about thousands of human-driven accidents as they arise, and train the autonomous driving system to devise workarounds.

3.5.1 OBJECT DETECTION/IDENTIFICATION

Object detection and identification, particularly used for avoidance and navigation, deals with a specific use case, in which cars, aircraft, robots, drones, and ships are using AI techniques to navigate their routes and avoid obstacles. This is critical for realizing self-driving cars or other vehicles. Multiple sensor technologies, including light detection and ranging (LIDAR), sonar, video cameras, distance, and position sensors, are able to sense objects and obstacles from as close as a few meters away to 200 meters away. This data is then analyzed in real time to identify and classify objects like trees, people, bicycles, street signs, and any other obstacles that it might see in its range.

Self-driving cars are seeing a lot of activity from companies like Uber, Didi Chuxing, and Google launching autonomous taxis, with traditional car original equipment manufacturers (OEMs) working to launch their own autonomous cars in the near future. Otto was recently acquired by Uber for $680 million for its autonomous truck driving algorithms. Amazon may also conduct some acquisitions in order to create a cost-effective drone-delivery service. In addition, established automakers continue their race to remain competitive in autonomous vehicles, with Ford, Volvo, and Mercedes being particularly active.

Neurala uses reinforcement learning to train cars how to navigate highly complex and dynamic environments using simulations. It has partnered with large automotive manufacturers and is developing an embedded "brain" capable of finding, recognizing, and classifying pedestrians, bicycles, infrastructure, other cars, and trucks that is embedded on board so that such critical processing does not rely on the cloud.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Drive.ai is seeking to simulate the widest range of unusual driving challenges possible, to train an autonomous vehicle to handle anything that arises. AImotive/AdasWorks provides a full-stack AI software solution that aims to reduce the cost of making a vehicle autonomous and can incorporate a variety of driving cultures and conditions around the world.

3.5.2 Sensor Data Fusion

Sensor data has been flowing from all manner of car parts and systems for years, but increasingly, this data will be collected and combined with other diverse sets of sensor data for analysis and decision-making. This is particularly relevant in contexts involving other domains, such as smart city applications of traffic management, road and infrastructure solutions, and weather. IoT companies that embed sensors and processors in the roads themselves will aim to infuse these nodes with AI to manage traffic holistically through inter-vehicle multi-point communication. Sensor data fusion via deep learning is a key enabler to ingesting, learning from, and predicting performance, safety, efficiencies, and requirements for truly interconnected municipal and logistic infrastructure.

3.5.3 Predictive Maintenance

Cars occasionally have issues and manufacturers are working to use technology to more efficiently react and be proactive in how they manage these issues. The use of machine and deep learning will be essential for enabling use cases supporting anomaly detection, predictive maintenance, and preemptive repair, update, and service solutions. Vehicle sensor data and vehicle parameters, as well as past data on repairs and repair diagnostics are the two main data sources needed to infer relationships between diagnostic trouble codes (DTCs) and repairs. Deep learning uses these data sets, potentially among others, to simulate and automate predictive maintenance.

Not only will these technologies help manufacturers identify issues, learn from parts, supplier, supply chain, user, or other related impacts, but such intelligence will make a significant difference compared to current car maintenance processes. Instead of *reacting* to an issue once it has occurred (threatening performance and sometimes safety), data coming off of car parts and systems will trigger communications to manufacturers, dealerships, and users, either scheduling appointments for repair, or deploying technicians or even software updates to resolve the issue.

Tesla is a leading manufacturer of smart cars and is already using machine and deep learning to drive a better and safer maintenance experience. In January of 2014, Tesla was forced to recall 29,222 Model S cars. The wall chargers were at risk of overheating. Tesla was able to deliver a software update that eliminated the problem in all 29,222 cars. Not only did this save drivers a pesky trip to the dealership, but Tesla gave customers full control over when they preferred to receive the 45-minute update. Currently, Tesla is using deep learning to monitor and predict parts and performance issues by compiling data across its entire fleet of cars to predict failure, and preemptively send alerts for maintenance.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.5.4 AUTOMOTIVE-RELATED USE CASES

The use cases for deep learning in the automotive industry are as follows:

- Automated on-road customer service
- Building generative models of the real world
- Driver face analytics and emotion recognition
- Gesture recognition
- Machine/vehicular object detection/identification/avoidance
- Truck platooning
- Predictive maintenance
- Sensor data fusion in machinery
- Simulating worlds for AI training
- Localization and mapping
- Vehicle network and data security
- Virtual testing and simulation for racing cars

Reference Section 6.3.4 for Tractica's 2016 to 2025 forecast of deep learning in automotive.

3.6 BUILDING AUTOMATION

A building controlled by a building automation system (BAS) is often referred to as an intelligent building, smart building, or (if a residence) a smart home. Building automation is enabled through devices that control a building's heating, ventilation, and air conditioning, lighting, and other systems. AI gives smart control systems the ability to learn about human habits and facility environments without being programmed.

Deep learning is being applied in building automation, leveraging sensors like motion detectors, photocells, temperature, air quality, smoke detection, cameras, and vibration. Companies are using these inputs to closely identify and track environmental dynamics and threats and recommend spatial optimization in and around buildings.

PointGrab is a company that provides sensing hardware and software that use deep learning and computer vision embedded into IoT devices for edge processing. Specifically, the company uses object tracking algorithms for background modeling, novelty detection, motion estimation, and non-rigid object detection, coupled with proprietary machine learning classifiers and training pipelines to support learning and modeling of office/work space management, staff planning, retail analytics, and occupant safety to track movement of building occupants, and to drive energy savings, smarter allocation, and cost savings for commercial environments.

Reference Section 6.3.5 for Tractica's 2016 to 2025 forecast of deep learning in building automation.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.7 BUSINESS SERVICES

Business Services is a broad category, including operational processes and management processes, as well as processes that support specific business functions (e.g., sales, marketing, HR, etc.). As organizations undergo "digital transformation," they are undergoing systematic digitization of historically analog workflows and tactics. The application of AI is part of this broader trend, and represents the shift from bringing processes online to learning from the massive amounts of information business processes and customers and employees generate, along with data on products and services. From accounting, compliance, enterprise resource planning (ERP), and supply chain management to customer relationship management and beyond, AI and deep learning are gaining traction across every part of the business technology landscape.

3.7.1 INTELLIGENT CUSTOMER RELATIONSHIP MANAGEMENT SYSTEMS

One of the most significant applications for AI and deep learning in the enterprise is its infusion into CRM suites. These are the tools that handle a wide array of tasks like contact management, lead generation, customer outreach, acquisition, tracking marketing engagement, sales information, content creation, ad targeting, and increasingly, documentation of customer support and issue resolution. The broad sets of data now being used in business systems, not only about customers and transactions, but also for products, locations, economic factors, sensor data, weather, and so forth, present organizations (and deep learning models) new context for innovation in customer engagement.

Machine and deep learning are now being applied to better *learn from and act on* these efforts. While machine learning is driving predictive lead scoring, close time, audience penetration, and case classification, deep learning is being applied to build rich customer personas, model sentiment and risk, enhance customer intelligence, and handle massive amounts of data to automate highly personalized customer service and support at scale.

Salesforce.com's recent acquisition of deep learning company MetaMind means that customers will see a wide range of deep learning applications across the company's CRM suite, which includes sales, commerce, apps, service, marketing, support, IoT, and community clouds. Salesforce will leverage MetaMind's deep learning networks for image recognition, ad targeting, customer preference and sentiment prediction, automated data entry, textual analysis, and automated customer support. Salesforce's Einstein product is designed to augment human agents by serving up only the high-priority leads for sales professionals to focus on, for example.

MarianaIQ is another deep learning company that specializes in business-to-business (B2B) account-based marketing. Its deep learning networks develop "personas" of individual customers and prospects to more efficiently prioritize and target, or not. Its ProspectIQ software identifies unknown prospects that the model assigns as high likelihood customers.

Conversica has developed a virtual sales assistant that learns and predicts high-priority and potential leads to pursue.

3.7.2 PREVENTION OF CYBERSECURITY THREATS

Deep learning applied to cybersecurity is one of the most compelling applications to one of the (if not the greatest) problems organizations (and governments) face today. Cybersecurity is an incredibly complex issue because the variables that make up a threat surface span every layer of architecture (e.g., firmware, hardware, software, connectivity, etc.); every user in the system; and an ever-growing range of types of threats (e.g., malware, distributed

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

denial of service (DDoS) attacks, and phishing, among others). In the face of growing sophistication of malicious actors, so much complexity amounts to ongoing vulnerability, along with challenging and costly risk mitigation.

A handful of companies is working to apply deep learning to cybersecurity. Israeli startup Deep Instinct is using deep learning to detect, predict, and prevent advanced persistent threats in real time by triggering actions (e.g., deleting, quarantining, blocking) on malicious files. Customers are provided dashboards that not only visualize monitoring and analytics, but depict what would have happened if threats were not intercepted. Deep Instinct also uses its proprietary database to model and predict *future* types of malware. The technology itself runs on a lightweight agent, which allows it to be used on any device, meaning endpoints are protected regardless of internet connectivity.

3.7.3 Intelligent Recruitment and Human Resources Systems

Recruitment is a high-cost, high-touch, and theoretically high-value business process, yet it has numerous issues today that cost business recruiters time, effort, and expense. With so much data now available about current and prospective employees (hiring trends, conversion, digital engagement, workflows, collaborations, sick days, vacation requests, etc.), companies are now beginning to deploy machine and deep learning to their recruitment and talent management solutions.

When it comes to recruiting, deep learning can be used with NLP and predictive analytics to better target better potential candidates. Models save recruiters the cost and hassle of using headhunters or social media, and use algorithms to search online resumes, websites, social networks, and beyond, correlating candidates' past experiences and relevant skills to job requirements.

For broader HR and talent management, use cases are diverse. AI and deep learning can be applied to examine past performance trends and predict future outcomes, turnover rates, changes to employee engagement, project completion issues, and collaboration improvements, and help automate personalized basic workflows like scheduling and training.

Job search company FirstJob now has a chatbot called Mya, which helps screen candidates, verify qualifications, and handle all communication, scheduling, and basic questions around company culture, policies, and benefits.

3.7.4 Business Services-Related Use Cases

The use cases for deep learning in the business services sector are as follows:

- Agent-based simulations for decision making
- Audio and video mining
- Automated report generation
- Chatbot-based brand/service interaction
- Chatbot-based e-commerce and sales
- Crowdsourced market research
- Enterprise chatbots for productivity and collaboration
- Intelligent CRM systems (contact management, customer acquisition and planning, customer service, predictive sales and marketing)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- Intelligent recruitment and HR systems (candidate finder, recruitment, predictive talent hiring)
- Prevention against cybersecurity threats
- Procurement management
- Project and stakeholder management
- Real-time news analysis and competitive intelligence
- Social media publishing and management

Reference Section 6.3.6 for Tractica's 2016 to 2025 forecast of deep learning in business services.

3.8 CONSUMER

Unlike some other emerging technologies, consumer-facing applications for AI and deep learning are among some of the most pervasive and advanced. Use cases in consumer markets are diverse and leverage numerous device types (e.g., mobile, robotic, smart home) and modes of interaction (e.g., web/click, voice, audio, etc.).

3.8.1 STATIC IMAGE RECOGNITION, CLASSIFICATION, AND TAGGING

Static image recognition, classification, and tagging is one of the most common use cases for deep learning in the enterprise and is being used in everything from image management, brand management, and quality assurance to product recommendations, content creation (and delivery), and beyond.

The primary purpose for consumer-oriented applications for image recognition and classification is to help users automatically segment, tag, and store images for better data mining and retrieval, which could be done through search or similar photo recommendations. Photo upload sites like Google Photos, Apple Photos, and Flickr all use AI image recognition and tagging techniques to automate photo classification and tagging.

Image classification and tagging is also being used to improve advertising by tagging and classifying images, or suggesting improvements. Computer vision specialist GumGum uses AI to embed ads or links into photos where it finds relevance and helps brands target and expand their advertising. It has used the technique to post ads about an upcoming TV series on targeted photos that featured the star of the show. Facebook uses AI to look for text on an image used for advertising, and labels it as "high text," "medium text," or "low text," helping advertisers achieve a higher success rate with ads that have low text.

Other novel consumer uses for image tagging include reading photo descriptions aloud for blind people, an approach that was pioneered by Facebook. Google is now able to automatically produce captions for images, predictive search rendering by device type, and even use algorithms to detect spam and prevent redirects.

3.8.2 SEARCH ENGINE QUERIES

Making information easy to find and access has been the most basic mission for search engines since their inception. But the vast amounts of data and information available (about both the user searching, and their search inquiry), means augmenting search engines is about optimizing content curation to specific user intentions and profiles. Google has even said that, since the beginning, its search engine was always intended to be the necessary

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

precursor to AI. Indeed, in January 2017, the head of Google's search product, Amit Singhal, announced his retirement, only to be succeeded by John Giannandrea, who has overseen AI for Google for years.

In 2015, Google rolled out a deep learning system called RankBrain to power responses to search queries. Through continuous refinement of some 12 billion web searches conducted per day, Google is using deep learning to constantly optimize its search product, and recent acquisitions of Deep Mind and api.ai indicate Google's continued development around the application of deep learning on search. Moreover, it has extended deep learning's application across other (searchable) media types and products, such as images, videos, social media, short messaging service (SMS) text, research papers, patents, statistics, music transcription, and beyond. The ability to consume, process, learn from, and predict associations, opportunities, and connections between so much information presents unprecedented capability, not just for Google, but for humanity.

3.8.3 Product Recommendations

Companies aiming to recommend products to prospects and customers have relied on machine learning for years. Amazon pioneered the "You Might Also Like" shopping experiences, which now incorporate everything from past purchasing history to social media connections, environmental data, and advertising campaigns. Indeed, with massive increases in data, more companies are now beginning to apply deep learning to "right time, right place" marketing and high-precision recommendations to incentivize people to buy.

Sentient Technologies specializes in mining images and customer interactions to support this use case. Its customer, Shoes.com, is currently using deep learning to reflect recommended products based on what customers select, in a 20-questions-inspired workflow. E-shoppers click the shoes most similar to what they are seeking, and the platform serves up personalized recommendations based on those initial selections. Instead of forcing shoppers to type out specific product feature requests (e.g., knee-high boot with 2-inch thick heel and pointy toe in red), the model uses image recognition to streamline the shopping process.

Similar techniques are also applied to entire websites, with Sentient powering customers like Sunglass Hut and Cosabella with specific advertising images, applying A/B testing of aesthetics and workflows, and then feeding the model and, therefore, learning from all associated customer, image, ad, and behavioral data.

Some companies are using deep learning to better recommend their own products, while others are applying this use case to open up new business models by selling others' products on their platforms. Houzz is a home remodeling platform that is using deep learning to scan photos of its proposed remodels and compare images to furniture and products in its database of some 11 million home photos. It now makes available about 6 million products across 15,000 merchants.

3.8.4 Consumer-Related Use Cases

The use cases for deep learning in the consumer sector are as follows:

- Automated tour guide and itinerary service
- Building generative models of the real world
- Computer-aided art

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- Contextual intelligence for mobile
- Face recognition (personal robots, robotics, and autonomous machines)
- Music recommendations
- Machine/vehicular object detection/identification/avoidance (cleaning robots, personal robots, robotics & autonomous machines)
- Predictive typing assistant
- Product recommendations
- Relationships and matchmaking
- Search engine queries
- Smart oven control with food recognition
- Social media feed curation
- Static image recognition, classification, and tagging
- Text-based automated bots
- Voice/speech recognition (personal assistants, cleaning robots, personal robots, robotics & autonomous machines)

Reference Section 6.3.7 for Tractica's 2016 to 2025 forecast of deep learning in consumer.

3.9 CONSTRUCTION

Construction has been relatively slow to adopt deep learning technologies to date, but the industry is ripe with opportunity. From design to on-site, deep learning is growing in experimentation particularly in the areas of image recognition, using robots and/or satellite imagery to assess project feasibility, detect changes within a bounded area, and project progress at a given site. Companies like Orbital Insight, Spaceknow, Descartes Labs, and RS Metrics are building solutions that allow anyone to analyze satellite imagery and perform geo-analytics using computer vision and AI.

Autodesk, a three-dimensional (3D) design software provider is now using AI and deep learning to expedite the design process, from large architecture and instructure, down to the smallest parts, bits, and screws. The model adopts a similar technique to product recommendation engines wherein users input broad ideas and the model returns suggestions based on those inputs, only it also takes into account critical factors such as structural integrity and generative designs.

3.9.1 CONSTRUCTION-RELATED USE CASES

The use case for deep learning in the construction industry is as follows:

- Satellite imagery for geo-analytics

Reference Section 6.3.8 for Tractica's 2016 to 2025 forecast of deep learning in construction.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.10 DEFENSE

The military and defense sector is one of the first adopters and financiers of AI. Today, applications for deep learning span the defense sector and aid in military, intelligence, and security programs' growing needs to mine, analyze, and learn from vast data, on land and in the air, as well as supporting the broader trend of security and defense applications growing more autonomous.

3.10.1 OBJECT DETECTION/IDENTIFICATION

Although the military has lots of data, there is a constant need for better, faster intelligence. Consider, for instance, that in 2011, during the height of the Iraq and Afghanistan Wars, the U.S. Air Force was processing 1,500 hours of full-motion video and 1,500 still images taken from aerial drones every day. The ability to [more] rapidly capture, analyze, and predict based on data, especially without relying on cloud processing, is the key enabler for autonomous decision-making.

Militaries are using deep learning to better enhance target recognition, search-and-rescue missions, and optimize delivery and support during crises. Like in other sectors, models are trained on large amounts of data to detect specific objects with high precision. Such capabilities can be applied in military settings involving object detection and classification for avoidance and navigation in the case of aircraft, drones, robots, tanks, ships, or other vehicles; search-and-rescue to provide humanitarian support; facial recognition; learning from signals or sensor data (e.g., radar, global positioning system (GPS), sound); and more.

A key enabler of these use cases in military contexts is high-performance embedded computing (HPEC) platforms, in which neural networks can run at the chip level instead of in the cloud. FPGAs, power-efficient GPUs, and advanced single instruction, multi-data (SIMD) processing units help surpass processing limitations required for real-time compute in data-intensive, connectivity-constrained, mission-critical contexts. This sort of compute allows, for example, drones to process object and image recognition on board, instead of sending it back to human data analysts, who could be halfway around the world. It would also support critical data processing in disaster zones or help circumvent malicious threats and compartmentalize sensitive intelligence.

Geospatial or satellite images are the lifeblood of the aerospace and defense industries, used for infrastructure mapping and to identify and track changes on the ground from space. Humans are a key part of the analysis of geospatial imagery, but increasingly, this is becoming automated with AI techniques like deep learning and computer vision being used to extract features, identify features, classify the images, and track changes.

3.10.2 AGENT-BASED SIMULATIONS

When it comes to strategy, warfare has only grown more and more complex. The importance of taking into account massive amounts of data, contexts, and evolving geopolitical forces is a task simply beyond the scope of human bandwidth. In 2009, officials at the Defense Advanced Research Projects Agency (DARPA) discussed the opportunity for deep learning, as it related to image recognition, multi-data set analysis, and decision-making:

> *Full exploitation of information is a major challenge… Human observation and analysis of [intelligence, surveillance, and reconnaissance] assets are essential, but the training of humans is both expensive and time-consuming. Human performance also varies due to individuals' capabilities and training, fatigue, boredom, and human attentional capacity.*

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

As a result, deep learning is being applied to simulate tactical moves in and refine military strategy in real time. Reinforcement learning helps make agents smarter and each agent plays out different strategies and the best strategy could be applied. The U.S. Department of Homeland Security's Synthetic Environment for Analysis and Simulations (SEAS) project is using deep learning to predict and evaluate future events and courses of action. Agent-based simulation is an emerging use case with applications in government, business, and beyond.

3.10.3 Sensor Data Fusion

Sensor data fusion has been deployed in military and defense settings for years, and is increasingly using deep learning to more accurately detect, classify, model, and "learn" from environmental context and impacts. In a military setting, sensor data fusion in the traditional sense would fall short if the radar is being jammed in a certain direction, while with an AI closed-loop technique, the radar can adjust the antenna in the jammer's direction to nullify the effect. In military or marine applications, sometimes a GPS is not very reliable, which makes sensor fusion a useful technique that uses geographic information system (GIS) data to determine a vehicle's location and predict its future location. Like in aerospace or energy, these applications are often mission-critical and precision, speed, and reliability are paramount to adoption.

3.10.4 Defense-Related Use Cases

The use cases for deep learning in the defense sector are as follows:

- Agent-based simulations for decision making
- Localization and mapping (aircraft, drones)
- Machine/vehicular object detection/identification/avoidance (aircraft, drones)
- Predictive maintenance (aircraft, drones, satellites)
- Prevention against cybersecurity threats
- Satellite imagery for geo-analytics
- Sensor data fusion in machinery
- Vehicle network and data security (aircraft, drones, satellites)

Reference Section 6.3.9 for Tractica's 2016 to 2025 forecast of deep learning in defense.

3.11 Education

Although born in academia, deep learning applications in the education sector are still somewhat limited beyond research in computer science. Given the massive amounts of data generated in the education process, a number of applications are being explored.

Alinea, a leading Danish publisher whose math content is used by the majority of students in grades 1 through 7 in Denmark, recently launched CampMat, an adaptive math learning product that tailors recommendations for students in grades 1 through 3 studying numbers and algebra. In partnership with adaptive courseware provider, Knewton, Alinea's content populates the platform, while machine learning powers a dynamic digital curriculum. Interactions and student data inform individualized instruction based on real-time analysis of what a student knows, how he or she learns, and his or her stated learning goals. CampMat also leverages game-based learning strategies to engage and motivate students.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.11.1 EDUCATION-RELATED USE CASES

The use cases for deep learning in the education industry are as follows:

- Automated grading of tests
- Spoken fluency evaluation
- Textual question answering

Reference Section 6.3.10 for Tractica's 2016 to 2025 forecast of deep learning in education.

3.12 ENERGY

Applications for deep learning are somewhat limited in the energy space today, particularly compared to other large industrial sectors. That said, there are numerous projects involving both deep learning and computer vision to better aid in various parts of the energy transmission and distribution value chain. Manufacturers of metering devices are leveraging robots; energy management systems are applying more machine and deep learning to process data and make predictions.

PowerScout is using deep learning to analyze satellite data to detect and determine homes that would be likely candidates for solar panels given their positioning and exposure to light. This helps optimize sales and marketing costs associated with targeting the right potential buyers. The company has trained two neural networks to determine 1) whether a home already has solar panels (or not), and 2) whether nearby vegetation would hamper installation or energy generation efforts. It is also developing an e-commerce site in order to use this data to let users run feasibility and estimated value and returns on solar panel installation for their homes. Based on "solar worthiness," the service then matches potential buyers with local installers, and offers tailored financing options. In the future, PowerScout hopes to use this data to optimize community solar sales by suggesting installations wherein multiple residents could take advantage.

3.12.1 ENERGY-RELATED USE CASES

The use cases for deep learning in the energy industry are as follows:

- Satellite imagery for geo-analytics
- Weather forecasting
- Nuclear/power plant safety

Reference Section 6.3.11 for Tractica's 2016 to 2025 forecast of deep learning in energy.

3.13 FASHION

The application of deep learning in fashion is not only about customer style and satisfaction, but about design as well. Techniques in which models are fed data sets and queried to develop outputs based on those data sets yields compelling creative assets, ranging from clothing to art, music, and beyond.

3.13.1 FASHION TREND PREDICTION

The ability to predict preferences, behaviors, and market movements has many mission-critical applications in defense, weather, finance, news, etc., but can also be applied to retail markets. One such area with high potential is in predicting fashion trends. The global fashion industry is valued at $3 trillion, accounting for 2% of the world's gross domestic product

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

(GDP), and employs millions of people internationally. It also encompasses a massive ecosystem of businesses, including designers, manufacturers, distributors, marketers, advertisers, etc. The challenge in this area is efficiently matching supply and demand, for all to benefit. Currently, fashion brands and retailers work with a limited amount of data to predict what products to order and when to discount or replenish them. If they predict wrong, the result is loss of income due to mark-downs, waste, and popular items selling out.

What has historically been developed based on traditional market research methods is now being assessed through diverse data streams fed to algorithms. By analyzing large amounts of data, such as the browsing and shopping history of every single one of a fashion brand's online customers, as well as those of its competitors, AI can tell a retailer how to align product drops to match demand, and even how to display products in a store to maximize sales.

Stitch Fix is an online platform that provides highly personalized styles to women. It provides its personal stylists tools and technology to help hand-select clothing and accessories that fit shoppers' preferences, lifestyles, and body shapes. The company introduced Deep Style to help collect and learn from product, customer, service, and workflow data in order to predict and deliver "just right" fashionable styles at the individual level. Specifically, it uses photographs to quantify the style and identify unique attributes of items in its collection. It is also using the model to associate one article of clothing with related accessories and color schema, as well as using all product and customer data to inform computer-generated clothing it can use to simulate new designs.

Figure 3.1 Stitch Fix Uses Deep Learning to Analyze Styles and Design New Clothing

In the first image, an actual shirt is analyzed in conjunction with other data sets to create a recommended shirt. In the second image, the model designs multiple "like" shirts, based on different variables for different user segments.

(Source: Stitch Fix)

Reference Section 6.3.12 for Tractica's 2016 to 2025 forecast of deep learning in fashion.

3.14 FINANCE

3.14.1 RISK ASSESSMENT AND COMPLIANCE

Financial institutions and other companies spend billions every year on fraud, counterfeit, and a host of other threats to credit, identity, and financial transactions. The World Bank estimates the amount of money laundered each year amounts to somewhere between $2 trillion and $3.5 trillion. To combat this tremendous problem, anti-money laundering (AML) compliance and penalties cost banks approximately $18 billion annually. The size of this

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

problem, coupled with the vast number of transactions and actors involved in international financial systems, not to mention the lack of data mutualization and duplicative efforts, means that today's AML procedures are extremely manual and labor intensive.

To combat these threats and help reduce costs, companies are developing AI-based methods for assessing customer risk and improving regulatory compliance. AI and deep learning are being applied to better learn legal requirements and variations, spot inconsistencies across millions of real-time transactions, identify policy violations, and mitigate against financial crimes, money or trade-based laundering, and assess liquidity risk.

Next Angles is a venture within Mphasis Corporation, which provides AI tools to mitigate know your customer (KYC)-AML and trade-based money laundering (TBML)-related risks by using algorithms to rapidly review large amounts of documentation, detect suspicious or inconsistent information, run sentiment analysis, and automate transaction monitoring. It also provides solutions that equip financial crime investigators with tools for more rapid and proactive data compilation, consolidation, analysis, and inference drawing. The platform "learns" over time and offers easier scale through a conversion of structured English documents into executable rules.

RAGE Frameworks, an automation technology and services provider, is supporting clients with assessing daily business and credit risk. Companies can set up the system to assess dozens of business drivers to ascertain such risk. For instance, enterprises can use RAGE to monitor the internet, stock investments, competitive intelligence, market developments, and a range of risk factors, and then apply linguistic analysis and learning to identify and alert them of various business risks as they evolve in real time.

Insurance companies are also looking at AI and deep learning tools to assess policy-holder risk. The 2013 Monsanto-Climate Corporate acquisition leverages agricultural and weather data to support such a business model. It is using deep learning to determine the underwriting risk of selling farmers insurance against weather-related losses.

3.14.2 Finance-Related Use Cases

The use cases for deep learning in the finance industry are as follows:

- Biometric identification
- Converting paperwork into digital data
- Patient data processing
- Employee expense management
- Risk assessment and compliance
- Tax filing and processing

Reference Section 6.3.13 for Tractica's 2016 to 2025 forecast of deep learning in finance.

3.15 Gaming

Perhaps more than any other industry, gaming has played a most significant role in the computational breakthroughs that have accelerated the technology's development. Andrew Ng's discovery that GPU chips used to support process-heavy video cameras could run neural networks in parallel opened the door to more rapid neural net processing.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.15.1 Create Dynamic and Interactive Video Game Experiences

Although gaming as an industry includes a wide range of games, including casinos, board games, fantasy sports, and beyond, the ability to create dynamic and interactive video game experiences is one of the original use cases for deep learning. Game developers are using neural networks to support all kinds of use cases, such as predicting player actions, inferring and recognizing player goals, developing adaptations to unpredictable player actions, and learning from simulated game environments.

Gaming environments are also crossing over from purely entertainment into industries as well, particularly in support of education, learning, and training students or employees on new skill sets. With the rise of virtual reality (VR) and AR hardware, new form factors will support immersive training and gaming becoming more mobile and less fixed.

Perhaps one of the most interesting aspects of deep learning applications within the virtual gaming context is what they offer, by way of simulation, to real world applications where models can learn to understand and navigate 3D spaces. This alleviates significant logistical, safety, and financial barriers. For example, many companies working on autonomous vehicle components and software are using neural networks to support simulated training environments. (Running over a digital cyclist and off a digital road is far less of a mishap in the virtual world than in the physical world after all.) 3D graphics and modeling in video games, for example, is one way researchers are working (using a development engine called Unity) to use deep learning algorithms to make better sense of the real world.

Figure 3.2 Xerox Research Scientists Simulate Driving Conditions Using Video Game Development Engine

Both screens depict a car driving down a simulated highway environment. Both scenarios are testing the vehicle's ability to successfully identify other cars, but compare different weather conditions and viewing angles on the bottom screen.

(Source: Massachusetts Institute of Technology)

3.15.2 Gaming-Related Use Cases

The use case for deep learning in the gaming industry is as follows:

- Dynamic and interactive video game experiences

Reference Section 6.3.14 for Tractica's 2016 to 2025 forecast of deep learning in gaming.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.16 GOVERNMENT

Government applications for deep learning are diverse. While the technology remains early in adoption, use cases span a wide range of public, municipal, operational, and conservation applications.

3.16.1 CROWD ANALYTICS

Crowds and public gatherings typically offer lots of information, including cause, size, demographic, movement patterns, etc. In 2015, IHS estimated there were over 245 million operational cameras active globally. When it comes to digitally analyzing this information, many approaches fall short. For instance, crowd analytics may be effective in a pre-configured setting (e.g., a plaza), but analyzing crowd formations in new or unknown areas fails. As public and commercial infrastructures install more and more cameras, and as computer vision and recognition techniques advance, deep learning finds new application in analyzing crowds. Research institutions in China and India have been working to develop training data and deep learning solutions capable of estimating crowd density and dynamics.

Such government surveillance tactics are one of many with which deep learning researchers and innovators are experimenting and include object detection and identification, behavioral analytics, sentiment analysis, facial recognition, even predicting social unrest or geopolitical events. In conjunction with surveillance footage, the sheer volume of video produced is fed into deep learning models to detect abnormalities, identify and trace moving objects, better manage large crowds, and ensure municipal and public safety.

3.16.2 TRAFFIC LIGHT MANAGEMENT

As cities and municipal infrastructure become increasingly connected through sensors and data analytics, AI and, eventually, deep learning will be critical tools to aid with learning from and better predicting traffic flow. Deep learning is well suited for this use case given the diverse and often unstructured and time-series data sets flowing in from a range of inputs influencing optimal lighting and energy utilization. Pedestrian traffic, private, commercial, and public vehicle movement and concentration, weather, and municipal services are just some of the diverse and huge data sources that will be analyzed to optimize street lighting, among a wide range of other smart city applications.

3.16.3 GOVERNMENT-RELATED USE CASES

The use cases for deep learning in the government sector are as follows:

- Agent-based simulations for decision making
- Behavioral analytics
- Converting paperwork into digital data
- Crowd analytics
- Dialect classification
- Face recognition
- Object detection for surveillance
- Predicting social unrest and geopolitical events
- Real-time video analytics

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- Sentiment analysis
- Street lighting
- Waste sorting and recycling
- Weather forecasting
- Crime reduction and prevention

Reference Section 6.3.15 for Tractica's 2016 to 2025 forecast of deep learning in government.

3.17 HEALTHCARE

Healthcare is one of the most data intensive, mission-critical, and economically unwieldy industries there is. Data associated with medical services, devices, patients, diseases, records, and so forth currently exists in siloed databases and is disparately applied to point solutions and acute problems. From clinical medicine to hospital administration and medical research, deep learning presents new opportunities to increase efficiency, quality of care, and even treatments and cures.

3.17.1 EFFICIENT, SCALABLE PROCESSING OF PATIENT DATA

The generation, input, processing, analysis, security, compliance, and utilization of patient data create massive challenges to healthcare organizations the world over. This is, of course, not to mention the ability to learn from and use such data predictively. Indeed, the greatest challenges when it comes to patient data center around processing and analytics.

More and more medical and research institutions are leveraging deep learning for data analysis in the name of driving faster and more precise treatment. Deep learning is also being applied to analyze medical images and aid doctors as they analyze images. The Pediatric Intensive Care Unit (ICU) of the Children's Hospital of Los Angeles is currently using RNN and CNN deep learning to analyze 10 years of electronic health records, across 20,000 patients in order to simulate and develop better treatments, create illness profiles, and observe patient outcomes over time. "Our overarching goals are to keep more kids alive, to reduce the length of their stays as well as morbidities and ancillary effects," explains David Ledbetter, of the Children's Hospital of Los Angeles. "But we also aim to be an augmentation to doctors by mining for collective wisdom: Wisdom from over roughly 10,000 years' worth of patient data as well as by analyzing the state-of-the-art information to recommend personalized treatments for particular patients at particular points to optimize their outcome."

Much of the latest innovation focuses on using decision trees and neural networks around patient data to improve fraud detection, claims processing, scanning (analog or digital) patient records, marketing, behavioral analysis, and preventive insurance. Unlike fixed statistical models, dynamic models using AI adapt to shifting parameters, making areas like fraud detection and claims processing self-learning and far more cost-effective than current models.

Longer-term applications for deep learning and patient data involve using diverse data sets for medical research, drug and treatment development, and preventative care. Integrating patient data with its AI health tool enables IBM's Watson Health to mine patient data to find relevant facts about family history, current medications, or any pre-existing conditions, providing alerts or early warning signs through its system. DeepMind Health in the United Kingdom has acquired data from the National Health Service (NHS) to allow its algorithms to look for early warning signs for specific conditions like Acute Kidney Injury (AKI).

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Another adjacent application for patient data is in genomic data mapping and analysis. Toronto-based startup Deep Genomics is using deep learning to better understand how genomic realities, alterations, and variations across individuals and populations manifest as diseases. Its goal is to better understand diseases, disease mutations, and gene therapies, eventually using these findings to inform precision medicine and personalized therapies.

Tractica expects patient data processing using AI to become much more commonplace by 2025, unleashing creative ways of understanding patient groups and health conditions, identifying hidden efficiencies in healthcare, and improving precision medicine.

3.17.2 Medical Image Analysis

Analyzing images is a strong application for deep learning within the realm of patient data processing. Historically, analyzing medical images has been difficult, highly prone to human error or oversight, and time-consuming and costly. Medical images like magnetic resonance imaging (MRIs), X-rays, computed tomography (CT) scans, and other diagnostic images are essential to better understanding and diagnosing a wide range of conditions. When it comes to diagnosing critical conditions, including cancer, neurodegeneration, and heart disease, the faster and smarter the speed, precision, and predictive capabilities, the better.

Deep learning is now being applied to automate the analysis and increase accuracy, precision, and understanding of images down to the pixel. A company called Enlitic uses deep learning networks that analyze medical imaging data, such as X-rays and MRIs, to identify even the smallest suspicious clues (e.g., tumors, hairline fractures, spots, etc.). Its networks increase diagnostic accuracy in less time and at reduced cost compared to traditional diagnostic methods. Enlitic's software also allows comparison of an individual patient's radiological data with millions of other patients who received the same diagnosis in order to identify and track treatment outcomes for the most similar cases.

3.17.3 Medical Diagnostic Assistance

Penetration of AI into medical diagnostics has the potential to not only enhance doctors' speed, accuracy, and preventative strategies, but to advance society's collective understanding of the body and medical treatment. For centuries, doctors have been using one-on-one medical diagnoses by matching patients' symptoms to various lists, common effects, or frameworks associated with diseases. Seasoned doctors surely offer their experience, intuition, and extra training to diagnoses and treatment plans—expertise that algorithms may never quite match—but there remain tremendous errors in medical diagnoses, or diagnoses come too late. At least 80% of cancers could be effectively treated if detected earlier.

Approaching this problem using deep learning involves feeding medical records and images into neural networks and algorithms begin to detect patterns and abstractions, not just across symptom-disease associations, but across diseases, patients, geographies, environments, etc. The ability to take in, retain, analyze, and learn from so much diverse data simply transcends human capability and bandwidth. Doctors of all types will increasingly begin to leverage AI-generated inputs in their diagnostics.

Researchers from Sutter Health and the Georgia Institution of Technology demonstrated that, upon analyzing electronic health records using neural networks, they were able to predict heart failure as early as 9 months before doctors.

Freenome is tackling the problem of cancer diagnosis by using deep learning to detect cell-free Deoxyribonucleic acid (DNA) sequencing of cancer in the blood. The model clustered

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

characteristics by location, which helps scientists and doctors pinpoint where cancer is growing in the first place – a critical part of the puzzle.

A number of adjacent use cases will also frame AI's ability to aid in medical research, diagnostics, treatments, etc., with clustering and phenotype discovery, bio-marker discovery, treatment recommendations, genomic mapping, virtual assistants for patients, and beyond. It is also worth noting that a host of bioethical and ethical issues could arise, particularly around genomic targeting, less clinical drug testing, explainability of systems, inadvertent erroneousness, etc. Moreover, the standards in medicine are very high, which contributes to a bias within the profession against innovation. Even if these digital diagnostic tools are able to reach a 99.999% success rate, they will never be perfect, and mistakes due to false readings could lead to medical malpractice lawsuits and product liability issues.

3.17.4 HEALTHCARE-RELATED USE CASES

The use cases for deep learning in the healthcare industry are as follows:

- Automated report generation
- Bio-marker discovery
- Clustering and phenotype discovery
- Computational drug discovery
- Converting paperwork into digital data
- Face recognition
- Genomic data mapping and analysis for personalized healthcare and precision medicine
- Market intelligence for life sciences
- Medical diagnosis assistance
- Medical image analysis (3D computer vision, radiology, eye diseases)
- Medical treatment recommendation
- Medication compliance for clinical trials and general usage
- Methods for monitoring vitals
- Patient data processing (administration, clinical medicine)
- Portable and low-cost ultrasound device
- Predicting illness and patient outcomes
- Text classification and mining for biomedical literature (clinical medicine, public health)

Reference Section 6.3.16 for Tractica's 2016 to 2025 forecast of deep learning in healthcare.

