Hype Cycle for Artificial Intelligence, 2018

Published: 24 July 2018 **ID:** G00357478

Analyst(s): Svetlana Sicular, Kenneth Brant

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.

Table of Contents

List of Tables

List of Figures

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 None of the profiles in this Hype Cycle is Obsolete Before Plateau, but not all will not survive, and many will morph into something different — this will depend on the choices and decisions that customers are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.

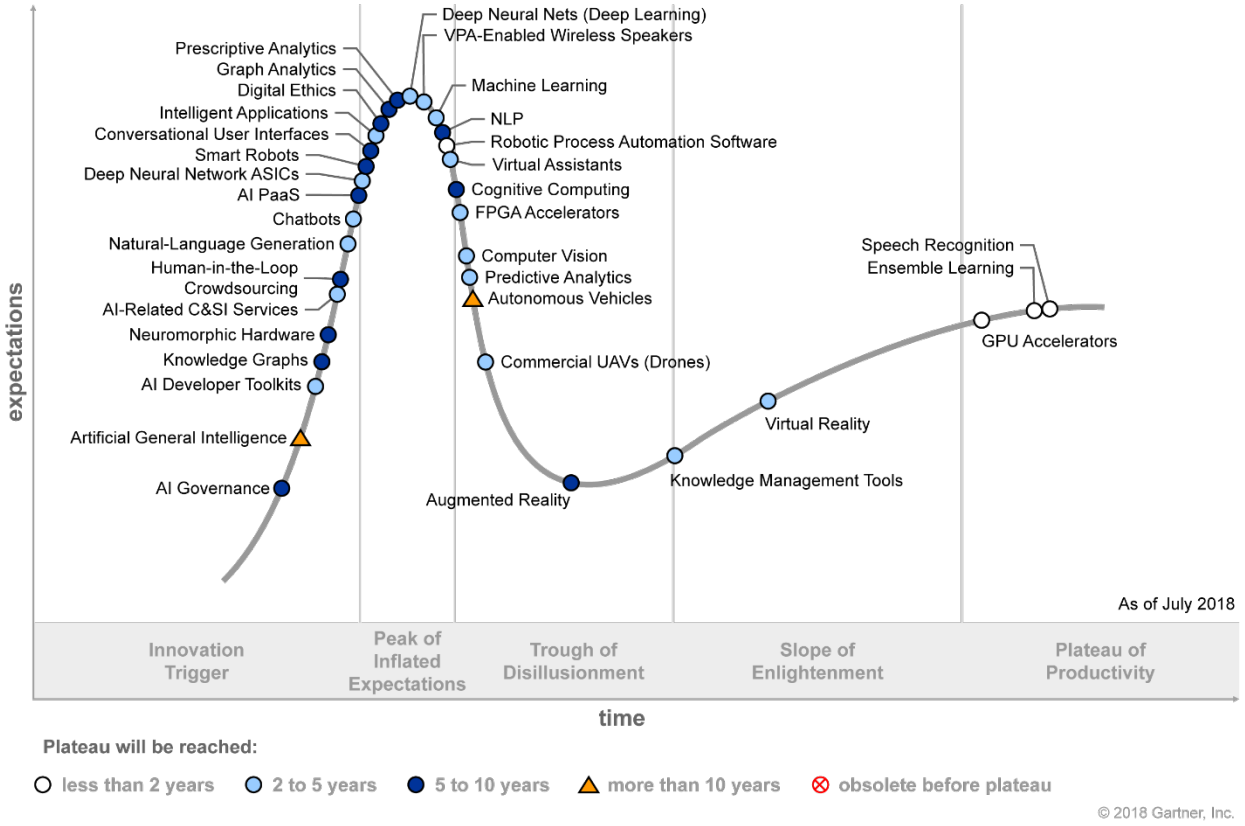
New Entrants

- **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)

