CAICT
中国信通院

2018 World AI Industry Development Blue Book

Gartner

2018 World AI Industry Development Blue Book

Preface

As the driving force of a new round of technology and industrial revolution, artificial intelligence (AI) is wielding a profound impact on the world economy, social progress and people's daily life. Therefore, the 2018 World AI Conference will be held in Shanghai from September 17 to 19 this year, jointly by the National Development and Reform Commission (NDRC), the Ministry of Science and Technology (MST), the Ministry of Industry and Information Technology (MIIT), the Cyberspace Administration of China (CAC), the Chinese Academy of Sciences (CAS), the Chinese Academy of Engineering (CAE) and the Shanghai Municipal People's Government, with the approval of the State Council, in order to accelerate the development of a new generation of AI, complying with the development rules and grasping the future trends. It is organized as a mobilization conference to make people join in planning for AI technologies and industrial developments. It also works as an important measure to promote the deep integration of the Internet, big data and AI with the real economy.

Themed "A NEW ERA EMPOWERED BY ARTIFICIAL INTELLIGENCE", this conference is featuring "Globalization, Sophistication, Specialization and Marketization". It boasts of being the world's top AI platform for cooperation and communication, where get together the most influential scientists, entrepreneurs and investors from the AI industry all around the world, as well as related government leaders and city administrators who will deliver speeches and engage in various high-level dialogues surrounding the leading-edge technologies, industrial trends and hot issues in the current AI field. This meeting is expected to share authoritative opinions and consensus, and exhibit new technologies, products, applications and ideas, in an effort to bring out the "Chinese Solutions" and the "Global Wisdom" in order to address the common challenges facing human beings and build a better life for people.

Committed to taking the lead in China's reform and opening-up and innovative developments, Shanghai has been working closely with related national ministries and commissions to jointly organize the 2018 World AI Conference. Hosting this event becomes a priority in the process of Shanghai constructing the "five centers" - international and economic center, finance center, trade center, shipping center and technical innovation center, and building the "four brands" of Shanghai - services, manufacturing, shopping and culture. Also, it is the practical measure to foster the reform and opening-up and optimize the business environment in Shanghai. This meeting provides the opportunity for Shanghai to accelerate industrial innovation and the extensive application of AI and take efforts to build national "AI highland", looking to become the innovation source of AI, the application demonstration zone, and the industrial and talents clusters.

As an important achievement of the conference, this Blue Book is released by China Academy of Information and Communications Technology (CAICT) and featuring research from Gartner, the world's leading information technology research and advisory company. With the opportunity of this conference, the Blue Book is expected to open a window for Chinese AI enterprises to communicate with global AI giants. And it makes a comprehensive and in-depth analysis of the developments of global AI industry and the trends of technologies routes. Hopefully, the practitioners and researchers who are working on AI and other readers who are interested in AI could get many inspirations from this report and use it as a reference to learn about AI.

The Organizing Committee of WAIC

In-depth Report on the Development of the World AI Industry, 2018

1 Overview and Description

1.1 Report Overview

This Blue Book will share the research results and practical experiences in the field of AI, comprehensively analyzing the current trend of the industrial developments and technologies in the world's AI industrial powers, and elaborating on the technological research, industrial inputs, application services in the AI field, in order to provide basic information and guidance for the developments of AI technology and industry. This Blue Book has been compiled by a group of editors together, by sorting out the AI technologies and the artificial intelligence developments in industry and application, and analyzing the technology fads, enterprise conditions, industrial applications and future trend of artificial intelligence, based on thorough and extensive investigation in the developments of the world AI industries, in accordance with the research data and relevant materials provided by authoritative departments.

This report mainly discusses on the world AI industry map, which is jointly formulated by the Research Institute of Information and Industrialization and the Data Research Center of CAICT. This map has systematically analyzed and drawn the world AI industry from the underlying technology to the vertical application, from the industrial structure to the industrial distribution, from the research institutions to the government policy, reviewing the trend of AI developments, in order to promote the AI industry and provide fundamental decision references for the development of the AI industry. Meanwhile, East China Branch of China Academy of Information and Communications Technology has further expanded the contents of the industrial map by designing the overall framework of this report, and expounding the environments of industrial development and technologies, the developments of the global AI enterprises, as well as the industrial applications in AI fields, which helps to reflect the situations of the current developments in AI industry more clearly.

1.2 Report Description

The world AI industry map includes the following sections: (1) the map of industrial chain (2) the map of industry distribution (3) the industry research institution and supporting policy.

1) The Map of Industrial Chain
The section of AI industrial chain falls into three layers: basic supporting layer, software algorithm layer and industry application layer.

The map of basic layer depicts the typical enterprises engaged in the basic industry in the world. It is drawn in three dimensions: computing hardware (cloud training, cloud inference, device inference, intelligent chips, intelligent sensors) computing system technology (cloud computing, big data, 5G communication and Internet of Things) and data (data acquisition, labeling and analysis).

The map of software algorithm layer describes the typical players working on the software algorithm industry across the world mainly on the level of algorithm theory (machine learning algorithm, brain-like algorithm, and knowledge graph), development platform (basic open source framework, technology open platform) and application technology (computer vision, natural language processing and human computer interaction).

The map of industry application layer pictures the main enterprises working on the application-oriented industry across two dimensions: industrial solutions ("AI+" industrial vertical application) and typical products (visual products, voice terminals, robots, intelligent cars, drones).

2) The Map of Industrial Distribution
This section has mapped out the distribution, scale and name of the main enterprises in the country where the AI industry is leading the world, as well as in various Chinese provinces and cities.

3) Industrial Research Institutions and Policy
This section has mapped out the mainstream AI industrial research institutions and alliances in China and across the world, and the industrial policy documents.

The structure of this report is divided into five parts: 1) industrial development environment 2) technical environment 3) The situations of world AI enterprise 4) The situations of world investment and financial 5) The situations of industrial developments

1) Industrial Development Environment
This part has comprehensively elaborated on the foundation and environment of the AI industrial developments , discussing the development progress of the artificial intelligence, the AI related policies in various countries all around the word, and the conditions of AI developments.

2) Technical Environment
This part has derived the current technical environments in the field of artificial intelligence in terms of patents, papers, research environments and technical competitions all around the world.

3) The Situations of World AI Enterprises
This part describes the situations of World AI enterprises in the aspects of structure, scale and regional distribution.

4) The Situations of World Investments
This section presents the distribution, scale and rounds of world investment and financing in the field of AI, analyzing the developments of developments of AI industry in AI industry, in the perspective of investment and financing, and the scale of the industry.

5) The Situations of Industrial Development
This section expounds the current situation and trend of technologies and applications in the global AI industry, as well as several application cases of model AI enterprises.

The specific research scope and data sources of this report are as follows:

1) AI Enterprise
The information of AI enterprises mentioned in this report came from the Monitoring Platform in the Data Research Center of CAICT, in which the AI enterprises refer to the enterprises that can provide AI products, services and related solutions. Enterprises can be divided into two dimensions: technical level and product or solution level. The technology dimension includes providers and manufacturers of general technologies such as algorithmic platform, basic hardware, speech and vision. The product or solution dimensions include manufacturers and solution providers of various types of AI products, as well as solution providers for vertical industries.

2) Investment and Financing
The data of investment and financing in this report come from the investment and financing websites such as CB insights, IT oranges, and sprouts. It is the result of matching and collating statistics based on the AI business directory.

3) Patent Data
The patent data in this report are derived from the research results of the Intellectual Property Center of CAICT. The Intellectual Property Center conducts search and statistics on AI patents worldwide based on professional databases such as PatSnap.

4) Paper Data
The paper data in this report are derived from the core collection of Web of Science, which is based on the results of the data retrieval center's AI keyword vocabulary.

5) Industrial Application Data
The industrial application data in this report come from Qixin which is the product of CCi Intelligence Co., Ltd. Besides, the data are also from the related forecasts of major websites and research institutions such as CAICT,PwC, MarketsandMarkets, Grand View Research, IFR, Roland Berger, ASKCI Consulting, and Forward Industry Research Institute.

6) Data Unit Description
The default currency for all market data in this report is RMB, if not specified.

This Blue Book does not seek to cover all aspects; it only gives an analysis and description over the current industrial and technological development environment, world enterprise situation and industrial application in the field of AI. It does not give much viewpoint statements on the AI field, trying to describe in a language and manner that is easy to understand. For the main contents of this Blue Book, experts and scholars from various fields are welcome to make suggestions. We are glad to listen to the opinions of experts from all sides and continue to make improvements.

2 Industrial Development Environment

2.1 AI Development Milestones

The history of AI can be roughly divided into three stages: the first stage (1956-1980) AI was born; the second stage (1980-2000) AI stepped into industrialization; the third stage (2000-present) AI ushers in the explosion.

The first stage (1956-1980) AI was born

Date	Iconic event
1956	The Dartmouth Conference in the US gathered the first batch of researchers to determine the name and mission of AI, which was called the birth of AI.
1957	Frank Rosenblatt, an experimental psychologist at Cornell University, implemented a neural network "perceptron".
1969	The International Federation of Artificial Intelligence was established and the first meeting was held in Seattle, Washington, US

The second stage (1980-2000) AI stepped into industrialization

Date	Iconic event
1980	Carnegie Mellon University designed an expert system called XCON for DEC, which was a huge success, and at that time it saved the enterprise USD 40 million each year.
1982	Japan planned to invest USD 850 million to develop AI computers (the fifth-generation computers), aiming to create machines that can talk to people, translate languages, interpret images, and reason like humans.
1986	Multi-layer neural networks and BP back-propagation algorithms have emerged to improve the accuracy of automatic recognition.
1988	The German Research Centre for Artificial Intelligence was established and is currently the world's largest non-profit AI research institution.
1997	Deep Blue, a chess-playing computer developed by IBM, defeated the world chess champion, a milestone event in the history of AI; under the influence of Moore's Law, computing performance began to increase dramatically.

The third stage (2000-present) AI ushers in the explosion

Date	Iconic event
2006	Geoffrey Hinton proposed a training algorithm in "Science" based on Deep Belief Networks (DBN) that can use unsupervised learning, making deep learning continue to heat up in academia.
2011	The IBM Watson system won at the US game show Jeopardy! against human players.
2012	The deep learning algorithm became well-known after the ImageNet Challenge, and was thereby widely used.
2016	AlphaGo developed by DeepMind defeated former World Go champion Lee Sedol.

2.2 AI Policies of Major Countries

The rapid development of AI will profoundly change the face of human society and the world. In order to seize the strategic opportunities of AI development, more and more countries and organizations have rushed to set up the national-level development plans.

US

Date	Organization	Policy released
1998	The Networking and Information Technology Research and Development Program (NITRD)	Next Generation Internet Research Act (P.L. 105-305)
2013	The White House	National Robot Project: A Roadmap for U.S. Robotics from Internet to Robotics 2013 Edition
2013.4	The White House	Brain Research through Advancing Innovative Neurotechnology
2015.10	National Economic Council (NEC)/Office of Science and Technology Policy (OSTP)	The new edition of the US National Innovation Strategy
2015.11	Center for Strategic and International Studies (CSIS)	Defense 2045
2016.10	National Science and Technology Council (NSTC) / Networking and Information Technology Research and Development Subcommittee (NITRD)	National Artificial Intelligence Research and Development Strategic Plan
2016.10	National Science and Technology Council (NSTC)	Preparing for the Future of Artificial Intelligence
2017.9	US Congress	SELF DRIVE ACT AV START ACT
2017.10	US Information Industry Council	AI Policy Principles

China

Date	Organization	Policy released
2015.7	State Council	Guiding Opinions on Actively Promoting the "Internet +" Action Plan
2016.3	State Council	Outline of the Thirteenth Five-Year Plan for National Economic and Social Development
2016.4	Ministry of Industry and Information Technology (MIIT)/ National Development and Reform Commission (NDRC)/ Ministry of Finance (MOF)	Robot Industry Development Plan (2016-2020)
2016.5	CPC Central Committee and the State Council	National Innovation Driven Development Strategy Outline
2016.5	National Development and Reform Commission (NDRC)/Ministry of Science and Technology (MST)/ Ministry of Industry and Information Technology (MIIT)/Cyberspace Administration of China (CAC)	The Three-year Implementation Program for "Internet +" Artificial Intelligence
2016.7	State Council	13th Five-year Plan on National Scientific and Technological Innovation
2017.3	State Council	Report on the Work of the Government
2017.7	State Council	Development Planning for a New Generation of Artificial Intelligence
2017.12	Ministry of Industry and Information Technology (MIIT)	Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020)
2018.4	Ministry of Education (MOE)	Innovative Action Plan for Artificial Intelligence in Colleges and Universities

Other countries and international organizations

Date	Organization	Policy released
2013.6	Japanese Cabinet	Japan Revitalization Strategy
2015.1	Ministry of Economy, Trade and Industry	New Strategy of Japan for Robots
2016.5	Japanese Cabinet	Science and Technology Innovation Comprehensive Strategy 2016
2016.6	Headquarters for Japan's Economic Revitalization	Japan Rejuvenation Strategy
2017	Japanese Government	Promotion Strategy for Next Generation Artificial Intelligence
2017.5	Ministry of Economy, Trade and Industry	Vision of New Industrial Structure
2017.6	Japanese Cabinet	Comprehensive Strategy on Science, Technology and Innovation for 2017
2018.6	Japanese Cabinet	Strategy for Comprehensive Innovation
2018.6	Japanese Cabinet	Future Investment Strategy 2018
2013	Korea Electronics and Telecommunications Research Institute	Exobrain Plan

Date	Organization	Policy released
2014.7	Ministry of Trade, Industry and Energy, Republic of Korea	The second Master Plan to develop intelligent robots (2014-2018)
2016.8	Korean Government	Nine National Strategic Projects
2017.7	Korean Parliament	Robotic Basic Act
2018.5	The Fourth Industrial Revolution Committee	Artificial Intelligence Research and Development (R&D) Strategy
2013	The UK Government	The Eight Great Technology
2016.12	The UK Government Office for Science	Artificial intelligence: opportunities and implications for the future of decision making
2017.1	The UK Government	building our industrial strategy
2017.10	The UK Government	Growing the Artificial Intelligence Industry in the UK
2010.7	German Government	Ideas. Innovation. Prosperity. High-Tech Strategy 2020 for Germany
2011.11	German Government	"Industry 4.0" as the Strategic Focus
2013.4	German Ministry of Education / Research "Industry 4.0 Working Group"	Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0-- Securing the Future of German Manufacturing Industry
2017.6	German Federal Ministry of Transport and Digital Infrastructure	Automated and Connected Driving Report
2013	French Government	French Robot Development Plan
2017.3	French Ministry for the Economy and Finance/ Ministry of National Education	Artificial Intelligence Strategy
2018.5	French Government	Make Sense to Artificial Intelligence
2017.5	Singapore National Research Foundation	"AI.SG" National Artificial Intelligence Program
2013.1	European Union (EU)	Human Brain Project
2013.12	European Commission / European Robotics Association	SPARC Program
2015.12	EU SPARC	Robotics 2020 Multi-Annual Roadmap
2016.6	European Commission	Proposed AI legislative motion
2016.10	European Parliament's Committee on Legal Affairs (JURI)	European Civil Law Rules in Robotics
2017.10	European Union (EU)	Horizon 2020
2016.8	World Commission on the Ethics of Scientific Knowledge and Technology (COMEST)	Draft Preliminary Report on Robot Ethics
2016.12	Institute of Electrical and Electronics Engineers (IEEE)	Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (1st Edition)
2017.12	Institute of Electrical and Electronics Engineers (IEEE)	Ethical Guidelines for Artificial Intelligence Design (2nd Edition)

2.3 Development Conditions

2.3.1 Evolution of the Algorithms

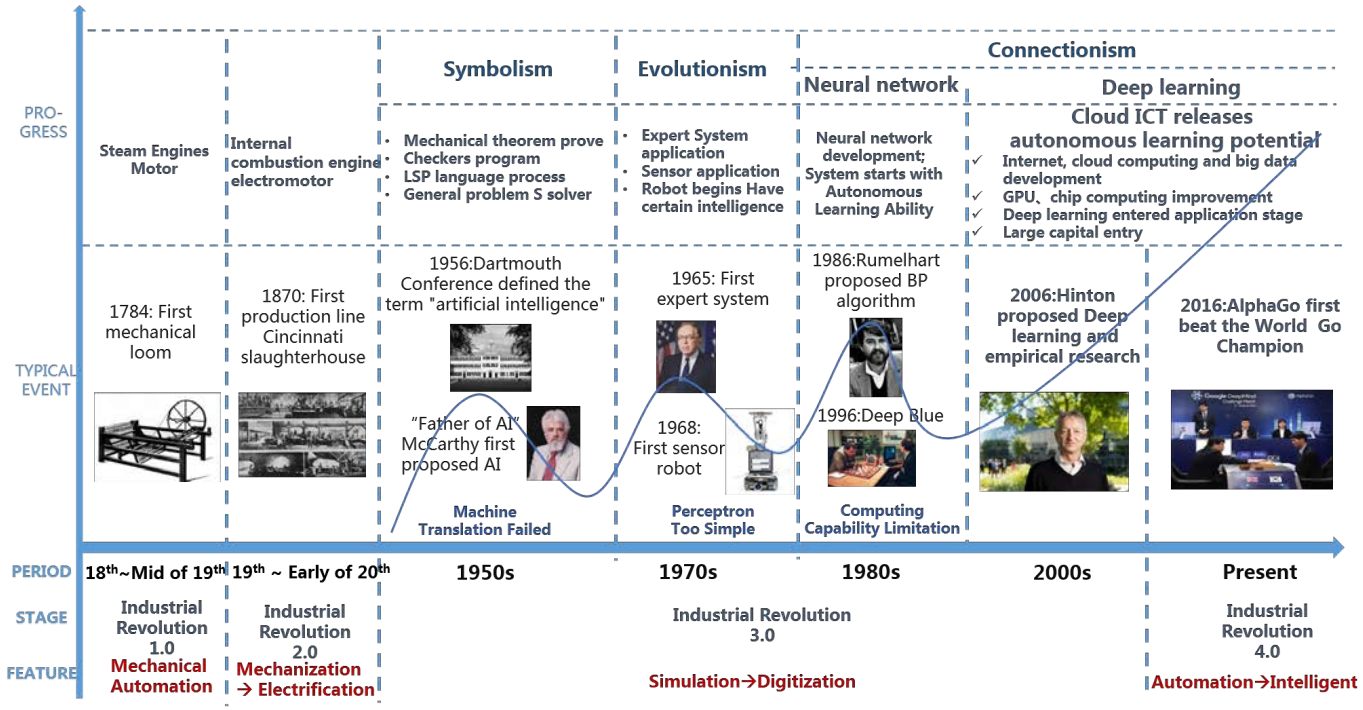
The development of AI algorithms has continued to innovate and the learning level has been increasing. The early research in the academic field focused on symbolic computing. The artificial neural network was completely negated in the early stage of AI development, then gradually recognized, and became a large-scale algorithm that leads the development trend of AI today, showing strong vitality. The current popular algorithms of machine learning and deep learning are actually further extensions of the theory of Symbolism, Evolutionism, and Connectionism.

Machine learning algorithms and deep learning algorithms are two hot points in AI. The open source framework has become the focus of the overall development of

Table 2-1 Deep Learning Framework

Framework	Organization	Language(s) supported	Introduction
TensorFlow	Google	Python/C++/Go/…	Open-source neural network library
Caffe	UC Berkeley	C++/Python	Open-source convolutional neural network framework
PaddlePaddle	Baidu	Python/C++	Open-source deep learning platform
CNTK	Microsoft	C++	Deep learning computing network toolkit
Torch	Facebook	Lua	Machine learning algorithm open source framework
Keras	Google	Python	Modular neural network library API
Theano	University of Montreal	Python	Deep learning library
DL4J	Skymind	Java/Scala	Distributed deep learning library
MXNet	DMLC community	C++/Python/R/…	Open-source deep learning library

Fig. 2-1 AI Setting off a New Wave of Technological Development



Source: CAICT (2018)

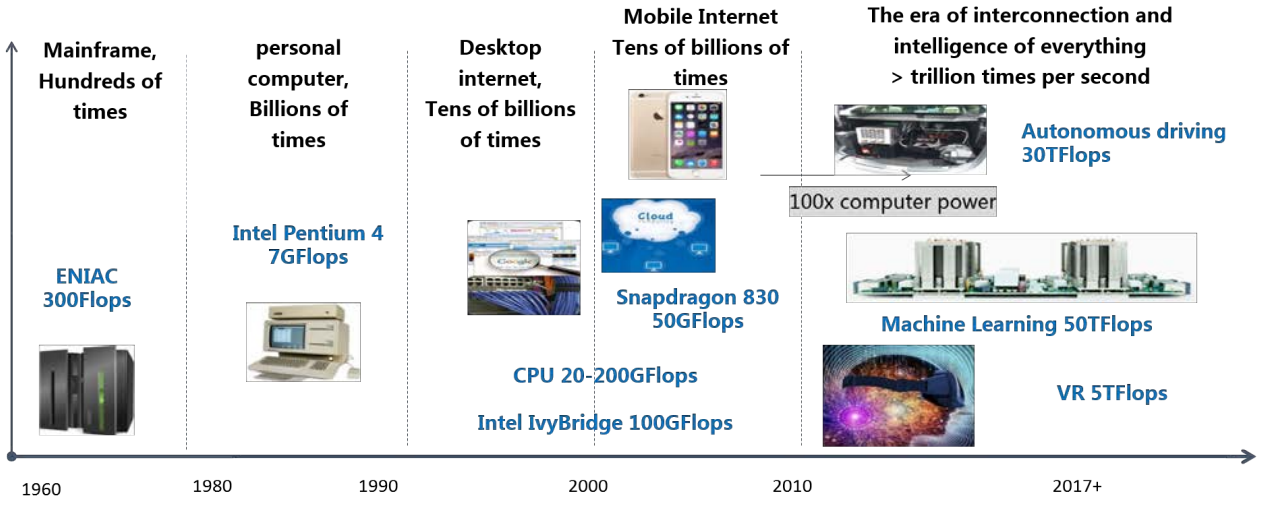
technology giants. The open source deep learning platform is an important driving force for the development of AI technology. The open source deep learning platform allows the public to use, copy and modify the source code. It has the characteristics of fast update and strong expansion, which can greatly reduce the development cost of enterprises and the purchase cost of customers. These platforms are widely used by enterprises to quickly build a deep learning technology development environment and promote the accelerated iteration and maturity of their own technology, so as to ultimately achieve the application of the products.

AI is still evolving rapidly and changing people's lives, and more AI algorithms are waiting for computer scientists to explore. Due to the long technology investment cycle, most AI enterprises in China still lack original algorithms, and they still need to plan ahead and pay attention to the talent reserve of AI algorithm level; combine academic research and industrial application scenarios, encourage innovation, actively explore talents in AI algorithm, and let AI researchers with strong potentials truly enter the industry.

2.3.2 Increase in Computing Power

The implementation of AI algorithms require strong computing power support, especially the large-scale use of deep learning algorithms, which put higher demands on computing power. AI has ushered in a real explosion since 2015, which is largely related to the widespread use of GPUs. Prior to this, hardware computing power could not meet the needs of AI computing power. When GPU and AI were combined, AI ushered in a real high-speed development. Therefore, the improvement of hardware computing power is one of the important factors for the rapid development of AI.

Fig. 2-2 Development of AI Computing Power



The evolution of computing power in general-purpose computers FLOPS: From Two digits -> G -> T ->P

Source: CAICT (2018)

Fig. 2-3 AI High Performance Computing Unit

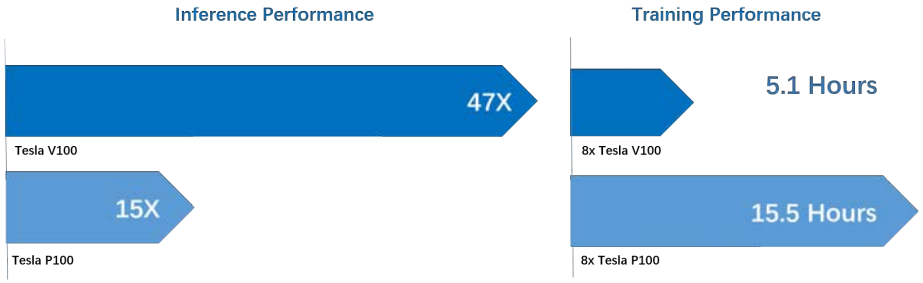
Source: CAICT, Internet

In recent years, the new high-performance computing architecture has become the catalyst for the evolution of AI technology. With the emergence of the deep learning boom in the AI field, the architecture of computing chips have gradually evolved toward the trend of deep learning application optimization, from the traditional Intel processor that takes CPU as the main and GPU as the auxiliary to a structure with GPU as the main and CPU as the auxiliary. In 2017, NVIDIA launched Tesla V100, the new generation of graphics processing chip, which is mainly used to study AI that is based on deep learning. For TensorFlow, the Google open source deep learning framework, Google launched a customized TPU for machine learning.

The development of AI needs urgently core hardware upgrades and computing innovation becomes the focus of deployment. The existing chip products cannot meet the high-throughput requirements of dense linear algebra and massive data in terms of basic capabilities. It is urgent to have high performance and versatility of the cloud, as well as high energy efficiency and low latency of the terminal.

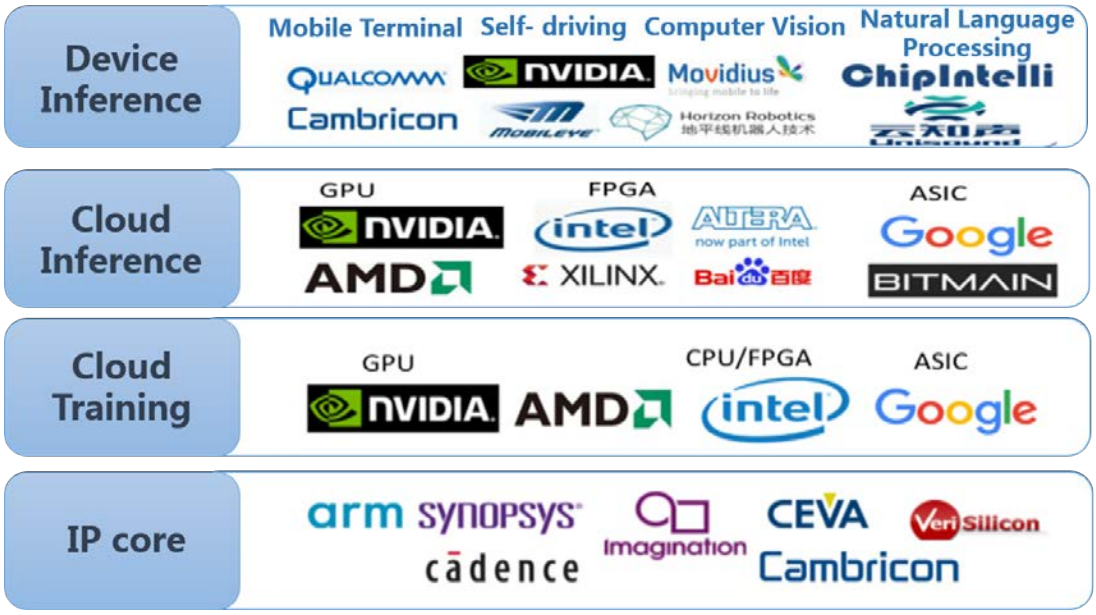
From the development stage of AI chips, general-purpose chips such as CPU, GPU and FPGA are the main chips in the AI field, and the dedicated chip ASIC for neural network algorithms is being launched one after another by Intel, Google, NVIDIA and many startups and is expected to replace the current general-purpose chip and serve as the main force of AI chips in the next few years.

Fig. 2-4 Tesla V100 Training and Reasoning Performance Comparison



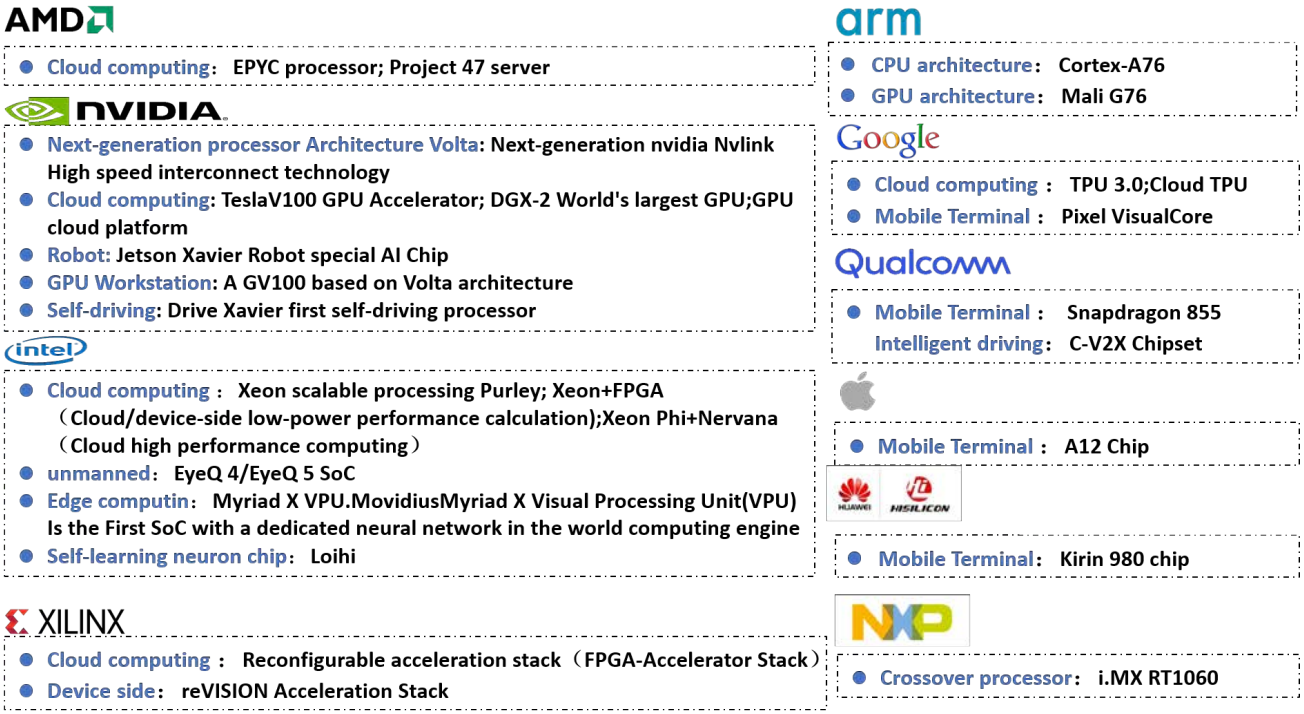
Source: NVIDIA official website

Fig. 2-5 AI Chip Industry Map



Source: CAICT (2018)

Fig. 2-6 Heating up Battle for AI Chips



Source: CAICT (2018)

2.3.3 Data Support

From the software age to the Internet, to today's big data era, the volume and complexity of data have undergone a quantitative to qualitative change, and big data has led the development of AI into an important strategic window.

Data is the cornerstone for the development of AI, and the core of AI lies in data support. From the status quo of development, AI technology has made rapid progress thanks to a good big data foundation, and massive data provides raw materials for training AI. According to the 2018 Q3 global digital statistics report of We Are Social, the number of Internet users worldwide has exceeded 4.1 billion, and the penetration rate of global unique mobile users has reached 67% of the total population.

Big data is a booster for the development of AI, because some AI technologies use statistical models to calculate the probability of data, such as images, text or voice, by exposing these models to the ocean of data, so that they are constantly optimized, or "trained". With the support of big data, the output of the deep learning algorithm will be more accurate as the amount of data processing increases.

Public data sets built by academic and research institutions are keeping enriched and promote the growth of start-ups. Public data sets are generally used in algorithmic testing and competency competitions, with high quality, providing quality data for innovation, entrepreneurship and industry competition, and bringing essential resources to start-ups.

The industry data set is the core competitiveness of the enterprises. The industry data set is closely integrated with the industry. The self-built data sets of each enterprise belong to its core competitiveness. The data service sector has developed rapidly, including data set construction, data cleaning, and data labeling.

Table 2-2 Partial Global AI Public Data Set

Type	Data set name	Characteristics
Natural Language Processing	WikiText	Wikipedia corpus
	SQuAD	The Stanford question answering dataset
	Common Crawl	PB level web crawler data
	Billion Words	Common language modeling database
Speech Recognition	VoxForge	Accented corpus
	TIMIT	Acoustic-Phonetic continuous speech corpus
	CHIME	Speech recognition data set containing ambient noise
Machine vision	SVHN	Image dataset in Google street view
	ImageNet	Commonly used image dataset based on wordnet composition
	Labeled Faces in the Wild	Face area image data set for face recognition training

Fig. 2-7 Industry database classification

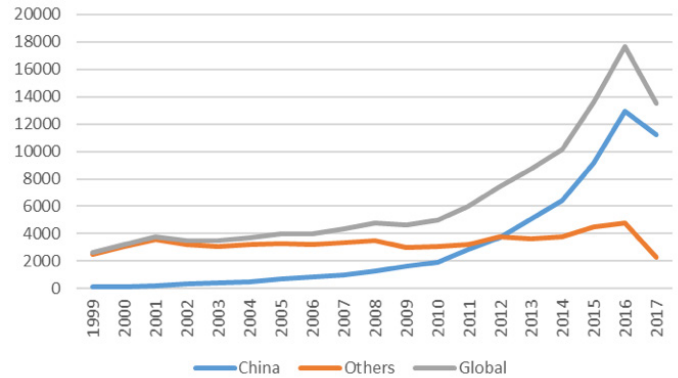
Source: CAICT (2018)

3 Technical Environment

3.1 Patent

From 1999 to 2017, the number of invention applications and authorized patents in the AI field, such as image recognition, biometrics, speech recognition, speech synthesis, natural language understanding and machine learning, had exceeded 100,000. In China, the number of patent applications and authorizations for AI has increased year by year since 2010, and has achieved rapid growth since 2014.

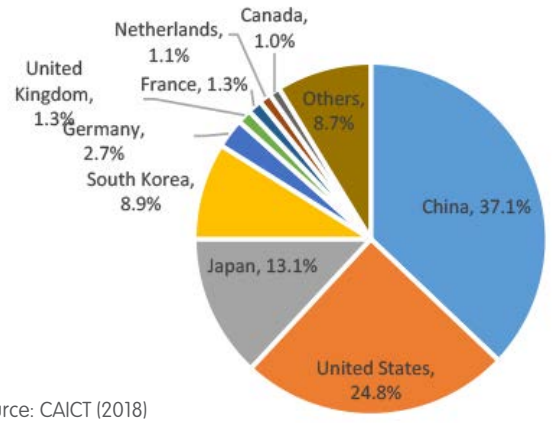
Fig. 3-1 Global AI Related Patent Application Trend



Source: CAICT (2018)

The number of patent applications in China, the US and Japan is ahead in the world, and China has overtaken the US with the highest patent applications in the field of artificial intelligence. The total number of patent applications of China, the US and Japan accounted for 75% of the global AI patents.

