The Last Mile in Analytics and Decision Making

Jeff McCrindle
Vice President, Yseop
Adjunct Professor, Graduate Business Intelligence, Saint Joseph's University
New York, NY
jmccrindle@yseop.com

ABSTRACT

Modeling and Simulation techniques have evolved tremendously over the years. But a significant gap remains between the creation of analytical insights and delivering effective, actionable information to non-analytics enabled stakeholders.

The last mile in analytics and decision making is language. Whether you have prescriptive or predictive analytics abilities, embraced Big Data, and have state-of-the-art analytics tools that process data in real time, you still need to explain your results, including the reasoning process, in a language the reader understands.

"No one ever made a decision because of a number. They need a story."

In January 2017, Forbes called Natural Language Generation (NLG) one of the hottest trends of the year. Gartner has described NLG as the last-step in the Data Discovery and BI & Analytics processes. Essentially, Natural Language Generation is software that knows how to write automatically. It connects to data and Big Data tools and explains the results of the analysis in plain English or other local language.

You no longer have to settle with static structured statistical analysis results reports. With NLG, executive summaries and detailed analysis reports that discuss performance can be automatically written with drill down explanations of trends and outliers. NLG solutions can be integrated with interactive dashboards where data can be selected and the descriptive insights focused on this selected data can be automatically updated.

The Last Mile in Analytics and Decision Making

Sometimes it seems that we forget the reason behind data collection and analysis. We get so focused on big data, data lakes, analytics tools, and data governance, we forget the point of data analysis to begin with. Fundamentally, we collect and analyze data to empower data-driven decision making to help our businesses succeed.

How could we lose sight of such a fundamental point? The answer lies in the complexity of data collection, analysis, and governance. Entire divisions of companies have spent much of the last decade trying to build their data infrastructure, so you can't blame them for losing sight of the ultimate goal.

The Last Mile: "No One ever made a decision because of a number. They need a story"

The quote from above comes from Daniel Kahemann who is a psychologist, Nobel Laureate, and one of the leading thinkers in the area of decision analysis, specifically, the study of how human beings make decisions.[1] If you agree with Kahemann, His basic premise in "Thinking, Fast and Slow" is that humans make decisions based on two "Systems":

1) An automatic, instinctual thought process that happens quickly and
2) A slower, more deliberate thought process

His basic premise is that our instinctual side often overpowers our slower, more deliberate thought process (for many reasons). This simple, yet novel idea seems quite familiar to many because and who hasn't heard a CEO say "I'm going with my gut" in response to numbers?

Fundamentally, data is about numbers, and human beings don't speak data, we speak English (or whatever language). The answer is so simple that the founding thinkers of Business Intelligence could see it, but for us today who are so focused on technology, we tend to miss it.

If I ask you to tell me how your company is performing, how do you respond? Do you email me a graph? Do you send me an excel file? No, you use data and graphics to tell me a story as well as, you provide a written explanation of your performance to me in plain English.

The last mile in business intelligence is language. Whether you have prescriptive or predictive analytics abilities, whether you've embraced a data lake and big data, and no matter what state-of- the-art analytics tools you use or how fast you process data in real time, you still need to explain that insight your results, including and the reasoning process, in English language so the reader understands. This explanation can take the form of a written report, which costs companies millions to do manually and which are slow to generate, negating the value of real-time data. Conversely, this can also take the form of a conversation, how many times has the C- suite called analysts to ask, "what does this mean?". This approach has its problems too. Your company can't hire enough analysts to explain all the data either verbally or in written form. So, what do you do?

The answer is Natural Language Generation (NLG).
In January 2017, Forbes called NLG one of the hottest trends of the year. Gartner has described NLG as the last- step in the Data Discovery and BI Analytics processes.[2]

Essentially, Natural Language Generation is software that knows how to write automatically. It connects to data and Big Data tools, explaining the results of the analysis in plain English. Moreover, NLG tools like Yseop Compose are fully self-service, so the analysts who manually write the reports, configure Yseop Compose so it scales their work, generating their reports just like them but at the speed of thousands of pages per second.

With Natural Language Generation, the last mile in the data- to- data- driven decision is fully automatable. But why haven't we heard more about NLG? Why isn't it integrated in all BI tools from the beginning? Interestingly, if you look back at the history of Business Intelligence, Natural Language Generation appears very early in the thinking process.

A Brief History of Data Analytics and Business Intelligence

The term "Business Intelligence" dates back to the 1860s when it appeared in R.M. Deveans' Cyclopædia of Commercial and Business Anecdotes. Deveans describes how a somewhat unscrupulous business man, named Sir Henry Furnace, kept ahead of his competition with a vast network of people who acted as data collectors. These collectors reported directly to him as quickly as possible, thus enabling him to make decisions based on the best possible data. No doubt these sort of Business Intelligence networks have existed as long as competition between enterprises.

