Deep Learning Enables a Quantum Leap in Content Processing

Published: 29 September 2017 **ID:** G00327965

Analyst(s): Alexander Linden

Deep learning is one of the biggest artificial intelligence technology wonders of the past decade. Most progress is happening around text, audio and image processing. For data and analytics leaders, it will enable a quantum leap in those areas of their operations that are especially content-rich.

Impacts

- Deep learning will enable a quantum leap in content processing that data and analytics leaders must prepare for: from image and speech recognition, and machine translation, to content creation and much more.
- Deep learning has superior capabilities to fuse/integrate data from various event streams; data and analytics leaders should be aware that it will, therefore, also enhance their corporate data science projects, especially those where multiple data types are involved.
- Initially, this corporate impact will materialize only for the most advanced organizations with fitting tasks; taking two to three years longer for the rest, once their nascent deep learning infrastructure has the chance to mature further.

Recommendations

Data and analytics leaders responsible for analytics and BI strategies should:

- Avoid building their own deep learning capabilities, in most circumstances, in favor of prepackaged options. They should, however, upskill their data science experts so that they can at least evaluate those prepackaged solutions.
- Create their own deep learning solutions in certain situations; especially in mass-scale application scenarios, which can best benefit from deep learning's superior data fusion and prediction capabilities.
- Progress with care, because deep learning is highly demanding in terms of both skill and infrastructure.

Strategic Planning Assumption

By 2020, deep learning will dominate machine learning, especially where big data, content processing and many data sources are involved.

Analysis

There is intense ongoing excitement in the field of machine learning, especially around deep learning. It has been compared with the so-called "Cambrian Explosion" in Earth's history — about 550 million years ago — when, during a small time window of a few million years, most of the earth's life forms appeared in a sudden rush (as evidenced by fossil records).

Similarly, deep learning architectures are emerging quite suddenly after 20 years of relative quietness; bursting into the limelight and solving problems that have resisted a sound technical solution for many decades. Many observers agree that this is akin to a "gold rush" for entrepreneurs and internet giants alike, and further contributes to a tremendous shortage of skills on the corporate side.

Data and analytics leaders, and IT and innovation leaders, must keep abreast of developments in order to understand where and how progress will happen (see Figure 1 for our summary of the impacts and top recommendations relating to this trend). They need to be well-prepared and should not waste time and effort on old technology threads — such as trying to create complex chatbots with classic natural-language processing (NLP) technology — many of which are quite likely to be superseded.

Figure 1. Impacts and Top Recommendations for Data and Analytics Leaders

Impacts	Top Recommendations
Deep learning will enable a quantum leap in content processing.	• Avoid creating your own deep learning capabilities; rely instead on an evolving array of packaged solutions with deep learning capabilities embedded.
Deep learning has superior capabilities to fuse data from various event streams.	• In scenarios such as demand forecasting, fraud detection, and failure and quality prediction, deep learning's superior data fusion capabilities can give a high ROI.
Initially, the benefits of deep learning will only materialize for the most advanced organizations with fitting tasks.	• Proceed with care, because deep learning is highly demanding from the perspectives of skill and infrastructure.

ROI = return on investment.

Source: Gartner (September 2017)