3.18 INFORMATION TECHNOLOGY

Businesses and other organizations rely on IT systems for storing, retrieving, and sending data. The main components of information systems are computer servers, network, storage, software, and support services. AI is used within IT departments in two ways: to improve the organization's operational processes and to improve support processes. Deep learning is

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

showing promise in the area of network/IT operations monitoring and management and data center automation. Data centers, in particular, consume a lot of energy, and any and all efficiencies gained in these contexts can save lots of money and unnecessary energy waste.

Google's acquisition of DeepMind for $625 million quickly paid for itself using the company's deep learning platform to understand data center dynamics, analyze cooling methods, increase efficiency, and search for more intelligent power management across Google's data centers. Google managed to secure an overall electricity savings of 10% to 15% across its data centers, representing a huge reduction in 18 overall operating costs. The same can be done for bandwidth allocation and packet delegation at the router levels. Companies are also beginning to use deep learning to support broader operational analytics, including, but not limited to:

- Determining optimal configurations
- Extracting real-time insights from Big Data application stacks
- Identifying and predicting performance problems and their causes (e.g., bottlenecks, balance loads, tune factors, etc.)
- Informing opportunities for optimizing efficiency
- Simulating "what-if" models to support chief information officer (CIO) decision-making on capacity planning

These capabilities help minimize time and costs associated with trial-and-error methods for protection and optimization, and are available as open-source tools:

- Network/IT operations monitoring and management
- Website creation

3.18.1 Information Technology-Related Use Cases

The use cases for deep learning in the information technology industry are as follows:

- Computer aided design
- Network/IT operations monitoring and management
- Simulating worlds for AI training
- Website creation

Reference Section 6.3.17 for Tractica's 2016 to 2025 forecast of deep learning in information technology.

3.19 Investment

The investment business is the professional management of various corporate and public securities (stocks, bonds, and derivatives), as well as other assets, such as real estate, in order to meet specified investment goals for the benefit of the investors. The business is divided between institutional investors who invest on behalf of insurance companies, pension funds, charities, and educational establishments, and retail brokers and money managers who invest on behalf of individual investors.

Every day, computers perform billions of calculations and make millions of electronic trades. Algorithmic trading, sometimes called "algo-trading," has been part of automating investment for years, but new advancements in deep learning are being applied to improve strategy and

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

performance. The most common application of algo-trading is to enhance trading strategies, including arbitrage, intermarket spreading, market making, and speculation. In this context, neural networks can uncover complex patterns, trends, and relationships unable to be detected by humans in high-input/high-speed environments. The idea is that, just as deep learning successfully identifies particular features in common to cat images, it may be able to identify particularly lucrative features of stocks as well.

Bridgewater Associates, Euclidean, Man (AHL) Group, and a number of other established investment hedge fund firms are investigating how and where they can apply deep learning. Meanwhile, a host of startups like Sentient Technologies, Clone Algo, Alpaca, and Binatix are working on using AI and deep learning to improve or automate investment as well.

Aidyia is a Hong Kong-based investment company applying evolutionary programming, chaotic dynamics, and probabilistic knowledge to algo-trading. The system ingests a range of inputs, such as price and volumes from around the world, news in numerous languages across multiple sources, and macroeconomic and company accounting data, and studies how multiple factors within these data sources have interrelated historically.

Given the high stakes, experts point to a number of remaining challenges in deep learning-enabled algo-trading, namely around the limitations of models to fully regard (or disregard) noise, random vectors, and high uncertainty prevalent in financial markets. Furthermore, the very commoditization of such algorithms would erode their competitive predictability, until, that is, algorithms themselves advance in evolutionary computation.

3.19.1 Investment-Related Use Cases

The use cases for deep learning in the investment industry are as follows:

- Algorithmic trading strategy performance improvement
- Market intelligence and data analytics for investment
- Satellite imagery for geo-analytics

Reference Section 6.3.18 for Tractica's 2016 to 2025 forecast of deep learning in investment.

3.20 Legal

The impact of emerging technology trends and implications on the law, and for attorneys and regulations, is vast. Not only are legal entities around the world contending with unprecedented contexts and cases involving data, devices, and networked services, they are trying to proactively safeguard consumers and not stifle innovation simultaneously. AI presents a host of new challenges and tools to all involved in case law.

3.20.1 Legal-Related Use Cases

The use case for deep learning in the legal industry is as follows:

- Automated report generation

Reference Section 6.3.19 for Tractica's 2016 to 2025 forecast of deep learning in legal.

3.21 Logistics

Logistics involves the process of planning, implementing, and controlling the efficient flow and storage of goods, services, and information from point of origin to point of consumption to meet customer requirements. Logistics services includes inbound and outbound

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

transportation management, fleet management, warehousing, materials handling, order fulfillment, logistics network design, inventory management, supply and demand planning, third-party logistics management, and other support services. Deep learning is being applied to inter-organizational supply chain management in a variety of ways, and also intersects with other industries, such as automotive, manufacturing, transportation, and finance.

3.21.1 Demand Forecasting

The ability to understand product demand has a direct impact on the economic viability of any business. In the past, companies have been reactive or merely formulaic in their approach to gauging supply orders to fulfill demand.

With the advent of machine and, increasingly, deep learning, companies are able to analyze, learn from, forecast, and predict demand with far greater accuracy and with regard to a wider range of forces. This is an area where AI will intersect with the IoT in that it will incorporate sensor data and radio frequency identification (RFID) tag monitoring, as well as support modifications at each level of supply chain. For example, deep learning can be applied to improve parts and labor sourcing, channel optimization, product inventory, quality assurance, fraud, risk modeling, weather forecasting, and predictive maintenance to support reliable supply chain operations.

3.21.2 Logistics-Related Use Cases

The use cases for deep learning in the logistics industry are as follows:

- Demand forecasting for warehouse and supply chain
- Machine/vehicular object detection/identification/avoidance
- Localization and mapping
- Satellite imagery for geo-analytics
- Supply chain & logistics (freight transport, retail)
- Weather forecasting

Reference Section 6.3.20 for Tractica's 2016 to 2025 forecast of deep learning in logistics.

3.22 Manufacturing

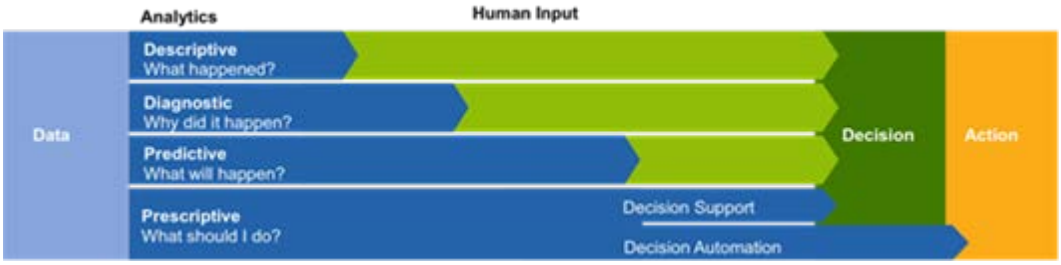
Manufacturing is a broad sector that includes the full process of engineering something from raw materials, components, or parts to the final product. The introduction of AI and deep learning into the processes and machinery that make up the manufacturing production cycle have applications in everything from design to robotics and distribution.

3.22.1 Predictive Maintenance

The ability to predict when, where, and why a machine, robot, device, or system will go down is a function of cost, efficiency, and sometimes even safety mitigation in industrial manufacturing environments. In manufacturing, predictive maintenance uses data inputs from disparate streams to predict failures. Unlike preventive maintenance or condition-based maintenance, which is triggered by the occurrence of one or more indicators, predictive maintenance helps predict failures beforehand.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Figure 3.3 Analysis and Automation Can Occur at Every Level of Maintenance



(Source: SkyTree)

Both predictive and condition-based maintenance use real-time data as feeds. While condition-based maintenance is much more widely used today, predictive maintenance is gaining popularity, especially for mission-critical assets. Techniques like sequence analysis can be used to understand failure patterns and follow-on failures, while machine learning and deep learning can be used to perform predictive models or recurrent event models.

Presenso is an Israeli-based cloud solution that runs neural networks on industrial factory and machinery data to deliver deep semantic insights on unstructured data, proactive asset management, and failure prediction based on adaptive learning. A number of startup and data analytics companies are working on predictive maintenance using AI in the fields of manufacturing, aerospace, and automotive. Examples include Machina Metrica, Pivotal, Falkonry, Simularity, Tellmeplus, Konux, and Augury, among many others.

The Mantis Project in the European Union (EU) is one such organization using deep learning to develop proactive maintenance service platform architectures that predict and prevent failures, estimate future performances, and schedule accordingly. Its work is geared specifically to industrial machines, aerial and land vehicles, and renewable energy assets.

3.22.2 Object Detection/Identification

The ability to "see" in factory and manufacturing settings is very often what has defined quality for parts and products produced. In the past, precision of parts and elements relied on humans, and, later, heavy machinery for preconfigured repetitive evaluation and sorting.

With advances in machine and computer vision, which are becoming deep learning enabled, the ability to more accurately and precisely detect and identify specific features automates tasks like fault detection, failure type detection, visual inspection, inventory monitoring, product testing, workplace safety, video analytics, and potentially additive manufacturing techniques like 3D printing in the long term.

Another impact of deep learning-based object detection is flexibility and reduced downtime. Replacing a robot or machine on the production line is costly, slow, and requires downtime to calibrate. A project by the University of Nottingham in England created smart algorithms to help machines self-optimize during start-up, which achieved 50% reductions in ramp-up time.

Meanwhile, these techniques are infusing industrial robots, and also support safer working environments for human employees. As data from thousands of cases flows in, neural networks help the robots quickly learn and predict thousands of non-automated manufacturing tasks.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Rethink Robotics is using object detection classification and avoidance techniques to have its manufacturing robots, Baxter and Sawyer, perform safely in the presence of humans. In the industrial and manufacturing robotics space, collaborative robots like Baxter and Sawyer are becoming much more prevalent, with leading robot manufacturers like ABB, Kuka, and Yasakawa all using computer vision-based object avoidance techniques to have robots work safely alongside humans.

3.22.3 Sensor Data Fusion

When analyzing massive amounts of data coming from various sources, traditional analytics and machine learning are insufficient to extract full value because certain patterns or results go unseen. This is because humans setting the assumptions, while programming metrics, indicators, alerts, etc., are bound by the limitations of their experience and biases.

Deep learning is being applied to analyze correlations and dependencies, identify separate clusters, extract hidden patterns across massive amounts of diverse and unstructured data, both from within manufacturing environments and outside (e.g., weather, markets, etc.), and predict future events. As manufacturers are able to integrate, analyze, and learn from the wide range of sensors in factory and product settings, what was once opaque to manufacturers becomes visible at very precise levels.

Sentience is using deep learning for sensor data fusion to model and analyze highly contextual elements in consumer-facing markets using mobile sensors. The company monitors mobile sensors, in conjunction with other public data (e.g., road data, venue maps, weather, etc.) to provide individual profiles on drivers, both in the context of professional drivers and for usage-based insurance.

3.22.4 Manufacturing-Related Use Cases

The use cases for deep learning in the manufacturing industry are as follows:

- 3D printing arm control
- Machine/vehicular object detection/identification/avoidance
- Predictive maintenance (IoT, robotics, autonomous machines)
- Real-time video analytics
- Localization and mapping
- Sensor data fusion in machinery
- Voice/speech recognition
- Product lifecycle management

Reference Section 6.3.21 for Tractica's 2016 to 2025 forecast of deep learning in manufacturing.

3.23 Media and Entertainment

Companies in the media industry process information by producing, collecting, and distributing content in the form of books and newspapers, motion pictures and music recordings, and radio and TV programming, as well as social media and software. This industry has been highly affected by the internet, machine learning, mobile, and programmatic services, among other software trends. With advancements in neural networks for image recognition, language analysis, and prediction engines, the application

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

impact on media and entertainment businesses is already diverse and well underway.

3.23.1 Content Distribution on Social Media

Content distribution has seen a major shift from the traditional models of TV, print, and radio to social media and the web. Social media-based content distribution is becoming a powerful influence on how media and entertainment companies are thinking about their channels for content distribution. The ability to specifically target content and filter down to a target audience in terms of age, gender, location, lifestyle, interests, and social media connections is a powerful tool for digital media strategists. However, for the most part, content distribution on social media is still a gamble, with a high degree of experimentation and A/B testing required before the right target audience or social media channel is identified.

AI and deep learning are now being applied to help address testing, targeting, and better predict and recommend the most appropriate content, displayed in the best way, at the best time, at the individual user level, at scale. CNNs are being used to capture and automatically annotate images (both generic and of specific brand assets), curate photos posted by social media users, facilitate searches, and surface more user-generated content. Perhaps the most lucrative future use of deep learning will come with the advent of automated tagging of video, audio, image, and other unstructured data.

Echobox provides an AI-based solution to help online publishers predict how well their content will do once it is shared on social media. Echobox has been working with French publishers, including Le Monde, Le Figaro, and Liberation, the three biggest news sites in France. On average, by using Echobox, French news publications have seen a 51% increase in page views over Facebook and a 97% increase from Twitter. The Echobox algorithm works by scanning all of a publisher's news stories in real time and assigning a score to each, in terms of how well a particular story would do. The system can then create a schedule order than can optimize click-throughs on social media throughout the day.

Tractica believes that this is going to be the future of content distribution online, with a centralized repository or engine that automates content distribution for media and entertainment companies. In addition, such a repository may be able to detect spam, protect intellectual property, and incorporate media experiences across future domains, such as autonomous cars.

3.23.2 Human Emotion Analysis

Human emotions are difficult to capture and measure, not to mention understand. Emotion analysis aims to detect and recognize types of feelings (e.g., anger, fear, stress, disgust, surprise, sadness, happiness, etc.) through the expression of text, language, body, and facial cues. Advances in both computer vision and NLP make it possible for machines to understand human emotions. Deep learning, often RNNs, are applied in NLP for feature extraction, and for classifying words into features to understand semantics and sentence structures. With computer vision, image, motion, facial recognition, and video analytics are used to assess emotion. According to the National Institute of Standards and Technology, today's facial recognition algorithms are 100X more accurate than those used in 1995.

Affectiva is one of the leading companies that has successfully commercialized emotion recognition and is applying it to various vertical markets, including advertising, market research, and media and entertainment. Affectiva offers a cloud-based service that reads facial expressions, essentially offering "emotions as a service." Affectiva has an emotion database of more than 4 million faces from people across 75 countries, which amounts to more than 40 billion data points. Its software is able to discern the difference between a

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

smile, surprise, dislike, attention, expressiveness, and other emotions. The technology in use is called Facial Action Coding System (FACS), which identifies 21 facial expressions, mapping them to 6 emotions, e.g., disgust, fear, joy, surprise, anger, and sadness.

Hershey has used Affectiva to test a new in-store device called a Smile Sampler that prompts users to smile in return for a treat. Affectiva has also been used on Hollywood movies to test how the audience reacts to certain scenes in a movie. Another big area where Affectiva is being used is emotion-aware gaming. The bio thriller game "Nevermind," which uses Intel's RealSense technology combined with Affectiva software, is able to change gameplay based on biofeedback, and help enhance the game using emotion data during playtesting. Affectiva software is also being used in the area of eSports or live professional gaming where emotion sensing is used to gauge audience engagement, as well as the emotional state of players.

Eyeris is another company specializing in emotion analysis for embedded systems. The company is using CNNs to learn prototypic micro-facial expressions in real time using advanced cameras and sensors. Blinking, yawning, crying, holding the phone, happiness, drowsiness, and stress are a few of the many emotional cues the company is working to capture, analyze, and use as the basis for new services. Eyeris is partnering with automotive manufacturers to power "affective computing" contexts in which analytics will understand drivers' moods and actions and offer "reactive support systems." The company is also working in areas of social robotics, healthcare, and virtual assistance.

Apple recently purchased Emotient, another emotion recognition software company, suggesting that the technology could be embedded into iPhone and Mac products at some point, which could allow for a new generation of emotion-aware applications and services to be launched through the Apple platform.

While very promising in commercial media, market research, and advertising applications, consumer protection advocates also caution the industry about the role of emotion recognition (and related tagging) as related to personally identifiable information and privacy protection. Digitizing and tracking emotions is indeed uncharted territory for businesses, consumers, governments, and legal structures.

3.23.3 News Curation for Consumers

In an age of more information and more access to more information than ever before, the role deep learning plays in news delivery is a growing, controversial, and important one. For one thing, the method in which news consumers go about consuming their news is increasingly digital, in particular through social media.

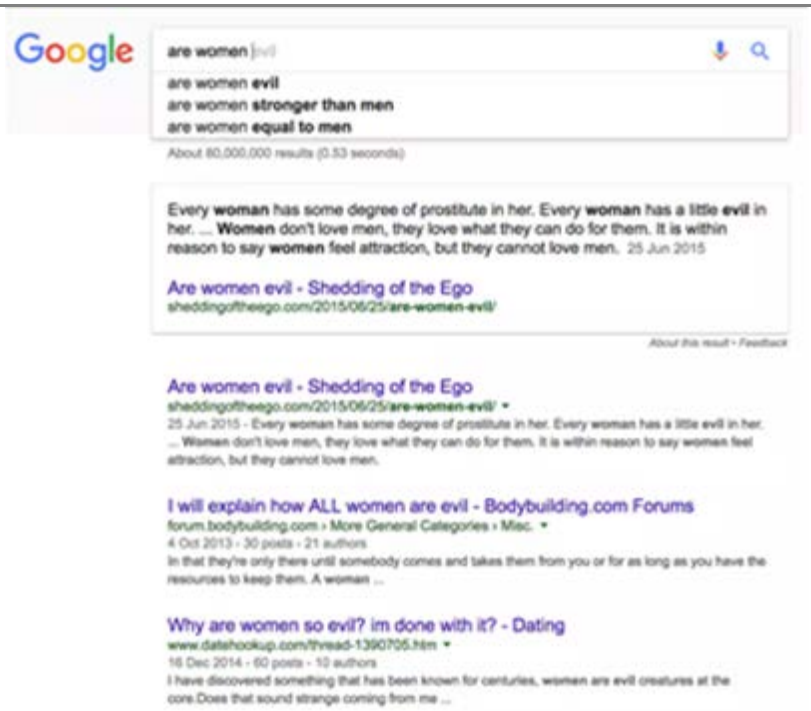
For news accessed on digital channels (e.g., news sites), search and advertising play a significant role in click-throughs and what content is served up in what order, sometimes even using different headlines for the same article. For news accessed via social media, content is typically driven by advertising and users' social graphs. The application of neural networks and search and content curation in social media introduce a host of efficiencies, but also significant issues. The contention lies in two areas.

First, advertising revenue is a fundamentally different incentive than public awareness. Social networks are transforming into publishing networks, with paid social media content on the rise. This means that access to news is at the mercy of advertising models. Many questions around this remain unanswered, such as how information, especially news, should or will be prioritized or deprioritized in favor of advertising, and who or what decides and monitors this.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

The second area to note is that of explainability (see Section 2.2.7); and the [in]ability to fully explain neural networks' decisioning. Two people can search the same query and receive entirely different search results; while this may be ideal in one scenario ("pizza near me"), it is problematic in other scenarios wherein informed decision-making has grave implications.

Figure 3.4 The Ordering of Search Results Has Influence



(Source: The Guardian)

3.23.4 MEDIA AND ENTERTAINMENT-RELATED USE CASES

The use cases for deep learning in the media and entertainment industry are as follows:

- Audio and video mining
- Film scene structure
- Font recognition and suggestions
- Gesture recognition
- Human emotion analysis
- Music production and generation
- News and feed curation for consumers
- Simulating crowds (films, games)
- Social media publishing and management
- Video editing

Reference Section 6.3.22 for Tractica's 2016 to 2025 forecast of deep learning in media and

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

entertainment.

3.24 OIL, GAS, AND MINING

The oil, gas, and mining industry is undergoing tremendous automation through sensor technology, analytics, networked services, and beyond. AI and, increasingly, deep learning have been introduced into every part of the mineral exploration process, including geology, geophysics, and reservoir engineering. This is especially effective here because the process is among the most data-intensive of any industry.

3.24.1 AUTOMATED GEOPHYSICAL FEATURE DETECTION

Geophysical feature detection is a critical part of the workflow in the oil and gas industry. Seismic surveys are carried out in the exploratory phase, and during various other phases from planning to field characterization before and during oil production. Once the data is gathered, the seismic traces are then processed and analyzed by human experts. Typically, this process can take several months.

Recently, Shell and MIT partnered together to use AI techniques to automate this process, and improve workflow efficiencies. Using deep learning, the raw seismic traces were analyzed to discover and locate subsurface faults in the underground structure, which are likely to contain hydrocarbons, before running migration and interpretation models. While there are still challenges in training and computational requirements, the study proved that geophysical feature detection could be automated.

Oil exploration capital expenditure is estimated to be around $100 billion per year, so any savings and efficiencies brought about by geophysical analysis is expected to be adopted widely across the oil and gas industry.

3.24.2 OIL, GAS, AND MINING-RELATED USE CASES

The use cases for deep learning in the oil, gas, and mining industry are as follows:

- Oil production optimization
- Automated report generation

Reference Section 6.3.23 for Tractica's 2016 to 2025 forecast of deep learning in oil, gas, and mining.

3.25 REAL ESTATE

Deep learning is being applied in real estate to better assess development opportunities. Using computer vision, developers are analyzing geographic images through drones and other technology to support models for valuation of properties and neighborhoods. In the past, property evaluation has been one of the most important parts of a real estate broker's job. With hundreds of variables to analyze to correctly determine the value of any parcel of real property, predicting real estate values is a perfect use case for AI.

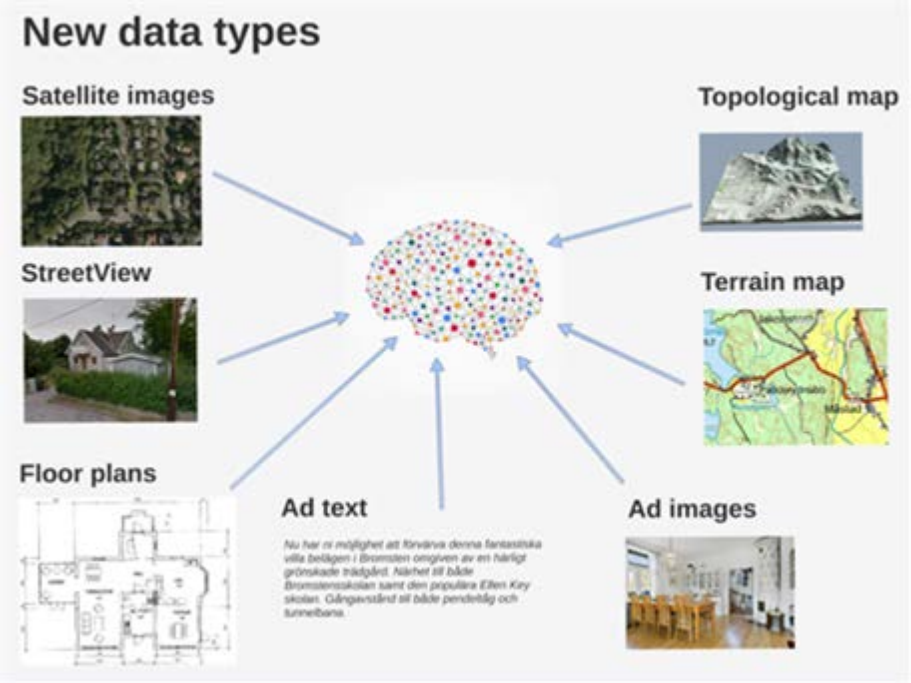
Seattle-based startup CityBldr has created a SaaS platform using AI to help determine the best use of all properties, to help developers find the most underutilized properties, and to help property owners understand the value of their properties as potential development sites by predicting what a developer might pay the property owner for the site's development value. The tool draws on 16 different public sources of data, including zoning codes, tax history, transit, and parcel data, and generates 3 proprietary data sources. In the end, it analyzes more than 180 variables on 118 million U.S. properties to determine how plots of

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

land can be improved to maximize their value.

Other applications include using deep learning to mine and extract key information from global contracts and real estate documents (e.g., leases, invoices, insurance policies, contracts, credit notes, etc.), sometimes in multiple languages. Data aggregation activities and particularly those involving unstructured data, such as in fiscal reporting, is another area where deep learning can be used to automate information extraction and expedite reporting. Leverton specializes in real estate data and document management, supporting large corporate real estate management companies.

Figure 3.5 Peltarion's Model Analysis Millions of Data Points to Product Real Estate Valuations



Swedish firm Peltarion uses deep learning to analyze data from historical data, object properties, demographic information, and nearby points of interest, using millions of data points to determine valuation. The platform is now used by most Swedish banks when administering home loans.

(Source: Peltarion)

3.25.1 REAL ESTATE-RELATED USE CASES

The use case for deep learning in the real estate industry is as follows:

- Real estate development optimization

Reference Section 6.3.24 for Tractica's 2016 to 2025 forecast of deep learning in real estate.

3.26 RETAIL

Since the emergence of e-commerce, retail has been undergoing digitization and automation in just about every capacity. From purely online applications, such as visual search-based e-commerce and in-store applications that bring the ease of "online" with real-life needs (e.g., clothes sizing and fitting), neural networks are a growing tool for retailers to gain intelligence

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

about their customers and improve their shopping experiences.

3.26.1 Crowd Analytics

Retailers have been monitoring and analyzing customer movement in brick and mortar environments for years, whether through video cameras, beacons, or other sensing technology. Deep learning introduces new capabilities to crowd analytics, not only involving image recognition and learning in computer vision-enabled cases, but also in using neural networks to analyze, learn from, and predict information about traffic patterns, in-store displays, energy allocation, etc.

ShopperTrak is using deep learning to pose hypothetical foot traffic scenarios by inputting data from specific days, product launches, promotions, weather patterns, or other contexts to model and predict foot traffic, both digital and brick and mortar. The model also uses back propagation to train itself over millions of simulations, by running predictions comparing outcomes to actual data, and then making adjustments accordingly.

Herta Security is a Barcelona-based company that uses deep learning for intelligent video analytics in malls, sports stadiums, airports, banks, and other retail environments. The company tracks and matches faces instantly and its system can be used to identify shoplifters and notify security personnel within 7 seconds.

3.26.2 Predictive Analytics for Retail

Retail is big business and includes a lot of Big Data. Bridging digital and physical worlds ("brick with click") demands analysis of extremely diverse and often unstructured data sets. Customer transaction and CRM data, browsing history, location data, sensor data, weather data, social media data, ad data, data from conversational commerce, and data across multiple websites are just a handful of the data retailers are mining to deliver highly personalized ads, product recommendations, marketing materials, purchasing options, campaigns, etc.

One of the greatest impacts of deep learning capabilities in retail is in supporting better analytics and real-time personalization through prediction and pattern detection. This includes customer-facing applications such as those listed above, as well as a range of operational, logistical, and even legal ones, including but not limited to:

- Identification of individual store trends (e.g., unmet demand)
- Customer churn prediction
- Store layout and inventory efficacy
- Fraud detection
- Anomaly detection

These tactics, often used in conjunction with machine learning, can be extremely powerful for retailers to better understand and predict every aspect of their businesses. Retailers must be highly cognizant of consumer privacy protections, particularly when involving facial recognition and/or third-party data sets to identify individuals. For example, some fast food restaurants today are using vision-based AI to reliably read license plates of cars passing their franchise locations, then combining this data with public third-party data to relate the license plate information to an individual, and based on that input and their recognized movement patterns, create hyper-personalized marketing communications.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.26.3 Retail-Related Use Cases

The use cases for deep learning in the retail industry are as follows:

- Behavioral analytics
- Crowd analytics
- Intelligent CRM systems
- Predictive analytics for retail
- Sentiment analysis
- Supermarket shelf analytics
- Visual search based e-commerce
- Weather forecasting

Reference Section 6.3.25 for Tractica's 2016 to 2025 forecast of deep learning in retail.

3.27 Sports

The term "sport," as used in contemporary sport management and in relation to the sports business industry, is a broad concept term used to denote all people, equipment and supplies, activities, businesses, and organizations involved in producing, facilitating, promoting, or organizing any activity, experience, or business enterprise focused on recreation, sports, or sports tourism. Deep learning has a number of applications in the sports industry, most notably specific to the optimization of predicting game outcomes, but also around player performance, team selection, play trajectories, and real-time video analytics in conjunction with computer vision.

3.27.1 Game Outcome Predictions for Betting

Predicting the outcome of sports, casino, or other games is a business worth an estimated $700 billion to $1 trillion a year. Businesses and individual fans alike place significant money on who will win in just about every type of sporting event, from football to horse racing to poker. Entire adjacent industries benefit from game outcomes, including sports tourism, sporting goods manufacturing, advertisers, content creators, recruiters, etc.

Deep learning is being applied in this context to augment game outcome predictions by analyzing diverse data sets and past historical data to "learn from" and more accurately predict who will win, and when.

Vantage Sports has partnered with New Data Sports in developing an algorithm for pick forecasting. The company tracks obscure metrics that other industry trackers like the National Basketball Association (NBA) do not measure, such as whether a pass was made to an open shot, or the number of times players contest shots (i.e., put their hands in the shooter's face to block the shot). The company tracks dozens and dozens of metrics around player behavior for every team. Vantage Sports recently conducted deep learning analysis on its data versus public data and found that insights analyzed from its data yielded a 54% positive rate for prediction compared to 49% of correct predictions using public data.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.27.2 Sports-Related Use Cases

The use cases for deep learning in the sports industry are as follows:

- Biomarker based athlete performance optimization
- Game outcome predictions for betting
- Sports statistics analysis and search
- Sports teams player selection

Reference Section 6.3.26 for Tractica's 2016 to 2025 forecast of deep learning in sports.

3.28 Telecommunications

The telecommunications industry covers a number of areas, including cable, wireless, switching, transmission, radio frequency (RF) and optical communications, media, and internet protocol (IP) networks. AI and deep learning are being applied broadly to this industry, in areas including but not limited to predictive maintenance, customer analysis, fraud prevention, network security, and beyond.

3.28.1 Predictive Maintenance

Current telecommunications services typically rely on heavy equipment, machinery, transformers, lines, boxes, poles, and a range of other infrastructure to maintain connectivity, reliability, and security. When parts go down, costs incurred are manifold: costs of machines, costs of maintenance required (i.e., labor, emergency rates), costs of downtime, and (often untold) costs of customer frustration and loss, especially when customers are businesses. The ability to manage so much capital outlay is critical.

Telecom and utilities providers are increasingly using deep learning to help address issues around predictive maintenance by analyzing massive amounts of data to better manage these risks. Deep learning is being used to preemptively diagnose *what could happen*; predict *what is likely to happen*; and even prescribe a recommended course of action or support. By analyzing metrics and data related to the lifecycle maintenance of technical equipment, companies can predict both timelines for probable maintenance events, anomalies, and upcoming capital expenditure requirements; recommend enhancements, allowing them to streamline their maintenance costs and risk mitigation; avoid critical downtime; and minimize brand damage.

3.28.2 Telecommunications-Related Use Cases

The use cases for deep learning in the telecommunications industry are as follows:

- Predictive maintenance
- Prevention against cybersecurity threats
- Improve customer experience management
- Fraud mitigation
- Intelligent CRM systems

Reference Section 6.3.27 for Tractica's 2016 to 2025 forecast of deep learning in telecommunications.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

3.29 TRANSPORTATION

Transport or transportation is the movement of people, animals, and goods from one location to another. Modes of transport include air, rail, road, water, and space. The field can be divided into infrastructure, vehicles, and operations. Transport infrastructure consists of the fixed installations, including roads, railways, airways, waterways, and canals, as well as terminals, such as airports, railway stations, warehouses, trucking terminals, and ports. Vehicles traveling on this infrastructure may include automobiles, bicycles, buses, trains, trucks, people, helicopters, watercraft, spacecraft, and aircraft.

Deep learning is increasingly growing in transportation applications, and overlaps in areas such as automotive and logistics, where sensor data fusion, weather forecasting, and computer vision are becoming key enablers for intelligent transportation systems. Classic deep learning use cases, such as object detection, identification, navigation, avoidance, search, and targeting, are essential functions for autonomous vehicles. Deep learning is also being used for vehicle network and data security analysis.

3.29.1 PREDICTING TRAFFIC DENSITY

One of the biggest applications for deep learning in the transportation space is in predicting traffic density and flow. In an age of smart and autonomous vehicles, accuracy in traffic flow is a key enabler of broader intelligent transportation systems. Improving accuracy in these systems is key for traffic operational efficiency, reducing carbon emissions, alleviating traffic congestion, helping road users make better decisions, and improving overall municipal efficiency.

Deep Drive is a research project out of the University of California at Berkeley that is currently using deep reinforcement learning in conjunction with microsimulation methods to support urban traffic optimization for future smart cities and smart cars. Using diverse data streams from both cars and cities (i.e., GPS-based measurements of speeds and delays, sensor data, static flow measurements with magnetic loops, Bluetooth re-identification, odometer data, and raw video feed), the project aims to support optimization and shared learning for autonomous vehicle traffic patterns. In other studies, even more data inputs from crowd-sourcing and social media are fed into models.

3.29.2 TRANSPORTATION-RELATED USE CASES

The use cases for deep learning in the transportation industry are as follows:

- Machine/vehicular object detection/identification/avoidance
- Predicting traffic density
- Sensor data fusion in machinery (ships, unmanned ships)
- Localization and mapping
- Vehicle network and data security
- Weather forecasting

Reference Section 6.3.28 for Tractica's 2016 to 2025 forecast of deep learning in transportation.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 4

TECHNOLOGY ISSUES

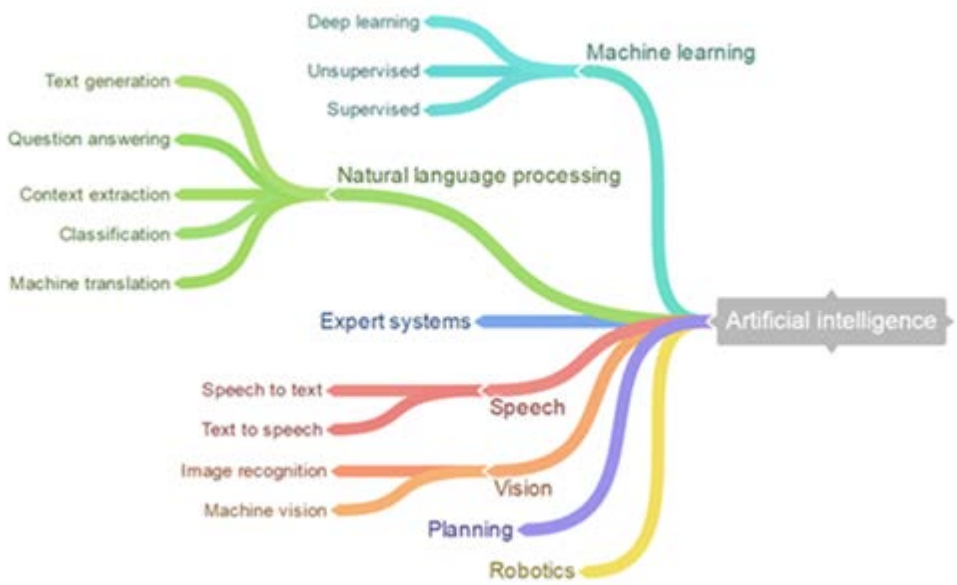
4.1 CONTEXTUALIZING DEEP LEARNING

Deep learning is part of the larger machine learning field and a subset of AI. Machine learning's founder defined it as a "field of study that gives computers the ability to learn without being explicitly programmed." Since its roots in the early 1900s, the field has grown to encompass multiple disciplines adjacent to computer science, including statistics, data science, mathematics, probability, neuroscience, and even design.

An important starting point when it comes to understanding deep learning is understanding where and how it fits into other technologies designed to increase machine intelligence. It is worth noting, however, that given the field's nascence and tremendous hype, border lines between technologies are often blurry and market constituencies are rarely in full agreement on where one ends and the next begins. Finally, the following overview of AI is geared toward conceptualizing deep learning, and does not encompass the full ontology of the field, as much is far beyond the scope of this report.

Artificial intelligence is a broader term that represents a variety of methods and tools that mimic cognitive *functions*, if not cognitive composition itself (see neural networks). Such cognitive functions might include planning, reasoning, problem solving, etc. Under the umbrella of AI, one finds a variety of approaches to simulate cognitive functions.

Figure 4.1 Artificial Intelligence Encompasses Numerous Technologies



(Source: Hacker Noon)

One is known as **knowledge-based systems** (KBS) or **expert systems**, in which large bodies of domain expertise are uploaded to computer memory. In this construct, cognitive tasks like learning or problem-solving are based on classic programming rules like "if-then"

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

and "inference logic" performed within the body of knowledge. The other, more broadly adopted approach is **machine learning**, in which learning occurs by the computer itself. Machines are programmed to extract relevant patterns from data and adjust program actions accordingly, derive predictions, and/or make recommendations.

From here, machine learning techniques are typically divided between **unsupervised** versus **supervised learning**.

- **Unsupervised Learning:** This occurs when untagged/disassociated data samples are fed to the system. Instead, the machine has to figure out how to structure the input itself based on statistical relationships of data. In this construct, there is no objective evaluation of the output structure; rather, clustering techniques, anomaly detection, and other methods identify both positive and negative deviations from "the norm." Unsupervised algorithms can draw inferences about new data based on inferences drawn from training data.
- **Supervised Learning:** This occurs when data samples introduced to the system are tagged or labeled, i.e., specific data are coupled with specific outputs. The output variables are known and often are linear; there is often a one-to-one ratio between inputs and outputs and, once the algorithm is trained, it can be presented with new data, make predictions based on that data (outputs), and adjust its outputs to new inputs. The goal is for the system to learn general rules that map inputs with outputs. (See Section 3.3.3 for reference to semi-supervised learning.)