In the 1950s Business Intelligence began to be industrialized. In 1958, an IBM researcher posited:

"An automatic system… to disseminate information to the various sections of any industrial, scientific or government organization. This intelligence system will utilize data-processing machines for auto-abstracting and auto-encoding of documents and for creating interest profiles for each of the "action points" in an organization. Both incoming and internally generated documents are automatically abstracted, characterized by a word pattern, and sent automatically to appropriate action points". [3]

Luhn, the author of the paper, called this system a Business Intelligence System. In his articles he lays out the exact technical requirements for an end- to- end Business Intelligence System that would provide executives with all the data they need to make decisions. But critically, Luhn didn't stop there, he argued the Business Intelligence System would also generate written documents and action points, essentially explaining the results of the data collection and analysis within the system. In short, at the advent of contemporary BI, Natural Language Generation was already seen at the last step in the process, the last mile in the BI Analytics process.

Criteria for Luhn's Original end to end Business Intelligence System

Let's dig a bit deeper. Using modern terminology, we can break down Luhn's system into a number of existing components.

1. Data Collection
Luhn is a little unclear on what he generally calls the "data processing" portion of this system and it makes sense, as his thinking was decades ahead of the technology. Here we would break down the data into two categories: structured and unstructured data. For the structured data, we would first talk about a MDM plan and data governance for data integrity and we would look at state- of- the- art systems like Teradata or Oracle for the actual databases. With regard to unstructured data, we would look at Natural Language Understanding systems like Nuance, which turn the unstructured data into structured data.

2. Data Analysis
Luhn again seems to assume that the analysis portion is obvious, although he does focus a little on the data analysis workflows. However, he doesn't focus on the volume of data nor does he seem to foresee what we would call Big Data. Today, the analytics market is full of powerful data analysis solutions and many of the leading BI vendors contain analytics engines more powerful than anything Luhn could have dreamed up.

3. Explanation of the Insights
Luhn discussed automated summaries or abstracts of data along with action items and the generation of written documents. The focus, for Luhn, is narrative describing the results of the analysis. Today, we know the technology that automating the writing of written documents is called Natural Language Generation (NLG). NLG is the last step in Luhn's Business Intelligence workflow, the last mile required to explain in a universally understandable way, the results of the data analysis.

Luhn's Business Intelligence System actually mimicked the way a human being works: we collect data, analyze it, and then explain the results in written or spoken form. Simple, right? Then how did we lose our way? Why isn't Luhn's BI tool in every business today?

Modern Business Intelligence Approach Misses the Last Step!

As technology advanced the market moved away from Luhn's vision. This change was driven by limitations in technology and by the complexities of large amounts of data. But most crucially the business users were told they needed to adapt to the tools, instead of tools that adapt to them. Concretely, the technical teams managing the BI adoptions looked for technical solutions with technical outputs. In the last decade, the BI market came to be dominated by data visualization tools.

Obviously, the market also solved the data analysis and aggregation problems through solutions (Teradata, Oracle, and others). With the growth of Natural Language Understanding, not only can companies make use of unstructured data, but a whole new submarket of data discovery has grown up. Despite all these advances, BI vendors and users alike have forgotten about the last step and the real point of analytics and BI: explaining insights from data as clearly and quickly as possible to support data-driven decisions.