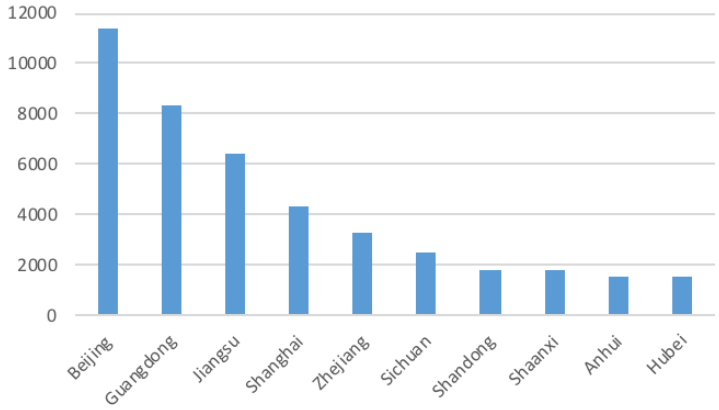
Fig. 3-2 Geographical Distribution of Global AI Patent Applications



Source: CAICT (2018)

China's AI patent applications mainly come from five provinces and cities, namely, Beijing, Guangdong, Jiangsu, Shanghai and Zhejiang. Among them, the number of AI patents applied and authorized by Beijing exceeds 10,000.

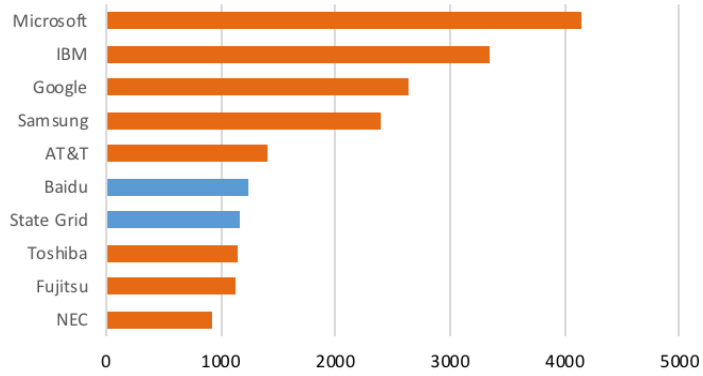
Fig. 3-3 Geographical Distribution of AI Patent Applications in China



Source: CAICT (2018)

From the perspective of the patentee, the US, Japan, Korea's technology giants in the field of artificial intelligence patent accumulation has a leading edge, Microsoft patent applications in the world's first, followed by IBM and Google.

Fig. 3-4 Main Applicants for Global AI Related Patents



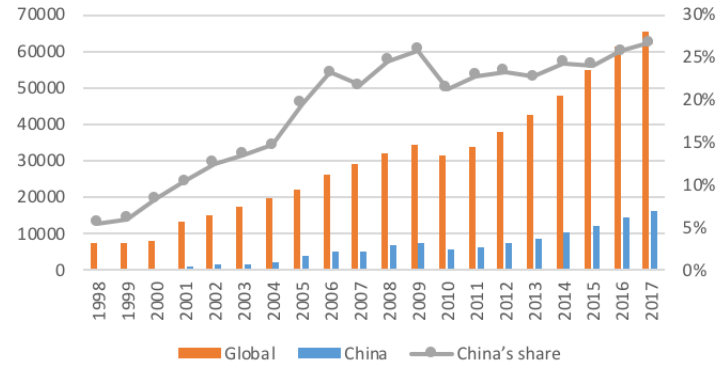
Source: CAICT (2018)

3.2 Paper

Between 1998 and 2018, the number of papers and journals in the AI field increased significantly in the world, with the total number exceeding 630,000 and a CAGR of 11.59%. In 2017, the core collection of Web of Science had a total of 65,100 AI papers and academic publications such as academic publications in journals. Among them, AI papers and academic publications such as academic publications in journals of China (including Hong Kong, Macao and Taiwan) reached 17,300, with a CAGR of 24.32%. The global ratio of artificial intelligence in China increased from 5.52% in 1998 to 26.63% in 2017, indicating that China's strength in the field of AI research has increased dramatically.

In terms of global growth trend, 1998-2017 was largely sustained, with an average annual growth rate of more than 10% per cent in both 2001-2007 and 2012-2016, and a decline in 2008-2011 and a decline in 2010.

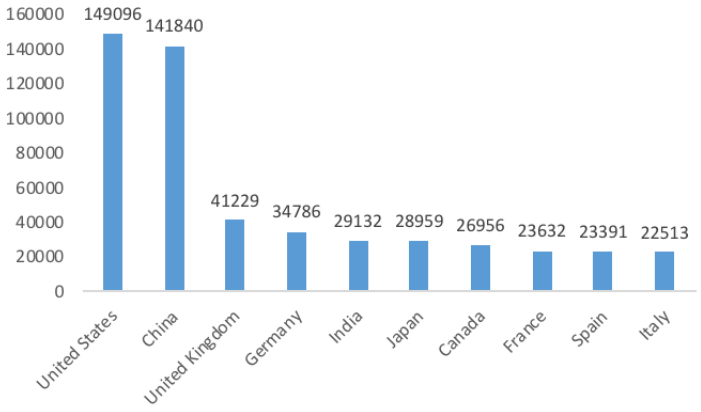
Fig. 3-5 Global/China AI Paper Publication Trend



Source: CAICT (2018)

From 1998 to 2018, the largest number of papers in the AI field in the world was 149,100 in the US, and China ranked second in terms of 141,800. The UK, Germany, and India ranked three to five.

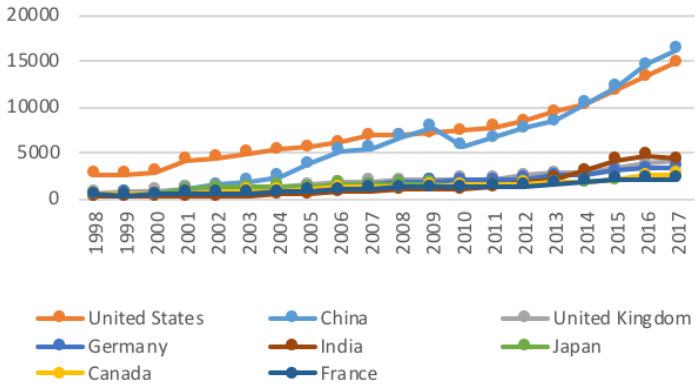
Fig. 3-6 TOP 10 Countries for AI Paper Output



Source: CAICT (2018)

The US has always been a leader in the world of artificial intelligence research, with a higher number of academic studies than in other countries. However, the number of academic publications on AI in China has grown rapidly, and in 2009 and 2014 years later, China surpassed the US in the first place. In addition, India has developed rapidly in the AI field research since 2013. In 2014, its number of AI papers published exceeded that of the UK, ranking third in the world.

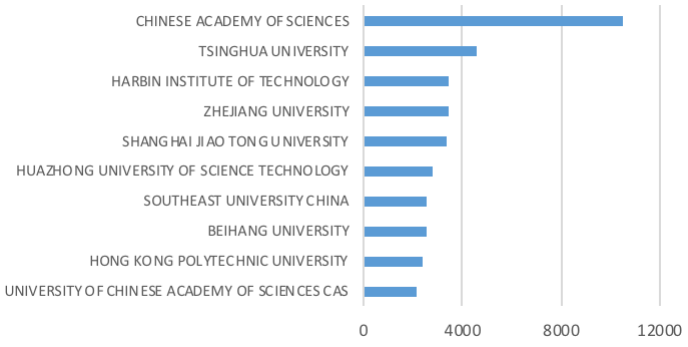
Fig. 3-7 Trend of AI Paper Output in Major Countries



Source: CAICT (2018)

In China, the most productive institutions for AI papers are the Chinese Academy of Sciences, Tsinghua University and Harbin Institute of Technology ranking the second and third. In terms of number, with more than 10,000 papers, the Chinese Academy of Sciences is far ahead of the second-ranked Tsinghua University (more than 4,500) and other institutions.

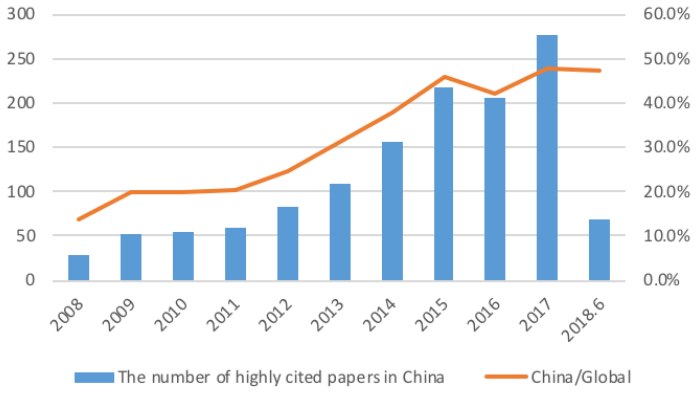
Fig. 3-8 Trend of AI Paper Output in China



Source: CAICT (2018)

Among the highly cited literature included in Web of Science, the number and proportion of China's AI related literature has increased rapidly since 2012, from less than 15% in 2008 to 47% in 2017, indicating that the quality of scientific research in the AI field in China has improved considerably.

Fig. 3-9 Trend of Highly Cited AI Paper Output in China

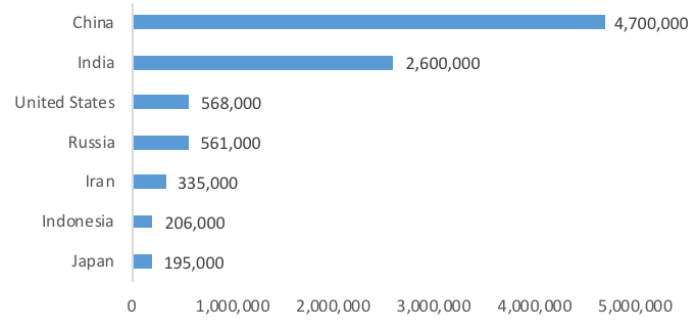


Source: CAICT (2018)

3.3 Research Environment

Global graduates of science, technology, engineering and Mathematics (STEM) are growing every year, and China ranks first in the world. According to the World Economic Forum, 4.7 million Chinese graduates were from the STEM field in 2016. In addition, China has a total of 30,000 STEM PhD graduates each year. After China, the second-ranked country is India. In 2016, India had 2.6 million STEM graduates, including local and overseas graduates. India is second only to China in the world not only in the total number of STEM graduates, but also in the number of students studying abroad, with 26% of these students majoring in computer science and mathematics.

Fig. 3-10 Number of STEM Graduates in the Countries (2016)



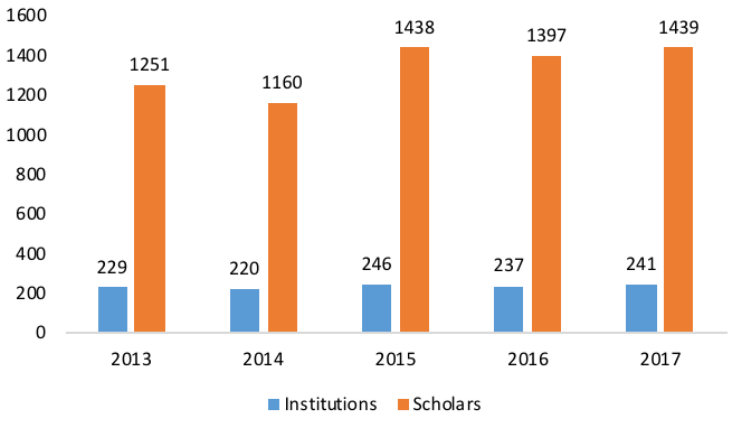
Source: The World Economic Forum, CAICT

Scientific research institutions represented by universities have strong scientific research strength in the AI field, and stand at the forefront of the development of AI theory and technology. Scientific research institutions with a certain scale of scientific research team, with more advanced laboratories and research facilities, and other institutions such as enterprises have a unique comparative advantage. Scientific research institutions can also continuously cultivate high-end artificial intelligence talents for the sustainable development of the industry to provide back-up forces.

The Digital Bibliography & Library Project (DBLP), jointly maintained by the University of Trier and the Schloss Dagstuhl – Leibniz Information Center, contains a large number of journals and literature in the computer field and provides literature search service based on metadata from the scientific literature in the computer field. In the past five years, DBLP has included 300,000-400,000 articles per year. As of now, there are more than 4.2 million DBLP index documents. Based on this, statistics can be made for the research institutions and scholars who have published literature on the AI field collected by DBLP every year, so as to understand the degree of attention paid by the academic community to the AI field.

According to the literature data of the DBLP index about the number of academic institutions and scholars who have literature published in the AI field in 2013-2017, in many countries, the attention paid to AI improved significantly in 2015; after cooling down in 2016, it reached a high point in 2017. See the figure below for the number of research institutions and scholars who have published papers in the DBLP index literature in recent years.

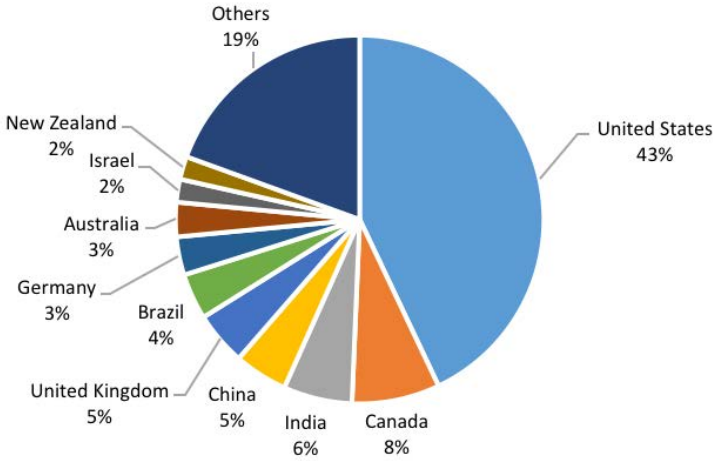
Fig. 3-11 Attention to AI Paid by the Scientific Research Community



Source: CS Rankings, CAICT

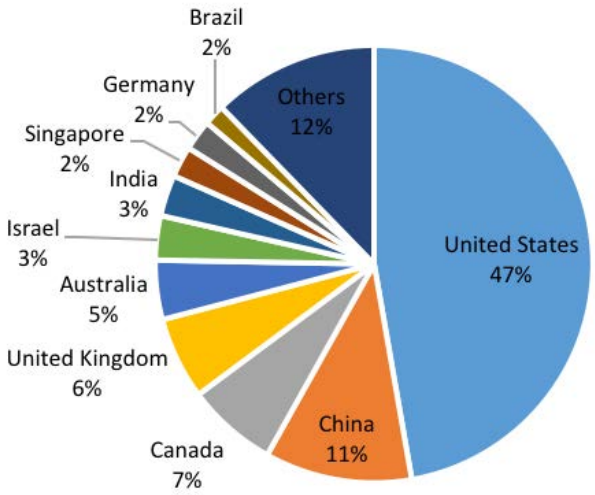
The US ranks first in AI technology in the world, and its leading advantage is obvious. The number of both the AI research institutions and scholars account for nearly half of the world. The number of Chinese scientific research institutions is comparable to that of the UK and India. Its number of scholars is less than that of the US, but is significantly higher than that of other countries and has shown an overall upward trend in recent years, with 2017 increasing by about 30% over 2013.

Fig. 3-12 TOP10 Countries for Number of Scientific Research Institutions with Publications



Source: CS Rankings, CAICT

Fig. 3-13 TOP10 Countries for Number of Scholars with Publications



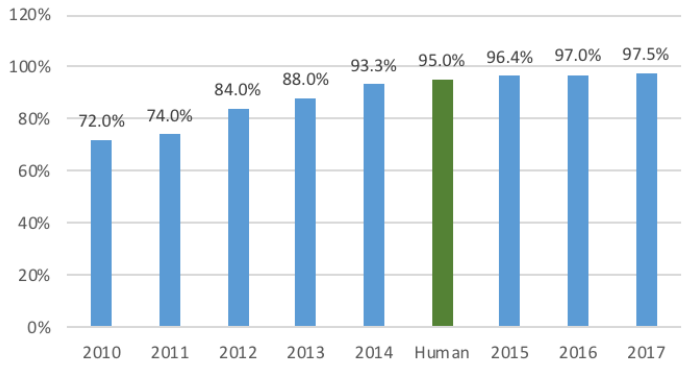
Source: CS Rankings, CAICT

3.4 Technical Competition

The technological development of AI is not only reflected in scientific research works such as patent papers, but also in various technology challenges held around the world. According to the competition data released on Kaggle, there are currently 19 competitions in progress, and 275 competitions have been completed. The content of the competition involves image recognition, speech recognition, object detection, classification, as well as predictive problems in a variety of scenarios. The number of participating teams in different competitions is also different. Some popular competition teams can reach tens of thousands, and the prize pool can reach millions of US dollars.

ImageNet's annual Large-Scale Visual Identity Challenge (ILSVRC) is an early and influential event in the field of computer vision. It is held once a year since 2010 and the competition process will classify and detect objects and scenes. In terms of the accuracy of classification of objects, the human level is 95%. Since 2015, the best AI system has outperformed the humans. In the 2017 ILSVRC competition, the classification accuracy of the AI system has reached 97.5%.

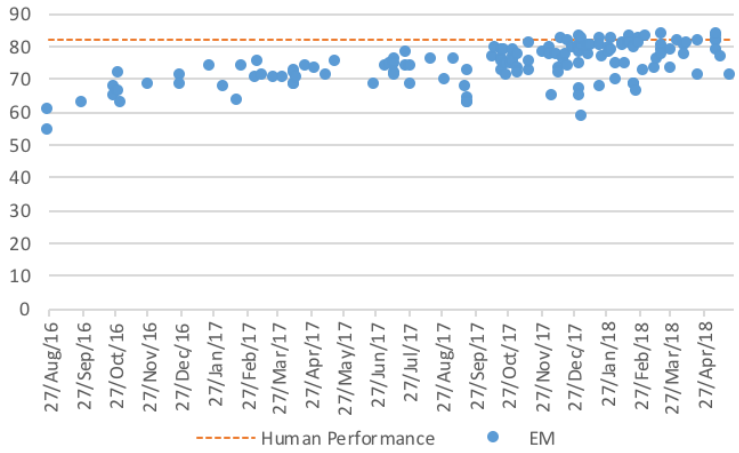
Fig. 3-14 ILSVRC Best Classification Accuracy



Source: CAICT (2018)

The more influential competition in the field of NLP is the system quiz contest based on the Stanford Question Answering Dataset (SQuAD). SQuAD is a data set of reading comprehension, which is composed by crowdsourcing workers who ask questions from many Wikipedia articles. The answer to each question is a combination of text and cross-paragraph content in the corresponding reading paragraph, or the problem itself has no solution. Since the release of SQuAD 1.0, the community has made great strides, and the optimal models have been comparable to human performance. The following is the ExactMatch (EM) of the optimal model evaluated on the v1.1 test set.

Fig. 3-15 Evolution Trend of SQuAD Optimal Model



Source: CAICT (2018)

In recent years, Chinese enterprises have also begun to pay attention to and actively organize and participate in AI challenge. At the CVPR conference this year, Baidu Apollo and the University of California at Berkeley jointly organized an autopilot seminar and defined a number of challenging tasks based on the ApolloScape large-scale data set, in which Megvii Technology, the unicorn enterprise in the visual field, defeated DeepMind and won the first place. In addition, in the video behavior recognition challenge, Chinese enterprises also performed well, taking the top three. In addition to the CVPR Challenge, in a growing number of top international challenges, Chinese enterprises and teams are doing better and better, repeatedly winning the title, indicating that China's AI technology is moving to the forefront of the world.

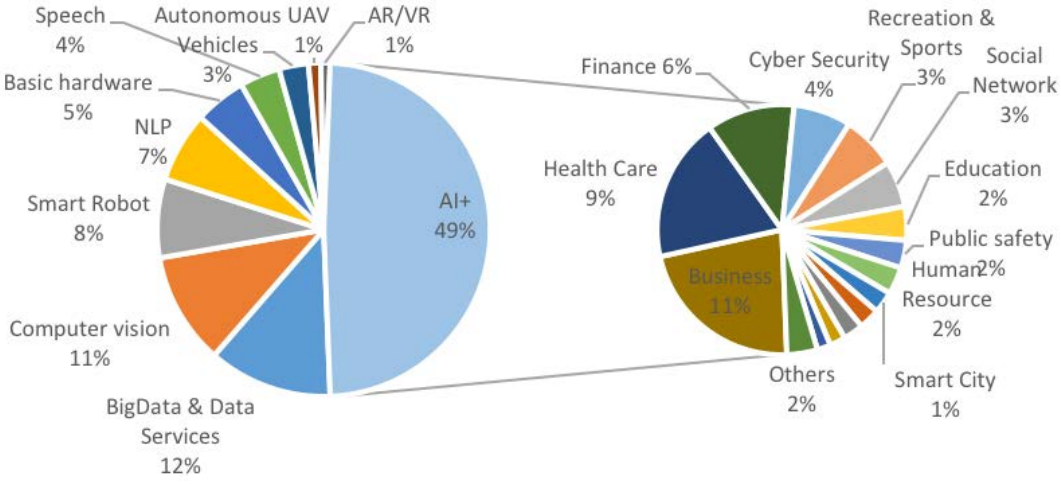
4 World AI Enterprises

4.1 Enterprises Structure

Globally, AI enterprises focus on AI+ (all vertical areas), Big Data & Data Services, vision and intelligent robotics. Among them, AI+ enterprises are mainly concentrated in business (mainly in the field of marketing and customer management), healthcare, and finance.

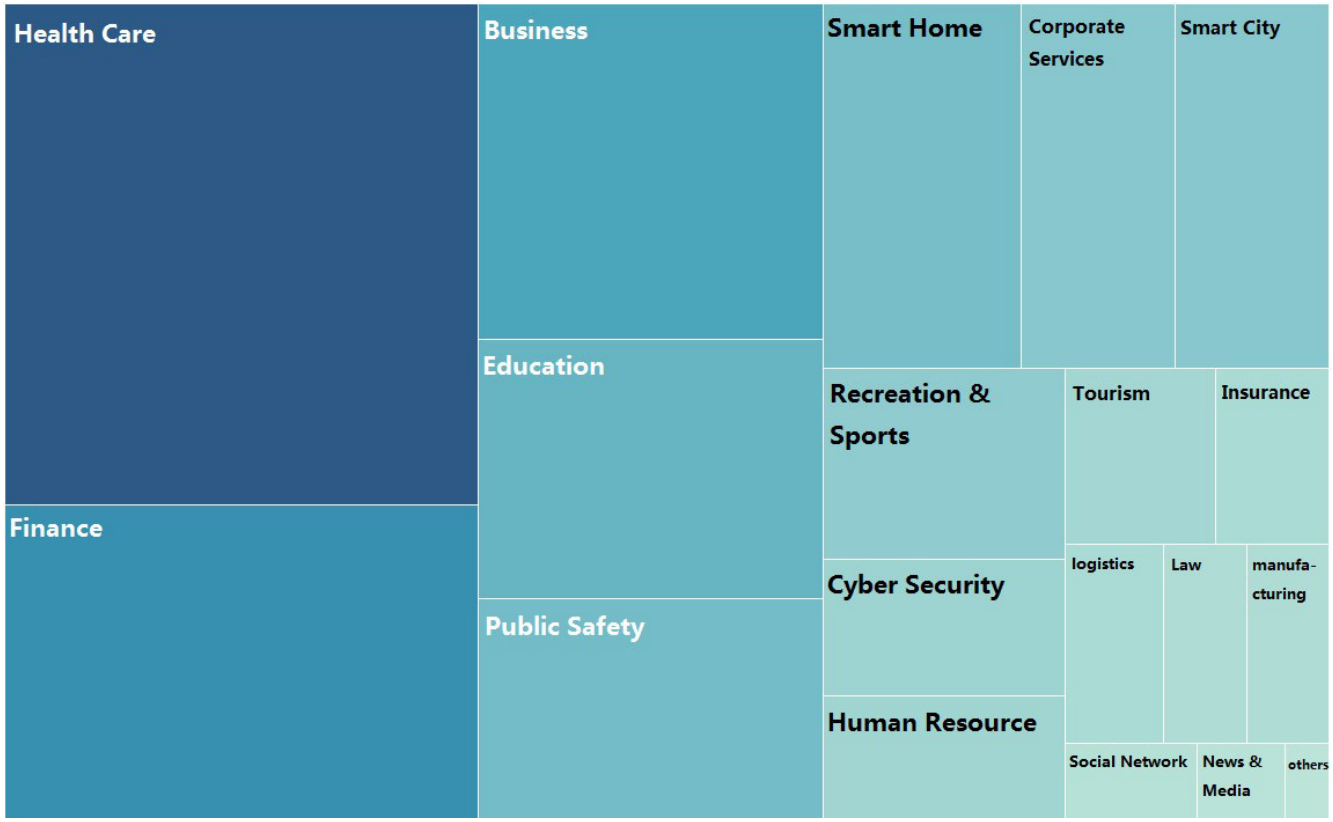
For China, the vertical areas of AI enterprises are also concentrated. Among all kinds of vertical industries, AI penetrates more in the fields of healthcare, finance, commerce, education and security. Among them, the healthcare field accounts for the largest proportion of 22%, followed by finance and business intelligence, accounting for 14% and 11%, respectively.

Fig. 4-1 Global AI Enterprise Structure



Source: CAICT (2018)

Fig. 4-2 China AI+ Enterprise Structure

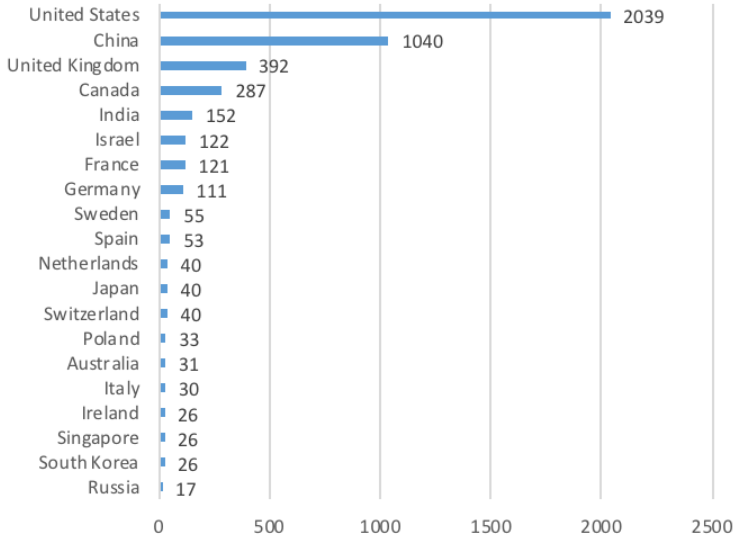


Source: CAICT (2018)

4.2 Business Scale

According to the real-time monitoring of the Data Research Center of the global ICT monitoring platform of CAICT, 4,998 AI enterprises were monitored worldwide as of the first half of 2018. Among them, the number of AI enterprises in the US was 2,039, ranking first in the world, followed by 1,040 in China (excluding Hong Kong, Macao and Taiwan), 392 in the UK, 287 in Canada and 152 in India. In addition, the number of artificial intelligence enterprises in Israel, France and Germany has also surpassed 100.

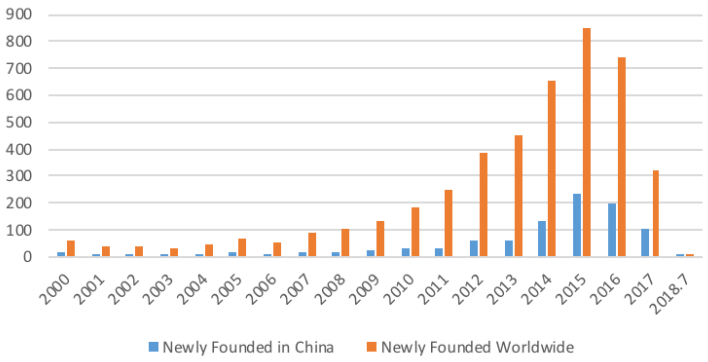
Fig. 4-3 Distribution of Global AI Enterprises



Source: CAICT (2018)

From the perspective of start-up time of enterprises, the global trend of AI entrepreneurship was concentrated in 2014 to 2016, with 2015 seeing the largest number of new AI enterprises in 2015, both globally and in China. During 2015, the number of newly established AI enterprises in the world reached 847, including 238 in China. Since 2016, the number of new startups in the world has begun to decrease, and the pace of entrepreneurship has slowed down. There were 738 new start-ups in the world, and by 2017 this number had dropped to 324.

Fig. 4-4 Time of Establishment of AI Enterprises

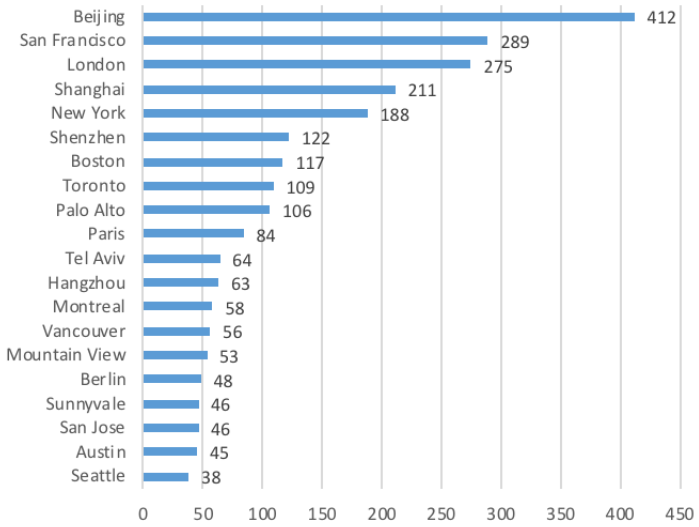


Source: CAICT (2018)

4.3 Business Area

Among the top 20 cities in terms of number of AI enterprises, 9 are in the US, 4 in China, 3 in Canada, and 1 respectively in the UK, Germany, France and Israel. Among them, Beijing has the largest number of AI enterprises (totally 412) in the world. The second is San Francisco and London, with 289 and 275 AI enterprises respectively. The number of AI enterprises in Shanghai, Shenzhen and Hangzhou has also entered the global Top 20.

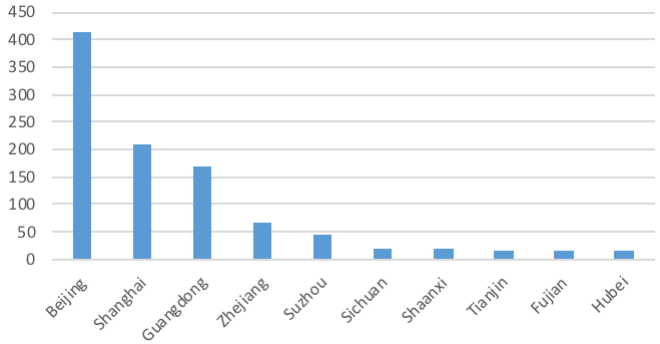
Fig. 4-5 TOP 20 Cities in Terms of Number of AI Enterprises



Source: CAICT (2018)

In China, AI enterprises are mainly concentrated in Beijing, Shanghai and Guangdong Province. Beijing ranks first, followed by Shanghai and Guangdong. In addition, Zhejiang Province and Jiangsu Province have also gathered a number of AI enterprises.

Fig. 4-6 Number of AI Enterprises in Major Provinces of China



Source: CAICT (2018)

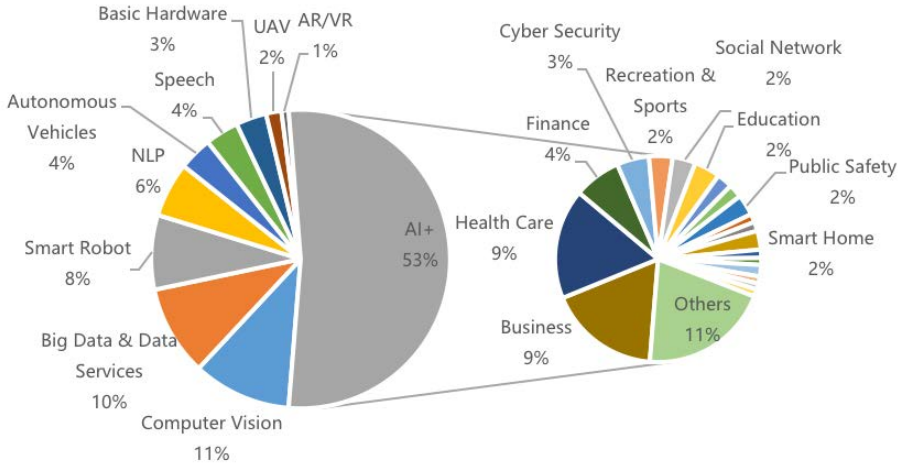
5 World Investment and Financing

5.1 Distribution of Investment and Financing

In the past five years, the global investment in AI industry focused on AI+ (vertical industry), vision, big data, data service and intelligent robotics. Among the various AI+ vertical industries, the most favored areas for capital are business intelligence, healthcare and finance.

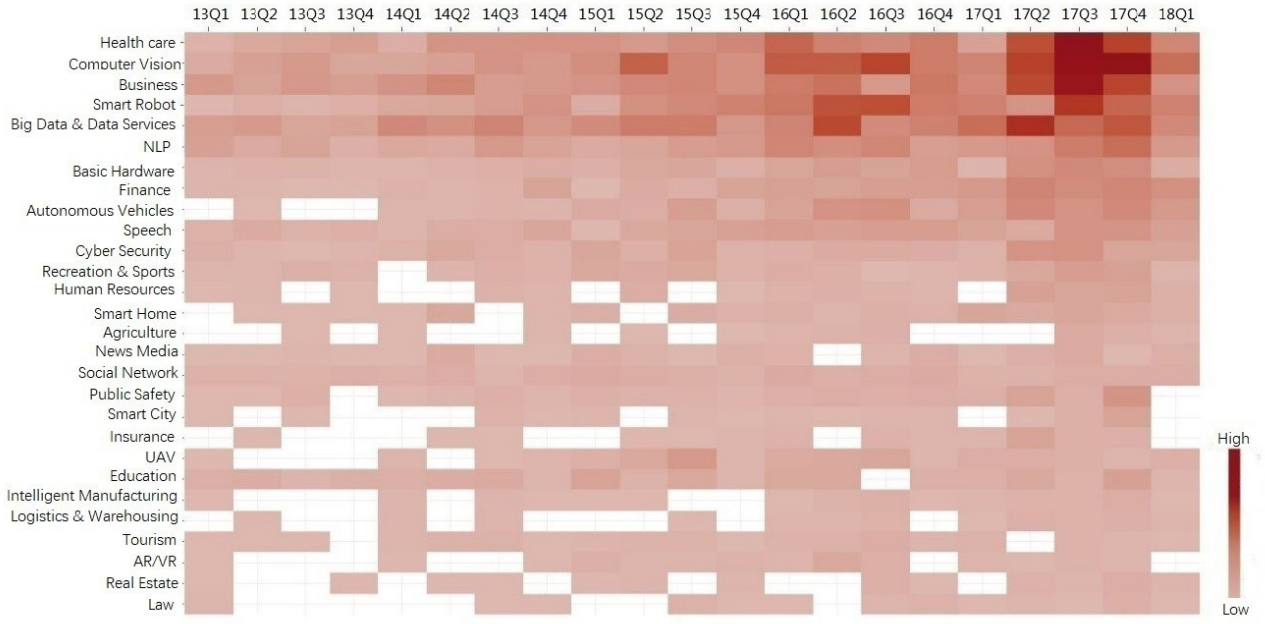
The quarterly investment and financing data show that the healthcare, vision, business intelligence and intelligent robotics sectors have maintained a high investment intensity since the second quarter of 2017, especially when it peaked in the third quarter, and then followed by a slight fall. The big data and data services field, as an evergreen tree in the AI field financing, compared with other areas in recent years, has attracted more capital investment. In general, it is also the field with the highest number of investments.