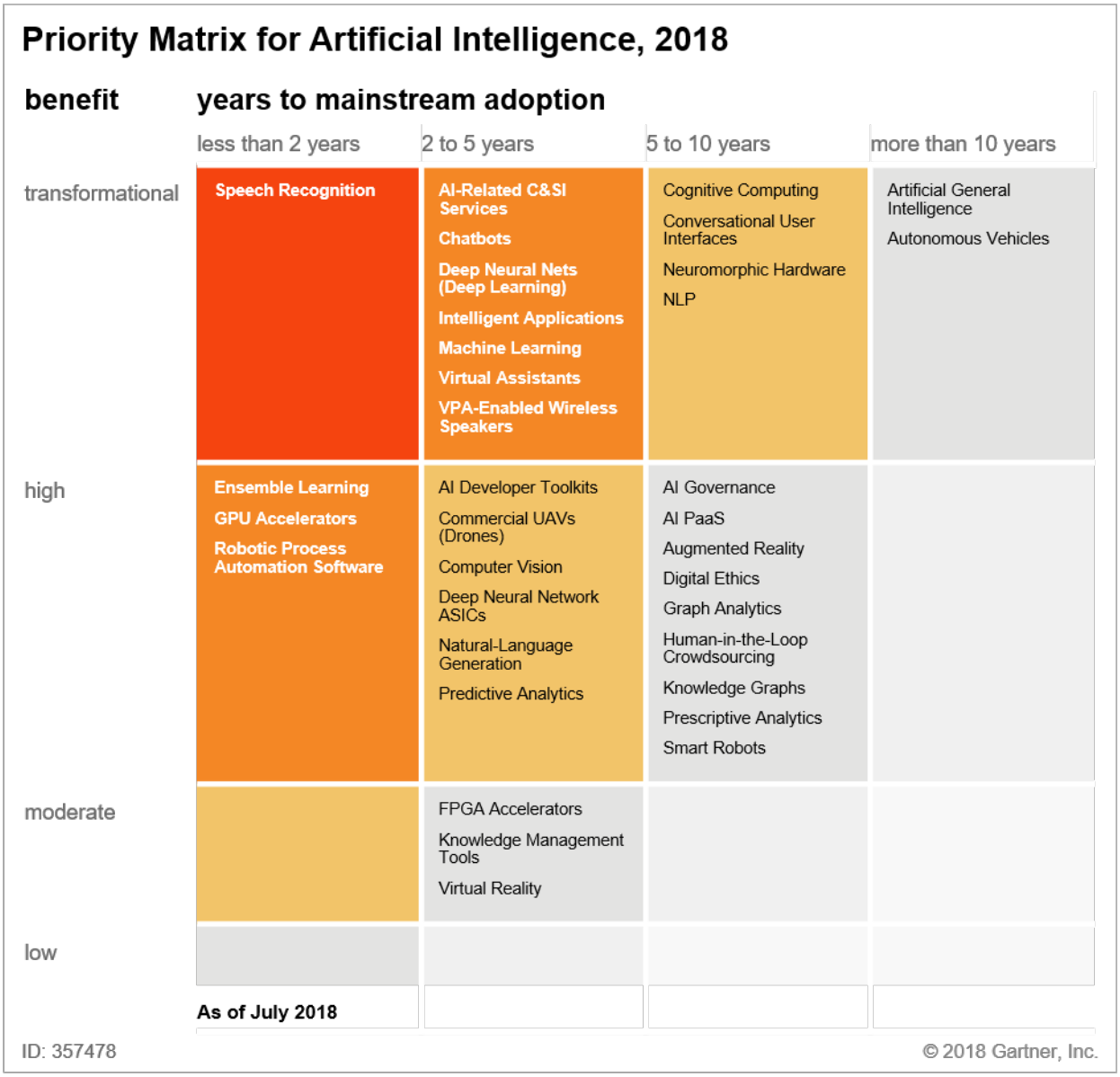
The Priority Matrix

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. 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.

