Hype Cycle for Data Science and Machine Learning, 2019

Published 6 August 2019 - ID G00369766 - 121 min read

ARCHIVED This research is provided for historical perspective; portions may not reflect current conditions.

By Analysts Shubhangi Vashisth, Alexander Linden, Jim Hare, Peter Krensky

Initiatives: Analytics, BI and Data Science Solutions

Data and analytics leaders should use this Hype Cycle to understand key trends and innovations, including those related to improving expert and citizen data scientist productivity, implementing new algorithms for cutting-edge use cases, and scaling and operationalizing data science projects.

Newer version of this document

28 July 2020 Hype Cycle for Data Science and Machine Learning, 2020

More on This Topic

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

- 2019 Hype Cycles: 5 Priorities Shape the Further Evolution of Digital Innovation: A Gartner Trend Insight Report

Analysis

What You Need to Know

Data science and machine learning continue as top priorities driven by strategic initiatives such as digital business transformation with intelligence at the center. According to survey data gathered for Gartner's 2018 "Magic Quadrant for Data Science and Machine Learning Platforms," 77% of senior executives view data science as delivering significant value or being essential to the success of their organizations. Virtually every aspect of data management, analytic content, application development and sharing of insights is beginning to incorporate machine learning (ML) and artificial intelligence (AI) techniques. These are used to automate or augment manual tasks, analytical processes and human insight to action for a range of user roles. Enhanced intelligence in all data and analytics platform components will democratize the skills required to exploit these capabilities and scale adoption across the enterprise to unprecedented levels.

This is a period of both education and experimentation. This Hype Cycle is especially relevant to data and analytics leaders, chief data officers and heads of data science teams who are implementing machine learning programs and looking to understand current and next-generation innovations. Technology providers, product marketers and strategists can also study this Hype

Cycle for product roadmaps. Some technologies will be worth investing in right now. Others may be too cutting-edge, expensive or overhyped to consider. And some that are climbing the Slope of Enlightenment may be immensely beneficial in specific use cases.

To succeed with data and analytics initiatives, enterprises must develop a holistic view of critical technology capabilities. There are eight Hype Cycles for 2019 that cover the technologies, architectures and frameworks for data and analytics. Together, they contain the necessary elements for data and analytics leaders to form this holistic view.

Hype Cycles Covering Data and Analytics

"Hype Cycle for Artificial Intelligence, 2019"

"Hype Cycle for Analytics and Business Intelligence, 2019"

"Hype Cycle for Back-Office Analytic Applications, 2019"

"Hype Cycle for Customer Experience Analytics, 2019"

"Hype Cycle for Data Management, 2019"

"Hype Cycle for Enterprise Information Management, 2019"

"Hype Cycle for Information Governance and Master Data Management, 2019"

The Hype Cycle

This year's Hype Cycle reflects four key trends impacting the data science and machine learning market:

- Augmented machine learning
- Data readiness and management
- Scaling and operationalization
- Decision management

The Peak of Inflated Expectations remains dense with many clustered technologies. The Trough of Disillusionment remains sparse, with a few widely used technologies such as Spark and Python. However, several technologies are climbing toward Plateau of Productivity, most notably, notebooks, which has moved rapidly owing to extensive usage. Widespread adoption of predictive analytics pushed it to the Slope of Enlightenment, but prescriptive analytics is visible on many agendas because business users who were exposed to predictions want to know also what to do about them.

Data science and machine learning talent remains in short supply. Yet, efforts to combat the shortage involve education, upskilling and innovations like citizen data science, augmented analytics and automated machine learning (autoML) — all of which are at the Peak of Inflated Expectations.

Advanced ML operationalization capabilities that can handle hundreds or thousands of models in production remain a key trend, dubbed as MLOps, which is a new entrant. Assessing impact of data science and ML technologies on business decisions is increasingly applicable to the entire ecosystem of decision models. Adaptive ML, a new entrant this year, reflects this trend and holds the promise of truly self-learning systems.

Natural language and computer vision capabilities continue to be enablers of new use cases. Transfer learning, advanced video/image analytics and generative adversarial networks are innovations supporting this trend. The popularity of open-source tools and libraries for data science and machine learning has reached critical mass and is now the de facto standard going forward. While Python and Spark have entered the Trough of Disillusionment, there are several new entrants to the Hype Cycle, including AI cloud services. There is a huge hype surrounding explainable AI, especially driven by industries that are highly regulated.

Several high-level trends continue to generate attention and enthusiasm for technologies in the spotlight:

New Entrants

- **Advanced video/image analytics:** This uses deep neural networks and advanced data modelling. It replaces traditional video/image analytics, which is rule-based and has already reached early mainstream adoption.
- **AI cloud services:** This is an important entry to track, as droves of developers can now incorporate machine learning and AI into their systems and applications.
- **Adaptive ML:** Adaptive ML is the capability of frequently retraining ML models when they are online, which would enable more autonomous systems that are responsive to ecosystem dynamics.
- **MLOps:** This is an important entry to track as it aims at streamlining the deployment, operationalization and execution of ML models. This capability is critical to the success of data science and machine learning projects and has overtaken the traditional model management, which is now at the Plateau of Productivity.
- **Synthetic data:** Synthetic data is utilized in use cases where the available data is limited, incomplete or cannot be sourced easily. Simulation and generative techniques can be used to increase the available training data.

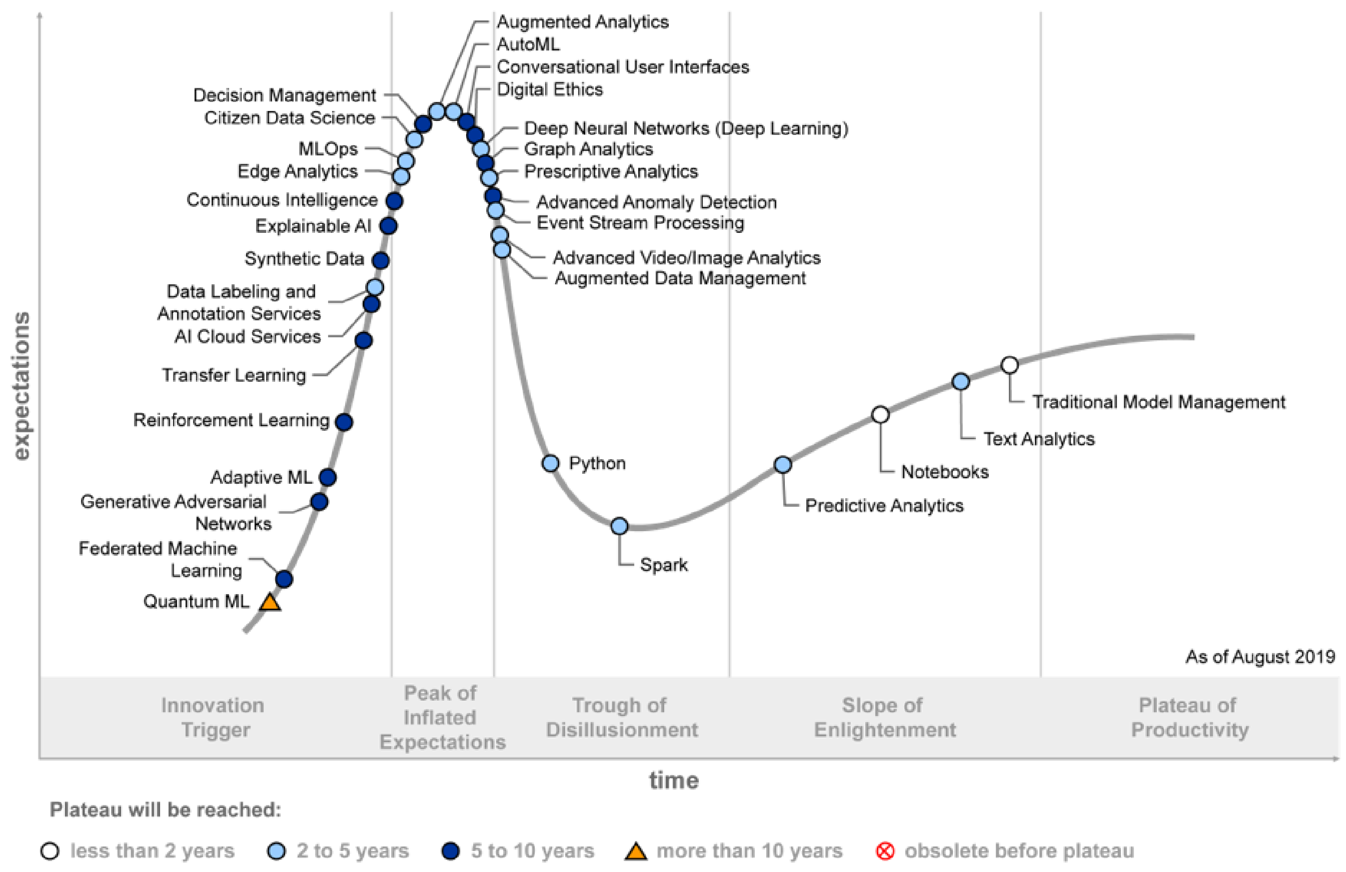
- **Transfer learning:** This remains a prominent discipline for reuse of previously trained machine learning models as an advanced starting point for new purposes in order to reduce the required learning time.
- **Reinforcement learning:** This technology was on the Hype Cycle in the past but was removed due to low interest. However, we have seen renewed interest in the technology through its applications in gaming and robotics, among others.
- **Federated ML:** This entry highlights an important innovation in (re)training ML algorithms in a decentralized environment. It holds the potential to deal with significantly higher amounts of data while addressing the challenges related with privacy.
- **Generative adversarial networks:** Also known as “GANs” these are composed of two neural network models that create original simulations of objects such as videos, music, text and others. This technology holds the potential to augment human talents for many creative tasks across industries.
- **Quantum ML:** This technology is currently embryonic, but captures a lot of interest as it uses algorithms that exploit quantum mechanical properties to outperform classical computing-based approaches to ML.
- **Explainable AI:** This innovation garners a lot of hype as it describes a model, highlights its strengths and weaknesses, predicts its likely behavior, and identifies any potential biases.

Name Changes

- **Traditional model management (formerly model management):** The new name better captures the current market scenario where we see a surge in MLOps and vendors that offer comprehensive capabilities for the latter.
- **Data labeling and annotation services (formerly human-in-the-loop crowdsourcing):** The name change reflects the broader market for such services.
- **Augmented data management (formerly machine learning-enabled data management):** This simplified name better reflects the current vernacular.

Figure 1. Hype Cycle for Data Science and Machine Learning, 2019

Hype Cycle for Data Science and Machine Learning, 2019



Source: Gartner
ID: 369766

The Priority Matrix

Figure 2 displays the associated Priority Matrix, which arranges each Hype Cycle entry in relation to two dimensions: maturity and business impact. It shows the level of benefit attainable relative to an innovation's progression along the Hype Cycle. Note, however, that impact is not the only factor to consider when selecting vendors and products, including open-source tools — applicability, budget, time to implement and receive payback, and return on investment are also important.

Innovations of "transformational" benefit will change the world of data science by using automation and embedded machine learning to simplify and accelerate the building, tuning and deployment of models. This will reduce dependence on traditional data scientists and lower the barriers to starting data science projects. As more deep neural networks are offered prepackaged as services, these will offer disruptive potential across industries.

Innovations of "high" benefit, such as autoML, MLOps and edge analytics, will have a significant impact on data science programs during the next two to five years. Predictive analytics and prescriptive analytics will help organizations mature beyond descriptive and diagnostic analytics.

Organizations will also use digital ethics thinking to proactively determine whether applying data science to particular use cases is the right thing to do.

Figure 2. Priority Matrix for Data Science and Machine Learning, 2019

Priority Matrix for Data Science and Machine Learning, 2019

benefit	years to mainstream adoption: less than 2 years	2 to 5 years	5 to 10 years	more than 10 years
transformational		Advanced Video/Image Analytics Augmented Analytics Citizen Data Science Deep Neural Networks (Deep Learning) Event Stream Processing	AI Cloud Services Continuous Intelligence Conversational User Interfaces Generative Adversarial Networks	Quantum ML
high	Notebooks Traditional Model Management	Augmented Data Management AutoML Edge Analytics MLOps Predictive Analytics Prescriptive Analytics	Adaptive ML Decision Management Digital Ethics Explainable AI Federated Machine Learning Graph Analytics Reinforcement Learning Synthetic Data	
moderate		Data Labeling and Annotation Services Python Spark Text Analytics	Advanced Anomaly Detection Transfer Learning	
low				

As of August 2019

Source: Gartner
ID: 369766

Off the Hype Cycle

Since this Hype Cycle is intended to provide insight into only the most important and prevalent concepts in the area of data science and machine learning, the following entries that featured in "Hype Cycle for Data Science and Machine Learning, 2018" have been removed from this year's edition. However, a few remain relevant and important in the field of analytics, and appear in other Hype Cycles:

- **Machine learning:** The past few years saw a surge in adoption of machine learning technologies and many organizations have moved past the pilot stage to deploy ML solutions in production and realize business benefits. All of the technologies captured in this Hype Cycle include data science and machine learning components, which is why this entry was dropped.

- **Artificial general intelligence:** This entry has been removed as the innovation is more relevant to the "Hype Cycle for Artificial Intelligence, 2019."
- **Guided analytics**: The hype about guided analytics has been subsumed by hype about augmented analytics and AutoML. This technology was obsolete before the Plateau last year.
- **Cognitive computing:** This entry has been removed because of a lack of client interest and adoption.
- **Ensemble learning:** This technology has moved beyond the Plateau of Productivity into mainstream adoption.
- **Data lakes:** This innovation profile remains relevant to clients for data management, however we also see a growing interest in other data management strategies captured in the "Hype Cycle for Data Management, 2019."
- **Data preparation:** As a part of the DS/ML project life cycle, this is reflected in multiple other technologies such as synthetic data, data labeling and annotation services among others.
- **Embedded analytics:** This innovation reflects the consumerization of analytics — embedded into applications where it provides more contextualized insights. Hence, it has been removed as it is more relevant to the "Hype Cycle for Analytics and Business Intelligence, 2019."
- **Optimization:** This entry has been subsumed into the broader category for prescriptive analytics.
- **Real-time analytics:** The hype about real-time analytics has been subsumed by event stream processing and continuous intelligence.

On the Rise

Quantum ML

Analysis By: Chirag Dekate

Definition: Quantum machine learning (ML) systems use algorithms that exploit quantum mechanical properties including superposition and entanglement to outperform classical-computing-based approaches for machine learning. Currently, only a select set of ML algorithms have been theoretically accelerated using quantum computers.

Position and Adoption Speed Justification: Quantum ML algorithms are at an early trigger stage, with active research and development underway. Early research in developing quantum ML indicates potential for applicability across a growing set of ML algorithms. These include: K-means, K-medians, Hierarchical clustering, principal component analysis, neural networks, support vector machines, nearest neighbors, regression and boosting. Considerable hardware and software challenges, however, still remain.