Fig. 5-1 Distribution of Numbers of Global AI Investment and Financing Projects (2013-2018 Q1)



Source: CAICT (2018)

Fig. 5-2 Distribution of Global Investment Intensity in Various AI Fields (2013-2018Q1)



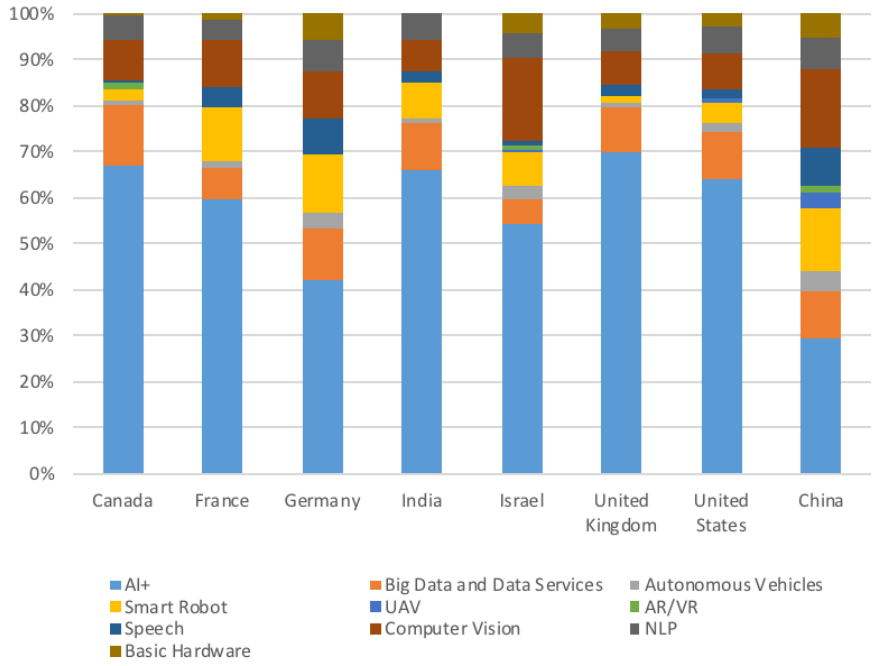
Source: CAICT (2018)

By comparing the number of financing projects in various AI fields in China, the US, the UK, Canada and other major countries, it is found that in China the distribution of fields is more balanced, while other major countries have more emphasis on vertical industry applications (AI+).

5.2 Investment and Financing Scale

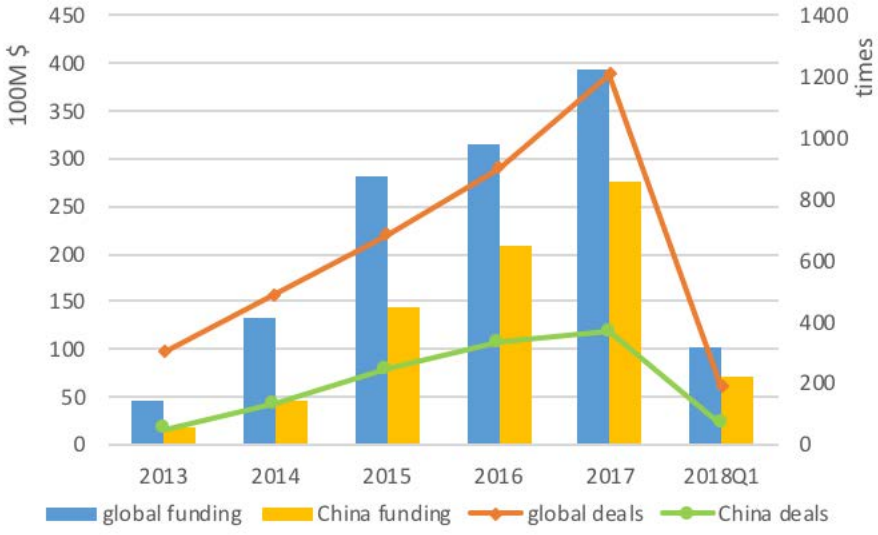
Since 2013, the global enthusiasm for investment and financing in the AI field has continued to rise. In 2017, the total scale of global AI investment and financing reached 39.5 billion USD, and China's investment and financing scale reached 27.7 billion USD, accounting for 70%, making it the world's most capital-absorbing country in the AI field. In contrast, the US accounts for 41% of the number of investment and financing projects, surpassing China, as the most active country in terms of investment and financing.

Fig. 5-3 Distribution of Number of AI Financing Projects in Major Countries (2013-2018Q1)



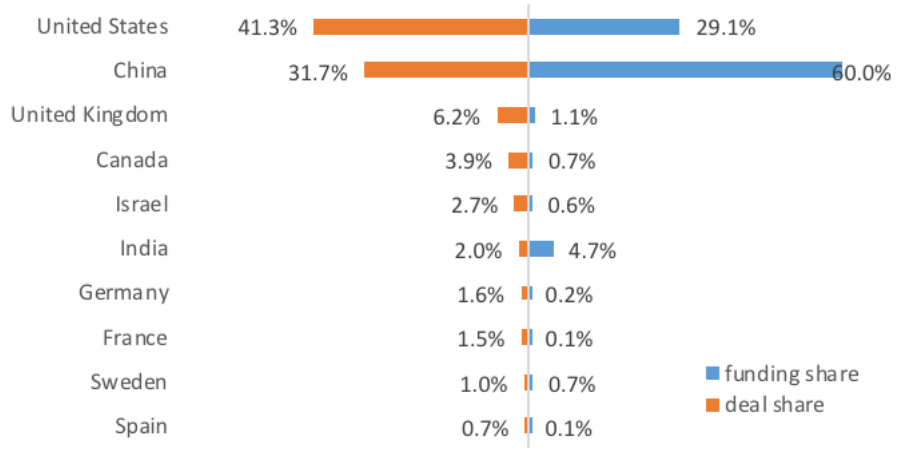
Source: CAICT (2018)

Fig. 5-4 Global (including China)/China AI Investment and Financing Trend



Source: CAICT (2018)

Fig. 5-5 Geographical Distribution of Global AI Investment and Financing (2013-2018 Q1)

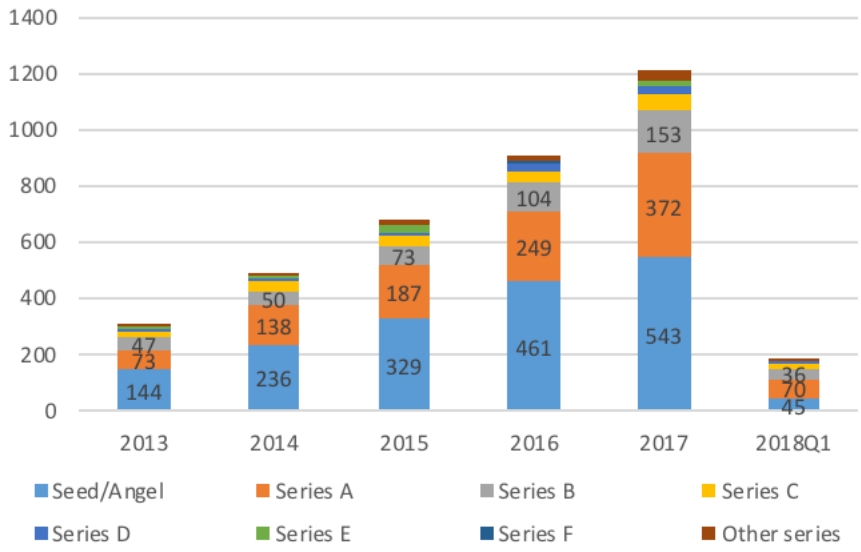


Source: CAICT (2018)

5.3 Distribution of World Rounds of Investment and Financing

In the past five years, globally, the number of financing rounds in the AI field showed a year-on-year growth trend, except for a slight fluctuation in the E and F rounds. In 2017, the annual growth rate was more than 40% for each of the A, B, C, D, and E rounds.

Fig. 5-6 Numbers of Global AI Investment and Financing Projects

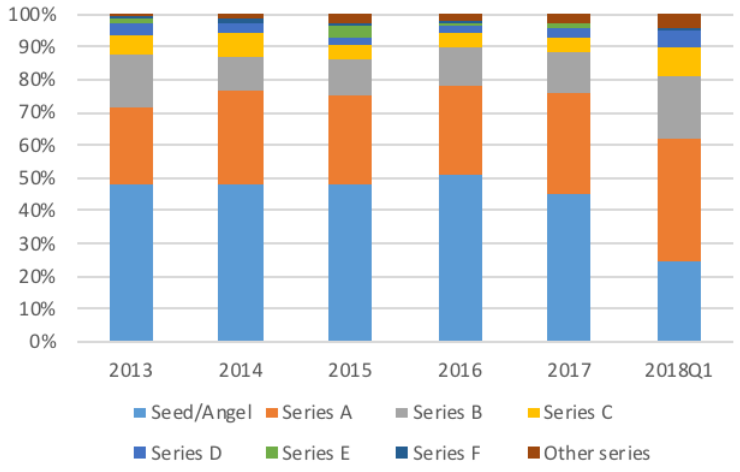


Source: CAICT (2018)

Early investment in the AI field in the world continued to be active, with the seed/angel round and A round investment being the highest, accounting for about 70% in total. With the development and maturity of the AI industry and technology, the proportion of B, C, and D rounds of financing has increased year by year, reaching about 20% in 2017.

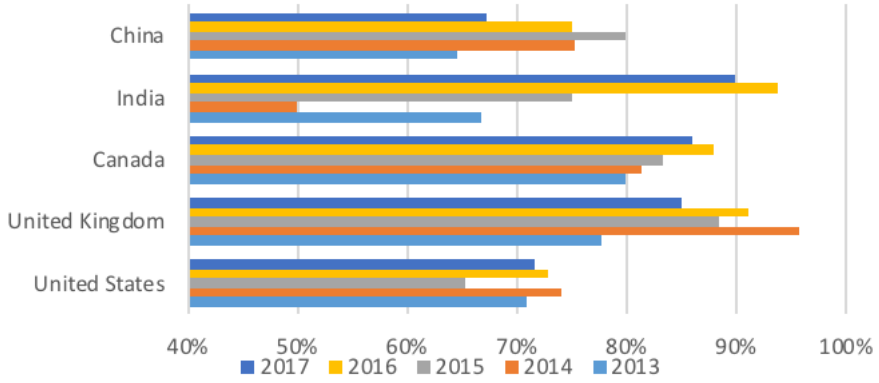
By comparing the trend of number investment projects in the seed round/angel round/A round in China, the US, the UK, Canada and India, it is found that the proportion of early investment in all these countries decreased in 2017, and China has declined obviously since 2016. These trends indicate that China's AI industry has experienced a period of entrepreneurial outbreaks and has begun to move toward a more mature stage of development.

Fig. 5-7 Proportion of Number of Global AI Financing Projects of the Rounds



Source: CAICT (2018)

Fig. 5-8 Total Proportion of Number of AI Financing Projects of Angel/Seed/A Round of Countries



Source: CAICT (2018)

6 Industrial Development

6.1 Industrial Development Technology

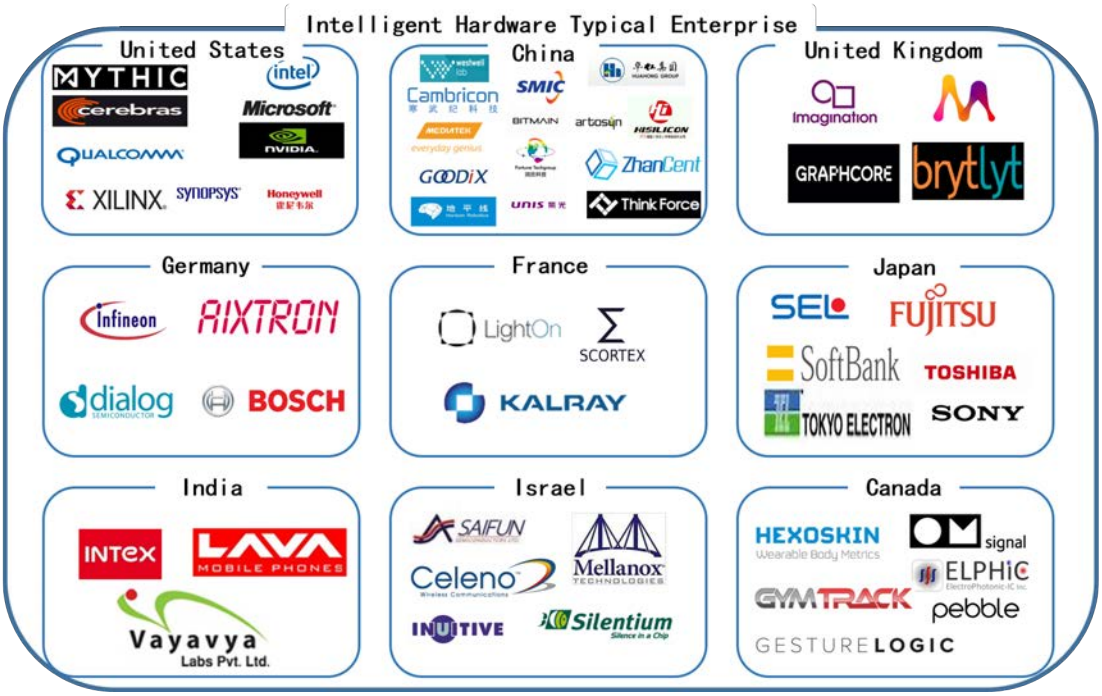
6.1.1 Intelligent Hardware

Intelligent sensors and intelligent chips are an important part of Intelligent hardware. If the intelligent chip is the central brain of AI, then the intelligent sensor belongs to the nerves with nerve endings. Different from traditional hardware, the intelligent sensor is a relatively independent intelligent processing unit with primary sensing processing capability that integrates traditional sensors, microprocessors and related circuits. The intelligent chip has high-performance parallel computing capability and supports mainstream artificial neural network algorithms. Intelligent sensors mainly include touch, visual, ultrasonic, temperature, distance sensors, etc.; intelligent chips mainly include GPU, FPGA, ASIC and brain-like chips.

The ResearchAndMarkets report shows that the global market value of intelligent sensors in 2017 was USD 26.906 billion. It is estimated that the total market size will reach USD 70.617 billion by 2023, and the CAGR will be 17.45% within the forecast period (2018-2023). The global AI chip market is expected to reach USD 10.8 billion by 2023, with a CAGR of 53.6% during the forecast period (2017-2023). *The Development Planning for a New Generation of Artificial Intelligence* predicts that by 2020, China's intelligent computing chip market will reach RMB 10 billion.

In the global intelligent hardware market, international giants such as Honeywell, BOSCH and ABB have comprehensively deployed a variety of types of intelligent sensors product; in China, there are also fingerprint sensors from Goodix and force sensors from ColliHigh, but the product layout is relatively simple. In terms of intelligent chips around the world, there are NVIDIA's GPU, Google's TPU, Intel's NNP and VPU, IBM's True North, ARM's DynamIQ, Qualcomm's Snapdragon series, Imagination's GPU Power VR and products of other mainstream enterprises; in China, there are Huawei HiSilicon's Kirin series, Cambrian's NPU, Horizon's BPU, Westwell Lab's Deepsouth and Deepwell, Unisound's UniOne, and Alibaba DAMO Academy's Ali-NPU.

Fig. 6-1 Typical Players for Intelligent Hardware



Source: CAICT (2018)

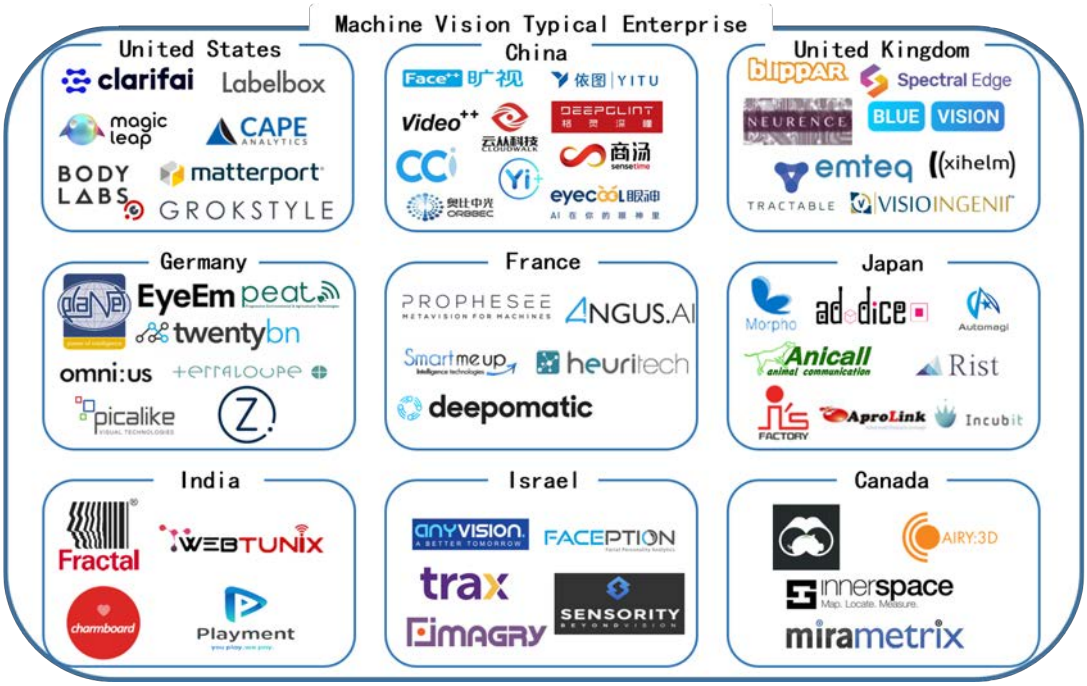
6.1.2 Machine Vision Technology

Compared with the traditional visual technology, AI empowers machine with the vision technology, which makes it have a visual perception and cognitive mechanism similar to human classification recognition of image features, with a series of advantages, such as fast speed, high precision and high accuracy.

From the point of view of technical ability, it mainly realizes the requirements of identification, classification, positioning, detection and image segmentation of object/scene within the picture or video and other functions in industrial applications, so it is widely used to realize video surveillance, automatic driving, vehicle/face recognition, medical imaging diagnosis, autonomous robot navigation, industrial automation systems, aviation and remote sensing measurement. According to the report of MarketsandMarkets, the global market for AI-based computer vision in 2017 was USD 2.37 billion, and is expected to reach USD 25.32 billion in 2023. During the forecast period (2018-2023), the CAGR is 47.54%. According to the report of Forward Industry Research Institute, in 2017, the size of China's computer vision market was RMB 6.8 billion. It is estimated that this figure will reach RMB 78 billion in 2020, with an average CAGR of 125.5%.

With the continuous integration of AI technology and the real industry, the image recognition capability of computer vision algorithms is getting stronger and stronger, and a large number of excellent computer vision enterprises have emerged in many countries. In the US, a number of multinational technology enterprises such as Amazon, Google, Microsoft and Facebook present the characteristics of an industry-wide layout covering the basic layer, the technical layer and the application layer; there are also some startups that focus on local application areas, e.g., Cape Analytics realizes smart valuation based on aerial photography of residence, and Photor, Steam, Oculus Home and Viveport serve as the three major VR content distribution platforms. In China, the technical experts of some top computer vision enterprises are mostly well-educated and well experienced, and the related industries get accumulation for many years. For example, SenseTime is currently providing AI+ shooting, AR special effects and AI identity verification services for major smartphone manufacturers; DeepGlint is deeply concentrating on visual algorithm technology and embedded hardware R&D technology; Yi+ is more about providing intelligent analysis and recommendation services for commercial visual content, while Cloudwalk Technology, Megvii Technology and YITU Technology have also different layouts.

Fig. 6-2 Typical Players in the Machine Vision Field



Source: CAICT (2018)

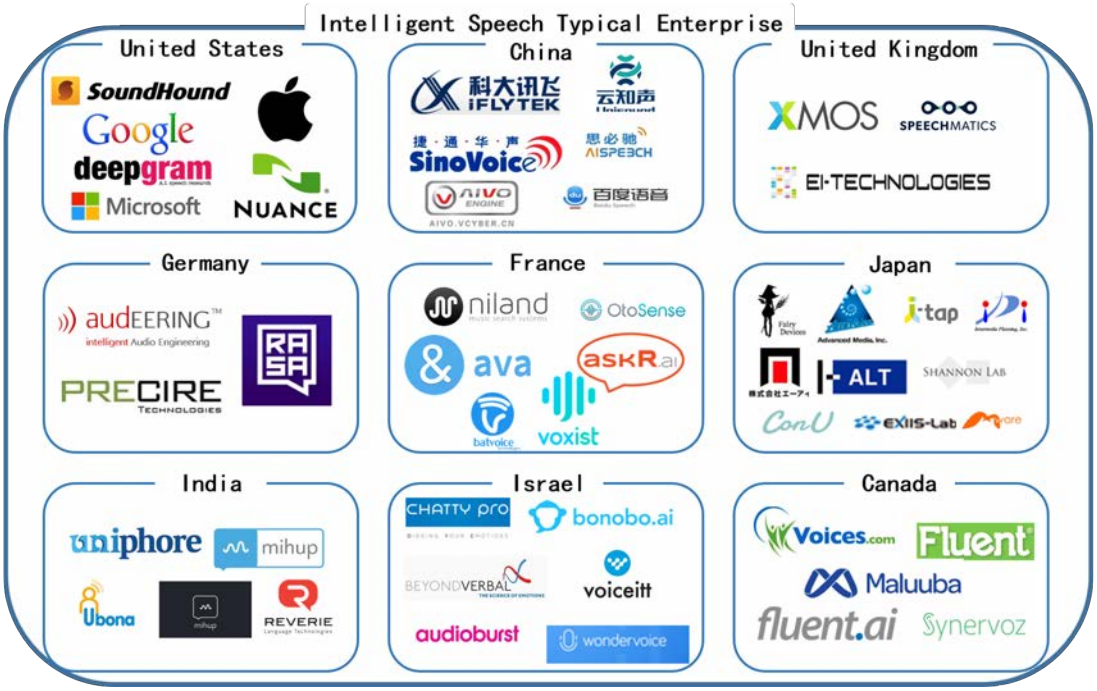
6.1.3 Intelligent Speech Technology

Intelligent speech technology is a technology that can realize the intelligent transformation of text or command and speech signals. It mainly includes speech recognition and speech synthesis. Speech recognition is like a "machine's auditory system" that transforms a speech signal into a corresponding text or command by recognizing and understanding. Speech synthesis is like a "machine's phonetic system", in which a machine can read the appropriate text or command and turn it into a personalized voice signal. The intelligent speech technology has been widely used in smart speakers, voice assistants and other fields because of its functions of human-computer speech interaction, speech control, voice-pattern recognition and so on.

According to data from ASKCI Consulting, in 2017, the global smart voice market amounted to USD 11.03 billion, a year-on-year increase of 30%. In 2017, the scale of China's intelligent speech market reached RMB 10.57 billion, an increase of 70% compared with 2016. With the expansion of the intelligent speech application industry and increasing market demand, it is expected that the scale of China's intelligent speech market will further increase in 2018, reaching RMB 15.97 billion.

At present, the application of intelligent speech technology on user terminals is the hottest. Many Internet enterprises have invested in human and financial resources to carry out research and application in this area, with the aim of rapidly capturing the customer base through the novel and convenient mode of voice interaction. In the US, Apple's Siri, Microsoft's PC-based Cortana, Microsoft's mobile-based Xiaobing, Google's Google Now, and Amazon's Echo are widely-known household applications; in China, iFLYTEK, AISpeech, Unisound, as well as the Internet giant BAT all involve themselves deeply in this area for the business layout.

Fig. 6-3 Typical Players in the Intelligent Voice Field



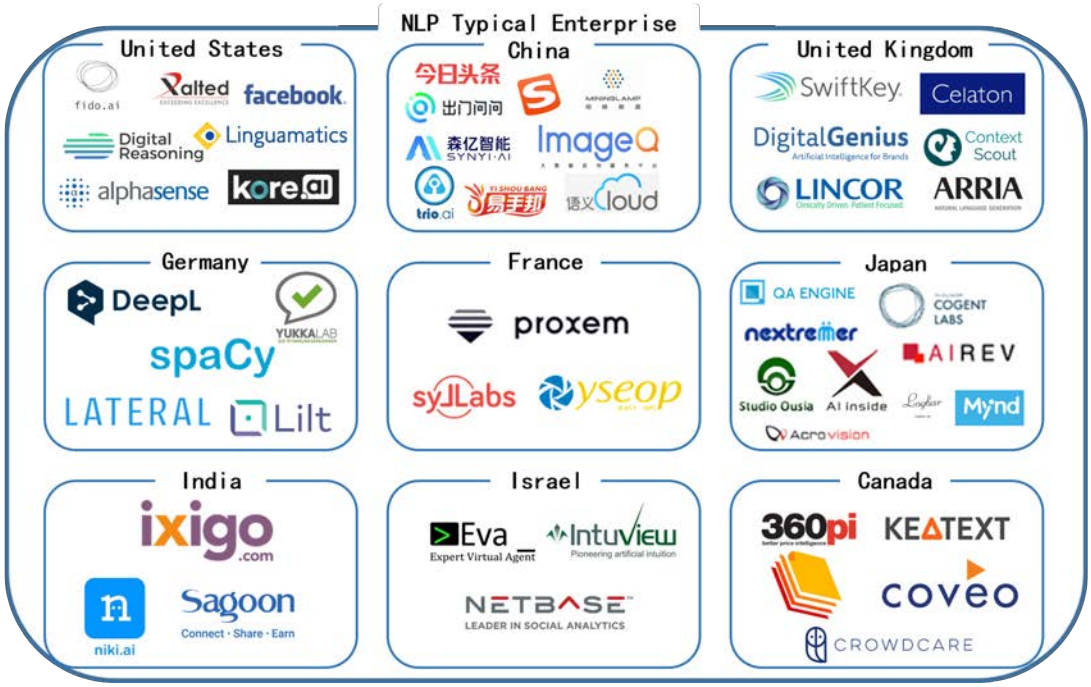
Source: CAICT (2018)

6.1.4 Natural Language Processing

Natural Language Processing encompasses a wide variety of research directions, including natural language understanding and natural language generation. To be clear, the former is to make the computer "understand" the idea or intention of the natural language text, while the latter is to make the computer "express" the thought or intention with the natural language text. From the perspective of application, NLP includes machine translation, public opinion monitoring, automatic summary, extraction of point of view, subtitle generation, text classification, question & answer (Q&A), text semantic comparison, etc.

According to the data of MarketsandMarkets, the global NLP market is expected to grow from USD 7.63 billion in 2016 to USD 16.07 billion in 2021, a CAGR of 16.1%. According to China's AI Development Report 2018, the scale of China's AI market reached RMB 23.7 billion in 2017, of which the NLP market was RMB 4.977 billion, accounting for 21%. At present, there are many related mature technology application products. For example, Amazon and Facebook of the US, and Toutiao of China use natural language technology to implement product reviews, social networking or news platform product review, community review, and news article topic classification and sentiment analysis over their own shopping website, social platform or news platform; Google, Baidu and Youdao are well-versed in and are constantly upgrading their online translation services; Logbar from Japan, iFLYTEK and Sogou from China have developed their own multilingual translation machine. In terms of the basic platform, there are Kore.ai, Linguamatics, etc. in the US, as well as Baidu Cloud, Tencent Wenzhi and LTP-Cloud in China. The application of the public opinion monitoring system includes iAcuity of Xalted in the US, the Wom-Monitor of Chaowen Tianxia in China, and Benguo public opinion monitoring of APEX Technologies.

Fig. 6-4 Typical Players in the Field of NLP



Source: CAICT (2018)

6.2 Industrial Development Application

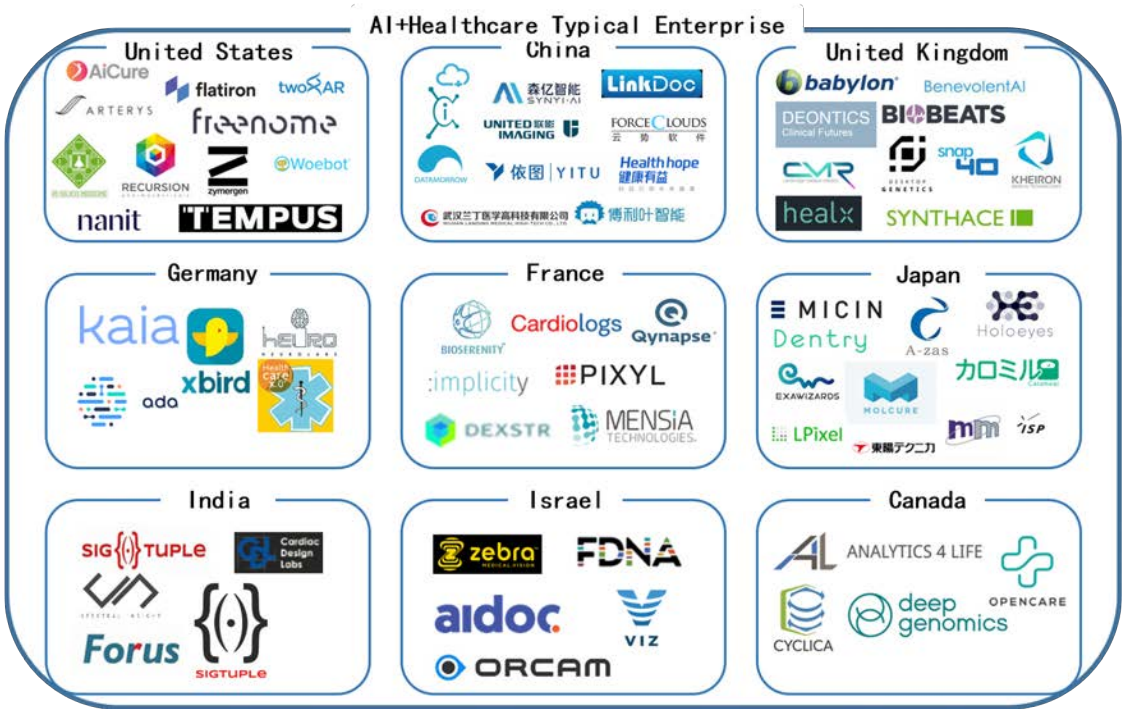
6.2.1 AI+ Healthcare

AI technology empowers the healthcare field, which significantly improves the efficiency of medical institutions and personnel, significantly reduces medical costs, and enables people to achieve scientific and effective daily health check-up and prevention and better manage their own health.

According to data from the ICT monitoring platform of CAICT, in recent years, AI+ healthcare ranks among the hottest areas of AI+ vertical application. From the perspective of application, intelligent healthcare mainly includes medical research, pharmaceutical research and development, intelligent diagnosis and treatment, and family health management. From the perspective of technical subdivision, it mainly includes the use of machine learning technology to achieve drug performance prediction, crystal form prediction, gene sequencing prediction, etc.; use of intelligent speech and NLP technology to achieve electronic health records, intelligent inquiry, guiding consultation, etc.; use of machine vision technology to realize medical image recognition, lesion identification, skin disease self-test, etc. According to McKinsey's forecast, by 2025, the global intelligent healthcare industry will reach a total of USD 25.4 billion, accounting for about one-fifth of the global AI market. China is at the tuyere of medical AI. According to data from Forward Industry Research Institute, the scale of China's medical AI market exceeded RMB 13 billion in 2017, and it is expected to reach RMB 20 billion in 2018. The space for medical AI is broad.

Currently, in medical research and pharmaceutical R&D, BergHealth and Numerate use data to drive drug discovery, vion and HBI Solution provide patient disease prediction and risk analysis for medical institutions; in the field of intelligent diagnosis, IBM Watson focuses deeply on the field of cancer, and through acquisition and cooperation, continues to accumulate medical data resources and expand capabilities in various fields. Alibaba's "Doctor You" series, Tencent's Miying, YITU Technology's "care.ai™" and PereDoc's intelligent image-assisted diagnosis and treatment platform realizes medical image-assisted diagnosis and treatment, and Fourier Intelligence's Fourier X1 gives birth to China's first exoskeleton robot. In the field of family health management, WellTok pays more attention to personal health management and lifestyle improvement, AiCure is committed to helping users to use drugs on time, and iCarbonX address itself to the building of a digital life health management platform.

Fig. 6-5 Typical Players in the Smart Healthcare Field



Source: CAICT (2018)

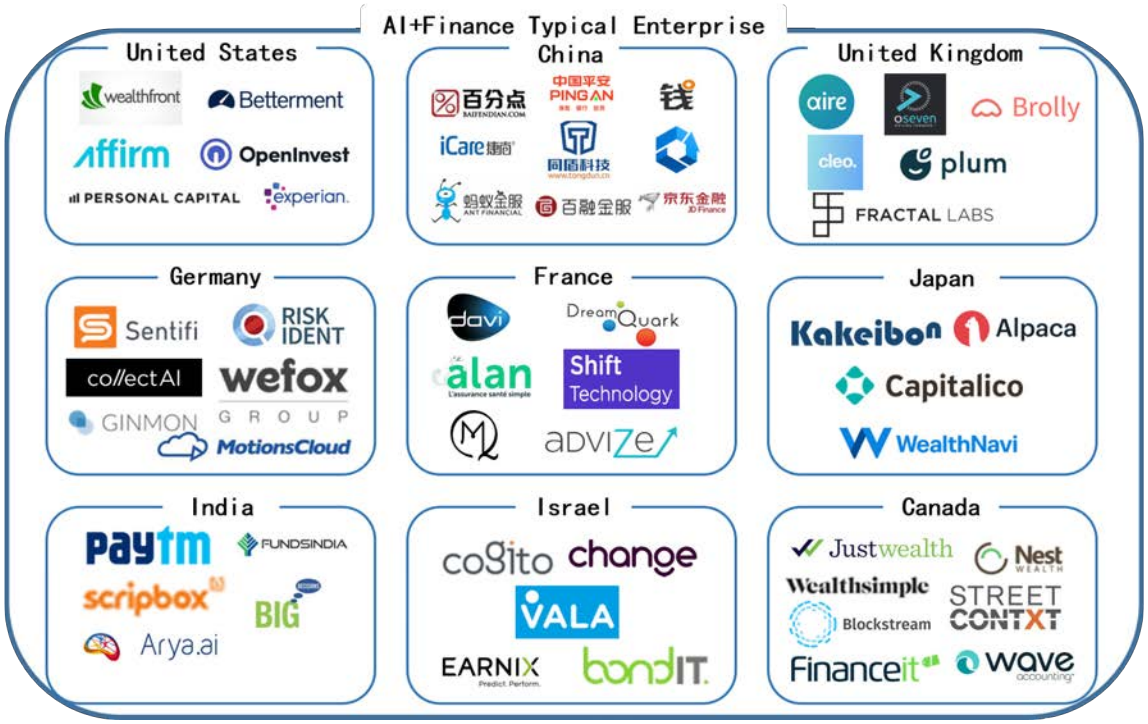
6.2.2 AI+Finance

AI technology empowers the financial field. From an application perspective, it mainly includes robo-advisor, credit/risk control, financial search engine, insurance, identity verification and intelligent customer service. Finance is one of the most data-dependent industries. AI technology is integrated with the financial industry to drive intelligent upgrade of financial technology through the AI technology based on big data. In the foreground, it can be used to provide users with more comfortable, convenient and safer services; in the middleground, it can provide decision-making support function for transactions, credits and analysis in financial services; in the background, it can improve the financial system for various risks. In the background, it can improve the ability of the financial system to identify, warn, prevent and control all kinds of risks. All in all, AI technology will deeply reconstruct the current financial industry's ecological structure, making financial services (banking, insurance, wealth management, lending, investment, etc.) more humane and intelligent.