Impacts and Recommendations

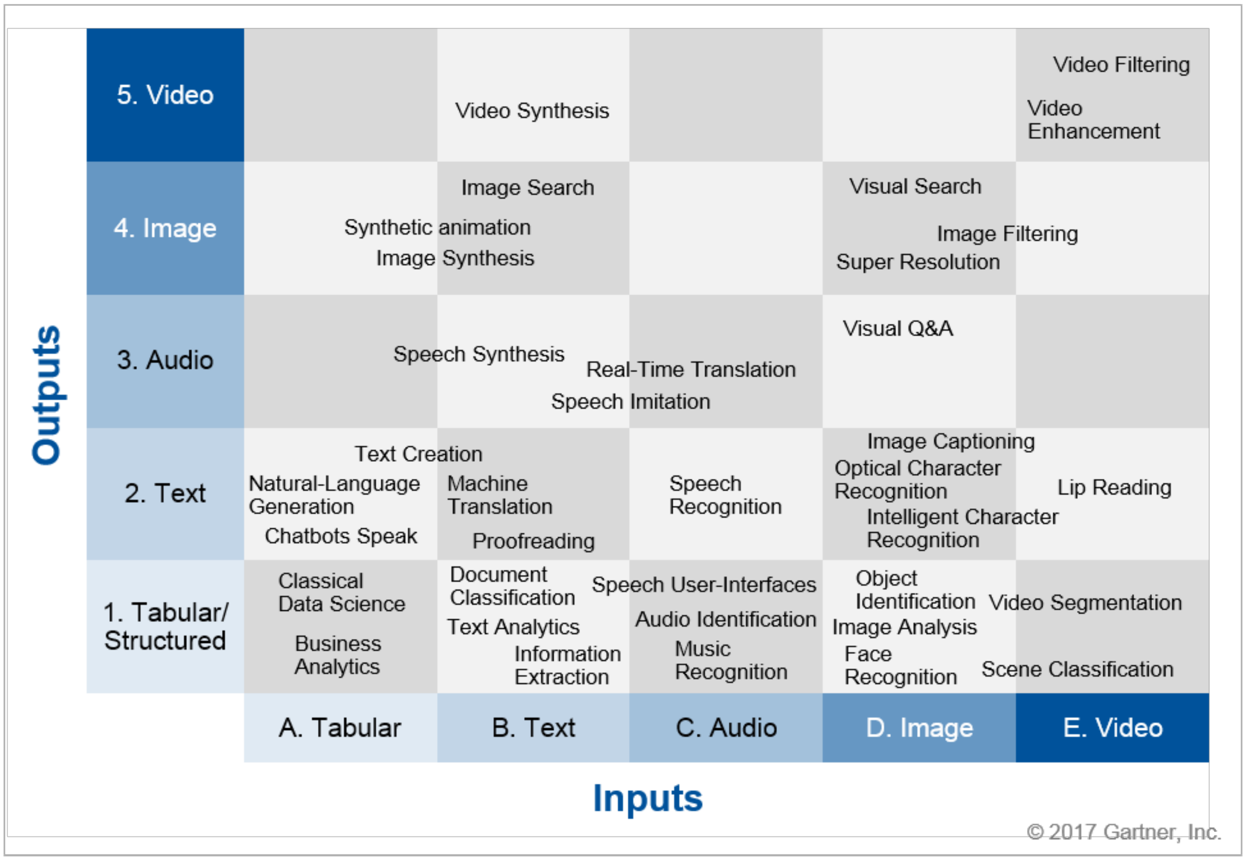
Deep learning will enable a quantum leap in content processing that data and analytics leaders must prepare for: from image and speech recognition, and machine translation, to content creation and much more

It is a common belief that supervised machine learning performs mainly classification and regression tasks where the output is just one value — for example, indicating a class membership (such as fraud or not fraud, churn or not churn). However, deep learning has shown remarkable progress in producing much richer content in its output layers and is therefore increasingly capable of performing transcoding or transformation.

Content is defined as data that is meant for human consumption; for example, it can be listened to (audio), or looked at (images and video) or read (text). As such, content is one of the big cornerstones of human civilization and communication. All content is data — but there are classes of data that cannot be classified as content (for example, sensory data or log-streams). In this research note, we focus on content, because this is where the deep learning revolution is showing most progress.

In Figure 2, we attempt an overview of a content processing matrix, where the horizontal and vertical axes describe the most prominent input and output types, respectively.

Figure 2. Gartner's Content Processing Matrix



Source: Gartner (September 2017)

Today's most prevalent case is positioned as A1 (business analytics and data science), where both the input and the output are simply tabular or (so-called) structured data. Here, traditional classification and regression tasks such as churn prediction, next-best-action, demand prediction and fraud detection reside.

However, the content processing matrix also contains several remarkable breakthroughs from the past five years:

1. Object identification (D1) for example, where in just six years the error rate for the ability to identify objects in images has dropped from about 25% to just 2% currently (better than humans).[1] This exemplifies one of the toughest computer science problems of the past 20 years. Relevant sample vendors can be looked up in the "computer vision" entry on the "Hype Cycle for Artificial Intelligence, 2017."
2. Speech recognition (C2), the conversion of spoken language as an audio stream into digital text, is the second discipline where massive progress has been made during the past two years. It has just recently been reported that some systems now match human performance; for example, a system from Microsoft,[2] whereas Baidu, Google and IBM were already closing in on

This research note is restricted to the personal use of SVarma2@scu.edu.

human-level performance. For sample vendors, see the "speech recognition" entry in the "Hype Cycle for Artificial Intelligence, 2017."