A few major barriers exist to the deployment of Quantum ML:

- **Nascent quantum computing ecosystem:** Quantum computing is still at a very early stage of development, and most of the research projects are very basic.
- **Data encoding:** Although quantum computing can hypothetically deliver dramatic boosts, one of the challenges is encoding input data. For quantum ML to work at scale, large amounts of data need to be encoded and loaded into the quantum system.

R&D today is focused on developing different quantum algorithms for ML kernels. Vendors like IBM have prototype ML algorithms implemented for a very select use cases. Developing scalable ML systems will require many qubits and algorithms.

User Advice: Quantum ML has the potential to deliver extreme capabilities for common ML algorithms including clustering, support vector machines and generative adversarial networks. Most of the innovation and R&D today is centered on developing quantum ML kernels. Quantum computing-based learning classifiers leverage Quantum ML as discriminators, i.e., training data mapped into a Quantum state. Few vendors like IBM have implemented small-scale quantum algorithms on their quantum environment. Quantum ML is theorized to work effectively in noisy intermediate scale quantum computers (NISQ). Quantum ML is not ready for mainstream adoption today, however these systems have the potential to become viable over the next five to 10 years. Prepare for Quantum ML by:

- Partnering with quantum mechanics experts to devise new ML algorithm kernels that can use superposition and entanglement to deliver quadratic or exponential speedup.
- Exploring quantum-as-a-service capabilities for validating hypothesis involving Quantum ML to minimize risk and maximize accessibility of quantum computing resources.

Business Impact: Quantum ML is in an embryonic stage, with most R&D activities clustered around devising quantum algorithms for key ML kernels. ML techniques include principal component analysis, support vector machines, clustering and GANs. In theory, applications based on these kernels can benefit, however the scale of the systems, struggle scaling algorithms and challenges associated data “loading” will limit adoption in the near term. Over the next decade, assuming the algorithms and quantum systems mature, we expect some applications involving compatible kernels can benefit.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: IBM; Microsoft; Xanadu

Recommended Reading: "Top 10 Strategic Technology Trends for 2019: Quantum Computing"

"2019 Strategic Roadmap for Compute Infrastructure"

Federated Machine Learning

Analysis By: Saniye Alaybeyi; Alexander Linden; Pieter den Hamer

Definition: Federated ML is the capability that enables local learning by (re)training ML models in a decentral runtime environment, without the need to share local data centrally or in a common runtime environment. Local learning may occur in smartphones, softbots, (semi)autonomous vehicles or IoT edge devices. In addition, federated ML enables collaborative learning by sharing local *model* improvements at a central level, that are then used to generate a new, improved common model that may be (re)deployed to decentral or local runtime environments.

Position and Adoption Speed Justification: Federated ML deserves a spot on the Hype Cycle starting this year due to proliferation of AI hardware and software technologies that support edge AI use cases, where the data is generated. The adoption of federated ML is currently mostly driven by; (1) the need to protect privacy of local data and (2) the difficulty to collect and store big data centrally. Other drivers include the collaborative and parallel solving of complex problems, the adaptability of ML to local context and conditions, the prevention of data sharing between competing actors, the parallel processing of extremely large datasets for ML and the leveraging of computing resources. A main challenge in federated ML is the interpretation and combination of local model improvements into a new central model. There are also concerns to what extent local model improvements may still be used to infer insights that are privacy sensitive. Related trends include multiagent systems and swarm intelligence.

User Advice: Apply federated ML to create and maintain decentral smart services or products, while protecting the privacy of their users and preventing the need to centrally collect massive amounts of data.

- Give a head start to decentral ML applications by deploying a common, centrally pretrained model, e.g., by doing simulation-based reinforcement learning, while still providing personalization and contextualization by locally retraining the model based on local data and feedback.
- Enable continuous improvement of decentral ML applications with collaborative learning by repeatedly collecting local model improvements to create a new improved central model and then redeploying this for decentral usage and fine-tuning.
- Keep in mind, many use cases today do not require federated ML. No major platform vendor has federated ML offering for end users.

- Keep a central reference model to ensure a "cognitive cohesion" across distributed models — i.e., by avoiding decentralized models to veer off too far from its original purpose.

Business Impact: The benefits of federated ML are likely to be high, as it helps to protect privacy, empowers ML and specifically DNNs to deal with larger amounts of data, and enables collaborative learning.

Federated ML is unlikely to reach the Plateau of Productivity in the next 5 to 10 years, with ongoing research and expected improvements. Sensor data resides and is collected at the edge/network edge. Due to scalability issues, excessive power consumption, connectivity and latency we see a move toward edge infrastructure in the form of federated computing architectures. This will impact AI initiatives as well.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Cray; Dell; Google; IBM; Uber

Recommended Reading: "Market Guide for Machine Learning Compute Infrastructures"

"Deploying IoT Analytics, From Edge to Enterprise"

"Raise Your Product Innovation Quotient With Edge AI"

"Use Edge AI to Drive Revenue Growth, Forecasting, Customer Engagement and Workforce Planning"

Generative Adversarial Networks

Analysis By: Brian Burke

Definition: Generative Adversarial Networks (GANs) are composed of two neural network models, a generator and a discriminator that work together to create original simulations of objects such as videos, images, music and text (poetry, stories, marketing copy) that replicate authentic objects or their pattern, style or essence with varying degrees of quality or realism.

Position and Adoption Speed Justification: Originally proposed by Ian J. Goodfellow in 2014, this technology is in a nascent state, with the majority of applications coming from research labs. Commercial applications have just started being explored. The algorithms require a lot of manual tuning to make them perform in the desired manner, and development of the technology is constrained by the extremely limited resources that have knowledge in this area. As commercial applications become more commonplace, the technology will improve as the benefits are significant.

GANs are commonly used to create of images of people who don't exist, to create fake political videos, to compose music and poetry. In 2018, an image produced by a GAN was sold at an auction for $432,500. While these 'novelty' applications are prominent, research is underway to apply these algorithms to far more valuable challenges such as generating marketing content, graphic designs, designing new parts (generative design), generating pharmaceutical compounds, creating simulated environments for training autonomous vehicles and robots and generating synthetic data to train neural networks and to protect privacy.

User Advice: In a GAN, the generator is essentially a deep learning classifier running in reverse to create an object. It then presents it to the discriminator which predicts if the object is real or not, providing this information to the generator. This feedback loop provides guidance to the generator to adjust the objects it is creating to become more like what the discriminator predicts is real. Through many iterations, the generated object becomes indistinguishable from a real object for the discriminator.

Technology innovation leaders in organizations that are technologically innovative should evaluate the potential for leveraging this technology today, and partner with universities to conduct proof of concepts where the potential benefits are significant. Tech innovation leaders should do their due diligence and consider the fact that while the core technologies are readily available in the public domain, the technology is brittle, resource hungry and requires significant (and rare) AI skills. They should also focus on other pressing issues such as explainability, as GANs are 'black boxes' and there is no way to prove the accuracy of the objects produced other than by subjective methods.

Business Impact: Commercial implementations of GAN technology are limited to experimental uses today. The powerful idea is that deep neural network classifiers can be modified to generate realistic objects of the same type. GANs have the potential to impact many creative activities from content creation (art, music, poetry, stories, marketing materials) to many types of design (architecture, engineering, drug, fashion, graphic, industrial, interior, landscape, lighting, process). GANs might also be used to create simulations where actual data may be difficult to obtain (training data for machine learning) or pose a privacy risk (health data) or be costly to produce (video game backgrounds). GANs have the potential to augment humans' talents for many creative tasks across many industries.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: Amazon; Autodesk; DeepMind; Microsoft; Neuromation; NVIDIA

Recommended Reading: "Market Insight: Creative AI — Assisted and Generative Content Creation"

"Embrace These 3 Key Trends in Content Marketing"

Adaptive ML

Analysis By: Pieter den Hamer; Erick Brethenoux

Definition: Adaptive machine learning (ML) is the capability of frequently retraining ML models when they're online in their runtime environment, rather than only training ML models when they're offline in their development and test environment. This capability allows ML applications to adapt to changing or new real-world circumstances that were not foreseen or available during development. Based on user feedback or data about the quality of the ML output, e.g., prediction errors, the model parameters are updated while online.

Position and Adoption Speed Justification: Adaptive ML is just emerging and in many cases still embryonic. It gets much closer to self-learning, or at least to more frequent learning, in contrast with current AI applications that after deployment only use static ML models that depend on infrequent redeployment of new model updates to improve themselves. Adaptive ML is technically challenging for a number of reasons, including:

- A closed loop between the output and input of the ML application is required, including a mechanism like user or expert feedback, or a quantitative measure to determine the quality of the ML output and the extent to which the ML model should be updated to enforce or improve current model parameter settings.
- Adaptive ML typically occurs frequently in its runtime environment — less frequent model updates can already be achieved by the current approach of offline training and periodic model update deployments. This implies that there is no time to fully retrain the model, using all available historical data and newly acquired closed loop data. Instead, the model must be retrained online and incrementally, using only new or most recent data, which requires incremental learning algorithms that are different from offline learning algorithms that typically rely on large batches of historical data.
- Adaptive ML must be tuned in terms of weighting new data versus older data that was used for earlier online or offline training and other challenges such as preventing overfitting and proper testing and validation, at least periodically.

Non-technical challenges include ethical, societal, reliability, liability, safety and security concerns that come with self-learning and autonomous systems.

User Advice: Adaptive ML is a key enabler of autonomous systems such as self-driving vehicles or smart robots that should be able to operate in their ever-changing contexts. Also, increasing pace and dynamics in business ecosystems will require more adaptive ML to power continuous intelligence, streaming analytics, decision automation and augmented intelligence in a myriad of industries and business areas.

Adaptive ML should not be considered by organizations to replace but to complement current ML. Most adaptive ML applications will start out with a model that was trained offline. The other way around, adaptive ML can be seen as a way to further improve, contextualize, personalize or fine-tune the quality of ML models, once online.

In addition, adaptive ML can be used to compensate for limited availability of historical or synthetic data, inhibiting offline training during development. Likewise, adaptive ML can be used to reduce the need for extensive simulations or trials for reinforcement learning, before deployment. In both cases, the adaptive ML application starts out with a minimal viable model that was pretrained offline, with the model then incrementally improved during the actual online usage.

Adaptive ML must be accompanied by a proper risk analysis and risk mitigation activities, if only to frequently monitor the quality and reliability of adaptive ML applications.

Adaptive ML relates to and works together with MLOps: apply MLOps to deploy new ML models from their development and test environment to their runtime environment, after which adaptive ML is applied to continue learning. Once deployed MLOps monitors quality and business relevance of the online adaptive ML application. If MLOps detect that the adaptive ML application is no longer meeting quality or other criteria, MLOps should take the adaptive ML application offline, overhaul it in a development environment, test and redeploy, after which adaptive ML takes over again and MLOps returns to monitoring.

Organizations should actively monitor the potentially significant impact of adaptive of ML on talent, infrastructure and enabling technology. For example, adaptive ML is likely to be more demanding in terms of computing power in runtime environments and will require the development of knowledge about new (incremental learning) algorithms and tools.

Business Impact: Adaptive ML will allow organizations to respond more quickly and effectively to changes in their business ecosystem by using more autonomous systems that are responsive to the dynamics. Adaptive ML will also help organizations to improve the quality of their ML applications by providing feedback after their deployment, thus implementing a more frequent "learning" mechanism. Adaptive ML is most relevant in areas in which context and conditions or in which the behavior or preferences of actors change frequently. Example, application areas include gaming, organized crime fighting and anti-terrorism, fraud detection, cyber security, quality monitoring in manufacturing, virtual personal assistants, semi(autonomous) cars and smart robotics.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Cogitai; IBM (Neural-Symbolic research); Microsoft; Tazi.ai

Recommended Reading: "Artificial Intelligence Hype: Managing Business Leadership Expectation"

"Building Your Continuous Intelligence Capability for Digital Transformation"

"Preparing and Architecting for Machine Learning: 2018 Update"

Reinforcement Learning

Analysis By: Alexander Linden; Erick Brethenoux; Shubhangi Vashisth

Definition: Reinforcement learning (RL) is a type of machine learning that is based on rewarding desired behaviors and/or punishing undesired ones. It is considered as one of the three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

Position and Adoption Speed Justification: Reinforcement learning has been around for more than three decades. RL can be considered when other ML approaches are simply not possible. There are three major aspects driving the revival of interest in RL:

1. The fascination of the RL framework, which involves much less training data and supervision than currently dominant supervised learning scheme.

2. Recent successes in computer-based game playing, e.g., DeepMind's AlphaGo and AlphaStar over top-ranked professional players of StarCraft by Blizzard Entertainment.

3. Commercial vendors coming up with RL components (Microsoft's acquisition of Bonsai, Amazon's reinforcement learning, and CogitAI) — but also semicommercial institutions activities like OpenAI.

However, RL is still in its early stage:

- Most current Data Science and Machine Learning (DSML) platforms don't have RL capabilities.
- The computational requirements are much more substantial than most similar supervised learning scenarios.
- Mostly, RL is only applicable in areas that can be fully simulated (e.g., in games and control situations) or that are quite stationary (meaning that the environment is not subject to constant change) or where massive amounts of relevant data or staff skills are available.
- Before RL can enter mass adoption, Gartner believes that significantly better simulation capabilities must be developed. Here again the field of DNN is likely to provide a major push in the next five to 10 years.

User Advice: Data and analytics leaders, as well as AI leaders, should understand that:

- Only a few practical applications are available for reinforcement learning — game playing, robotics, predictive maintenance and control engineering are the most prominent.

- RL still requires deep expertise. The application of RL is currently more risky than more traditional techniques and requires experienced data scientists and/or ML experts.
- RL can reduce the need for labeled data, but typically requires a simulation or at least a very stationary environment.
- The technology is still very immature, so being selective while introducing reinforcement learning on your development or deployment roadmaps is critical.

Business Impact: Reinforcement learning has primary potential for gaming and automation industries. It has the potential to deliver incremental efficiency improvements on complex automated processes. It may also lead to significant breakthroughs in robotics, vehicle routing, logistics and other industrial control scenarios.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: Amazon SageMaker RL; CogitAI; Facebook; Google Dopamine

Transfer Learning

Analysis By: Stephen Emmott; Anthony Mullen

Definition: Transfer learning is the reuse of previously trained machine learning (ML) models as an advanced starting point for new purposes in order to reduce the learning time required to attain acceptable performance. Learning can be transferred through unsupervised, inductive and transductive methods. Transfer learning is distinct from other approaches to reducing learning cycle time and improving performance, such as ensemble learning.

Position and Adoption Speed Justification: The reuse of ML models is not new. The theory, techniques and tools behind transfer learning stretch back many years. However, despite the hype of AI touching all enterprises, only a minority have proactively deployed it beyond what is embedded in enterprise software products. Developing and training AI is challenging and risky, with the key barriers to adoption being the skills of staff, understanding benefits and use cases, and scope or quality of data. Given the investment required to succeed, enterprises are looking for ways to leverage existing models. This might be sourced from internal efforts (custom models trained on internal data) or external sources. These could include successful implementations of transfer learning in natural language processing (word2vec), audio/speech (ASR for English — used to accelerate development of other languages) and computer vision (Imagenet).