According to the PwC 2017 Global Digital IQ Survey, the information utilization rate in the global financial service sector is only 26%, which is a low level compared with other industries. According to the report of MarketsandMarkets, the global market size of AI in financial technology is expected to grow from USD 1.338 billion in 2017 to USD 7.306 billion in 2022, a CAGR of 40.4%. *The White Paper on the Development of a New Generation of Artificial Intelligence (2017)* predicts that China's intelligent financial industry will reach USD 800 million by 2020.

At present, robo-advisor enterprises are mainly transformed from licensed securities, fund or asset management operators, e.g., Wealthfront and Betterment of the US, as well as Koudai Fortune and JD.com robo-advisor of China. In the financial intelligent customer service field, Digital Genius, Qiyukf, sobot and other enterprises focus on improving the user experience; in the field of credit/risk control, most of the players rely on government information, enterprise information or personal information to form industry barriers based on big data intelligence analysis, such as Zest Finance and Affirm, both being the US financial technology enterprises with consumer finance and mobile payment data, and Qixin, which has enterprise multi-dimensional real-time dynamic and full-scale business data credit platform; other enterprise applications, such as Rong 360, Data.GOV and DBpedia focus on financial search engines; SenseTime, Cloudwalk Technology, YITU Technology and Megvii Technology have entered the identity authentication market by relying on their leading face recognition core technology.

Fig. 6-6 Typical Players in the Global Smart Finance Field



Source: CAICT (2018)

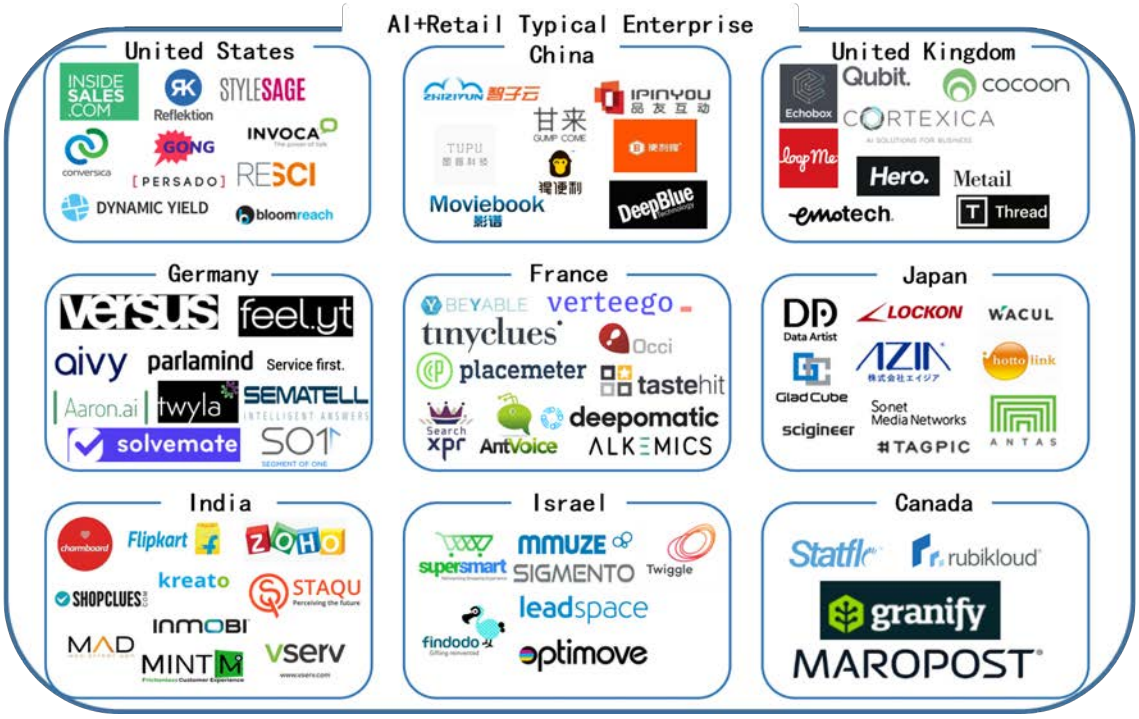
6.2.3 AI+Retail

AI technology empowers the retail industry. Intelligent retail drives the new market retail business pattern with big data and intelligent technologies, optimizing resource allocation and efficiency across the entire industry chain from production to distribution to sales, thus enabling intelligent upgrades of industry services and performance. Its commercial applications include intelligent marketing recommendation, intelligent payment system, intelligent customer service, unmanned storehouse/unmanned vehicle, unmanned store, intelligent distribution and more.

The MarketandMarkets report shows that the global intelligent retail market is expected to grow from USD 13.07 billion in 2018 to USD 38.51 billion in 2023, with a CAGR of 24.12% over the forecast period (2018-2023). In China, according to the National Bureau of Statistics, at the end of 2017, the total retail sales of consumer goods in China amounted to RMB36.6262 trillion, an increase of 10.2%. Roland Berger predicts that by 2030, AI technology will bring about RMB 420 billion in cost reduction and gain increase to the Chinese retail industry.

Nowadays, the global intelligent retail industry participants are mainly e-commerce giants and startups. As for the business application scenarios, it is still mainly on the sales side. For example, in the US, for the unmanned retail store, there are Standard Cognition, the unmanned convenience store, and Amazon's Amazon Go. In China, there are Alibaba's Taocafe, as well as JD.com's X unmanned supermarkets; there are also related products of famous startups such as DeepBlue Technology, F5 Future Store, and Bingo Box. In the aspect of customer service robots, China has Cheetah Mobile's Fanbot retail robots, Keenon Robotics's Peanut guiding robots, and SIASUN's Sunbot-I sales promotion and guiding robot, which have been applied in various application scenarios. In the smart retail supply chain scenario, UPS in the US tested UAV distribution in Florida; Wal-Mart's Pickup Tower is being deployed across the US. In China, Meituan has launched an unmanned distribution open platform; JD.com is building an all-environment intelligent retail logistics with unmanned distribution stations, unmanned warehouse "Asian One" and large cargo drone "Jinghong".

Fig. 6-7 Typical Players in the Smart Retail Sector



Source: CAICT (2018)

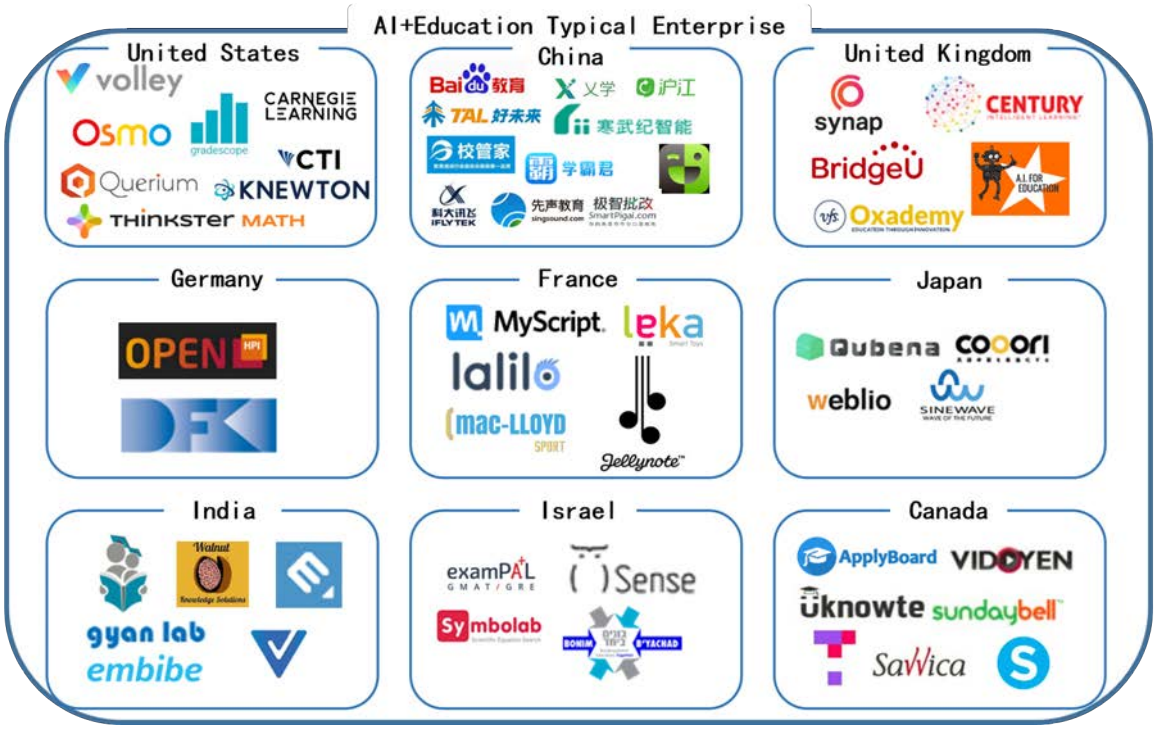
6.2.4 AI+ Education

AI technology empowers education. It pays attention to the individualized education of students, helps teachers to teach students according to their aptitude, enhances the quality of teaching and learning, and promotes equalization and affordability of education. At present, the intelligent education comprehensively covers the industrial chain of "teaching, learning, examination, evaluation and management", and has accelerated in the subdivision market, such as preschool education, K12, higher education, vocational education and online education. From the perspective of application, intelligent education can be divided into four parts: learning management, learning evaluation, teaching guidance, and teaching cognitive thinking. From the perspective of subdivision, it includes educational evaluation, photo-taking & answering, intelligent teaching, intelligent education, intelligent scoring, AI adaptive learning and other application scenarios.

According to the report of MarketsandMarkets, in 2017, the global AI technology market size in the education industry was USD 373.1 million, and is expected to reach USD 3,683.5 million by 2023. The CAGR during the forecast period (2018-2023) is 47.0%. Global market Insights also released a new research report that predicts that the value of the AI industry in the education market will exceed USD 6 billion in 2024. Among them, the Asia-Pacific intelligent education market, including China, will have a CAGR of more than 51%, making it the most profitable region.

At present, in the field of AI adaptive learning, whether it is Knewton in the US, Century Tech in the UK, Smart Sparrow in Australia, and Yixue Education and TAL Education of China, the AI education platform is used to help students quickly grasp knowledge points and improve their learning result. In the field of intelligent evaluation, China's Xuebajun and iFLYTEK and other enterprises have launched intelligent scoring system; in the field of teaching guidance, China's Singsound and Liulishuo and other enterprises have introduced language tutoring system, Tabtor, Carnegie Learning, Front Row and other enterprises of the US have launched the intelligent tutor system to simulate one-on-one tutoring, which forms the effect of close expert guidance. LightSail and Newsela of the US have made personalized and intelligent recommendations for students' reading, and developed students' reading ability and interest.

Fig. 6-8 Typical Players in the Intelligent Education Field



Source: CAICT (2018)

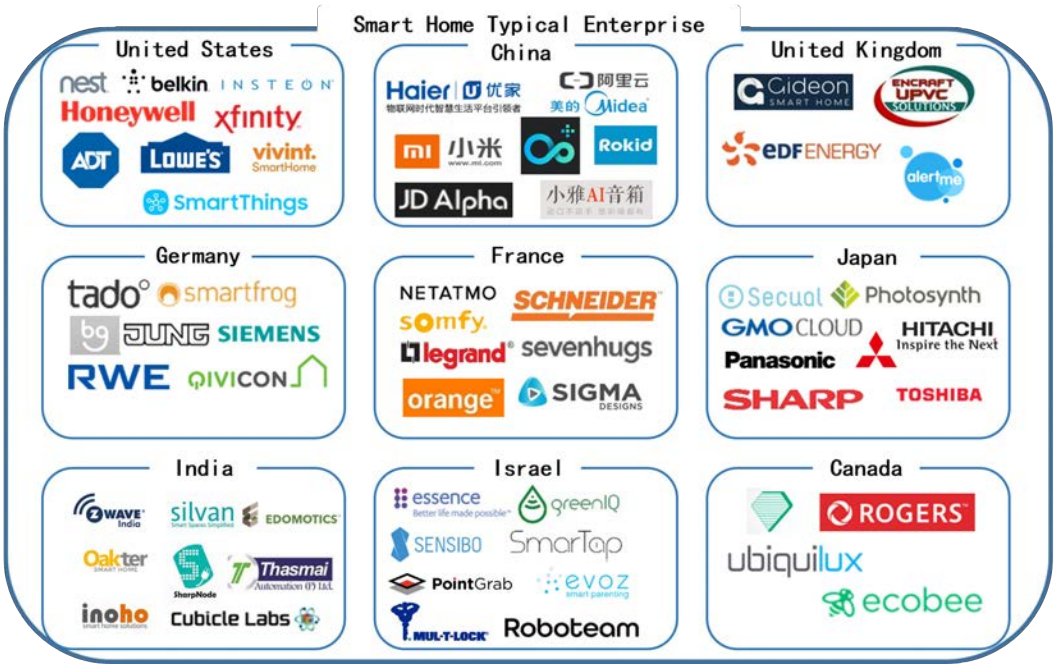
6.2.5 AI+ Home

AI technology empowers the home, helping the home ecosystem to develop from perception to cognition, making home life safer, more comfortable, more energy efficient, more efficient and more convenient. In the future, smart home will gradually implement adaptive learning and control functions to meet the individual needs of different families. Smart home is an IoT-based home ecosystem that includes the intelligent lighting system, intelligent energy management system, intelligent audiovisual system, and intelligent security system.

According to the latest Strategy Analytics report, the global smart home market reached USD 84 billion in 2017, up 16% from 2016's USD 72 billion, and is expected to reach USD 96 billion in 2018. According to data from ibaogao.com, the size of China's smart home market in 2017 was RMB 91.66 billion, and it is expected to expand to RMB 139.6 billion in 2018.

In recent years, smart home has shown strong vitality on a global scale. As the largest market for smart home, the US pays attention to smart home with smart speakers as the central control, including Amazon Echo, Google Home and other hot sale products. In China, on the one hand, major players in the market have released various types of smart speaker products, such as Alibaba's "Tmall Genie Smart Speaker", Xiaomi's "Mi AI Speaker", iFLYTEK-JD.com's "DingDong Smart Speaker", Baidu's "Xiaodu Smart Speaker", Tencent's "Tingting Smart Speaker", Rokid's "Rokid Smart Speaker", and Himalaya's "Xiaoya Smart Speaker". On the other hand, major players are also actively building smart home with IoT platforms, such as Xiaomi's MioT, Huawei's HiLink, and Haier's U+.

Fig. 6-9 Typical Players in the Smart Home Sector



Source: CAICT (2018)

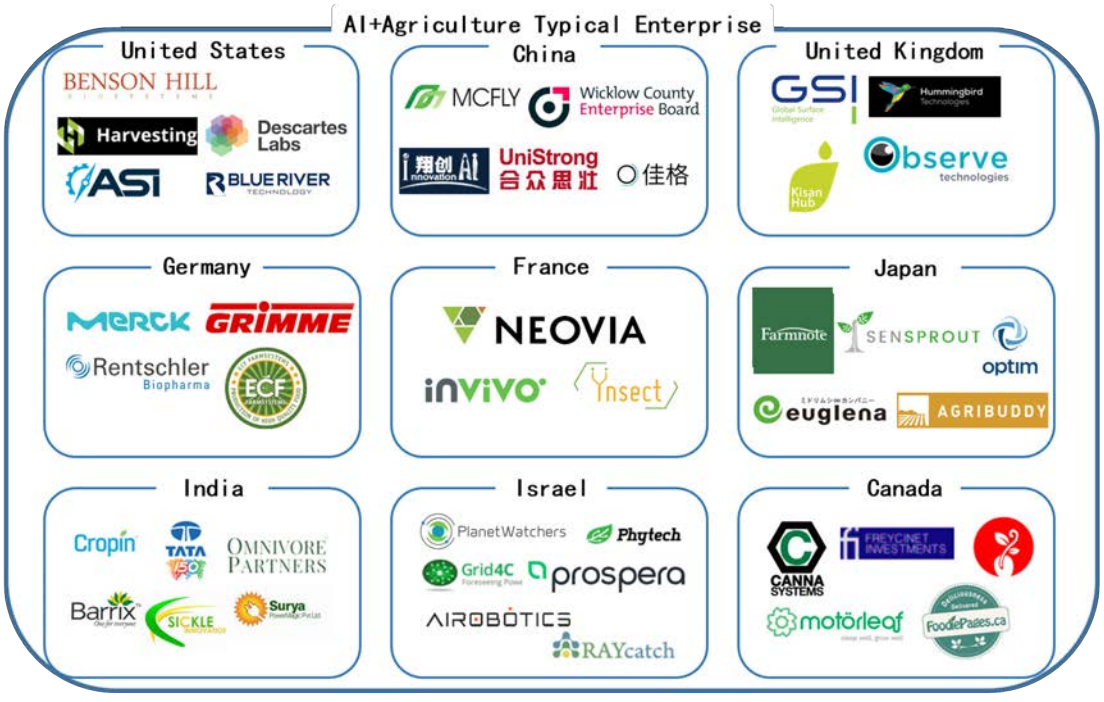
6.2.6 AI+Agriculture

AI technology empowers the agricultural sector, enabling agriculture to effectively cope with the effects of extreme weather, reduce resource consumption, optimize resource allocation, reduce costs, and optimize time and resource allocation for maximum yield and efficiency. From the perspective of application, intelligent agriculture mainly includes agricultural robots, precision farming, drone analytics, and livestock monitoring.

According to the report of MarketsandMarkets, the global intelligent agriculture market amounted to USD 6.7 billion in 2017. It is expected to reach USD 7.53 billion in 2018 and USD 13.5 billion in 2023. The CAGR in the forecast period (2018-2023) is 12.39%. Among them, the value of AI technology in the agricultural market in 2016 was USD 432.2 million, and the value is expected to be USD 2.6285 billion by 2025. The CAGR during the forecast period (2017-2025) is 22.5%.

As the world's first agricultural power, the US has always led the development of intelligent agriculture industry. In precision farming, enterprises such as Prospera, Arable and Trimble use cameras, sensors, micrometeorological data or positioning technology to monitor and analyze crops; in the field of agricultural intelligent devices, Blue River has created "LettuceBot" robots, "see-and-spray" system and the unmanned aerial system. At the same time, the intelligence of China's agricultural industry is also accelerating its transformation. For example, McFly's intelligent agricultural monitoring drone, UniStrong's "Huinong" Beidou navigation agricultural automatic driving system, and GAGO's large-scale application of AI technology in farming and pig raising are typical cases. At the same time, some technology giants have also begun to make deployment in the intelligent agriculture sector. In April 2018, JD.com's "JingDong Farm" made its debut; in June, Alibaba Cloud's ET agricultural brain came out.

Fig. 6-10 Typical Players in the Intelligent Agriculture Sector



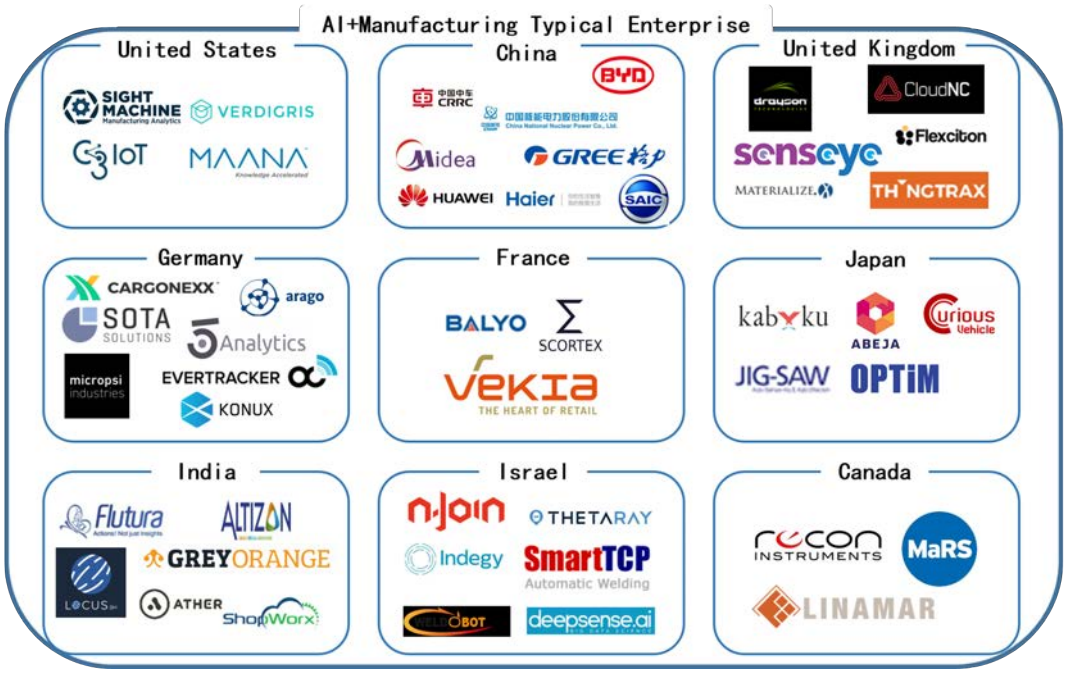
Source: CAICT (2018)

6.2.7 AI+ Manufacturing

AI technology empowers the manufacturing sector to significantly optimize the manufacturing cycle and efficiency, improve product quality, and reduce labor cost. The intelligent manufacturing industry chain has a wide range of scenarios. Its typical application scenarios include intelligent product and equipment; intelligent factory, workshop and production line; intelligent management and service; intelligent supply chain and physics; intelligent software development and integration; intelligent monitoring and decision-making, etc. Data from Market Research shows that the global intelligent manufacturing market reached USD 202.82 billion in 2017 and is expected to reach approximately USD 479.01 billion in 2023. The CAGR during the forecast period (2018-2023) is approximately 15.4%. According to data from Forward Industry Research Institute, in 2017, the output value of China's intelligent manufacturing industry has reached about RMB 1.5 trillion. It is expected that in the next few years, China's intelligent manufacturing industry will maintain an average CAGR of around 11%. By 2023, the industry market will reach RMB 2.81 trillion, and the industry has huge room for growth.

In recent years, intelligent manufacturing has become the main battlefield for industrial upgrading of the countries, and some developed countries have gone far in this respect. For example, the intelligent workshop in Phoenix, Germany, the AWS cloud ecosystem of C3 IoT in the US, and the intelligent manufacturing unit of Harley-Davidson in the US. In the field of smart device monitoring, there are also enterprises such as KONUX in Germany, Scortex in France, and BrainsTechnology in Japan. Some Chinese enterprises are also building smart factories and increasing the intensity of enterprise transformation and upgrading. For example, Eston's industrial robot intelligent factory established in Nanjing, Trumpchi's smart factory in Hangzhou, and the digital factory of CRRC Nanjing Puzhen. In the traditional household appliance manufacturing industry, Midea, Haier, Gree and other enterprises are actively shifting toward the intelligent manufacturing model.

Fig. 6-11 Typical Players in the Intelligent Manufacturing Sector



Source: CAICT (2018)

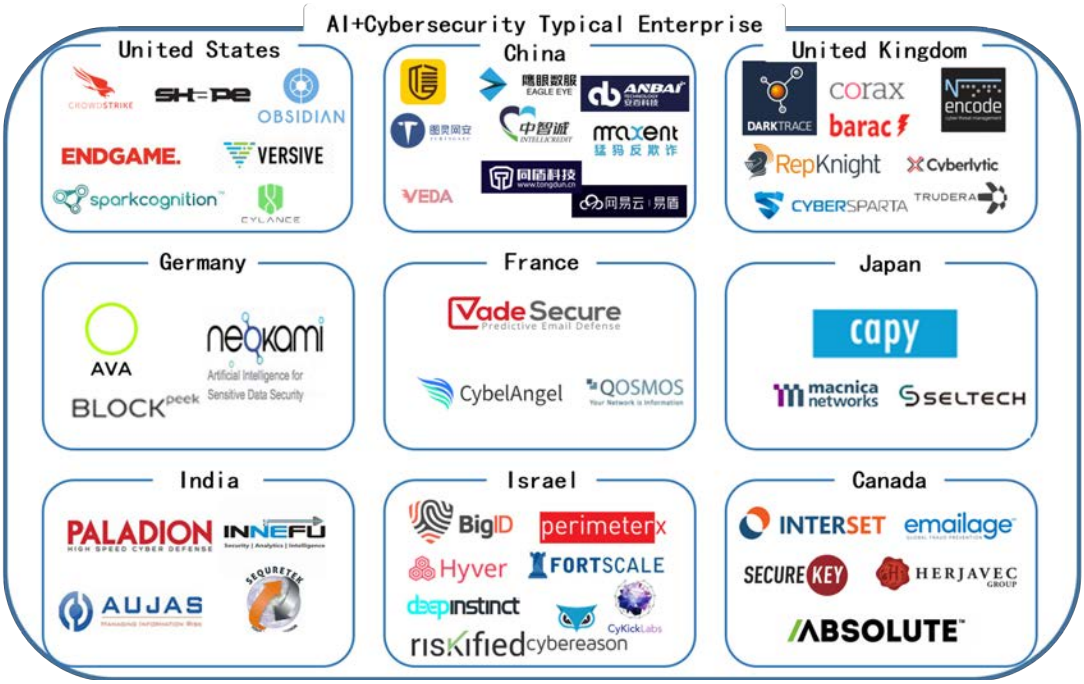
6.2.8 AI+ Cybersecurity

AI technology empowers the entire network, helping manufacturers, enterprises, and individuals to effectively address the growing number of cybersecurity issues such as cyber fraud and malicious attacks. AI has some unique advantages in the field of network defense, which make it a breakthrough for AI cybersecurity protection. The main applications of intelligent cybersecurity include: network monitoring and prevention (including real-time identification, response and defense against network attacks, and security vulnerabilities and system failure prediction, cloud security, etc.); prevention of the execution of malware and files; improvement of operational efficiency of security operation center; network traffic anomaly detection; application security detection; network risk assessment.

According to Technavio's latest market research report, the global intelligent cybersecurity market was USD 4.96 billion in 2017, and the estimated global AI-based cybersecurity market will grow at a CAGR of over 29% during 2018-2022.

Governments attach great importance to cyber security. For example, IBM and CrowdStrike of the US, and RepKnight of the UK provide cybersecurity protection service such as blocking malware and detecting phishing and data leakage by providing a cybersecurity protection platform. Fraud and risk detection are another important application in the field of cyber security. Enterprises such as DataVisor and Drawbridge from the US, and Feedzai from Portugal have their related business layouts. In recent years, China's intelligent cybersecurity enterprises have seen rapid growth. Among them, TuringSec, 360, NetEase Cloud and other enterprises provide network defense and anti-virus service, while Eagle Eye Tech, Maxent, Ahi Fintech, Tongdun Technology and other enterprises provide data monitoring and anti-fraud service.

Fig. 6-12 Typical Players in the Field of Intelligent Cybersecurity



Source: CAICT (2018)

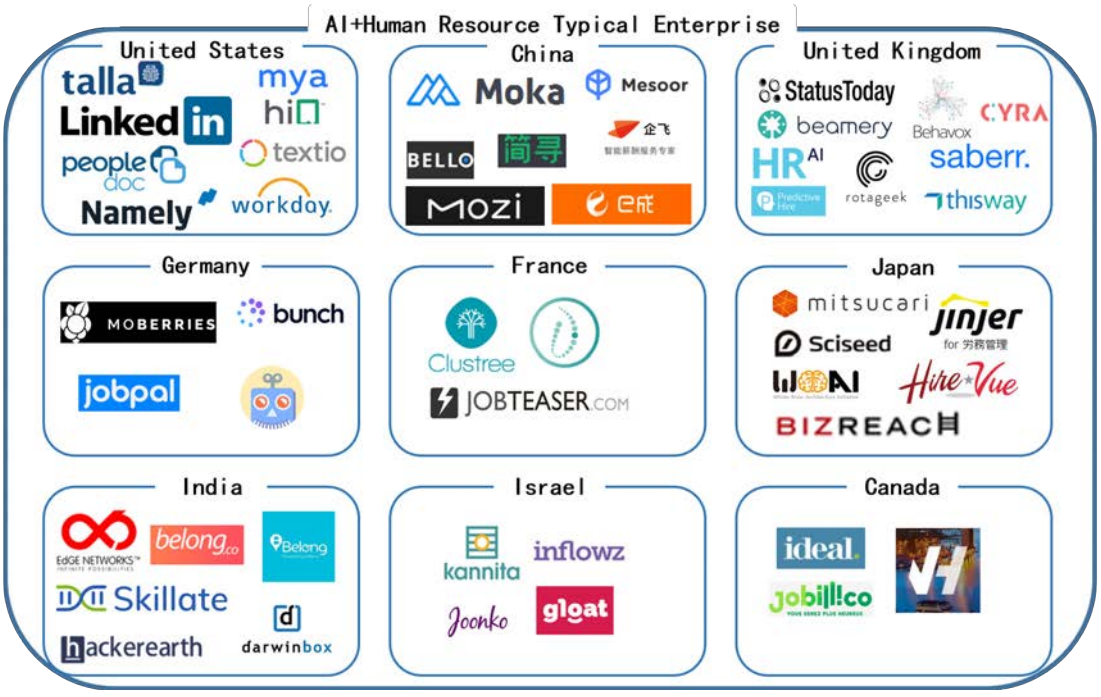
6.2.9 AI+ HR

AI technology empowers the HR field, which helps HR to automate the process of management process, greatly improve work efficiency and compliance, as well as reduce staff recruitment and management costs and personal bias. The main application contents include pre-recruitment talent channel maintenance, talent forecast analysis, job matching, resume screening, AI chat support, etc.; appointment for interview, interview result inquiry, handling employment procedures, etc. during the recruitment process; new employee training, Q&A interaction, knowledge learning and career planning support; employee behavior and efficiency analysis, salary analysis, mental health analysis, team culture analysis, etc. after the recruitment. According to Grand View Research, the global HR management market size was valued at USD 12.6 billion in 2016. It is forecast to reach USD 30 billion by 2025, and the CAGR during the forecast period (2017-2025) is 10.4%. According to ASKCI Consulting, the scale of China's HR service market increased from approximately RMB 158.4 billion in 2013 to RMB 343.6 billion in 2017, with a CAGR of 21.9% and is expected to increase to RMB 842.7 billion in 2022.

The international HR recruitment sector attaches great importance to the application of AI. The international HR recruitment sector attaches great importance to the application of AI, with the US playing a leading role. IBM's Watson contains three frameworks of HR, which, through the deep learning of models, finds professionals, seek insight into the needs of competitors, and evaluate job experience and skill perception.

Ultimate Software offers a range of HR solutions from recruitment to retirement. Japan's HR giant Recruit is known for its talent appraisal technology. Its business model consists of two parts: Career Adviser (CA), which serves for personal consultation for job seekers, and Recruit Adviser (RA), which serves for the recruiting enterprise. Many Chinese enterprises have applied AI technology to the daily HR work. For example, Emotibot's intelligent HR assistant robot, Beichoo Technology's HR super assistant "Xiaobei", and the "MOBOT" developed by Mozi AI can share most of the work of recruiters. Ifchange uses AI to provide scientific judgment and decision-making basis, and helps enterprises to fully increase their human capital.

Fig. 6-13 Typical Players in the Field of Intelligent HR



Source: CAICT (2018)

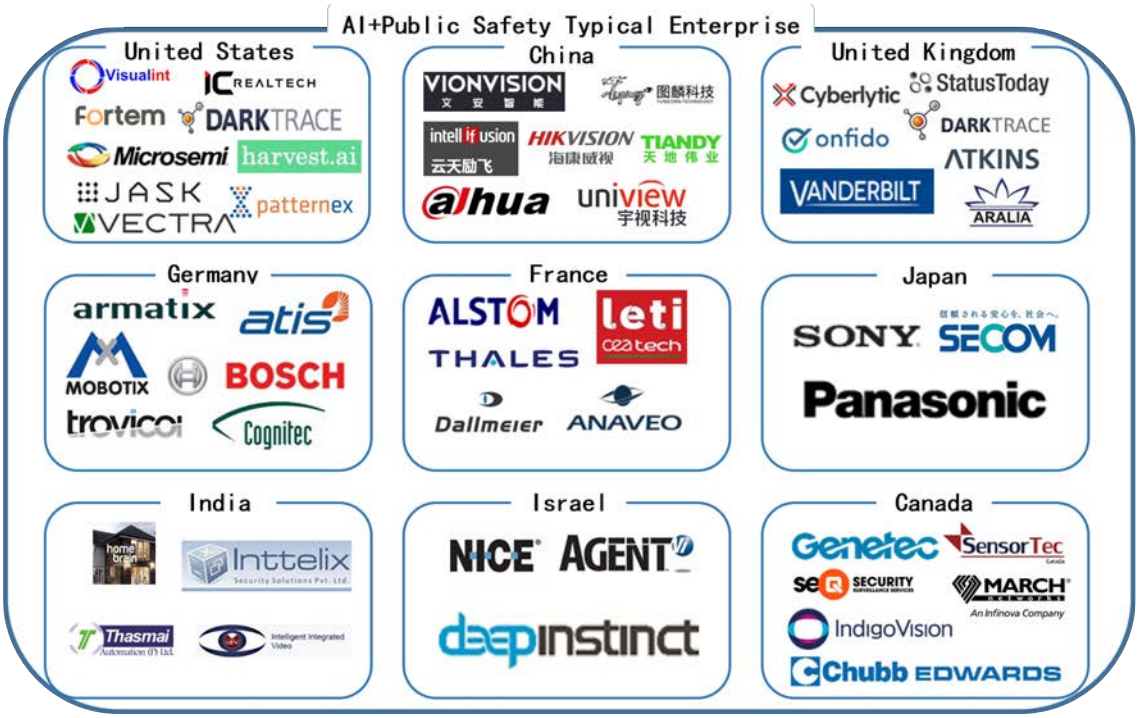
6.2.10 AI+ Public Safety

AI technology empowers the security field, filling the gap that traditional security is unable to meet the industry's requirements for the accuracy, breadth and efficiency of security systems. Intelligent security is the area that gets earliest large-scale application of AI and continues to generate commercial value. Its functions put into application are mainly in object tracking, detection and abnormal behavior analysis, video quality diagnosis and summary analysis, face recognition and feature extraction analysis, vehicle identification and feature extraction analysis, etc.