3. Machine translation (B2), where Google provided a huge leap forward with its Google Neural Machine Translation (GNMT) project — making progress in just under a year that wasn't possible during the previous five years. The Neural Machine Translation project provides state-of-the-art, sentence-level translation capabilities that were rolled out for public use in October 2016. Other vendors following suit include Baidu, DeepL and Systran.[3]

4. Lip reading (E2), where Google's DeepMind has shown that its system performs far better than professional lip-readers.[4]

Also, speech synthesis or text-to-speech (B3) is progressing rapidly (see Table 1 for a concise definition of the various kinds of content processing). This has been a niche — largely in existence to help the visually handicapped or to read email while driving in a car. Recently, however, due to deep learning, the performance has made a quantum leap and the voices are becoming much more realistic (for example, Amazon Polly, Baidu's Deep Voice and Samsung's Innoetics). Furthermore, voices can be imitated (speech imitation; B3/C3 in Figure 2) to sound like a politician or any other popular speaker, and they can be parameterized to obtain characteristic sounds such as optimistic, sceptic, mellow, aggressive and so on.[5]

This research note is restricted to the personal use of SVarma2@scu.edu.

Table 1. Quick Glossary of Content Processing Types

Content Processing	Description	Real-World Scenarios Affected
Document Classification	Maps text to a set of classes (e.g., spam or not, "about sport" or "about politics").	Content filtering, sentiment analysis, help desk/call center routing.
Image Search	Maps a textual description (e.g., a list of keywords) to find fitting images from an image repository.	Graphical design, product search.
Lip Reading	Takes a video of someone speaking and converts this into the text that is spoken.	Speech recognition in very noisy environments; spying on people in the distance.
Machine Translation	Takes text in the source language and produces the text with the same meaning in the target language.	International communication.
Object Identification	Takes an image as input and creates a structural description of the objects found in the image.	Retail (e.g., recognizing which articles are missing on a shelf).
Speech Synthesis/ Text to Speech	Takes a text and produces an audio file containing the spoken text.	Handsfree consumption of text and emails, enhanced customer interactions.
Visual Search	Uses an image to find similar images from a given image repository.	Product search, graphical design, person search.
Voice Imitation	Takes text or audio and produces audio that sounds like a specific person.	Advertising.

Source: Gartner (September 2017)

All this success is rooted in deep learning's capability to exploit even very weak signals (on the pixel-level) — where each signal considered alone has almost no meaning, but when together with other features constitute something meaningful.

As with the example above, other recognition disciplines such as optical character recognition or intelligent character recognition (see D2 in Figure 2), which convert images of *typed or written* text into digital text, are also benefiting significantly from deep learning technologies and are reaching a super-human level of ability.