Figure 2. Priority Matrix for Artificial Intelligence, 2018



Source: Gartner (July 2018)

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

Recommended Reading: "Seek Diversity of People, Data and Algorithms to Keep AI Honest"

"Best Practices for Driving Successful Analytics Governance"

"Reset Your Information Governance Approach by Moving From Truth to Trust"

"A Chief Data Officer's Guide to an AI Strategy"

"The CIO's Guide to Digital Ethics: Leading Your Enterprise in a Digital Society"

"Choose the Right Center of Excellence for Your Artificial Intelligence Strategy"

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

Recommended Reading: "Smart Machines See Major Breakthroughs After Decades of Failure"

"How to Define and Use Smart Machine Terms Effectively"

AI Developer Toolkits

Analysis By: 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

Recommended Reading: "AI Will Alter Application Development — Things to Do Now"

"Where You Should Use Artificial Intelligence — and Why"

"2018 Planning Guide for Application Platform Strategies"

"Toolkit: Enabling Data Literacy and Information as a Second Language"

"Integrating Machine Learning Into Your Application Architecture"

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

Recommended Reading: "Magic Quadrant for Data Management Solutions for Analytics"

"Magic Quadrant for Insight Engines"

"Magic Quadrant for Operational Database Management Systems"

"Hype Cycle for Data Science and Machine Learning, 2017"

"Automate Knowledge Management With Data Science to Enable the Learning Organization"

"Digital Workplace Graphs Promise to Improve Productivity and Collaboration, but Risks Exist"

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

Recommended Reading: "Cool Vendors in Novel Semiconductors for Neural Networks, 2016"

"Market Guide for Compute Platforms"

"Three Elements of a Scalable Enterprise Machine Learning Infrastructure Strategy"

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

Recommended Reading: "Market Guide for AI-Related Consulting and SI Services for Intelligent Automation"

"Market Trends: Five Emerging Artificial Intelligence Use Cases Fuel IT Services Opportunities for Intelligent Automation"

"Smart Machines: Consulting and System Integration Services Market Forecast and Opportunities"

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

Recommended Reading: "Predicts 2018: Intelligent Automation Services Will Force Providers to Reinvent Themselves"

"Forecast Analysis: IT Services, Worldwide, 1Q18 Update"

"Explore the Potential of Artificial Intelligence in Integration Platforms"

"Market Trends: Driving Customer Management Business Growth Through 2019"

"Market Guide for Innovation Management Tools"

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

Recommended Reading: "Magic Quadrant for Analytics and Business Intelligence Platforms"

"Cool Vendors in Analytics"

"Predicts 2018: Analytics and BI Strategy"

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

Recommended Reading: "Architecture of Conversational Platforms"

"Market Insight: How to Collaborate and Compete in the Emerging VPA, VCA, VEA and Chatbot Ecosystems"

"Conversational AI to Shake Up Your Technical and Business Worlds"

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)

Recommended Reading: "Market Guide for AI Portfolio Cloud Service Providers"

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

Recommended Reading: "Market Guide for Compute Platforms"

"Three Elements of a Scalable Enterprise Machine Learning Infrastructure Strategy"

"Find the Right Accelerator for Your Deep Learning Needs"

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

Recommended Reading: "How Virtual Assistants, Immersive Experiences and Robots Will Impact Your Organization"

"10 Critical Components Driving the Robot and Drone Revolution"

"Five Ways Vendors Can Grow Their Personal Assistant Robot Business Opportunity"

"Top 10 AI and Sensing Technology Capabilities for Personal Assistant Robots in 2020"

"Top 10 Strategic Technology Trends for 2018: A Gartner Trend Insight Report"

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

Recommended Reading: "Conversational AI to Shake Up Your Technical and Business Worlds"

"Architecture of Conversational Platforms"

"Market Insight: How to Collaborate and Compete in the Emerging VPA, VCA, VEA and Chatbot Ecosystems"

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

Recommended Reading: "A Guide to AI in Postmodern ERP Strategies"

"Transform Talent Acquisition With Artificial Intelligence and Machine Learning"

"Applying AI in the Digital Workplace"

"Use Digital Workplace Programs to Augment, Not Replace, Humans With AI"