According to analysis of Mordor Intelligence, the market size of global video surveillance system was USD 34.9623 billion in 2017, and is expected to reach USD 82.6153 billion in 2023. The CAGR in the forecast period (2018-2023) is 15.41%. According to CAIJING.COM.CN, the output value of China's security industry reached RMB 450 billion in 2017, but the output value of all AI security products is less than RMB 2 billion, indicating a less than 1% penetration rate of AI in the security industry.

The establishment of intelligent security systems is inseparable from the integration of software algorithm and hardware system. In terms of system hardware, in the world, there are Axis Communications AB, the video processing chip vendor in Sweden, as well as ADT and OPTEX, the security hardware providers in the US. In China, Hikvision, Dahua, NetPosa and other enterprises are the market leaders in the relevant fields. In terms of software algorithms, Israel's Agent Video Intelligence, Canada's Genetec, the US' Google, Facebook and Microsoft, and China's SenseTime, Megvii Technology, YITU Technology, Cloudwalk Technology and other enterprises have excellent image analysis algorithms.

Fig. 6-14 Typical Players in the Field of Intelligent Public Safety



Source: CAICT (2018)

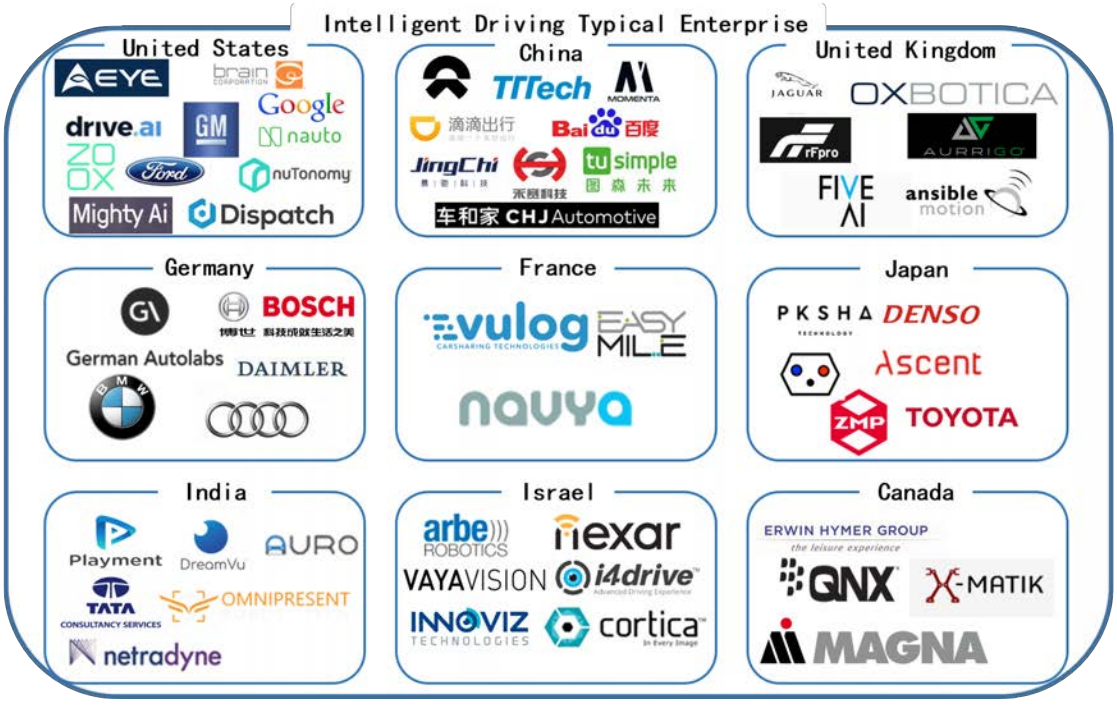
6.2.11 Intelligent Driving

AI technology empowers traditional driving, which can effectively improve production and transportation efficiency, alleviate labor shortage, achieve safety, environmental protection and high efficiency, and lead the overall upgrading and reshaping of industrial ecology and business models. Intelligent driving is a complex industry chain that covers chips, software algorithms, high-definition maps, security controls, and more.

By the end of 2017, China's per capita car ownership was about 0.156, less than 1/5 of the US, and the market has huge room for growth. The policies of countries in the world and the industrial layout of major enterprises have made intelligent driving a hot research field. At present, the auto-driving cars are generally divided into L0 to L5, L4 and L5 can be collectively referred to as "Unmanned Driving", and the current autonomous driving technology is basically at the L2-L3 level. The US IHS Automotive report predicts that by 2025, the global autonomous vehicle sales number will be close to 600,000, and in 2035 it will reach 21 million. The market will maintain a CAGR of 48% during the forecast period (2025-2035). According to the report of Sootoo, the size of China's intelligent driving market had reached RMB 68.1 billion in 2017. It is expected to reach RMB 89.3 billion in 2018, with an annual growth rate of 31.1%.

Globally, autonomous driving mainly includes mainframe vendors (BMW, GM, Audi, etc.), suppliers (Autoliv's Veoneer, Bosch, etc.), technology enterprises (Google's Waymo, Baidu's Apollo, Israel's Mobileye, etc.) and travel enterprises (Tusimple, Pony.ai, JingChi.ai, etc.). Mobileye, BMW, Volvo and Ford have all announced that they will achieve commercial application of at least L4-class self-driving cars in 2021. However, due to factors such as maturity of relevant technologies, improvement of laws and regulations, and infrastructure support, there are still many uncertainties, and some relatively closed driving scenarios with low complexity and few external interference factors are considered to be the most hopeful scenario in which the driverless car application will be first achieved. At present, the Tesla model series (L2-L3 car) and Audi a8 (L3 car) have been mass-produced. In March 2018, autowise.ai announced the trial operation of the world's first self-driving clean-up team in Shanghai. In July,Baidu announced mass-production of the world's first L4 self-driving bus "Apolong" manufactured in cooperation with Jinlong Bus.

Fig. 6-15 Typical Players in the Field of Intelligent Driving



Source: CAICT (2018)

6.2.12 Intelligent Robotics

AI technology empowers robots, enabling robots to have human-like perception, coordination, decision-making and feedback capabilities. From the perspective of application, it mainly includes intelligent industrial robots, intelligent service robots and intelligent specialized robots. The current mainstream intelligent industrial robots generally have functions of packaging, positioning, sorting, assembly and detection; intelligent service robots generally have functions such as family companion, business service, healthcare, retail sales, and rehabilitation of disabled people; intelligent specialized robots generally have functions of reconnaissance, search & rescue, fire fighting, decontamination, and demolition.

Data released by IFR show that the global robot market had reached USD 50 billion in 2017. In China, according to data from *Analysis Report on Robot Industry in China 2018*, robot ontology and system integration market exceeded RMB 120 billion, a year-on-year increase of 25.4%. Among them, the industrial, service and special robot market has reached RMB 43.5 billion, and is expected to reach RMB 71.9 billion by 2020, with a CAGR of 18.24% over the forecast period (2017-2020). The four international giants in the robotics field (ABB, Fanuc, Yaskawa and Kuka) have already occupied an absolute market share in the intelligent industrial robot industry. As an emerging industry, the intelligent service robot industry is more diversified in market demand as it directly faces the consumers, and the market competition is more regional. Home cleaning robots such as iRobot, LEGO's educational programming robots, CYBERDYNE's medical assistant robots, Ninebot's walking robots, and Roobo's commercial service robots are used in different service areas such as home, education, healthcare, travel, and business. In the field of intelligent special robots, there are reconnaissance robots such as ReconRobotics and fire-fighting robots of CITIC HIC.

Fig. 6-16 Typical Players in the Field of Intelligent Robotic

Source: CAICT (2018)

Research from Gartner

Hype Cycle for Artificial Intelligence, 2018

AI is almost a definition of hype. Yet, it is still early: New ideas will surface and some current ideas will not live up to expectations. This Hype Cycle will help CIOs and IT leaders trace essential trends and innovations to determine scope, state, value and risk in their AI plans.

Analysis

What You Need to Know

Now is the deciding time for the future of AI. Only 4% of CIOs worldwide report they have AI projects in production. Every decision about AI influences AI's long-term direction. AI gives hope and fulfills sci-fi fantasies; it is both utopian and dystopian. Anxiety about implementing AI is increasing. The term "artificial intelligence" is on gartner.com's top 10 lists for emerging searches, high-growth searches and most popular searches. Data and analytics leaders across many industries are seeking a breakthrough, which they should target in the long run. However, the immediate impact of AI is within practical applications.

The Hype Cycle

AI is overhyped as a socioeconomic phenomenon. The media, governments, corporations and individuals each have an opinion about AI, mostly based on vague ideas of what it really is. This Hype Cycle views AI as a pervasive paradigm and an umbrella term for many innovations at the different stages of value creation. The traffic jam at the Peak of Inflated Expectations is increasing, as early implementers grow in numbers, but production implementations remain scarce. A long line of high-promise innovation profiles at the Innovation Trigger phase are approaching the traffic jam at the Peak of Inflated Expectations, indicating that the AI hype will continue. None of the profiles in this Hype Cycle is Obsolete Before Plateau, but not all will not survive, and many will morph into something different — this will depend on the choices and decisions that customers are making today.

To find short-term wins and hone the long-term vision for AI, CIOs, IT leaders and AI champions should track major AI trends:

- **Conversational AI** is on many corporate agendas, spurred by the worldwide success of Amazon Alexa, Google Assistant and others represented by VPA-enabled wireless speakers at the pinnacle. To develop chatbot and voice-enabling strategies, implementers should pay attention to the time before plateau of virtual assistants, chatbots, NLG, NLP and speech recognition, among others.

- **Machine learning,** and related DNNs, ensemble learning, and predictive and prescriptive analytics, are becoming a common capability, with new tools and approaches pouring into the market. DNNs remain a focal point for implementers and scientists, but the key is to find the right problems for deep learning to solve. Meanwhile, in his Test-of-Time award talk, a renowned AI researcher, Ali Rahimi, asked, "Has machine learning become alchemy?" in his call for greater rigor within the machine learning community.

- **Compute infrastructure** drives AI progress and is being tailored for AI. It will frontier AI advancement. GPU accelerators, FPGA accelerators, deep neural network ASICs and neuromorphic hardware showcase different compute ideas, and more approaches are looming in the future. CIOs and IT leaders should balance cost and performance for use-case-driven capabilities in their compute infrastructure strategies.

- **AI is coming to the masses of application developers and software engineers,** most of whom don't even suspect yet that they will be the main AI implementation force in two to five years. Although it is early, CIOs and IT leaders should encourage developers to experiment with AI developer toolkits and AI PaaS, as well as plan developers' upskilling to get this contingent ready for its new role in AI strategies.

New Entrants

- **AI Governance:** Concerns about validity, explainability and unintended bias of AI came to the foreground this year. Many Gartner clients want to understand what it takes to govern AI even before they start AI initiatives.

- **AI Developer Toolkits:** The need for AI is massive, and data scientists are a small group compared to software developers. Vendors make a concerted effort to enable this very large group to perform basic AI development functions via familiar concepts presented in developer toolkits.

- **Knowledge Graphs:** The rising role of content and context for delivering insights with AI technologies, as well as recent knowledge graph offerings for AI applications have pulled knowledge graphs to the surface.

- **AI PaaS:** The AI PaaS hype is heating up, with the leading cloud service providers' competition using AI PaaS as a lure to their clouds and as a tool to attract developers and data scientists.

- **Chatbots:** Chatbots have increased in hype and are up for major growth over the next years, but also they are set up for a backlash once they reach the Trough of Disillusionment.

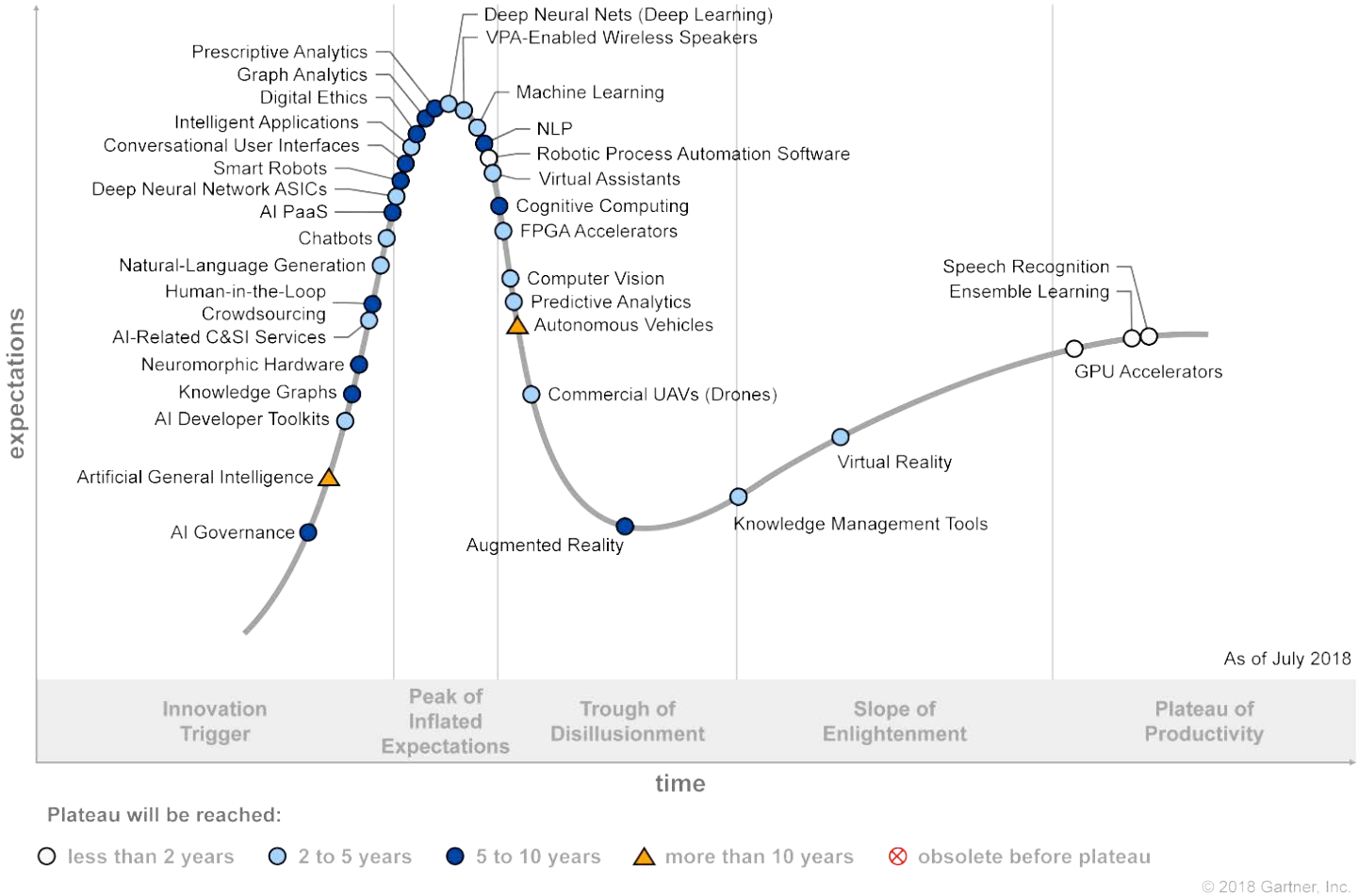
- **VPA-Enabled Wireless Speakers:** AI hype would be incomplete without Amazon Alexa, Google Assistant and the likes. Although, these are just speakers, they sound and some even look intelligent.

- **Intelligent Applications:** These applications signify a trend of embedding AI in enterprise applications, as well as encapsulating AI in domain applications.
- **RPA Software:** RPA has accelerated to the top 10 most popular searches on Gartner.com. Hype Cycle readers should be familiar with RPA capabilities and understand that AI is a small part of them.

Name Changes

- **Deep Neural Nets (Deep Learning), formerly Deep Learning:** The second name change in two years reflects the vibrant innovation, disruption, and debate around these algorithms and frameworks.

FIGURE 1
Hype Cycle for Artificial Intelligence, 2018



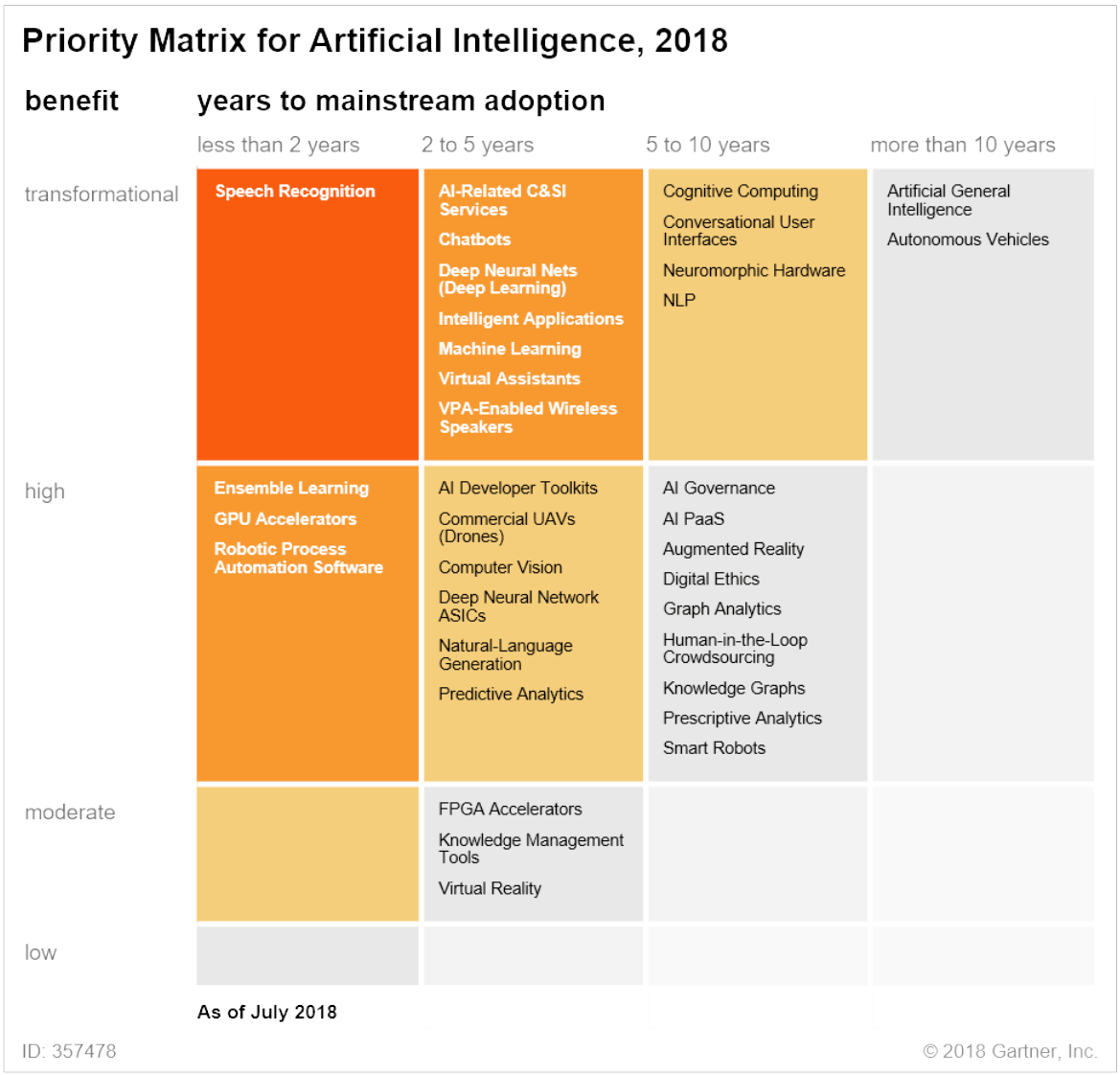
Source: Gartner (July 2018)

The Priority Matrix

Except for just three innovations, everything else on this Hype Cycle is high-impact or transformational. More than a half of the Hype Cycle entries could reach the Plateau of Productivity (20% market penetration) in less than five years. These are indeed big expectations! High-benefit innovations bring business efficiencies and require ongoing education. Transformational innovations are game changers — they call for new skills and present high risk and reward. AI techniques themselves, when applied to risk taking, make risks literally "calculated," as well as more accurate and more frequent, so every step becomes smaller and failures are not so dramatic.

The key to AI success is "narrow AI" — that is, not artificial general intelligence, but narrow use cases with defined benefits. The priorities in Figure 2 should help CIOs and IT leaders to identify profiles relevant to their plans and to stage their commitments appropriately. Innovations at the peak are already useful if approached without inflated expectations. For practical efficiency, start with the profiles approaching the Plateau of Productivity.

FIGURE 2
Priority Matrix for Artificial Intelligence, 2018



Source: Gartner (July 2018)

For competitive advantage, start with the profiles on the Innovation Trigger that will go through the Hype Cycle fast, such as chatbots, natural-language generation, AI-related C&SI services and AI developer toolkits.

It is still early for AI. Many innovations in this Hype Cycle are not yet fully understood. For example, DNNs (which started the current AI hype) are opaque, computer vision algorithms are overfitted, and autonomous vehicles cause debates about responsibilities in case something goes wrong. That is why AI governance, digital ethics and human in the loop draw the attention of more AI-advanced organizations and should be on the radar of AI strategists.

Off the Hype Cycle

This year's Hype Cycle does not include lower-level cases that are represented by higher-level concepts. These are:

- **Virtual customer assistants** (part of the virtual assistants innovation profile)
- **Cognitive expert advisors** (part of the cognitive computing innovation profile)
- **Level 3 and level 4 vehicle autonomy** (part of the autonomous vehicles innovation profile)
- **Deep reinforcement learning** (part of the deep neural networks [deep learning] innovation profile)
- **Intelligent apps** (part of the intelligent applications innovation profile)
- **Artificial intelligence for IT operations (AIOps) platforms** (part the of augmented analytics innovation profile that is included in "Hype Cycle for Analytics and Business Intelligence" and "Hype Cycle for Data Science and Machine Learning")

Other changes to this year's Hype Cycle are:

- **Bots:** Chatbots are included instead of bots to more accurately represent the hype.
- **Algorithm marketplaces:** The hype did not live up to the expectations. This profile was absorbed into API marketplaces, an innovation profile in the "Hype Cycle for Data Science and Machine Learning."
- **Learning BPO:** While AI requires continuous learning, and it will also take away routine in favor of creative tasks that require constant education, learning BPO will be one of many upskilling and training approaches. This innovation profile remains on the "Hype Cycle for Business Process Services and Outsourcing."

On the Rise

AI Governance

Analysis By: Svetlana Sicular; Frank Buytendijk

Definition: AI governance is the process of creating policies, assigning decision rights and assuring organizational accountability for risks and investment decisions for the application and use of artificial intelligence in the context of predictive models and algorithms. AI governance is part of adaptive data and analytics governance. It addresses the perceptive, predictive and probabilistic nature of AI.

Position and Adoption Speed Justification: Until recently, AI has been mostly in the domain of scientists and researchers, with little focus on the practical implications of AI upon the wider enterprise. With AI having now reached the perimeter of practical enterprise application, data and analytics leaders are beginning to raise the question of how AI governance will be conducted, before they start implementing AI. They are asking how to balance the business value promised by AI against the need for appropriate oversight, risk management and investment management. As yet, there aren't many answers, but enterprise practitioners are already making steps toward establishing AI governance. New organizations — like the AI Now Institute and Partnership on AI — are being formed to prevent AI-related biases, discrimination and other negative implications.

User Advice: AI governance is set on three cornerstones — trust, transparency and diversity. At its core, it builds upon the principles of data and analytics governance, but the fundamental difference of AI governance is in the probabilistic nature of AI, and in how AI is used to drive advanced forms of prediction. AI-based systems (often using machine learning) have emergent designs, while classical information systems have deliberate designs. Data and analytics governance focuses on trust, but analytics governance puts an emphasis on transparency. AI governance extends these concepts to trust, transparency and diversity of data, algorithms and the people in the AI teams. AI governance favors diversity to counteract bias and predictive errors.

Data and analytics leaders should:

- Ensure trust in the current data sources to avoid one-sided information.
- Demand new, different and even contradictory data to combine with what you already use to minimize risks of AI biases.
- Identify transparency requirements for data sources and algorithms.
- Promote transparency of AI solutions and communication around AI to minimize different interpretations of AI results.
- Diversify algorithms to meet the complexity of the problems that AI is solving.

- Challenge the expected outcomes. If the outcomes are fully expected, the problem you are solving is too simple.

- Create feedback loop, "guard rails" with "circuit breakers" and human oversight to prevent AI mistakes.

- Maximize the benefit from AI by establishing organizational roles and responsibilities, starting with a center of excellence that allows to share skills, resources and knowledge.

- Define governance process to evaluate business returns to either pivot to another AI project or iterate. Vary governance approaches: In some cases, no or little governance accelerate initial innovation, but this should be a conscious decision within your AI governance framework.

Business Impact: AI reflects what it learns: It can be cool, or it can be creepy (or both, unintentionally). Typically, an AI team decides what to learn, how to do it and how to ensure the best outcomes. The goal of governance is to ensure that such decisions maximize value and minimize risk. Some "black box" models are acceptable if you can prove the validity of the outputs, while some other models require transparency in order to meet regulations or preserve organizational reputation (for example, when it comes to decisions like credit, employment or housing). Data sources for AI often contain incomplete or unintentionally biased information — this is the main cause of erroneous AI outcomes. The awareness of AI risks is currently limited to those publicized by the media, such as fake news and self-driving cars, while every industry may encounter its own AI problems. Unfortunately, 95% currently neglect AI governance, unaware of the potential risks.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Artificial General Intelligence

Analysis By: Tom Austin

Definition: Artificial general intelligence (AGI) — also known as "strong AI" and "general-purpose machine intelligence" — would handle a very broad range of use cases, if it existed. It does not, though it is a popular subject of science fiction. Current AI technologies do not deliver AGI. Despite appearing to have human-like powers of learning, reasoning and adapting, they lack commonsense, intelligence, and extensive means of self-maintenance and reproduction. Special-purpose AI — "weak AI" — does exist, but only for specific, narrow use cases.

Position and Adoption Speed Justification: Tangible progress on AI has been limited to weak AI. AGI's position and adoption speed on this Hype Cycle therefore remain unchanged. (We changed this entry's name from "general-purpose machine intelligence" in 2017 to reflect the popularity of the term "AGI.")

Today's AI technology cannot be proven to possess the equivalent of human intelligence (the lack of agreement about a test to prove such intelligence is itself a problem). It may, at some point, be possible to build a machine that approximates human cognitive capabilities, but we are likely decades away from completing the necessary research and engineering.

The subject of AGI often arises in discussions of "cognitive computing" — a term that means different things to different people. For some it denotes a set of AI capabilities, for others a specialized type of hardware (as in neuromorphic or other highly parallel, short propagation path processors). It can also describe the use of information and communication technology to enhance human cognition, which is how Gartner uses the term.

User Advice: Focus on business results enabled by applications that exploit special-purpose AI technologies, both leading-edge and older.

Leading-edge AI is enabling what are currently considered "amazing innovations," including deep-learning tools and related natural-language processing capabilities. These innovations are doing what we previously thought technology could not do. They are, however, typically research tools that are only just emerging from research labs, undergoing turbulent changes in direction, and not fully understood in terms of engineering principles. Over time, we will learn their limitations and develop workable engineering guidelines. As the amazement wears off and ennui sets in, we will treat them as "aging innovations."

Look for business results enabled by applications that exploit aging innovations (including expert systems and other symbolic AI approaches, as well as simpler forms of machine learning), amazing innovations (typically more powerful but less understood technologies), or both. Examples of such applications include autonomous means of transportation, smart advisors and virtual assistants focused on various goals (such as improved wealth management) and responsibilities (such as sales or budget management). Most use both amazing and aging innovations.

Special-purpose AI will have a huge and disruptive impact on business and personal life. End-user organizations should ignore AGI, however, until researchers and advocates demonstrate significant progress. Until then, ignore any suppliers' claims that their offerings have AGI or artificial human intelligence — these are generally illusions created by programmers.

Business Impact: AGI is unlikely to emerge in the next 10 years, although research will continue. When it does finally appear, it will probably be the result of a combination of

many special-purpose AI technologies. Its benefits are likely to be enormous. But some of the economic, social and political implications will be disruptive — and probably not all positive.

There are currently no vendors of systems that exhibit AGI, but many companies are engaged in basic research. Examples are DeepMind (owned by Google), OpenAI and Vicarious.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

AI Developer Toolkits

Analysis By: Eric Hunter, Svetlana Sicular

Definition: Artificial intelligence (AI) developer toolkits are applications and software development kits (SDKs) that abstract data science platforms, frameworks and analytic libraries to enable software engineers to deliver AI-enabled applications. They cover four maturing categories: cloud-based AI as a service (AIaaS); toolkits for virtual assistants (e.g., Apple Siri, Amazon Alexa and Google Assistant), device development kits; and AI serving SDKs. Software engineers use them to incorporate AI into new or existing applications.

Position and Adoption Speed Justification: Vendors have worked aggressively to deliver developer-oriented AI toolkits and SDKs during the past 18 to 24 months. Representative offerings include:

- Cloud-based AIaaS platforms (e.g., Google AutoML, AWS SageMaker and Azure ML Studio)
- Toolkits for Virtual Assistants (e.g., Amazon Alexa Skills Kit, Apple SiriKit, Baidu DuerOS Open Platform, Google DialogFlow and Cortana Devices SDK)
- Device development kits (e.g., AWS DeepLens and Microsoft Vision AI)
- AI serving SDKs (e.g., Apple's CoreML and Google ML Kit).

Across all categories, vendor offerings require distinct deployment considerations and have varied feature coverage differences, but we expect greater consistency in the future.

Cloud-based AIaaS platforms reduce data science complexities for more developer-friendly adoption, as compared with native PaaS platforms. Model life cycle support varies widely by vendor across data preparation, feature engineering, model selection and training, hyperparameter tuning and model deployment phases.

AI developer toolkits support a limited set of native use cases, such as image recognition (including faces and landmarks), text analytics and image labeling. Developers can also deploy custom-built models and optionally update those models from cloud services at model runtime. Although Core ML and ML Kit have unique model formats, numerous conversion utilities continue to be released for models from numerous formats, including ONNX and MXNet. Commercial vendors have also introduced services (such as IBM's Watson Services for CoreML) to extend AI serving SDK support.

Device development kits position custom hardware devices (such as cameras) with developer-friendly APIs and SDKs to encourage platform developer adoption. As platform support is incorporated into broader market offerings, direct platform vendor kit offerings will diminish.

User Advice: Application development leaders must evaluate AI developer toolkits and balance their present-day benefits and capabilities. IT leaders adopting these offerings to incorporate AI capabilities and features into applications should:

- Abstract adopted vendor offerings where possible to minimize portability constraints and lock-in.
- Avoid directing disproportional investments or effort in migrating established applications to a new platform for a small set of differentiating features.
- Ensure deployed capabilities are aligned to direct end-user benefits that cannot be easily achieved without AI.
- Understand that such offerings do not enable software engineers to replace an experienced data scientist.
- Leverage established information management best practices for data management and privacy.
- Adopt offerings in alignment with larger organizational cloud and mobile development standards and strategies.

Vendor offerings are being released at a rapid pace in the market with a desire to attract new development communities. Early adopters should carefully evaluate and stress test employed offerings, along with fully understanding the going concern support for each specific function.

Business Impact: CIOs, application development leaders, and data and analytics leaders should prepare for software developers to become a key contingent for AI development and implementation. The demand for AI is significant and is increasing at a rate beyond which experienced data scientists can meet alone. Gartner notes that 60% of data science talent is concentrated in 50 cities worldwide and, in those cities, there is a finite set of employers.

Adoption of developer toolkits will continue to increase. As these offerings continue to mature, Gartner expects offerings to:

- Expand support for edge and device-centric AI models through lightweight runtime frameworks
- Mature into distinct categories in future Hype Cycles
- Increase support for higher-level, focused AI use cases across specific business verticals and consumer demands
- Continue to reduce adoption barriers in the deployment of AI capabilities for software engineers and citizen data scientists
- Increase user gravity and stickiness to broader, vendor-based cloud and platform offerings, including platform as a service (PaaS)

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Amazon; Apple; Baidu; Google; IBM; Microsoft

Knowledge Graphs

Analysis By: Stephen Emmott; Svetlana Sicular; Alexander Linden

Definition: Knowledge graphs encode information ("knowledge") as data arranged in a network ("graph") of nodes and links ("edges") rather than tables of rows and columns. Nodes hold data or their labels; edges link nodes together, representing relationships between them. This results in sequences of "triples" — i.e., node-edge-node, or Mary-manages-John — which can accommodate, throughout the graph's life cycle, multiple and varied data schemas without the need for redesign. Once encoded, information can be recalled, or synthesized, in response to queries.

Position and Adoption Speed Justification: The rising role of content and context in delivering insights through the use of AI technologies has pulled knowledge graphs to prominence Google's Knowledge Graph and Microsoft Graph are examples of the knowledge graph's growing popularity due to its promise to enrich your data with missing data. Specialist vendors are offering graph-based products to new markets and well-known vendors are accommodating the technology in their platforms and products.

Knowledge graphs are ideally suited to storing data extracted from the analysis of unstructured sources, such as documents, using natural-language processing (NLP) and related text analysis techniques. They are also capable of storing structured data, including metadata that implicitly provides structure and context. For this reason, graphs enable the storage of data, the means to structure and contextualize this by building relationships within the data, and the ability to subject the information it encodes to processing in support of varied use cases.

User Advice: Application leaders should employ knowledge graphs to connect disparate concepts and enrich their data with missing information. Using graph analysis, organic and dynamic relationships between digital assets, data sources, processes, people and interactions can be discovered and exploited automatically. A key aspect in this respect is entity extraction, whereby entities — people, events, etc. — can be identified through analysis of unstructured data prior to ingestion, and subsequent disambiguation within the knowledge graph once contextualized.

Knowledge graphs silently accrue "smart data" — i.e., data that can be easily read and "understood" by AI systems. Although available as stand-alone products from niche vendors, the knowledge graphs' benefits are typically realized through the wider platforms and applications they service. Application leaders should evaluate how vendors apply knowledge graph concepts to determine how vendor solutions could benefit their digital business platform.