Transfer learning provides an approach to adoption that has the potential to overcome the barriers to adopting AI, provided that there are sufficient numbers of applicable models available, the tools to reuse the data are in place, and there is enough subsequent retraining on new data to attain successful use. Although challenging, early adopters and providers of AI are now starting to generate models and tools that make transfer learning viable for the majority.

User Advice: Enterprises at all levels of AI maturity should start to use AI for new purposes after asking the question: "Which models can we leverage?" and seek to source these models both internally and externally. Those with a more mature level of AI adoption should additionally ask how their current models might be reused in related domains and similar tasks and facilitate their use through their enterprises' AI center of excellence (CoE).

Making transfer learning habitual is a critical success factor to accelerating enterprisewide ML-based AI adoption. Aside from ensuring the mandate for transfer learning, IT leaders need to ensure the availability of models and tools to facilitate its use. Maintaining internal directories of AI models — with the help of your CoE — and external sources of models or products embedding models is also key.

The tools used to create and train models should be checked for their support of transfer learning, which should also be a requirement for any new tools. A mandate for transfer learning can help to organize approaches to AI into a hierarchy based on whether it is:

- Embedded in enterprise software
- Embedded in a point software application
- An external component integrated into software via APIs
- A custom solution sourced externally
- A custom solution sourced internally

At each level in the hierarchy, you should seek support for transferable learning commensurate with the level of customization required. For example, at the level of an "external component integrated into software via APIs" Google's AutoML facilitates additional training with customer data.

Transfer learning can be facilitated using both internal as well as third-party datasets, but this use is excluded without knowledge of their existence. Within your directory of models, include external datasets, and work with your data analytics leaders to utilize metadata management initiatives to identify internal datasets.

Business Impact: Transfer learning will impact all use of ML, both in terms of how organizations apply this, and the technology they acquire with it embedded. Presently, the primary motivators for enterprises adoption of AI are the improvement of customer service and the automation of repetitive or manual tasks. Enterprises that are prioritizing the application of AI to improved customer experience are already benefiting from transfer learning through speech-to-text and translation services. The shift from symbolic to nonsymbolic and hybrid (a combination of both) techniques in the processing of natural language, combined with transfer learning, is enabling (and will continue to enable) the extraction of richer insights from text-based content, including more accurate sentiment analysis and understanding.

Transfer learning promises the possibility to attain acceptable task performance with less training data. Such capabilities will augment customer-facing staff and elevate the performance of the bots that work alongside them. These capabilities will also lead to progressive automation of work in adjacent information-rich domains, accelerating automation in the enterprise. A consequence of this will be easing the burden of managing many and similar narrow models in the context of the slow arc to general artificial intelligence.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: DeepMind; Google; Hyperscience; Indico; Leverton; Microsoft; NVIDIA

Recommended Reading: "Clarify Strategy and Tactics for Artificial Intelligence by Separating Training and Machine Learning"

"Maverick* Research: Use Simulations to Give Machines Imagination"

AI Cloud services

Analysis By: Van Baker

Definition: AI cloud services are hosted services that allow development teams to incorporate the advantages inherent in AI and machine learning. These services deliver capabilities in language, vision and automated machine learning. Services are often available via API access. They can be used to enhance the value of existing enterprise assets or add capabilities to applications that were previously unavailable. Examples of these services include NLP, sentiment analysis, image recognition and automated machine learning model creation.

Position and Adoption Speed Justification: AI cloud services emerged in the market, relatively, recently but will have a significant impact. The main categories for the services are automated machine learning (AutoML), language and vision services. While low-level services are available, offers often are focused on higher-level building blocks making development significantly easier. This is in part driven by the scarcity of data science and machine learning professionals. Providers have focused the efforts to enable applications with these features on the developer resources in the enterprise. Use of these hosted services will allow development teams to add conversational capabilities, identify people and assets in images or video feeds and automate the building of machine learning models. These machine learning models can be deployed in the public cloud or in containers to classify unstructured data or predict values bringing greater asset utilization to the enterprise as well as other application enhancements.

User Advice: Development teams should be using services such as natural language processing, image recognition and automated machine learning model creation to add incremental capabilities to applications deployed in the enterprise for both workforce and customer-facing applications. These services can add value to any data, structured and especially unstructured, in the enterprise by automatically adding metadata. This in turn makes assets more discoverable, searchable and as a result accessible to the business. Additional machine learning models can be created via these services to predict values such as risk, customer value and supply chain disruptions just to give a few examples. In short, these services can enhance the ability of apps to create business value. Use cases span all parts of the enterprise from security to asset management and transactional capabilities such as:

-Lead scoring — Ranking of customers based on potential to close or total value of the business or risk of churn

-Asset classification — Evaluation of unstructured data such as analysis of contract language or metadata labelling of image assets

-Image recognition — Automated ordering of replacement parts for field service personnel or two-factor authentication of individuals entering a secure facility

-Loss potential — Analysis of factors to assign a risk score to outstanding receivables or to customers seeking insurance coverage

Business Impact: AI (including machine learning) is among the single most impactful technologies affecting businesses today. These services and resulting models allow businesses to unlock value in their data repositories that give more precision to business decisions, facilitate automation of business processes and accelerate workflows to enhance the responsiveness of the business to opportunities and risks surfacing in the business. The ability of language services to make business information available for the asking or text analytics to discover hidden value in text documents or video content analysis to identify where specific people or information appear in a video are just a few examples. Machine learning models will increasingly be able to make more accurate decisions augmenting the workforce with more information applied to complex problems. Most applications will incorporate machine learning and AI cloud hosted services that enhance the value to the workforce, the customers, the partners, and as a result the business too.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Alibaba Group; Amazon; Baidu; DataRobot; Google; H2O.ai; IBM; Microsoft; Salesforce

Recommended Reading: "Market Guide for Hosted AI Services"

"Artificial Intelligence Primer for 2019"

"Magic Quadrant for Data Science and Machine Learning Platforms"

Data Labeling and Annotation Services

Analysis By: Anthony Mullen

Definition: Data annotation and labeling services support enterprises in labeling/annotating data for artificial intelligence (AI) projects. These services and associated platforms route and allocate this work to both internal staff and external third-party knowledge workers.

Position and Adoption Speed Justification: The proto provider in this space was Amazon Mechanical Turk (MTurk) launched in 2005 designed to coordinate labor to perform micro-jobs that computers were unable to perform. As AI adoption has picked up among enterprises, the need for labeled data has dramatically increased in order to remove the bottleneck in developing AI solutions. As a result, offerings in this space have grown to help companies turn their unstructured data into structured data.

Baseline offerings in this space are access to pools of prequalified knowledge workers who can label and annotate training data such as street scenes, speech, music, photos, documents and other assets. Many providers are now beginning to adopt a combination of machine learning techniques and human workers to accelerate the classification and annotation of training data.

Increasingly these solutions are enlarging from preproduction focus to real-time human-in-the-loop solutions designed in real time to call upon a pool of workers (internal and external) to handle automation exceptions where model confidence is low, e.g., classifying and answering customer support questions. Further, annotations, classifications and content provided by third-party knowledge workers can be synchronized back to enterprise platforms such as content management systems, CRM, conversational platforms and knowledge management systems. Challenges remain around the skills that third-party knowledge workers have to annotate the data but are ameliorated somewhat by the development of reputation systems and pre-qualification tests.

While tech heavyweights like Facebook, Amazon, Google and Microsoft have used these providers for a while, many end users are quite unaware that such services exist.

User Advice:

- Ensure the provider you choose has methods to test their pool of knowledge workers for domain expertise and measures around accuracy.

- Carefully estimate the amount of labeled data that will be necessary when investigating deep learning models.
- Allow data scientists to focus their time on more valuable tasks and lighten their load in classifying and annotating data by using these services.
- Use providers with real-time human-in-the-loop solutions for production systems like chatbots and recommenders to handle low confidence thresholds, spikes in demand or access to real-time knowledge not present in the enterprise.
- Design development and production workflows to leverage a mixture of knowledge workers — both internal and external staff.

Business Impact: While the supervised learning approach is predominant in AI, these services will continue to grow in usage. Scenarios that do not require deep domain knowledge of a field or datatype can expand annotation more quickly by using external knowledge workers. While there are many applications for this capability in preproduction environments, the real-time human in the loop solutions where models are continually trained and calibrated, such as chatbots or recommendation engines, will provide ongoing benefit. Business users need to join the human-in-the-loop workflows to route and train handover and moderation tasks to subject matter experts.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Alegion; Amazon SageMaker Ground Truth; Apache Hive; CloudFactory; Directly; Figure Eight; Globalme; Labelbox; Mapillary; Prolific

Recommended Reading: "Individuals, Groups and Society in the Loop of Artificial Intelligence Design and Development"

"Maverick* Research: Use Simulations to Give Machines Imagination"

"Clarify Strategy and Tactics for Artificial Intelligence by Separating Training and Machine Learning"

Synthetic Data

Analysis By: Anthony Mullen; Alexander Linden

Definition: Synthetic data is a class of data that is artificially generated, i.e., not obtained from direct measurements. Generation can use different methods such as statistically rigorous sampling from real data, semantic approaches, Generative Adversarial Networks (GANs) or by creating simulation scenarios where models and processes interact to create completely new datasets of events.

Position and Adoption Speed Justification: One of the major problems with AI development today is the burden in obtaining real-world data and labelling it so that AI may be trained effectively. For example, retail environments contain endless combinations of product, store layout and observable shopper behavior. Also, autonomous vehicles require millions of hours of training data. Synthetic data addresses the problem of volume and variety for sparse, nonexistent or difficult to get data. There are three primary methods we see in the marketplace today:

1. Sampling to generate data according to some distribution

2. GAN architectures that mimic real data

3. Agent-based techniques in simulations

Synthetic data, today at least, is not a panacea — it can have bias problems, miss natural anomalies, be complicated to develop or may not contribute any new information to the existing real-world data. Data quality is tied to the model that developed it. Yet, Gartner believes, the pros outweigh the cons and the vendor landscape is rapidly expanding to meet the demand for more training and test data in AI. Increased adoption of simulation techniques will accelerate this trend.

User Advice:

- Identify areas in your organization where data is missing, incomplete or expensive to obtain, thus, currently blocking AI initiatives. In regulated industries, such as pharma or finance, exercise caution and adhere to rules.

- Where personal data is required but data privacy is a requirement consider synthetic variations of the original data or synthetic replacement of parts of it.

- Begin with the sampling approach and leverage data scientists to ensure statistical validity of the sample and distribution of the synthetic data.

- Mature toward the simulation-driven approach, emphasizing creating agents and processes within a simulation framework to generate permutations of interactions that result in synthetic data.

- Above all, leverage specialist vendors while the technology matures.

Business Impact: Synthetic data promises to accelerate training data development, improve precision, reduce costs and cover more scenarios resulting in a more quickly developed and resilient AI solutions. Further, synthetic data can act as a democratizer for smaller players as they try to compete with data-laden tech heavyweights. Privacy restrictions are an additional major driver of this technology. As a result, we see the benefit as “high.” Today, we see increased adoption of synthetic data approaches across industries — in particular automotive, robotics, content generation, retail, fraud, finance, defense, and logistics — along with early shoots of use in development of synthetic speech and natural language data for NLP applications. Enterprises should expect a rapid growth in the use of these techniques over the next three years.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: AiFi; Apprente; Archarithms; Bitext; Kinetic Vision; Mapillary; May Mobility; Neuromation; Simudyne; TwentyBN

Recommended Reading: "Predicts 2019: The Democratization of AI"

"Maverick* Research: Use Simulations to Give Machines Imagination"

"Top 10 Strategic Technology Trends for 2019: AI-Driven Development"

"How to Develop the Right Technical and Human Architectures for Digital Business"

"Elevating Test Data Management for DevOps"

Explainable AI

Analysis By: Saniye Alaybeyi

Definition: AI researchers define "explainable AI" as an ensemble of methods that make black-box AI algorithms' outputs sufficiently understandable. Gartner's definition of explainable AI is broader — a set of capabilities that describes a model, highlights its strengths and weaknesses, predicts its likely behavior, and identifies any potential biases. It can articulate the decisions of a descriptive, predictive or prescriptive model to enable accuracy, fairness, accountability, stability and transparency in algorithmic decision making.

Position and Adoption Speed Justification: Not every decision an AI model makes needs to be explained. There is considerable hype that is associated with explainable AI today. Although some vendors have introduced early explainable AI capabilities, most are using it for marketing purposes. Therefore, we decided to put explainable AI at prepeak on the Hype Cycle. Gartner anticipates that organizations do and will continue to achieve a lot of fantastic results without the need for full explainability. Depending on the business context, however, privacy, security, algorithmic transparency and digital ethics may demand different levels of requirements for explainability. For example:

- AI that makes decisions about people, such as rejecting a loan application, may require explainability. By law, providers of algorithms must give the user a reason for the rejection.

- According to the EU's GDPR, which took effect in May 2018, users affected by an algorithmic decision may ask for an explanation.
- AI that makes decisions in a closed loop with important consequences, such as autonomous driving, also has a high need for explainability due to ethical and possibly legal reasons.
- The Financial Stability Board (FSB) identified the lack of interpretability of AI and machine learning methods as "a potential macrolevel risk." The same board indicated that a pervasive use of these AI models that lack explainability may result in unintended consequences.
- Explainable AI comes up often during Gartner end-user client inquires, as well as in the news and media. During Gartner vendor briefings, vendors are also starting to claim they have explainable AI available to their customers. The Defense Advanced Research Projects Agency (DARPA) now projects that explainable AI will emerge in transportation, security, medicine, finance, legal and military applications.

User Advice:

- Foster ongoing conversations with various line-of-business leaders, including legal and compliance, to gain an understanding of the AI model's interpretability requirements, challenges and opportunities from each business unit. Integrate these findings into the development of the enterprise information management strategy.
- Build partnerships with IT, in particular with application leaders, to explain how the AI model fits within the overall design and operation of the business solution, and to give stakeholders visibility into training data.
- Start with using AI to augment rather than replace human decision making. Having humans make the ultimate decision avoids some complexity of explainable AI. Data biases may still be questioned, but human-based decisions are likely to be more difficult to be challenged than machine-only decisions.
- Create data and algorithm policy review boards to track and perform periodic reviews of machine learning algorithms and data being used. Continue to explain AI outputs within changing security requirements, privacy needs, ethical values, societal expectations and cultural norms.