Supervised learning methods include a range of techniques, such as decision trees, support vector machines, cross-validation, inductive logic programming, and most importantly for the purposes of this report, **(artificial) neural networks**. In neural networks, layers of neurons (information nodes) mimic the brain in that they are connected to each other not by axons, but by weighted edges that are constantly refined by mapping inputs to tagged outputs. When neural networks have multiple processing layers (usually between 10 and 100, although conceivably millions), they are considered "deep." Deep learning typically involves supervised learning, but methods are evolving to train deep neural networks on unsupervised data as well, particularly when it comes to processing signals from lots of tagged data as in the case of speech recognition and image processing. (See Section 3.2 for an in-depth description of deep learning.)

Deep learning can *support learning and prediction within, but can also exist independently from* a number of other segments of AI, including but not limited to:

- Natural Language Processing
- Speech Generation
- Computer Vision
- Computer Audition
- Intelligent Agent
- Robotics
- Strategic Planning

It is also useful to highlight is that deep learning or neural networks on their own have limits, and are often used in conjunction with adjacent techniques. While Google's/DeepMind's AlphaGo program (in which AI learned and then defeated the one of the world's [human] champions at the game Go) was an undeniably impressive demonstration of neural

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

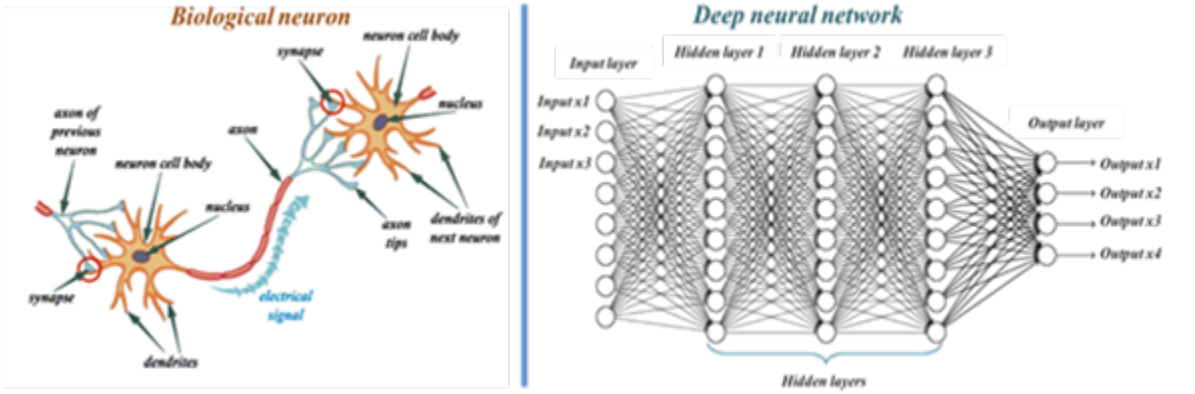
networks, it was in fact the product of a *hybrid* approach, in which a variety of classical techniques were used including tree search, genetic algorithms and reinforcement learning.

As a branch of machine learning, deep learning has gained greater prominence over the last few years, primarily due to the drastic improvement in performance seen in areas such as image recognition and speech recognition and the ensuing use cases made possible. The big breakthrough with deep learning came in 2012 when Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton from the University of Toronto won the ImageNet image classification competition where they achieved an order of magnitude reduction in error rates compared to previous years using a two-layer DNN. Also in 2012, Google saw a halving of its error rate for a cat recognition algorithm, with the algorithm able to identify cats by simply being fed 10 million cat videos from YouTube.

4.2 What Is Deep Learning?

Deep learning is a subset of AI that consists of artificial neural networks. It aspires to biomimicry in that its architecture roughly mirrors that of the human brain: information is processed through multiple layers to compute an outcome. In contrast to the biological brain wherein neurons can connect to *any* other local neuron, artificial neural networks have discrete layers and directions of data propagation. Unlike other machine learning algorithms, which only have one or two layers, deep learning is "deep" because it has multiple layers, typically between 10 and 100 layers.

Figure 4.2 Biological Neurons versus Artificial Deep Neural Networks

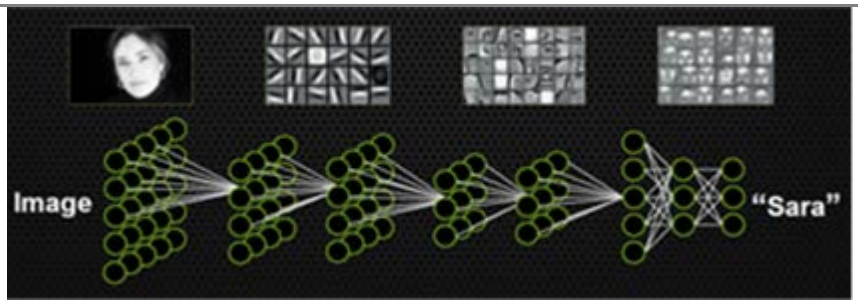


(Sources: Urantia.org; RSIPvision.com)

The inputs are combined together into an input layer. Each layer is responsible for the detection of one characteristic, and computations at each level base assumptions/build upon previous levels, which allows the network to "learn" more nuanced and abstract characteristics. From here, the deep learning algorithm creates hidden layer after hidden layer of neural nets. Layers on the top are more abstract, having learned and classified based on inputs from the lower layers. It has been shown that the performance of deep learning neural network algorithms is directly proportional to the number of layers. As the number of layers in the network increase, the performance has been demonstrated to increase on a linear scale.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Figure 4.3 Schematic Representation of a Deep Neural Network



This image depicts how deep neural networks capture more complex, nuanced features as data reaches deeper layers.

(Source: NVIDIA)

Deep learning networks are created through training, in which numerous sample data sets are shown to the algorithm. It is important to note that the algorithm itself is determining the different weights given to different inputs. No programming is required. The statistical relationships between the training data drive the structure of the neural net. As the algorithm is refined, connections are validated and weights of the connections between neurons are adjusted. Upon deployment, such "learned" connections enable the network to respond to examples without labels.

4.3 Practical Components of Implementation

When it comes to implementing deep learning technology, enterprises face a range of questions and considerations. Although open-source and more off-the-shelf platforms offer the technology to a wider audience, the reality is that deploying neural networks is a complicated endeavor at almost every level. Below are areas that Tractica's research found to be most critical in their potential impacts on deployment, use case development, and business model economics.

4.3.1 Criteria for Deep Learning Application

Deep learning algorithms are very good at dealing with very large data sets, or a complex set of inputs, which have multiple features that can be used to identify the data. In any deep learning or machine learning model, the AI is trained on specific features of the data, which allows it to learn specific characteristics of the data and to make predictions or correlations.

Deep learning is particularly suited for applications involving:

- **Domain Knowledge:** Tasks based on learning a large body of knowledge in order to formulate inferences and decision-making (e.g., legal, financial, medical)
- **Domain Extension:** Where complex domain knowledge is applied in order to extend or suggest new insights about the domain itself (e.g., information about existing medication and patients applied to new drug development)
- **Perception:** Tasks requiring perceptive capabilities, such as sight, sound, and touch, as well as motion detection, depth perception, tactile awareness, and other modes of "experiencing" physical realities (e.g., computer vision)
- **Communication:** Tasks that involve language-based communication modes, such as language processing, generation, automatic translation, or intelligent agents

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- **Complex Planning:** In which algorithmic models are able to detect patterns, predict, plan, and simulate based on ingestion of massive amounts of data
- **High Volumes of Unstructured Data:** Deep learning excels at identifying patterns in unstructured data, which is not organized in a pre-defined matter or does not fit in a pre-defined data model (e.g., PDF files, Word documents, audio, or video files)

For a thorough examination of deep learning use cases, organized by industry, please reference Section 4.

4.3.1.1 Applying Deep Learning for Feature Engineering and in Conjunction with Other Technologies

One of the major benefits of deep learning (used independently or in conjunction with other AI techniques) has been the reduction in time and effort spent on feature extraction and, subsequently, feature engineering. In contrast to rule-based systems, wherein systems engineers (with heavy physical, systems, and domain expertise) define features, deep learning minimizes or eliminates such efforts. This is a critical enabler for a few reasons:

- Feature engineering is energy, time, and cost-consuming
- It is not scalable given the massive and growing amounts of (real-time, IoT, unstructured, diverse) data generated, and therefore relevant to models
- Role of humans is to define core architecture, not manually account for new scenarios, parameters, or inputs

In the case of an image recognition or classification algorithm using deep learning, there could be millions of features that the model needs to account for or understand. In such instances, deep learning is known to provide better results, with the neural network learning about the features on its own, rather than having someone hand-engineer the features.

In a combination of deep learning and machine learning, one would use machine learning algorithms to perform basic clustering or regression learning types of tasks during which features of the data are handcrafted, while deep learning would be used on a larger or more complex set of data. For example, in the use case of content distribution on social media, machine learning algorithms could perform a basic clustering of the content types that need distribution, or use regression learning to identify the best social media channels for a specific type of content using historical data. But deep learning would perform a more rigorous analysis of the type of content, extract features from the video or image, and then perform machine learning to make suggestions for time of day, social media channels, geographies, demographics, etc.

It is important to reiterate that deep learning will increasingly be used in conjunction with other AI techniques, such as machine learning, NLP, computer vision, or machine reasoning.

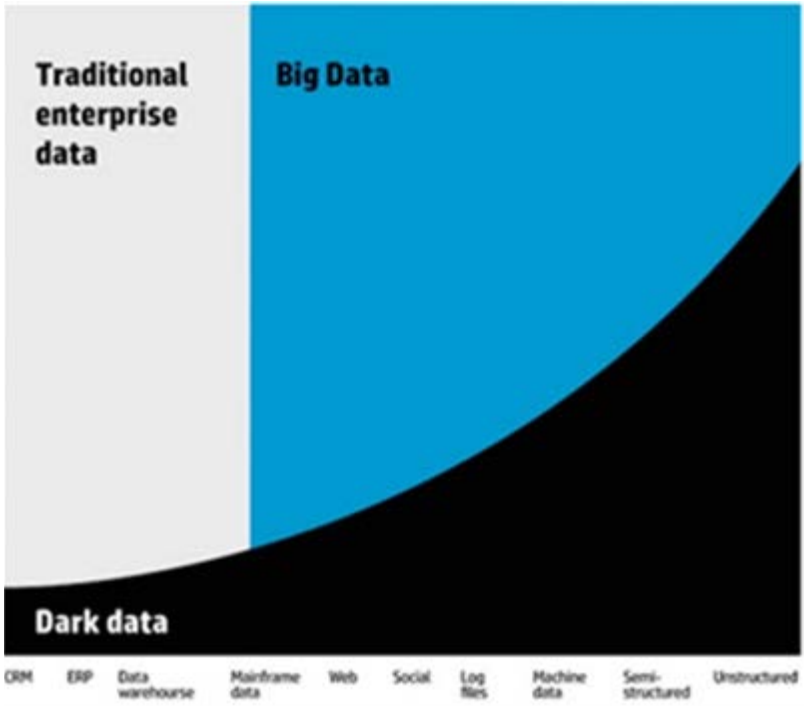
4.3.2 Technical Obtaining of High-Integrity/Accuracy Data for Training

One of the greatest challenges of implementing deep learning applications is obtaining quality data for training. “Algorithms do nothing without data, they’re just code,” explains Stuart Feffer, CEO of Reality.ai. Deep learning, in particular, is very data hungry: with each layer of the neural network, there are multiple nodes, and each represents coefficients that require statistically significant amounts of data to learn. The mathematics behind this are what enables neural networks to perceive abstraction in nuance, but also what demands greater data.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

The demand for data is not just one of volume, but of integrity and usability. Today, most enterprise data is dark data, which organizations collect and store, but never actually use. IDC estimates that 90% of unstructured data is dark data. Old versions of documents, images, presentations, qualitative feedback, email correspondences, financial statements, customer and account information, log files, and employee information are just a few examples of unstructured data, which Computerworld estimates accounts for approximately 80% of enterprise data.

Figure 4.4 Dark Data Accounts for the Majority of Enterprise Data



(Source: KDnuggets News)

The goal of harnessing dark data is *making sense of everything unable to be captured through structured data alone*; finding hidden patterns, improving business analytics, competitive intelligence, and decision making, identifying issues and opportunities to improve customer experience, energy efficiencies, product development, safety, and productivity. While deep learning offers unprecedented opportunity for organizations to "shine a light" on dark data, revealing nuances and insights that structured data simply do not, achieving this may be the most complicated aspect of implementation. Sometimes this data is not stored in relational databases; often, data is stored in multiple disconnected databases; for multi-national firms, such data may exist untranslated in many languages; and often, it exists in the deep web making it very difficult to search. "Deep learning helps redeem Hadoop; Hadoop stores unstructured data well, deep learning analyzes it well," summarizes Chris Nicholson, co-founder of SkyMind.

Because networks must be trained, significant high bandwidth-consuming computation is required, meaning that the learning phase is typically performed in data centers running non-stop. Information is weighted and optimized numerous times in order to minimize errors. The result of each training is gauged as a snapshot, with the expectation that each snapshot improves upon the previous one. This is an especially critical part of the development

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

process for applications involving HPEC environments, such as military or aerospace.

Another issue that arises given deep learning's "hunger for data" is that of data pooling. According to numerous companies Tractica interviewed, it is often the case that companies bringing deep learning to market are combining data from multiple customers in order to make bigger pools for training. While opportune in some industries, other industries like insurance avoid such pooling for fear of liability or enabling competitors.

In addition to data hunger, bigger data also signals the need for more training and associated costs of obtaining, cleaning, and storing data in the first place. There are techniques like one-shot learning and memory augmented neural networks that apply small data sets to neural networks and achieve good performance. Another technique that is likely to help with the problem of getting good data is generative adversarial networks (GANs), which are able to generate sample data which is almost as good as or better than real-world data. One experiment by Tim Salimans, Ian Goodfellow, and Wojciech Zaremba involved GANs using semi-supervised learning on image databases MNIST, SVHN, and CIFAR-10, in which the discriminator produced and additional output indicating the label of the input. This experiment achieved a 99.14% accuracy with only 10 labeled images per class with a fully connected neural network. While still early in development, these efforts are part of a larger push to use neural networks to mitigate intense costs of training agents.

Broadly speaking, the challenge inherent to obtaining the volume and integrity of data required to train models is that of proving value. It is difficult to foresee the full value of a system without a full build. As organizations struggle to understand deep learning, experimenting with the technology *without clearly defining the problem* is risky. Technology vendors warn against setting expectations too high at the outset, creating a sort of chicken and egg problem. Without starting with a premise, you will not know how valuable the data will be until you build out a solution. As one interviewee offered, "you don't want to spend millions of dollars to see that when it snows, people buy snow shovels."

For smaller data sets (and companies), Tractica's research finds a growing use for model repurposing, or the reuse of pre-trained deep learning models. Google is one such leader in this space, both externally with its TensorFlow offering and internally. A network built to recognize cats can be repurposed and trained using CT scans to recognize specific types of cancer. One engineer on Google's Translate team recently took a network used to judge artwork and applied its computer vision learning to drive an autonomous radio-controlled car.

4.3.3 Human Supervision and Semi-Supervision

Beyond the obvious role of humans in strategic planning, departmental alignment, procurement, and other governance-related planning, the role of humans in deep learning development and deployment (as well as ongoing maintenance) is extensive.

During the implementation process, significant labor hours are required for data cleansing, grooming, scrubbing, and processing. Across the board, enterprises agree that continuously training and deploying updated models is one of the greatest challenges. The goal of the training for a neural network is to define parameters for functions that enable the network to identify and recognize patterns and assign value. Selecting a deep learning framework instead of building one from scratch) can save tremendous amounts of time and money. Although deep learning can function in certain unsupervised settings, humans are essential for the initial "phase zero" of deployment: ensuring the highest quality and highest integrity (and low bias) data as possible.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Ongoing efforts to maintain updated, safe, compliant, and reliable software, and to manage inferences, will also require humans for the foreseeable future. Models must be constantly updated and trained; "one of the biggest imperatives enterprises overlook," says Kumar Srivastava, Vice President (VP) of Product & Strategy at BNY Mellon Silicon Valley Innovation Center. The lifecycle of a neural network reflects that of a human; the training of the network is the most education-intensive part of its life as it sets a foundation (if an incomplete one) for accomplishing cognitive tasks later in life. Taking this analogy a step further, Luca Rigazio of Panasonic offers, "as humans we are able to reason about the future, but then we go out in the world and gather additional information to validate and improve ourselves. That kind of virtuous cycle of experience and learning mastered by the human brain, is not yet fully possible for machines." Therefore, humans are required to facilitate this.

Although most deep learning today is trained through supervised learning, most expect R&D to pave the way for deep learning to more easily handle unsupervised learning in the future. In the meantime, **semi-supervised learning** is emerging, in which models make use of both [a small amount] of labeled data with a large amount of unlabeled data. Research in this area shows that the use of unlabeled data, (typically less expensive to obtain,) alongside minimal amounts of labeled data, can considerably improve accuracy (and, potentially, development costs). This method also shows promise to increase computational efficiency, whereby adding a decoder triples computation during training, but not actual training *time*, because labeled data expedites the utilization of available information.

Of course, "humans in the loop" as a term is typically coined in the context of human-machine partnership or "enhancement." In this context, employees or consumers play the essential end-user role in model training simply by engaging with a dashboard or using the service. Doctors leveraging diagnostic recommendations, for example, are more likely to be able to immediately influence accuracy than a developer. Call center agents using service applications inadvertently influence algorithmic learning simply by registering customer outcomes.

4.4 Various Hardware and Software Configurations of Deep Learning

4.4.1 Deep Learning Frameworks and Development

A growing number of frameworks are being developed in order to facilitate the mathematics involved in training networks. It is a sort of middleware for cognitive computing, which implements basic training and provides the model to be used directly by an application or used directly with other cognitive middleware (e.g., implementing vision tasks or recognizing natural language). Deep learning frameworks are growing in access and variety, with dozens of new additions submitted to open-source communities every week. The following figure 4.5 is a roundup of some of the most commonly used deep learning frameworks.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Figure 4.5 Common Deep Learning Frameworks

Name	Features	Creator	Language	Applications
Caffe	• Built for speed and modularity • Not general purpose • Expressive architecture + extensible code encourages applications • Large user community "Caffe Zoo"	• Berkeley Vision and Learning Center (BVLC) and by Caffe User Community	• C • C++ • MATLAB • Command Line Interface	• Academic research • Industrial applications in speech, multimedia, especially computer vision
Caffe2	• Built on Caffe but lighter weight • Built for expression, modularity, and speed for quick demo applications • Large user community "Caffe Zoo" • Cross-platform libraries	• Facebook	• C++ • Python	• Demonstration and production, both cloud and mobile
TensorFlow	• For research and deployment of large scale machine learning models • Minimal/low level additional frameworks and custom code often required • Comparatively slower	• Google (Google Brain)	• C++ • Python	• Used extensively for Google • Supports scaling of deep learning to clusters
Torch	• Built for speed, flexibility, and modularity • Collection of neural network libraries • Facebook open-sourced its deep learning modules and extensions for Torch	• EPFL (Swiss University)	• C • C++ • Lua	• Academic research • Industrial applications in vision, speech, multimedia, gaming • Used in Facebook Research
Theano	• General purpose framework Minimal/low level; • additional frameworks and custom code often required • Extensive documentation	• University of Montreal	• Python	• Academic research • Strong for data exploration and multi-dimensional arrays • Low level for other Python packages (e.g. Blocks, Keras, Lasagne, OpenDeep, etc.)
Keras	• Built for simplicity • High level library configurable on top of Theano or Tensorflow • Modularity • General Purpose	• François Chollet	• Python	• Prediction, feature extraction • Image classification
MATLAB	• Broader machine learning library and multi-paradigm numerical computing environment, offering numerous deep neural networks specific tools • Ex. Matlab Deep Learning, DeepLearn Toolbox, Lightnet, Deep Belief Networks, MatConvNet, Matrbm, Deepmat, and many others	• Mathworks	• C, C++ • C# • Java • Fortran • Python	• Image classification, transfer learning, • Computer vision
Cognitive Toolkit (CNTK)	• Build for vertical and horizontal scaling • Includes feed forward DNNs, CNNs, and RNNs • Uses a graph model • Efficient/fast • Can scale across many GPU-based machines • Supports feed-forward, convolutional, and recurrent networks for speech, image, and text workloads	• Microsoft (Microsoft Research)	• C++ • C# • Python • Command Line Interface	• Speech, image, text recognition • Used in Cortana, Bing and Skype Translator
Deep Learning 4J (DL4J)	• Extensive documentation • Integrated with Hadoop and Spark • Customizable at scale • Plug and play • Distributed CPUs and GPUs • GPU support for scaling on AWS • Adapted for micro-service architecture	• SkyMind	• Java Clojure Scala	• Industrial applications • Business environments • Used by Accenture, Booz Allen, Chevron, and IBM
Mxnet	• Built for efficient memory usage and mobile processing • Flexible programming model • Native Distributed Training Supported • Portable from the Cloud to the Client • MultiLingual	• Distributed Deep Machine Learning Community	• Python • R • C++ • Julia	• Used in Amazon's reference library for deep learning • Speech recognition • Computer vision • Industrial applications in image recognition
Paddle	• Flexible - supports a wide range of neural network architectures and optimization algorithms • Efficient • Scalable - easy to use many CPUs/GPUs and machines to speed up your training • Easy to deploy	• Baidu	• C++	• Ranking search results • Targeted advertising • Click-through rates prediction • Image classification • Detecting computer viruses • Recommendation • Translation

(Source: Tractica)

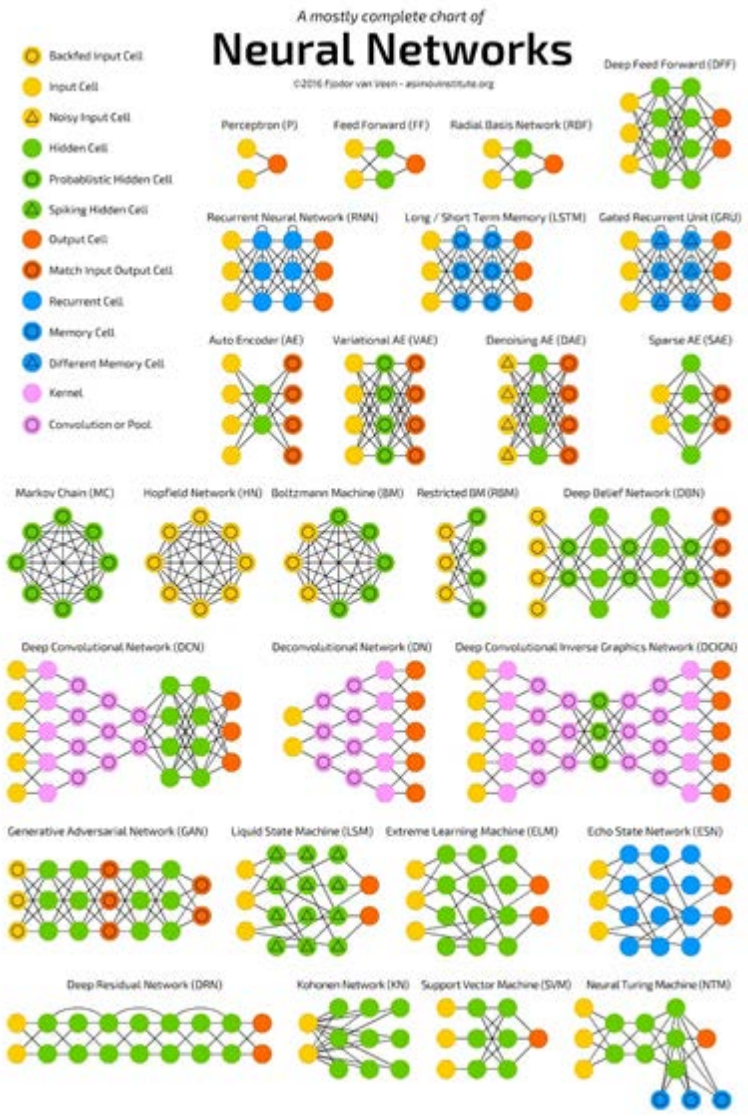
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

A variety of different neural network configurations fall under the deep learning umbrella for consideration based on use case. Four of the most commonly found types of software-defined neural networks include:

- DNNs
- CNNs
- RNNs
- Long Short-Term Memory (LSTM) Networks

A more comprehensive summary of neural network types appears in Figure 4.6 below.

Figure 4.6 Summary of Neural Network Types



(Source: The Asimov Institute)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Although the majority of implementations rely on a small handful of these (primarily CNNs, RNNs, and LSTMs), enterprises should also remember the pace of innovation is likely to drive down computational costs and limitations associated with diversification. “The pace of innovation and proliferation of ideas around neural networks will erase a lot of the limitations we face today,” says Anthony Stevens, Offering Manager for IBM Watson’s Deep Learning solution. “The cutting edge of neural networks is occurring where different neural network types are being combined together. Over the next 5 to 10 years, we expect major breakthroughs as the industry will expand from 5 or so common neural networks to at least 20, which can then be combined.”

Despite significant barriers in talent, most middleware, including deep learning frameworks and configurations can be *accessed* by any fluent developer; an individual, academic, enterprise, or otherwise. When implementing deep learning, enterprises should consider resources required for internally developed architectures using pre-built frameworks and packages versus deploying those “out-of-the-box,” as enabled by technology vendors. Consider the following options:

- Use case definition and prioritization
- Resources (talent, time, energy, data grooming, etc.)
- Difficulty and complexity to build, weighed against business outcome
- Vendor lock-in, multi-vendor management
- Integration with existing technologies
- Data pooling needs or restrictions
- Speed to pilot/experiment
- Scalability

4.4.2 Chip and Device-Based Deep Learning

Since 2009, GPUs have been playing a significant role in the growing application and accessibility of deep learning. GPUs have been the dominant hardware platform for deep learning applications and will continue to drive advances, especially for high-performance deep learning systems. Although NVIDIA dominates this market, competitors are beginning to emerge. Recently, AMD introduced Radeon Instinct, a GPU dedicated not to gaming, but to machine intelligence. Instead of powering high-power graphics, computing power is used to build neural networks’ underlying machine learning applications. The chip uses an open-source library for GPU accelerators called MIOpen.

At the same time, Tractica sees the emergence of alternative HPEC and hardware platforms like FPGAs, ASICs, and specialized processor architectures as potentially competing with GPUs on various combinations of performance, cost, and power consumption. Startups like Graphcore, Wave Computing, and Knupath are introducing new neural architectures customized for AI and deep learning applications. While some of the aforementioned companies remain in the early stages with very low or no actual market share at this time, their emergence signals a notable development in the deep learning chipset landscape as it moves forward.

Established semiconductor vendors are, in some ways, constrained by their legacy silicon architectures, providing room for new startups in deep learning hardware. As AI algorithms change to account for applications like autonomous driving or personalized medicine with dynamic inputs, there is a case for having memory storage on the processor itself. The

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

evolving nature of algorithms and workloads will determine which architecture is best suited for deep learning.

- **FPGAs** are integrated circuits that run programmable logic components supporting very specific (custom) tasks. They may support "soft" hardware, as they can be re-written after manufacturing. They run algorithmic processing quickly and tend to use less energy than general purpose CPUs.

- **Digital Signal Processors (DSPs)** are specialized processors designed for operational needs of digital signal processing.

- **ASICs** typically target specific applications by emulating logic through gates while offering little programmability beyond their core function. They are also not able to be reprogrammed once in the field. Tractica's research finds that the current state of ASICs in deep learning constructs renders it something of a misnomer, considering such gates on chips would need to expand beyond a certain number of (neural) layers and, currently, gates are limited on ASIC chips.

- **Tensor Processing Units (TPUs)**: Google developed its own ASIC specifically for machine learning and its TensorFlow framework. TPUs handle higher volumes of reduced precision computations, without hardware for texture mapping.

- **Application-Specific Standard Processors (ASSPs)** are integrated circuits that run a specific function applicable to a wide market, such as video or audio encoding. Unlike ASICs, which are designed by or for specific companies, ASSPs are sold off-the-shelf.

Processors will increasingly be "right-sized" to align capabilities and cost with specific workloads, and whenever possible, inference will be moved to edge devices for most applications. Inference in the AI context refers to using a trained model to make inferences or provide outputs on real-world data. This is usually done at the application or client end point, rather than on the server or cloud. Inference requires fewer hardware resources, and depending on the application, can be performed using a CPU or non-specialized hardware. This could be FPGA, ASIC, DSP, or other processor architectures.

What so much activity in this area signals is increasing demand for local processing. With more and more mobile and IoT devices permeating every industry, the amount of data and the requirement to efficiently process this data requires agility at the firmware level, rather than relying exclusively on the cloud or on-premise solutions. Critical applications, such as autonomous vehicles, military, healthcare, or industrial devices, will simply require real-time edge processing to be viable, in a way search queries or web-based applications do not. Over time, Tractica expects that other chipset types (including FPGA, ASIC, DSP, etc.) will grow within the market for inference applications. That said, given the velocity and variety of deep learning application development, there is enormous growth opportunity for both GPUs and CPUs.

"In the not so distant future," explains Anthony Stevens, Offering Manager for IBM Watson's Deep Learning solution, "we'll see devices running on a combination of GPU, CPU, and neural network chips with specific models trained to run on specific devices in order to support processing requirements and agility."

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

4.4.3 An Increasingly Open-Source Market

Open sourcing deep and machine learning capabilities characterizes the state of the market, one that is led by enterprise giants and is advancing quickly. The primary reason these giants are open sourcing is to drive adoption among enterprises, that is, to "recruit" talent and applications using the configurations and tools on which they are building.

While open sourcing research and architecture innovations has not always been a common strategy of such enterprises (e.g., Apple), the pivot toward open-sourcing signals that big technology companies are not just competing with each other directly, they are also competing to coalesce communities around their technologies at the infrastructural level. Google, Microsoft, IBM, Amazon, Facebook, Baidu, and others are all open sourcing various components of their technology up to developers. (See Section 4.4.1 for a list of frameworks, many of which are associated with specific technology companies.)

Vendor lock-in is a particular issue in this space considering the massive amounts of AI data to which models require access. Considering Big Data is a natural resource for developing algorithms, some companies are open sourcing data as well in order to accelerate AI applications for all.

4.5 Will Deep Learning Enable a New Programming Paradigm?

The longer-term implication and disruption of AI is its potential to fundamentally shift how organizations develop software. Today, 99% of software is developed in a rules-based deterministic sequence, based on step-by-step instructions and static parameters like binary logic, distance, geography, time, and beyond. This creates a sense that things are the product of their underlying instructions.

What machine learning and AI enable is a shift from deterministic to predictive, wherein models are dynamic and based on history and learning of patterns. As primitive a replication of the brain as they are, deep learning systems do represent a radical departure from past computer systems. In a conventional system, a human instructs a program what to do with data. Deep learning introduces the opposite: an outside-in view in which code does not just determine behavior, behavior also determines code. Their structure is determined by the data that they are given, not by the instructions from a programmer. Instead of human-generated code defining the entire program, outcome, or experience of technology, it will, as Oren Etzioni, CEO of the Allen Institute for Artificial Intelligence, noted, become the "scaffolding." This turns traditional design systems on their heads.

The transfer from human to machine design is also compounded by the issue of explainability in which our ability to understanding the inner workings of deep neural networks is largely absent and, at best, opaque. (See Section 2.2.7 for a deeper discussion on explainability.) To the extent we humans are able to peer into deep neural networks, what we find is an enormous multilayer set of calculus problems that are constantly deriving relationships between billions of data points and estimating outcomes.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 5

KEY INDUSTRY PLAYERS

Following is a summary of key industry players representing some of the most significant movements, investments, pilot projects, and incumbents in the deep learning market today. A table of additional players follows these profiles.

5.1 AI-ONE

Year Founded: 2003
Ownership: Private
Headquarters: La Jolla, California
CEO: Walter Diggerman
Number of Employees: 25

Company Profile:
ai-one, founded in Zurich in 2003, develops, licenses, and supports software development kits (SDKs) that enable developers to build machine learning applications. It typically works with clients to develop solutions using deep learning techniques at the application level as part of specific workflows.

ai-one's main product, Nathan, is a neural network that enables developers to build AI into almost any application. The core "brain" is a holosemantic dataspace (HSDS) that detects the contextual meaning of data by detecting patterns. The current version of Nathan learns the meaning of any text or unstructured data. Another important offering is the Analyst Toolbox, an end-user application for text analytics that can be cloud-based or on-premise and provides API access for workflow automation. The engine of the Analyst Toolbox is the BrainDocs application, which provides the platform for developers to expedite proofs of concept, by processing document libraries, building agents, and analyzing results. Other offerings include Topic-Mapper, a solution for processing language in the form of text or data; UltraMatch, a solution for image analysis and image matching; and ai-one Graphalizer API, a solution for working with data in the form of sensor data, such as financial trading, biometric, and industrial process control data.

Use cases for ai-one's technology offerings include intelligent agents for data mining and classifications, deep learning systems, communication agents using natural language generation, and auto-tagging and classification of unstructured text. Currently, the company's licensees offer products in semantic search, insurance, airline security, retail, forensics, aerospace, and computational biology.

5.2 AMAZON

Year Founded: 1994
Annual Revenue: $135.99 billion (2016)
Ownership: Public (AMZN)
Headquarters: Seattle, Washington
CEO: Jeffrey Bezos
Number of Employees: 341,400 (2016)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Company Profile:
Founded in 1994, Amazon.com is an international e-commerce website for consumers, sellers, and content creators. The company is the largest e-tailer in the United States. AWS offers more than 70 services for computer, storage, databases, analytics, mobile, IoT, and enterprise applications. AWS has grown rapidly and dominates the cloud services market.

Amazon AI's stated philosophy is "high quality, best-in-class deep learning systems with deep functionality." Amazon has long used deep learning to power its product recommendation system, as well as its Alexa assistant. Moving forward, it will also be used in image recognition, text-to-speech, and natural language and speech recognition. An upcoming application of deep learning may be Amazon's proposed Amazon GoStore, where customers will be able to walk in, pick up items they want to buy, and walk out. Amazon Go will use sensors, computer vision, and deep learning to track customers and automatically register which items get picked up. Amazon uses deep learning within its own services and it also offers access to its deep learning technologies through AWS. The company aims to make it easy for developers to run their own deep learning and machine learning workloads to build AI platforms on top of AWS.

Amazon Machine Learning (AML) is a managed service for building machine learning models and generating predictions. It uses the same technology as Amazon's internal algorithms. AML offers visual aids and analytics to assist in making models. Once the models are created, AML provides APIs to obtain predictions from them. In November 2016, MXNet was announced as the officially adopted deep learning framework, citing its horizontal scaling capabilities. The company released its own deep learning framework, DSSTNE (pronounced "destiny"), as open-source software in May 2016. The framework is for production only, not for general purpose, research, or testing, and is specialized for search and recommendation. DSSTNE is designed to work with sparser data sets than TensorFlow and other deep learning frameworks.

In November 2016, Amazon launched three new AI services. Amazon Rekognition, which can perform image recognition, categorization, and facial analysis; Amazon Polly, a deep learning-driven text-to-speech (TTS) service; and Amazon Lex, a natural language and speech recognition program. Rekognition was made possible through Amazon's acquisition of deep learning startup Orbeus in April 2016.

5.1 APPLE

Year Founded:	1976
Annual Revenue:	$215.64 billion (2016)
Ownership:	Public (AAPL)
Headquarters:	Cupertino, California
CEO:	Timothy Cook
Number of Employees:	116,000 (2016)

Company Profile:
Founded in 1976 by Steve Jobs, Steve Wozniak, and Ronald Wayne, Apple Inc. is one of the largest technology companies in the world. Apple designs and sells a variety of consumer electronics and services, including computer software, personal computers (PCs), online services, media players, smartphones and wearable devices. Apple is an avid protector of customer privacy and, to date, has been doing most of its machine learning on devices, rather than in the cloud. Specific ways Apple is using deep learning include Siri, fraud detection, extending battery life between charges on devices, choosing news stories, recognizing faces and locations in photos, etc.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

The company has made a plethora of AI acquisitions over the past 2 years, deepening its commitment in this technology area, although relative to its competitors, many perceive Apple as a latecomer to the field of deep learning. Within the last 2 years, Apple has been ramping up efforts through acquisitions, AI executive hires, and signals toward a more open approach to sharing research findings.

The most recent purchase was Turi. Turi lets developers build apps with machine learning and AI capabilities that automatically scale and tune. Other AI acquisitions include Tuplejump and Emotient in 2016, as well as Perceptio and VocalIQ in 2015. In October 2016, Apple hired its first director of AI research, Ruslan Salakhutdinov, an associate professor at Carnegie Mellon University in Pittsburgh.

Apple has deployed a suite of Basic Neural Network Subroutines (BNNS) designed to make it easier to build AI applications for mobile and desktop devices. This API is a collection of functions to construct neural networks on a wide range of Apple OSes. Apple has also introduced a new photo application with deep learning used to help search for images, put group-relevant photos into albums, and gather photos, videos, and locations into mini snapshots.

Compared to its competitors, Apple has been (not uncharacteristically) tight lipped about its activities and foci in the AI and deep learning space. In December 2016, at an invite-only lunch, it revealed a number of areas of image processing, activity recognition, intelligent assisting and language modeling, health and vital signs analysis, and, most notably, volumetric detection of LIDAR and prediction with structured outputs—both key technological elements for autonomous driving. Another announcement made was Apple's claim that it can build neural networks that are 4.5X smaller than the original neural network with no loss in accuracy and twice the speed. It does this by using a large, robust neural network to teach the smaller "student" network decisions it would make. Such lightweight replication and efficiency is critical for Apple to continue leadership in mobile hardware. Maintaining such processes locally on the phone means data does not have to be encrypted and sent over wireless networks. It also announced that its algorithm efficiency running on GPUs was so strong, it could process twice as many photos per second as Google's (3,000 compared to 1,500, respectively).

In October 2016, Tim Cook announced plans to open a new R&D base in Yokohoma, Japan. This will supplement its other R&D bases focused on AI. The CEO said the new Japanese R&D center would be for "deep engineering." That same month, the company announced it will allow its AI teams to publish research papers for the first time, and the subsequent publication of its first public AI research paper entitled: *Learning from Simulated and Unsupervised Images through Adversarial Training*. Pressed on its openness, Apple recently stated it would begin to publish its research more openly and work more collaboratively with the broader research community, although details were not provided.