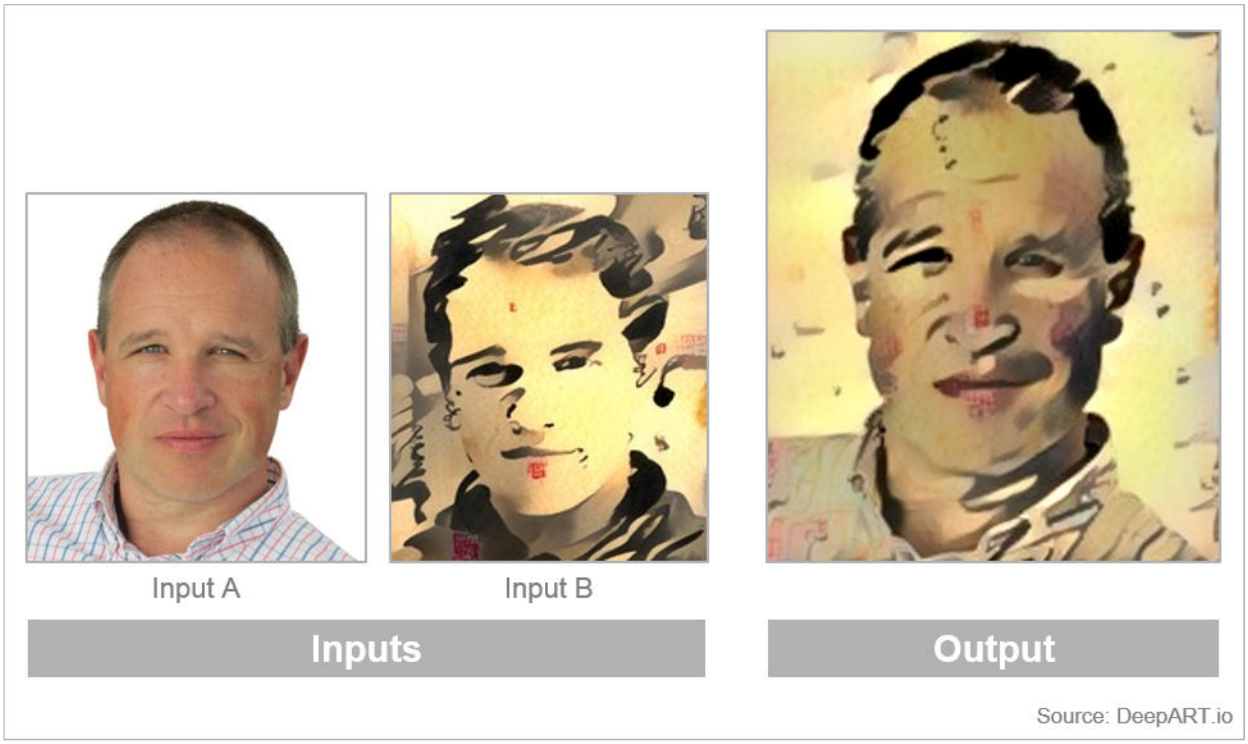
Generally, Gartner believes that NLP will enjoy massive growth. So far, niche text analytics (B1) is poised for an explosion in novel capabilities. This is largely driven by a recent breakthrough in vectorising words, which describes how words are being mapped onto numerical vectors.[6] It was the major driver behind the success of the Google Neural Machine Translation project and it must be a wake-up-call for the text analytics and NLP industry, which has been stagnating for the past five years.

In the Matrix (Figure 2), chatbots — also called conversational user interfaces — stretch across B1 (for the information extraction part) and A2 (for the text creation part) — For more information on conversational user interfaces, see the relevant profile in the "Hype Cycle for Artificial Intelligence, 2017."

In the same way, computer vision (see the corresponding profile in "Hype Cycle for Artificial Intelligence, 2017") will enjoy similar tremendous growth. Examples of this technology include:

- Super resolution imaging, where a system can enhance the resolution and the detail based on its knowledge about context.[7]
- Style transfers, where deep learning systems "learn" the style of certain images and apply them to another image (see Figure 3).

Figure 3. Content Synthesis



From left to right: A deep neural net can take a sample image (from the author in this case) and an artist's portrait and produce a novel output based on both input pictures.

Source: With Permission of DeepART.io

It will only be a question of time until — for most categories in the content processing matrix — machine capabilities surpass human capabilities; not only in speed but, more interestingly, in accuracy, consistency and price (for quite a number of these tasks, it is already true).

This research note is restricted to the personal use of SVarma2@scu.edu.

Overall, according to Gartner's estimate, there are about 10,000 ongoing projects in this space: one-third of them at the top 10 IT/internet giants (such as Amazon, Baidu, Facebook, Google, IBM, Microsoft, SAP and Yandex); another third at the thousands of startups in that domain; and approximately another third in academia and similar for-profit or governmental think tanks. This entire domain is somewhat like a "gold rush," with the most highly skilled deep learning specialists being pulled in — resulting in even greater scarcity of available talent in the already dried-up market for data science skills.

Recommendations for data and analytics leaders:

- Avoid, in general, creating your own deep learning capabilities. Instead, rely on an evolving array of packaged solutions that have deep learning capabilities embedded.
- Most benefit from content identification and enhancement can be gained in the following industries: advertisement, entertainment, government, military, retail, transportation and surveillance.
- In many contexts, humans are still superior; for example, in audio transcription with a lot of background noise. Here, human-based online services such as CastingWords, Rev services and services from One Hour Translation will give superior accuracy.