Business Impact: End-user organizations may be able to utilize some future interpretability capabilities from vendors to be able to explain their AI outputs. But eventually, AI explainability is the end-user organization's responsibility. End users know the business context their organizations operate in, so they are better-positioned to explain their AI's decisions and outputs in human-understandable ways. The need for explainable AI has implications for how IT leaders operate, such as consulting with the line of business, asking the right questions specific to the business domain, and identifying transparency requirements for data sources and algorithms. The overarching goal is that models need to conform to regulatory requirements and take into account any issues or constraints that the line of business has highlighted. New policies around the inputs and boundary conditions on the inputs into the AI subsystem, how anomalies are handled, how models are trained and the frequency of training need to be incorporated into AI governance frameworks. Many questions about the suitability of the AI model will rely on a clear understanding of the goals of the application(s) being designed.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: H2O.ai; IBM; Microsoft; simMachines

Recommended Reading: "Top 10 Data and Analytics Technology Trends That Will Change Your Business"

"Build Trust With Business Users by Moving Toward Explainable AI"

"Predicts 2019: Digital Ethics, Policy and Governance Are Key to Success With Artificial Intelligence"

At the Peak

Continuous Intelligence

Analysis By: W. Roy Schulte; Pieter den Hamer; Melissa Davis

Definition: Continuous intelligence is a design pattern in which real-time analytics are integrated into business operations, processing current and historical data to prescribe actions in response to business moments and other events. It provides decision automation or decision support. Continuous intelligence leverages multiple technologies such as augmented analytics, event stream processing, optimization, business rule management and machine learning.

Position and Adoption Speed Justification: The current hype is focused on holistic, integrated continuous intelligence solutions that share real-time information from multiple sources with multiple departments and applications to support multiple business functions. Examples include real-time 360-degree views of customers, supply chain networks and enterprise nervous systems in airlines, railroads and other transportation operations. Simpler kinds of continuous intelligence are already common in point systems such as mobile device navigation, monitoring the health of machines, contact center monitoring, pop-up web ads, high frequency trading and package tracking. The hardware and software technologies for holistic, integrated continuous intelligence, including inexpensive sensors, publish-and-subscribe messaging systems such as Apache Kafka, event stream processing platforms and augmented analytics, are available and affordable.

However, many companies lack the skills necessary to develop their own custom-built solutions so holistic continuous intelligence will take five to 10 years to achieve 50% penetration of the target audience.

User Advice: Data and analytics leaders should consider continuous intelligence for new business processes and when making significant changes to existing processes. It applies to situations in which real-time data from the last few seconds or minutes significantly improves business decisions. It is not relevant where equally good decisions can be made with data that is hours, days, weeks or older. It goes beyond real-time descriptive, diagnostic and predictive analytics by supplying prescriptive information about the best available action to be taken in response to the situation. The potential role of continuous intelligence should be discussed with business managers and subject matter experts early in the requirements-gathering process. If continuous intelligence is implemented, it will fundamentally affect the design of business processes and their data and analytics. Companies can reduce the effort of achieving holistic continuous intelligence by subscribing to SaaS offerings, or acquiring packaged applications or devices that provide internal continuous intelligence on a point basis. However, holistic continuous intelligence will still entail custom design and integration with multiple applications, including independently owned and operated systems. This will require multidisciplinary collaboration among business domain experts, change managers, architects and developers. It may leverage messaging systems, event stream processing platforms, decision management tools, intelligent business process management suites (iBPMS), IoT platforms or other development, middleware and analytics products.

Business Impact: Continuous intelligence plays a major role in digital business transformation projects. A key benefit is improved situation awareness and a common operating picture across business functions by providing real-time dashboards and alerts. Equally important is the capability to trigger automated responses by sending signals to machines or initiating business processes in cases where the decision on what to do can be automated. Systems with continuous intelligence leverage real-time context data to support decisions for customer support, customer offers, risk, or allocating resources in the most efficient manner possible. However, enterprises that do not already have staff expertise in messaging, stream analytics, machine learning and decision management disciplines may need to hire outside service providers or train their staff on the new disciplines.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Confluent; FICO; RedPoint Global; SAS; Software AG; TIBCO Software; Unscrambl; Vitria; XMPro; ZineOne

Recommended Reading: "Make Your Customer Engagement Hub Real Time With Continuous Intelligence"

"How to Architect Continuous Intelligence Solutions"

"Innovation Insight for Continuous Intelligence"

"Building Your Continuous Intelligence Capability for Digital Transformation"

"How Companies Succeed at Decision Management"

"How to Move Analytics to Real Time"

Edge Analytics

Analysis By: Eric Hunter; W. Roy Schulte; Jim Hare

Definition: "Analytics" is the discipline that applies logic (i.e., "rules") and mathematics ("algorithms") to data to provide insights for making better decisions. "Edge" analytics means that the analytics are executed in distributed devices, servers or gateways located away from corporate data centers or cloud servers closer to where data from "things" (commonly sensors) is being generated.

Position and Adoption Speed Justification: Edge analytics moved further along the Hype Cycle toward the Peak of Inflated Expectations driven by increased expectations for edge analytics via machine learning and advances in the hybrid cloud. Five drivers for edge analytics use cases include latency/determinism, local interactivity, data/bandwidth considerations, privacy/security, or limited autonomy. Edge analytics offerings primarily support decentralized deployments of device-isolated insights. However, as connectivity advances and the demand for cross-device analytics increase, edge analytics will be tasked not only with providing edge-resident insights, but also to support conversion and compression to move data to hybrid cloud platforms for aggregation.

An increasing number of IoT platform and analytics vendors are adding the ability to deploy and run small-footprint analytics packages on edge devices — supporting both endpoints and aggregation devices like an IoT gateway. It reflects the shifting balance between edge and cloud computing. Public cloud providers are further accelerating this trend with announcements from Amazon Web Services (AWS Outposts), Microsoft (Azure Stack), Google (Anthos) and IBM (OpenShift). This trend is being driven by several factors including rightsizing connectivity to the edge, real-time analytics and data privacy considerations.

User Advice: Analytics leaders should consider edge analytics across the following five imperatives:

- **Limited Autonomy.** An individual device, asset or even a larger distributed site must provide analytic insights even in the midst of disconnection from cloud or data center infrastructure and resources

- **Privacy/Security.** Regulations or data privacy laws require that data be kept within the location of origin or the organization deems the transfer of data to introduce too many security vulnerabilities
- **Latency/Determinism.** Network connectivity does not have the ability to support desired latency or stability requirements
- **Local Interactivity.** Cross-device interdependencies as part of a larger system require edge-resident analytics
- **Data/Bandwidth.** It would cost too much to upload the full volume or fidelity of generated data, and there is no benefit to moving device-level data to a central location for aggregated analysis. Another scenario includes edge analytics for support of centralized cloud or data center analytic strategies by converting/compressing edge-generated data alongside for network transmission

Business Impact: Running analytics at the edge will become commonplace for both data and analytics and IoT architectures by the time it reaches the plateau.

Advantages include:

- **Faster response times.** Many sensors deliver digital and analog data at very low millisecond or sub millisecond intervals. When that data is sent to a central location for analysis, delays are introduced, and it loses its value for real-time requirements.
- **Reduced network bottlenecks.** Minimizes the risk of congesting device networks with full-fidelity or high-bandwidth data transmission (video, millisecond interval sensor reads and analog sensors).
- **Data filtering.** This reduces the data management and storage overhead by using edge analytics to look for just the actionable data. As a result, only the necessary data is analyzed or sent on for further analysis.
- **Reliability.** The remote location can remain in operation even if the network, cloud servers or data centers are unavailable.
- **Reduced communications cost.** Device-only edge analytics eliminate communication costs while edge device-based conversion and compression can dramatically lower costs versus sending raw analog or full-fidelity digital data to a central cloud or data center

Disadvantages include:

- **Increased complexity.** Remote, distributed analytics deployment and management make the deployment and management more complicated than for aggregated data in a single location — particularly when devices are heterogeneous in nature — lacking in standardization.

- **Reduced data granularity.** There is a potential loss of useful insight by discarding raw data stored locally as "data exhaust."
- **Lack of cross-device analytics.** Unless device data is transmitted to a consolidated location from the edge, leaders can lose the ability to deliver cross-device insights and analytics.
- **Device maintenance and technical currency.** Edge analytics will require more capable devices that increase demands for monitoring and maintenance of device health along with introducing demands for keeping devices up-to-date with both software and hardware revisions.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amazon Web Services; Arundo; Element; FogHorn; Iguazio; Microsoft Azure; Particle; PTC (ThingWorx); SAS; Sisense

Recommended Reading: "Top Strategic IoT Trends and Technologies Through 2023"

"Top 10 Strategic Technology Trends for 2019: Empowered Edge"

"The Edge Completes the Cloud: A Gartner Trend Insight Report"

"Deploying IoT Analytics, From Edge to Enterprise"

"4 Technology Sources for an AI-Enabled Enterprise"

"What Tech CEOs Must Know About Edge Computing in 2019"

MLOps

Analysis By: Erick Brethenoux; Jim Hare; Pieter den Hamer

Definition: Machine Learning Operationalization (MLOps) aims at streamlining the deployment, operationalization and execution of ML models. It supports the release, activation, monitoring, performance tracking, management, reuse, maintenance and governance of ML models.

Position and Adoption Speed Justification: Machine learning provides significant business value for organizations when ML models are deployed (i.e., embedded) within business processes — that is, when analytical assets are consumed by people or processes. Making sure that ML models can be legally traced, monitored and secured requires discipline and management capabilities. Deploying analytical assets within operational processes in a repeatable, manageable, secure and traceable manner requires more than a set of APIs and a cloud service. The democratization of ML techniques in the last few years has seen the proliferation of model development practices but unfortunately a majority of these models are neither operationalized nor deployed at scale. This capability is becoming critical for the survival of data science teams and this urgency will push MLOps toward the Plateau of Productivity in two to five years. More vendors are adding and expanding model management capabilities to their data science platforms (see "Critical Capabilities for Data Science and Machine Learning Platforms"), and stand-alone products are also available.

User Advice: To systematically leverage their data science and machine learning efforts while strengthening their analytics strategy, data and analytics leaders should:

Ensure the business value of analytical deployments while prioritizing use cases by establishing close, ongoing dialogue with, and explicit buy-in from, their business counterparts. The earliest the dialogue, the most successful the model operationalization will be.

Ensure the integrity (technical and business), transparency and sustainability of deployed analytical models by establishing a systematic operationalization process.

Maximize operationalization success by securing the help of domain experts, process engineers, IT professionals and business practitioners, in addition to existing data science talent.

Minimize the heavy technical debt and complex maintenance procedures of deployed machine learning models by monitoring and revalidating their business value on an ongoing basis.

Business Impact: MLOps tools and functionality should provide the following for maximum ongoing and sustainable business impact:

- Catalog — A "centralized" way to store and secure data and analytical assets to make it easier for analysts to collaborate and to allow them to reuse models or other assets as needed (this could be a secured community or a collaboration space).
- Governance — Protocols to ensure adherence to all internal and external security procedures and regulations: not just for compliance reasons, but, because an increasing amount of data gets aggregated, to address potential privacy issues.
- Capabilities — Automated versioning, fine-grained traceable model scoring and change-management capabilities (including championing/challenging, data sources and model dependencies features) to closely test, monitor and audit analytical asset life cycles from a technical as well as a business performance perspective (through KPIs); to maximize the business impact of operationalized models.
- Coherence — Establish simple protocols to provide functional bridges between the development and operationalization cycles; enhance the cooperation and consultations between the development and operationalization teams; provide efficient "translation" services between data science, IT and the lines of business; improving the transparency of deployed analytical assets.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Sample Vendors: Algorithmia; Datatron; Open Data Group; ParallelM; Saagie; Seldon

Recommended Reading: "How to Operationalize Machine Learning and Data Science Projects"

"Magic Quadrant for Data Science and Machine Learning Platforms"

"Critical Capabilities for Data Science and Machine Learning Platforms"

Citizen Data Science

Analysis By: Carlie Idoine; Shubhangi Vashisth; Rita Sallam

Definition: Citizen data science is an emerging set of capabilities and practices that allow users to extract advanced analytic insights from data without the need for extensive data science expertise. Central to citizen data science are augmented analytics capabilities including automated, streamlined data access and data engineering, augmented user insight through automated modeling and pattern detection including feature engineering, model selection and validation, automated deployment and operationalization, and a focus on collaboration and sharing.

Position and Adoption Speed Justification: Current modern analytics and BI approaches enable business users to do manual data preparation, data exploration and some pattern detection. However, building data science and machine learning models requires expert data scientists who are expensive to hire and in short supply. Citizen data science tools guide the user through the end-to-end modeling process by automating some manual and bias-prone tasks like feature selection. Many augment the user's discovery capabilities by automatically generating and prioritizing statistically meaningful insights for users.

User Advice: For data and analytics leaders:

Look for opportunities for citizen data science to complement and collaborate with existing user-oriented modern analytics, and BI and expert data science initiatives. Define the citizen data scientist as a legitimate role within the organization. Define its "fit" relative to other analytic roles, and identify business analysts and other specialists who fit the citizen data scientist profile.

Assess your organization's readiness for business-user-accessible advanced analytics by evaluating your data, technology, process and skills in terms of alignment with business outcomes and skills. Implement a program for developing citizen data scientists from existing roles such as the business and data analyst.

Monitor the capabilities (technology) and roadmaps of your modern BI and data science platforms, as well as emerging startups as they mature in terms of the data preparation required, types of data that can be analyzed, the types of algorithms supported and the augmented analytic features supported.

Educate business leaders and decision makers about the potential impact of a broader range of users leveraging and understanding data science and machine learning. At the same time, stress the need for education, user enablement, responsible use, governance and collaboration between citizen data scientists and specialists to avoid negative consequences.

Improve highly skilled data science productivity with citizen data science by defining and providing guidance for the interactions and responsibilities of both disciplines. Recognize that you still need specialist data scientists to validate and operationalize models, findings and applications.

Business Impact: Citizen data science forms the foundation of next-generation analytics. It will make insights from data science and machine learning more accessible and pervasive in the enterprise. Citizen data scientists can be leveraged to fill the data science and machine learning talent gap that is currently being experienced due to the shortage and high cost of data scientists. Citizen data science also has the potential to make expert data scientists more efficient.

Incorporating citizen data scientists into specific phases of the analytic life cycle can enable more scalable and focused use of data science and machine learning resources across the organization. Leveraging citizen data scientists in the exploratory phase of a project, for example, can enable the highly skilled data scientists to focus their expertise on the more cutting-edge model building phases. Citizen data scientists can also operate as “business translators,” an often missing and much needed role for data science teams today.