For example, Microsoft and Google embed knowledge graphs in their cloud office environments — Office 365 and G Suite respectively. By capturing signals from the usage of these environments, their graphs are able to ingest data about the use of applications, enabling working relationships between employees, as well as thematic connections between digital assets to be gathered. This supports collaboration and sharing, search and discovery, and the extraction of insights through analysis. Other platforms and applications — such as text analytics and insight engines — also include the underlying graph technology upon which to build knowledge graphs and enhance functionality. In contrast, stand-alone products are graph-based applications dedicated to the management of data using a graph-based approach in support of other products — see "Magic Quadrant for Data Management Solutions for Analytics," for example.

Business Impact: Organizations can expect significant value from knowledge graphs in many areas, with the following being prominent:

- **Collaboration/sharing** — Interrelated data is contextualized data, thereby aiding its discovery and findability via implicit and indirect connections.
- **Investigation and audit** — With the capability to capture and disambiguate entities that map to entities in the real world, relationships can be explored to identify fraud, supply chain risks or patterns of collaboration.
- **Analysis/reporting** — Once structured in the form of a knowledge graph, unstructured data can be queried, thereby preprocessing it for analysis.

- **Interoperability and automation** — Autonomous reading and "understanding" of data supports integrating and operationalization of data for different enterprise applications.

- **Data reuse/cross-industry collaboration** — Being linked conceptually chunks data and metadata, which can be shared more easily and hence foster reuse.

However, it is too early to tell whether knowledge graphs will deliver on the broader promise.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Facebook; Google (Cloud Platform); Intelligent Views; Maana; Microsoft (Microsoft Graph); Mindbreeze; Neo4j; Semantic Web Company (PoolParty); TopQuadrant

Neuromorphic Hardware

Analysis By: Chirag Dekate; Martin Reynolds

Definition: Neuromorphic hardware comprises semiconductor devices conceptually inspired by neurobiological architectures. Neuromorphic processors feature non-von-Neumann architectures and implement execution models that are dramatically different from traditional processors. They are characterized by simple processing elements, but very high interconnectivity.

Position and Adoption Speed Justification: Neuromorphic systems are at the very early prototype stage. IBM has delivered a TrueNorth-based system to Lawrence Livermore National Laboratory. BrainChip's Spiking Neuron Adaptive Processor technology and Hewlett Packard Enterprise's Labs Dot Product are other early entries, Intel's "Loihi" chip tackles a broader class of AI workloads: Loihi offers a higher degree of connectivity than competing implementations. Qualcomm, an early exponent of neuromorphic processors, has shifted its focus to conventional processors.

There are three major barriers to the deployment of neuromorphic hardware:

- GPUs are more accessible and easier to program than neuromorphic silicon.

- Knowledge gaps: Programming neuromorphic hardware will require new tools and training methodologies.

- Scalability: The complexity of interconnection challenges the ability of semiconductor manufacturers to create viable neuromorphic devices.

At the moment, these projects are not on the mainstream path for deep neural networks (DNNs), but that could change with a surprise breakthrough in programming techniques.

User Advice: Neuromorphic computing architectures can deliver extreme performance for use cases such as deep neural networks because they operate at very low power and are potentially capable of faster training than the GPU-based DNN systems deployed today. Furthermore, neuromorphic architectures can enable native support for graph analytics. Most of the neuromorphic architectures today are not ready for mainstream adoption. However, these architectures will become viable over the next five years, and will deliver new opportunities. I&O leaders can prepare for neuromorphic computing architectures by:

- Creating a roadmap plan by identifying key applications that could benefit from neuromorphic computing.

- Partnering with key industry leaders in neuromorphic computing to develop proof of concept projects.

- Identifying new skillsets that need to be nurtured for successful development of neuromorphic initiatives.

Business Impact: Neuromorphic hardware faces the largest barriers in advancing DNN, but also may unlock the most powerful results. There are likely to be major leaps forward in hardware in the next decade, if not from neuromorphic hardware, then from other radically new hardware designs.

Neuromorphic systems promise lower power, but will likely operate across smaller input sets. As such, they will likely first appear in edge devices, where they will process images and sound. These devices may also execute lower levels of a DNN at the edge, reducing bandwidth and central processing constraints.

We are in the midst of an extremely rapid evolution cycle, enabled by radically new hardware designs, suddenly practical DNN algorithms and huge amounts of big data used to train these systems. Neuromorphic devices have the potential to drive the reach of DNNs further to the edge of the network, and potentially accelerate key tasks such as image and sound recognition inside the network. They will require significant advances in architecture and implementation to compete with other DNN architectures.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: BrainChip; Hewlett Packard Enterprise; IBM; Intel; Micron

AI-Related C&SI Services

Analysis By: Susan Tan

Definition: Artificial intelligence (AI)-related consulting and system integration

(C&SI) services are a subset of intelligent automation services to help clients ideate use cases, design business or IT processes, select technologies, curate data, build and train models, deploy solutions, assess and mitigate risks, and adapt talent mix to successfully incorporate intelligent solutions. Intelligent solutions must involve one or more advanced technologies, such as machine learning, deep learning and natural-language processing.

Position and Adoption Speed Justification: Organizational buyers are engaging service providers to explore the inclusion of AI in solutions. A large majority of these engagements (68% according to a survey of 24 service providers) are in ideation, exploration and proof-of-concept. To accelerate time-to-value, service providers are using rapid, phased approaches, platforms and prebuilt assets and/or pretrained models to deliver intelligent solutions.

While the market is emerging, many leading SIs are already working with their clients on intelligent solutions, often including AI with other more proven technologies. Their track record has proven success using AI to achieve targeted business outcomes such as increased productivity, increased consistency, reduction in error rates and improvement in customer retention and revenue, which should improve the confidence of other clients using such services and lead to higher adoption in the next two to three years.

However, obstacles to organizations adopting AI-related C&SI services at this stage include:

- The technology is new and some aspects — for example, security, privacy, risks, liabilities — are still unknown.
- Limited understanding of the capability, limitations and implications of AI.
- Lack of internal skills and competencies to initiate an AI program and roadmap.
- Availability of ready-to-use data for training AI, long lead time for training AI and lack of process standardization and documentation.
- Limited understanding of how to scale and integrate AI into existing systems and workflows.
- Fear of the impact of intelligent solutions on jobs and tasks.

Due to limited internal capabilities, when organizations are ready to apply AI, a high-proportion turn to service providers for consulting and implementation.

User Advice: Clients looking to engage AI-related C&SI service providers should:

- Use a "start small, achieve benefits, then scale up" approach by focusing on a narrow domain use case and the associated business outcomes where AI approaches can add value beyond traditional techniques.
- Contract a time-boxed engagement for service providers to help build a minimum viable product or automate defined tasks in a single knowledge domain to make training the AI faster and reaping the benefits quicker.
- Avoid "moonshot" projects that take years of training and validation, unless it results in the potential to disrupt an entire industry or bring disproportionate competitive advantage as a consequence. In this case, ensure you have contract terms that prevent service providers from divulging such intellectual property (IP) to competitors and ensure you have executive commitment at the highest level.
- Engage service providers to help you understand the impact of AI on your organization's processes and workforce, and take steps to mitigate risks, institute change management and apply responsible AI ethics such as avoiding bias and unintended consequences, and developing AI systems that are secure, transparent and explainable.
- Favor service providers that have invested in building AI-leveraged or pretrained solutions, such as AI-infused predictive maintenance for your industry.
- Decide what data and IP you need to protect explicitly in contracts to avoid service providers using them in their solutions for other clients.
- Ensure service providers bring the right mix of interdisciplinary consultants with relevant experience, including technical, domain and industry/process knowledge, while understanding that the newness of the technologies means few have direct AI solutioning experience.
- Get references and discuss with them how their implementation went and what were areas they did not anticipate to avoid repeating the same mistakes.

Business Impact: AI-related C&SI services can be applied to any business process or model. A recent Gartner study found organizations are deploying the following use cases:

- Predictive analytics: Providing insights, detecting anomalies, providing personalization, predicting likely events by using learning systems that use data mining and pattern recognition across large amounts of data.
- Automating tasks and replacing human judgment using a combination of AI and RPA. An example of a business process

optimization use case is the use of intelligent search which can summarize contract data — intelligent technology ingests structured and unstructured data, extracts relevant key clauses in contracts and policies that requires attention, shrinking the amount of text to be read and enables employees to concentrate their time on relevant clauses.

- Chatbots or virtual agents that use text or voice to communicate with users in natural language to scale call centers quickly and have more compelling personalized conversations, often in multiple languages.

- Products with embedded sensors and AI technologies to make them smarter so they learn about their environments.

- Mechanical equipment such as drones and robots with AI that can perform tasks in dangerous or remote environments, and/or learn from its environment and its experience to perform better.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Sample Vendors: Accenture; Atos; Deloitte; EPAM; Fujitsu; IBM; Infosys; Luxoft Holding; Mindtree; PwC

Human-in-the-Loop Crowdsourcing

Analysis By: Svetlana Sicular; Gilbert van der Heiden

Definition: Human-in-the-loop crowdsourcing is the complementary use of humans and algorithm-based automation to solve a problem or perform a task, where the human input further improves the automated AI or data management solution.

Human-in-the loop crowdsourcing has three key characteristics:

1 The ability to reach prequalified people at scale.

2 The ability to aggregate human (crowd) contributions into meaningful results.

3 Engaging contributors for a specific, mostly information-centric, task (not as full-time employees).

Position and Adoption Speed Justification: Human-in-the-loop crowdsourcing is a big shift from entrusting problem solving to the known personnel (in-house or outsourced). Currently, academia and market leaders — including Google, Facebook, Amazon, Microsoft, IBM, eBay, Baidu and many others — routinely incorporate this approach. Over the past year, adoption has been substantially accelerated, mostly driven by the machine learning requirements of data labeling and the quality of training data. For example, the growing machine learning business made CrowdFlower rebrand itself to Figure Eight. Data science problem solving is also on the rise: Kaggle reported accelerated member growth in 2018.

While the market potential is high, human-in-the-loop crowdsourcing faces many barriers to adoption — including low awareness of its benefits, and "perceived" (rather than real) concerns about quality, security and confidentiality. Adoption will grow alongside the maturity of the overall AI market, as organizations will realize that human-in-the-loop crowdsourcing is a viable (and probably the most reliable) solution to improving accuracy of machine learning models.

User Advice: Companies working on AI and machine learning should employ human-in-the-loop crowdsourcing as an enabler of AI solutions. This approach yields more-fluid costs and a wider access to problem solving, model training, classification and validation capabilities compared to internal or traditional outsourced capabilities.

Data and analytics leaders should use human-in-the-loop crowdsourcing when:

- Rules are hard to describe for automation (mostly for data collection, verification and enhancement), such as labelling images or data enrichment with data from unspecified sources. Humans can find the right information and their input can serve as a training dataset for further improvement of the algorithm.

- The problem cannot be solved efficiently by machines, for example, when a machine learning algorithm reaches the limit of its accuracy, humans can further improve the output (such as content moderation, detecting subtleties in the text or validation of information retrieval and search results).

- Tasks or projects require rare skills (for example, data science competency). Such cases usually involve competitions and access to validated experts (via Kaggle, Topcoder, Experfy, Aigency or Gigster, for example). The beauty of this approach is in the self-selection of the participants. Another example of rare skills is knowledge of a narrow market or a specific niche, like validating usability of a solution in different geographies with their cultural differences.

- Traditional solutions require too much setup, for example, a one-time job for coding a user interface for all kinds of mobile devices.

Data and analytics leaders should allow time to adjust to crowdsourcing capabilities. They should also consider the potential risks, including labor-related legal implications, IP protection and inconsistent quality.

Business Impact: The business impact of human-in-the-loop crowdsourcing ranges from modest to high:

- Modest, because in some instances, organizations have been able to achieve savings by procuring niche solutions via crowdsourcing.
- High, because it enables AI, machine learning and information quality that would not have been accomplished without this innovative approach.

A wide range of industries will find human-in-the-loop crowdsourcing indispensable. The approach will greatly benefit analytics teams interested in applying human intelligence to unstructured text, image, audio and video data for AI, machine learning and information quality, as well as those who are looking for one-time solutions or rare skills, such data science. The tasks could include conditioning of training data for algorithms, metadata extraction, proofreading, image recognition, content creation and classification, data collection, product categorization, refining product descriptions, text translation, creating photos of real estate properties, and audio transcriptions.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Alegion; Amazon Web Services; Experfy; Figure Eight; Kaggle; Playment; Topcoder; TwentyBN

Natural-Language Generation

Analysis By: Rita L. Sallam; Cindi Howson

Definition: Natural-language generation (NLG) automatically generates a natural-language description of insights in data. Within the analytics context, the narrative changes dynamically as the user interacts with data to explain key findings or the meaning of a chart or dashboard. NLG combines natural-language processing with machine learning and artificial intelligence to dynamically identify the most relevant insights and context in data (trends, relationships, correlations).

Position and Adoption Speed Justification: Whereas text analytics focuses on deriving analytic insights from textual data, NLG is used to synthesize textual content by combining analytic output with dynamic selection-driven descriptions.

Although still in the early stages of adoption, NLG is being used effectively to reduce the time and cost of conducting repeatable analysis, such as for operational and regulatory reporting, earnings reports in the financial services sector, benefits statements and weather forecasts in the government sector, and personalized messages in the advertising sector. It is also used for data products such as sports analytics (personalized "fantasy football" analysis and reports), customer communications, and marketing and research information services.

The combination of NLG with modern analytics and business intelligence (BI) — used to create analytics content including analytics applications — is one of the most promising applications improving insights derived from analytics for all users. Modern analytics and BI platforms have made significant advances in visualizing data in interactive dashboards and storyboards, and have collaboration capabilities for sharing and socializing findings. However, many users have varying degrees of analytics skill to correctly interpret and act on statistically significant relationships in visualizations. Moreover, without NLG, the annotation and presentation of findings is manual and static.

With the addition of NLG, augmented analytics platforms — for example, those of Salesforce (Einstein Discovery) and search/natural-language-query-based platforms such as ThoughtSpot and AnswerRocket — can automatically generate a written or spoken context-based narrative of findings in the data. This accompanies visualization, storyboard and batch reports to fully inform the user about what is most important and actionable. BI teams can now also integrate stand-alone NLG engines (such as those of Automated Insights, Narrative Science and Yseop) with modern analytics and BI or data science platforms to explain findings from analytics to information consumers and citizen data scientists. Narrative Science, Automated Insights, and Yseop all now offer APIs for their platforms. Qlik, Microsoft (with Power BI), Tableau, Sisense, Information Builders, and MicroStrategy have integrated with one or more of these NLG vendors. Other partnerships are emerging. Integration of NLG with analytics and BI platforms and virtual personal assistants — such as Amazon Alexa, Apple Siri, or Google Assistant for conversational analytics — will further drive adoption.

Easy configuration and multilanguage support will be necessary for broad adoption. We expect that, due to NLG's potentially beneficial impact on the expansion of analytics to a broader audience, NLG will be a feature of most modern analytics and BI platforms by 2019. We already see this happening, with most modern analytics and BI vendors offering or planning to offer NLG through integration with third-party NLG vendors or organic development.

User Advice: Data and analytics leaders should:

- Integrate NLG with existing modern analytics and BI and data science initiatives, or explore emerging augmented data discovery tools that embed NLG with automated data preparation and pattern detection.

- Assess their organization's readiness for business-user-accessible advanced analytics in terms of alignment with business outcomes.

- Monitor the NLG capabilities and roadmaps of their analytics and BI and data science platforms, as well as of startups.

- Be aware of a solution's maturity, particularly in terms of data integration and preparation requirements, the platform's self-learning capabilities, upfront set-up and configuration required, the range of languages supported, the extent of narration for a single chart or across a dashboard, and the accuracy of the findings and narration.

- Understand potential drawbacks relating to multilingual user scenarios, as NLG requires specific libraries for each language in use. Additionally, industry-specific use cases need to be considered carefully with respect to jargon, tone and specialized ontologies.

- Recognize that NLG could be attractive to government organizations that are required to have their analytics and BI solutions comply with the Americans with Disabilities Act (in the U.S.), and similar mandates in other countries.

Business Impact: NLG supports a number of productivity-enhancing use cases that reduce the need for writers (such as of financial reports, sports analysis or product recommendations) outside analytics.

The combination of NLG with automated pattern/insight detection and self-service data preparation has the potential to drive the user experience of next-generation augmented analytics platforms. It could also expand the benefits of advanced analytics to a wider audience of business users and citizen data scientists.

Context-based narration will reinforce mobile BI use cases, where a lack of screen space is a major impediment to information consumption. It will also expand the use of conversational analytics that combine NLQ, chatbots and NLG via virtual personal assistants. Moreover, it will reduce the time and cost involved in creating regular operational and regulatory batch reports.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: AnswerRocket; Arria NLG; Automated Insights; Marlabs; Narrative Science; Salesforce (BeyondCore); ThoughtSpot; Yseop

Chatbots

Analysis By: Magnus Revang; Anthony Mullen; Brian Manusama

Definition: A chatbot is a stand-alone conversational interface that uses an app, messaging platform, social network or chat solution for its conversations. Chatbots vary in sophistication, from simple, decision-tree-based marketing stunts, to implementations built on feature-rich platforms. They are always narrow in scope. A chatbot can be text- or voice-based, or a combination of both.

Position and Adoption Speed Justification: Chatbots have really increased in hype over the last couple of years. But still, only 4% of enterprises have deployed conversational interfaces, which includes chatbots. However, 38% of enterprises are planning or actively experimenting, according to the Gartner 2018 CIO Survey. This sets chatbots up for tremendous growth over the next few years, but also sets it up for a large backlash once it reaches the top of the Hype Cycle.

Chatbots in social media, service desk, HR or commerce, as enterprise software front ends, and for self-service, are all growing rapidly. Still, the vast majority of chatbots are simple, relying on scripted responses in a decision tree and relatively few intents. Related to chatbots are virtual agents, which are broader in scope and sophistication, require more infrastructure and staffing to maintain, and are designed for a longer relationship with its users outside of single interactions. Users will interact with hundreds of chatbots, but few virtual agents.

Enterprises with successful chatbot installations are already looking at the challenge of managing multiple chatbots from different vendors performing different use cases. It is likely that more enterprises will seek out platform offerings and middleware offerings as the space matures. The space is currently oversaturated with companies and offerings, the vast majority of which will not manage to keep up with the pace of innovation as alternatives to decision trees, such as fact extraction and process mapping, become more common — and voice and multimodality become more viable. Looking at the investments, attention and research by big software companies in this space, we are looking at a rapid evolution until we reach productivity in about four years.

User Advice:

- Start proofs of concept for chatbots today — the window of opportunity for experimentation is still here, but will likely close by the end of 2018. The lessons from those experimental projects will be invaluable as the technology evolves.

- Treat vendors as tactical, not strategic — acknowledge that you'll most likely want to switch vendors two to three years from now.

- Focus on vendors offering platforms that can support multiple chatbots

Business Impact: Chatbots are the face of artificial intelligence and will impact all areas where there is communication between humans today. Customer service is a huge area in which chatbots are already impacting. Indeed, it will have a great impact on the number of service agents employed by an enterprise, and how customer service itself is conducted. For chatbots as application interfaces, the change from "the user having to learn the interface" to "the chatbot is learning what the user wants" has great implications for onboarding, training, productivity and efficiency inside the workplace. To summarize, chatbots will have a transformational impact on how we interact with technology.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amazon; Facebook; Google; Gupshup; iFLYTEK; IBM; Microsoft; OneReach; Oracle; Rulai

At the Peak

AI PaaS

Analysis By: Jim Hare; Bern Elliot

Definition: Cloud artificial intelligence and machine learning platform services are known collectively as AI platform as a service (PaaS). They provide AI model building tools, APIs and associated middleware that enable the building/training, deployment and consumption of machine learning models running on prebuilt infrastructure as cloud services. These cover vision, voice and general data classification and prediction models of any type.

Position and Adoption Speed Justification: The AI PaaS hype is rapidly increasing, with the leading cloud service providers, including Amazon Web Services (AWS), Google, IBM and Microsoft, clamoring to become the platform of choice. Over the last several years, AI applications utilizing cloud services have continued to gain traction and acceptance in the market both by data scientists and developers alike. AI PaaS offerings are primarily focused on the three key areas of machine learning, natural-language processing and computer vision. The AI cloud approach is beginning to disrupt the more established on-premises data science and machine learning platform market, especially as organizations experiment and build AI prototypes. The availability of specialized hardware instances with AI-optimized chips and large amounts of data storage makes the cloud an ideal environment for organizations to build and deploy AI applications without the risks, costs and delays of conventional on-premises procurement. Cloud service providers are also offering packaged APIs and tools that make it easier for developers to integrate AI capabilities into existing applications. The promise of using cloud services to more quickly and easily build and deploy AI solutions will push AI PaaS to the Peak of Inflated Expectations. This will be followed by some level of disillusionment as organizations experience and understand the limitations of AI PaaS offerings.

User Advice: Enterprise architecture and technology innovation leaders responsible for AI-enabled applications should take these steps:

- Consider AI PaaS over on-premises options to reduce the overhead of packaging and for easier deployment and elastic scalability.

- Improve chances of success of your AI strategy by experimenting with different AI techniques and PaaS providers, using the exact same dataset, and then selecting one that best addresses your requirements.

- Increase your organization's AI project success by selecting AI cloud services that balance your data science, developer and infrastructure expertise.

Business Impact: AI PaaS offerings are focused on the three key AI portfolio services of machine learning, natural-language processing and computer vision:

- Machine learning: Packaged ML services offered by the AI cloud service providers unify the end-to-end ML workflow. They extend the capabilities of an isolated ML engine by providing integrated access to all phases of the project — from data preparation to deployment in a managed training and execution environment accessible through APIs. For technical professional teams with little to no data science expertise, features like automated algorithm selection and training-set creation will offload some of the complexity of the project and leverage existing expertise on operating cloud services.

- Natural-language processing: Organizations can use pretrained NLP systems to create cloud-based chatbots for a variety of use cases. Major AI PaaS vendors provide a language-processing catalog as part of their conversational platform that can be used to deliver applications through a natural-language interface.

- Computer vision: This enables organizations to apply facial detection, recognition and analysis to unlock new sources of image-based data. Pretrained systems require no data science expertise and allow developers to gain unique and new insight by invoking an API.

The combination of the above as cloud services will accelerate digital business technology platform viability in the short term.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Amazon Web Services; Google (Cloud AI); IBM (IBM Cloud); Microsoft (Azure AI Platform)

Deep Neural Network ASICs

Analysis By: Chirag Dekate; Martin Reynolds; Alan Priestley

Definition: A deep neural network (DNN) application-specific integrated circuit (ASIC) is a purpose-specific processor that accelerates DNN computations.

Position and Adoption Speed Justification: Deep neural networks (DNNs) are statistical models that detect and classify patterns in input data such as sound and images, or text patterns such as sentences. There are two phases in DNN systems. In the training phase, the DNN iterates across a large dataset and distills it down to a small DNN parameter set. In the inferencing phase, the DNN uses this parameter set to classify an input such as an image, speech or text. Today, a vast majority of training and inferencing tasks use GPUs. DNN ASICs can deliver significantly higher performance and lower power consumption than CPUs or GPUs when accelerating neural networks.

Google has deployed DNN ASICs (known as Tensor Processing Units [TPU, TPU2, TPU3]), at scale, providing inferencing across its businesses for, for example, speech and image recognition. The TPU2 and TPU3 also accelerate the training process, a task formerly delegated to GPUs. Google does not make the TPU2 available other than through a cloud-based service. Other cloud vendors are following suit.

Other dedicated silicon is coming. Graphcore has developed a custom processor to deliver extreme performance for DNN-based applications and plans to launch the next-generation "Colossus" processor in 2018. Their marketing materials suggest close to an order of magnitude improvement over GPUs, although performance improvements move faster than presentations. Intel is also developing an ASIC code named "Lake Crest," optimized for DNN, based on the technology it acquired from Nervana Systems in 2016.

User Advice: The benefits of DNN ASICs in performance and energy consumption are significant. However, widespread use of DNN ASICs will require the standardization of neural network architectures and support across diverse DNN frameworks. Plan an effective long-term DNN strategy comprising DNN ASICs by choosing DNN ASICs that offer or support broadest set of DNN frameworks to deliver business value faster. Compare the return on investment of a GPU-based solution against an ASIC solution, and plan to retire the GPU solution if your business will perform better with a dedicated neural network processor.

Business Impact: Hardware acceleration will enable neural-network-based systems to address more opportunities in a business, through improved cost and performance. Use cases that can benefit from DNNs include speech-to-text, image recognition and natural-language processing.

IT leaders deploying deep neural network applications should include DNN ASICs in the planning portfolio. We expect this market to mature quickly, possibly within the three-year depreciation horizon of new systems.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Adolescent

Sample Vendors: Google; Graphcore; Intel

Smart Robots

Analysis By: Annette Jump; Kanae Maita

Definition: Smart robots are electromechanical form factors that work autonomously in the physical world, learning in short-term intervals from human-supervised training and demonstrations or by their supervised experiences on the job. They sense environmental conditions and recognize and solve problems. Some can interact with humans using voice language, while some have a specialized functional form, like warehouse robots. Others have general forms and/or humanoid appearances. Due to advanced sensory capabilities, smart robots may work alongside humans.

Position and Adoption Speed Justification: Smart robots have had significantly less adoption to date compared with their industrial counterparts (predefined, unchanged task) — but they received great hype in the marketplace, which is why smart robots are positioned climbing the Peak of Inflated Expectations. In the last 12 months, we have seen some of the established robot providers expanding their product line and new companies entering the market (particularly from China). Therefore, the market is becoming more dynamic, opening to new technology providers and technologies, and the barriers for entry are slightly dropping.

Hype and expectations will continue to build around smart robots during the next few years, as providers execute on their plans to

expand their offerings and deliver solutions across the wider spectrum of industry-specific use cases and enterprise sizes. Hype is quickly building for smart robots as a result of several key vendors' actions during the past few years:

- Amazon Robotics (formerly Kiva Systems) deployed robots across Amazon warehouses.
- Google has acquired multiple robotics technology companies.
- Rethink Robotics launched Baxter and Sawyer, which can work alongside human employees.
- SoftBank Robotics introduced the humanoid Pepper and created the Pepper for Business Edition.
- In early 2018, LG introduced CLOi, a series of robots developed for commercial use in hotels, airports and supermarkets. Also, various hotels in the U.S. and two Shangri-La hotels in Singapore now use robots for delivering room service.

User Advice: Users in light manufacturing, distribution, retail, hospitality and healthcare facilities should consider smart robots as both substitutes and complements to their human workforce. Begin pilots designed to assess product capability, and quantify benefits. Examine current business- and material-handling processes into which smart robots can be deployed; also, consider redesigning processes to take advantage of the benefits of smart robots with three- to five-year roadmaps for large-scale deployment. Smart robots could also be a quality control (QC) check at the end of the process, rejecting product with faults and collecting data for analysis.

Business Impact: Smart robots will make their initial business impact across a wide spectrum of asset-centric, product-centric and service-centric industries. Their ability to do physical work, with greater reliability, lower costs, increased safety and higher productivity, is common across these industries. The ability for organizations to assist, replace or redeploy their human workers in more value-adding activities creates potentially high — and occasionally transformational — business benefits. Typical and potential use cases include:

- Medical materials handling
- Disposal of hazardous wastes
- Prescription filling and delivery
- Patient care
- Direct materials handling
- Stock replenishment
- Product assembly
- Finished goods movements
- Product pick and pack
- E-commerce order fulfillment
- Package delivery
- Shopping assistance
- Customer care
- Concierge

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Aethon; Amazon Robotics; ARxIUM; Google; iRobot; Panasonic; Rethink Robotics; Savioke; SoftBank Robotics; Symbotic

Conversational User Interfaces

Analysis By: Magnus Revang; Van L. Baker

Definition: Conversational UI (CUI) is a high-level design model in which user and machine interactions primarily occur in the user's spoken or written natural language. Typically informal and bidirectional, these interactions range from simple utterances through to highly complex interactions, with subsequent highly complex results. As design models, CUI depends on implementation via applications or related services or on a conversational platform.

Position and Adoption Speed Justification: CUIs have seen an explosive growth in interest over the last couple of years, with chatbots, messaging platforms and virtual assistants, especially home speakers such as Amazon Echo and Google Home, all contributing to the increased hype. Still, only 4% of enterprises have a conversational interface solution in production, while a further 38% is experimenting or planning, according to Gartner CIO Survey 2018. Expected growth will be greatly fueled by enterprises entering production from those planning and experimentation phases.

The promise of CUIs is a dramatic shift in responsibility between the user and the interface — where the responsibility shifts from the user having to learn the software, to the interface learning what the user wants. This promise warrants a transformational impact — even if current CUIs are far from living up to this promise.

Since 2017, there has been an explosion in the availability of conversational platforms used to implement CUI. These tools have made it a lot easier for developers to build CUIs. We have, as a consequence, also seen CUIs being implemented inside popular applications as an alternative to GUI, and even in application suites. We expect application suite vendors to bring to market CUIs in front of their business

applications — which can quickly lead to hundreds of different chat interfaces being available to employees of a large enterprise — on multiple messaging platforms. The emerging pattern of chatbots acting as a guide or concierge in front of these conversational interfaces will likely gain a lot of traction over the next year.

Most CUI implementations are still primitive, and thus are not able to respond to complex queries. Increases in capabilities will, at first, largely come from improvements in natural-language understanding (NLU) and speech recognition, which will bring CUIs closer to the promise and hype. Additional capabilities around context handling, user identification and intent handling will likely arrive within the next year, but will still not be good enough to avoid a disillusionment phase in two-to-three years' time. Only at that point will we see a standardization around design methodologies for creating flows and personality in interactive conversations.

User Advice: CUIs shift the responsibility for learning from the user to the software, so the software learns what the user wants. The impact on training, onboarding and expansion of use cases is profound. The need for literacy-related training and tools will thus significantly diminish during the next decade. Plan on CUIs becoming the dominant model.

Be wary, however, of committing to CUIs too deeply. Conversational interfaces can make machines smarter and improve the ability of people to handle novel situations (people and machines collaborating will be better than either working alone), but they also carry an extra burden. For well-developed, repetitive skills that can be performed almost effortlessly, injecting conversation can degrade performance — unless the technology is able to recognize the repetitive patterns and invoke many steps of a routine process with a single, user-generated command.

Avoid retrofitting CUI front ends to existing applications unless this improves usability and user delight.

Prepare for new roles in the enterprise. Dialogue designer, AI trainer, digital coach, humanizer and AI interaction designer are all titles Gartner is seeing in the market to support the creation of conversational experiences.

Business Impact: CUIs are the interaction pattern of many chatbots and virtual assistants — both will be significant contributors to the impact of CUIs, especially in high-touch communicative fields of customer service and Q&A-type interactions with significant volume.

Outside of this, CUIs will appear primarily in new applications. Enterprise IT leaders should be on the lookout for (and biased toward) CUIs to improve employee (and customer) effectiveness, as well as to cut operating expenses and time spent learning arcane computer semantics.

There will also be some retrofitting. Over the next five years, we do not expect large enterprises to invest heavily in retrofitting existing systems of record where the employee base is experienced and stable, and the feature set well-known to the user base. However, where there is high employee turnover or significant rapid changes in feature sets, or where enterprises face a continuing burden of providing computer literacy training, IT leaders need to consider creating people-literate front ends to make it easier for employees to adapt and excel.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Amazon; Baidu; Facebook; Google; IBM; IPsoft; Microsoft; Oracle; Salesforce; SAP

Intelligent Applications

Analysis By: Jim Hare; Helen Poitevin

Definition: Intelligent applications are enterprise applications with embedded or integrated AI technologies to support or replace manual human-based activities with intelligent automation and improved decision making.

Position and Adoption Speed Justification: AI has become the next major battleground. Every application and service will incorporate AI at some level over the next few years. Enterprise application vendors are beginning to embed AI technologies within their offerings as well as introducing AI platform capabilities — from ERP to CRM to HCM to workforce productivity applications. AI has the potential to be organizationally transformational and is at the core of digital business. Customer-facing and back-office enterprise applications are a vital component of that transformation effort because they provide the digital foundation upon which most of the endeavors rest. AI will run unobtrusively in the background of many familiar application categories while giving rise to entirely new ones.

There is an AI "land grab" from both large vendors making "big bets" and from startups seeking to gain an edge. They all aim to support or replace manual human-based activities with intelligent automation. For example, the main enterprise software vendors are emphasizing sales, service, marketing, human resources and ERP as particularly valuable areas for applying AI techniques.

Intelligent applications will use AI in the following ways:

- Analytics: AI can be used to create more predictive and prescriptive analytics that can then be presented to users for further evaluation, or plugged into a process to drive autonomous action. AI is also being used for augmented analytics.

- Process: AI can drive more intelligent actions by an application. For example, you can use AI for intelligent invoice matching or analysis of email documents to improve service flow. In the future, this can be extended further to identify patterns of work, from which process models can be built and executed.

- User Experience: Natural-language processing used to create VPAs is one application of AI to the user experience. Further examples include facial recognition and other AI applications for understanding user emotions, context or intent, and predicting user needs.

User Advice: Enterprise application leaders should:

- Explore how AI can alter your organization's processes and operations by adding more intelligent automation, dynamic workflows, and guided decision making.

- Challenge your packaged software providers to outline how they'll be using AI to add business value in new versions in the form of advanced analytics, intelligent processes and advanced user experiences.

- Acclimatize employees to the idea of automation and "bots" by deploying robotic process automation (RPA).

- Develop a deeper understanding of and expertise in AI by piloting initiatives during the next two years in areas where significant opportunity exists to mine data quickly and efficiently to uncover underlying insights. Build upon these early successes and apply the lessons learned from any failures.

- Be aware of "AI washing" as more and more startups, and even aging solutions, claim AI as part of their solution. Ask them how they use AI to deliver advanced analytics, intelligent processes and new user experiences.

- Prioritize investments in highly specialized and domain-specific intelligent applications delivered as individual point solutions. They can help solve problems in domain areas such as customer service, talent acquisition, collaboration, engagement and more.

Business Impact: Intelligent enterprise applications that leverage AI can offer the following benefits:

- Reshape how tasks are performed by allowing workers to focus on more value-adding activities through the use of automation — via bots, sensors and machine learning.

- Deliver business efficiencies at scale via packaged AI technologies available in enterprise applications.

- Enable organizations to derive better performance outcomes from their assets.

For example, in the area of Human Capital Management (HCM), AI is increasingly being added to HCM applications to match talent supply and demand, predict recruitment success, or optimize recruitment marketing. Predictors include the fit of a particular candidate for a job, the likelihood that a candidate would be open to exploring a new job opportunity, and behavioral profiles through analysis of voice or video interviews. Candidate-facing chatbots are becoming increasingly common in enabling further automation of this process, such as recommending which jobs to apply for and answering questions or conducting initial candidate screening.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Google Docs; Microsoft Office 365; Oracle Applications; Salesforce Einstein; SAP Leonardo; ServiceNow; Workday

Digital Ethics

Analysis By: Jim Hare; Frank Buytendijk; Lydia Clougherty Jones

Definition: Digital ethics comprises the systems of values and moral principles for the conduct of electronic interactions, and the use and sharing of data between people, businesses, governments and things. The scope of digital ethics is broad and includes security, cybercrime, privacy, social interaction, governance, free will, and society and economy at large.