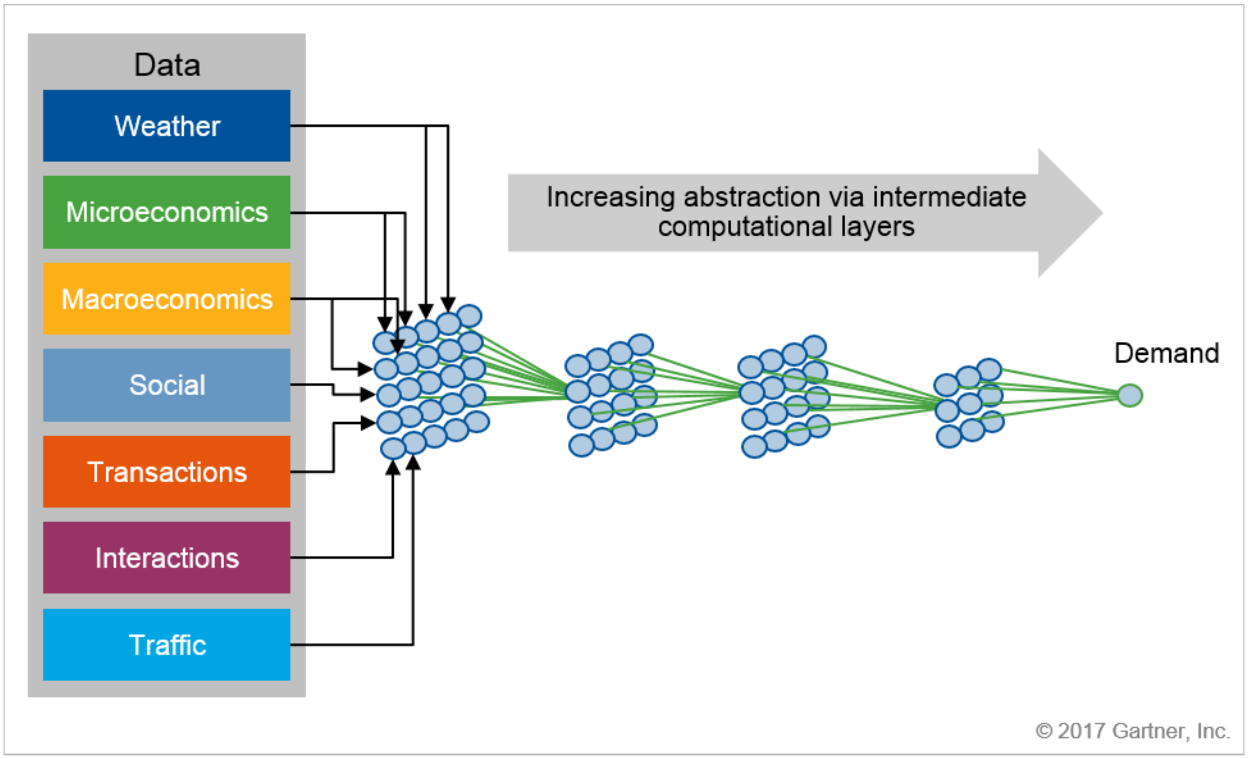
Deep learning has superior capabilities to fuse/integrate data from various event streams; data and analytics leaders should be aware that it will, therefore, also enhance their corporate data science projects, especially those where multiple data types are involved

The major characteristic of deep learning is its ability to learn suitable representations of the raw input data. Theoretically, therefore, it can itself derive domain-specific features during learning, which even experienced data scientists would find notoriously difficult. The engineering of good features is not only highly domain specific, but also ad hoc and prone to error. Even worse, it has proven to work against the plan, because lots of relevant information can be inadvertently removed from the raw data as well. The notion of having the deep learning methodology figure out the features itself from the raw data has been referred to as "end-to-end deep learning" and is currently considered one of the most promising ways to accomplish data fusion.

Here, the particular data fusion will be figured out through the machine learning process. This will likely be beyond human comprehension and only applicable to situations where decision makers can tolerate the lack of transparency of systems in favor of best-in-class performance (that is, precision). In Figure 4, we illustrate this data fusion process for the example of demand prediction.

This research note is restricted to the personal use of SVarma2@scu.edu.

Figure 4. Deep Learning's Use of Data Fusion for Demand Prediction



Source: Gartner (September 2017)

Typically, demand prediction is treated as a time-series prediction problem, where a time window of previous transactions is used to predict another upcoming time window of transactions (demand). In other words, only transaction data would be used — all the other available data is mostly neglected, because:

- It is difficult to obtain the data
- It traditionally requires integrating, preparing and blending the data before analysis
- It is also difficult for classical machine learning to appropriately make sense of the different data segments.