Citizen data science will be a key driver of analytics adoption during this decade and the next. Gartner anticipates that, within the next several years, citizen data science will rapidly become more prevalent as an approach to enabling and scaling data science capabilities more pervasively throughout organizations. Gartner predicts that, by 2020, more than 40% of data science tasks will be automated, resulting in increased productivity and broader usage by citizen data scientists. Gartner also predicts that, by 2024, a scarcity of data scientists will no longer hinder the adoption of data science and machine learning in organizations.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Aible; Alteryx; Big Squid; dotData; DataRobot; Prevision.io; Salesforce (BeyondCore); SAS; SparkBeyond; Tellius

Recommended Reading: “Maximize the Value of Your Data Science Efforts by Empowering Citizen Data Scientists”

"Build a Comprehensive Ecosystem for Citizen Data Science to Drive Impactful Analytics"

"How Citizen Data Science Can Maximize Self-Service Analytics and Extend Data Science"

"Pursue Citizen Data Science to Expand Analytics Use Cases"

Decision Management

Analysis By: W. Roy Schulte; Erick Brethenoux; Pieter den Hamer

Definition: Decision management is the discipline of designing, building and maintaining systems that produce structured decisions. In this context, a "decision" is a judgment, conclusion or determination of what to do. A decision is structured if it can be documented in an explicit decision model, which is a representation of the decision-making process including data inputs, algorithms and results. Decision making that is invoked by other applications uses prescriptive analytics such as optimization, business rules and machine learning.

Position and Adoption Speed Justification: Companies are ramping up their use of decision management to cope with the increasing demand for decision automation, and the growing complexity of decisions in business. Design approaches have evolved to focus first on requirements and decision models before developing analytics or business rules. Business rule management systems (BRMSs) are evolving into decision management suites to support decision modeling, machine learning (ML) and optimization.

Recent advancements include:

- Knowledge graph representations at scale
- Agent-based decision intelligence techniques
- Augmented analytics
- Formal standards for decision modeling, such as Object Management Group's Decision Model and Notation (DMN)

Many IT leaders, including data and analytics leaders, have limited understanding of the available techniques and tools, or their value. Many applications that could benefit from decision management do not use it yet, and it will take more than five years to reach the Plateau of Productivity.

User Advice: Decision management typically applies to continuous intelligence and other operational decisions, but it is also relevant to some strategic and tactical decisions. Decision-making software solutions may be discrete applications, but are more often subsystems, such as callable decision services or embedded code segments that are component parts invoked by larger applications.

Data and analytics leaders should:

- Use decision management for complex or repeatable decisions that involve substantial calculations, frequently changing logic, or traceable decision making.
- Use decision management for decision automation for decisions that can be fully specified in a model and implemented in software with no need for direct human involvement at decision time.
- Use decision management for decision support for decisions that require both analytics and human judgment.
- Develop a basic understanding of business rule processing (decision tables, for example), ML, optimization and other prescriptive analytics to guide the selection of software.
- Use data brokers and other sources of context data to obtain the information needed for more accurate and more precise decisions.
- Improve complicated decisions that involve trade-offs among alternative courses of action by using *decision-problem* models (those based on management science), which calculate outcomes from different inputs.
- Improve repeatable decisions that are primarily based on policies, heuristics or ML by using *decision-composition* models that document the act of decision making, and by exploring new explainable AI techniques.

Business Impact: Decision management improves the design and software engineering process for systems that use prescriptive analytics by focusing attention on the business goals, requirements, decisions and subdecisions before jumping into detailed analytics and rules. Early use of decision management was concentrated in insurance and loan underwriting, mortgage approval, resource allocation, logistics, and public-sector applications, such as approving permits and determining welfare and taxes. More recently, it spread into other data- and logic-intensive applications, particularly customer experience management, compliance, fraud detection and risk management.

Decision management:

- Strengthens collaboration among business leaders, subject matter experts, business analysts, management scientists, data scientists and software developers.
- Makes it easier to develop sophisticated decision-making systems that combine rules, ML and optimization techniques.
- Helps make decisions more accurate, consistent, transparent and auditable.

- Supports continuous decision improvement.
- Improves compliance with corporate and legal requirements.
- Facilitates sharing business policies and rules among departments and external business partners.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Actico; Decision Management Solutions; Enova Decisions; Experian; FICO; IBM; Red Hat; Sapiens International; SAS; Signavio

Recommended Reading: "How Companies Succeed at Decision Management"

"Decision Intelligence Is the Near Future of Decision Making: A Gartner Trend Insight Report"

"Develop Good Decision Models to Succeed at Decision Management"

"Innovation Insight for Optimization"

"Innovation Tech Insight for Business Rules Management Systems"

Augmented Analytics

Analysis By: Rita Sallam; Carlie Idoine

Definition: Augmented analytics uses machine learning to automate data preparation, insight discovery, data science, and machine learning model development and insight sharing for a broad range of business users, operational workers and citizen data scientists. It is expanding insights by using AI and ML techniques to deliver analytics to everyone in the organization with less time, skill and interpretation bias of current manual approaches.

Position and Adoption Speed Justification: Visual-based data discovery — a key feature of current modern analytics and business intelligence BI platforms — has been transformative in the way it enables business users to generate analytics insights (in comparison with traditional BI technologies). However, many of the activities associated with preparing data, finding patterns in data building data science and machine learning models on complex combinations of data, and sharing insights with others, remain highly manual. As a result, it is not possible for users to explore every possible pattern combination, let alone determine whether their findings are the most relevant, significant and actionable.

Relying on business users to find patterns manually may result in users exploring their own biased hypotheses, missing key findings and drawing their own incorrect or incomplete conclusions, which may adversely affect decisions and outcomes.

Augmented analytics includes:

Augmented data preparation, which uses machine learning automation to augment data profiling and data quality, harmonization, modeling, manipulation, enrichment, metadata development and cataloging.

Augmented analytics in analytics and BI platforms, which is a key feature of next-generation modern analytics and BI platforms. It enables business users and citizen data scientists to automatically find, visualize and narrate relevant findings, such as correlations, exceptions, clusters, segments, outliers, predictions, and prescriptions without having to build models or write algorithms. Users explore data via visualizations, search and natural language query technologies, supported by natural-language-generated narration interpretation of results.

Augmented data science and machine learning, which automates key aspects of advanced analytic modeling, such as feature and model selection. Some platforms offer model explanation and will automate aspects of model management. This reduces the requirement for specialized skills to generate, operationalize and manage an advanced analytics model.

Augmented analytics capabilities will advance rapidly along the Hype Cycle to mainstream adoption, as a key feature of data preparation, modern analytics and BI and data science platforms. More importantly, automated insights from augmented analytics will also be embedded in enterprise applications — for example, those of the HR, finance, sales, marketing, customer service, procurement and asset management departments — to optimize the decisions and actions of all employees within their context, not just those of analysts and data scientists.

Augmented analytics will also be a key feature of conversational chatbots for analytics. This is an emerging paradigm that enables business people to generate queries, explore data, and receive and act on insights in natural language (voice or text) via mobile devices and personal assistants.

User Advice: Data and analytics leaders should:

- Embrace augmented analytics as part of a digital transformation strategy to deliver more advanced insights to a broader range of users — including citizen data scientists and, ultimately, operational workers — without expanding the use of data scientists. Pilot to prove the value and build trust.

- Monitor the augmented analytics capabilities and roadmaps of modern analytics and BI, data science platforms, data preparation platforms, and of startups as they mature. They should do so particularly in terms of the upfront setup and data preparation required, the types of data that can be analyzed, the types and range of algorithms supported, languages supported, integration with existing tools, explainability of models, and the accuracy of the findings.

- Explore opportunities to use augmented analytics to complement existing modern analytics and BI, data science initiatives and embedded analytic applications, where automating algorithms to detect patterns in data could reduce the exploration phase of analysis and improve highly skilled data science productivity.
- Recognize that citizen data scientists must collaborate with, and be coached by, specialist data scientists that still need to validate models, findings and applications.

Business Impact: Expanded use of machine learning automated and human-augmented models will translate into less error from bias, which is inherent in manual exploration processes. It will reduce the time users spend on exploring data, while giving them more time to act on the most relevant insights from data. It will also give front-line workers access to more contextualized analytical insights and guided recommendations to improve decision-making and actions.

Gartner predicts that, by 2020, due in large part to the automation of data science tasks, citizen data scientists will surpass data scientists in the amount of advanced analysis produced. This growth, enabled by augmented analytics, will complement and extend existing modern analytics and BI and data science platforms, as well as enterprise applications, by putting insights from advanced analytics — once available only to data science specialists — into the hands of a broad range of business analysts, decision makers and operational workers across the enterprise, driving new sources of business value.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: DataRobot; Outlier; Paxata; Prevedere; Salesforce (Einstein Analytics); SparkBeyond; Tellius; ThoughtSpot; Trifacta

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

"Magic Quadrant for Data Science and Machine Learning Platforms"

"How Citizen Data Science Can Maximize Self-Service Analytics and Extend Data Science"

"Pursue Citizen Data Science to Expand Analytics Use Cases"

"Augmented Analytics Is the Future of Data and Analytics"

"Top 10 Data and Analytics Technology Trends That Will Change Your Business"

AutoML

Analysis By: Jim Hare; Alexander Linden

Definition: Automated machine learning (AutoML) is the capability of automating the process of building, deploying and/or managing machine learning models.

Position and Adoption Speed Justification: AutoML has emerged as a way for traditional and citizen data scientists (developers and business analysts) to build and deploy models easier and faster using automation and embedded machine learning. AutoML solutions aim to automate some or all steps of the machine learning pipeline processing including:

- Data preprocessing
- Feature engineering
- Feature extraction
- Feature selection
- Algorithm selection
- Hyperparameter optimization
- Model deployment
- Model management/monitoring

The hype in autoML is driven by the promise that it can reduce the demand for data scientists by enabling domain experts to automatically build predictive models with limited knowledge of statistics and ML. The term is also being amplified by vendors adding autoML capabilities to their data science platforms and proactively marketing it to differentiate their products or target new users. Over the past couple of years, the leading AI cloud service providers (AWS, Google, Microsoft) have introduced and evolve their autoML capabilities. Google released a branded product called "Google AutoML" targeted at developers with limited machine learning expertise; however, the term autoML for this innovation profile is not related directly to this product. ML platform providers like SAS, H2O.ai, RapidMiner and KNIME have also added autoML capabilities as part of their ML platforms.

In addition to commercial offerings, there are a wide of variety of open-source autoML tools such as auto-sklearn, Auto-WEKA, Machine.js, H2O.ai and TPOT. There are competitions such as ChaLearn AutoML and Shoodoo that give prize money to top data scientists who design the perfect machine learning "black box" capable of performing all model selection and hyperparameter tuning without any human intervention.

The reality is that a general purpose end-to-end autoML solution that covers the full ML pipeline — from data selection to model deployment — doesn't yet exist. Most autoML capabilities available today focus on helping data scientists with a subset of the ML pipeline such as feature engineering, model selection, or tuning. And, autoML only works for supervised ML use cases. Even those that claim to they can automatically build a machine learning pipeline for citizen data scientists are only able to do it for very well-defined, narrow use cases. AutoML also requires clean input data to provide high quality results.

Gartner expects that autoML will rapidly move from the peak to the Trough of Disillusionment as (1) the hype dissipates and it becomes a standard feature of every ML platform. And (2), as end-user organizations run into challenges using the capability and experience its limitations.

User Advice: Data and analytics leaders responsible for data science teams should:

- Become familiar with the different types of autoML capabilities (open source and commercial) and how they are can be used to help and support different parts of the machine learning pipeline.
- Set expectations that autoML is not a magical solution to replace data scientists. Some vendors promote that their autoML offerings can replace the need for data scientists rather than being used to assist and free them of repetitive tasks. The reality is that autoML is just another tool to help as part of the pipeline. For example, data cleansing still accounts for a big part of a data scientist's workload, which AutoML doesn't really address today.
- Understand the limitations of autoML. For example, using autoML for feature engineering may give you suboptimal results since the secret ingredient to obtaining high-quality models in many real-world problems continues to be domain knowledge. Also, autoML overlooks the more challenging tasks of unsupervised and reinforcement learning, focusing only on supervised tasks that require labelled data as input.
- Be aware of data science platform vendors overhyping their autoML. Some vendors market that they can automatically generate end to end machine learning pipelines but this is typically limited to very well defined, known use cases.
- Understand that data is the key to get the best results using autoML. The autoML solutions assume that users have the right data as an input for autoML. The choice of training data will directly impact the accuracy and quality of the autoML results. And, most only work with structured data while some solutions have been extended to handle unstructured data, namely, text and image. Choosing the wrong data can also introduce bias.

Business Impact: Today, autoML offers the following benefits:

- Enables domain experts (developers, business analysts) to build ML models without needing to code.
- Helps data scientists more quickly identify the best features from the available training data.
- Saves time by recommending the best algorithm for a given use case.

- Reduces the time to find the best model hyperparameter tuning settings.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Amazon SageMaker; Big Squid; dotData; DataRobot; Google (Cloud Platform); H2O.ai; KNIME; RapidMiner; Sky Tree

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

"Critical Capabilities for Data Science and Machine Learning Platforms"

"Augmented Analytics Is the Future of Data and Analytics"

"How to Choose the Right Data Science and Machine Learning Platform"

"Top 10 Strategic Technology Trends for 2019: Augmented Analytics"

Conversational User Interfaces

Analysis By: Magnus Revang; Van Baker

Definition: Conversational user interface (CUI) is a high-level design model in which the user and machine interactions primarily occur in the user's spoken or written natural language. Sophistication of the CUI can vary from understanding just simple verbal utterances to handling complex multiturn interactions.

Position and Adoption Speed Justification: CUIs can exist as a front end to application or business process, but also as a description of the interface employed by chatbots and virtual assistants. It's being popularized through products like the Amazon Echo that uses the Amazon Alexa Virtual Personal Assistant (VPA) and Google Home that uses Google Assistant VPA. Enterprises are coming on board, with chatbots being the No. 1 use case for AI technology in enterprises.

The promise of CUIs is a shift in responsibility between the user and the interface. In traditional user interfaces (UIs), the user is an operator of the technology and is largely responsible for the effects of using the technology. In a CUI, this responsibility shifts as the CUI is responsible for taking the user input and determining the intention of the user. Conceptually, the CUI has taken over some of the responsibility that was once reserved for the user. This makes CUIs the first widespread adoption of agent user interfaces.

CUIs will evolve their conversational capabilities through advances in natural language understanding (NLU) and in more advance dialogue management. Additionally, we will see the introduction of multimodal interactions, where speech, text, video and point-and-click interactions are all part of the input used to determine the intention of the user.

User Advice: The conceptual shift away from the user as the operator toward the user as conversing with an agent that will execute on a determined intention — has greater impact on the enterprise than most realize. Training, onboarding, escalations, productivity, empowerment and responsibility all change with this new model and need to be embraced as part of CUI projects. Treat CUIs as transformative and plan on it, and by evolution AUIs becoming the dominant interaction model in the future.

Underlying technology supporting CUIs, either front ends delivered as part of software or custom developed CUIs like chatbots and virtual agents built on top of conversational platforms, will greatly improve over the coming 12 to 24 months. Plan on any adoption to be tactical, with vendor selection changing at that point.

Prepare for CUIs to communicate with each other. Larger architectures connecting different use cases for CUIs, like virtual agents for customer service, human resources, IT help desk to front ends for enterprise software, business intelligence tools and similar, will be a central challenge for organizations in the next three to five years. This will lead to a variety of architectural models like CUI-to-CUI communication and specialist tooling entering the market.