Position and Adoption Speed Justification: Digital ethics jumped several positions toward the Peak of Inflated Expectations due to the recent wake of well-publicized negative press, rising public discourse, and new regulatory compliance including data privacy considerations. Current themes such as "artificial intelligence," "fake news" and "digital society" are triggers driving the increased need for digital ethics. Innovations such as the Internet of Things, 3D printing, cloud, mobile, social and AI are moving faster than business, governments and society can organize around it or even comprehend. The probability that unintended consequences will occur is high as the use of technology creates distance between morals and actions. For business and the technologies used in business, a morally

agnostic stance is a position that simply cannot and should not be sustained. Digital ethics require societal, economic, political and strategic debate, new types of governance, and new processes and technologies to control new technologies.

User Advice: Privacy rules and data protection provide a legal minimum in handling data that is insufficient. Instead, take a "care ethics" approach to the application of digital technologies in the business world to reconcile principles and consequences. The core question of care ethics is, "How do we take responsibility for the consequences of our actions, even if they are unintended?" (see "ethics of care" on Wikipedia). In the digital world, the concept of care ethics is not only about people, but also about how businesses and even technologies act. Care ethics teaches that ethics is about taking responsibility when confronted with situations you feel are not OK. Apply "care" ethics by following these call to actions:

- Be empathetic — put yourself in the other person's shoes; develop a sense of right and wrong that goes past just being afraid of punishment or hoping to generate a product sale whether legally or in terms of customer loyalty.
- Take responsibility — taking responsibility is essential for taking the lead within your ecosystem, and being the interface to the customer or citizen. In emerging digital environments, taking responsibility over the use of digital technologies, even if legally not required, builds and improves trust.
- Display competence — build the capacity and expertise to be able to quickly and adequately address problems. Don't simply acknowledge the need to care and accept the responsibility; you also need to be able to follow through.
- Promote trust — trust is needed to make the other three calls to action work. It is great to take responsibility, but if your stakeholders do not trust you to do so, your offer will not be accepted.

Business Impact: The four areas of business impact, listed in increasing order of "moral development" are:

- **Submitting to compliance** — staying within the boundaries of the law.
- **Mitigating risk** — being mindful of not using technology in ways that can upset stakeholders, or cause reputational or financial risk in other ways.
- **Making a difference** — making ethical use of technology as a proposition that sets you apart in the market. For example, this could be in terms of data for good initiatives or social purpose.
- **Follow your values** — there is a direct correlation between the use of technology and delivering value to customers, other stakeholders and yourself.

Actively engage and participate in online data ethics and data for good initiatives such as DataEthics and DataKind.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Graph Analytics

Analysis By: Mark A. Beyer; Rita L. Sallam; Alexander Linden

Definition: Graph analytics is a set of analytic techniques that allows for the exploration of relationships between entities of interest such as organizations, people and transactions. Graph analytics are typically portrayed via a visualization for business user consumption. Graph analytics consists of models that determine the "connectedness" across data points to create data nodes/clusters and their demarcation points. Nodes are connected explicitly or implicitly, indicating the levels of influence, frequency of interaction, or probability.

Position and Adoption Speed Justification: Graph analytics are steadily climbing to the Peak of Inflated Expectations, primarily due to a lack of broad awareness. The growing adoption of graph analytics is driven largely by the need to find insights across an exponentially larger amount of heterogeneous data, and the demand to analyze it. Once highly complex models are developed and trained, the output is more easily stored because of the expanded capabilities, computational power and adoption of graph databases, which present an ideal framework for storing, manipulating and analyzing graphs.

Analytics experts are beginning to claim specialization in graph analytics, and some traditional analytics vendors are offering capabilities that enable users to build interactive network graphs — as additional features of their products. Importantly, the utilization of graph analytics is necessary in order to develop knowledge graphs — a highly useful output of graph analytics. Commercialization of graph analytics is still at quite an early stage, with a small number of emerging players. However, the unique method of storing and processing data within many graph databases, combined with the need for new skills related to graph-specific knowledge, may limit growth in adoption. For example, knowledge and experience with the Resource Description Framework (RDF), SPARQL Protocol and RDF Query Language (SPARQL), and emerging languages such as Apache TinkerPop or the recently open-sourced Cypher.

User Advice: Data and analytics leaders should evaluate opportunities to incorporate graph analytics into their analytics portfolio and strategy. This will enable them to address the high-value use cases less-suited to traditional SQL-based queries and visualizations (such as computing and visualizing the shortest path, or the relationship between and influence of two nodes or entities of interest in a network). They should also consider using graph analytics to enhance pattern analysis.

The user can interact directly with the graph elements to find insights, and the analytic results and output can also be stored for repeated use in a graph database.

Business situations in which graph analytics constitute an ideal framework for analysis and presentation include:

- Route optimization
- Market basket analysis
- Fraud detection
- Social network analysis
- CRM optimization
- Location intelligence
- Supply chain monitoring
- Load balancing
- Special forms of workforce analytics, such as enterprise social graphs and digital workplace graphs
- Recency, frequency, monetary (RFM) analysis of related networks of objects, assets and conditions

There are also more specialized applications, such as:

- Law enforcement investigation
- Epidemiology
- Genome research
- Detection of money laundering

Business Impact: Graph analytics is highly effective at both assessing risk and responding to it to analyze fraud, route optimization, clustering, outlier detection, Markov chains, discrete-event simulation and more. The engines used to expose fraud and corruption can also be used to identify it within the organization and answer issues of liability in a proactive manner. A most recent example of identifying networks of relationships was the International Consortium of Investigative Journalists (ICIJ) research revealing the Panama Papers. Relational analytics is typically ideal for structured, static data in columns and rows in tables. Graph analytics, by contrast, is a new kind of "lens" for exploring fluid and indirect relationships between entities across multistructured data. It can deliver the kind of insight that is difficult to reach with SQL-based relational analytics.

Graph analytics processing is a core technology underlying many other advanced technologies, such as virtual personal assistants, smart advisors and other smart machines. Platforms such as those of Cambridge Semantics, Digital Reasoning, Ayasdi and Maana also use graph analytics to identify important findings.

Graph analytics can extend the potential value of the data discovery capabilities in modern business intelligence and analytics platforms. Once a graph process is completed, it can be visualized — using size, color, shape and direction — to represent relationship and node attributes.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Ayasdi; Cambridge Semantics; Centrifuge Systems; Databricks; Digital Reasoning; Emcien; Intel (Saffron); Maana; Palantir; SynerScope

Prescriptive Analytics

Analysis By: Peter Krensky; Carlie J. Idoine

Definition: The term "prescriptive analytics" describes a set of analytical capabilities that specify a preferred course of action to meet a predefined objective. The most common examples of prescriptive analytics are optimization methods (such as linear programming), a combination of predictive analytics and rules, heuristics, and decision analysis methods (such as influence diagrams). Prescriptive analytics differs from descriptive, diagnostic and predictive analytics in that the output is a recommended (and sometimes automated) action.

Position and Adoption Speed Justification: Although the concepts of optimization and decision analysis have existed for decades, they are now re-emerging along with a greater awareness of and experience with data science, better algorithms, cost-effective cloud-based computing power and available data. In addition, the focus on the business prioritization of providing actionable, proactive insight — as opposed to the more traditional reactive reporting — has further fueled the resurgence. Prescriptive analytics augments a user's decision making by recommending a course of action to achieve a defined objective. Some use cases are very mature. These include optimization in supply chain and logistics, or combining predictive scores with business rules for credit and lending decisions, database marketing and churn management. Many new use cases continue to emerge. It is, therefore, still early days for broad adoption and awareness, but many more organizations today are expressing an interest in prescriptive techniques.

Because prescriptive analytics often leverages and extends predictive methods, its adoption tends to be higher among companies that have already built predictive capabilities. Although it is a necessary competence, prescriptive analytics does not automatically result in better decision making. With improvement in analytics solutions, data quality, skills and broader use of predictive analytics, prescriptive analytics will continue to advance, reaching the Plateau of Productivity in five to 10 years.

User Advice: Data and analytics leaders should:

- Start with a business problem or decision where there are complicated trade-offs to be made, multiple considerations and multiple objectives.
- Look for packaged applications that provide specific vertical or functional solutions, and service providers with the necessary skills.
- Understand the breadth of prescriptive analytics approaches and decision models available, and which best cater to the nature of your specific business problems and skills.
- Gain buy-in and willingness from stakeholders — ranging from senior executives to front-line workers carrying out the recommended actions — to rely on analytic recommendations.
- Ensure that your organizational structure and governance will enable the company to implement and maintain functional, as well as cross-functional, prescriptive analytics recommendations.

Business Impact: Prescriptive analytics can be applied to strategic, tactical and operational decisions to reduce risk, maximize profits, minimize costs, or more efficiently allocate scarce or competing resources. Importantly, prescriptive analytics can be deployed to improve performance because it recommends a course of action that best manages the trade-offs among conflicting constraints and goals. Significant business benefits are common and are obtained by improving the quality of decisions, reducing costs or risks, and increasing efficiency or profits.

Common use cases include customer treatments, loan approvals, claims triage, and many kinds of optimization problems such as supply chain or network optimization and scheduling. Prescriptive analytics can also be a business differentiator for planning processes, whether it be financial or production or distribution planning, allowing users to explore multiple scenarios and compare recommended courses of action.

Although prescriptive analytics has been traditionally relegated to strategic and tactical time horizons, more advanced capabilities can support real-time or near-real-time decision making. This can support automation of some operational decisions.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Sample Vendors: AIMMS; Decision Lens; FICO; Gurobi Optimization; IBM; River Logic; SAS; Sparkling Logic; Veriluma

Deep Neural Nets (Deep Learning)

Analysis By: Alexander Linden; Chirag Dekate

Definition: Deep neural nets (DNNs) are large-scale neural networks, often with many processing layers. They underpin most recent advances in artificial intelligence (AI) by enabling computers to process much more complex data than before, such as video, image, speech and textual data.

Position and Adoption Speed Justification: The internet giants deploy systems based on DNNs across their businesses. Examples of well-developed DNN systems underpin Amazon Alexa's speech-to-text capability, Google's search capability, image recognition and self-driving cars, and Facebook's face-tagging technology.

DNNs are, however, tricky to build and train. To achieve consistently good results, you need large quantities of labeled data, data science expertise and special-purpose hardware. Most enterprises struggle to obtain enough labeled data to support their DNN initiatives. Furthermore, data science experts are scarce, as the IT and internet giants have hired aggressively. Additionally, optimized computational resources for DNNs require a great deal of capital expenditure.

The most widely implemented DNNs are convolutional neural networks (CNNs) and recurrent neural network (RNNs). CNNs are used, for example, for image classification and speech to text. RNNs are good for, among other things, extracting meaning from snippets of speech. Additionally, hyperscalers are developing generative adversarial networks (GANs), a technology that is most useful in gameplay situations, but that will no doubt be pressed into service for business applications.

The level of hype about DNNs is not very different from last year.

User Advice: Data and analytics leaders of modernization initiatives should:

- Explore DNNs: These technologies could help them solve previously intractable classification problems, especially relating to images, video and speech.
- Start with tools from cloud providers: Wherever possible, begin by using tools available from the major cloud providers.

They have enormous resources invested in image, speech and facial classification systems, and in training and data. Their systems will likely outperform almost anything you build and deploy yourself.

- Develop and acquire skills: Improve your machine learning experts' skills through training. Engage with academic teams. Use crowdsourcing providers like Algorithmia, Experfy, Kaggle and TunedIT. Although it's currently difficult to compete with the big cloud companies, there is a good stream of graduates skilled in this area, and talent will become easier to acquire.

- Focus on data for deep learning as a long-term investment: DNNs are within your field of competency, and the value of the right data will grow over time. Don't assume that DNNs will derive insights from any type of data through unsupervised learning. So far, results have mostly been achieved using supervised learning.

Business Impact: DNNs have transformational, and therefore disruptive, potential for all industries. The challenge for those wanting to realize this potential is to identify the business problems to solve, and to secure availability of enough experts and reasonably good datasets. DNNs demonstrate superior accuracy to past state-of-the art algorithms in detecting fraud, determining quality, predicting demand and other classification problems that involve sequences (using, for example, video, audio or time series analysis).

The basis of a DNN's potential is its ability to produce granular representations of highly dimensional and complex data. A DNN can, for example, give promising results when interpreting medical images in order to diagnose cancer early; help improve the sight of visually impaired people; enable self-driving vehicles; colorize black-and-white photographs; add missing elements to photographs; and recognize and understand speech (which, in time, may make most devices conversational devices).

Completely new product categories are likely in fields such as personal assistance and surveillance.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amazon; Baidu; deepsense.ai; Google; H2O.ai; Intel; Microsoft; NVIDIA; Skymind

VPA-Enabled Wireless Speakers

Analysis By: Fernando Elizalde; Werner Goertz; Annette Jump

Definition: Gartner defines VPA-enabled wireless speakers as cloud-enabled, far-field voice capturing devices connecting the user to a virtual personal assistant (VPA) service such as Alexa, Google Assistant, Siri, Cortana, WeChat and others. With the advent of screen-enabled VPA speakers in 2017, multimodal interactions were brought to the VPA experience. Our definition refers to stand-alone products and does not include implementations of VPA into products such as lighting fixtures, home appliances or cars.

Position and Adoption Speed Justification: Even though the conversational experience delivered by VPAs is still far from perfect, the higher-than-expected consumer adoption of VPA speakers justifies upgrading the position at the pinnacle of the Hype Cycle this year. Microsoft's Cortana is now enabled in Harman Kardon's device and Apple began shipping its HomePod product, adding to forerunners Amazon Echo and Google Home. Originally marketed as consumer and connected home products, as a source of general information, manage connected home devices and play digital entertainment, new use cases in business environments, such as hospitality services have been developed. The evolution of VPAs into domain experts (such as digital doctors, lawyers and insurance agents) will open additional use cases and digital business models. At the end of 2017, the Alexa for Business announcement launched VPA speakers' entry into the managed enterprise environment.

User Advice: Technology service providers should capitalize on the current hype around VPA speakers by voice-enabling their devices and services, and join one or more of the ecosystems. New recurring revenue streams (for example, from licensing digital content or in-skill purchases) can be monetized, and consumer relationships and affinities are there to be exploited. Develop new interactive, 24-hour user experiences, piggybacking on the further proliferation (and international expansion) of hands-free, far-field voice and video. Recognize the communal and ambient nature of this device category, and deliver use value that differentiates between individual users and respects privacy and confidentiality, paying particular attention to confidentiality in the enterprise context. Improve consumer experience in relevant use cases such as home shopping by adding a camera and display to aid visualization. Continue to improve the user experience beyond early adopters, for example, language support and accent recognition, relevance, setup, troubleshooting and address privacy concerns. During the past year, many skills and use cases have emerged that leverage emotion detection and emotive analysis. Understanding your clients' emotional states can be leveraged for improved marketing.

Business Impact: The business impact of VPA speakers and the ecosystems built around them are truly transformational. Through cloud integration with connected home solutions, VPA speakers have the potential to become the focal center of control for all smart devices in the home.

VPA speakers will have an impact on e-commerce experiences, especially with repetitive purchases. The ability to cast shopping request images to a connected screen or the addition of screens onto the device itself facilitates the shopping experience and reduces transactional friction. The quality of the built-in speakers of these devices, and their ability to stream music from online services and Bluetooth players can easily displace the appeal for wireless music systems.

The advent of VPA speakers affects and accelerates digital business opportunities in enterprise segments, such as the following:

- Customer portals and interactive voice response (IVR) system replacements
- Home healthcare and, especially, elder care
- Multifactor authentication (MFA) for building access, asset management and homeland security applications

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Ainemo; Amazon; Apple; Centralite Systems; Google; Lenovo; LG; Sony

Machine Learning

Analysis By: Shubhangi Vashisth; Alexander Linden; Carlie J. Idoine

Definition: Machine learning is a technical discipline that aims to solve business problems utilizing mathematical models that can extract knowledge and pattern from data. There are three major subdisciplines that relate to the types of observation provided: supervised learning, where observations contain input/output pairs (also known as "labeled data"); unsupervised learning (where labels are omitted); and reinforcement learning (where evaluations are given of how good or bad a situation is).

Position and Adoption Speed Justification: Machine learning is still one of the hottest concepts in technology, given its extensive range of effects on business. The drivers of its continued massive growth and adoption are the growing volume of data and the complexities that conventional engineering approaches are unable to handle. An increasing number of organizations are exploring use cases for machine learning and many are already in the initial phases of pilot/POC. Tech providers are adding embedded machine learning capabilities into their software. Despite the heightened interest in the technology, most organizations are still dabbling in their approaches to machine learning. Finding relevant roles and skills needed to implement machine learning projects is a challenge for such organizations. As the volume and sources of data increase, the complexity of systems will also grow and, in such scenarios, traditional software engineering approaches would produce inferior results. In the future, advances in many industries will be impossible without machine learning.

User Advice: For data and analytics leaders:

- Start with simple business problems for which there is consensus about the expected outcomes, and gradually move toward complex business scenarios.
- Utilize packaged applications, if you find one that suits your use case requirements. These often provide superb cost-time-risk trade-offs and significantly lower the skills barrier.
- Nurture the required talent for machine learning, and partner with universities and thought leaders to keep up to date with the rapid pace of advances in data science. Create an environment conducive to continuous education, and set explicit expectations that this is a learning process and mistakes will be made.
- Track what initiatives you already have underway that have a strong machine learning component — for example, customer scoring, database marketing, churn management, quality control and predictive maintenance — to accelerate machine learning maturation through cross-pollination of best practices. Monitor what other machine learning initiatives you could be a part of and what your peers are doing. The choice of machine learning algorithms is also influenced by the ability to explain how the algorithm arrived at a certain outcome.
- Assemble a (virtual) team that prioritizes machine learning use cases, and establish a governance process to progress the most valuable use cases through to production.
- Focus on data as the fuel for machine learning by adjusting your data management and information governance for machine learning. Data is your unique competitive differentiator and high data quality is critical for success of machine learning initiatives. Although the choice of fundamental machine learning algorithms is fairly limited, the number of algorithm variations and available data sources are vast.

Business Impact: Machine learning drives improvements and new solutions to business problems across a vast array of business, consumer and social scenarios:

- Automation
- Drug research
- Customer engagement
- Supply chain optimization

- Predictive maintenance
- Operational effectiveness
- Workforce effectiveness
- Fraud detection
- Resource optimization

Machine learning impacts can be explicit or implicit. Explicit impacts result from machine learning initiatives. Implicit impacts result from products and solutions that you use without realizing they contain machine learning.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Alteryx; Amazon Web Services; Domino Data Lab; Google Cloud Platform; H2O.ai; IBM (SPSS); KNIME; Microsoft (Azure Machine Learning); RapidMiner; SAS

NLP

Analysis By: Bern Elliot; Erick Brethenoux

Definition: Natural-language processing (NLP) provides an intuitive form of communication between humans and systems, i.e., NLP includes computational linguistic techniques aimed at parsing, interpreting (and sometimes generating) human languages. NLP techniques deal with the pragmatics (contextual), semantics (meanings), grammatical (syntax) and lexical (words) aspects of natural languages. The phonetic part is often left to speech-processing technologies that are essentially signal-processing systems.

Position and Adoption Speed Justification: Enterprise NLP usage is increasing as capabilities improve, along with new use cases based on conversational agents and automatic machine translation, among others. Existing syntactic- and semantic-based methods are increasingly augmented and displaced with deep neural networks (DNNs) approaches.

Visible accomplishments include technologies that:

- Improve natural-language parsing (via Google's SyntaxNet, an open-source, DNN-based, natural-language parsing framework for TensorFlow).
- Translate in real time from one spoken language to another (as in Microsoft's Skype Translator).
- Build large-scale knowledge graphs (illustrated by the work of Google, IBM and Microsoft).
- Offer answers instead of a list of page links (as in Google's information cards).

However, human language is complex; and while NLP solutions have made progress, there are many subtleties and nuances that require human intervention to enable proper interpretation. These limitations are slowing adoption. For instance, dialogue capabilities are weak, DNNs are experimental and fragile, and understanding, inferences, context and synthesis pose significant challenges. Additionally, many NLP solutions require specialists in order to ensure continued accuracy of the grammars and models.

User Advice: NLP offers enterprises significant opportunities to improve operations and services. For many enterprises, the strongest and most immediate use cases for NLP are related to improved customer service (impacting cost, service levels, customer satisfaction and upselling) and employee support (including making them smarter and more effective in their work).

Initial projects should start with modest goals in order to demonstrate success. As experience is obtained, projects should iterate and scope can increase.

Additional current NLP opportunities exist for enterprises but are not as mature, or will require effort before they provide consistent returns on investment. Translation or transcription services, for instance for meetings or documents, offer opportunities to improve operations and lower costs. However, these NLP-based solutions are less accurate than similar human-based options and may benefit in some cases from human involvement.

As enterprises enhance their NLP implementations, new skills should be explored. Computational linguists, for example, are versed in the manipulation of various linguistic techniques and the impact of natural communications on users. Upskilling of data scientist talents might also be necessary given the increasing use of data science techniques in NLP applications.

Finally, the quality of NLP solutions offering knowledge-based consolidation, content mapping, search enhancements and text summarization will vary. As a result, enterprise planners should test and verify the effectiveness of these solutions before making significant commitments. If enterprises invest in specialized grammars, care should be taken that these be compatible across vendor solutions.

Business Impact: To obtain clear near-term ROI and to build enterprise knowledge and skills in the area of NLP, planners should leverage NLP applications such as:

- Virtual assistants and chatbots to improve interactions, including employee and customer services in select environments.

- Text analytics to extract and summarize the focus of textual reports and preview what questions are most common before building chatbots.
- Basic transcription and translation services.
- Language-generation applications that produce natural-language descriptions of tabular data, making it easier for many to understand.
- Keyword tagging in documents, making it easier to determine relevant sections or to extract other information such as intent and entities.
- Content moderation services that examine user-generated content (text or images), to flag potentially offensive content or to identify fake news in social media.
- Sentiment analysis to identify the feeling, opinion expressed in statements — from negative to neutral, to positive.
- Search improvements by better understanding the intent of a search query as well as by summarizing content that is retrieved.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Bitext; Clarabridge; CognitiveScale; Digital Reasoning; Google; IBM Watson; Microsoft; Narrative Science; SAS; Yseop

Robotic Process Automation Software

Analysis By: Cathy Tornbohm

Definition: Robotic process automation (RPA) is a combination of user interface recognition technologies and workflow execution. It can mimic the mouse clicks and keystrokes of a human using screen and keyboard to drive applications and execute system-based work. It can sometimes also be designed to automate application to application. While called robotic process automation, this tool is not a physical object. It can be paired with other tools, BPMS or AI, for example. It is a type of automation that needs structured data to work.

Position and Adoption Speed Justification: The RPA concept is experiencing much hype in consulting, business process outsourcing (BPO) and shared-service centers as it can have an impact on replacing humans in tasks that are rule-based and repetitive rekeying or data collation. This work has often been relocated to less expensive nearshore or offshore delivery locations.

User Advice: RPA is a "glue" type of technology, which will allow you to stick systems together. In order to do more sophisticated activities than just automate work, you need to be able to read handwriting or structure unstructured data or process activities performed by chatbots or machine learning activities. Whether you may be charged more money or not will depend on the RPA vendor or machine learning vendor offerings. Users with manual rule-based processes, where data entry is not automated, should consider finding out where RPA makes sense, and whether their BPO and IT suppliers have built RPA tools or are piloting RPA in other accounts. The process automation technology can then be evaluated to see if it can be programmed and used to replace employees. If the users' current BPO or consulting provider is not piloting such technologies, it is in the users' interest to encourage their provider to do so or to explore running pilots internally.

Users contracting for new labor-intensive BPO deals should consider RPA, but remember that it alone is not a complete solution. Automation of existing processes is akin to the old "mess for less," achieved by simply offshoring inefficient processes. Therefore, users should look for solutions that incorporate both process transformation and automation — reducing or eliminating tasks and activities where possible, and perhaps automating when not.

One criticism of RPA is that it is a temporary fix in lieu of implementing straight through systems. Users should consider other technology solutions that may be more preferable as longer-term solutions.

Pilots and proofs of concept to date have proved that it is possible to rapidly develop automation capabilities and migrate work from humans to software. There are many advantages to finding relatively quick ways to remove people from the process, including the following:

- RPA tools do not make keying errors.
- Automated work is time-stamped, trackable and auditable.
- RPA software can cost less than employees for comparable workloads — as long as the RPA tool is utilized optimally.
- Automation potentially reduces the need for multilingual capabilities.
- RPA tools can be designed to operate 24/7.
- User interface recognition is less able to work in Citrix environments.
- RPA can help address the problematic issue of staff attrition.

- RPA can help mitigate the impact of wage inflation seen in most offshore locations.

Use automated business process discovery tools to accelerate the discovery of automation opportunities for RPA and to inform changes/enhancements to existing scripts to cover more exceptions. Complement RPA with content analytics services (which may be machine-learning-powered) to expand the types of content/channels that the RPA tools can interact with.

Business Impact: If organizations have been outsourcing or offshoring heavily labor-based, data entry or consolidation of data work, then RPA could decrease the need for as many people, increase quality and lower overall process costs. RPA allows organizations to automate work from being manual and to look at new ways to automate work to deliver business outcomes. Potential savings will depend on your organization's IT legacy applications regarding what hasn't already been automated.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: Automation Anywhere; Blue Prism; Infosys; Kofax; Kryon Systems; NICE; Pegasystems; Softomotive; UiPath

Virtual Assistants

Analysis By: Van L. Baker

Definition: Virtual assistants (VAs) help users or enterprises with a set of tasks previously only made possible by humans. VAs use AI and machine learning (such as natural-language processing, prediction models, recommendations and personalization) to assist people or automate tasks. VAs listen to and observe behaviors, build and maintain data models, and predict and recommend actions. VAs can be deployed in several use cases, including virtual personal assistants, virtual customer assistants and virtual employee assistants.

Position and Adoption Speed Justification: The VA space is increasingly dominated by conversational interfaces such as Apple's Siri, Google Assistant, Microsoft's Cortana, IPsoft's Amelia, Nuance's Nina, Amazon Alexa, and IBM's Watson Assistant. Increasingly, behavior and event triggers will enhance VAs. Devices such as Amazon's Echo and Google Home, together with the broad deployment in cellular phones, have put VAs in a position of importance in the consumer's mind. We also continue to see more business-oriented VAs being created, with tools such as Dailogflow Enterprise Edition, Alexa for Business and Watson Assistant. Adoption grows as users get more comfortable with them, technologies improve and the variety of implementations multiplies:

- Unobtrusive, VA-like features — such as Gmail's Smart Compose with recommended sentence completion, and the discovery features in Microsoft's Graph that find unknown resources — are embedded in existing products.
- Use-case-specific VAs have also emerged — such as personal financial advisors, health and wellness coaches, and calendaring agents.
- Chatbots that are subsets of VAs are increasingly used to answer customer questions about products and services.
- VAs can act on behalf of consumers, employees and businesses, but the use cases are all based on the same, constantly improving, language-centric artificial intelligence (AI) technologies.

User Advice: App development leaders should develop a VA strategy that includes voice and text enablement, because VAs will deliver significant benefits to the enterprise's workforce and its customers.

- Anticipate that VAs will proliferate as people and businesses move to conversational user interfaces. Individuals may use several different VAs, while businesses migrate from one deployment to multiple VAs that are composed of groups of specialist chatbots, with narrowly scoped intents, working together with a master chatbot to coordinate the classification of requests.
- Businesses that haven't begun the process of deploying VAs to interact with customers and employees should start now, because customers and employees are increasingly expecting conversational interfaces to be available to address help desk and customer service desk issues.
- Adopt the VAs that are emerging in cloud office suites first, followed by SaaS offerings such as those from SAP, ServiceNow and Salesforce, and consumer application environments such as Facebook.
- Look for opportunities to leverage VAs to make users more productive with their business apps and mobile platforms in targeted, well-defined use cases.
- Incorporate analytics to measure the impact of VAs on behavior and performance. Closely monitor the use of VAs, especially in virtual customer assistant (VCA) use cases, and implement an architecture where handoff to human agents is automated to ensure customer satisfaction.
- Utilize VAs in different use cases: including customer support and engagement, and employee support and enablement, as well as employees' use of personal virtual assistants for services such as HR.

Business Impact:

- VAs have the potential to transform the nature of user behavior, and customer

and employee service, as well as the way work is done and how workplace activities are structured.

- There are many providers of VAs and the quality varies dramatically, so expect rapid changes to the provider landscape.

- VAs can be built using tools and hosted AI services licensed from providers or created using professional services. Performance of the VA is dependent on the quality of the dataset used to add domain-specific information, and the quality of the hosted-language-oriented AI services.

- Security and the collection of personal information are still concerns, but users are growing more comfortable in their interactions with VAs. VAs that are embedded will be the first to gain traction; but as enterprises deploy the technology, so VAs will be broadly used by employees and customers.

- As they mature, VAs may act for the user, forming a relationship with the user over time. VAs shift the responsibility for understanding the process from the user to the system, by corresponding with the user.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: 1-Page; Amazon; Apple; Google; IBM; IPsoft; Microsoft; Nuance; Oracle; [24]7.ai

Cognitive Computing

Analysis By: Kenneth F. Brant

Definition: Cognitive computing is a class of technology enabling the improved performance of a human in a wide range of cognitive tasks. These systems are interactive, iterative and stateful in dialogue, able to recall previous interactions; they are also contextual and able to adapt to changes in information and/or goals. We recognize "cognitive computing" as a promotional term overused by vendors in the marketplace today; we do not believe these systems are truly capable of cognition; they merely mimic/extend the cognitive abilities of humans.

Position and Adoption Speed Justification: Cognitive computing rapidly climbed to the Peak of Inflated Expectations due to the pervasive promotion of the term by major vendors seeking differentiation in the latest generation of the AI marketplace. While some classes of AI such as autonomous vehicles and virtual customer assistants may replace human workers, cognitive computing enhances them. Usability still suffers from difficulty in assembling the right bundles of technology matched to rich bodies of data, lack of skills to train rather than code systems, and organizational and cultural acceptance. Thus, while the hype and expectations will continue to build, there is considerable disillusionment with cognitive computing still ahead. We expect these obstacles will be resolved for the mainstream adopters during the next five years. This will occur as users demand solutions for making sense of patterns in the Internet of Things (IoT), digital business development and big data, coupled with significant investment and innovation by large and startup vendors.

User Advice: Beware of the hype surrounding cognitive computing at this stage of its development, with vastly overstated expectations along the lines of an artificial general intelligence.

Realize that cognitive computing is not a single technology; it is a broad class of technologies that share an approach to augment human cognition, ranging from:

- Virtual assistants (VAs) that will assist with email and administrative issues

- Cognitive expert advisors (CEAs) that pair with specialist knowledge workers to solve very narrow problems and make profound discoveries

- Computer vision (CV), augmented reality (AR) systems that enhance humans' sensory abilities

Develop a mission statement with clear objectives and planning for performance improvement via cognitive computing as part of a five-year technology adoption plan. Make sure to include overall employment goals, new policy considerations, impacts on workers, and sufficient time and resources to implement communication and change management programs.

Resist the temptation to select "winners" at this stage, and make experimental trials involving many vendors.

Employ Mode 2 development, cognitive ergonomics and design thinking to cognitive computing adoption plans.

Business Impact: Cognitive computing can impact the business in broad and deep dimensions. VAs, for example, will impact productivity horizontally and across many job categories, including performers of routine work. Meanwhile, CEAs will impact primarily vertical-specific use cases in the banking, insurance, healthcare and retail sectors, and in the narrow fields of nonroutine, knowledge work. CV/AR will enhance human perception, decision making and productivity in utilities, mining, construction, manufacturing, and maintenance repair and overhaul functions.

Some of the business benefits you should seek to verify and quantify in cognitive-computing-based business models and trials include:

- Higher output per dollar of selling, general and administrative (SG&A) expenses
- Faster cycle times
- Improved productivity of field maintenance workers
- Reduced risk and opportunity costs due to poor/late decisions
- Greater return on R&D investments
- Improved employee safety and satisfaction

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Accenture; CognitiveScale; Deloitte; Digital Reasoning; Google; IBM; Intel Saffron; IPsoft; Microsoft; SparkCognition

Sliding Into the Trough

FPGA Accelerators

Analysis By: Chirag Dekate; Alan Priestley

Definition: Field-programmable gate array (FPGA) accelerator is a server-based reconfigurable computing accelerator that delivers extremely high performance by enabling programmable hardware-level application acceleration.

Position and Adoption Speed Justification: FPGA accelerators feature a large array of programmable logic blocks, reconfigurable interconnects and memory subsystems that can be configured to accelerate specific algorithmic functions. This allows FPGA processors to offload tasks from the main system processor. FPGAs are not programmed with a procedural language. Instead, they are configured with a circuit design language that is alien to the typical programmer. "VHDL" is a language which is difficult for most software engineers to learn, which makes FPGA programming difficult.

In the data center, FPGAs can be used in a range of use cases that require applying consistent processing operations to large volumes of data, such as high-frequency trading (HFT), hyperscale search and DNA sequencing. For example, Microsoft is leveraging FPGAs for search analytics and networks, and Edico Genome's FPGA-based DRAGEN Bio-IT Platform enables high-performance genome sequencing workflows.

FPGAs are typically configured using hardware programming languages such as RTL and VHDL that are very complex to use, this has held back widespread adoption. However, the major FPGA vendors (Intel and Xilinx) are working to address this with libraries and toolsets that enable FPGAs to be configured using software-centric programming models. Adoption is also becoming easier, helped by new frameworks such as OpenCL that lower the time and skills required to use FPGAs. Emerging workloads like deep learning (inference) are driving interest in FPGAs. Intel's integration of FPGAs with mainstream server CPUs and easier access to development platforms exemplified by Amazon Web Services' (AWS's) FPGA-enabled instance types are also driving adoption of FPGAs with the data center.

Today, the biggest growth opportunity for FPGAs in the data center is in the inference portion of deep learning workloads. Given the evolving nature of this new use case and the surrounding software ecosystem, the FPGA accelerator position has been moved to postpeak — 25%.

User Advice: FPGA accelerators can enable dramatic performance improvements within significantly smaller energy consumption footprints than comparable commodity technologies. I&O leaders need to evaluate applicability of FPGA accelerators by:

- Identifying subset of applications that can be meaningfully impacted using FPGAs.
- Outlining costs associated with skill set and programming challenges.
- Evaluating the availability of FPGA-based hardware for use in data center server deployments — either FPGA-based PCIe add-in cards or servers with processors that integrate FPGAs.
- Leveraging cloud-based FPGA services to accelerate development.

I&O leaders should use FPGA accelerators when:

- Preconfigured solutions exist that can help dramatically transform key workloads (e.g., financial trading analytics, genome sequencing, etc.).
- Algorithms will evolve requiring frequent updates at the silicon level that can be utilized by broader applications (example, Microsoft Project Catapult).