5.2 BAIDU

Year Founded:	2000
Annual Revenue:	$10.16 billion (2016)
Ownership:	Public (BIDU)
Headquarters:	Beijing, China
CEO:	Robin Li
Number of Employees:	43,500 (2016)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Company Profile:
Baidu was founded in 2000 and is the leading Chinese-language internet search engine (Baidu.com). The company possesses about 80% of the search market in China, handling 6 billion searches a day. Baidu is a leader in the deep learning space and the company prides itself in having deep learning fully integrated into its company, and taking on a significant role in AI and deep learning research. Baidu's deep learning technology is used in computer vision, image processing, NLP, etc.

PaddlePaddle is Baidu's deep learning framework, based primarily on C++ and open-sourced in August 2016. PaddlePaddle is being used by more than 30 of its offline and online products and services, covering sectors from search to finance to health. It also helps work out how long food will take to reach customers in Baidu's takeout business. Baidu uses it for predicting click-through rates, classifying images, optical character recognition, ranking search results, and detecting computer viruses.

AI work is done under Baidu Research, until recently with well-known AI evangelist and pioneer, Andrew Ng as its chief scientist. Research areas include image recognition, speech recognition, NLP, robotics, and Big Data. Baidu Research is composed of three interrelated teams: Big Data Lab, the Institute of Deep Learning (IDL), and Silicon Valley AI Lab. The IDL's focus areas include image recognition, machine learning, robotics, human-computer interaction, 3D vision, and heterogeneous computing.

Baidu Research's DeepBench, is the first open-source benchmarking tool for evaluating deep learning performance across different hardware platforms. DeepBench will help organizations of all sizes identify the optimal hardware platform for their deep learning applications. Melody, an AI-powered conversational bot, which incorporates deep learning and NLP to provide relevant information to doctors, was introduced in October 2016 and the company publicly released Chinese language APIs containing a deep learning speech synthesis tool in November 2016.

The next frontier for Baidu's deep learning application is in devices and robotics. In November 2016, China Unicom and Baidu agreed to partner on mobile internet, AI, and other areas amid plans for ownership reform of state-owned enterprises. In January 2017, Baidu released its AI OS DuerOS. Baidu is using DuerOS to power a home robot, Little Fish.

5.3 CLARIFAI

Year Founded:	2013
Ownership:	Private
Headquarters:	New York, New York
CEO:	Matthew Zeiler
Number of Employees:	33

Company Profile:
Clarifai was founded in 2013 and is an image and video recognition technology provider, available as APIs, that lets computers "see." The company was founded by Matthew Zeiler, who studied under Geoffrey Hinton, a noted expert in computer neural networking with the Machine Learning Group at the University of Toronto.

Clarifai uses both machine and deep learning for organizing photo collections, searching large untagged image collections, and targeting ads to images. The platform, which has been fully operational for 3 years, can now "automatically" understand video and image. It recently launched a custom training feature in which users can fully customize and train the platform to recommend anything they care about within a few seconds—a process that used to take

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

weeks. Its technology is used in e-commerce, online communities, online advertising, stock photography, travel, real estate, weddings, medical imaging, healthcare, media and publishing, and digital asset management. Customers include Unilever, Trivago, Ponds, Curalate, 500px, Disqus, and BuzzFeed.

Currently, users can sign up for free with limited API calls. Upon passing a threshold, paying users are then metered by use, with lower rates for higher scale. The company has also branched out to offer custom training, visual search services, and a photo storage app for iOS. Clarifai does not plan to make its software available for companies to run in on-premises data centers.

Clarifai has raised $41.3 million in two rounds. The latest round was $30 million in 2016. This round will be used to grow all business and research functions.

5.4 DECLARA

Year Founded: 2012
Annual Revenue: $3.64 million
Headquarters: Palo Alto, California
CEO: Ramona Pierson
Number of Employees: 100

Company Profile:
Founded in 2012, Declara uses machine learning and NLP to personalize and enhance individual and organizational learning. The core of Declara's technology is its CognitiveGraph platform, which analyzes and observes how users interact with data, which could include searches, tweets, posts, blogs, videos, likes, recommendations, messages, and web content. This analysis is used to create a personalized learning profile based on a user's data, history, and interests.

While Declara's past focus has been on the education market, it also offers platforms for large corporations and government departments. In August 2016, Declara was selected as the platform for California's educator collaboration network (Collaboration in Common), serving all 500,000 teachers, educators, and administrators in the state. There have also been national-level implementations with teachers across Australia and Mexico, and Declara was chosen as the platform for the Puerto Rico Department of Education. Declara has raised $31.1 million in funding in 3 rounds. The latest round was in 2014.

5.5 DEEP INSTINCT

Year Founded: 2014
Ownership: Private
Headquarters: Tel Aviv, Israel
CEO: Guy Caspi
Number of Employees: 65

Company Profile:
Founded in 2014, Deep Instinct specializes in cybersecurity protection using deep learning. Its staff includes veterans of special cyber units in the Israel Defense Forces, deep learning researchers, and executives from global security companies. The company claims to be the first to apply deep learning to cybersecurity, and targeting U.S. Fortune 500 companies, cybersecurity is currently its sole focus.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Deep Instinct uses deep learning technology to offer real-time detection and prevention of zero-day threats and advanced persistent threat (APT) attacks on mobile devices, endpoints, and servers. The company does not use open-source deep learning libraries; rather, it has developed its own proprietary models to help recognize malicious code and detect malware. Because prediction models run on the device level in real time, the technology is also able to take action (e.g., deleting, blocking, quarantining malware) before it is too late. In addition to real-time detection, Deep Instinct continuously trains its engine so it is able to predict and recognize new types of malware. This improves the prediction model and gives a higher level of confidence in identifying malicious files. Its software can be run on a networked server or in standalone mode.

In partnership with FireLayers, the company also offers cyberthreat real-time detection and prevention solutions for enterprise cloud applications based on deep learning. According to Israel Venture Capital Research, the company has raised $35 million in funding.

5.6 DIGITALGENIUS

Year Founded: 2013
Ownership: Private
Headquarters: London, United Kingdom/San Francisco, California
CEO: Mikhail Naumov
Number of Employees: 27

Company Profile:
DigitalGenius was founded in October 2013 and originally focused on being a chatbot company for brands. The company quickly realized the limitations for standard NLP and has since shifted to machine and deep learning. The company believes in using human intelligence and AI together, thus coming up with its Human+AI Customer Service Platform. The platform serves human agents to automate routine conversations and deliver faster and more targeted answers for customer inquiries.

DigitalGenius leverages its relationships with enterprises to mine millions of historical customer service chat logs, which represent millions of appropriate responses to a wide range of customer queries. Using this data, DigitalGenius trains its NLP algorithms. In 2016, it launched its Human+AI Customer Service Application on the Salesforce AppExchange. The application was built on a Salesforce platform to work with Salesforce Service Cloud Lightning.

The Human+AI Customer Service Platform is deployed on a SaaS basis within customer service departments and clients include airline, consumer packaged goods, automotive, insurance and banking industries. Specific customers include KLM, Unilever, and L'Oréal. While historically focused on the customer service space, the company is hoping to move its technology into the sales/marketing/legal arenas.

It is a member of Level39, a European accelerator. The company has raised $7.35 million in two rounds. The latest round was in April 2016 when the company secured $4.1 million.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

5.7 FACEBOOK

Year Founded:	2004
Annual Revenue:	$27.64 billion (2016)
Ownership:	Public (FB)
Headquarters:	Menlo Park, California
CEO:	Mark Zuckerberg
Number of Employees:	17,048 (2016)

Company Profile:
Established in 2004 by a group of college students, Facebook is an online social networking platform where users can connect and share information with individuals worldwide. The company has 1.5 billion active users a month. Facebook has especially been applying deep learning in the areas of text analytics, facial recognition, spam filtering, newsfeed improvement, language translation, and targeted advertising. Facebook's deep learning efforts have been coming out of its Facebook Artificial Intelligence Research (FAIR) group, headed by New York University (NYU) Professor and deep learning pioneer, Yann LeCun. The group actively engages with the research community and open sources much of its software. Facebook's use and development of deep learning technologies is becoming increasingly widespread and the company is a leader in this space.

Facebook offers many deep learning-powered products and libraries. DeepText can extract meaning from words by learning to analyze them textually. DeepFace offers facial recognition, which is Facebook claims is more accurate than humans, reportedly 97% accurate. FBLearner Flow uses deep learning analysis to run simulations of 300,000 machine learning models every month, to allow engineers to test ideas and pinpoint opportunities for efficiency. Just introduced in 2016, FastText is an open-source set of libraries supporting text representation, classification, and learning word vector representations.

In 2016, Facebook announced it was also leveraging deep learning in the development of automatic alternative text. Designed to help the blind or vision impaired also enjoy Facebook, automatic alternative text generates an audible rich description of a photo using object recognition technology. Recent deep learning announcements include the release of a video tool that can take video or an image from a cell phone and convert the image in real time into a selection of artistic styles, such as Van Gogh. Called Caffe2Go, the algorithm is a real neural net running on a phone in real time. CEO Mark Zuckerberg also used deep learning in the development of Jarvis, his self-made AI assistant.

When it comes to community-facing efforts, Facebook offers open-source extensions of both Torch and Caffe frameworks, and a number of associated modules. Torch runs on Lua and is a relatively flexible and robust framework for convolutional neural nets and has a native interface for temporal convolution. Caffe runs on C++ and is one of the first industry-grade deep learning frameworks specialized for computer vision. Facebook tends to use Torch for research and Caffe for deployments. Recently, Facebook released extensions to Caffe to support efficient inference and converting models from Torch to Caffe for deployment.

Broadly speaking, AI and deep learning are inherent in Facebook's longer-term strategy; from global connectivity to sustained social media dominance and virtual reality, machine learning is essential for scale. Facebook is also using neural networks to identify both urban and rural population centers to map out utility polls for connectivity. Its philosophy of open-source is also a nod to its commitment: the company believes opening up tooling and methods to everyone helps accelerate AI development for all.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

5.8 GOOGLE

Year Founded:	1998
Annual Revenue:	$90.27 billion (2016)
Ownership:	Public (GOOGL)
Headquarters:	Mountain View, California
CEO:	Sundar Pichai
Number of Employees:	72,053 (2016)

Company Profile:
Founded in 1998 by two Stanford PhD students, Larry Page and Sergey Brin, Google is now a multinational technology corporation specializing in internet-related services and products, including online advertising technologies, search, cloud computing, software, and hardware. In 2015, Alphabet Inc. was created as the parent company of Google and several other companies, including London-based DeepMind. The company's stated mission is "to organize the world's information and make it universally accessible and useful." Deep learning has been leveraged to accomplish this mission throughout Google. The company is an industry leader in AI and deep learning, with cutting-edge technologies rolling out on a frequent basis. The company has helped grow the AI industry by open sourcing these technologies.

AI is found throughout the organization, but the Google Brain team was set up specifically to explore neural networks. Members of this team set their own research agenda. TensorFlow is an open-source framework for deep learning that grew out of the Google Brain group. TensorFlow is used to train neural networks to detect and decipher patterns and correlations. Google Brain is also responsible for a visualization tool called DeepDream, which lets users see what neural networks are seeing when they perform deep learning.

DeepMind, acquired in 2014, greatly reinforced Google's presence in the deep learning space. DeepMind, for the most part, is kept relatively separate from Google and works on items with longer research cycles than Google Brain. DeepMind Lab is a fully 3D game-like platform tailored for agent-based AI research. DeepMind Lab was used internally for quite some time, but is now open-sourced. DeepMind's deep reinforcement learning project, AlphaGo, defeated the world's Go champion, Lee Sedol, in March 2016. In December 2016, AlphaGo went up against online players and beat them. DeepMind has identified smartphone assistants, healthcare, and robotics as its ultimate targets.

Google advancements and announcements in the deep learning space happen on a continual basis. Currently, deep learning is found in a multitude of Google products, such as Google Assistant, Google Translate, Google Maps, YouTube, and Smart Reply (Gmail).

In September 2016, the company announced Google Neural Machine Translation, which operates entirely through neural networks and, in October, DeepMind revealed a new system, known as a differential neural computer (DNC), which can independently learn from its own experiences. In November of the same year, the company announced it would give $4.5 million over the next 3 years for AI research at the Montreal Institute for Learning Algorithms and would open up a deep learning and AI research group at its offices in Montreal. AMD stated it would be supplying Google with new GPUs for cloud computing and machine learning and, in early 2017, the Google Cloud Platform will offer GPUs worldwide for the Google Compute Engine and Cloud Machine Learning users. Most recently, a team from Google Brain demonstrated that machines can learn how to protect their messages from others.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Aside from DeepMind, Google's acquisitions in this space include API.ai, Moodstocks, Timeful Jetpac, and Granata Decision Systems. Dark Blue Labs and Vision Factory were acquired in 2014.

5.9 H2O.AI

Year Founded: 2011
Ownership: Private
Headquarters: Mountain View, California
CEO: SriSatish Ambati
Number of Employees: 70

Company Profile:
Founded in 2011, H2O.ai is the company behind H2O, an open-source machine learning platform that enables companies to build models on large data sets. H2O's platform provides tooling for end-to-end solutions for AI and over 75,000 data scientists have used its products. Aside from its flagship platform, the company offers Sparkling Water, a combination of H2O with Apache Spark; Steam, an end-to-end platform that streamlines the whole process of building and deploying smart application; and Deep Water, which encompasses native implementation of deep learning models (TensorFlow and Caffe) for GPU-optimized backends.

While the technology can be used by many vertical industries, the company is particularly focused on finance, insurance, and healthcare. Customers include Capital One, Progressive Insurance, Zurich North America, and Transamerica. In 2016, H2O announced a partnership with IBM whereby H2O's machine learning capabilities will be available through IBM's Data Science Experiment.

While historically known for offering much of its technology free and open-sourced, the company has signaled a change in strategy. In September 2016, H2O.ai stated its focus would move toward larger, fewer deals and deeper engagements with select finance and insurance customers. In 2017, the company plans to offer new cloud services.

H2O has raised $33.6 million in 4 rounds from 10 investors. Its most recent funding was in November 2015 when the company raised $20 million.

5.10 IBM

Year Founded: 1911
Annual Revenue: $79.92 billion (2016)
Ownership: Public (IBM)
Headquarters: Armonk, New York
CEO: Virginia Rometti
Number of Employees: 380,300 (2016)

Company Profile:
Founded in 1911, IBM is a global leader in technology innovation and the largest technology and consulting employer in the world. Long known for its strong legacy in middleware and analytics, IBM is increasingly investing in its cognitive computing systems. Watson, no longer just the winner of Jeopardy game shows, has pivoted to be an entire ecosystem of products and services, with deep learning teams working to improve its performance on different problems. Aside from Watson, IBM is involved in both the hardware and software spaces of deep learning.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Many products are offered that are helping companies deploy AI more rapidly. The company has developed a series of new servers specifically designed to drive AI, deep learning, high-performance data analytics, and other compute-heavy workloads. A deep learning software kit, Power AI, was launched in November 2016 to take advantage of these servers that have NVIDIA's NVLink capabilities. Power AI consists of five open-sourced AI tools and includes deep leaning frameworks, such as Caffe, Torch, and Theano. IBM has been working on its TrueNorth computer chip for years, which is capable of engaging in deep learning and making decisions based on associations and probabilities. The chip has demonstrated it is especially good at image recognition. Watson allows customers to apply cognitive computing features across a variety of applications and industries. APIs are offered for language, vision, speech, and data. Watson Machine Learning Service (formerly IBM's Predictive Analytics Service) intelligently and automatically builds models from data and open machine learning libraries. Project DataWorks, another Watson initiative, is designed to make it simple for business leaders and data professionals to collect, organize, govern, and secure data or business insights. It uses Machine Learning and Apache Spark.

With the acquisition of AlchemyAPI in 2015, Watson gained access to image, text analysis, and sentiment analysis using deep learning, as well as computer vision services, also based on deep learning. Some applications for its deep learning products and services include medical image recognition, video processing/analysis, text analysis, virtual agents, weather forecasting (Deep Thunder), personalized recommendations, music interpretation and creation, and virtual agents. A wide range of verticals leverage its services, including healthcare, commerce, education, IoT, supply chain, finance, and marketing. In the future, IBM is planning hardware and software for inferencing. Drones, robots, and autonomous cars use inferencing engines for navigation, image recognition, or data analysis. Inferencing chips are also used in data centers to boost deep learning models.

5.11 INDICO

Year Founded: 2013
Ownership: Private
Headquarters: Boston, Massachusetts
CEO: Slater Victoroff
Number of Employees: 20

Company Profile:
Founded in 2013, Indico was formed out of Olin College, an undergraduate engineering school located in Massachusetts. The company provides deep learning algorithms to developers, such as platform APIs (custom collections, CrowdLabel, text features, and image features), text APIs (sentiment analysis, text tags, keywords, political analysis, etc.), and image APIs (facial-emotion recognition, content filtering, object recognition). The custom collection of APIs handles unstructured data and specializes in smaller sets of data with transfer learning.

Its technology is used in many different industries, including retail, fashion, marketing, finance, media, image filtering, software consulting, legal, and public relations. Some specific customers include Mavrck, CO Everywhere, and interlinkONE.

Moving forward, Indico will be moving deeper into the financial sector by building a machine learning software program for companies working in insurance and asset management. In August 2016, the company announced a partnership with Manulife's Lab of Forward Thinking (LOFT). Manulife's LOFT will use Indico's platform to develop an AI and deep learning tool to analyze unstructured financial data.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Indico has raised $4.3 million in four rounds from multiple investors. The latest round was $1.2 million in November 2016.

5.12 INTEL

Year Founded:	1968
Annual Revenue:	$59.39 billion (2016)
Ownership:	Public (INTC)
Headquarters:	Santa Clara, California
CEO:	Brian Krzanich
Number of Employees:	106,000 (2016)

Company Profile:
Founded in 1968, Intel is the world's biggest computer chip company, controlling 80% of the market for microprocessors that go into desktop and notebook computers. It also makes chips for servers, smartphones, and tablets as well as embedded semiconductors for the industrial, medical, and automotive markets. Intel has recently begun to jumpstart its investment in the field of AI and deep learning, in both hardware and software, with a few key acquisitions.

The Intel Deep Learning Inference Accelerator (DLIA) combines traditional Intel CPUs with FPGAs, semiconductors that can be reprogrammed to perform specialized computing tasks. Aside from hardware, Intel offers libraries (Intel Math Kernel Library, a DNN) as well as toolkits (Intel Deep Learning SDK).

In 2015, Intel acquired Saffron, a cognitive computing platform that specializes in processing and analyzing enterprise, industrial, and device data, and modeling predictions for the performance and behavior of processes and infrastructure.

In August 2016, Intel acquired Nervana Systems to apply deep learning at the chip level. Nervana offers a full stack hosted deep learning platform (Nervana Cloud), as well as development work on deep learning hardware (Nervana Engine). The Nervana platform aims to deliver up to 100X reduction in the time to train a deep learning model over the next 3 years compared to GPU solutions. As part of the Xeon Phi chip family, the company is working on an AI chip code-named Knights Mill, which it alleges will be 4X faster in deep learning tasks than the current Xeon Phi chip. The Intel Nervana Graph Compiler was introduced to accelerate deep learning frameworks on Intel silicon.

Additionally, in September 2016, the company announced it would be purchasing Mobidius, which will bring algorithms tuned for deep learning, depth processing, navigation and mapping, and natural interactions, as well as broad expertise in embedded computer vision and machine intelligence.

Across these acquisitions, one can see Intel pushing into deep learning across multiple parts of the stack. While today the focus is on classic verticals for deep learning (e.g., text and object classification), Intel's longer-term play is to support specialized hardware that enables data ingestion, sensor fusion, and reasoning. One such area is autonomous driving; Intel plans to introduce a roadmap of new automotive products that will incorporate deep learning, and it announced a partnership with BMW and Mobileye in this area. Intel will test its first silicon, code-named "Lake Crest," in early 2017. Lake Crest is optimized specifically for neural networks to deliver high performance for deep learning.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

5.13 IRIS AUTOMATION

Year Founded: 2015
Ownership: Private
Headquarters: San Francisco, California
CEO: Alexander Harmsen
Number of Employees: 1-10

Company Profile:
Iris Automation was founded in 2015 based on the need for a high-precision vision solution for drones. Co-founded by Alex Harmsen and James Howard in Vancouver, BC, the company relocated to California in 2016.

Iris Automation is using its own deep learning technology, as well as computer vision, to develop avoidance systems for industrial drones. This system assists the drone in situational awareness, an area that has been lacking with drone technology. Its first product is a combination of off-the-shelf chips and other components, and proprietary software that can learn and tell a drone's autopilot system when there is any obstacle nearby, and how to adjust instantly to avoid it. The company hopes to build its systems so that they can be added to drones made by any manufacturer.

The technology will have application in the mining exploration, search and rescue, pipeline inspection, agricultural surveying, forestry, and potentially package delivery markets. The company has raised more than $500,000 in funding.

5.14 MICROSOFT

Year Founded: 1975
Annual Revenue: $85.32 billion (2016)
Ownership: Public (MSFT)
Headquarters: Redmond, Washington
CEO: Satya Nadella
Number of Employees: 114,074 (2016)

Company Profile:
Founded in 1975 by Bill Gates and Paul Allen, Microsoft is a multinational technology company that develops, licenses, and supports a range of software products, services, and devices. The company also designs, manufactures, and sells PCs, tablets, gaming and entertainment consoles, phones, and other intelligent devices that integrate with its cloud-based offerings. Microsoft's goal is to make AI available to everyone through agents, applications, services, and infrastructure. Its commitment in this area was cemented by the formation of the Microsoft AI and Research Group, which brings together more than 5,000 computer scientists and engineers focused on the company's AI product efforts.

Deep learning plays a critical role in almost all its services and products, as well as its own day-to-day processes. Consumers see Microsoft's deep learning in Cortana, Bing, Office 365, SwiftKey, Skype, Translate, Dynamics 365, and HoloLens. Its toolkits are typically in speech recognition, search rankings, photo search, translation systems, etc. The Cortana Intelligence Suite has multiple corporate, manufacturing, medical, and IoT applications. Sales teams are using neural nets to recommend which prospects to contact next or what kinds of product offerings to recommend. Churn prediction and prevention is also assisted by deep learning. Microsoft's deep learning offerings hinge on the company's Cognitive Toolkit, formerly known as the Computational Network Toolkit (CNTK). This open-source deep learning framework supports the implementation of AI on different languages, such as

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

C++ or Python. Microsoft's Cognitive Services are a series of cloud APIs that provide AI capabilities in cognitive computing areas, such as vision, speech, text, and knowledge.

On Azure, Microsoft's cloud computing platform, the company offers a suite for machine learning and advanced analytics called the Cortana Intelligence Suite. These services can be used, along with the Cognitive Toolkit or any other deep learning framework, to deploy intelligent applications. The Data Science Virtual Machine (DSVM) is available in the Azure Marketplace and comes pre-loaded with a range of deep learning frameworks and tools for Linux and Windows. The deep learning toolkit for the DSVM is a solution for the Windows DSVM that installs several GPU-accelerated tools for deep learning, CUDA, cuDNN, the GPU driver, and several samples. Microsofts's Zo chatbot, introduced in December 2016, and its widely popular Xiaoice (only available in China) are both powered by deep learning.

Microsoft has found particular success using deep learning in its speech recognition and translation, which is bolstered by the Cognitive Toolkit. In October 2016, Microsoft Artificial Intelligence and Research announced that it had, for the first time, created a technology that recognizes words in a conversation as well as a person does. In December 2016, the company announced Microsoft Translator, a new tool that is in preview, which delivers live, in-person speech translation via internet-connected smartphones, tablets, and PCs.

The company continues to bolster its presence in the market through acquisitions. In January 2017, Microsoft acquired Maluuba, a Canadian startup focused on artificial general intelligence (AGI) in language. Through this acquisition, AI researcher Yoshua Bengio will be joining Microsoft as an advisor. Other notable AI purchases include Genee in August 2016 and Equivio in 2015.

In November 2016, OpenAI, the non-profit AI research firm backed by Elon Musk, signed an agreement to run most of its large-scale experiments on Microsoft's Azure. OpenAI will use Azure for its experiments in deep learning and AI, and Microsoft will collaborate with the company on advancing research and creating new tools and technologies. In December of the same year, Microsoft, Cray, and the Swiss National Supercomputing Centre (CSCS) announced work on a project to speed up the use of deep learning algorithms on supercomputers.

5.15 MOBILEYE

Year Founded:	1999
Annual Revenue:	$358.2 million (2016)
Ownership:	Public (MBLY)
Headquarters:	Jerusalem, Israel
CEO:	Ziv Aviram
Number of Employees:	470 (2016)

Company Profile:

Mobileye was founded in 1999 by Professor Amnon Shashua and Mr. Ziv Aviram and is one of the leading suppliers of software supporting Advanced Driver Assist Systems (ADAS). The company primarily serves OEMs, insurance companies, lease companies, consumers, fleet management systems providers (telematics), and off-highway industrial vehicles.

The company has evolved to include AI, deep learning, and crowdsourcing to create the hardware and software needed to enable ADAS and, eventually, fully autonomous vehicles. Deep learning is used in Mobileye's computer vision system. Specifically, Mobileye has implemented deep learning to improve the path detection algorithms of its EyeQ chips. Sensing algorithms use supervised learning, while its "driving policy" algorithms use

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

reinforcement learning, which is a process of using rewards and punishments to help the machine learn how to negotiate the road with other drivers. The technology trains the vehicle to take the information that the sensors provide it, and the data that the computer chip is processing, and make decisions based both on past learning and informed predictions on how other drivers will react to decisions based on simulations. This technology can also deploy 360° surround-view mono-vision sensing to build high-definition, crowdsourced maps.

Unlike many of its competitors, Mobileye's technology is based on optical vision systems and motion detection algorithms. The company's technology is capable of pedestrian collision warning, lane departure detection, forward collision warning, headway monitoring, intelligent headlight intensity adjustment, and speed limit indication, among others, such as traffic sign detection and rear camera applications.

Mobileye has received numerous awards for its technology. As of August 2016, the company's stock (NYSE: MBLY) had a market capitalization of $11.46 billion. In January 2017, BMW Group, Intel, and Mobileye announced that a fleet of approximately 40 autonomous BMW vehicles will be on the roads by the second half of 2017.

5.16 NVIDIA

Year Founded:	1993
Annual Revenue:	$6.91 billion (fiscal 2017)
Ownership:	Public (NVDA)
Headquarters:	Santa Clara, California
CEO:	Jen-Hsun Huang
Number of Employees:	10,299 (fiscal 2017)

Company Profile:
Founded in 1993, NVIDIA invented the GPU in 1999 and still dominates this market. Almost every major technology company is partnered with NVIDIA in some way and uses its hardware. NVIDIA has recently repositioned itself as "the AI computing company" and the company has been out in front of the pack in terms of accelerating deep learning. NVIDIA has been developing deep learning software, libraries, and tools for a number of years and most of today's deep learning solutions rely almost exclusively on NVIDIA GPU-accelerated computing to train and speed up AI services, such as image, handwriting, and voice identification.

NVIDIA's GPUs, built for deep learning, are available in desktops, notebooks, servers, and supercomputers around the world, as well as in cloud services from Alibaba, Amazon, Baidu, Google IBM, Microsoft and Tencent. The NVIDIA Titan X is the fastest accelerator for deep neural network training on a desktop PC based on the NVIDIA Pascal architecture. NVIDIA claims its DGX-1 is the world's first purpose-built system for deep learning with fully integrated hardware and software. The Tesla P100 is an accelerator for HPC, advanced analytics, and deep learning based on the Pascal architecture, while the Tesla P40 is known for its speed. The Tesla P4 is a lower-power, small form-factor GPU accelerator optimized for video transcoding, image recognition and processing, and deep learning inference. The company is continuously improving upon its chips to increase efficiency and speed for deep learning.

In May 2017, the company announced its launch of the Tesla V100 Volta, a next-generation offering based on its graphics architecture that boasts more than 21 billion transistors and 5,120 computing cores, including 640 Tensor cores built to deliver some 120 teraflops of performance, specialized for deep learning. This is a sharp blow to competitors, as these

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

chips speed up deep learning training by 12x and inferencing by 6x relative to current solutions, according to NVIDIA.

The NVIDIA Deep Learning SDK provides tools and libraries for designing and deploying GPU-accelerated deep learning applications. It includes libraries for deep learning primitives, inference, video analytics, linear algebra, sparse matrices, and multi-GPU communications. NVIDIA Digits assists with image classification. NVIDIA SHIELD and SPOT are services primarily geared toward voice recognition, smart home interactions, and automation.

The company is going big in the autonomous driving space and offers end-to-end solutions. Drive PX2 offers scalable in-vehicle AI supercomputing for autonomous driving. NVIDIA's unified AI computing architecture enables training deep neural networks in the data center on the NVIDIA DGX-1, and then seamlessly runs them on NVIDIA DRIVE PX 2 inside the vehicle. This leverages NVIDIA DriveWorks software and allows cars to receive over-the-air updates to add new features and capabilities throughout the life of a vehicle. In January 2017, NVIDIA announced a partnership with Mercedes-Benz, Audi, Bosch, Toyota, and Volvo to bring a self-driving car to production within the next 12 months.

Other industries leveraging NVIDIA's deep learning technologies include robotics, healthcare, finance, government, advertising, and energy. NVIDIA is also involved in nurturing startups in the AI field through Inception, a virtual incubator program and its Deep Learning Institute partners with online education providers to offer online and in-person deep learning training events.

5.17 OpenAI

Year Founded:	2015
Ownership:	Private
Headquarters:	San Francisco, California
CEO:	Sam Altman/Elon Musk/Greg Brockman (Co-Founders)
Number of Employees:	45

Company Profile:
OpenAI was founded in 2015 as a research-based AI company focused on collaboration with other entities and researchers. With Elon Musk as one of its co-founders, it was generated from the idea that AI should be leveraged in a positive way to benefit humanity. The company will make its patents and research open to the public.

OpenAIGym, a toolkit for developing and comparing reinforcement learning algorithms was launched in April 2016. It supports teaching agents for applications such as games, robotics, and language-based tasks. Universe was released in December 2016. Universe is a software platform for measuring and training an AI's general intelligence across games, websites, and other applications.

OpenAI is using Microsoft's Azure as its primary cloud provider. NVIDIA's GPUs are being used in a machine learning project with the goal of allowing a robot to recognize speech and to use the data appropriately. OpenAI has $1 billion in funding committed by its co-founders and other technologists, as well as Amazon, Infosys, and YC Research.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

5.18 RIPJAR

Year Founded: 2012
Ownership: Private
Headquarters: Cheltenham, Gloucestershire, England
CEO: Tom Griffin
Number of Employees: 25

Company Profile:
Ripjar was founded in 2012 by engineers from the U.K. Government Communications Headquarters (GCHQ). The company has developed a data intelligence platform for analyzing, interacting with, and visualizing data in real time.

Ripjar's Strategic Intelligence Platform can be deployed directly onto a company's hardware or over the internet as a service. The platform applies analytics, such as NLP, machine learning techniques, and visual analytics, to data sets. The connections are then processed very quickly to present intelligence for decision-making. Ripjar's Data Studio, Explorer, and Command Center are also part of its platform. These products can be used in the areas of cybersecurity, reputation management, defense and intelligence, customer and legal intelligence, and internal company monitoring for risk and compliance issues. The company has raised over $22,000 in funding, with its last round in May 2016.

5.19 SALESFORCE

Year Founded: 1999
Annual Revenue: $8.39 billion (2016)
Ownership: Public (CRM)
Headquarters: San Francisco, California
CEO: Marc Benioff
Number of Employees: 20,000 (2016)

Company Profile:
Founded in 1999, Salesforce is a global leader in CRM. Salesforce's Customer Success Platform provides cloud services for sales, service, marketing, community, analytics, apps, and the IoT. A relative latecomer to the deep learning space, Salesforce has essentially bought itself into the market through a series of acquisitions over the past few years. The MetaMind acquisition in April 2016 was an important one and its AI efforts are now being led by Dr. Richard Socher, the former founder of MetaMind. In addition to MetaMind, Salesforce has made many other AI acquisitions. In December 2016, Salesforce entered into an agreement to purchase Twin Prime, a machine learning provider. PredictionIO and Implisit were purchased in 2016. Other notable acquisitions were RelateIQ in 2014 and Tempo AI and Minhash in 2015.

The result of such acquisitions is its Salesforce Research machine learning team of more than 175 data scientists. This team aims to develop offerings that are made available both as services that developers can use and build into their apps, as well as offered as an "add-on" to the software that Salesforce sells.

MetaMind uses multi-task learning and has also combined natural language and image recognition networks into a single system that can answer questions about images. Salesforce leveraged this technology, as well as technology gained through the acquisition of PredictionIO to introduce the Einstein Platform. This platform uses machine learning, deep learning, predictive analytics, NLP, and smart data discovery. Einstein embeds advanced AI capabilities in the Salesforce Platform in fields, objects, workflows, components, and more,

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

enabling customers to build AI-powered apps that get smarter with every interaction. Within the platform, each cloud (service, analytics, IoT, sales, marketing, commerce, and community) will gain extra functionalities, such as case classification, lead scoring, predictive analytics, content identification, smart data discovery, recommendations, and automated rules optimization. Predictive vision and sentiment analysis will also enable developers to train deep learning models to recognize and classify images and sentiment in text.

In November 2016, Salesforce Research announced advances in several areas, including the development of a neural network called a dynamic coattention network, which can answer questions about text by using a coattentive encoder and a dynamic decoder. Additionally, the research team developed a joint many-task model for handling a variety of NLP tasks in a single deep model and made advances in text and translation processing.

The company is developing machine and deep learning techniques to enhance its existing platform, a broad suite of enterprise applications with plenty of room to benefit itself, and Salesforce's customers through greater software intelligence. Unlike other competitors, Salesforce is focusing solely on the enterprise sales, marketing, and service spaces, and is unlikely to branch into other verticals or parts of the stack with its AI efforts.

5.20 SENTIENT TECHNOLOGIES

Year Founded: 2007
Ownership: Private
Headquarters: San Francisco, California
CEO: Antoine Blandeau
Number of Employees: 80

Company Profile:
Sentient Technologies was founded in 2007 by entrepreneurs Antoine Blandeau and Babak Hodjat, who were involved in developing Apple's Siri. The company remains active in research in the deep learning space and is particularly interested in end-to-end training, modular models, and multiple layers of abstraction.

The company's AI technology is capable of scaling to millions of cores and is further differentiated by using distributed evolutionary computation. Combined with deep learning technologies, Sentient can analyze image-based interactions in order to deeply understand a user's intent and preferences. It then combines content and recommends actions based on that understanding. Its two main products are Sentient Aware (product discovery, intelligent recommendations, adaptive/living websites, and marketing) and Sentient Ascend (automation and acceleration of website testing and optimization). Aware was launched with Canadian-based retailer Shoes.com.

Sentient is also known for its work on developing an AI-driven equity fund, in the works using private investor funds since 2014. A subsidiary, Sentient Investment Management, has been formed to develop and apply its technologies to trading and investment strategies. Users adjust risk setting and the AI runs the trading while serving up specific recommendations around instruments, exits, exposure, etc. According to Sentient, it aims to bring the intelligent hedge fund to market sometime in 2017. Beyond the financial industry, Sentient is currently targeting e-commerce and online content, as well as the medical field. Sentient Technologies is extremely well funded, with approximately $143 million raised. Its most recent round clocked in at $103 million in November 2014.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

5.21 SKYMIND

Year Founded: 2014
Annual Revenue: $2.8 million
Ownership: Private
Headquarters: San Francisco, California
CEO: Chris Nicholson
Number of Employees: 15

Company Profile:
Founded in 2014 by Chris Nicholson and Adam Gibson, Skymind's mission is to make deep learning accessible and easy to use for enterprises. The company offers the only commercial-grade, open-source, distributed deep learning library written for Java and Scala, Deeplearning4j (DL4J). Integrated with Hadoop and Spark, DL4J is specifically designed to run in business environments on distributed GPUs and CPUs.

The Skymind Intelligence Layer (SKIL) is the company's open-source enterprise distribution that packages up SL4J and other open-source tools from Skymind and features a graphical user interface. It can be run in public clouds or in on-premises data centers, on standard x86 chips, optionally those with GPUs, or Intel Xeon Phi chips or IBM Power8 chips. Skymind also offers services such as training, consulting, and support.

Specific applications include face/image recognition, voice recognition, speech-to-text, spam filtering, fraud detection, and recommendations. Industries using Skymind's solutions include telecom (fraud), data centers (assistance with OpenStack), finance (fraud/anomaly detection), commerce (recommendation systems), and CRM. Orange SA is already using Skymind and SKIL to build and deploy production code for deep learning projects.

The company has raised a total of $3.31 million in three rounds. Its most recent funding was $3 million in September 2016.

5.22 SPARKCOGNITION

Year Founded: 2013
Ownership: Private
Headquarters: Austin, Texas
CEO: Amir Husain
Number of Employees: 70

Company Profile:
SparkCognition, founded in 2014, was launched by serial entrepreneur Amir Husain. The company is a cognitive security analytics provider, which applies machine learning and AI to cloud security and the IoT. Its products are focused on using AI to enhance cybersecurity and leverage machine learning to identify and prevent equipment failures before they happen.

SparkCognition has three main products. SparkSecure is a cloud service or an on-premise application that applies machine learning and AI algorithms to find trends, patterns, and anomalies. The company also provides MindFabric cognitive analytics platform, a data science platform for non-data scientists. Patent-pending Cognitive Fingerprinting machine-learning algorithms learn from this data. It works in conjunction with IBM Watson, which has incorporated the technology into its analytics toolset. SparkPredict applies algorithms to huge bodies of sensor data, identifies impending failures long before they occur, and alerts operators to sub-optimal operation before it can cause any harm. DeepArmor is a beta

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

project to protect networks from malware attack. It uses machine learning, NLP, and AI algorithms to analyze files and provide signature-free security.

The company's two primary verticals are the IoT and cybersecurity. Clients include many Fortune 500/1000 organizations and one of the largest wind farms in North America uses it to predict wind turbine failures and for fleet optimization. It also serves oil and gas, utility, and IT industries.

In 2016, SparkCognition announced it had begun working with Carnegie Mellon University's Software Engineering Institute (SEI) to develop an automated threat identification and remediation system that works in conjunction with IBM's Watson. Later that year, the company announced a partnership with NI and IBM to collaborate on the Condition Monitoring and Predictive Maintenance Testbed. Overall, the company has raised $10 million in funding with its last round occurring in April 2016 for $6 million.