The advent of data visualization tools was both a quantum leap in the Business Intelligence market. But it also exposed that Natural Language Generation, which could write reports explaining the data and actions items, wasn't yet available. Admittedly, in the early 2000's, NLG technology wasn't saleable, so the growth in data visualization made sense even as late as 2010, but why are companies still relying only on graphics?
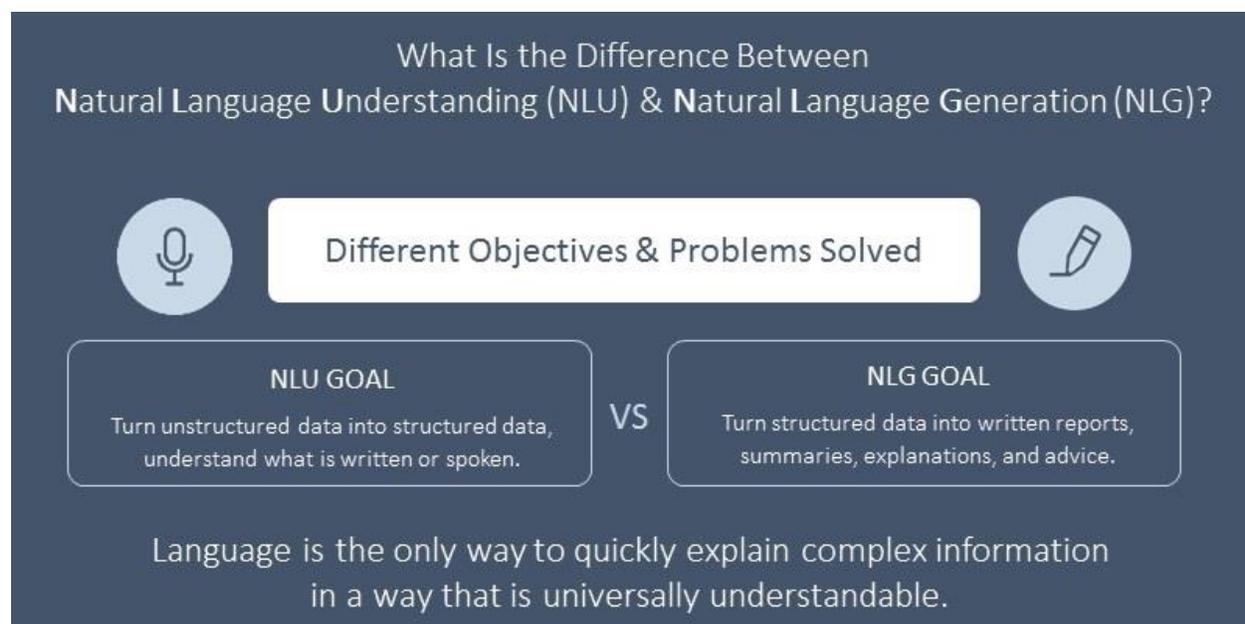
Figure 1. Natural Language Understanding vs. Natural Language Generation

There are two main types of language technologies: Natural Language Understanding (NLU) and Natural Language Generation (NLG). The difference between the terms is subtle: **NLU converts language into data while NLG generates language from data**. While most people have heard of NLU, few have come across NLG. This is simply because most of our mobile phones come loaded with NLU technology – **Apple's Siri** or **Microsoft's Cortana** for example. However, despite the fact that many of us unknowingly interact with NLU technology on a daily basis, most people couldn't tell you what NLG is. Part of NLG's relative anonymity is due to the lack of companies selling NLG tech (only 3 with customers in the US today) and the limitations of conventional NLG systems.
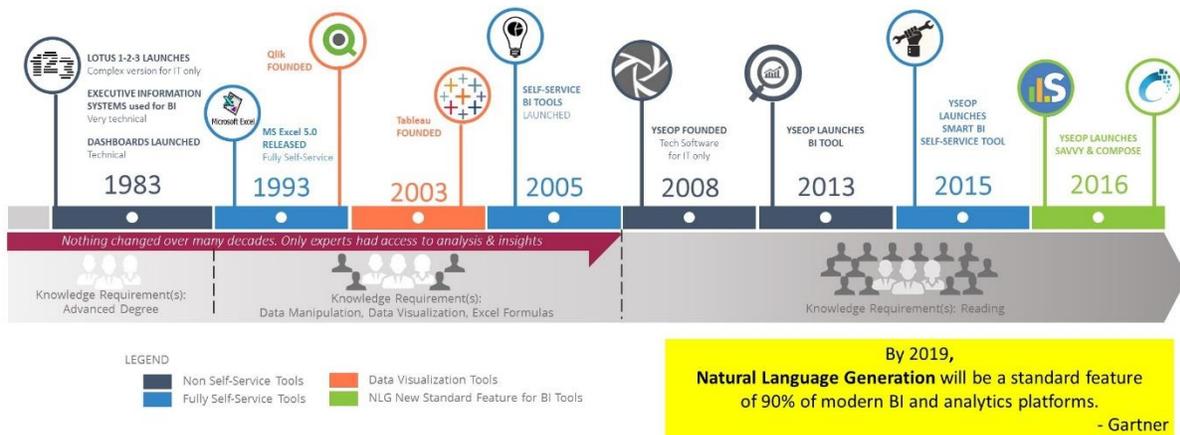
Figure 2. A Brief History of BI Tools

A Picture Isn't Worth a Thousand Words
Despite all of the advances in business analytics over the last 25 years, what has not changed or evolved is the insight process. Yes, it is mildly faster because the data can be brought together and visualized quicker, but the process is the same: an expert looks at the data and writes a narrative explaining what the data means, what actions to take, and why.

Taken as a whole, the history of analytics can be characterized as the success of self-service software. A case in point is when Excel replaced Lotus 1-2-3 as a result of Microsoft's self-service capacity and ease of use. As the BI market has matured, leading companies have become self-service providers. Now, in the burgeoning automation of insight market, companies are already looking for self-service solutions that are able to integrate with their existing BI tools.

Data Overload
Data overload has been given some creative names over the last few years, from data asphyxiation and data smog to the aptly descriptive cognitive overload. No matter what you call it, data overload is a problem that will only grow as companies collect more and more data. The problem goes deeper than computing, storage, or data transfer. It is a fundamental problem that is engrained in the human psyche.