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 three to 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 is 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: 5% to 20% of target audience

Maturity: Adolescent

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

Recommended Reading: "Architecture of Conversational Platforms"

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

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.

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. Government commissions and industry consortiums are actively developing guidelines for ethical use of AI. See "Ethics Guidelines for Trustworthy AI" (https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai).

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 "Data Ethics Enables Business Value"). 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: Digital ethics should be treated as a tangible business practices discipline rather than an academic discussion. Key areas where it should be applied include social and mobile technologies, and social interaction; cloud and security; big data and privacy; autonomous technologies and freedom; artificial intelligence/robotization and the value of work; and predictive algorithms and decision-making.

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 data and 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 Data for Good (see "How to Use Data for Good to Impact Society").

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Recommended Reading: "Top 10 Strategic Technology Trends for 2019: Digital Ethics and Privacy"

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

"How to Use Data for Good to Impact Society"

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

"How to Apply Gartner's Digital Humanism Manifesto"

"Data Ethics Enables Business Value"

"Modernize Data Privacy to Put the Personal Back Into Personalization"

"Workplace Analytics Needs Digital Ethics"

"The #DigitalSociety Requires a Digital Social Contract"

Deep Neural Networks (Deep Learning)

Analysis By: Alexander Linden; Chirag Dekate; Svetlana Sicular

Definition: Deep neural networks (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 better complex data, such as video, image, speech and textual data.

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

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 — which are difficult to obtain and may even require a great deal of capital expenditure. DNNs are reaching the market in a prepackaged form: as AI services for language and image processing, as well as specific applications, for example, medical diagnostics, chatbot speech capabilities or video forensics. However, the end users are mostly unaware of the DNNs' role in these solutions and their performance remains brittle from a lack of transfer of learning capabilities.

The level of hype about DNNs has worn out a bit. While it is still an active and vibrant area to invest for VC and academia — the commercial traction is falling slightly behind expectations due to the overall challenges with AI projects. Also, the time of one breakthrough every month seems to be over for now.

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** prepackaged capabilities from vendors like Clarifai, Matroid, Deep Instinct, Noble, Google's Vision API, Amazon Rekognition, etc.

- **Then proceed with platforms provided by 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 the corresponding training data. For certain narrow use cases, 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 specialist consultancies and academic teams. Use crowdsourcing providers like Experfy and Kaggle. 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:** 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. Utilize data annotation services to curate datasets for deep learning.

- **Explore the use of synthetic data** to potentially expand the range of training data while retaining datasets with a high-level of quality that are diversified, labeled and less biased.

Business Impact: DNNs started the current AI hype. They have transformational, and therefore disruptive, potential for all industries. DNNs' algorithms will be developed by a small number of organizations. For most companies, DNNs will come prepackaged as services and applications. DNNs demonstrate superior accuracy to past state-of-the art algorithms in detecting fraud, monitoring quality of goods and processes, and predicting demand and other machine learning problems that involve sequences (using, for example, video, audio or time series analysis). The challenge for those wanting to realize the DNNs' potential is to identify the business problems to solve, have large enough datasets and to secure availability of the experts to implement DNN solutions of choice.

The basis of a DNN's potential is its ability to produce more nuanced representations of highly dimensional and complex data. A DNN can, for example, give promising results when interpreting medical images 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 parts of speech (which, in time, may make most devices conversational devices).

Completely new product categories are likely to be found 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; Clarifai; DimensionalMechanics; Google; H2O.ai; Matroid; NVIDIA; Skymind; TwentyBN

Recommended Reading: "4 Ways to Obtain AI Solutions Using Deep Neural Nets"

Graph Analytics

Analysis By: Mark Beyer; Rita Sallam

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

Position and Adoption Speed Justification: Graph analytics climbed slightly over the Peak of Inflated Expectations due to increased awareness in early 2019. The growing adoption of graph analytics is driven largely by the need to identify and explore insights into relationships for specific business use cases. This requires analysis across an exponentially larger amount of heterogeneous data, which is not well-suited to relational storage and analysis. These highly complex models are developed and used within machine learning with the output stored in graph databases. GraphDBs present an ideal framework for storing, manipulating and analyzing the widely varied perspectives in the graph model due to their graph-specific processing languages, capabilities and computational power. At the same time, established AI techniques (such as Bayesian networks) are increasing the power of knowledge graphs and the usefulness of graph analytics by introducing further nuance in representational power.

Graph analytics processing is a core technology underlying many other advanced technologies, such as virtual personal assistants, smart advisors and other smart machines. Various platforms also use graph analytics to identify important findings.

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. Many providers are introducing graph engines embedded in their platforms, which means that some of the adoption curve will remain almost hidden. 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. The method of storing and processing data within graph databases differs from traditional relational data, creating a demand for new skills related to graph-specific knowledge, which may limit growth in adoption. Examples of the new skills required include knowledge and experience with the Resource Description Framework (RDF), Property Graphs, 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 that are 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. In a more recent development, metadata analysis has shown graph analysis to be specifically high value.

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.

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.

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 similar issues within the organization and answer issues of liability in a proactive manner. They can also be used to identify peculiarly successful operating units within a larger organization to analyze if their patterns can be repeated. Once a graph process is completed, it can be visualized — using size, color, shape and direction — to represent relationship and node attributes.

A now classic example of identifying networks of relationships is the International Consortium of Investigative Journalists (ICIJ) research, which revealed the Panama Papers.

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. Additionally, graph analytics enable causality and dependency analyses, therefore increasing transparency in predictive models.

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

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Cambridge Semantics; Centrifuge Systems; Databricks; Digital Reasoning; Emcien; Maana; Palantir; Symphony AyasdiAI; SynerScope

Recommended Reading: "Combine Predictive and Prescriptive Analytics to Solve Business Problems"

"Best Practices for Designing Your Data Lake"

Prescriptive Analytics

Analysis By: Carlie Idoine; Peter Krensky

Definition: "Prescriptive analytics" are a set of capabilities that specify a preferred course of action to meet a predefined objective. The most common types of prescriptive analytics are optimization methods, a combination of predictive analytics and rules, heuristics, and decision analysis methods. Prescriptive analytics differs from descriptive, diagnostic and predictive analytics in that the technology explores multiple outcomes and provides a recommended (and sometimes automated) action.

Position and Adoption Speed Justification: From a “purist” perspective, the term “prescriptive analytics” is a broad category with little hype. The broad category encompasses components with varying positions across the Hype Cycle and various levels of maturity. Such components include optimization, rules plus predictive techniques and decision intelligence.

The concepts of optimization and decision analysis have existed for decades. However, they have sustained a recent resurgence due to maturing and expanding data science initiatives, better algorithms, more cost-effective cloud-based computing power and a substantial increase in 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. Many use cases are well-established, and some organizations are fairly productive with these techniques. 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. New use cases continue to emerge. As adoption steadily broadens, expectations for prescriptive analytics continue to exceed reasonable expected value. Even longstanding use cases can fall victim to common data science challenges such as data quality, bias and talent shortages.

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 two to five 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: Adolescent

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

Recommended Reading: “Combine Predictive and Prescriptive Techniques to Solve Business Problems”

“Forecast Snapshot: Prescriptive Analytics Software, Worldwide, 2019”

“The 2019 Top Supply Chain Technology Trends You Can’t Ignore”

“Market Guide for Supply Chain Analytics Technology, 2018”

“The Evolving Capabilities of Analytics and Business Intelligence Platforms”

Advanced Anomaly Detection

Analysis By: Peter Krensky; W. Roy Schulte

Definition: Anomaly (or outlier) detection means identifying objects, groups of objects, events or event patterns that deviate from the expected or norm. Anomaly detection is “advanced” if it leverages sophisticated mathematical techniques — machine learning or other advanced analytics — to precisely detect more-subtle anomalies, provide earlier notice of likely future anomalies, or streamline the design and development of systems that detect (and even act on) anomalies.

Position and Adoption Speed Justification: Anomaly detection has been the subject of academic research and limited commercial use for many years and its sophistication and commercial appeal continue to grow. However, hype for the technology is slightly past its peak as solutions mature and myriad other concepts in the AI ecosystem compete for attention. Advances in machine learning, stream analytics, behavioral analytics, in-memory computing, DBMS technology and hardware technology all improve the accuracy and timeliness of anomaly detection systems and reduce the human effort required to design, deploy and manage the systems.

Virtually all organizations have aspects of operations where it is valuable to distinguish between routine conditions and matters that require extra attention. Much of advanced anomaly detection is a competitive advantage today, but we expect most of the current technology driving it will be widespread and even taken for granted within 10 years because of its broad applicability and benefits. Advanced anomaly detection is already being embedded in many enterprise applications. Still, many organizations across all industries have yet to encounter the hype around advanced anomaly detection, and it will take time to reach the Plateau of Productivity as data and analytics leaders invest in more prominent and mission-critical technologies first.

User Advice: Risk managers, IT infrastructure and operations managers, manufacturing engineers, business operations and supply chain managers, healthcare professionals, public safety managers, customer experience and relationship managers and anyone else who uses monitoring or testing systems should make anomaly detection capabilities (both traditional and advanced) a major part of the selection process for the tools they will use.

- In many situations, managers and subject matter experts should obtain advanced anomaly detection technology as a feature in some larger application or system, because it is typically not a product category in its own right.
- If no vendor provides packaged or configurable advanced anomaly detection relevant to the task at hand, developers should use a data science and machine learning platform (see "Magic Quadrant for Data Science and Machine Learning Platforms").
- Where relevant, use advanced anomaly detection techniques on data about economic conditions, competitive actions, channel disruptions and upstream value-chain disturbances.
- Subject matter experts who are familiar with the operation being monitored must be consulted to determine which anomalies require responses and what those responses should be.
- Give favorable consideration to products that can make the root causes of anomalies visible, or that can learn what is normal (a baseline or "fingerprint") without extensive human involvement by using machine learning to process historical data on past performance or by leveraging knowledge graph capabilities to detect causality and event dependencies.
- Leverage the services of a data scientist to evaluate the performance requirements — the trade-off between false alarms (false positives) and failed detections (false negatives), for example — and compare them to the capabilities of the tools.

Business Impact: Advanced anomaly detection is relevant in situations where the existence, nature and extent of disruptions cannot be anticipated and predicted. Systems that use advanced anomaly detection are more effective than those that use simpler techniques.

They can:

- Detect subtle anomalies that might otherwise escape notice.
- Provide earlier warning of impending problems.
- Give more time to capitalize on emerging opportunities.
- Reduce the human effort required to develop a monitoring or measuring application.
- Reduce the time required to implement complicated anomaly detection systems.

Advanced anomaly detection is well-established in certain applications, such as cybersecurity, fraud detection for payment cards, financial market surveillance, and control systems and health monitoring systems for physical assets such as machines, vehicles and power plants. Its use is increasing in applications as diverse as enterprise security, unified monitoring and log analysis, application and network performance monitoring, business activity monitoring (including business process monitoring), Internet of Things predictive equipment maintenance, supply chain management, and corporate performance.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Sample Vendors: Anodot; Ayasdi; Darktrace; Datanomers; Featurespace; Netskope; Oversight; ThetaRay; WeDo Technologies

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

"Market Guide for User and Entity Behavior Analytics"

"Market Guide for Online Fraud Detection"

"Market Guide for Endpoint Detection and Response Solutions"

Event Stream Processing

Analysis By: W. Roy Schulte; Nick Heudecker; Pieter den Hamer

Definition: An event stream is a sequence of event objects arranged in some order, typically by time. Event stream processing (ESP) is computing that is performed on event objects for the purpose of stream data integration or stream analytics (also called complex-event processing [CEP]). Stream analytics can be executed either as new data arrives (using event-driven ESP platform software); shortly after it arrives (using real-time, on-demand queries); or long after it has been stored (using on-demand queries on historical data).

Position and Adoption Speed Justification: Three factors are driving the expansion of ESP:

- The growth of the Internet of Things (IoT) and digital interactions is making event streams ubiquitous.
- Business is demanding continuous intelligence for better situation awareness and faster, more personalized decisions.
- Vendors are bringing out new products, many of them open source or partly open source, giving the impression of lower acquisition costs.

Companies have access to more streaming data from internal sources (such as sensors, meters, control systems, corporate websites and transactional applications); and from external sources (such as social computing platforms, news and weather feeds, other data brokers, government agencies and business partners). ESP technology is maturing rapidly. It will eventually be adopted by multiple departments within every large company. ESP will reach the Plateau of Productivity within five years, largely by being embedded in SaaS solutions and off-the-shelf packaged applications.

User Advice: Data and analytics leaders should:

- Use ESP when traditional DBMS architectures cannot execute fast enough to provide real-time information from high-volume event streams..
- Acquire ESP functionality by using a SaaS offering or an off-the-shelf application that has embedded CEP logic (but only if a product that addresses your specific business requirements is available).
- Build your own application on an ESP platform if an appropriate off-the-shelf application or SaaS offering is not available.
- Build your own application on an ESP platform if your company has multiple applications that need some of the same overlapping data or CEP logic, to avoid redundant or stove-piped ESP solutions.
- Use free, community-supported, open-source ESP platforms if your developers are familiar with open-source software and languages such as Java, Scala or Python, and license fees are the primary consideration. Use vendor-supported closed-source platforms or products that mix an open-source core with value-added closed-source extensions for mainstream applications that require enterprise-level support and more complete sets of features.

- Use on-premises ESP in preference to cloud event processing services for ultra-low-latency applications such as high-frequency financial trading, and for applications where most of the data originates on-premises.
- Use ESP technology that is optimized for stream data integration to ingest, filter, enrich, transform and store event streams in a file or database for later use.

Business Impact: Stream analytics provided by ESP platforms:

- Can support situation awareness through dashboards and alerts by analyzing multiple kinds of events in real-time.
- Enable smarter anomaly detection and faster responses to threats and opportunities.
- Can help shield business people from data overload by eliminating irrelevant information and presenting only alerts and distilled versions of the most important information.

ESP is one of the key enablers of continuous intelligence and other aspects of digital business. It has transformed financial markets and become essential to smart electrical grids, location-based marketing, supply chain, fleet management and other transportation operations. Much of the growth in ESP usage during the next 10 years will come from three areas where it is already somewhat established: IoT, customer experience management and fraud detection applications. More than 40 ESP products or PaaS services are available on the market.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: EsperTech; EVAM; IBM; Microsoft; Oracle; SAP; SAS; Software AG; The Apache Software Foundation; TIBCO Software

Recommended Reading: "Market Guide for Event Stream Processing"

"Adopt Stream Data Integration to Meet Your Real-Time Data Integration and Analytics Requirements"

"The Five Levels of Stream Analytics — How Mature Are You?"

"Technology Insight for Event Stream Processing"

"Innovation Insight for Continuous Intelligence"

Sliding Into the Trough

Advanced Video/Image Analytics

Analysis By: Nick Ingelbrecht

Definition: Advanced video/image analytics is the application of data science methods (including deep neural networks) to automatically identify significant information contained in image and video streams (a series of images or pixels).