5.23 TENCENT

Year Founded: 1998
Annual Revenue: $21.9 billion
Ownership: Public
Headquarters: Shenzhen, Guangdong, China
CEO: Huateng (Pony) Ma
Number of Employees: 30,641

Company Profile:
Founded in 1998, and listed on the Hong Kong Stock Exchange in 2004, Tencent is a Chinese provider of internet, mobile, and telecommunication services. More specifically, the company offers instant messaging, mail, online media, games, wireless internet value-added services (such as SMS), and online advertising. Tencent's WeChat social messaging app is extremely popular and widely used in China. In 2016, the company announced a deeper push into the AI space with the creation of an AI Lab.

The quickly growing AI Lab will focus on machine learning, NLP, computer vision, and speech recognition. The AI solutions developed by Tencent will be open sourced to its partners through Tencent Cloud. In December 2016, the company launched an AI platform, Angel, with plans to open source it in 2017. Tencent Angel is the third generation of AI platform, developed by Java and Scala, with distributed computing framework for machine learning. Tencent has been using machine learning in its products and services, such as video streaming, social advertisement, personalized news recommendations, search, and data mining.

The company has a huge investment portfolio and has been investing in companies in the AI space, such as Skymind and Diffbot.

5.24 TERADEEP

Year Founded: 2014
Ownership: Private
Headquarters: Santa Clara, California
CEO: Didier LaCroix
Number of Employees: 6

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Company Profile:
Founded in 2014, TeraDeep is a pioneer of deep learning on FPGAs. It was founded as a spinoff from Purdue University and is working to commercialize the technology developed at the university. Originally, its focus was a deep learning camera and hardware package that was taught to recognize objects or perform tasks. TeraDeep has since developed machine learning/deep learning capabilities for the IoT/mobile and embedded devices.

The company has developed a suite of advanced deep learning algorithms, optimized to run on the company's TeraDeep Accel acceleration platform. The fourth generation of TeraDeep Accel was recently developed and is the result of the close collaboration with Xilinx and Micron. The platform is a combination of cloud-based machine-learning algorithms and deep learning neural network software, proprietary datasets, running on embedded devices, custom processor accelerators, and mobile apps to visualize the analytics. Capabilities include visual scene understanding, content tagging, and object detection/recognition/ localization in videos, images, audio, and speech. Its technology can be used for security and safety applications or any industry that needs to process large numbers of images or videos. TeraDeep once partnered with Aipoly to develop technologies that would assist the vision impaired. TeraDeep provided the company with a neural network trained on its private dataset of 10 million images.

In December 2016, TeraDeep announced plans to use Amazon Elastic Compute Cloud (EC2) F1 instances to bring high-performance, low-power video analytics to AWS cloud customers. The low-latency, power-efficient TeraDeep Amazon Machine Image (AMI) relies on Amazon EC2 F1's FPGA-optimized architecture.

TeraDeep benefited from an undisclosed amount of funding in March 2016. Xilinx Technology Ventures was the prime investor.

5.25 UBER

Year Founded:	2009
Annual Revenue:	$972.22 million
Ownership:	Private
Headquarters:	San Francisco, California
CEO:	Travis Kalanik
Number of Employees:	6,700

Company Profile:
Founded in 2008, when CEO Travis Kalanick could not find a cab in Paris, the company has now grown from an app to request transportation in a few metropolitan areas, to a major player in urban logistics. Uber is leveraging deep learning to weed through its masses of data to produce more efficient customer and driver experiences.

The company has been using machine learning to improve the way prices respond to supply and demand and to better match riders and drivers; however, the company is broadening its focus. Its goal is to create machine learning as an internal service with a graphical interface and APIs so that every team in the company can use it. Models are being built from the large amounts of data Uber has and the model is learning from continuous experience.

In November 2016, Uber rolled out some big changes to its mobile app, with machine learning behind the redesign. The app redesign has focused on greater personalization, which has some extensive machine learning work on the back end. Uber is using machine learning algorithms layered on top of its historic trip data to make more accurate estimated time of arrival (ETA) information, taking into account traffic patterns, etc. Machine learning

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

will be used to determine the best routes, optimize multiple pickups with UberPOOL, compute when a car and UberEats will arrive, and assist with fraud detection. A potential future application is assisting Uber's efforts in the self-driving space. Uber has an Advanced Technology Center in Pittsburgh focusing on this area and in August 2016, Uber acquired Otto, an autonomous trucking company.

Uber will continue to leverage deep learning throughout its offered experience, and its December 2016 acquisition of Geometric Intelligence and the subsequent formation of Uber AI Labs is another signal of its commitment to the space. Uber AI Labs will be directed by Geometric's founder Gary Marcus and will take on developing forms of machine learning that need less data; training AI systems using not only data, but also explicit rules; and designing machine learning systems that explain their decisions. These are technologies that were being developed by Geometric Intelligence prior to the acquisition.

Uber has raised over $15 billion in multiple rounds of funding. Its latest round was for $3.5 billion in June 2016, which is the single largest round of funding ever for a private U.S. company.

5.26 VICARIOUS

Year Founded:	2010
Annual Revenue:	$973,290
Ownership:	Private
Headquarters:	Union City, California
CEO:	D. Scott Phoenix
Number of Employees:	40

Company Profile:
Vicarious was founded in 2010 by Scott Phoenix and Dileep George. The company is building a general AI that can adapt to whatever task it is confronted with, aspiring to achieve "human- level intelligence."

Vicarious is focused on visual perception problems like recognition, segmentation, and scene parsing. Its underlying framework combines the advantages of deep architectures and generative probabilistic models. The long-term goal is to build a unified algorithmic architecture. The company has been relatively quiet without any public commercial applications being released yet, and there are no current plans for productization.

Vicarious has raised $72 million in funding, with the last round taking place in August 2015 for an undisclosed amount, a sum that it says avoids the constraints of publications, grant applications, or development cycles.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Table 5.1 Additional Industry Participants

Company	Location	Emp.	Markets Served	Owner-ship	Website
Affectiva	Waltham, MA	15	Consumer, Healthcare, Online Education, Gaming, Market Research	Private	www.affectiva.com/
Agrima Infotech	Kerala, India	120	Robotics, Mobile, Security, Risk Management	Private	http://agrimainfotech.com/
AiMotive	Budapest, Hungary	120	Auto	Private	https://aimotive.com/
Algolux	Montreal, Canada	25	Auto, Robotics, Mobile, AR/VR, Medical, Computing	Private	https://algolux.com/
Alpaca	San Mateo, CA	25	Finance	Private	www.alpaca.ai/
Alyuda Research	Cupertino, CA	120	Finance, Energy, Retail, Healthcare, Insurance, Science	Private	www.alyuda.com/
Arya	Mumbai, India	25	Banking, Retail, Health, Insurance, Oil & Gas	Private	http://arya.ai/
Atomwise	San Francisco, CA	10	Healthcare	Private	www.atomwise.com/
AxonAi	Harrison-burg, VA	10	Business, Healthcare, National Security, Legal, R&D	Private	http://axonai.com/
Bay Labs	San Francisco, CA	10	Healthcare	Private	baylabs.io/
Behold.ai	New York, NY	5	Healthcare	Private	http://behold.ai/
Captricity	Oakland, CA	120	Insurance, Healthcare, Government, Non-Profit	Private	captricity.com/
Cerebras Systems	Portola Valley, CA	20	N/A – Stealth Mode	Private	http://cerebras.net/
Comma.ai	San Francisco, CA	5	Education, Electronics, Automotive	Private	http://comma.ai/
Cortexica	London, England	25	Ecommerce, Retail	Private	www.cortexica.com/
Cortica	Tel Aviv, Israel	75	Advertising, Publishing	Private	www.cortica.com/
CrowdFlower	San Francisco, CA	60	E-Commerce, Customer Service, CRM	Private	https://www.crowdflower.com/
Deep Genomics	Toronto, Canada	25	Medical	Private	www.deepgenomics.com
Deepomatic	Paris, France	25	E-Commerce, Media	Private	www.deepomatic.com/

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Company	Location	Emp.	Markets Served	Owner-ship	Website
Deepgram	Mountain View, CA	10	Call Centers, Media	Private	www.deepgram.com/
DeepLift	Canada	N/A	Genomics	Private	http://deeplift.ai/
deepsense.io	Menlo Park, CA	25	Big Data	Private	https://deepsense.io/
DeepVision	Cordoba, Argentina	10	B2B, Computer Vision	Private	www.deepvisionai.com/
DeepVision	Palo Alto, CA	5	Auto, Robots, Drones, Smart Cameras	Private	http://deepvision.io
Drive.ai	Mountain View, CA	30	Auto & Robotics	Private	www.drive.ai/
DT42 (DeepThought 42)	Taipei City, Taiwan	5	IoT, Security	Private	http://www.dt42.io/
Ditto Labs	Cambridge, MA	25	Digital + Social Media, Advertising, Marketing	Private	http://dittolabs.io/
Enlitic	San Francisco, CA	25	Healthcare/Information Technology and Services	Private	www.enlitic.com
Ersatz Labs	Pacifica, CA	4	Consulting	Private	www.ersatzlabs.com/
Eyeris Technologies	Palo Alto, CA	25	Auto	Private	http://emovu.com/e/
Freenome	S. San Francisco, CA	20	DNA/Medical Diagnosis	Private	www.freenome.com/
GraphCore	London England	40	Software Development Services, Hardware	Private	www.graphcore.ai/
GridSpace	San Francisco, CA	25	Finance, Insurance, Transportation, Business Services, Communications, High Tech	Private	www.gridspace.com/
Gastrograph	New York, NY	15	Food+ Beverage	Private	www.gastrograph.com/
Horizon Robotics	Beijing China	50	Smart Home, Automobiles, Safety	Private	www.horizon-robotics.com
Hypothizer	Delhi, India	N/A	Geospatial	Private	www.hypothizer.com/
HyperVerge	Palo Alto, CA	25	Consumer, Image Recognition	Private	hyperverge.co/
Imagenii/ Pomika	San Antonio, TX	5	AR, E-Commerce, Advertising	Private	http://pomika.com/ www.imagenii.com/
Imagia	Montreal, Quebec, Canada	20	Healthcare	Private	www.imagia.com/
Intuition Machine	German-town, MD	5	Business Intelligence	Private	www.intuitionmachine.com/

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Company	Location	Emp.	Markets Served	Owner-ship	Website
Kensho	Cambridge, MA	75	Finance	Private	www.kensho.com
KickView	Centennial, CO	5	Aerospace & Defense, Wireless, Data Science & Automation	Private	https://kickview.com/
Luminoso	Cambridge, MA	40	Business Intelligence	Private	www.lumino.so/
LunIT	Seoul, Korea	25	Healthcare	Private	www.lunit.io/
MarianaIQ	Palo Alto, CA	35	B2B/Marketing	Private	www.marianaiq.com/
MedyMatch	Tel Aviv, Israel	20	Healthcare	Private	https://medymatch.com/
minds.ai	Sunnyvale, CA	30	Software Development	Private	http://minds.ai/
MindMeld (Expect Labs)	San Francisco, CA	25	Consumer Virtual Assistance	Private	www.mindmeld.com/
Mobvoi	Beijing, China	100	Consumer, Language	Private	www.chumenwenwen.com/en/site/index.html
Netra	Boston, MA	10	Advertising, Market Research, E-Commerce	Private	www.netra.io/
Netradyne	San Diego, CA	30	Commercial Vehicles, Fleet Safety	Private	www.netradyne.com/
Neurala	Boston, MA	25	Robots, Cameras, Drones	Private	www.neurala.com/
NeuralWare	Carnegie, PA	5	Business Intelligence	Private	www.neuralware.com/
Neuramatix	Kuala Lumpur, Asia	14	Bioinformatics, Language, Robotics	Private	http://neuramatix.com/
Numenta	Redwood City, CA	20	Research	Private	http://numenta.com/
Nnaisance	Lugano, Switzerland	10	Research	Private	https://nnaisense.com/
Orbital Insight	Mountain View CA	30	Geo-analytics, Satellite Imaging	Private	https://orbitalinsight.com/
Oben	Pasadena, CA	20	VP/AR/IoT	Private	https://oben.me/
Osaro	San Francisco, CA	12	Robots, Drones, IoT, Autonomous Vehicles, Digital Advertising	Private	www.osaro.com/
Palantir	Palo Alto, CA	2,250	Enterprise Intelligence	Private	www.palantir.com/
Pavlov	San Francisco, CA	5	AI Agents/Customer Service	Private	https://pavlov.ai/
Pilot.ai	Redwood City, CA	25	Consumer Electronics	Private	http://pilot.ai/

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Company	Location	Emp.	Markets Served	Owner-ship	Website
Pluto AI	Palo Alto, CA	5	IoT, Water Management	Private	www.plutoai.com/
Preferred Networks	Tokyo, Japan	25	Retail, Automobile, Healthcare, Advertising, Life Sciences, Manufacturing, Public Safety, Network Security and Optimization, Public Transportation	Private	www.preferred-networks.jp/en/
Qualcomm	San Diego, CA	30,500	Telecom and Semiconductor hardware	Public	www.qualcomm.com
Quantified Skin	San Francisco, CA	25	Healthcare	Private	http://quantifiedskin.com/
RSVP Tech	Waterloo, Canada	25	Robotics/Natural Language	Private	www.rsvptech.ca/
Securonix	Addison, TX	350	Security/Cybersecurity	Private	www.securonix.com/
Sensory Inc.	Santa Clara, CA	100	Wearables, Mobile, Consumer Electronics, Banking, Medical, Automotive, Entertainment/ Toys	Private	www.sensory.com/
Sentiance	Antwerp, Belgium	40	Mobility, Commerce, Connected Home, Health	Private	www.sentiance.com/
SightHound	Winter Park, FL	20	Home/Business Surveillance	Private	www.sighthound.com/
SignalSense	Seattle, WA	25	Security	Private	www.signalsense.com/
Sigtuple	Bangalore, India	25	Healthcare	Private	www.sigtuple.com/
SiteZeus	Tampa, FL	25	Retail, Real Estate, Brokerages	Private	https://sitezeus.com/
Staqu Technologies	Haryana, India	25	Ecommerce	Private	www.staqu.com/
Stiya	Sunnyvale, CA	5	Real Estate	Private	www.stiyalive.com/
StocksNueral/ Deep Gnosis	Moscow, Russia	N/A	Finance	Private	https://stocksneural.net/
SenseTime	Hong Kong	150	Banking/Security	Private	http://sensetime.com/
Tesla	Palo Alto, CA	13,058	Autonomous Systems/Ground Navigation	Public	www.tesla.com/
The Curious AI Company	Helsinki, Finland	25	Research	Private	https://thecuriousaicompany.com/
TrueAI	London, England	5	BI, Sales & CRM	Private	www.trueai.io/

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Company	Location	Emp.	Markets Served	Owner-ship	Website
twentybn	Berlin, Germany	20	Autonomous Vehicles, Robotics, Video Monitoring	Private	www.twentybn.com/
Tractable	London, England	5	Insurance	Private	www.tractable.io/
Viv	San Jose, CA	35	Consumer Virtual Assistance	Samsung (Public)	http://viv.ai/
Wave Computing	Campbell, CA	175	Computer/ Hardware/Big Data	Private	http://wavecomp.ai/
WayBlazer	Austin, TX	30	Travel	Private	http://wayblazer.com/
X.ai	New York, NY	90	Business, Virtual Assistance	Private	https://x.ai/
Vuno	Seoul, Korea	10	Healthcare	Private	www.vuno.co/home/
Zebra Medical	Kibbutz Shefayim Israel	30	Healthcare	Private	www.zebra-med.com/

(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 6

MARKET FORECASTS

6.1 FORECAST METHODOLOGY

The model used for estimating the size of the deep learning market follows a use case-based approach. The model is set up to mainly calculate the deep learning software revenue. For the purposes of this study, Tractica has identified 154 use cases for deep learning through various sources that include startups, technology companies, university research labs, and industry conferences. The use cases are then organized and classified under industry types, sub-industries, technologies, and timeline of adoption. For each use case, individual metrics are identified that help to estimate the addressable market size of the use case. Careful consideration has been given to the type of use case and appropriate metric, to ensure that there is no double counting and that use cases that have applications across different industries and sub-industries are identified and chosen appropriately.

The use case metrics are classified as top-down or bottom-up metrics, with top-down representing a dollar-denoted metric and deep learning revenue as a percentage of a larger addressable market. Bottom-up is where the metric uses units such as the number of mobile phones, cars, or airplanes, and the units are multiplied by unit cost for the technologies that are used. Appropriate scaling up and scaling down of costs have been applied based on the units being in the billions or in the hundreds, with a higher scale of units denoting a lower unit cost. Individual metric values were identified across the five world regions of North America, Europe, Asia Pacific, Latin America, and the Middle East & Africa. In the final step of the model, penetration rates are used in conjunction with the metrics to derive the deep learning market size for each of the individual case studies, across the regions.

The model does not account for new or future use cases that could emerge during the forecast period. Tractica believes that most of the new use cases that could emerge will be subsets or branches of the existing use cases that have been included in the model. Tractica also aims to keep track of new emerging use cases and will be publishing updates to the model and forecasts on a regular basis.

6.2 GLOBAL DEEP LEARNING MARKET FORECASTS

Tractica forecasts that annual revenue for deep learning software will increase from $655 million worldwide in 2016 to 34.9 billion in 2025, representing a CAGR of 56%. Annual revenue for deep learning applications, taking into account software, services, and hardware, will increase from $3.3 billion in 2016 to $160 billion in 2025, at a CAGR of 54%

Such a high CAGR is consistent with a market that is relatively new, even if the technology is not. According to Tractica's modeling across six distinct subsets of AI, deep learning accounts for the largest segment; 48% of the overall AI revenue in 2016 growing to 57% of the revenue by 2025. Both deep learning and AI are among the most funded areas of research in computer science, but it is important to note that other technologies are expected to emerge during this report's forecast period.

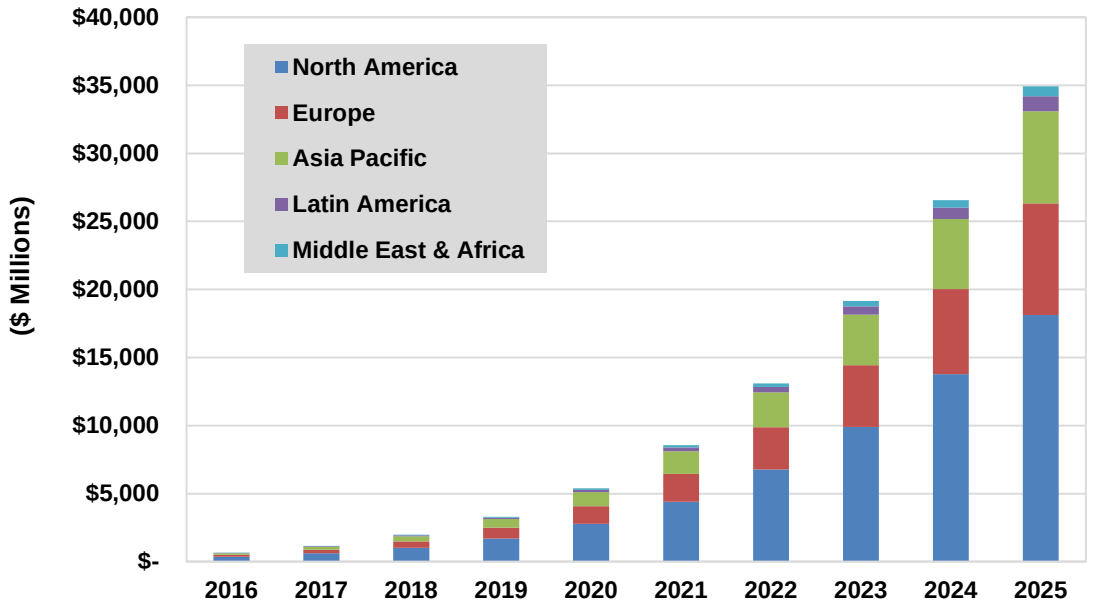
As with many software technologies, the most robust adoption in the early years will occur in North America and deployments will subsequently spread to the rest of the world, although, for some use cases, less developed regions may leapfrog more mature markets.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.2.1 DEEP LEARNING SOFTWARE REVENUE

Tractica's forecast for deep learning software is illustrated in Chart 6.1 below. North America will remain the leading region for deep learning adoption throughout the forecast period, however other world regions will demonstrate strong growth, as well, with the highest growth rates occurring in Asia Pacific.

Chart 6.1 Deep Learning Software Revenue by Region, World Markets: 2016-2025



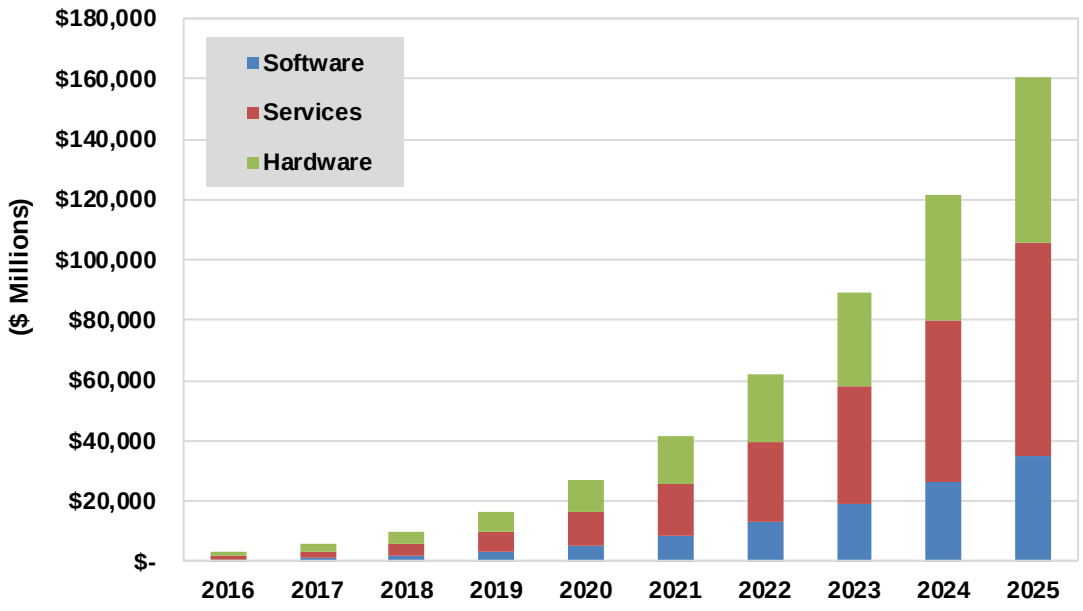
(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.2.2 TOTAL REVENUE FOR DEEP LEARNING SOFTWARE, SERVICES, AND HARDWARE

The chart below forecasts global revenue generated by deep learning-driven software, services, and hardware.

Chart 6.2 Deep Learning Total Revenue by Segment, World Markets: 2016-2025



(Source: Tractica)

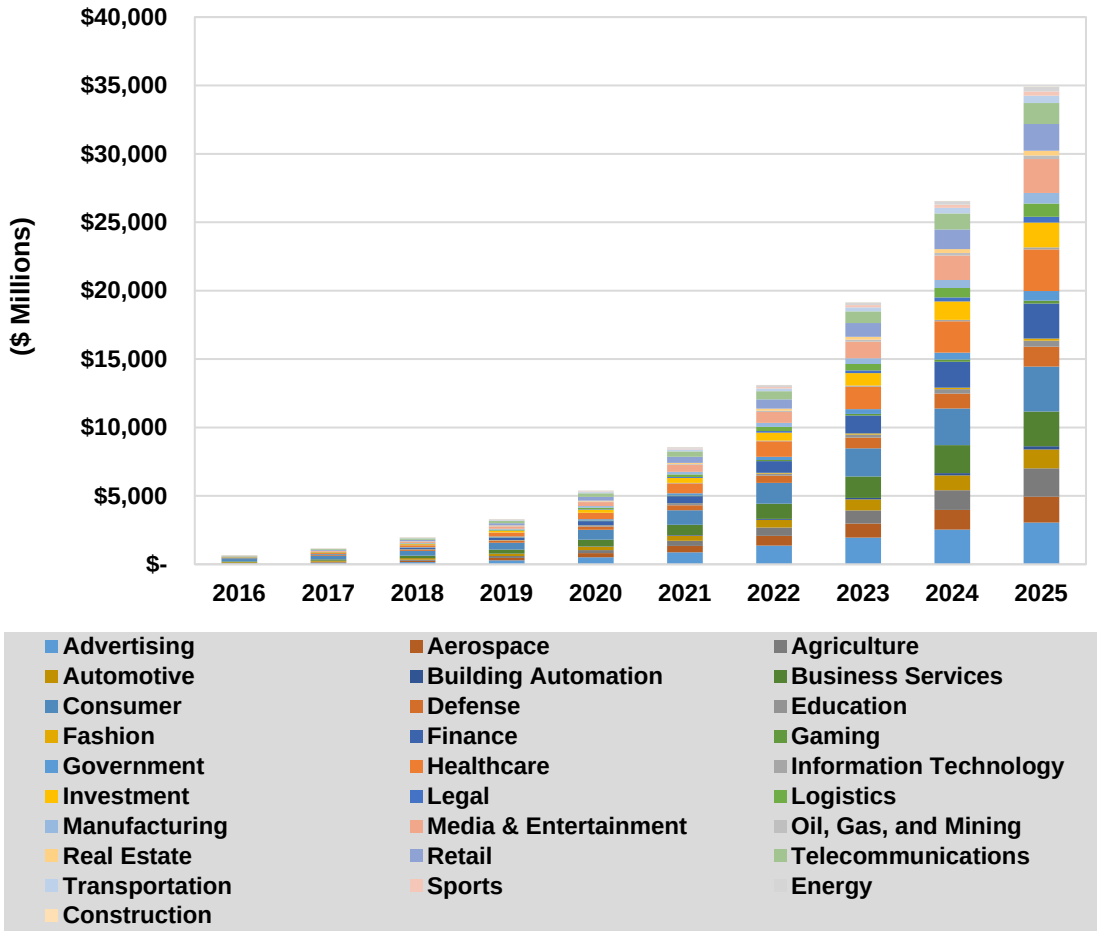
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.2.3 DEEP LEARNING SOFTWARE REVENUE BY INDUSTRY

Like AI more broadly, deep learning is also gaining in adoption across a wide range of industries. In particular, Tractica's research finds leading adopters of deep learning in the automotive, advertising, business services, defense, and consumer sectors. Consumer, in particular, consists of a diverse set of consumer-facing use cases and represents the greatest adopter of deep learning applications to date by revenue.

Over time, Tractica expects other industries, such as healthcare, agriculture, finance, and manufacturing, to grow significantly. By 2025, deep learning is expected to be a much more fragmented market in terms of industry mix, with no single industry clearly dominating the revenue forecast. The fragmentation increases as we move through the forecast period. What is certain is increased adoption in those industries with access to massive (and growing) amounts of data, for which deep learning will reliably support vision and language-based tasks, and analyze very large data sets with far greater efficiencies.

Chart 6.3 Deep Learning Software Revenue by Industry, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

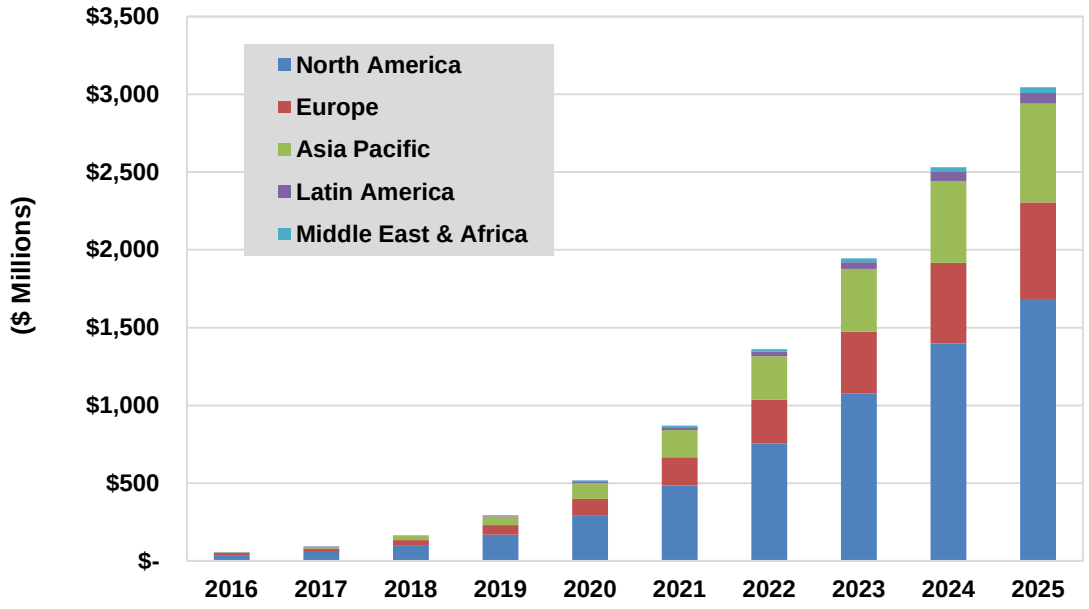
6.2.4 DEEP LEARNING IN THE ADVERTISING INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in advertising will increase from $55.3 million worldwide in 2016 to $3 billion in 2025. This is a market already well underway in its adoption of AI and one of the most data-intensive industries, generating millions of petabytes every day. Alphabet, Google's parent, currently controls 12% of all money spent globally on media advertising; Facebook is another leader in hyper-targeted advertising; both are leading the development and practical application of deep learning.

Use cases considered in this forecast were:

- Ad insertions into images and video
- Human emotion analysis
- Interactive window displays
- Performance reporting and analytics of ad campaigns
- Static image recognition, classification, and tagging
- Targeted advertising using multi-domain customer data (social, web, context)
- Video content analysis
- Voice/speech recognition

Chart 6.4 Deep Learning Software Revenue in the Advertising Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

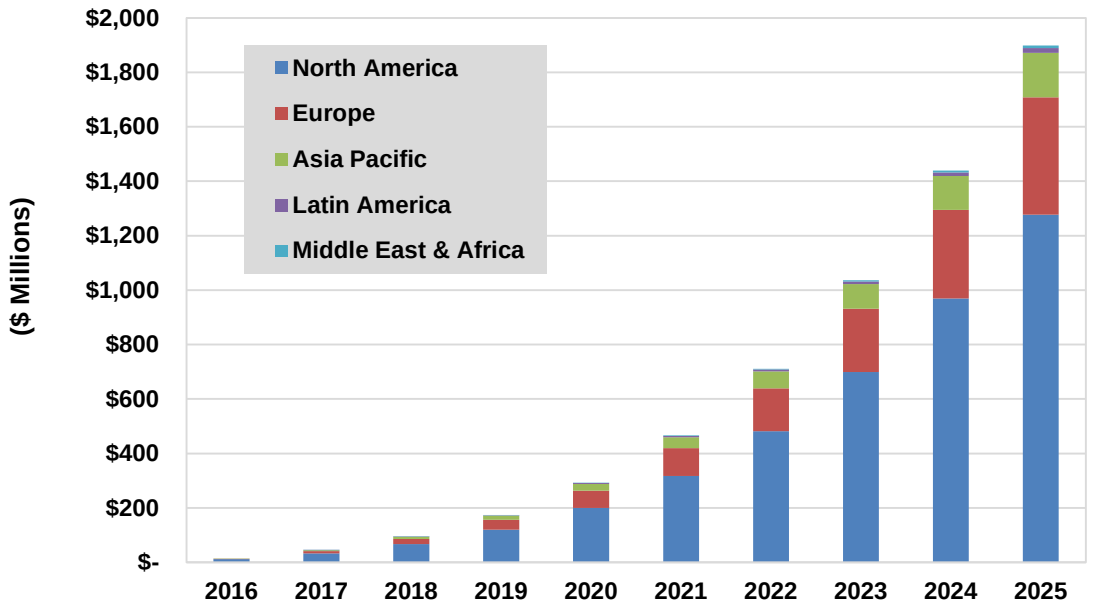
6.2.5 DEEP LEARNING IN THE AEROSPACE INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in aerospace will increase from $13.4 million worldwide in 2016 to $1.9 billion in 2025. Aerospace and avionics are industries that have been working with machine learning and AI applications for years, but are advancing rapidly, given improvements to data processing speed.

Use cases considered in this forecast include:

- Localization and mapping (commercial aircraft, consumer drones)
- Machine/vehicular object detection/identification/avoidance (commercial aircraft, consumer drones)
- Predictive maintenance (commercial aircraft, consumer drones, satellites)
- Sensor data fusion in machinery (commercial aircraft, consumer drones, satellites)
- Vehicle network and data security (commercial aircraft, consumer drones, satellites)
- Weather forecasting

Chart 6.5 Deep Learning Software Revenue in the Aerospace Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

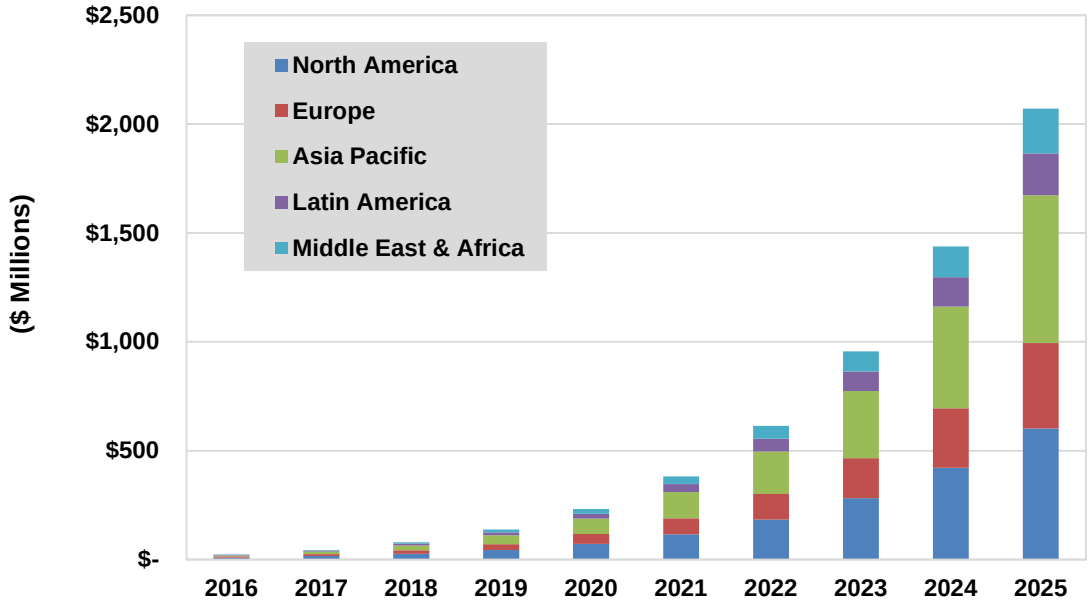
6.2.6 DEEP LEARNING IN THE AGRICULTURE INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in agriculture will increase from $ 23.8 million worldwide in 2016 to $2.1 billion in 2025. Although many estimate agricultural producers will need to increase production yield by some 70% to sustain the world's population by 2050, and technology will be central to achieving to this, adoption of AI and deep learning will be highly fragmented. About 1 billion of the world's people remain subsistence or small-scale farmers. They are among the poorest people on the planet and the overwhelming majority of them will be untouched by AI or any electronic technology. Regions that do adopt AI will also produce the most crop surpluses.

Use cases considered in this forecast include:

- Food safety
- Machine/vehicular object detection/identification/avoidance
- Satellite imagery for geo-analytics
- Sensor data analytics
- Sensor data fusion in machinery
- Localization and mapping
- Weather forecasting
- Weed identification

Chart 6.6 Deep Learning Software Revenue in the Agriculture Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.2.7 DEEP LEARNING IN THE AUTOMOTIVE INDUSTRY

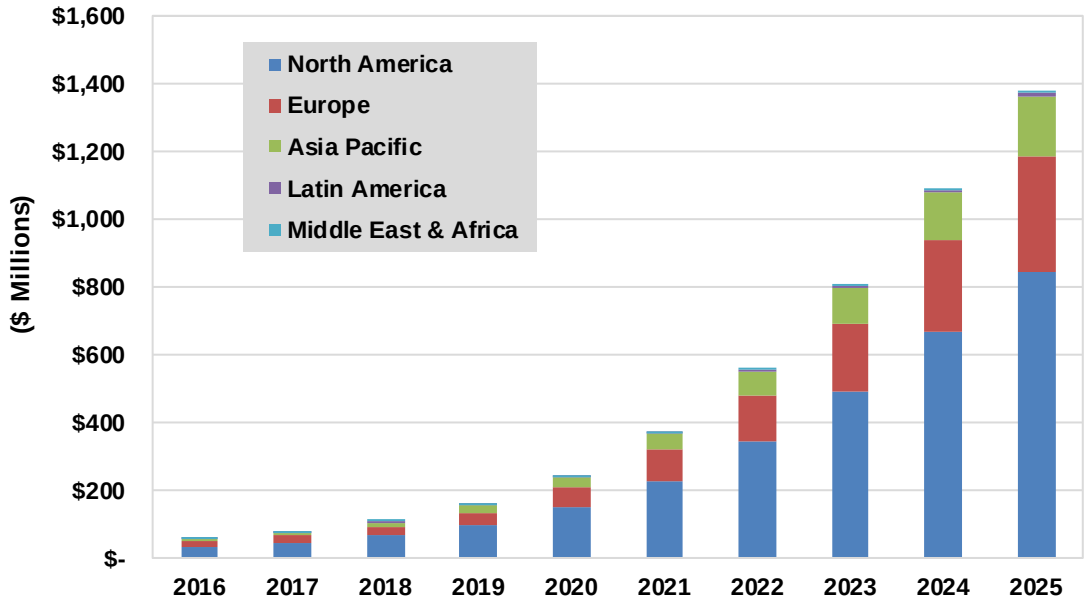
Tractica forecasts that the annual revenue for enterprise applications of deep learning in advertising will increase from $ 61.2 million worldwide in 2016 to $1.4 billion in 2025. The automotive industry is a primary adopter of deep learning technology, particularly in the development of autonomous vehicles. Rather than attempt to introduce a completely automated vehicle to market, the industry has chosen to release AI products piecemeal as "safety products" that will lead eventually to a self-driving car. In addition to object detection and machine vision, training and testing of cars on different road and traffic conditions is where the majority of the focus is, with an arms race between traditional auto manufacturers and internet companies in rolling out the first fully autonomous self-driving car.