For many areas of demand prediction (such as consumer demand for food, beverages or other leisure products), the demand can depend on many sorts of events — such as bad weather, major construction, social events, and microeconomic/macroeconomic factors.

Enter deep learning; which, through sheer brute force of gradient-based numerical optimization, attempts to find the most helpful reshaping/combination of the various data elements (see the different colored blocks in Figure 4) in its deeper and deeper layers.

Of course, no deep learning specialists can guarantee if such an architecture will indeed produce superior results; there is no sound mathematical theory, available. However, by using heuristics and

This research note is restricted to the personal use of SVarma2@scu.edu.

upcoming massive compute infrastructures, one can help — via automated machine learning — to produce such optimized network architectures.[8]

Recommendation for data and analytics leaders:

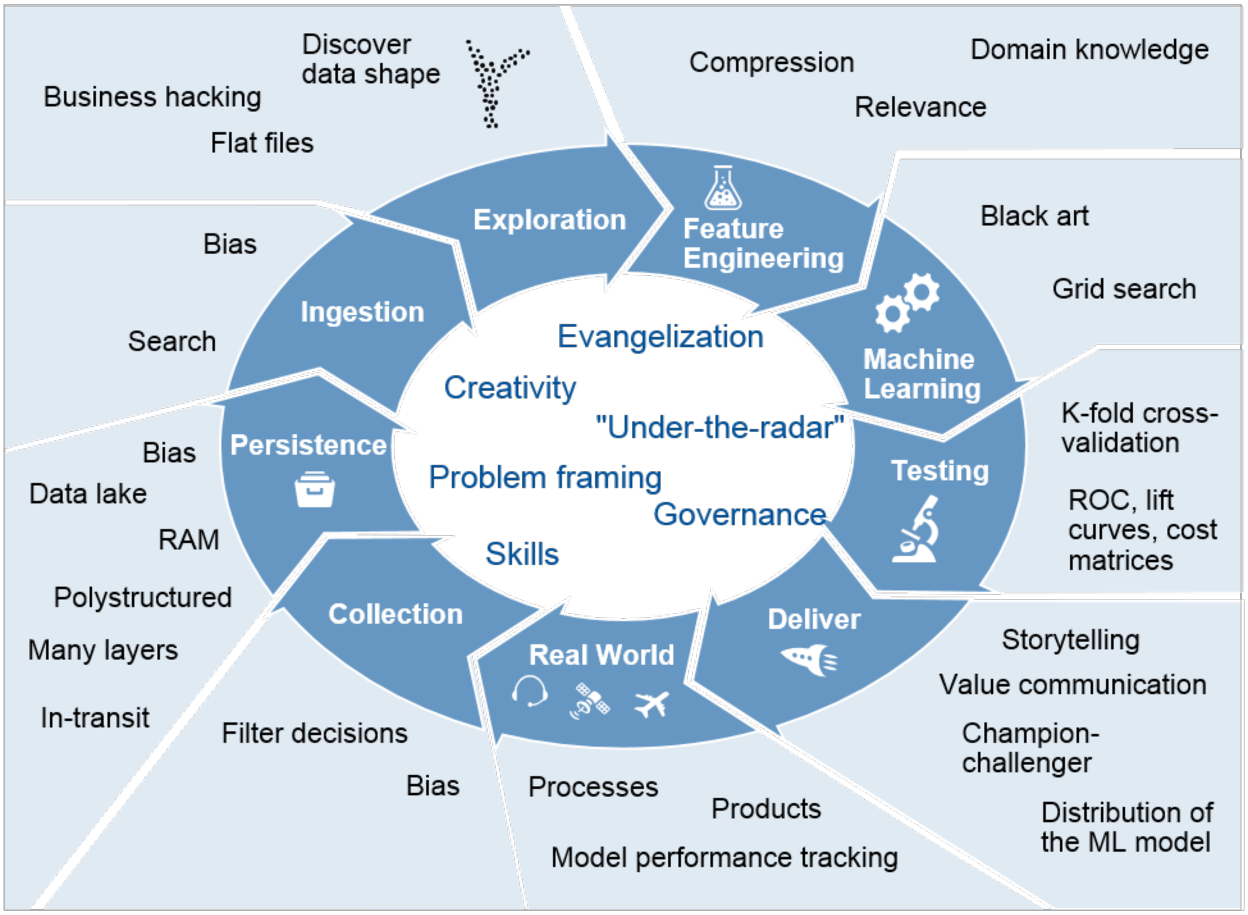
- Certain application scenarios can benefit greatly from deep learning's superior data fusion and prediction capabilities. These are most likely to be found in advanced manufacturing (in, failure and quality prediction for example), demand and fraud detection, and time-series prediction.