"How to Get Started, and Grow, With AI Enhancements in CRM"

"The AI Leader's First 100 Days"

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

Recommended Reading: "Digital Ethics, or How to Not Mess Up With Technology, 2017"

"The CIO's Guide to Digital Ethics: Leading Your Enterprise in a Digital Society"

"How to Apply Gartner's Digital Humanism Manifesto"

"Workplace Analytics Needs Digital Ethics"

"The #DigitalSociety Requires a Digital Social Contract"

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

Recommended Reading: "Combine Predictive and Prescriptive Analytics to Drive High-Impact Decisions"

"Best Practices for Designing Your Data Lake"

"Critical Capabilities for Retail Assortment Management Applications"

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

Recommended Reading: "Forecast Snapshot: Prescriptive Analytics, Worldwide, 2017"

"Predicts 2018: Analytics and BI Strategy"

"Market Guide for Supply Chain Analytics Technology, 2018"

"The Evolving Capabilities of Analytics and Business Intelligence Platforms"

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

Recommended Reading: "Innovation Insight for Deep Learning"

"Cool Vendors in AI Core Technologies"

"Cool Vendors in Data Science and Machine Learning, 2017"

"Artificial Intelligence Primer for 2018"

"Predicts 2018: Analytics and BI Strategy"

"Preparing and Architecting for Machine Learning"

"Magic Quadrant for Data Science and Machine-Learning Platforms"

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

Recommended Reading: "Market Trend: VPA Speakers, Worldwide, 2017"

"Market Insight: Selling Connected Home Solutions to Multifamily Housing Developers"

"Market Insight: How Sensors Drive New Interactions in the Future Connected Home"

"Forecast Snapshot: VPA-Enabled Wireless Speakers, Worldwide, 2016-2021"

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

Recommended Reading: "Magic Quadrant for Data Science and Machine Learning Platforms"

"Critical Capabilities for Data Science and Machine Learning Platforms"

"Machine Learning: FAQ From Clients"

"How to Start a Machine-Learning Initiative With Less Anxiety"

"Five Ways Data Science and Machine Learning Deliver Business Impacts"

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

Recommended Reading: "Market Guide for Social Analytics for Marketing Leaders"

"Cool Vendors in AI for Conversational Platforms, 2017"

"Cool Vendors in AI Core Technologies"

"A Framework for Applying AI in the Enterprise"

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

Recommended Reading: "Robotic Process Automation: Eight Guidelines for Effective Results"

"Market Guide for Robotic Process Automation Software"

"Use Cases for Robotic Process Automation: Providing a Team of 'Virtual Workers'"

"Achieving Excellent Business Outcomes via Business Process Outsourcing and Captive Shared-Service Centers"

"Cool Vendors in Smart Machines, 2016"

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

Recommended Reading: "Market Guide for Virtual Customer Assistants"

"When Will AI Virtual Support Agents Replace Your IT Service Desk?"

"Use Master Chatbots to Improve Conversational Experiences"

"Four Channels for Conversational Technologies"

"Top 10 Strategic Technology Trends for 2018: Conversational Platforms"

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

Recommended Reading: "Market Guide for Robotic Process Automation Software"

"Innovation Insight for Artificial Intelligence in Life and P&C Insurance"

"Market Guide for Compute Platforms"

"Emerging Applications of AI for Healthcare Providers"

"Market Insight: The Road to AI — A Journey to Smarter Security and Risk Decision Making"

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)

Recommended Reading: "Market Guide for Compute Platforms"

"Three Elements of a Scalable Enterprise Machine Learning Infrastructure Strategy"

"SWOT: Xilinx, PLD/FPGA Market, Worldwide"

"Market Insight: Top Five Data Center Trends Driving Future Semiconductor Vendor Business"

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

Recommended Reading: "Cool Vendors for AI in Computer Vision"

"Digital Disruption Profile: Computer Vision Sharpens Focus on AI Strategy"

"Two Ways That Consumer Goods Companies Are Using Computer-Vision-Enabled AI Tech to Improve Their Unified Commerce Strategy"