Use cases considered in this forecast include:

- Automated on-road customer service
- Building generative models of the real world
- Driver face analytics and emotion recognition
- Gesture recognition
- Machine/vehicular object detection/identification/avoidance
- Truck platooning
- Predictive maintenance
- Sensor data fusion in machinery
- Simulating worlds for AI training
- Localization and mapping
- Vehicle network and data security
- Virtual testing and simulation for racing cars

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Chart 6.7 Deep Learning Software Revenue in the Automotive Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

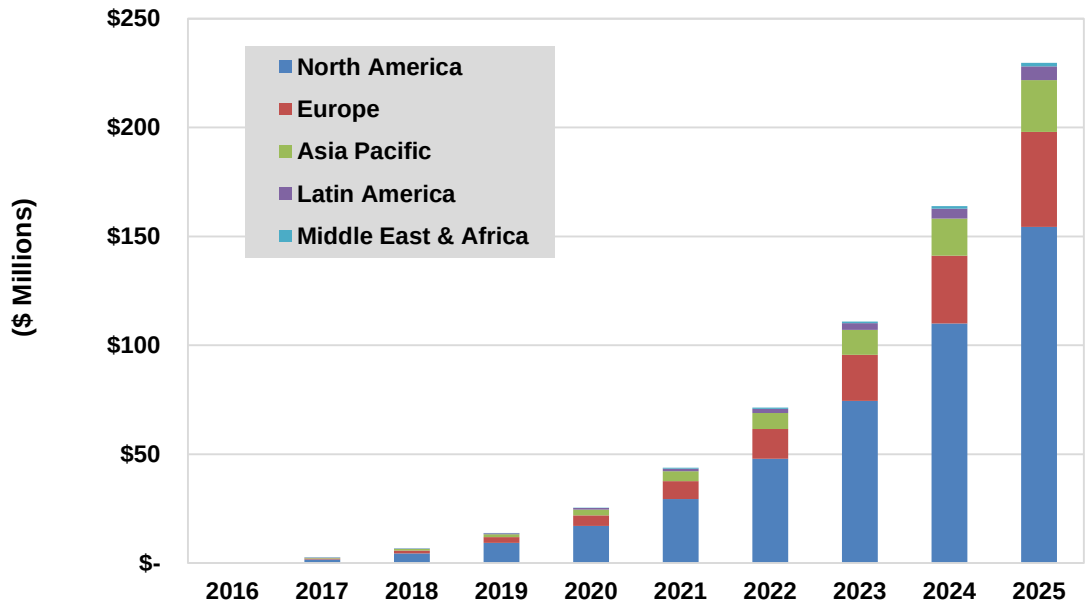
6.2.8 DEEP LEARNING IN THE BUILDING AUTOMATION INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in building automation will increase from $2.4 million worldwide in 2017 to $229.7 million in 2025.

The use case considered in this forecast is:

- Building automation and energy management

Chart 6.8 Deep Learning Software Revenue in the Building Automation Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.2.9 DEEP LEARNING IN THE BUSINESS SERVICES INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in business services will increase from $ 50.2 million worldwide in 2016 to $2.5 billion in 2025.

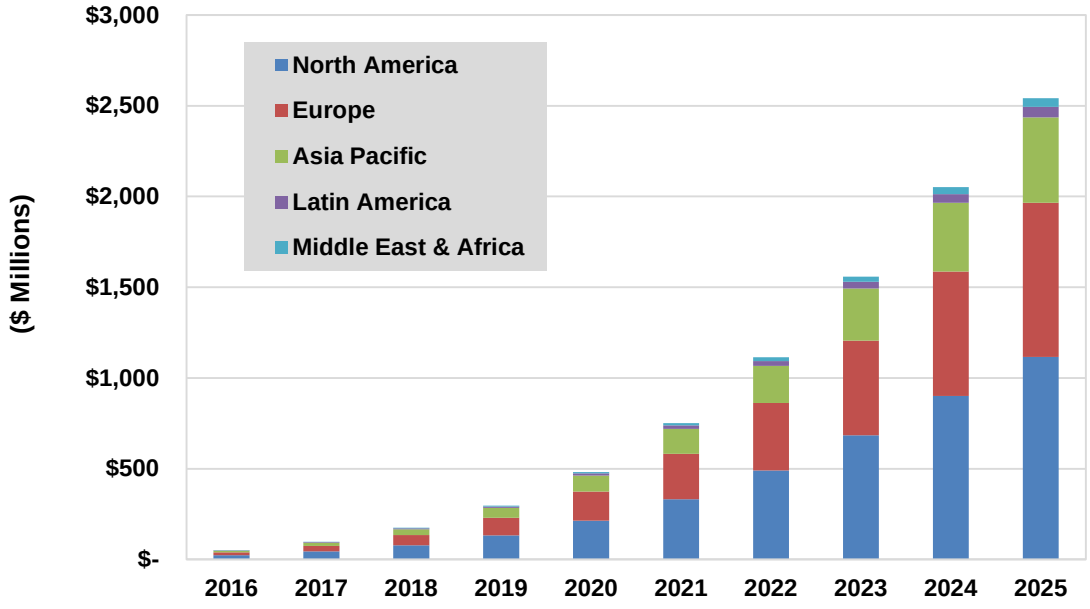
A business process is a set of activities that will accomplish a specific organizational goal. AI and deep learning are being applied to enhance, automate, and predict various aspects in these activities across business functions, from support processes, accounting, compliance, ERP, supply chain management, and human resource management to risk management and CRM.

Use cases considered in this forecast include:

- Agent-based simulations for decision making
- Audio and video mining
- Automated report generation
- Chatbot-based brand/service interaction
- Chatbot-based e-commerce and sales
- Crowdsourced market research
- Enterprise chatbots for productivity and collaboration
- Intelligent CRM systems (contact management, customer acquisition and planning, customer service, predictive sales and marketing)
- Intelligent recruitment and HR systems (candidate finder, recruitment, predictive talent hiring)
- Prevention against cybersecurity threats
- Procurement management
- Project and stakeholder management
- Real-time news analysis and competitive intelligence
- Social media publishing and management

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Chart 6.9 Deep Learning Software Revenue in the Business Services Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

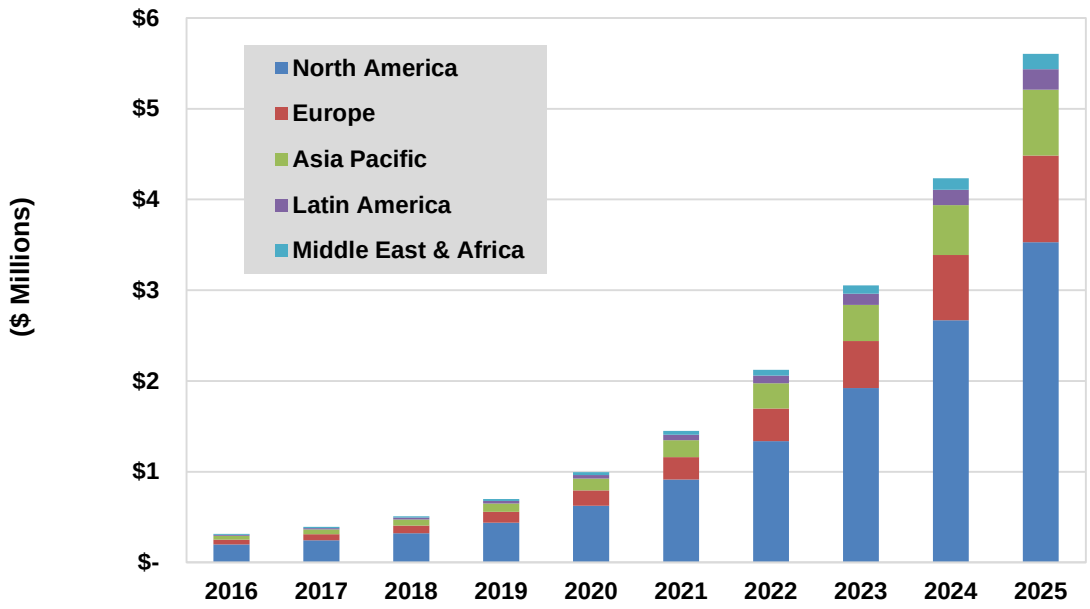
6.2.10 DEEP LEARNING IN THE CONSTRUCTION INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in construction will increase from $0.3 million worldwide in 2016 to $5.6 million in 2025.

The use case considered in this forecast is:

- Satellite imagery for geo-analytics

Chart 6.10 Deep Learning Software Revenue in the Construction Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.2.11 DEEP LEARNING IN THE CONSUMER SECTOR

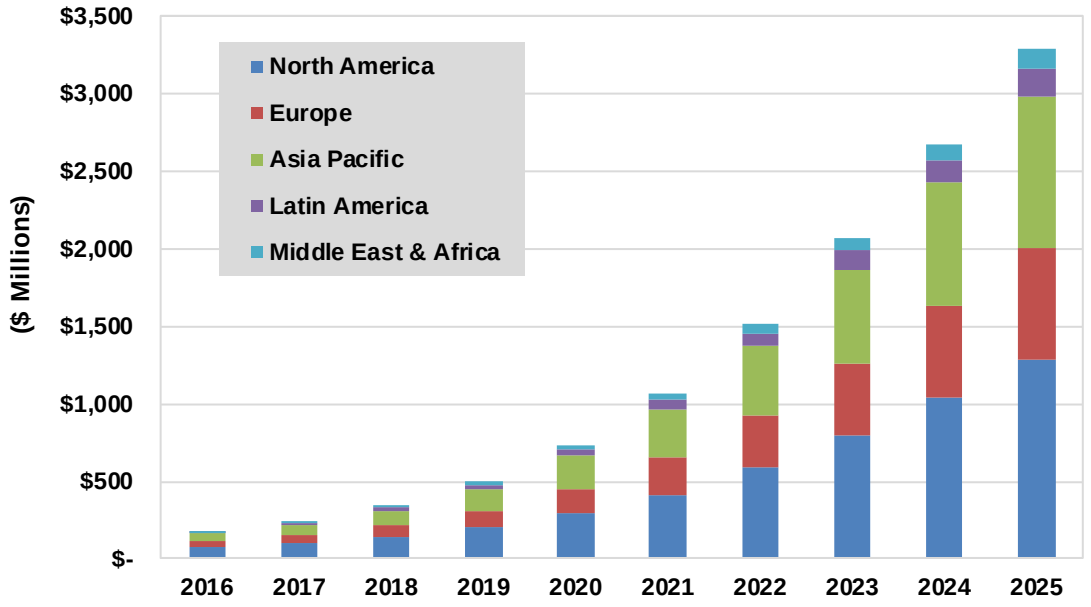
Tractica forecasts that the annual revenue for applications of deep learning in consumer markets will increase from $179.2 million worldwide in 2016 to $3.3 billion in 2025. Consumer markets represent the largest market share of deep learning revenue today, in part due to the broad range of applications. From consumer robotics to search engine functionality to online dating, consumers are slowly beginning to interface (albeit unknowingly) with deep learning in almost every realm of their lives. Adoption will be fragmented overall depending on use case, but most application development will take place in North America.

Use cases considered in this forecast include:

- Automated tour guide and itinerary service
- Building generative models of the real world
- Computer-aided art
- Contextual intelligence for mobile
- Face recognition (personal robots, robotics & autonomous machines)
- Music recommendations
- Machine/vehicular object detection/identification/avoidance (cleaning robots, personal robots, robotics & autonomous machines)
- Predictive typing assistant
- Product recommendations
- Relationships and matchmaking
- Search engine queries
- Smart oven control with food recognition
- Social media feed curation
- Static image recognition, classification, and tagging
- Text-based automated bots
- Voice/speech recognition (personal assistants, cleaning robots, personal robots, robotics & autonomous machines)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Chart 6.11 Deep Learning Software Revenue in the Consumer Sector by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

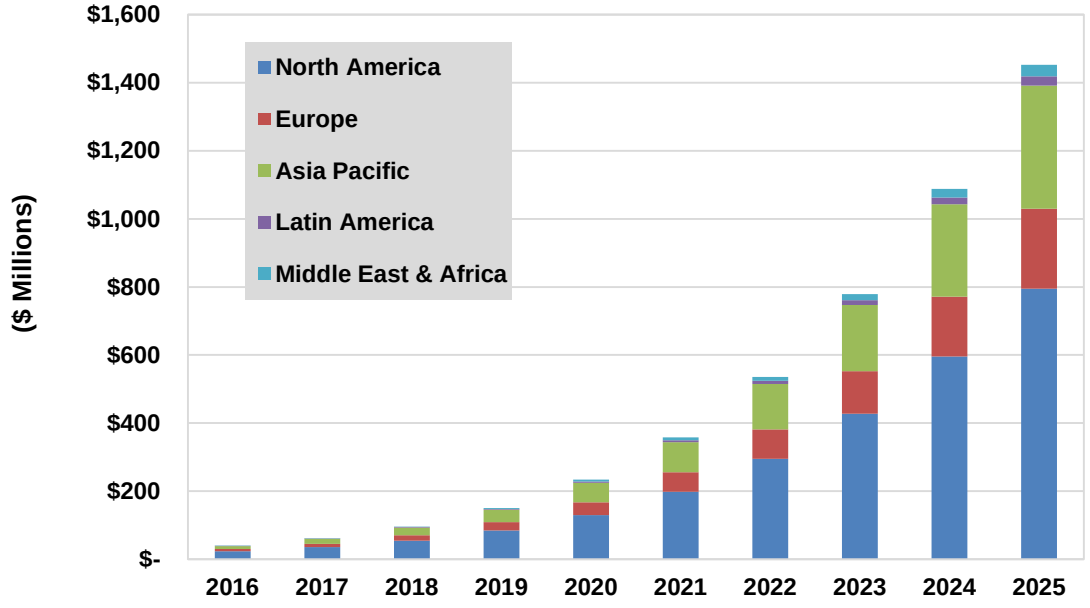
6.2.12 DEEP LEARNING IN THE DEFENSE SECTOR

Tractica forecasts that the annual revenue for applications of deep learning in defense will increase from $ 39.4 million worldwide in 2016 to $1.5 billion in 2025. AI and deep learning applications are gaining traction in defense industries given their applicability in areas around decision-making, machine perception, sensor data fusion, and beyond. Tractica expects continued growth in North American markets, while Asia Pacific will gain significantly as advancements in robotics accelerate there.

Use cases considered in this forecast include:

- Agent-based simulations for decision making
- Localization and mapping (aircraft, drones)
- Machine/vehicular object detection/identification/avoidance (aircraft, drones)
- Predictive maintenance (aircraft, drones, satellites)
- Prevention against cybersecurity threats
- Satellite imagery for geo-analytics
- Sensor data fusion in machinery
- Vehicle network and data security (aircraft, drones, satellites)

Chart 6.12 Deep Learning Software Revenue in the Defense Sector by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

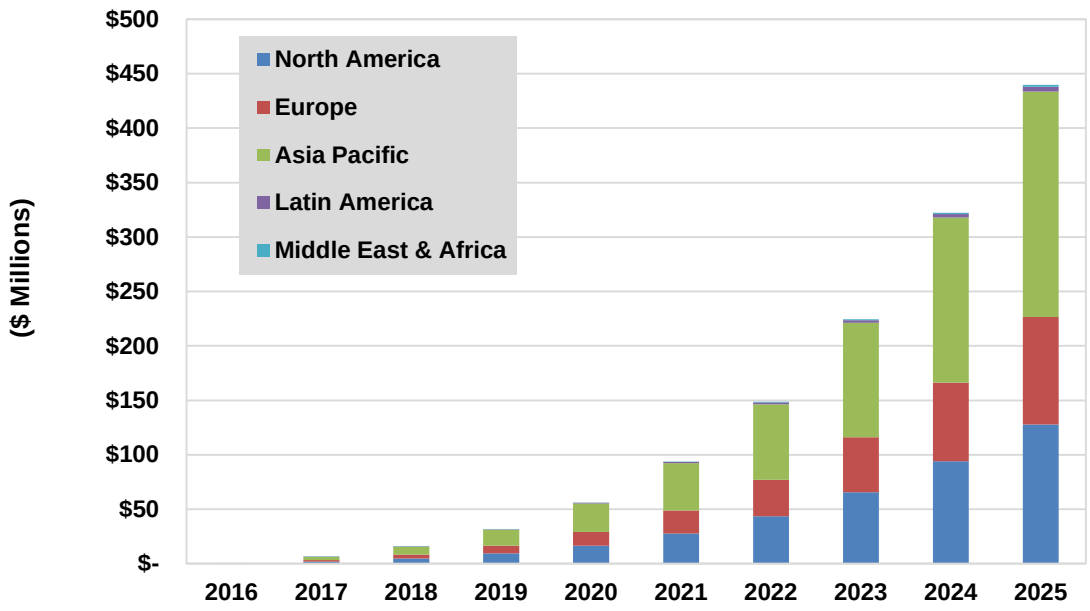
6.2.13 DEEP LEARNING IN THE EDUCATION INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in education will increase from $0.4 million worldwide in 2016 to $440 million in 2025. Despite its academic roots, deep learning has been adopted very slowly by educational institutions. In North American and Asia Pacific markets, deep learning is sometimes applied to academic research problems, but most of the AI in use today is being deployed by online education companies that hope to disrupt the traditional education markets. Use will vary widely by region and will never be fully adopted.

Use cases considered in this forecast include:

- Automated grading of tests
- Spoken fluency evaluation
- Textual question answering

Chart 6.13 Deep Learning Software Revenue in the Education Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

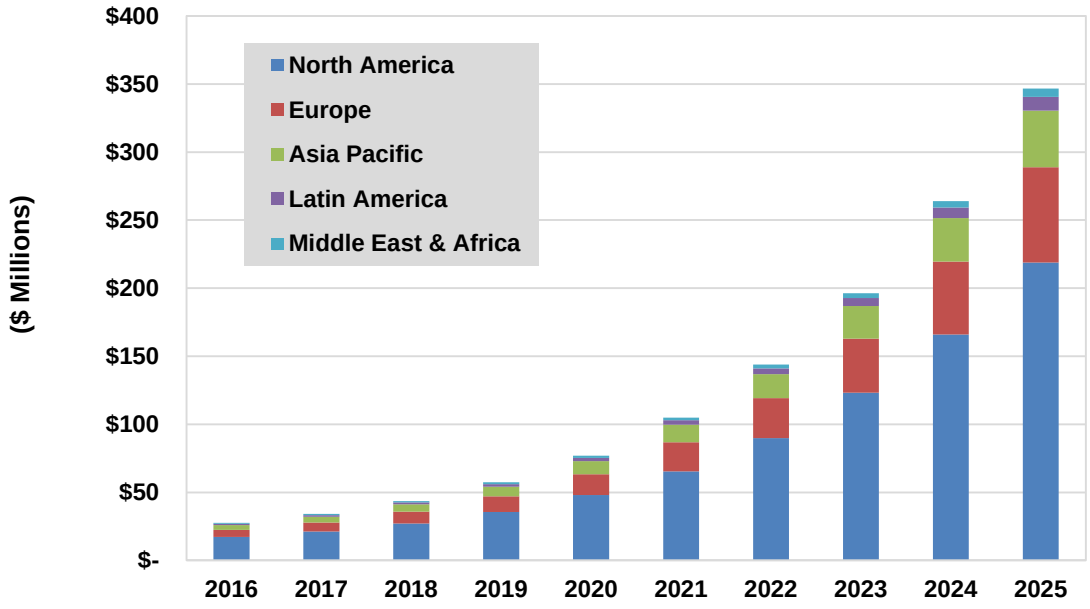
6.2.14 DEEP LEARNING IN THE ENERGY INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in energy will increase from $27.7 million worldwide in 2016 to $346.6 million in 2025.

Use cases considered in this forecast include:

- Satellite imagery for geo-analytics
- Weather forecasting
- Nuclear/power plant safety

Chart 6.14 Deep Learning Software Revenue in the Energy Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

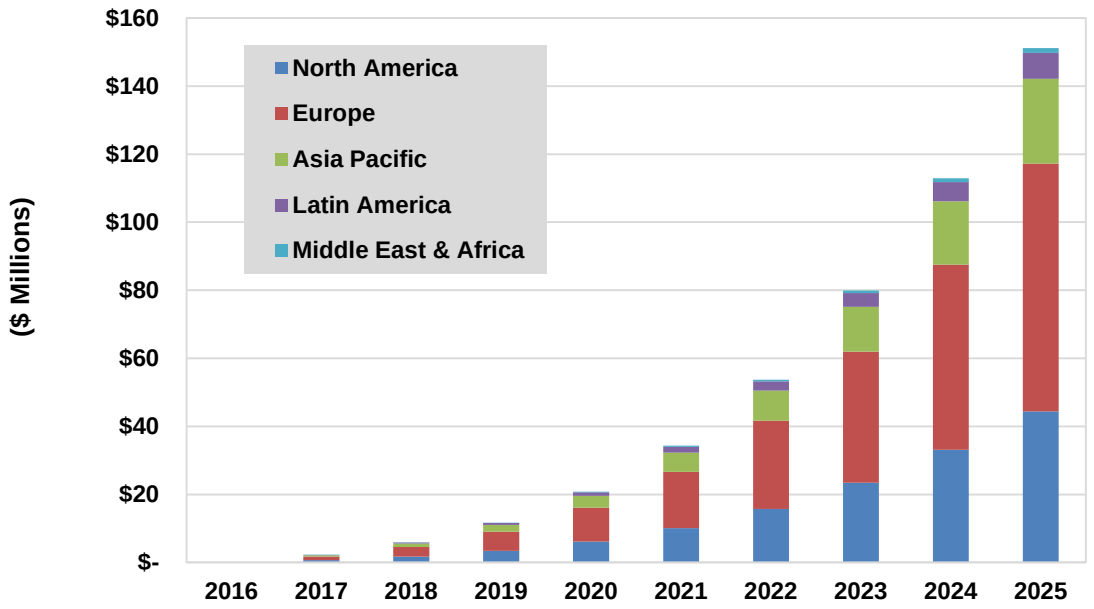
6.2.15 DEEP LEARNING IN THE FASHION INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in fashion will increase from $2.2 million worldwide in 2017 to $151.2 million in 2025.

The use case considered in this forecast is:

- Fashion trend prediction

Chart 6.15 Deep Learning Software Revenue in the Fashion Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

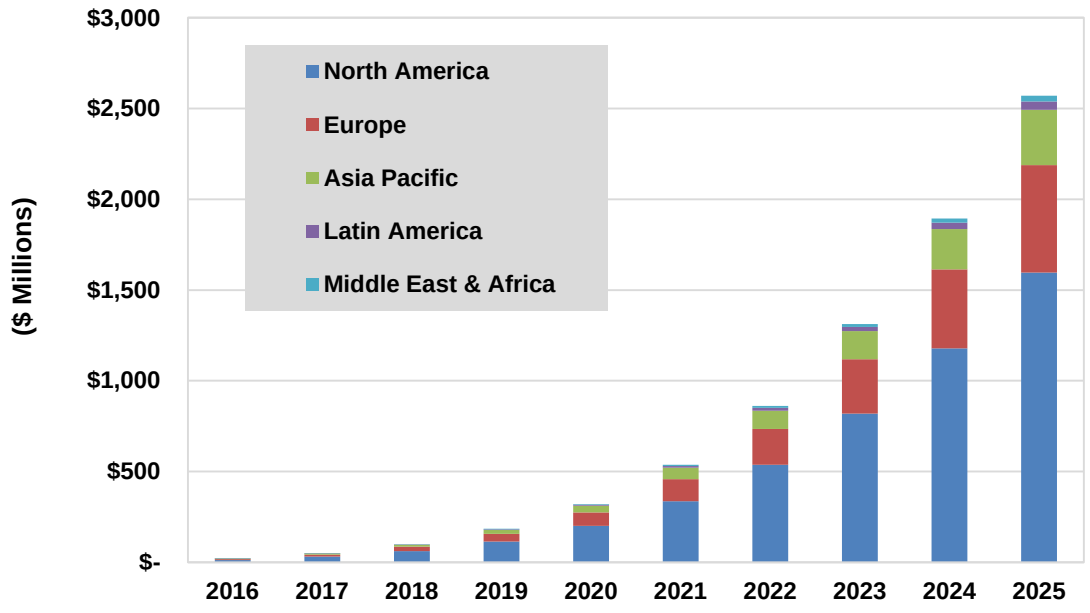
6.2.16 DEEP LEARNING IN THE FINANCE INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in finance will increase from $21.2 million worldwide in 2016 to $2.6 billion in 2025. Deep learning is being applied to the finance industry to aid in mining unstructured data, fraud prevention, and enable "smarter" and more efficient transactions processing. Given the massive amounts of data this industry generates every day, Tractica expects significant uptake in finance, particularly in North America.

Use cases considered in this forecast include:

- Biometric identification
- Converting paperwork into digital data
- Patient data processing
- Employee expense management
- Risk assessment and compliance
- Tax filing and processing

Chart 6.16 Deep Learning Software Revenue in the Finance Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

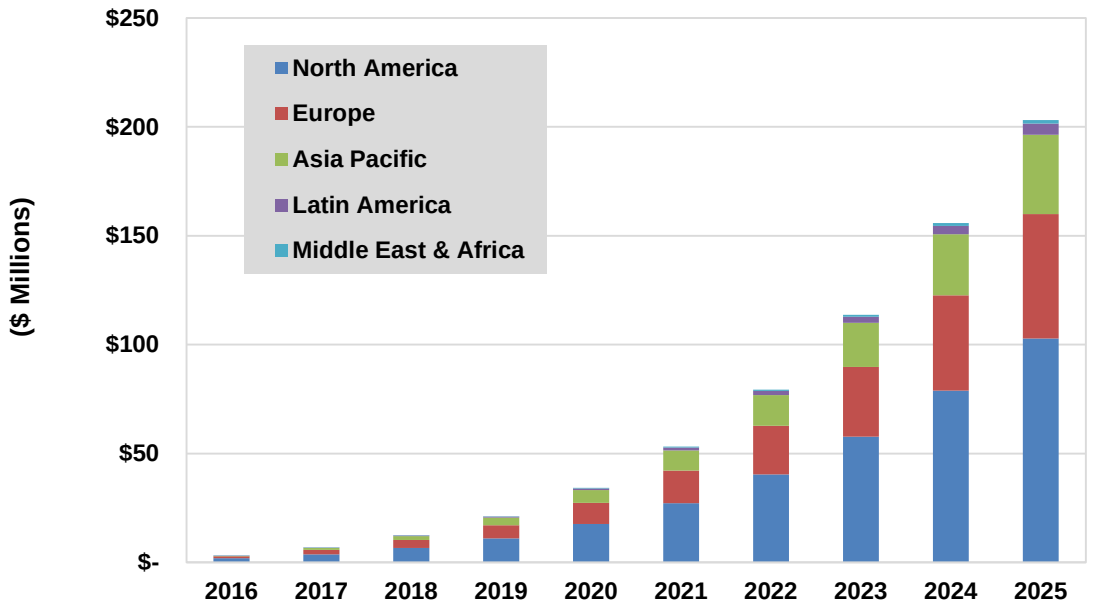
6.2.17 DEEP LEARNING IN THE GAMING INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in gaming will increase from $3.1 million worldwide in 2016 to $203 million in 2025.

The use case considered in this forecast is:

- Dynamic and interactive video game experiences

Chart 6.17 Deep Learning Software Revenue in the Gaming Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

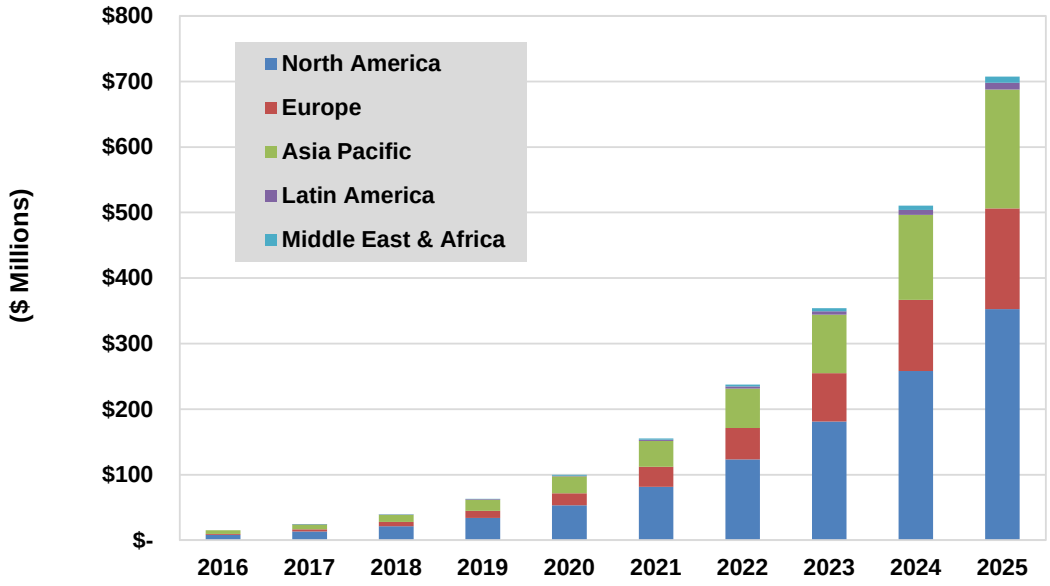
6.2.18 DEEP LEARNING IN THE GOVERNMENT SECTOR

Tractica forecasts that the annual revenue for applications of deep learning in government will increase from $15 million worldwide in 2016 to $707.3 million in 2025.

Use cases considered in this forecast include:

- Agent-based simulations for decision making
- Behavioral analytics
- Converting paperwork into digital data
- Crowd analytics
- Dialect classification
- Face recognition
- Object detection for surveillance
- Predicting social unrest and geopolitical events
- Real-time video analytics
- Sentiment analysis
- Street lighting
- Waste sorting and recycling
- Weather forecasting
- Crime reduction and prevention

Chart 6.18 Deep Learning Software Revenue in the Government Sector by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.2.19 DEEP LEARNING IN THE HEALTHCARE INDUSTRY

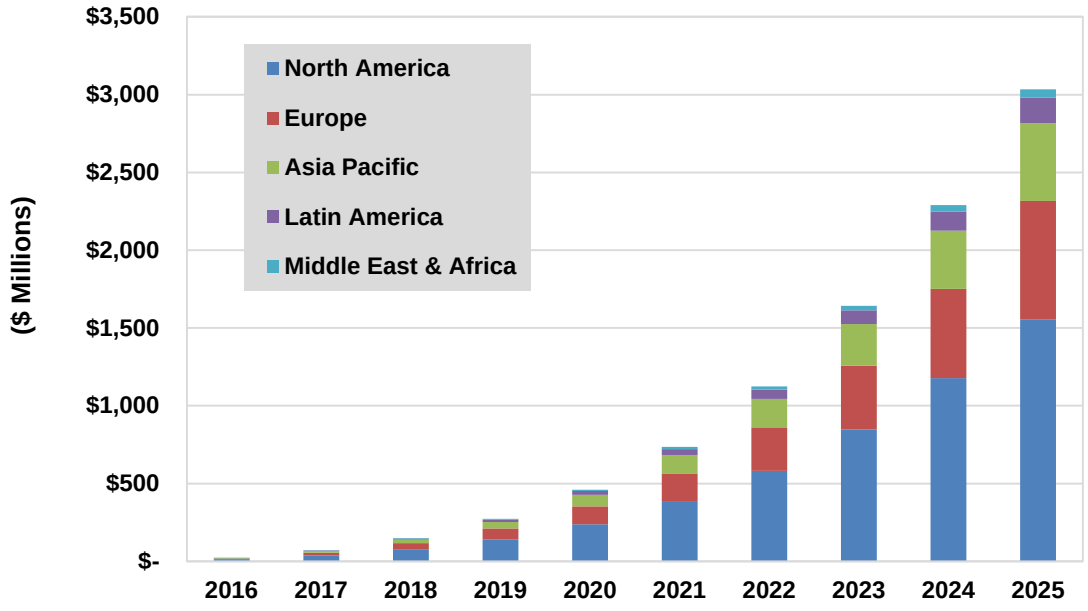
Tractica forecasts that the annual revenue for enterprise applications of deep learning in healthcare will increase from $19.9 million worldwide in 2016 to $3 billion in 2025.

Use cases considered in this forecast include:

- Automated report generation
- Bio-marker discovery
- Clustering and phenotype discovery
- Computational drug discovery
- Converting paperwork into digital data
- Face recognition
- Genomic data mapping and analysis for personalized healthcare and precision medicine
- Market intelligence for life sciences
- Medical diagnosis assistance
- Medical image analysis (3D computer vision, radiology, eye diseases)
- Medical treatment recommendation
- Medication compliance for clinical trials and general usage
- Methods for monitoring vitals
- Patient data processing (administration, clinical medicine)
- Portable and low-cost ultrasound device
- Predicting illness and patient outcomes
- Text classification and mining for biomedical literature (clinical medicine, public health)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Chart 6.19 Deep Learning Software Revenue in the Healthcare Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

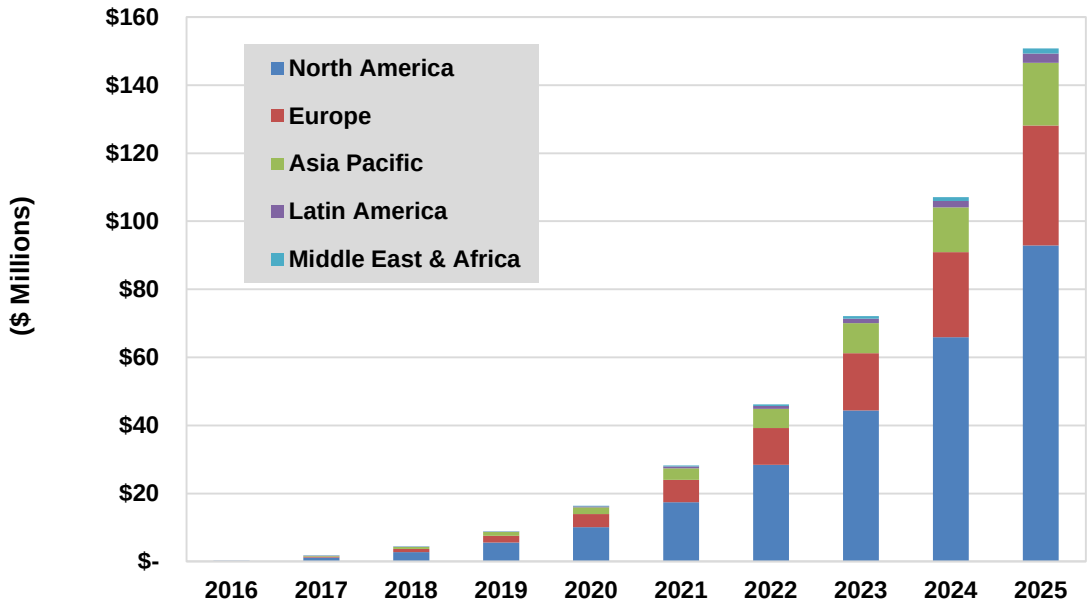
6.2.20 DEEP LEARNING IN THE INFORMATION TECHNOLOGY INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in information technology will increase from $0.2 million worldwide in 2016 to $150.8 million in 2025.

Use cases considered in this forecast include:

- Computer-aided design
- Network/IT operations monitoring and management
- Simulating worlds for AI training
- Website creation

Chart 6.20 Deep Learning Software Revenue in the Information Technology Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

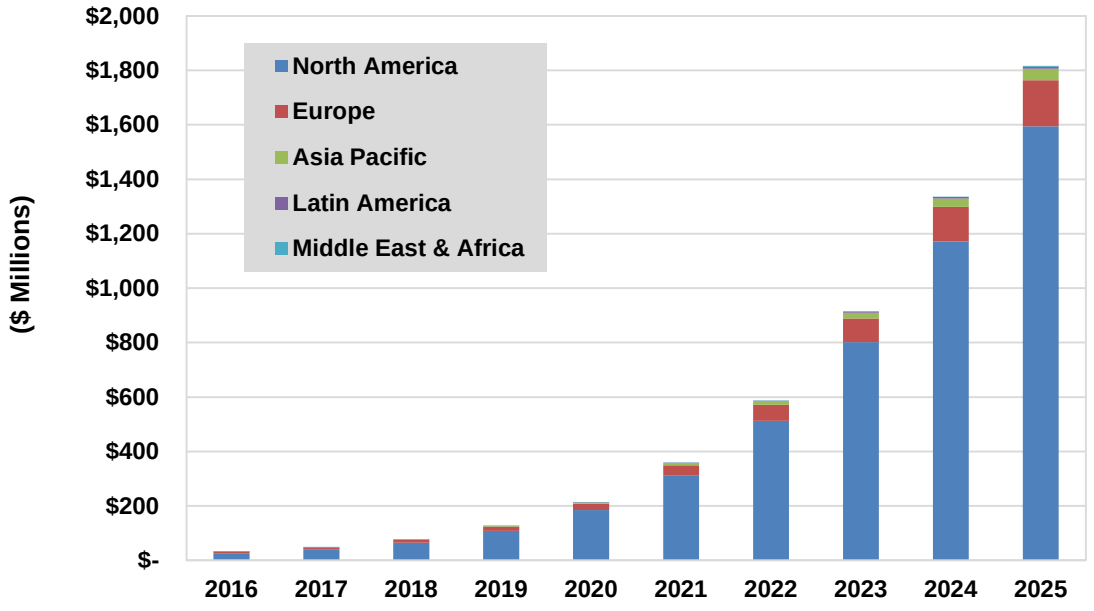
6.2.21 DEEP LEARNING IN THE INVESTMENT INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in investment will increase from $32 million worldwide in 2016 to $1.8 billion in 2025.