Initially, this corporate impact will materialize only for the most advanced organizations with fitting tasks; taking two to three years longer for the rest, once their nascent deep learning infrastructure has the chance to mature further

Despite all the upcoming prospects, general adoption of deep learning by organizations will remain behind the hype for quite a while, for several reasons:

- The infrastructure investments are considerable. In the cloud, organizations can play with and pilot deep learning quite easily; for serious development of deep learning capabilities, however, the bills for the deep learning cloud providers pile up pretty rapidly. Furthermore, most Gartner clients are still not willing to put most of their relevant data into the public cloud, even though this is what it would take to benefit most from deep learning.
- On-premises deployment is complicated, due to the immaturity of the current deep learning infrastructure. Any serious deep learning deployment would require additional hardware to be purchased, installed and managed (see "Find the Right Accelerator for Your Deep Learning Needs").
- Deep learning is far more experimental than most other technologies. It is more hit-and-miss than any other technology we know and failures are abundant. No mathematical theory (see "Innovation Insight for Deep Learning") can state upfront whether a satisfactory solution to a given problem with a corresponding dataset exists or not. Top-notch skills in this area are extremely rare, because they are mostly snapped up by high-flying startups or the internet giants.
- Clients simply have bigger fish to fry. Figure 5 illustrates that machine learning involves fairly complex pipelines and requires numerous stakeholders to work together. These pipelines are, as yet, quite immature and still evolving. Improvements at the pipeline and governance levels are perceived by most teams as being more beneficial than engaging in risky deep learning adventures — especially given the extra requirements and the considerable risk — when deep learning may well not be as fruitful for many corporate applications today.

Figure 5. Most Machine-Learning Pipelines Require Much Improvement



ML = machine learning; ROC = RAID on chip.

Source: Gartner (September 2017)

Recommendations for data and analytics leaders:

- Find time for some initial piloting and testing of deep learning in the cloud, especially if your organization has a large data science team. This technology is poised to stay, so experience in it will help in many ways — such as better planning and better testing of available integrated deep learning capabilities.
- Invest only selectively in your own deep learning capabilities, given the significant requirements in terms of skills and infrastructure and the high level of risk.
- Evaluate partnerships (with academia, research think tanks and top consulting firms, for example), because these can pay off tremendously. They can provide not only a platform for mitigating the significant experimental risks, but also a potential way of recruiting experts directly away from ongoing projects.

Acronym Key and Glossary Terms

NLP	natural-language processing
ROI	return on investment

Gartner Recommended Reading

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

"Innovation Insight for Deep Learning"

"Magic Quadrant for Data Science Platforms"

"Machine Learning: FAQ From Clients"

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

"Find the Right Accelerator for Your Deep Learning Needs"

Evidence

[1] Object identification: "China's Rise in the Global AI Race Emerges as It Takes Over the Final ImageNet Competition." Forbes.

[2] Speech Recognition:

- "The Microsoft 2017 Conversational Speech Recognition System." Microsoft (PDF).
- "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin." Cornell University Library.

[3] Machine Translation: "How Google Translations Are Getting More Natural. Neural Machine Translation Is the Game Changer in Google Translate's Pursuit of Accuracy and Fluency." Livemint.

[4] Lipreading: "Google's DeepMind AI Can Lip-Read TV Shows Better Than a Pro." New Scientist Also "AI Has Beaten Humans at Lip-Reading." MIT Technology Review.

[5] Lyrebird.

[6] Word2vec. Wikipedia.

[7] Super-resolution imaging. Wikipedia.

[8] Automate ML: "The Current State of Automated Machine Learning." KDnuggets News.

Note 1 There Is Reportedly a Zoo of Deep Neural Network Architectures Out There

However, it becomes apparent that these will not only perform classification and regression, but also all kinds of content processing; refer to "The Neural Network Zoo" from the Asimov Institute.

This research note is restricted to the personal use of SVarma2@scu.edu.

GARTNER HEADQUARTERS

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

Regional Headquarters
AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

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

This research note is restricted to the personal use of SVarma2@scu.edu.