"Smart Vision: Seven Steps to Get Started With Image and Video Analysis in Your Organization"

"Smart Vision Systems Promise a Lot, but Are Difficult to Adopt Successfully"

"Market Trends: Use Artificial Intelligence Computer Vision in Personal Technology Products"

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

Recommended Reading: "Combine Predictive and Prescriptive Analytics to Drive High-Impact Decisions"

"Machine-Learning and Data Science Solutions: Build, Buy or Outsource"

"Magic Quadrant for Data Science and Machine-Learning Platforms"

"Critical Capabilities for Data Science and Machine Learning Platforms"

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

Recommended Reading: "Hype Cycle for Connected Vehicles and Smart Mobility, 2017"

"Hype Cycle for Automotive Electronics, 2015"

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

Recommended Reading: "Changing Drone Regulations Are a Major Market Opportunity for Technology and Service Providers"

"10 Critical Components Driving the Robot and Drone Revolution"

"Ready-for-Development Semiconductor Solutions Massively Reduce Drone Market Entry Barriers"

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

Recommended Reading: "Competitive Landscape: Augmented Reality Tools for Enterprise, 2018"

"Market Guide for Augmented Reality"

"Market Insight: AR and VR Adoption Limited by Current Technology"

"Create More Compelling Immersive Experiences With Artificial Intelligence"

"Forecast: Wearable Electronic Devices, Worldwide, 2017"

"Competitive Landscape: HMDs for Augmented Reality and Virtual Reality"

"Three Key Development Practices to Implement Effective Enterprise Augmented Reality Applications"

"Best Practices for Using Augmented Reality in Mobile Apps"

"Immersive Technologies Offer Infinite Possibilities"

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

Recommended Reading: "Use These Three Techniques to Achieve Success With ITSM Knowledge Management"

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

Recommended Reading: "Top 10 Strategic Technology Trends for 2018: Immersive Experience"

"Disruption Profile: Immersive AR and VR Technologies Transform Computing Experiences"

"Predicts 2018: Immersive Technologies and Devices Will Transform Personal and Business Interactions"

"Best Practices for Virtual Reality in Higher Education"

"Market Insight: Mixed-Reality Immersive Solutions Are the Ultimate User Experience for Everyone"

"Getting Started Developing Virtual Reality Experiences"

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

Recommended Reading: "Market Guide for Compute Platforms"

"Three Elements of a Scalable Enterprise Machine Learning Infrastructure Strategy"

"Find the Right Accelerator for Your Deep Learning Needs"

"Forecast: Discrete GPUs, Worldwide, 2016-2022, 2Q18 Update"

"SWOT: Nvidia, Data Center GPU Market, Worldwide, 2017"

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

Recommended Reading: "Magic Quadrant for Data Science and Machine-Learning Platforms"

"Critical Capabilities for Data Science and Machine Learning Platforms"

"Seek Diversity of People, Data and Algorithms to Keep AI Honest"

"Augmented Analytics Is the Future of Data and Analytics"

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 (see "Competitive Landscape: Speech-to-Text Applications"). 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

Recommended Reading: "Competitive Landscape: Speech-to-Text Applications"