Use cases considered in this forecast include:

- Algorithmic trading strategy performance improvement
- Market intelligence and data analytics for investment
- Satellite imagery for geo-analytics

Chart 6.21 Deep Learning Software Revenue in the Investment Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

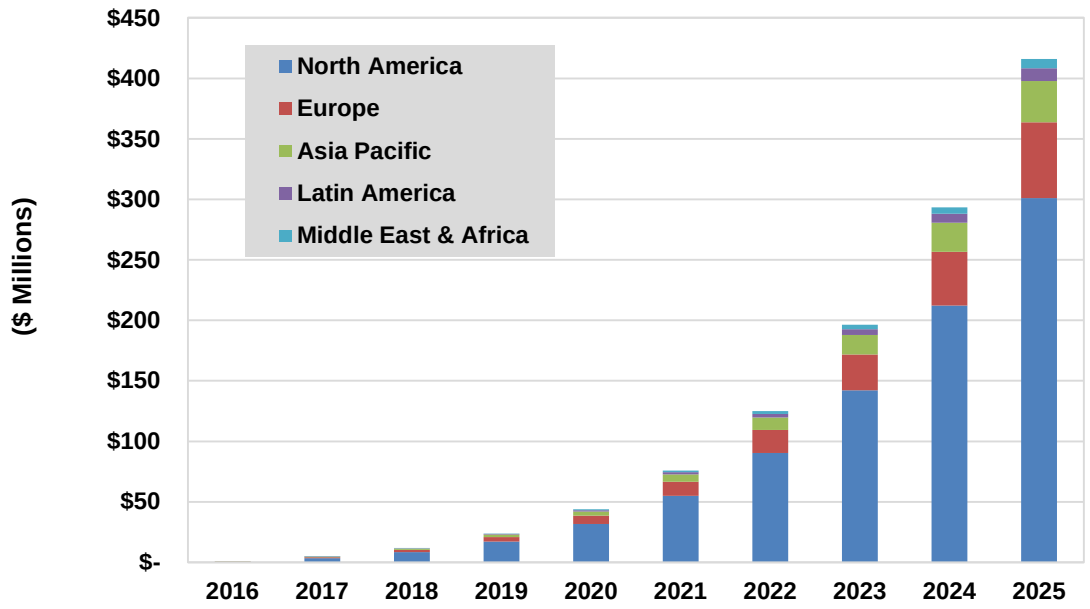
6.2.22 DEEP LEARNING IN THE LEGAL INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in legal will increase from $0.9 million worldwide in 2016 to $416 million in 2025.

Use cases considered in this forecast include:

- Automated report generation

Chart 6.22 Deep Learning Software Revenue in the Legal Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

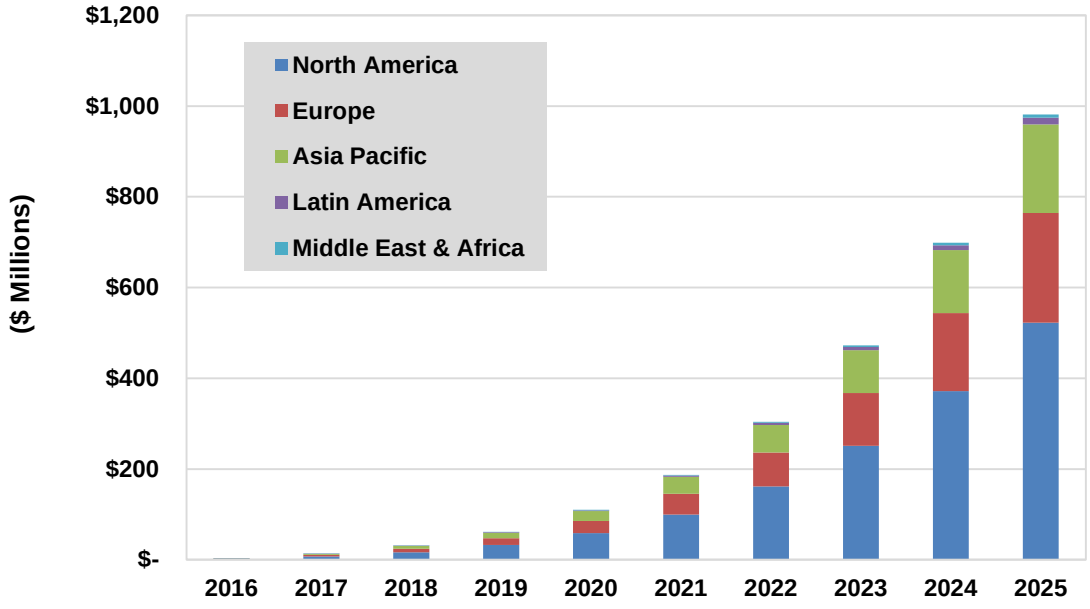
6.2.23 DEEP LEARNING IN THE LOGISTICS INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in logistics will increase from $3.4 million worldwide in 2016 to $981 million in 2025.

Use cases considered in this forecast include:

- Demand forecasting for warehouse and supply chain
- Machine/vehicular object detection/identification/avoidance
- Localization and mapping
- Satellite imagery for geo-analytics
- Supply chain & logistics (freight transport, retail)
- Weather forecasting

Chart 6.23 Deep Learning Software Revenue in the Logistics Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

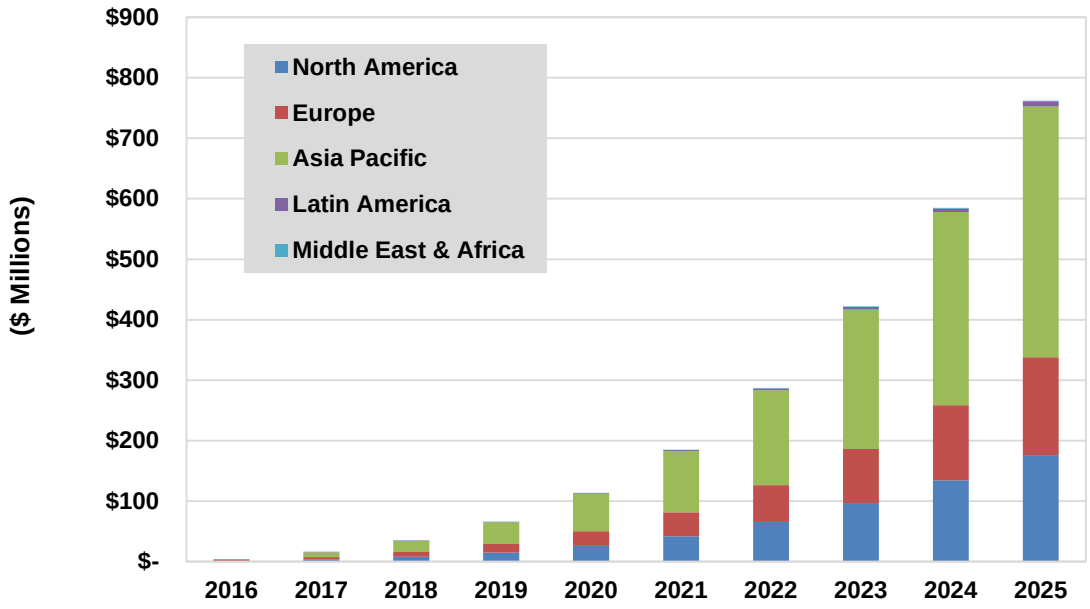
6.2.24 DEEP LEARNING IN THE MANUFACTURING INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in manufacturing will increase from $4.1 million worldwide in 2016 to $762 million in 2025.

Use cases considered in this forecast include:

- 3D printing arm control
- Machine/vehicular object detection/identification/avoidance
- Predictive maintenance (IoT, robotics, autonomous machines)
- Real-time video analytics
- Localization and mapping
- Sensor data fusion in machinery
- Voice/speech recognition
- Product lifecycle management

Chart 6.24 Deep Learning Software Revenue in the Manufacturing Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

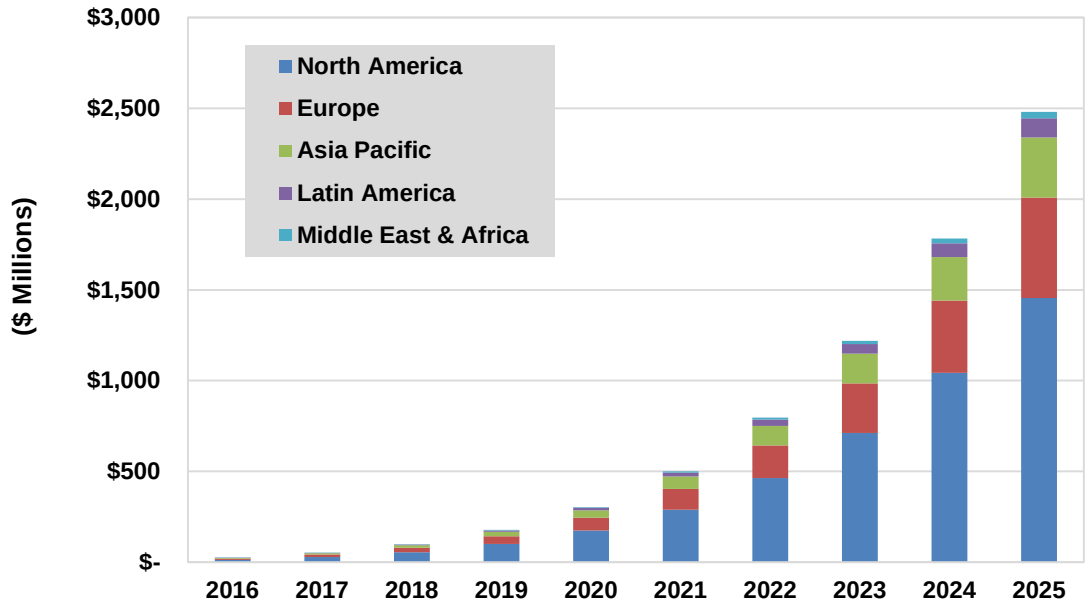
6.2.25 DEEP LEARNING IN THE MEDIA & ENTERTAINMENT INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in media & entertainment will increase from $24.7 million worldwide in 2016 to $2.5 billion in 2025.

Use cases considered in this forecast include:

- Audio and video mining
- Film scene structure
- Font recognition and suggestions
- Gesture recognition
- Human emotion analysis
- Music production and generation
- News and feed curation for consumers
- Simulating crowds (films, games)
- Social media publishing and management
- Video editing

Chart 6.25 Deep Learning Software Revenue in the Media & Entertainment Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

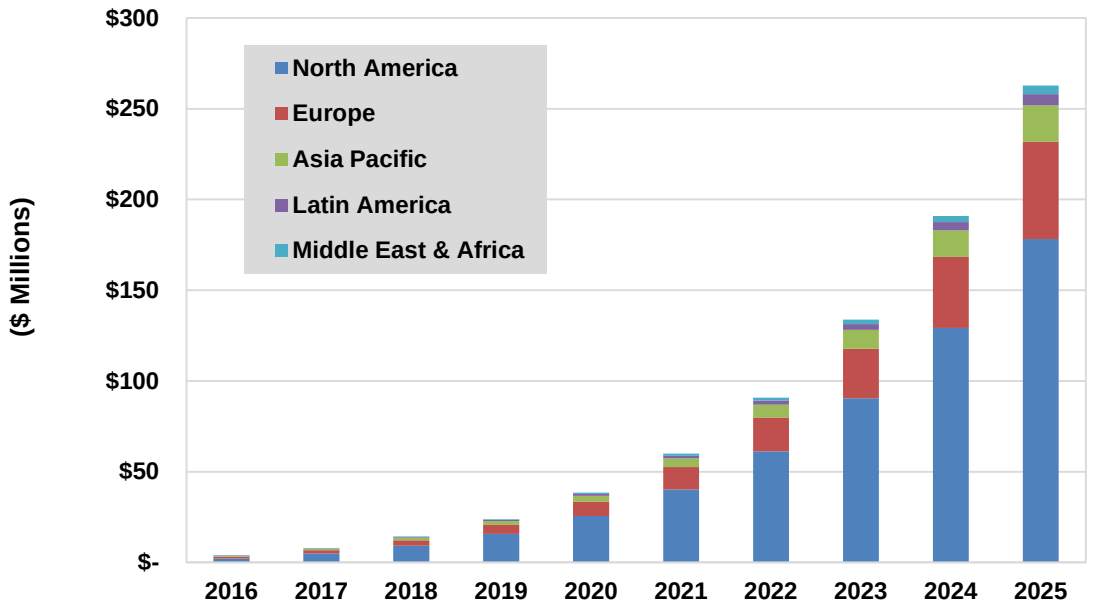
6.2.26 DEEP LEARNING IN THE OIL, GAS, AND MINING INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in oil, gas, and mining will increase from $4 million worldwide in 2016 to $262.8 million in 2025.

The use case considered in this forecast is:

- Automated report generation

Chart 6.26 Deep Learning Software Revenue in the Oil, Gas, and Mining Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

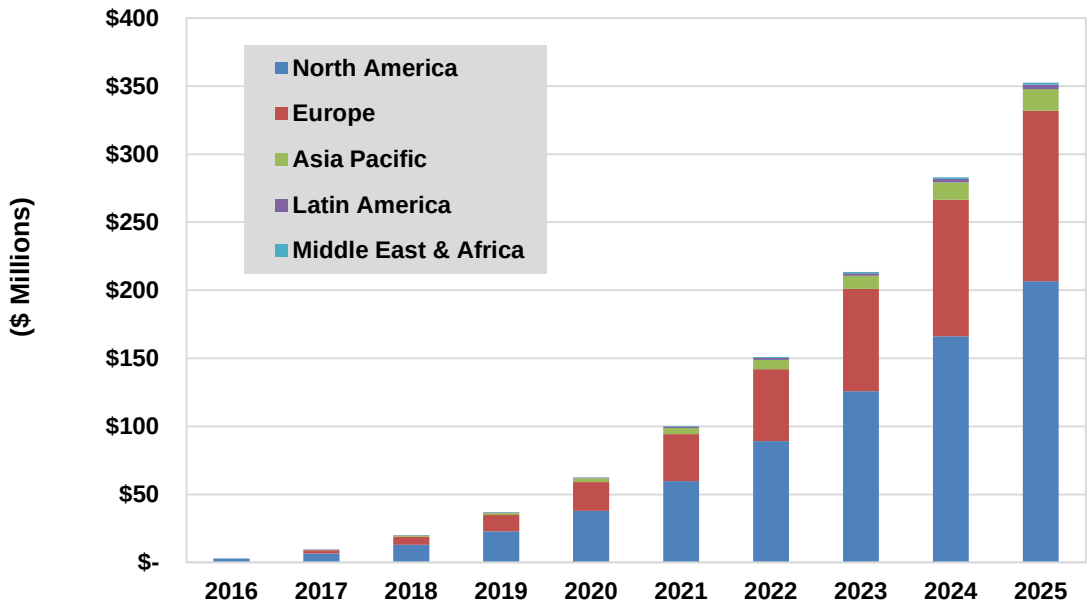
6.2.27 DEEP LEARNING IN THE REAL ESTATE INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in real estate will increase from $3 million worldwide in 2016 to $352.6 million in 2025.

The use case considered in this forecast is:

- Real estate development optimization

Chart 6.27 Deep Learning Software Revenue in the Real Estate Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

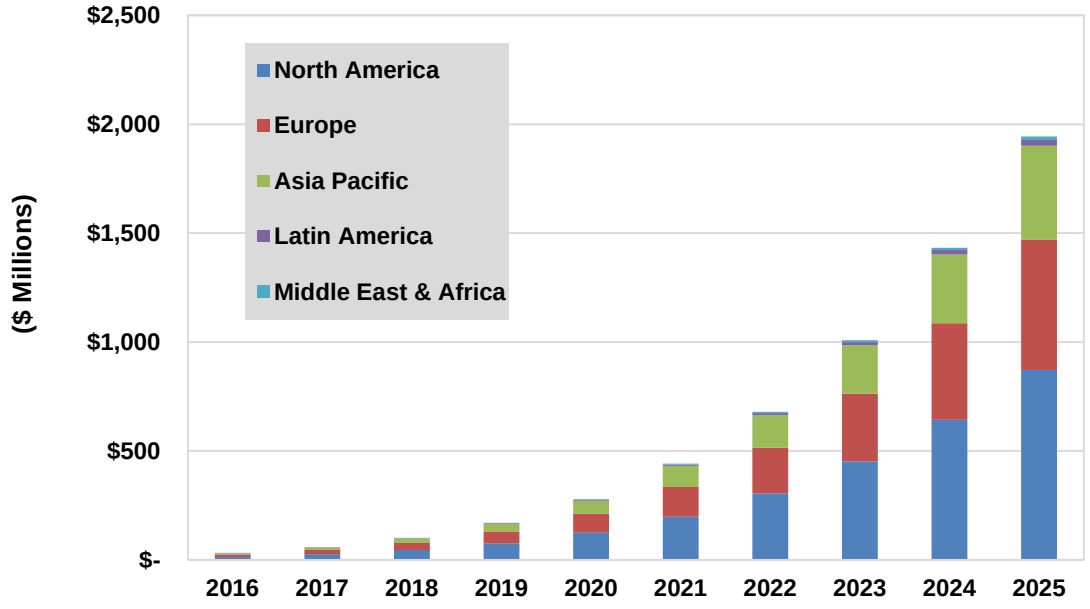
6.2.28 DEEP LEARNING IN THE RETAIL INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in retail will increase from $30.8 million worldwide in 2016 to $1.9 billion in 2025. The common thread running across these retail use cases is that of retailers using AI to become lean, agile, and more responsive in order to compete with e-commerce giants like Amazon.

Use cases considered in this forecast include:

- Behavioral analytics
- Crowd analytics
- Intelligent CRM systems
- Predictive analytics for retail
- Sentiment analysis
- Supermarket shelf analytics
- Visual search based e-commerce
- Weather forecasting

Chart 6.28 Deep Learning Software Revenue in the Retail Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

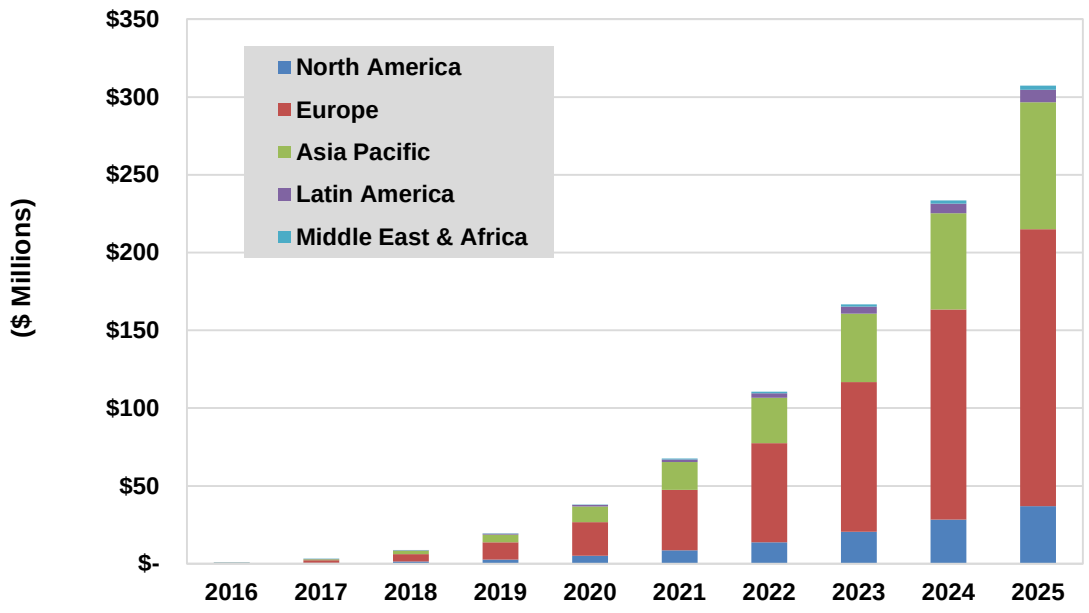
6.2.29 DEEP LEARNING IN THE SPORTS INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in sports will increase from $0.39 million worldwide in 2016 to $307 million in 2025.

Use cases considered in this forecast include:

- Biomarker-based athlete performance optimization
- Game outcome predictions for betting
- Sports statistics analysis and search
- Sports teams player selection

Chart 6.29 Deep Learning Software Revenue in the Sports Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

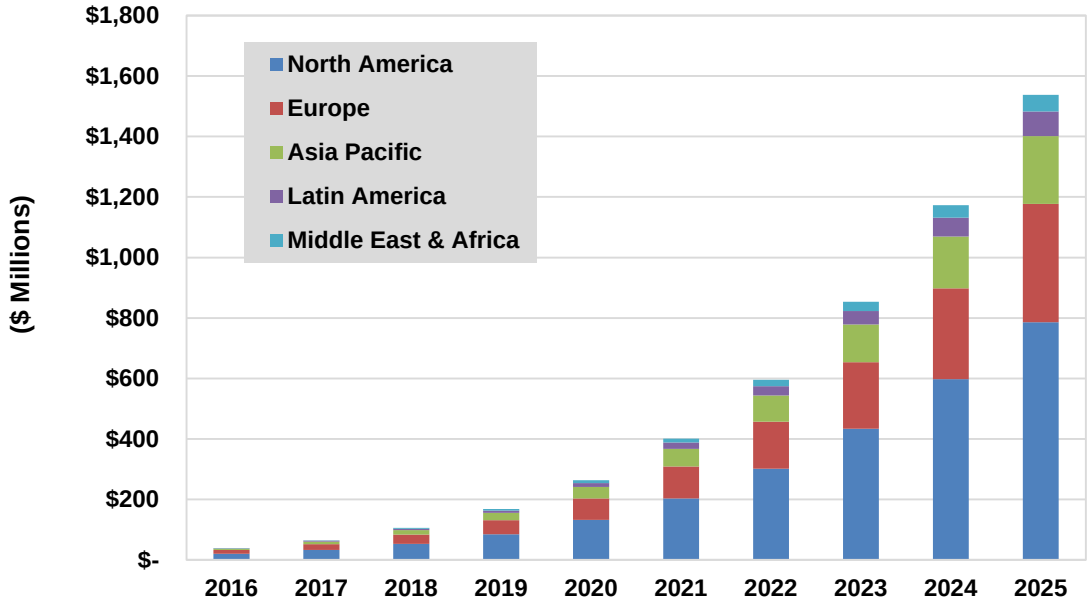
6.2.30 DEEP LEARNING IN THE TELECOMMUNICATIONS INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in telecommunications will increase from $38.9 million worldwide in 2016 to $1.5 billion in 2025. The general adoption of AI within telecoms is part of the general move toward software-defined networks (SDNs) from fixed-rule hardware systems, allowing network operations to be flexible with the changing needs of the customer, and have the same flexibility and leanness as some of the internet companies.

Use cases considered in this forecast include:

- Predictive maintenance
- Prevention against cybersecurity threats
- Improve customer experience management
- Fraud mitigation
- Intelligent CRM systems

Chart 6.30 Deep Learning Software Revenue in the Telecommunications Industry by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

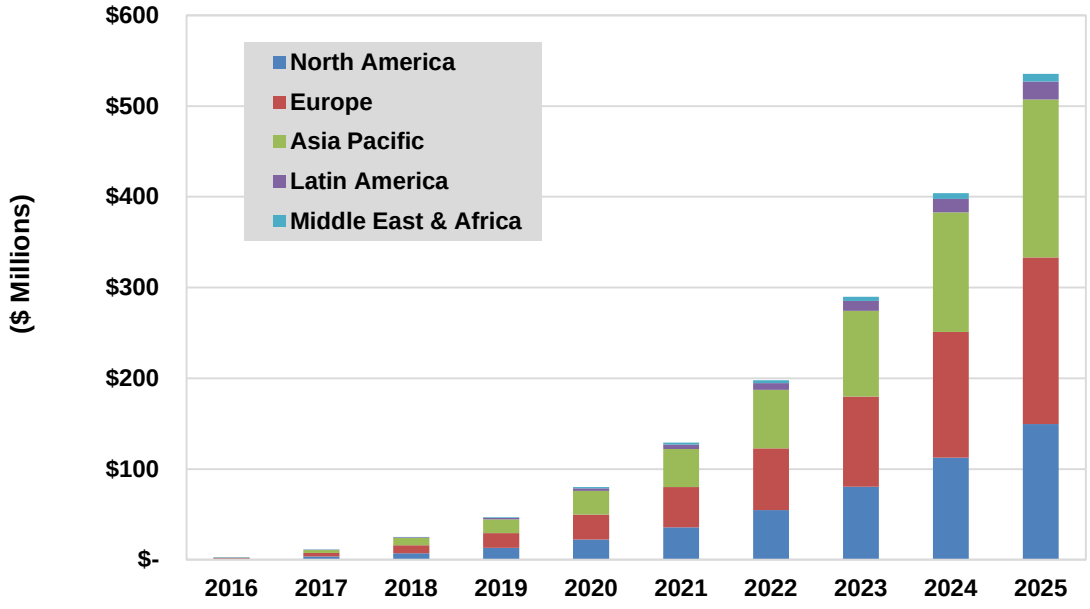
6.2.31 DEEP LEARNING IN THE TRANSPORTATION INDUSTRY

Tractica forecasts that the annual revenue for enterprise applications of deep learning in transportation will increase from $2.5 million worldwide in 2016 to $535 million in 2025.

Use cases considered in this forecast include:

- Machine/vehicular object detection/identification/avoidance
- Predicting traffic density
- Sensor data fusion in machinery (ships, unmanned ships)
- Localization and mapping
- Vehicle network and data security
- Weather forecasting

Chart 6.31 Deep Learning Software Revenue in the Transportation Industry by Region, World Markets: 2016-2025



(Source: Tractica)

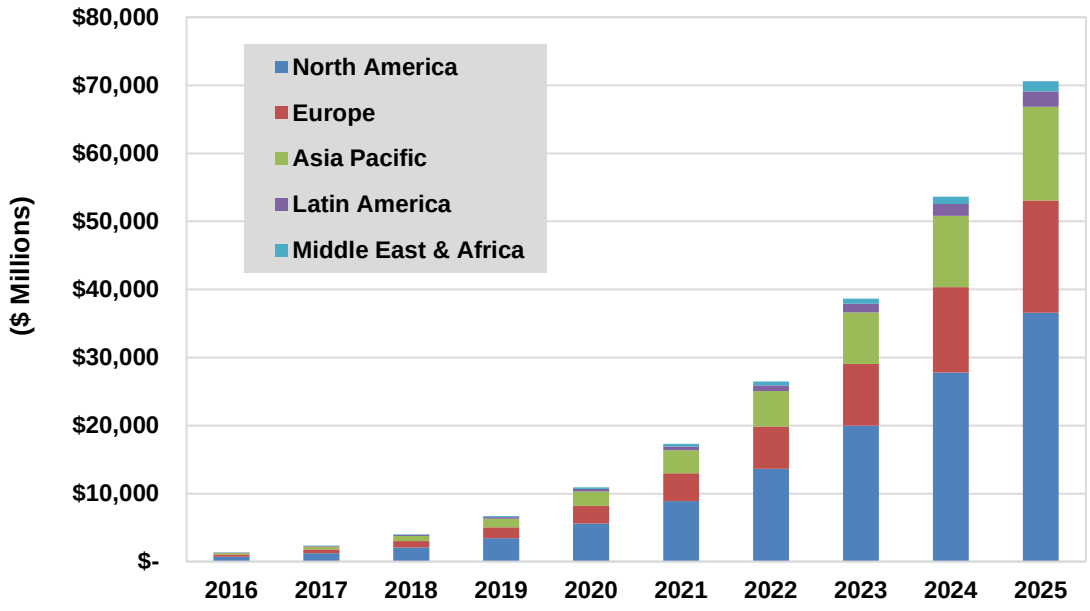
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.3 DEEP LEARNING-DRIVEN SERVICES REVENUE

Deep learning systems will require a higher level of professional services, particularly as it grows in applications involving decision-making and any mission-critical task. Because they guide other systems and condition human behavior in ways other parts of the software stack do not, the revenue spent to supporting them will be even greater than what is spent on ERP or other transaction systems. As a result, revenue spent to install, customize, integrate, and support AI and deep learning systems will likely be many times greater than the original purchase price of the AI software.

Tractica estimates overall deep learning-driven services revenue will grow from $1.3 billion in 2016 to $70.6 billion by 2025.

Chart 6.32 Deep Learning Services Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

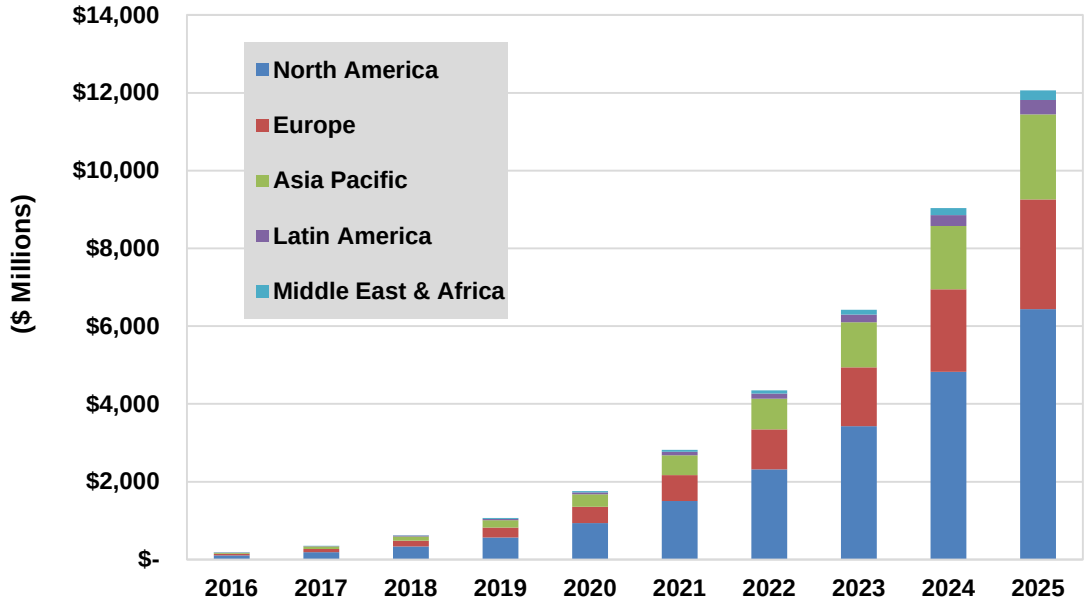
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.3.1 DEEP LEARNING-DRIVEN INSTALLATION SERVICES

Thanks to the advent of cloud computing, installation services will be concentrated in areas of political and geophysical stability where easy access to low-cost energy is available. Most installation services will be provided in North America, Western Europe, and Asia Pacific. On the other hand, geopolitical reality and national interests will require that AI systems be spread across all national boundaries, so there will be some demand everywhere in the world. The ratio of installation services will initially be the largest in North America, but eventually, Asia Pacific will surpass it and Western Europe.

Tractica estimates deep learning-driven installation services revenue will grow from $182 million in 2016 to $12.1 billion by 2025.

Chart 6.33 Deep Learning Installation Services Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

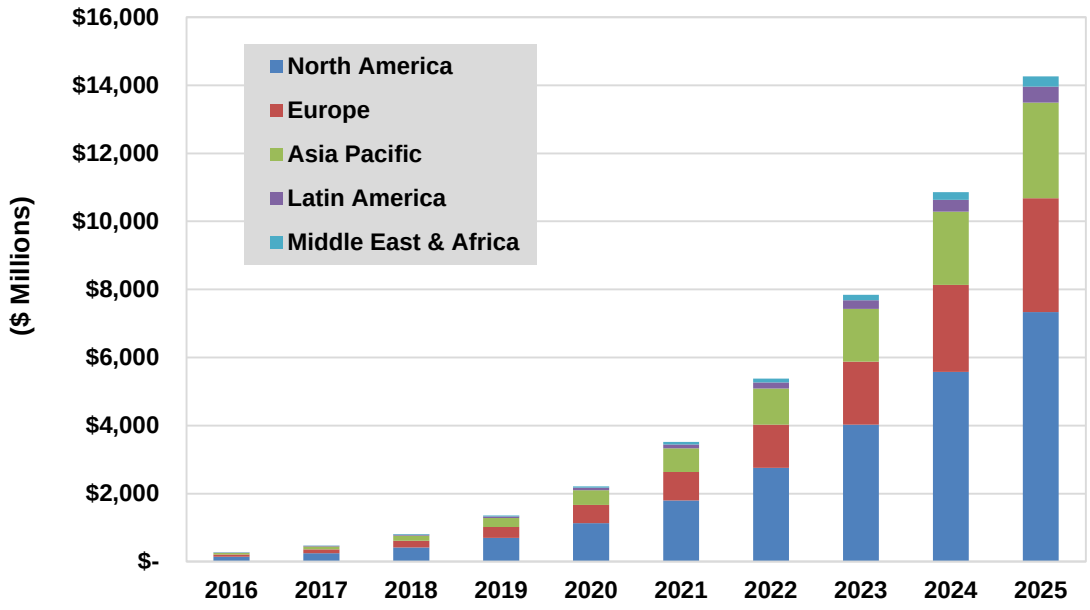
6.3.2 DEEP LEARNING-DRIVEN TRAINING SERVICES

Deep learning systems, in particular, require a high level of complex training to implement. This is true at both the technology and human levels; training deep learning models is highly compute-intensive, as is training humans on value and methods for integrating (and trusting) deep learning models into workflows. In order to adapt both machines and people to this way of doing business, training revenue, in particular, will be significantly higher than traditional systems.

Training is an area of services that will grow the fastest and will become the single biggest services category. Since the architecture of backwards-chaining neural nets used in deep learning systems requires significant amounts of data and GPU time to work, most training work will be done in the cloud, even if the AI system itself remains on-premise. Tractica expects most training services will be provided by computing clouds based in relatively concentrated areas in North America, Europe, and Asia Pacific.

Tractica estimates deep learning-driven training services revenue will grow from $269 million in 2016 to $14.3 billion by 2025.

Chart 6.34 Deep Learning Training Services Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

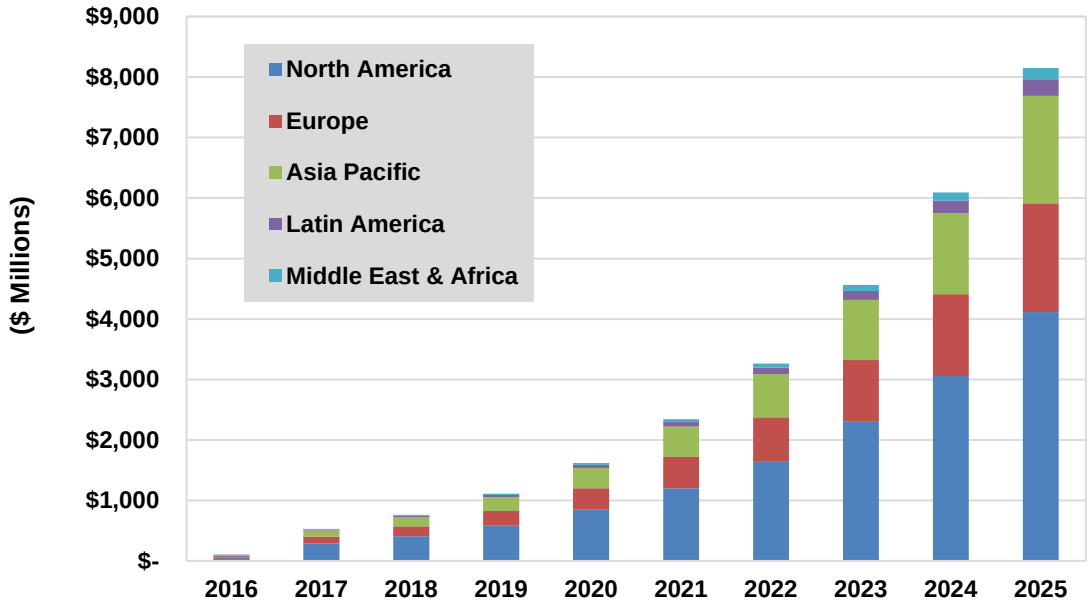
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.3.3 DEEP LEARNING-DRIVEN CUSTOMIZATION SERVICES

Given the highly sensitive and, in some cases, mission-critical application opportunities for deep learning, Tractica expects a fairly significant level of customization will be required. These complex customization services will drive a significant amount of ancillary revenue associated with AI deployments. Most customization services will need to be provided on-premise.

Tractica estimates deep learning-driven customization services revenue will grow from $102.5 million in 2016 to $8.2 billion by 2025.

Chart 6.35 Deep Learning Customization Services Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

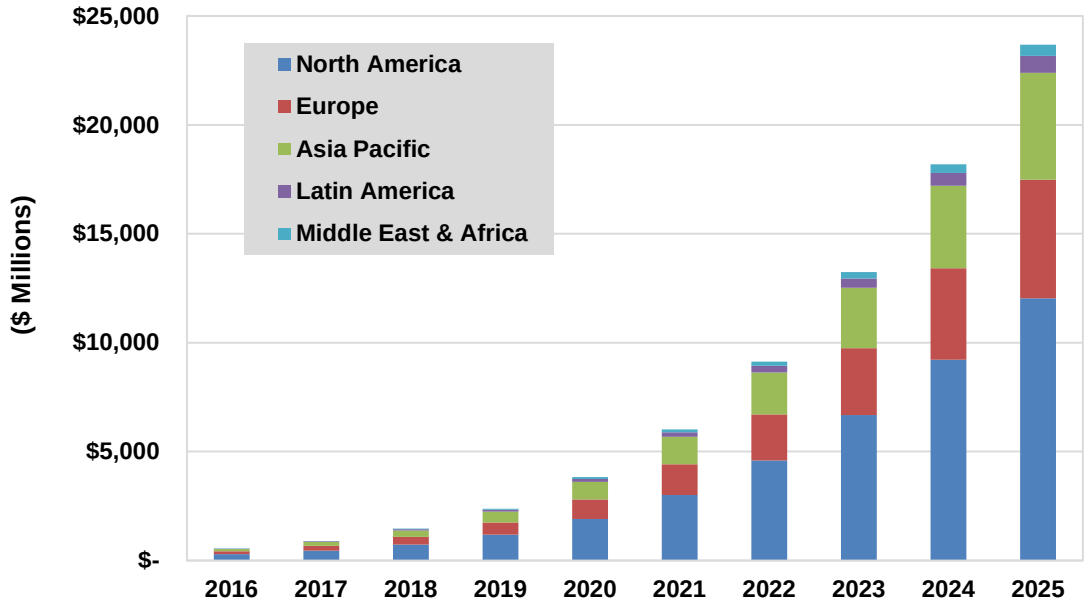
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.3.4 DEEP LEARNING-DRIVEN APPLICATION INTEGRATION SERVICES

As AI and deep learning systems become more integral to central business operations, they will be more and more integrated with other systems. Integration services will need to be provided both on-premise and in the cloud depending on the nature of the services.