Position and Adoption Speed Justification: Video/Image analytics have advanced through the Gartner Hype Cycle, supported by proven use cases and competitive products. Advanced video analytics using deep neural networks and advanced data modelling are gaining traction in security, retail, automotive and other specialist vertical markets, but are still overhyped by many vendors. By contrast, traditional video analytics using geometric and rule-based systems have already reached early mainstream adoption. Advanced applications range through facial recognition in crowds, autonomous vehicles and, in the retail industry, shelf and shopper analysis.

Mainstream adoption remains elusive due to:

- A lack of plug-and-play solutions
- Integration challenges
- Fragmented markets and diverse buying centers
- Proprietary algorithms and patent pools
- Lack of independent standardization and performance benchmarks
- High-end systems are expensive to maintain and support.
- Exponential increases in video/image traffic are driving demand for automated analysis, especially edge inferencing.
- The application of deep neural networks for advanced video/image analytics has raised the bar in terms of nuisance alarm management and new functionality, especially in human behavior recognition, and complex classification tasks.

Product maturity will drive mainstream adoption by 2023, enabled by more embedded edge inferencing in cameras, along with packaged applications and self-configuring, integration-ready third-party applications.

User Advice:

Data and analytics leaders should:

- Work with Gartner to survey the competitive landscape and shortlist capable vendors, especially if there are complex and varied analytics requirements.

- Work with key stakeholders on prioritizing business use cases and define the types of analytics required and levels of performance/accuracy.
- Analyze the potential to use advanced video/image analytics to answer broader business questions — for example, exploiting a video surveillance investment to health and safety benefits to employees or improvements in customer, citizen or student experiences.
- Hold vendors accountable for meeting agreed project outcomes rather than the performance of individual system elements.
- Pilot video/image analytics solutions prior to production in order to test vendor claims, and stage a "vendor shootout" for large or high-stakes deployments in the real-world deployment environment during different lighting and environmental conditions.
- Adopt an analytics-driven total system approach that balances the trade-offs between different system elements, and resist internal pressures for piecemeal upgrades and bolt-on investments.
- Use business process mapping (BPM) tools to understand the impact of new analytics to create cost-optimized workflows, especially when deployed in conjunction with command and control systems and information security management and decision management suites.
- For customized projects, determine the available data assets, and build an appropriate problem taxonomy that will address your business objectives.

Business Impact: The benefit rating for advanced video analytics is transformational as these technologies will become essential to interpret automatically the world in both visual and nonvisual spectrums (including infra-red, laser and spectral analysis). Anomaly detection, object and behavior recognition and complex scene understanding are being enabled by advances in image and video analytics. Beyond, security, automotive, robotics, retail and commercial, advanced video/image analytics has huge and growing applications potential in healthcare, manufacturing, supply chain/logistics, banking and finance, government and media industries.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Amazon Web Services; Cortexica; Deepomatic; Google (DeepMind); Herta; iOmniscient; Microsoft Azure; Trax; viisights

Recommended Reading: "Innovation Insight for Video/Image Analytics"

"Market Trends: Facial Recognition for Enhanced Physical Security — Differentiating the Good, the Bad and the Ugly"

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

"Market Insight: Vision Processing Units — Enabling IoT Endpoints With AI-Based Computer Vision"

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

"Competitive Landscape: Computer Vision Platform Service Providers"

Augmented Data Management

Analysis By: Donald Feinberg; Merv Adrian

Definition: Augmented data management refers to the application of AI and ML for optimization and improved operations. AI and ML are applied — based on the existing usage data — to tune operations and to optimize configuration, security and performance. They are also applied to create, manage and apply policy rules within the different products, such as metadata management, master data management, data integration, data quality and database management systems.

Position and Adoption Speed Justification: Artificial intelligence (AI) and machine Learning (ML) are not new in data management products — significantly replacing manual data management tasks with automated capabilities, altering job roles, product design and overall data management processes. Since the inception of cost-based query optimization in the late 1970s at IBM, rudimentary ML (or rules management) has been used for query optimization in many DBMS products. Initially based on the collection of operational and usage statistics, this has now dramatically expanded across large numbers of deployed instances, especially in the cloud. With the increased availability of ML and AI libraries, vendors have begun using modern AI and ML (such as self-learning) well beyond query optimization. Augmented data management products have the ability to examine large samples of operational data, including actual queries, performance data, schema and metadata. These solutions can not only tune and optimize the use of the products themselves based on actual usage, including failures and poor performance, but also suggest and implement new designs, schemas and queries. They can even infer the semantics and associations of the data in order to recommend structural improvements.

These capabilities are currently being implemented in data integration, data quality, master data management (MDM), metadata management and DBMS software. This profile therefore represents an aggregate view and position of what is happening across data management. Much of the effort to use AI and ML in these products is in its infancy, but growing fast. Its slow movement on the Hype Cycle (despite ML's longevity in the DBMS) is due to the many use cases other than query optimization. Also, many of the newer augmented data management products are available only in cloud platforms — such as Amazon Aurora and Oracle Autonomous Database, and data management software that has made the transition to the cloud. In these platforms, enormous volumes of user data on a consistent infrastructure improve the applicability of the results and offer opportunities for the continuous training and retraining of models. As a result, they are being aggressively used to drive competitive improvements in cloud-based offerings. We believe the improvements, initially available in cloud platforms, will neutralize the distinction between cloud and on-premises over the next few years.

User Advice: For data and analytics leaders focused on data management capabilities, we recommend you:

- Question the vendors of your data management tools about their roadmap for the introduction of AI and ML into their products.
- Begin testing the components of augmented data management products (where visible) to understand their capabilities and the validity of the automated functionality. Audit the results: with any new functionality, there is the risk of introducing errors and reduced performance.
- Create a business case for using these new tools, and be sure to include the benefits realized from the resources that will be released for other functions of greater business value.
- Make augmented capabilities a “must have” selection criterion for new purchases of data management products.
- Begin seeking data management solutions that share design and performance metadata for use.

Business Impact: Augmented data management will offer benefits in the following areas:

- Metadata management — Increasingly, AI and ML are used to explore and define metadata from the data, helping the analysts to evaluate metadata more rapidly, accurately and with reduced redundancy.
- Data integration — To automate the integration development process, by recommending or deploying repetitive integration flows.
- MDM — MDM solution vendors will increasingly focus on offering AI and ML-driven configuration and optimization of record matching and merging algorithms as a part of their information quality and semantics capabilities.
- Data quality — AI and ML will be used to extend profiling, cleansing, linking, identifying and semantically reconciling master data in different data sources, to create and maintain “golden records.”
- DBMS — In addition to enhancing cost-based query optimization, AI and ML will be used to automate many of the current manual management operations, including the management of configurations, storage, indexes and partitions, and database tuning.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Amazon Web Services; IBM; Informatica; Oracle; Orchestra Networks; Panoply; SAP; SnapLogic; Teradata

Python

Analysis By: Erick Brethenoux; Alexander Linden; Jim Hare

Definition: Python is a general-purpose, dynamic and object-oriented programming language that is very popular across multiple disciplines, especially in data science and machine learning. It is implemented within a stand-alone interpreter; however, versions of Python also run within the Java and Microsoft .NET platforms.

Position and Adoption Speed Justification: Python is a top language for data science. It continues to secure followers due to its growing user community, ease of use and friendliness to application development. Python has become a de facto programming language leveraged by all data science platforms. Most data science and machine learning platforms (as well as some analytic and BI platforms) have added Python support as a standard capability. Anaconda, for instance, has specialized in Python, offering a federated development capability centered around Python libraries and the community of developers who continually grow the large amount of Python code. Google's TensorFlow also offers first-class Python binding, and many cutting-edge deep-learning and other data science libraries are either based on Python or offer first-class Python binding. However, as organizations are hitting the operationalization wall (i.e., facing difficulties to put their models into production with the lack of Python operationalization libraries), even Python's popularity cannot prevent it slipping toward the Trough of Disillusionment.

User Advice: Data and analytics leaders should:

- Look for data scientists skilled in Python and/or R, as they are the two most popular languages used in development processes in advanced analytics and machine learning. Compared with R, Python is more cutting edge in a variety of disciplines, such as link analysis, deep learning, natural language processing and scientific computing. However, R is still currently more widely adopted and used by experienced data scientists. Scala becomes an important language when considering Apache Spark as a deployment capability.
- Implement guidelines for when and how to use Python. While Python is great for machine learning, it is not a universally great choice for everything. It doesn't excel in large and complex application scenarios. Its enterprise-readiness is similar to R, but is not equal to C#, C++ or Java.
- Implement guidelines for when and how to use which versions of Python. Version 3.7.x is the latest version, but the version 2.x libraries are still widely used and supported.

- Hire complementary talent in more conventional programming languages. Despite Python's popularity in model development, a large majority of operationalized machine learning models are written in other more traditional languages.

Business Impact: For almost every data science team, Python is the cutting-edge, go-to programming language. It is increasingly used by developers and IT departments to develop machine learning models. Open-source capabilities epitomized in Python will continue to be important to the future of data science and machine learning. Python is making skills more transferable, helping to address the data science resources gap. It is also resulting in faster innovation cycles as communities of thousands of users and enthusiasts are able to collaborate to make cutting-edge innovations much faster than is possible with proprietary software approaches. If the Python community succeeds, in the coming two years, in providing data science teams with robust machine learning production libraries, Python still has the opportunity to bypass the Trough of Disillusionment to steadily climb the Slope of Enlightenment.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: Anaconda; BigML; Dataiku; Domino Data Lab; Enthought; Google; H2O.ai; IBM; Microsoft; Project Jupyter

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

"Critical Capabilities for Data Science and Machine Learning Platforms"

"How to Choose the Right Data Science and Machine Learning Platform"

Spark

Analysis By: Svetlana Sicular

Definition: Apache Spark is an open-source in-memory distributed computing framework for large-scale data processing and analytics. It uses a memory-centric processing model to support performance for highly iterative tasks, such as query processing, data streaming, machine learning and graph analysis. Spark can be deployed both on-premises and in the cloud.

Position and Adoption Speed Justification: Apache Spark initially gained popularity because it aimed to simplify many data engineering tasks that required a stack of tools. Spark is now embedded in many data management and analytics products. The focus of Spark is shifting from data processing to machine learning and analytics, and from on-premises to the cloud. This causes changes to the Spark ecosystem and challenges the implementers to clearly define Spark's role.

While Spark continues to attract data scientists and data engineers, it has become another tool in rapidly proliferating toolboxes of data pipelines' development and machine learning. Meanwhile, the creators of Spark (mostly employed by Databricks) are introducing new commercial and open-source capabilities, such as MLflow and Delta Lakes, outside Apache Spark, leaving this open-source framework less directed. Hadoop distributions, although marked as obsolete before plateau in 2017, continue to be part of market offerings and include Spark as their dominant feature. While Spark is well-understood for its Hadoop-related data processing, its graph and machine learning capabilities are the less clear areas of Spark applicability. Time to plateau is currently estimated at two to five years, but Spark may also become obsolete before plateau as a separate package if vendor competition slows the Apache Spark open-source project, or if a disruptive alternative to Spark draws talent and attention from the current Spark ecosystem.

User Advice:

- Use Spark as a general-purpose engine if you have large datasets or if your vendor relies on Spark in its own solution. D&A leaders who seek flexibility of programming languages for large-scale data processing and analysis in various ways, such as SQL, ETL, data streaming and machine learning, should employ Spark as a general-purpose distributed compute engine, even given the volatility in the machine learning space.

- It takes time and effort to get the Spark design right, but once implemented, Spark is a relatively low-risk, portable method of experimentation with machine learning, when underlying data is represented in large datasets. Spark could be also viewed as a gateway to other popular software frameworks, for example, TensorFlow, Keras or PyTorch. Spark is less effective for small datasets and for general-purpose BI and analytics due to limits in result presentation and retrieval.

- Spark is available on all major cloud platforms, and therefore, it could be the means of portability among the clouds. However, D&A leaders should monitor the rapidly changing cloud landscape for alternative solutions that could be disruptive to Spark or could represent a best fit for their use cases, rather than a general-purpose solution.

- Apache Spark doesn't have its own storage, and disruptive storage or platform offerings could overturn Spark's popularity.

Business Impact: As Spark is available in multiple cloud offerings, the fact that Spark is open-source is overshadowed by commercial solutions that support it. It leads customer selections of data engineering and ML tools to a "fit for purpose" reasoning, rather than to a choice of Spark as a starting point. The shift from big data to ML in the Spark community makes the long-term impact of Apache Spark uncertain, with no obvious leader in ML so far.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amazon Web Services; Cloudera; Databricks; Google Cloud Platform; IBM; MapR Technologies; Microsoft Azure

Recommended Reading: "An Introduction to and Evaluation of Apache Spark for Big Data Architectures"

"Magic Quadrant for Data Science and Machine Learning Platforms"

"Critical Capabilities for Data Science and Machine Learning Platforms"

"Magic Quadrant for Data Management Solutions for Analytics"

"Market Guide for Database Platform as a Service"

Climbing the Slope

Predictive Analytics

Analysis By: Peter Krensky; Alexander Linden; Carlie Idoine

Definition: Predictive analytics is a form of advanced analytics that examines data or content to answer the question, "What will 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. Levels of project underperformance and ROI failure are low and this technology has quickly crossed 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 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 or an AI cloud service. However, packaged applications pretrained models do not exist for every analytics use case. Packaged applications and AI cloud services 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 (buy, build, 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; DataRobot; H2O.ai; IBM; KNIME; MathWorks; Microsoft; RapidMiner; SAS; TIBCO Software

Recommended Reading: "Combine Predictive and Prescriptive Techniques to Solve Business Problems"

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

"Magic Quadrant for Data Science and Machine Learning Platforms"

"Critical Capabilities for Data Science and Machine Learning Platforms"

Notebooks

Analysis By: Peter Krensky; Alexander Linden

Definition: Notebooks for data science are user interfaces that enable users to create active documents where they can write and run code, display results (via various visualizations) and annotate, collaborate and share their work. Most are web-based and many also offer a desktop version. Notebooks can be used for many different data science tasks, including ad hoc data ingestion, preparation, exploration and machine learning.

Position and Adoption Speed Justification: Notebooks have continued a recent and rapid rise in adoption and deployment. This rapid proliferation is attributable to the sustained popularity of numerous coding languages (primarily Python and R) in conjunction with a larger and more active data science and machine learning talent base. Most notebooks are open source, they can often support multiple languages in a single document (some also include a converter for working between languages). Notebooks have passed clear across the Trough of Disillusionment and are already climbing toward the Plateau of Productivity. This is largely due to their great appeal to the rapidly growing class of data scientists and citizen data scientists, and their clearly demonstrated value as agile tools integrated with the ever-expanding open-source stack. The proper use and scale for notebooks was quickly identified as the technology came to prominence, and cautionary tales and project failures around notebooks are few. The organic popularity of notebooks has been accompanied by a wave of vendor support. Notebooks have first-class support and integration on the majority of data science and machine learning platforms. New users will continue to be attracted by notebooks' visual and conversational approach and their nimble footprint as a very agile technology. Hosted notebooks, where multiple users can share an environment and resources, are also gaining in popularity.