Business Impact: FPGAs can deliver extreme performance and power efficiency for a growing number of workloads. FPGAs are well-suited for AI workloads as they excel in low precision (8 bit and 16 bit) processing capabilities in exceptionally energy-efficient footprints. While programmability continues to be a major challenge, limiting broader adoption of FPGAs, I&O leaders should evaluate FPGA-based solutions for genome sequencing, real-time trading, video processing and deep learning (inference). I&O leaders can further insulate against risks by utilizing cloud-based infrastructures for provisioning FPGAs (example, Amazon

Elastic Compute Cloud F1, Microsoft Azure, Baidu Cloud).

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Early mainstream

Sample Vendors: Amazon Web Services; Baidu; Bigstream; Intel; Microsoft Azure; Xilinx (DeePhi Tech)

Computer Vision

Analysis By: Tuong Huy Nguyen; Brian Blau

Definition: Computer vision (CV) is a process that involves capturing, processing and analyzing real-world images and videos to allow machines to extract meaningful, contextual information from the physical world. There are numerous different and important CV technology areas, including machine vision, optical character recognition, image recognition, pattern recognition, facial recognition, edge detection and motion detection.

Position and Adoption Speed Justification: Building algorithms and models to solve vision problems have existed for more than half a century. The convergence of deep neural networks, availability of large swaths of data and massively parallel processors has breathed new life by significantly advancing the field of computer vision — enabling supervised and unsupervised learning, identification, classification, prediction and operation of CV models. For example, 30 years ago, object classification was a difficult, manual task. Results over the past eight years from the ImageNet Challenge best demonstrates current progress in this field. Miscalculation rates have decreased by 30% — meeting and occasionally exceeding human levels of identification. In turn, this has led to the rise of "new players" in CV (outside of academia) such as Amazon, Baidu, IBM, Microsoft and Google. Adoption is still limited, but interest is ramping up quickly for a number of reasons: (1) the interest and hype in using DNNs (although CV can and will continue to use geometric and rule-based systems in many circumstance) and the associated artificial intelligence hype; (2) CVs broad applicability across numerous domains such as robotics, autonomous vehicles, drones, augmented reality and virtual reality; (3) most enterprises are challenged over what to do with all the image/video data they are collecting and how to automate the processing of that image data; (4) computer vision is a special use case and natural extension for the Internet of Things (IoT). It's the external sensor that extends and expands the reach of IoT.

User Advice: Technology innovation leaders: use computer vision to augment your workforce and automate the processing of image and video data. For example, it can be used in automation, such as bin picking, in assistive technologies for people with disabilities and act as expert advisors in fields that require analysis of images and video. In light of ongoing security and privacy concerns as well as legislation such GDPR, you will also need to evaluate exposure to legal liabilities associated with the collection and processing of image/video data.

Business Impact: Visually enabling devices will transform how they can interact with the environment. Vision makes an excellent complement to other sensor data, such as geolocation, inertia and audio. As such, it also enhances humans' abilities to interact with the digital and physical world. This has fueled broad interest in this discipline for applications such as autonomous vehicles, robotics, drones, augmented — mixed and virtual reality — security, biometrics and more.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Amazon; Apple; Baidu; Clarifai; Facebook; Google; IBM; Microsoft; Tencent

Predictive Analytics

Analysis By: Peter Krensky; Alexander Linden; Carlie J. Idoine

Definition: Predictive analytics is a form of advanced analytics that examines data or content to answer the question, "What is going to happen?" or more precisely, "What is likely to happen?" It is characterized by techniques such as regression analysis, multivariate statistics, pattern matching, predictive modeling and forecasting.

Position and Adoption Speed Justification: The excitement surrounding predictive analytics continues to drive more interest and adoption at all maturity levels. However, recent disillusionments connected to the wave of open-source adoption are beginning to surface in the form of overinflated vendor promises and failed early notebook projects that were excessively ambitious. Still, this technology is unlikely to spend significant time in the Trough of Disillusionment as the rate of evolution and underlying value of predictive analytics drives the technology rapidly toward the Plateau of Productivity in the near future.

From those just getting started with predictive analytics, to enterprises with mature data science labs, organizations are evangelizing the value and potential impact of predictive models. Interest is also driven by improved availability of data, lower-cost compute processing (especially in the cloud)

and proven real-world use cases. Predictive models are no longer just produced by data science platforms; predictive analytics is being embedded in more business applications than ever before. Client searches on gartner.com for "predictive analytics" continue to trend steadily upward.

User Advice: Predictive analytics can be quite easy to use if delivered via a packaged application. However, packaged applications do not exist for every analytics use case. Packaged applications may also often not provide enough agility, customization or competitive differentiation. In these situations, organizations are advised to build solutions either through an external service provider, or with typically highly skilled in-house staff using a combination of open-source technologies and a data science platform. Many organizations increasingly use a combination of these tactics (build, buy, outsource) and some vendors have hybrid offerings. Finally, to secure the success of predictive analytics projects, it is important to focus on an operationalization methodology to deploy these predictive assets.

Business Impact: By understanding likely future outcomes, organizations are able to make better decisions and anticipate threats and opportunities, being proactive rather than reactive (for example, predictive maintenance of equipment, demand prediction, fraud detection, and dynamic pricing). Interest and investment continues to grow in both new use cases, and more traditional applications of predictive analytics (for example, churn management, cross-selling, propensity to purchase, database marketing, and sales and financial forecasting).

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Alteryx; H2O.ai; IBM; KNIME; MathWorks; Microsoft; RapidMiner; SAS

Autonomous Vehicles

Analysis By: Carsten Isert

Definition: Autonomous or self-driving vehicles can navigate and drive certain parts or the whole distance from a starting point to a specified destination without human intervention by using various onboard sensing and localization technologies, such as lidar, radar, cameras, GPS and map data, in combination with AI-based decision-making capability. While self-driving cars are getting most of the attention at present, the technology can also be applied to nonpassenger vehicles for transportation of goods.

Position and Adoption Speed Justification: A number of signs of moving into the Trough of Disillusionment have occurred in the past year. In early 2018 several accidents with automated vehicles occurred, including the death of a pedestrian. In addition, some announced milestones have passed without the promised launches.

But there has also been progress. The first commercial services are expected to be launched in 2018.

The efforts of automobile manufacturers and technology companies to develop self-driving vehicles have been prominently featured by mainstream media, leading to unrealistic and inflated expectations for the technology. Artificial Intelligence (AI) is a critical technology for enabling autonomous vehicles, and development of machine learning algorithms for autonomous vehicles has accelerated. Key challenges for the realization of autonomous vehicles continue to be centered on cost reductions for the technology and industrialization, but they increasingly include regulatory, legal and societal considerations, such as permits for operation, liability, insurance and the effects of human interaction.

Autonomous vehicle technology has disruptive potential not only for applications in smart mobility and logistics, but also for shipping, mining, agricultural, industrial, security and military operations.

Continued advancements in sensing, positioning, imaging, guidance, mapping and communications technologies, combined with AI algorithms and high-performance computing capabilities, are converging to bring the autonomous vehicle closer to reality. However, in 2018, complexity and cost challenges remain high, which is impacting reliability and affordability requirements.

In terms of investments, 2018 has not seen the major numbers as those from 2017.

Initially, the pace of technology innovations in individual country, state and global legislation will likely result in specific, limited-use deployments of self-driving vehicles in the short term (for example, low-speed operations in a campus environment or designated area within a city, and high-speed operations on certified highways) as demonstrated by the planned commercial launch of Waymo in Arizona.

Overall, the reduced funding, the accidents and evidence gathered in discussion with industry experts leads to the prediction of autonomous driving entering the trough.

User Advice: The adoption of autonomous vehicle technology will develop in three distinct phases — automated driver assistance, semiautonomous and fully driverless vehicles. Each phase will require increasing levels of technical sophistication and reliability that rely less and less on human driving intervention. Automotive companies, service providers, governments and technology vendors (for example, software, hardware, sensor, map data and

network providers) should collaborate on joint research and investments to advance the required technologies, as well as work on legislative frameworks for self-driving cars.

Furthermore, educate all constituencies of the benefits of self-driving vehicles. Consumer education is critical to ensure that demand meets expectations once autonomous vehicle technology is ready for broad deployment. Specific focus must be applied to the transitional phase, where autonomous or semiautonomous vehicles will coexist with an older fleet of nonautonomous vehicles.

Autonomous vehicles will have a disruptive impact on some jobs, such as bus, taxi and truck drivers. Develop policies and programs to train and migrate employees that will be affected by automation to other roles.

Business Impact: The main implications of self-driving vehicles will be in economic, business and societal dimensions. Automotive and technology companies will be able to market autonomous vehicles as having innovative driver assistance, safety and convenience features, as well as an option to reduce vehicle fuel consumption and to improve traffic management. The interest of nonmobility companies (such as Waymo and Baidu) highlights the opportunity to turn self-driving cars into mobile computing systems that offer an ideal platform for the consumption and creation of digital content, including location-based services, vehicle-centric information and communications technologies.

Autonomous vehicles are also a part of mobility innovations and new transportation services that have the potential to disrupt established business models. For example, autonomous vehicles will eventually lead to new offerings that highlight mobility-on-demand access over vehicle ownership by having driverless vehicles pick up occupants when needed. Autonomous vehicles will deliver significant societal benefits, including reduced accidents, injuries and fatalities, as well as improved traffic management, which could impact other socioeconomic trends. For example, if people can use travel time for work or entertainment while being driven in an autonomous vehicle, living near a city center to be close to work won't be as critical, which could slow the process of urbanization.

When autonomous driving enters the Trough of Disillusionment, it might be a good opportunity for new market entrants.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Audi; BMW; Daimler Group; Ford Motor Co.; General Motors; Nissan; Tesla; Uber; Volvo Cars; Waymo

Commercial UAVs (Drones)

Analysis By: Aapo Markkanen

Definition: Commercial unmanned aerial vehicles (UAVs, also known as drones) are small helicopters, fixed-wing airplanes, multirotors and hybrid aircrafts that have no human pilot on board. They are either remotely controlled by human pilots on the ground or outfitted for autonomous navigation. Unlike their consumer or military counterparts, they are used for commercial purposes.

Position and Adoption Speed Justification: In 2018, commercial UAVs have entered the Trough of Disillusionment. In the technological sense, such drones are relatively mature and capable of increasingly sophisticated tasks. However, their adoption is often held back by regulations that restrict many use cases. In particular, flying drones beyond visual line of sight (BVLOS), above people or in restricted airspace, such as close to airports, are types of operations that are heavily regulated in most countries. Additionally, the high cost of vertically specialized end-to-end drone solutions — which cover the devices, the supporting software and the flight operations — holds back large-scale adoption among end users. Gartner expects commercial UAVs to approach the Plateau of Productivity in two years, assuming that regulatory conditions and certain technology elements continue to improve as expected. Especially, autonomous flights will represent a boost to the market, but their enablement requires both regulatory changes and technological advancements.

User Advice: Overall, a corporate drone program should have both short-term and long-term objectives. This is because commercial UAVs can deliver major operational benefits already today, but once the relevant aviation regulations become more permissive, their potential can shoot up practically overnight. Organizations considering drones, therefore, should not solely plan on the basis of available technology, but also factor in the local regulatory outlook. It makes sense to proactively identify relevant regulatory changes and take advantage of them as early as possible. The Low Altitude Authorization and Notification Capability (LAANC) initiative in the U.S., aiming to accelerate the waiver approvals for flying in restricted airspace, is one such example. Today, the leading use cases include aerial photography, mapping and surveying, volume measurement, and remote inspection. All of these can be considerably enhanced by the right of use analytics, so as part of their UAV planning, the adopters should also take into account how they can exploit the captured data in the best possible way. Use cases involving physical tasks — such as delivering objects or repairing assets — are currently largely in their nascency, but they can be expected to become gradually more viable over the next two to three years.

However, benefits of commercial UAVs in applications that rely on completion of physical tasks will take longer to materialize than in the ones that focus on data capture and analysis.

Business Impact: Most of all, commercial UAVs are a technology to enhance the capabilities of the roles such as surveyors, inspectors, drivers and cameramen who traditionally perform labor-intensive tasks in potentially unsafe conditions. As such, drones offer productivity improvements by reducing and/or redeploying head count, while enabling real-time data capture and improving employee safety. Examples of industry verticals where commercial UAVs can particularly add value include agriculture, construction, emergency services, extractive industries, media and entertainment, as well as utilities. In most of the verticals, the value of commercial UAVs is in reducing operational expenditure and improving safety, but there are also revenue-generating opportunities in industries such as cinematography, surveying and logistics.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Boeing; Cyberhawk Innovations; DJI; Lockheed Martin; PrecisionHawk

Augmented Reality

Analysis By: Tuong Huy Nguyen; Brian Blau

Definition: Augmented reality (AR) is the real-time use of information in the form of text, graphics, audio and other virtual enhancements integrated with real-world objects and presented using a head-mounted-type display or projected graphics overlays. It is this "real world" element that differentiates AR from virtual reality. AR aims to enhance users' interaction with the environment, rather than separating them from it.

Position and Adoption Speed Justification: Current technology is best-suited for purpose-built, specialized solutions. As such, position and adoption speed will vary by vertical and industry. Current horizontal tasks seeing the most traction are task itemization, visual design and video guidance. This profile represents a homogeneous view of AR implementations across market segments.

Market interest remains fairly steady according to Google Trends, even as high-profile developments in the AR space continue to fuel interest and hype in this area. These developments include Apple's launch of ARKit, Google's launch of ARCore and Magic Leap's long-rumored HMD.

AR is currently struggling with mismatched expectations (vendors promising solutions beyond current capabilities) and poor implementations (for example, solutions delivered without immersive development knowledge or workflow integration, or not mapped to business value or need). B2C implementations are still struggling to show consumers value. Better and more transparent hardware, coupled with more compelling use cases, is needed before further progress can be made.

Based on Gartner inquiry and industry news, B2B AR continues to gain traction as more enterprises are discovering and seeing the value of using AR in their workflow. HMD sales reflect the burgeoning pilot deployments. Advancements in HMD hardware will provide more compelling hands-free use cases for AR, as well.

User Advice: Decide on the audience for your AR solution. Internal- and external-facing solutions are not transposable. Restrict initial trials to a specific task or goal. Set benchmarks against unaugmented solutions to understand risks and benefits. Set the business goals, requirements and measurements for your AR implementation before choosing a provider. Rich and robust offerings can bring value only if you have a clear intention for the deployment. For external-facing implementations, use AR as an extension of your brand and experience. For internal-facing implementations, use AR as a tool that will enhance employee job function. This could include, for example, delivering context-specific information at the point of need for mobile workers, reduction of head count in plant and maintenance operations, or enhancing business processes via AR-based training and instruction.

Business Impact: By leveraging device sensors, AR acts as a digital extension of users' senses, and it serves as an interface for humans to the physical world. It provides a digital filter to enhance the user's surroundings with relevant, interesting and/or actionable information.

AR bridges the digital and physical world. This has an impact on both internal- and external-facing solutions. For example, internally, AR can provide value by enhancing training, maintenance and collaboration efforts. Externally, it offers brands, retailers and marketers the ability to seamlessly combine physical campaigns with their digital assets.

As such, AR is broadly applicable across many markets, including gaming, industrial design, digital commerce, marketing, mining, engineering, construction, energy and utilities, automotive, logistics, manufacturing, healthcare, education, customer support, and field service.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Blippar; Catchoom; DAQRI; Google; Microsoft; Ubimax; Upskill; Wikitude

Knowledge Management Tools

Analysis By: Rich Doheny; Kenneth Gonzalez

Definition: Infrastructure and operations (I&O) leaders use knowledge management (KM) tools to create, modify and access IT knowledge bases. KM tools are often linked to portals that support self-service, so that end users can access relevant intellectual assets themselves. The products are defined by their ability to federate, store and access information about IT and non-IT services. KM tools are available as stand-alone options or integrated components of broader IT service management (ITSM) tools.

Position and Adoption Speed Justification: KM provides significant untapped potential for many IT organizations to optimize, drive efficiencies and realize economies of scale in ITSM. Done correctly, it can greatly improve I&O effectiveness and business user self-sufficiency. In addition, KM can provide a vital component in enabling future automation as a repository of information to teach emerging technologies, including chatbots and virtual support agents. Many intermediate and advanced ITSM vendors are enhancing their products' capabilities in the area of KM. As a result, the market for stand-alone tools targeting the ITSM use case has seen consolidation.

KM tools are becoming more commonplace in IT organizations, and Gartner estimates that market penetration is between 20% and 50%. Many organizations struggle to realize the ROI and true value, due to cultural issues, behavioral challenges and a lack of understanding regarding the successful implementation of the underlying KM practices.

User Advice: Knowledge management tools should be an integral part of an I&O strategy, whether through stand-alone options or as part of an ITSM suite. Integration is necessary to reap the benefits of a knowledge base, and buyers should assess which platform best suits their needs. Don't overemphasize the tools' potential for success. Tools enable processes, but are only as good as the processes, procedures and policies you have in place. Formal knowledge management governance mechanisms are crucial to ensure that the content is reviewed, updated and corrected on an ongoing basis.

Business Impact: Used optimally, a good knowledge base can create significant efficiencies across I&O, although are often targeted for incident handling, request fulfillment, training, impact assessments and self-service implementation. Knowledge tools can drive down support costs as well as free up IT service desks and other resources to be deployed elsewhere. The effective use of KM can also pay off in terms of the qualitative perspective, driving customer satisfaction and overall customer perception.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Sample Vendors: ComAround; Upland Software

Climbing the Slope

Virtual Reality

Analysis By: Brian Blau

Definition: Virtual reality (VR) provides a computer-generated 3D environment that surrounds a user and responds to an individual's actions in a natural way, usually through immersive head-mounted displays (HMDs). Gesture recognition or handheld controllers provide hand and body tracking, and haptic (or touch-sensitive) feedback may be incorporated. Room-based systems provide a 3D experience while moving around large areas or can be used with multiple participants.

Position and Adoption Speed Justification: Immersive VR applications are more advanced than other graphical simulations and the Time to Plateau of five to 10 years is consistent with awareness, exposure to the technology and overall adoption is mainly in the consumer market, but growing in business.

VR is usually experienced using HMDs. The well-known devices on the market in 2018 are the Oculus Rift and Oculus Go, Sony PlayStation VR, HTC VIVE, Samsung Gear VR and Google Daydream. VR is mature enough for enterprise use, but caution is required as while the devices are capable, the success of VR usage depends on the quality of the device and user experience. Most VR user engagement come from video games or watching video, which can be 360-degree surround or TV and movie content. VR HMD deployments are slow but growing. New areas for VR include retail and e-commerce, and improved HMD quality and system ease of use.

User Advice: Virtual reality can be used in a variety of business scenarios:

- Complex simulation and training applications
- Military simulation and training, such as flight simulators
- Telepresence in scenarios such as remote medicine
- Equipment operator training
- Entertainment and social experiences, such as video games or 360 surround video or interactive movies
- Product marketing to extend in the brand interaction or in product design
- Architectural walkthroughs and scientific visualization, such as genome mapping
- Modeling, such as geomodeling in the oil industry

While VR can be amazingly sophisticated and beneficial, the level of customization can come at a high cost. Recent advances in HMD technologies may help ease these obstacles, so developers should focus on building effective and quality experiences. Standards for artificial intelligence scripting, object metadata and social identity data are becoming more popular, due to increased use of personal and social networking technologies, which will help developers make VR more personalized and intelligent. Technologies such as cloud graphics processing and mobile video games, as well as the proliferation of broadband access, will allow application developers to integrate VR more easily into their products.

VR developers should consider targeting immersive video game development, interactive movies and new storytelling experiences, live events and business-focused scenarios where using advanced visualization and HMDs can benefit the task or customer interaction point due to their ability to offer higher degrees of visual fidelity and personalization over what flat-screen-based systems can provide.

Business Impact: VR can support a wide variety of simulation and training applications, including rehearsals and response to events. VR can also shorten design cycles through immersive collaboration and enhance the user interface experience for scientific visualization, education and entertainment. Businesses will benefit due to VR's immersive interfaces, helping create task efficiencies or reducing costs associated with new product design, or can enhance the understanding of information through advanced graphical visualization and simulation technologies.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Google; HTC; NextVR; Oculus VR; Samsung Electronics; Sony; Valve; WorldViz

Entering the Plateau

GPU Accelerators

Analysis By: Chirag Dekate; Martin Reynolds; Alan Priestley

Definition: GPU-accelerated computing is the use of a graphics processing unit (GPU) to accelerate highly parallel compute-intensive portions of the workloads in conjunction with a CPU.

Position and Adoption Speed Justification: GPUs are highly parallel floating-point processors designed for graphics and visualization workloads. Over the last decade, NVIDIA and others have added programmable capability to GPUs, enabling applications to access deep, fast-floating-point resources. GPUs also have very high-bandwidth memory subsystems. For many highly parallel, repetitive, compute-intensive applications, these capabilities deliver dramatic performance improvements.

Compute-intensive applications including molecular dynamics, computational fluid dynamics, financial modeling and geospatial applications can utilize GPUs today. Programming GPUs can be challenging, as execution order and code optimization are critical. However, toolkits like NVIDIA's CUDA can dramatically lower the programming challenges. GPU computing has moved forward on the Slope of Enlightenment to account for new use cases and the evolutionary nature of deep neural networks (DNNs).

We anticipate that DNN technologies will mature quickly, supported by open frameworks from the large cloud providers. These frameworks include TensorFlow, Torch, Caffe, Apache MXNet and Microsoft Cognitive Toolkit.

User Advice: GPU-accelerated computing can deliver extreme performance for highly parallel compute-intensive workloads in high-performance computing (HPC) and DNN training. GPU computing is also available as a cloud service, and may be economical for applications where utilization is low but urgency of completion is high. Cloud GPUs shift the balance of supercomputing from on-premises toward the cloud.

I&O leaders can accelerate compatible applications using GPU solutions by:

- Selecting GPU compute platforms that offer the most mature software stack.
- Optimizing infrastructure costs by evaluating cloud-hosted GPU environments for proof of concept (POC) and prototype phases.

I&O leaders should use GPU accelerators when applications require extreme performance and have high degrees of compute parallelism (example, many high-performance computing and deep learning applications, etc.).

Business Impact: High-performance computing and deep learning are essential to many digital business strategies. For this fast-growing workload, traditional enterprise ecosystems based on CPU-only approaches will not suffice. Leverage mature GPU technologies for select HPC applications and deep learning infrastructures. Programmability challenges have been largely solved in GPU-accelerated computing by frameworks like CUDA. I&O leaders can minimize risk by using cloud-hosted GPU environments for testing and evaluation.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Mature mainstream

Sample Vendors: AMD; Cray; Dell; Hewlett Packard Enterprise; IBM; Lenovo; NVIDIA; Supermicro

Ensemble Learning

Analysis By: Peter Krensky; Alexander Linden

Definition: Ensemble learning techniques are machine learning algorithms wherein a set of predictive models is created and its outputs combined to become the single output of the entire ensemble. This methodology draws heavily on the "wisdom of the crowd" principle, where the diversification of opinions or model outputs is key. The most well-known patterns of bagging, boosting and stacking techniques include random forests and gradient boosting. Such techniques regularly achieve very high accuracy for supervised learning problems.

Position and Adoption Speed Justification: Adoption of ensemble techniques continues to steadily grow. All major data science vendors offer this technology as part of their portfolios. Ensemble learning has become a widely accessible approach for both data scientists and citizen data scientists. As the use of ensemble techniques becomes even more commonplace within data science teams, the technology will reach the Plateau of Productivity in the next 12 months.

User Advice: For even a moderately skilled data science professional, ensemble techniques are relatively easy to apply to scenarios involving high precision. They are often able to achieve a 5% to 30% reduction in error rates, which may result in a substantial impact on modeled metrics. Ensemble learning is especially valuable for novel projects where it is difficult to identify a model of best fit. However, deployment of ensemble techniques can be a computational burden to current infrastructures.

Ensemble techniques may not be an option in regulated industries, where predictive models must be entirely explainable and transparent.

Data and analytics leaders should understand the advantages and disadvantages of ensemble learning:

Pros:

- A proven method for improving the accuracy of a model that works in most cases.
- Makes models more robust and stable.
- Can be used to capture linear as well as nonlinear relationships in the data.

Cons:

- Can be time-consuming in terms of performance, and generally not well-suited for real-time applications.
- Selection of models for creating an ensemble is a skill that can take time to master.

Business Impact: Almost every predictive analytics use case and machine learning task can benefit significantly from the application of ensemble techniques. Success stories of applied techniques continue to bolster ensemble learning's reputation for increasing predictive accuracy. Ensemble methods are frequently deployed in analytics competitions such as the KDD Cup and Kaggle competitions, and acquit themselves finely.

Data and analytics leaders should inquire with their data science teams as to how and when they are employing ensemble techniques. These techniques are robust against outliers and overfitting. Predictions based on ensemble methods can be used as rank-ordering scores or interpreted as regression functions, making them especially useful for a wide range of tasks across financial services and marketing applications, where customer behavior needs to be predicted.

Ensemble techniques can offer an invaluable new perspective on a model and provide validation for existing models already in production. Ensemble learning is also a well-established idea creation tool for data science teams working in the business exploration and advanced prototyping use cases.

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Alteryx; DataRobot; H2O.ai; IBM; KNIME; Microsoft; RapidMiner; SAP; SAS

Speech Recognition

Analysis By: Anthony Mullen

Definition: Speech recognition technology translates human speech into text for further processing.

Position and Adoption Speed Justification: Speech recognition performance has rapidly accelerated in the last three years. Heavyweights such as IBM, Microsoft, Google, Amazon and Baidu all demonstrated rapid scientific improvements in 2016-2017, claiming equal or better performance than human transcription.

Services continue to improve especially around developer and process support, for example, in 2018, Google overhauled their Cloud Speech to Text API to improve performance by providing multiple machine learning models to suit different use-case contexts (e.g., phone call, voice commands, video) as well as improving punctuation to make transcriptions more readable.

In tandem with algorithmic advances, speech-to-text applications have been propelled by hardware progress, the adoption of conversational agents such as chatbots and virtual assistants by enterprises and consumer adoption of speech interactions on smartphones, game consoles and, in particular, virtual personal assistant speakers such as Amazon Echo and Google Home. The use of speech-to-text technologies is also growing for connected home and automotive domains and embedded solutions running on edge devices without the need for cloud to create new usage scenarios.

User Advice: From a human computer interface standpoint, speech recognition is applicable and useful where users have:

- An interest or motivation, e.g., injuries or disabilities.
- Their "hands busy, eyes busy" and need data entry or system control performed via voice alongside other tasks, i.e., in warehouses, factories, hospitals, shop floors, cars or homes.
- A need for sustained, voluminous or repeated input such as office, medical and legal dictation, particularly in applications where speech shortcuts can be used to insert commonly repeated text segments.

- Domain knowledge but not system knowledge. i.e., interactions are expressed through natural speech, rather than proprietary system commands and interfaces.

Typical use-case scenarios of adoption include:

- **Supporting Users.** Consumer electronics providers should consider the use of speech recognition services for applications, smartphones, smart homes and cars, either licensing technology to work online/offline for their own devices or using cloud services to enrich the experience and presence of their devices and services.

- **Supporting Customers.** Speech recognition for telephony and contact center applications enable enterprises to automate call center functions such as travel reservations, order status checking, ticketing, stock trading, call routing, directory services, auto attendants and name dialing. Further applications include the use of speech to text for marketing and commerce interactions.

- **Supporting Employees.** Existing enterprise application developers should consider the use of speech recognition and natural-language entry as a method of simplifying UIs and increasing productivity. There are an increasing amount of use cases for speech to text in the workplace from meeting room support and transcription, sales support, voice access to analytics and reports to hands-free warehousing and virtual employee assistant (VEA) use cases. Further, there are legal imperatives, such as GDPR, compliance and redaction, that require businesses to be able to obtain transcripts of voice calls.

Vendors in this space can generally be split broadly into two camps — general purpose platforms and specialists that provide a managed service. Generalist platforms tend to cover many languages and target general purpose speech. Specialists offer tailored solutions designed to perform well for a specific business context and lexicon using custom dictionaries and semantic tools to work with DNN models to improve disambiguation.

Making speech to text work for most organizations entails more than simply activating an off-the-shelf solution. Organizations should plan for an extended period of human involvement to monitor, train and improve performance — especially around modelling proprietary business terms, dialects and noisy/complex environments.

Business Impact: Unlike other elements of the natural-language processing chain, speech to text (and text to speech) can be considered to be a stand-alone commodity where its modules can be plugged into a variety of natural-language workflows.

After a series of breakthroughs with the technology and while the rapid pace has eased, there is still a regular cadence of innovation and improvement in areas such as edge-based speech to text, hybrid models using semantic and DNN techniques and GPU/TPU hardware. These gains were largely driven by deep learning. Using techniques like convolutional neural networks (CNNs), long short term memory (LSTM), recurrent neural networks (RNNs). Also, end-to-end neural architectures using connectionist temporal classification (CTC) loss (championed by Baidu) are improving time to train models.

Tech heavyweights like Google, Apple and Microsoft also collect large troves of training data from opt-on programs with consumers and this ongoing cycle of training data and improved algorithms will see the issue of speech to text as a largely solved problem within the next two years. Specialist vendors with custom language models designed for verticals will continue to be essential to organizations looking to embed this technology deep into their business.

Benefit Rating: Transformational

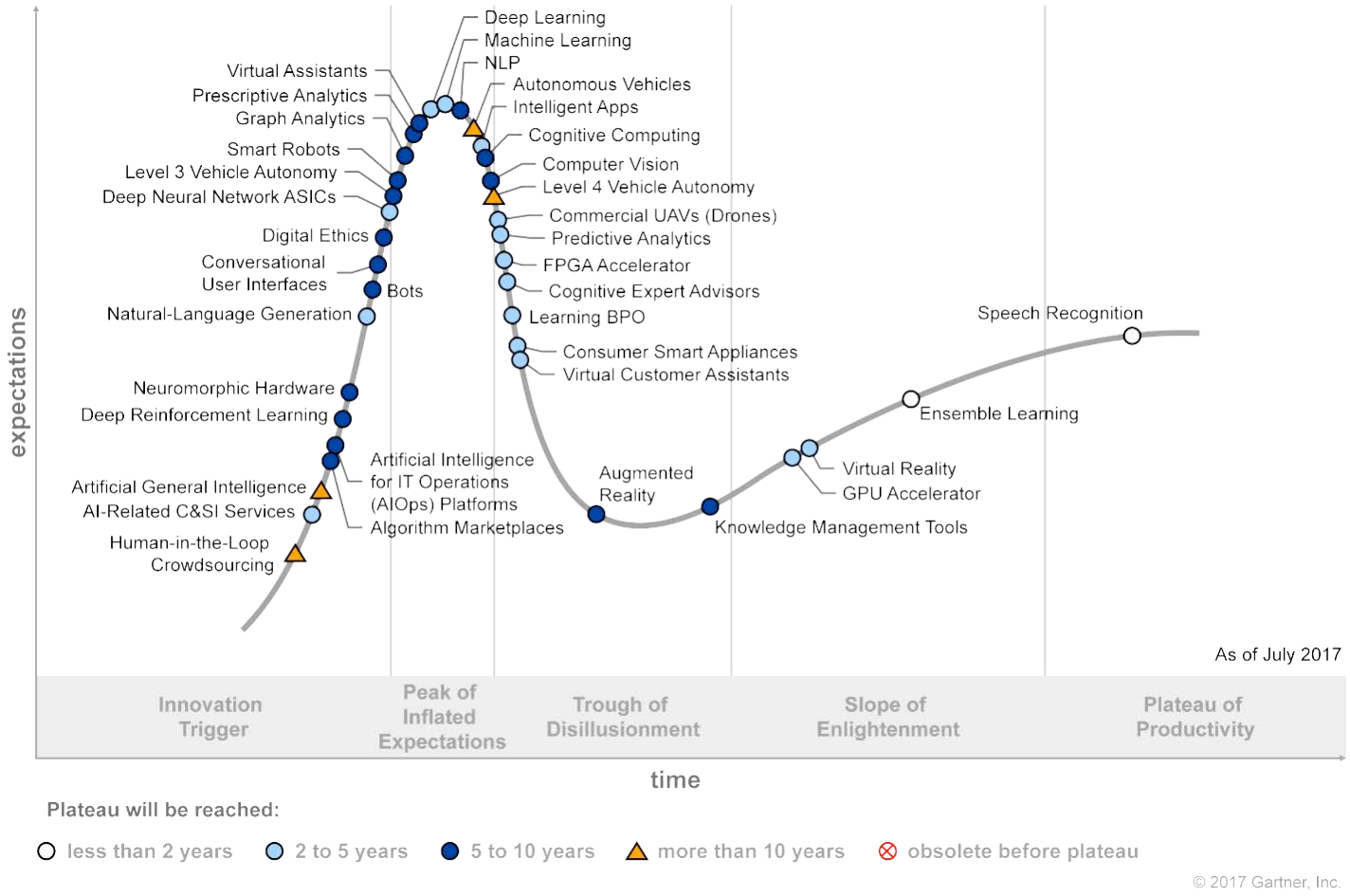
Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Sample Vendors: Amazon; Baidu; Google; iFLYTEK; IBM; Intelligent Voice; Microsoft; NICE; Nuance

Appendixes

Figure 3. Hype Cycle for Smart Machines, 2017



Source: Gartner (July 2017)

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Table 1. Hype Cycle Phases

Phase	Definition
Innovation Trigger	A breakthrough, public demonstration, product launch or other event generates significant press and industry interest.
Peak of Inflated Expectations	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the technology is pushed to its limits. The only enterprises making money are conference organizers and magazine publishers.
Trough of Disillusionment	Because the technology does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
Slope of Enlightenment	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the technology's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
Plateau of Productivity	The real-world benefits of the technology are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.
Years to Mainstream Adoption	The time required for the technology to reach the Plateau of Productivity.
Source: Gartner (July 2018)	

Table 2. Benefit Ratings

Benefit Rating	Definition
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings
Source: Gartner (July 2018)	

Table 3. Maturity Levels

Maturity Level	Status	Products/Vendors
Embryonic	• In labs	• None
Emerging	• Commercialization by vendors • Pilots and deployments by industry leaders	• First generation • High price • Much customization
Adolescent	• Maturing technology capabilities and process understanding • Uptake beyond early adopters	• Second generation • Less customization
Early mainstream	• Proven technology • Vendors, technology and adoption rapidly evolving	• Third generation • More out-of-the-box methodologies
Mature mainstream	• Robust technology • Not much evolution in vendors or technology	• Several dominant vendors
Legacy	• Not appropriate for new developments • Cost of migration constrains replacement	• Maintenance revenue focus
Obsolete	• Rarely used	• Used/resale market only
Source: Gartner (July 2018)		

Evidence

Evidence for this note was garnered from:

- "Forecast: The Business Value of Artificial Intelligence, Worldwide, 2017-2025"
- Gartner 2018 CIO Survey conducted among 3,138 respondents
- Gartner 2018 Artificial Intelligence Consumer Perceptions Survey conducted among 4,019 respondents.
- Gartner search analytics.
- Gartner client inquiry analytics.

Source: Gartner Research Note G00357478, S. Sicular, K. Brant, 24 July 2018

A New Era Empowered By Artificial Intelligence