Tractica estimates deep learning-driven application integration services revenue will grow from $556 million in 2016 to $23.7 billion by 2025.

Chart 6.36 Deep Learning Application Integration Services Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

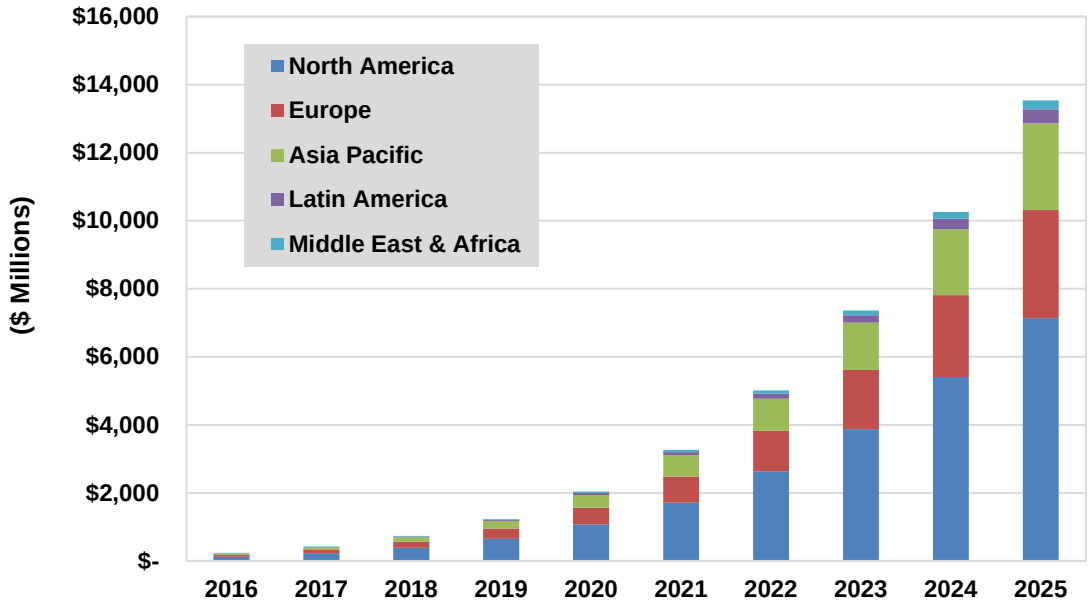
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.3.5 DEEP LEARNING-DRIVEN MAINTENANCE AND SUPPORT SERVICES

Maintenance and support will be key to the ongoing operation and enhancement of AI and deep learning systems. Tractica expects the ongoing maintenance and support of deep learning systems will receive a disproportionate share of IT resources relative to traditional enterprise systems, particularly if and when deep learning systems become critical to the success of the organization itself.

Tractica estimates deep learning-driven maintenance and support services revenue will grow from $231 million in 2016 to $13.5 billion by 2025.

Chart 6.37 Deep Learning Maintenance and Support Services Revenue by Region, World Markets: 2016-2025



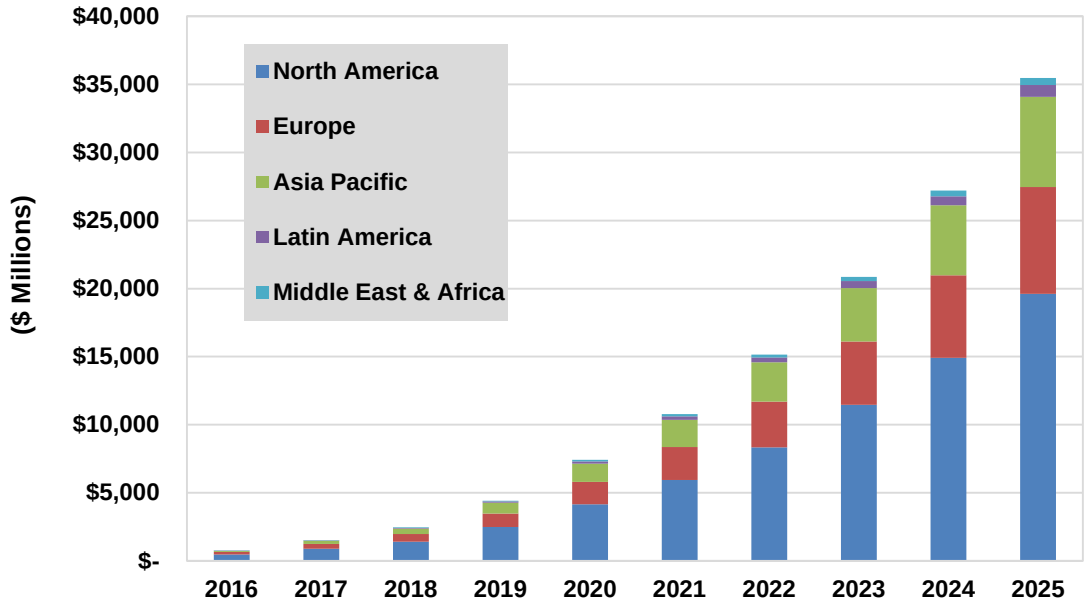
(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.4 DEEP LEARNING-DRIVEN CLOUD SERVICES REVENUE

Cloud computing plays a critical role in AI, especially in the training of AI systems, but given the extremely sensitive nature of deep learning-driven decision making, many organizations will not be comfortable putting these systems in the cloud. As a result, AI systems are likely to be the last systems to move off-premise and promise to give internal data systems new life. Tractica estimates deep learning-driven cloud services revenue to grow from $757 million in 2016 to $35.5 billion by 2025.

Chart 6.38 Deep Learning-Driven Cloud Services, World Markets: 2016-2025



(Source: Tractica)

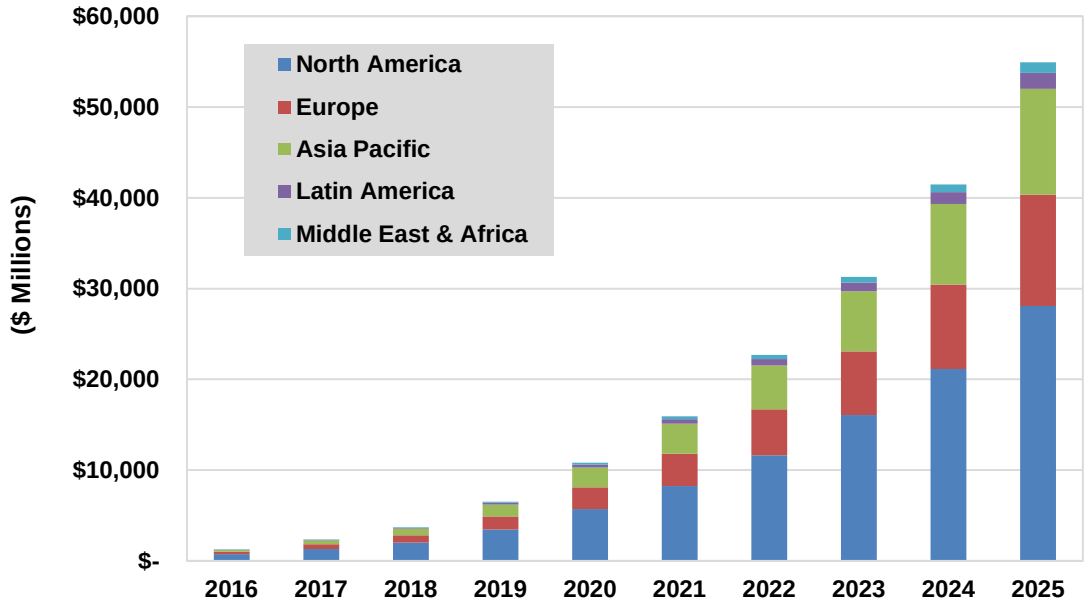
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.5 DEEP LEARNING-DRIVEN HARDWARE REVENUE

Vendors of servers, networks, and storage, as well as cloud providers, can expect to see the adoption of deep learning contribute significantly to both their top and bottom lines. In the deep learning space, this is especially true for GPU vendors today, and over time, for CPU, ASIC, and FPGA components. Given the data and compute-intensive realities of deep learning systems, Tractica expects sales of hardware to grow at a much higher rate than the mission-critical systems these hardware categories have seen in the past.

Pure hardware components, including CPUs/ASICs/FPGAs, GPUs, network products, and storage devices are estimated to grow from $1.3 billion in 2016 to $41.4 billion by 2025.

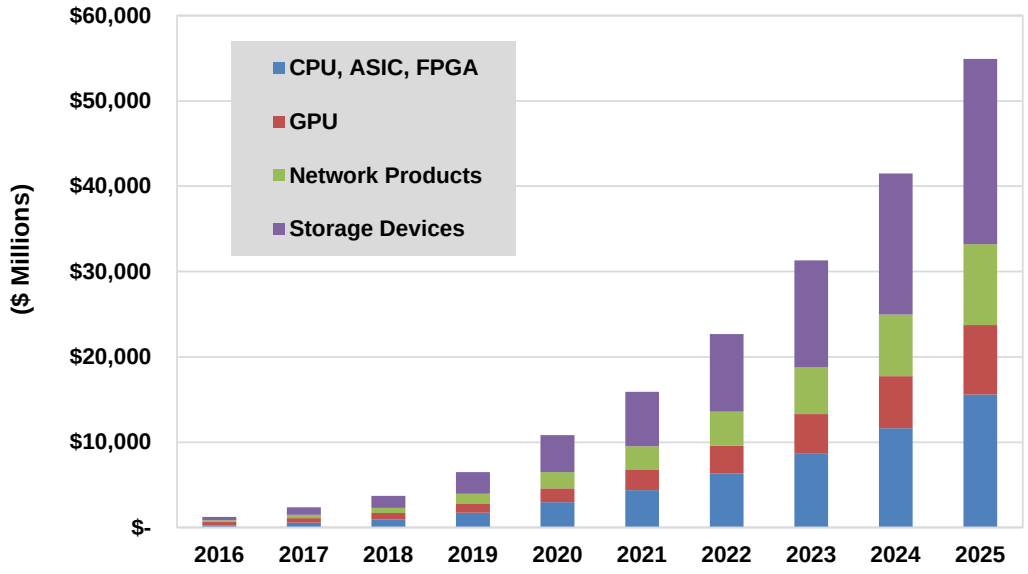
Chart 6.39 Deep Learning Hardware Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Chart 6.40 Deep Learning Hardware Revenue by Product Category, World Markets: 2016-2025

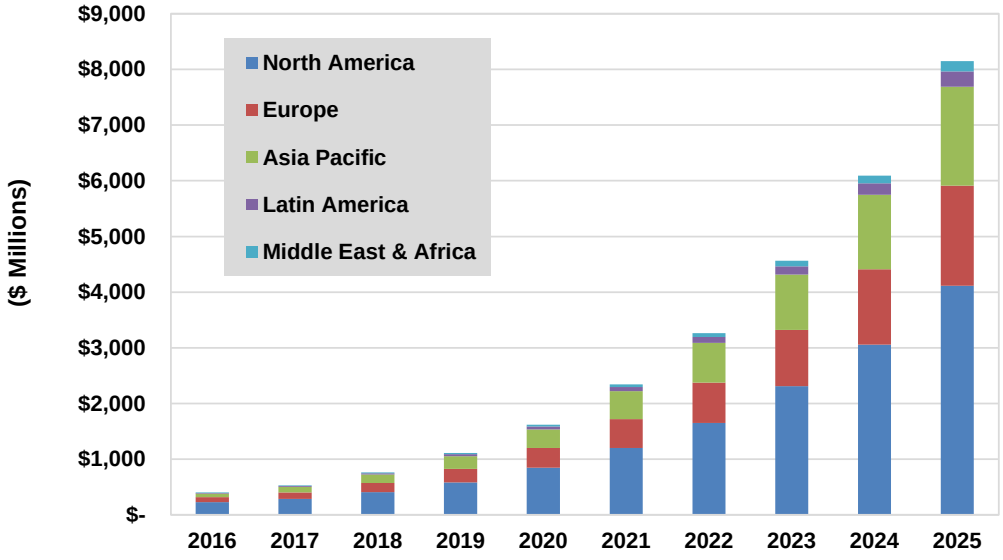


(Source: Tractica)

6.5.1 DEEP LEARNING-DRIVEN CHIPSET REVENUE

Thanks to Andrew Ng and his team's discovery at Stanford that the GPU's speed at floating point calculations could be harnessed to train and run neural nets in parallel, deep learning systems are forecast to have an impact on the sale of GPU chips, as they become more and more adopted across industries.

Chart 6.41 Deep Learning-Driven GPU Revenue by Region, World Markets: 2016-2025



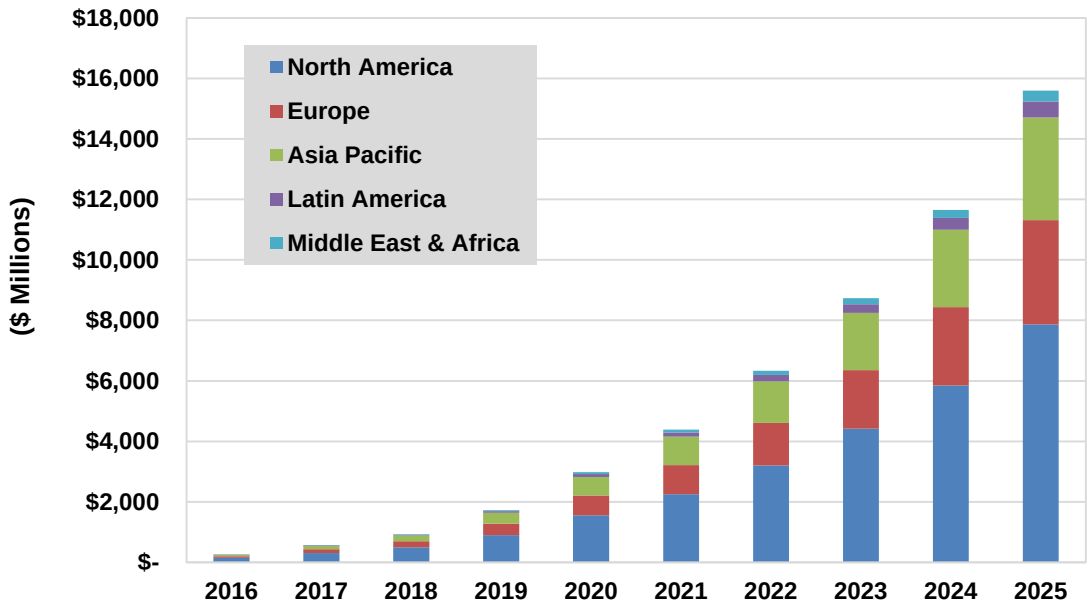
(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

AI systems will mostly need to be trained in the cloud, but once that process is completed, they will, for the most part, operate on-premise where they will require more compute resources, as both the responsibility and the size of AI application increases.

The GPU market in 2016 for deep learning hardware is estimated to be $403 million, with revenue growing to $8.2 billion by 2025 at a CAGR of 39.7%. NVIDIA makes up almost 100% of the GPU market today with the majority of the revenue being spent on training. The CPU/ASIC/FPGA market, which is dominated by CPUs today, is led by Intel's chips, largely used for inference, with revenue in 2016 estimated to be $258 million. ASICs are expected to grow their share as new custom architectures start to increase their presence in the AI hardware market giving competition to GPUs. The CPU/ASIC/FPGA category will see its revenue reach $15.6 billion by 2025.

Chart 6.42 Deep Learning-Driven CPU, ASIC, FPGA Revenue by Region, World Markets: 2016-2025



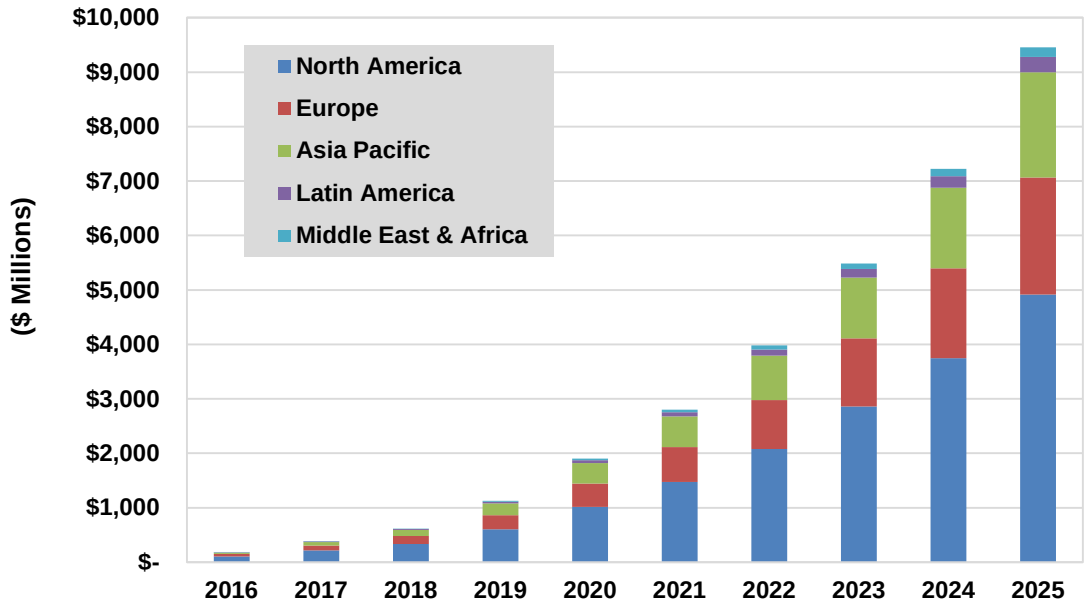
(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.5.2 DEEP LEARNING-DRIVEN NETWORK PRODUCT REVENUE

Deep learning systems will require significant networking resources to integrate such systems with other enterprise IT systems, as well as to other like systems and to the cloud. Tractica estimates deep learning-driven network product revenue will grow from $183 million in 2016 to $9.5 billion by 2025.

Chart 6.43 Deep Learning-Driven Network Products Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

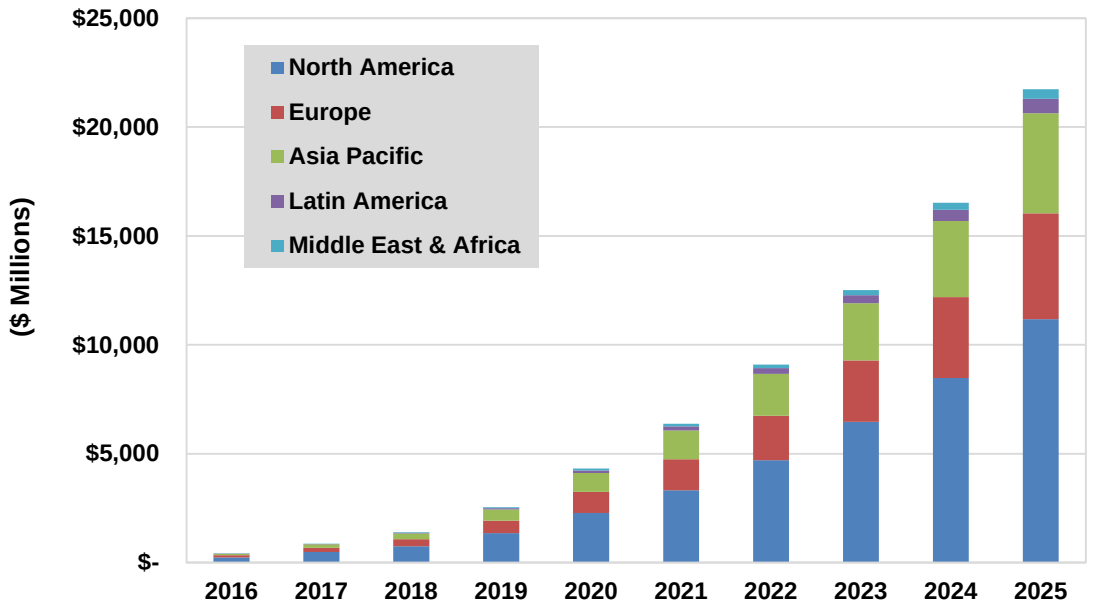
© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

6.5.3 DEEP LEARNING-DRIVEN STORAGE DEVICE REVENUE

The data-intensive nature of deep learning systems makes them prime candidates for high-end storage devices, as data processed in deep learning systems will drive significant demands on storage capabilities.

Tractica estimates deep learning-driven storage device revenue will grow from $415 million in 2016 to $21.7 billion by 2025.

Chart 6.44 Deep Learning-Driven Storage Device Revenue by Region, World Markets: 2016-2025



(Source: Tractica)

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 7

KEY FINDINGS, RECOMMENDATIONS, AND CONCLUSION

7.1 KEY TAKEAWAYS

- **Deep learning is best suited for handling massive data sets,** perception tasks, expedited feature extraction, and applications with extensive domain knowledge.
- **Despite aspirations for biomimicry, human intelligence is not the proxy for deep learning success**; understanding the limitations of AI is the first step toward realizing how useful it might be in improving and identifying suitable business applications.
- **Recent advancements accelerate innovation, but applications remain narrow**. Today's state-of-the-art AI is still very much "narrow AI," which works in specific settings and applications, rather than AGI that can work across a variety of applications, without any prior knowledge or training.
- **Deep learning suffers from an introspection problem.** A human's ability to see into and understand why neural networks arrive at specific outcomes is opaque at best, a significant hurdle for adoption in certain scenarios.
- **Deep learning is not a silver bullet, but an enabling technology** that will drive the greatest disruption when applied in conjunction with other technologies, namely machine perception, language processing, Big Data, the IoT, cybersecurity, etc.

7.2 RECOMMENDATIONS

Deep learning is a subset of machine learning and is a powerful tool applied in some but not all scenarios. It has the potential to disrupt numerous industries, workflows, jobs, and mechanisms for knowledge generation and sharing. Tractica recommends that almost every business, but especially businesses that are transaction-oriented or that deal with large amounts of data, begin piloting projects using deep learning to understand how it is different from other technologies. Begin with the following steps:

- **Invest in Understanding:** Invest in time, guidance, and talent to educate internal stakeholders and leadership about machine and deep learning, particularly areas of application, differentiation, overhype, controversy, and risk.
- **Define the Problem:** Begin not with deep learning, but with current pain points and problems. Know your highest-impact decision bottlenecks. Deep learning is best applied to very specific questions and scoped problems, rather than general issues or experiments.
- **Prioritize Data Integrity:** The old IT axiom, "trash in-trash out" is never truer or more important than in the context of deep learning. Regardless of familiarity with deep learning, all enterprises should be prioritizing data cleansing, standardizing, consolidation, and formalizing processes to maintain and optimize data integrity of internal and external data sources.
- **Develop Talent:** Regardless of the serious lack of talent in deep learning skills, enterprises should be forging new skill sets internally and aligning with external organizations (e.g., universities, consortia, open-source communities, etc.) to recruit talent.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

- **Build Collaborations:** Tap into open-source communities, consortia, partnerships, universities, etc., in order to foster collaborative ideation and development for deep learning initiatives.

- **Monitor, Manage, and Secure:** Enterprises must constantly monitor and provide ongoing maintenance to deep learning models, as well as to other relevant operational analytics. For applications, set, monitor, and evolve key performance indicators (KPIs), and assess risks. Formalize relevant security requirements, such as identity authentication, access controls, auditing, and privacy assessments, related to both model development and performance.

- **Provide Training, Support, and Communications:** It is also essential to coordinate necessary training and communications plans for the role of machine learning in employee, partner, and end-user workflows and experiences.

7.3 CONCLUSION

Deep learning enables new capabilities and ways of thinking, both for machines and humans. But despite its potential, and perhaps because of its nature, the technology is also subject to overhype, oversell, under-delivery, and controversy. As we teach machines to perceive and think, it is critical that we design, build, apply, and scale mindfully, with individual and institutional regard for risks, unintended consequences, societal benefit, and human empowerment. Collaboration with broader ecosystems will not just support better enterprise applications, but will enable more intelligent and secure interactions with the world and each other – deeper learning in its most basic sense – for all.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 8

ACRONYM AND ABBREVIATION LIST

Acute Kidney Injury ..... AKI

Advanced Driver Assistance System ..... ADAS

Advanced Persistent Threat ..... APT

Amazon Web Services ..... AWS

Anti-Money Laundering ..... AML

Application Programming Interface ..... API

Application-Specific Standard Processors ..... ASSP

Application-Specific Integrated Circuits ..... ASIC

Artificial General Intelligence ..... AGI

Artificial Intelligence ..... AI

Association of Computing Machinery ..... ACM

Augmented Reality ..... AR

Basic Neural Network Subroutines ..... BNNS

Building Automaton System ..... BAS

Business-to-Business B2B

Central Processing Unit ..... CPU

Chief Executive Officer ..... CEO

Chief Information Officer ..... CIO

Compound Annual Growth Rate ..... CAGR

Computed Tomography ..... CT

Convolutional Neural Network ..... CNN

Customer Experience CX

Customer Relationship Management ..... CRM

Deep Belief Network ..... DBN

Deep Neural Network ..... DNN

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Defense Advanced Research Projects Agency ..... DARPA

Deoxyribonucleic Acid ..... DNA

Diagnostic Trouble Code ..... DTC

Differential Neural Computer ..... DNC

Digital Sign Processor ..... DSP

Distributed Denial of Service ..... DDoS

Enterprise Resource Planning ..... ERP

Estimated Time of Arrival ..... ETA

European Union ..... EU

Facial Action Coding System ..... FACS

Field Programmable Gate Array ..... FPGA

Geographic Information System ..... GIS

Gigabyte ..... GB

Global Positioning System ..... GPS

Government Communications Headquarters (U.K.) ..... GCHQ

Graphics Processing Unit ..... GPU

Gross Domestic Product ..... GDP

High-Performance Computing ..... HPC

High-Performance Embedded Computing ..... HPEC

Holosemantic Dataspace ..... HSDS

Human Resources ..... HR

Internet Protocol ..... IP

Internet of Things ..... IoT

Information Technology ..... IT

Intensive Care Unit ..... ICU

Key Performance Indicator ..... KPI

Know Your Customer ..... KYC

Knowledge-Based System ..... KBS

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Knowledge Discovery in Databases ..... KDD

Light Detection and Ranging ..... LIDAR

Local Interpretable Model Agnostic ..... LIME

Long Short-Term Memory ..... LSTM

Machine-to-Machine M2M

Magnetic Resonance Imaging ..... MRI

Massachusetts Institute of Technology ..... MIT

National Basketball Association ..... NBA

National Energy Research Computing Center ..... NERSC

National Health Service ..... NHS

National Oceanic and Atmospheric Administration ..... NOAA

Natural Language Processing ..... NLP

Neural Information Processing Systems ..... NIPS

New York University ..... NYU

Operational Technology ..... OT

Operating System ..... OS

Original Equipment Manufacturer ..... OEM

Personal Computer ..... PC

Public Relations ..... PR

Radio Frequency ..... RF

Radio Frequency Identification ..... RFID

Recurrent Neural Networks ..... RNN

Red, Green, Blue ..... RGB

Research and Development ..... R&D

Single Instruction Multiple Data ..... SIMD

Small to Medium-Sized Business ..... SMB

Software as a Service ..... SaaS

Software-Defined Network ..... SDN

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Software Development Kit.....SDK

Synthetic Environment for Analysis and Simulations (U.S. Department of Homeland Security)..... SEAS

Television .....TV

Tensor Processing Unit .....TPU

Text-to-Speech..... TTS

Three-Dimensional.....3D

Trade-Based Money Laundering .....TBML

User Experience..... UX

Vice President ..... VP

Virtual Reality ..... VR

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 9

COMPANY DIRECTORY

ai-one
5711 La Jolla Boulevard
La Jolla, CA 92037, USA
www.ai-one.com
+1.858.531.0674

Amazon
410 Terry Avenue
N. Seattle, WA 98109, USA
www.amazon.com
+1.206.266.1000

Apple Inc.
1 Infinite Loop
Cupertino, CA 95014, USA
www.apple.com
+1.408.996.1010

Baidu.com Inc.
No. 10, Shangdi 10th Street
Haidian District, Beijing, 100085, China
www.baidu.com
+86.10.5992.7396

Clarifai
32 W. 22nd Street
New York, NY 10010, USA
www.clarifai.com

Declara
977 Commercial Street
Palo Alto, CA 94303, USA
www.delcara.com
+1.877.216.0604

Deep Instinct
501 Folsom Street, #400
San Francisco, CA 94105 USA
www.deepinstinct.com
+1.855.522.2223

Digital Genius
7 World Trade Center, 250 Greenwich Street
New York, NY 10007, USA
www.digitalgenius.com
+1.212.266.0090

Facebook
1 Hacker Way
Menlo Park, CA 94025, USA
www.facebook.com
+1.650.543.4800

Google
1600 Amphitheatre Parkway
Mountain View, CA 94043, USA
www.google.com
+1.650.253.0000

H20.ai
1185 Terra Bella Avenue
Mountain View, CA 94043, USA
www.h2o.ai
+1.650.429.8337

IBM
1 New Orchard Road
Armonk, NY 10504, USA
www.ibm.com
+1.914.499.1900

Indico
129 South Street
Boston, MA, 02111, USA
https://indico.io
+1.857.991.1000

Intel
2200 Mission College Boulevard
Santa Clara, CA 95054, USA
www.intel.com
+1.408.764.8080

Iris Automation
2 Shotwell Street
San Francisco, CA, 94103 USA
www.irisonboard.com
+1.640.469.4085

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052, USA
www.microsoft.com
+1.425.882.8080

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Mobileye
Hotzvim, 13 Hartom Street
P.O. Box 45157
Jerusalem 9777513, Israel
www.mobileye.com
+972.2.541.7333

NVIDIA
2701 San Tomas Expressway
Santa Clara, CA 95050, USA
www.nvidia.com
+1.408.486.2000

OpenAI
San Francisco Bay Area, CA, USA
https://openai.com

Ripjar
Eagle Tower, Montpellier Drive
Cheltenham GL50 1TA, United Kingdom
www.ripjar.com
+44.1242.312052

Salesforce.com
The Landmark at One Market, Suite 300
San Francisco, CA, 94105, USA
www.salesforce.com
+1.415.901.7000

Sentient Technologies
1 California Street, #2300
San Francisco, CA 94111, USA
www.sentient.ai
+1.415.422.9886

SkyMind
44 Tehama Street
San Francisco, CA 94105, USA
www.skymind.io
+1.415.696.4031

SparkCognition
4030 W. Braker Lane, #4
Austin, TX 78759, USA
http://sparkcognition.com
+1.844.205.7173

Tencent
Tencent Building, Kejizhongyi Avenue
Hi-tech Park, Nanshan District
Shenzhen, China 518057
www.tencent.com
0755.86013388

TeraDeep
51 E. Campbell Avenue
Campbell, CA 95008, USA
www.teradeep.com
+1.203.645.0736

Uber Technologies Inc.
555 Market Street
San Francisco, CA 94103, USA
www.uber.com
+1.415.986.2715

Vicarious
San Francisco Bay Area, CA, USA
www.vicarious.com

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 10

TABLE OF CONTENTS

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 11

TABLE OF CHARTS AND FIGURES

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

SECTION 12

SCOPE OF STUDY

This report examines the practical application of deep learning within consumer, enterprise, and government markets. Other adjacent AI technologies, including classical machine learning, computer vision, natural language processing, machine reasoning, and strong AI, are not covered, other than to say that deep learning enables machine perception and can be used in conjunction with thse other technologies. One of the challenges of creating forecasts in such a highly innovative field is that a technology not yet invented is likely to play a significant role during the time period (2016 to 2025) covered by this study.

Within that scope, the report discusses the most common use cases for deep learning, spanning 28 different industry segments, and identifies those applications best suited for commercial use. The report also considers the impact that such technologies will have on the demand for services and hardware to support AI-based applications. The accompanying databook forecasts revenue for deep learning applications across geographic region and industry during the period from 2016 through 2025.

SOURCES AND METHODOLOGY

Tractica is an independent market research firm that provides industry participants and stakeholders with an objective, unbiased view of market dynamics and business opportunities within its coverage areas. The firm's industry analysts are dedicated to presenting clear and actionable analysis to support business planning initiatives and go-to-market strategies, utilizing rigorous market research methodologies and without regard for technology hype or special interests including Tractica's own client relationships. Within its market analysis, Tractica strives to offer conclusions and recommendations that reflect the most likely path of industry development, even when those views may be contrarian.

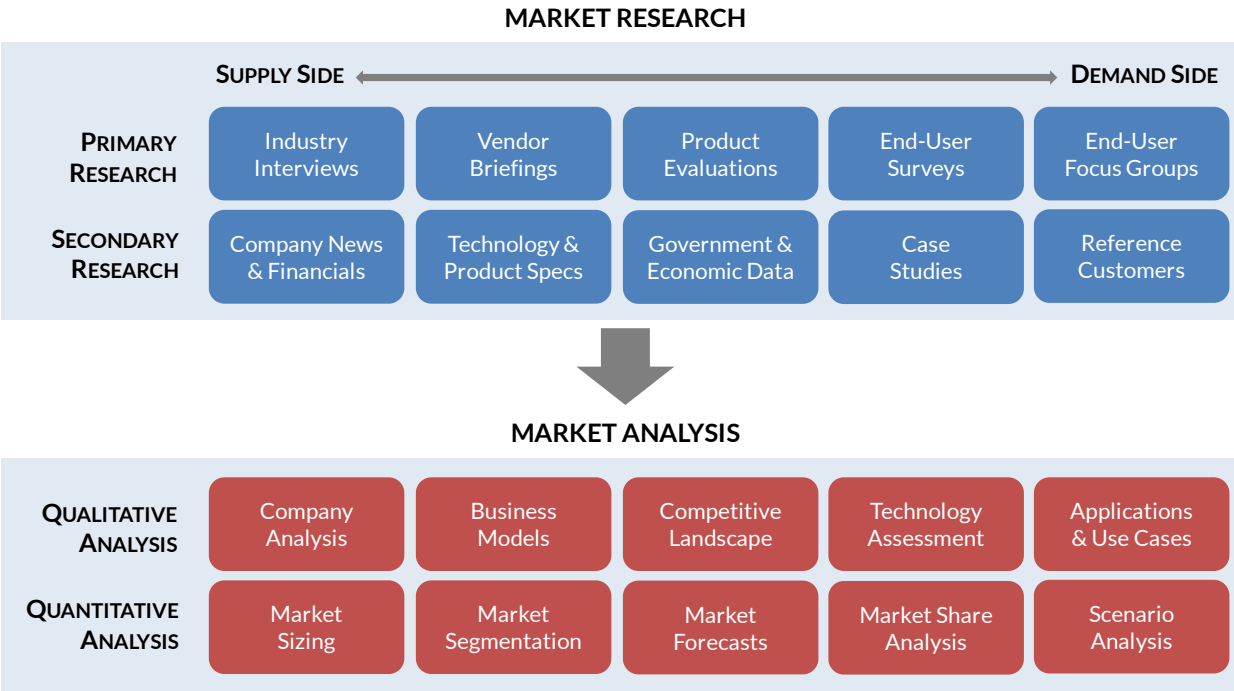
The basis of Tractica's analysis is primary research collected from a variety of sources including industry interviews, vendor briefings, product demonstrations, and quantitative and qualitative market research focused on consumer and business end-users. Industry analysts conduct interviews with representative groups of executives, technology practitioners, sales and marketing professionals, industry association personnel, government representatives, investors, consultants, and other industry stakeholders. Analysts are diligent in pursuing interviews with representatives from every part of the value chain in an effort to gain a comprehensive view of current market activity and future plans. Within the firm's surveys and focus groups, respondent samples are carefully selected to ensure that they provide the most accurate possible view of demand dynamics within consumer and business markets, utilizing balanced and representative samples where appropriate and careful screening and qualification criteria in cases where the research topic requires a more targeted group of respondents.

Tractica's primary research is supplemented by the review and analysis of all secondary information available on the topic being studied, including company news and financial information, technology specifications, product attributes, government and economic data, industry reports and databases from third-party sources, case studies, and reference customers. As applicable, all secondary research sources are appropriately cited within the firm's publications.

All of Tractica's research reports and other publications are carefully reviewed and scrutinized by the firm's senior management team in an effort to ensure that research methodology is sound, all information provided is accurate, analyst assumptions are carefully documented, and conclusions are well-supported by facts. Tractica is highly responsive to feedback from industry participants and, in the event errors in the firm's research are identified and verified, such errors are corrected promptly.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Chart 12.1 Tractica Research Methodology



(Source: Tractica)

NOTES

CAGR refers to compound average annual growth rate, using the formula:

CAGR = (End Year Value ÷ Start Year Value)$^{(1/steps)}$ – 1.

CAGRs presented in the tables are for the entire timeframe in the title. Where data for fewer years are given, the CAGR is for the range presented. Where relevant, CAGRs for shorter timeframes may be given as well.

Figures are based on the best estimates available at the time of calculation. Annual revenues, shipments, and sales are based on end-of-year figures unless otherwise noted. All values are expressed in year 2017 U.S. dollars unless otherwise noted. Percentages may not add up to 100 due to rounding.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC

Published 2Q 2017

© 2017 Tractica LLC
1111 Pearl Street, Suite 201
Boulder, CO 80302
Tel: +1.303.248.3000
Email: info@tractica.com
www.tractica.com

This publication is provided by Tractica LLC ("Tractica"). This publication may be used only as expressly permitted by license from Tractica and may not otherwise be reproduced, recorded, photocopied, distributed, displayed, modified, extracted, accessed or used without the express written permission of Tractica. Notwithstanding the foregoing, Tractica makes no claim to any Government data and other data obtained from public sources found in this publication (whether or not the owners of such data are noted in this publication). If you do not have a license from Tractica covering this publication, please refrain from accessing or using this publication. Please contact Tractica to obtain a license to this publication.

© 2017 Tractica LLC. All Rights Reserved. This publication may be used only as expressly permitted by license from Tractica LLC and may not otherwise by accessed or used, without the express written permission of Tractica LLC