- User Advice: Data and analytics leaders should encourage experienced data scientists to experiment with shareable notebooks, in order to improve collaboration and enable others to visually follow workflows.
- Data scientists can use notebooks to explain methodologies, call out impactful insights and identify areas ripe for further analysis. Data science projects can continually evolve in multiple versions, with users following the thought processes of the original author as they edit different code blocks and tweak visualizations.
- Business analysts and code-oriented citizen data scientists should seek out data scientists in their organization, using notebooks to learn their methods and emulate their work.
- Data and analytics leaders should ensure that their teams standardize on a single type of notebook, to encourage collaboration and coherence.

Business Impact:

- Knowledge transfer: Data science notebooks can be used to create intelligible presentations of complex data science projects, making them more accessible to users who are less technically sophisticated.
- Auditing and validation: Helping data scientist teams to replicate and confirm their results. This will lead to more-accurate analysis and models and, ultimately, a greater volume of data science projects.
- Collaboration: Enabling the analytical minds distributed throughout an organization to collaborate — sharing methodologies and findings. However, sustaining collaboration between different personas and stakeholders via notebooks is challenging.
- Integration across user communities: Users can always use the optimal coding language for a given task, taking advantage of a language's individual strengths.
- Portability: Work done in data science notebooks (particularly the highly popular Jupyter Notebook) is easily transferrable across different vendor platforms.

- Agility: Data science notebooks are an ideal tool for an "act-observe-plan-repeat" cycle and can be easily modified and deployed in new areas.

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Apache Software Foundation; Databricks; Google; MathWorks; Project Jupyter; Two Sigma; Wolfram Research

Recommended Reading: "Five Ways Data Science and Machine Learning Deliver Business Impacts"

"Pursue Citizen Data Science to Expand Analytics Use Cases"

"Magic Quadrant for Data Science and Machine Learning Platforms"

"How to Use Storytelling to Sell Your Data Science Projects"

"How to Choose the Right Data Science and Machine Learning Platform"

Text Analytics

Analysis By: Stephen Emmott; Alexander Linden; Marko Sillanpaa

Definition: Text analytics is the process of deriving business insight or automation from text. This process can include determining and classifying the subjects of texts, summarizing texts, extracting key entities from texts, and identifying the tone or sentiment of texts.

Position and Adoption Speed Justification: Text analytics addresses a diverse range of uses from general capabilities, to extracting data from textual content, to industry-specific and line-of-business use cases. Typically, vendors supporting text analytics do so in the form of applications that, although built on a general capability, are tailored to specific categories of use case (for example, voice of the customer in the context of customer relationship management). As well as dedicated applications, text analytics capabilities are embedded into many other products (contract life cycle management suites, for example).

A surge in the volume of volume of textual data, especially from sources other than traditional "documents" (such as instant messages and automatically extracted metadata) has fueled the evolution of text analytics. Another strong driver is the desire to complement insights gleaned from analysis of structured numerical data with text-based facts for more robust predictive modelling.

Text analytics is a proven technology that is well adopted, albeit in multiple guises. However, several factors hinder the emergence of more pervasive, easy-to-use business solutions for text analytics. The technology is still maturing, and differentiation between the many overlapping vendors is too nuanced for those organization without in-house expertise. Although easier to use, it is still challenging to incorporate solutions into an organization's wider digital platform, given the diversity of use cases and specialist skills needed to utilize and gain benefit.

With scope for further evolution, notably the increasing use of machine learning to process text, its current position reflects steady progress up the Slope of Enlightenment.

User Advice: Text analytics is a natural language technology and a form of artificial intelligence (AI). It is therefore important to position it as such.

Given this context:

- Identify and prioritize use cases that text analytics can address.
- Review the text analytics market to acquaint yourself with its vendors, products and capabilities.
- Start with text analytics applications that are provided as prepackaged products designed for nontechnical business users to administer for well-established use cases. These could include voice of the customer (VoC) or employee (VoE).
- Select products based on how well they suit specific business scenarios, and be clear when identifying these.
- Engage with business and analytics leaders to identify relevant initiatives and activities elsewhere in the organization, and especially those involving the use of unstructured data for analytical purposes.
- Look for the ability to integrate with other applications that use unstructured data, such as content services or conversational agents.
- With more use cases and a need to ingrate text analytics into your wider digital business platform, consider text analytics platforms or components.
- Assess references from companies of a similar size, or demand comprehensive proofs of concept from vendors where references are unavailable or ill-matched.
- Allow a realistic lead time to recruit text analytics talent. If the requisite skills are in short supply within the team, consider working with a third-party analytics service provider that offers text analytics capabilities.

Business Impact: In many use cases, text analytics, when combined with various other analytic capabilities, may be of significant benefit to an organization in the following areas:

- Preprocessing unstructured data for analysis (for example, ingesting data from forms captured via OCR for onward processing).

- Automated document matching and classification (analyzing documents and matching metadata to them from a controlled vocabulary).
- Discovery and insight (indexing reports in preparation for natural language question and answering).
- Sentiment (analyzing notes, social medical, or transcripts to identify the author's attitude about a subject).

Use different combinations of technologies for different business use cases:

- Healthcare — medical records analysis by mapping key medicate terms into a graph for analysis.
- Insurance — identifying fraudulent claims by analyzing the narratives and identifying common individuals across claims.
- Finance — gain insights on investments by monitoring public information sources and social media.
- Legal — supporting contract review by extracting key terms and obligations from complex contracts.
- Retail — identifying fraud patterns by analyzing warranty claims.
- Marketing — monitoring brand loyalty and sentiment by analyzing social media feeds and customer feedback.
- Law enforcement — forensic analysis of a body of documents by identifying key subjects and dates, and developing a chain of events.
- Digital publishing — Identifying related articles and developing a summary relevant to an article in progress.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Amenity Analytics; Cambridge Semantics; Clarabridge; Digital Reasoning; IBM; KNIME; Lexalytics; Megaputer; OdinText; SAS

Recommended Reading: "Artificial Intelligence Primer for 2019"

"Market Guide for Text Analytics"

“Toolkit: Text Analytics Vendors”

“Four Data Preparation Challenges for Text Analytics”

“How Chief Data Officers Can Succeed by Driving Analytic Value”

“Understanding Your Customers By Using Text Analytics and Natural Language Processing”

Traditional Model Management

Analysis By: Peter Krensky

Definition: Traditional model management consists of proven techniques and technology to streamline the prioritization, creation, operationalization and execution of predictive models. It supports version and access control, model performance tracking, scheduled or condition-based recalibration of models and also serves as a library to facilitate end-users’ access and reuse of completed models. The hype around innovative technologies and roles to support model operationalization is captured in the MLOps profile.

Position and Adoption Speed Justification: Traditional model management has been around for many years and is entering the Plateau of Productivity. Their importance persists as organizations apply predictive analytics and machine learning on a greater scale, and as the regulatory requirements around predictive models increase in certain industries. More vendors are adding and expanding model management capabilities to their data science platforms (see “Critical Capabilities for Data Science and Machine Learning Platforms”), and stand-alone products are also available. Existing offerings continue to mature as customers expect these capabilities, because they work in heterogeneous modeling environments with multiple tools and platforms. The commonly established volume and complexity of models in production within most organizations cements the criticality of sound model management, though hype is largely shifting to concepts like MLOps and other deployment challenges that arise within modern data science teams.

User Advice: Data and analytics leaders should deploy traditional model management capabilities in the following scenarios:

- For organizations with many models, model managers should be used to monitor their life cycle, particularly for retirement or renewal. In addition, model managers can prioritize and promote models smoothly from the test environment to production.
- For organizations with models in production that are essential to core digital business, traditional model managers are a must-have capability for governance and monitoring.
- For regulated environments, model managers should be used to support compliance and auditability.

- For heterogeneous model environments (multiple tools), model managers should be a focal point for publishing, deployment and visibility .

- For organizations with multiple commissioning groups, research model managers should enable collaboration, filtering and prioritization of the model creation agenda.

Tools vary in sophistication. Organizations should look at the spectrum of needs for their model management. Analytics decision makers should ask if they need something basic, such as version control, or more sophisticated capabilities such as automatic recalibration of models.

Business Impact: Data and analytics leaders should look to MLOps capabilities to support the next generation of machine learning models. For handling large portfolios of legacy models, traditional model management tools and functionality should provide the following:

- Catalog — A “centralized” way to store and secure analytical assets to make it easier for analytics professionals to collaborate and to allow them to reuse models or other assets as needed (this could be a secured community or a collaboration space).

- Governance — Protocols to ensure adherence to all internal and external procedures and regulations: not just for compliance reasons, but, because an increasing amount of data gets aggregated, to address potential privacy issues.

- Capabilities — Automated versioning, fine-grained traceable model scoring and change-management capabilities (including championing/challenging features) to closely test, monitor and audit analytical asset life cycles from a technical as well as a business performance perspective (through KPIs).

- Coherence — Establish simple protocols to provide functional bridges between the development and operationalization cycles; enhance the cooperation and consultations between the development and operationalization teams; provide efficient “translation” services between data science and the lines of business; improve the transparency of deployed analytical assets.

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Alteryx; FICO; IBM; KNIME; MathWorks; Microsoft; RapidMiner; SAP; SAS

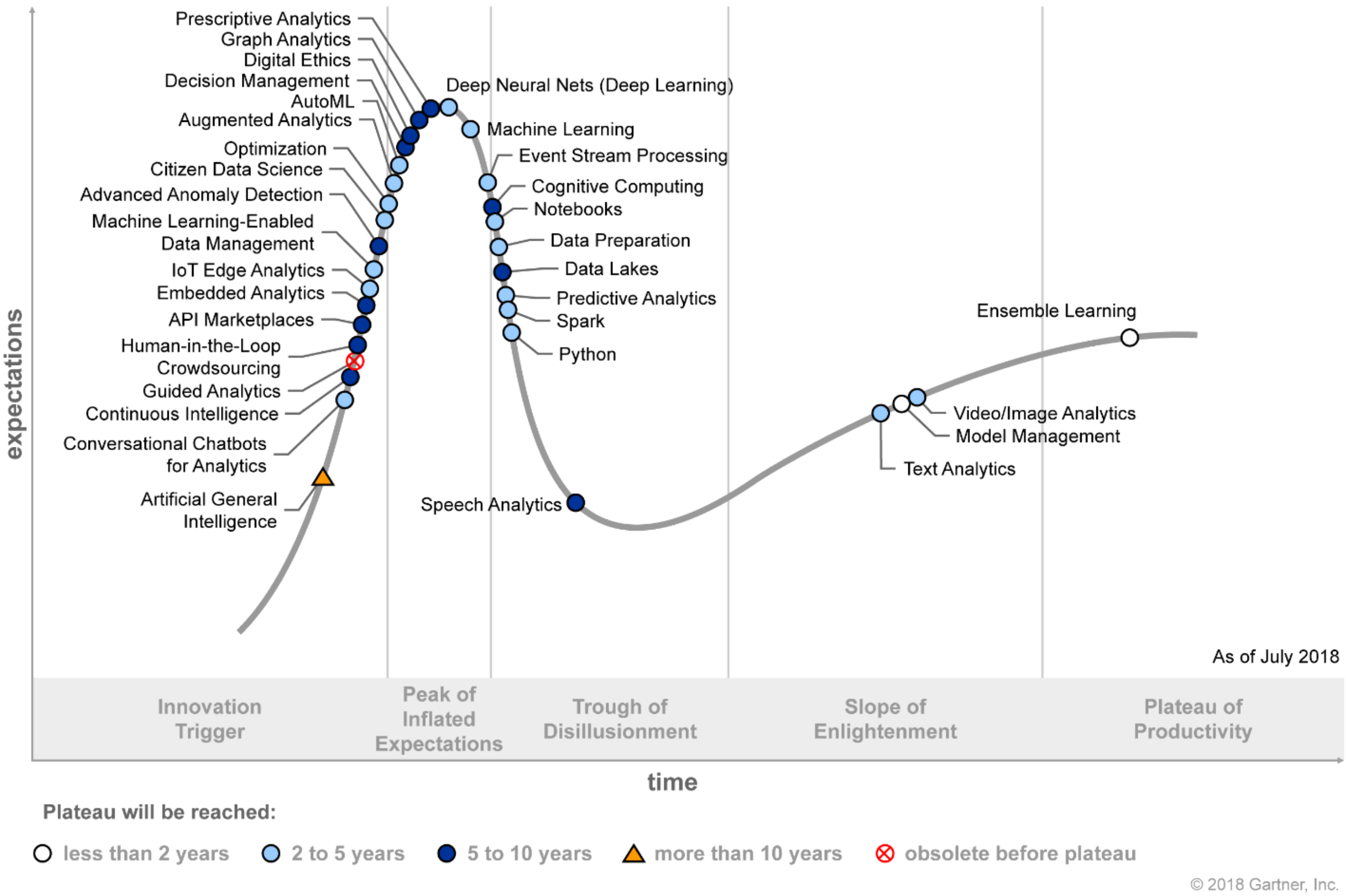
Recommended Reading: “Magic Quadrant for Data Science and Machine Learning Platforms”

“Critical Capabilities for Data Science and Machine Learning Platforms”

“Ignition Guide to Building a Data and Analytics Governance Program”

Appendixes

Figure 3. Hype Cycle for Data Science and Machine Learning, 2018



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.

Phase	Definition
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 (August 2019)

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 (August 2019)

Table 3: Maturity Levels

Maturity Level	Status	Products/Vendors

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 technologyVendors, 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 (August 2019)

Document Revision History

Hype Cycle for Data Science and Machine Learning, 2020 - 28 July 2020

Hype Cycle for Data Science and Machine Learning, 2018 - 23 July 2018

Hype Cycle for Data Science and Machine Learning, 2017 - 28 July 2017

Hype Cycle for Data Science, 2016 - 25 July 2016

Hype Cycle for Advanced Analytics and Data Science, 2015 - 6 July 2015

Recommended by the Authors

Understanding Gartner's Hype Cycles

Hype Cycle for Data Science and Machine Learning, 2017

Hype Cycle for Analytics and Business Intelligence, 2018

Hype Cycle for Artificial Intelligence, 2018

Machine Learning: FAQ From Clients

Magic Quadrant for Data Science and Machine-Learning Platforms

Critical Capabilities for Data Science and Machine Learning Platforms

Artificial Intelligence Hype: Managing Business Leadership Expectations

Maximize the Value of Your Data Science Efforts by Empowering Citizen Data Scientists

Recommended For You

Tool: Customer Foresight Enablement Guide to Discover Durable Customer Changes

Achieve Successful Managed Mobile Services Deals by Avoiding These Pitfalls

Revisit Sales Role Design to Increase Seller Effectiveness

Critical Steps for Product Managers: Measuring Business Unit Performance

Building Inclusive Leadership To Enable Future Success objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."