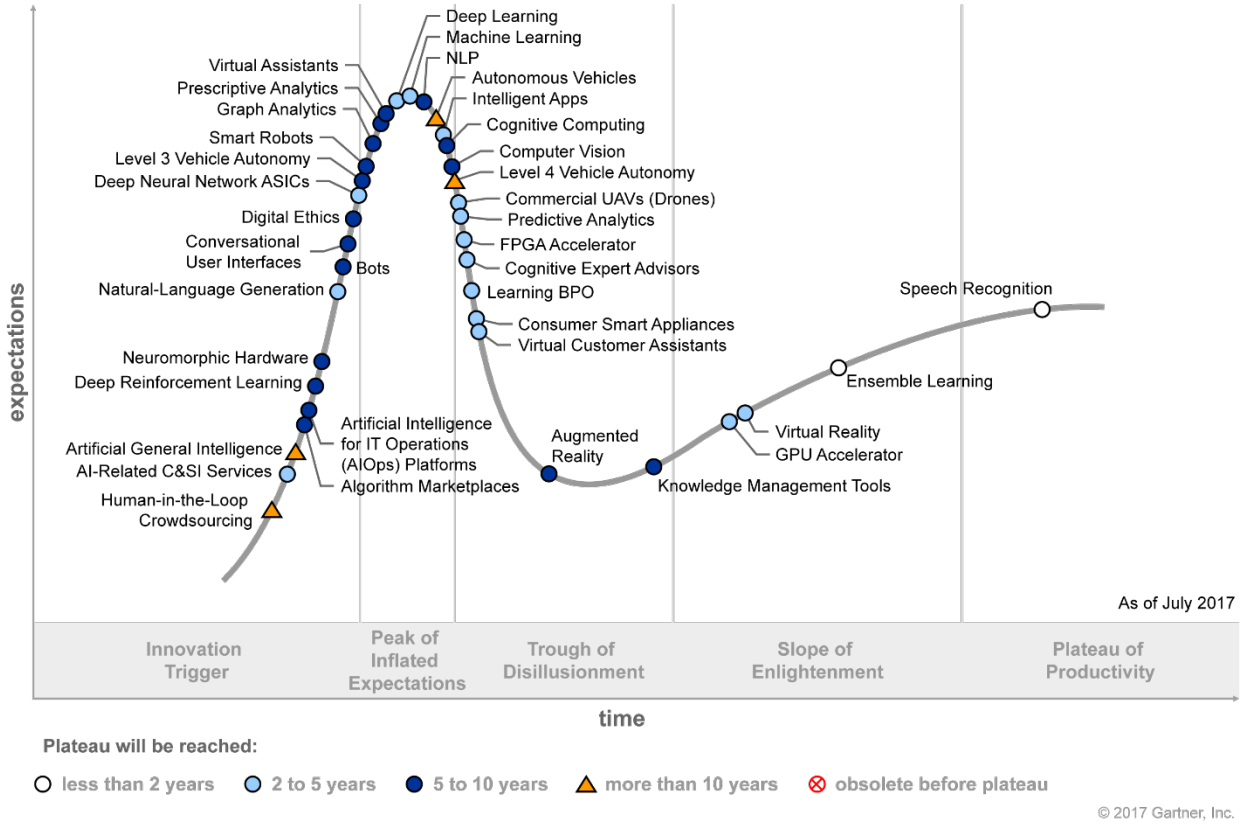
"Deep Learning Enables a Quantum Leap in Content Processing"

"Architecture of Conversational Platforms"

"Market Insight: How to Collaborate and Compete in the Emerging VPA, VCA, VEA and Chatbot Ecosystems"

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)

Gartner Recommended Reading

Some documents may not be available as part of your current Gartner subscription.

"Understanding Gartner's Hype Cycles"

"A Framework for Applying AI in the Enterprise"

"Where You Should Use Artificial Intelligence — and Why"

"Craft an Artificial Intelligence Strategy: A Gartner Trend Insight Report"

"Applying Artificial Intelligence to Drive Business Transformation: A Gartner Trend Insight Report"

"Deliver Artificial Intelligence Business Value: A Gartner Trend Insight Report"

"Ten Ways AI Will Appear in Your Enterprise — No One Source Can Meet All Your Business Needs"

"The AI Leader's First 100 Days"

"Predicts 2018: AI and the Future of Work"

"AI Technical Maturity for Enterprise Architects and Technology Innovators"

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.

More on This Topic

This is part of an in-depth collection of research. See the collection:

- Don't Let Artificial Intelligence's Immaturity Stop You Exploiting Its Potential: A Gartner Trend Insight Report

GARTNER HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
USA
+1 203 964 0096

Regional Headquarters
AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

For a complete list of worldwide locations,
visit http://www.gartner.com/technology/about.jsp