The Limits of Data Visualization in the Age of Big Data
In the beginning, maybe 5 to 10 years ago, data visualization tools were able to keep up with the amount of data they faced. For example, providing insight for a single graph generated daily with 1,000 data points is a reasonable task for a data analyst. However, analysing 20 graphs that refresh every 20 minutes with 20,000 data points can be very trying for even a team of expert data analysts.

The history of BI and some of the challenges we all face with Big Data provide a context to look in detail at why Business Intelligence fails. Many Business Intelligence managers feel their ROI (return on investment) is lower than it ought to be. At the same time, many everyday users of BI tools feel lost. As one user put it "the only person in my company who really understands our dashboard is the person who built it." A disconnect is forming between a company's investment in big data analytics and the everyday usage of BI tools.

"The only person in my company who really understands our dashboard is the person who built it"

Disengagement of the Everyday User
Too many BI efforts are managed with an "if you build [a dashboard] they will come" attitude. The reality is that every day BI users aren't data analysts and they aren't BI experts. The point of the BI tools is to help them leverage the data to enable them to make better business choices. When the BI tools fail to help the everyday user do his/her job, then the user will disengage with the system. Or, to put it another way, they will stop using the BI software altogether. This happens for a number of reasons, from lack of internal communication to lack of intuitive software. However, the biggest issue is that dashboards today assume the user is a data expert, and many successful business people are not data experts.

Shortage of Experts

Globally, we are in the midst of a data analyst shortage. In fact, McKinsey recently found there is a major shortage of data analysis professionals in the United States. By 2018, they predict a shortage of at least 140,000 skilled data analysts and an additional shortage of 1.5 million data experts with the capability to utilize the analysis to make better choices for the business[4]. This means that there are fewer and fewer people available to look at complex data and draw insight from it. However, the shortage of experts is actually a much bigger problem than simply a shortage of talent.

Imagine that you have 50 users of a BI tool: you cannot afford to assign one data analyst to each user to help draw insights from the data. The users must be able to draw accurate conclusions from the data on their own without the help of data analysts.

"Siloed" Expertise

Setting aside the data experts, businesses are also faced with the growth of a new phenomenon: the business silo. Across your business, you hire and cultivate the best employees who work with you continually on the company's best practices, because it is those best practices that separate you from competition. However, as a result of prevailing management techniques, expertise has a tendency to become "siloed".

The silo effect leads to overall poor business performance, but that is a topic for another discussion. From a Business Intelligence perspective, what it means is that the users of BI software in department A have a very different skill set from the users of software in department B. Therefore, BI tools must not only be self-service and adaptable to users with varying skill sets, but also incorporate the insights from expertise across the business.

The key is to use tools that capture and deploy your expertise across departments.

Bad Data

In the field of data analysis and software, professionals like to say: "garbage in, garbage out." This is a well-known problem and most businesses have already tackled it, implementing industry-standard best practices for data management, storage and collection. Generally, bad data problems today arise from antiquated legacy systems. This is why when looking for BI solutions, it is critical find a solution that integrates with all your data, even if the data is contained in multiple places and multiple types of systems.

Manual Data-to-Insight Process

Many businesses today, despite investing in real-time data streams and state-of-the-art data aggregation tools, still manually draw insights from data. This task is frequently left to the end users or data analysts. This problem overlays many of the others we've discussed so far, but what is the point of a real-time data stream if you can't draw insight in real time? And given the concerns about data overload and shortage of experts, can your business afford to follow this manual process?

Manually drawing insight from data is <u>too expensive, too slow, and lacks consistency</u>. The truth of it is that the combination of too much data and too little time is a recipe for mistakes.

Figure 3. Automating creation of analytical insights is the last mile

The conventional response to this bottleneck is to invest in dashboards, but this connects back to the problem we discussed about data overload. Fundamentally, dashboards and charts can show trends, but only words can explain.

The conclusion is that businesses need to automate the last step in this process: the drawing of insights. Moreover, this automation must be able to articulate in plain language the results of analysis. Could Artificial Intelligence and Natural Language Generation (NLG) offer an answer?

Seismic Changes in Artificial Intelligence & What They Mean for BI
Artificial Intelligence (AI) is often understood incorrectly through the lens of science fiction. The truth is that AI today is more about real science than science fiction. But what is AI? At its core, AI is a computer science dedicated to the automation of repetitive tasks, like a robot in a car factory or a computer understanding the spoken word. Therefore, for years AI was divided into robotics and language processing technologies.

Conventional Natural Language Generation
There is no shortage of conventional NLG systems on the market today. In fact, **NLG has existed since the 1980s.** These systems work from a series of templates. For example, *"The weather in <INSERT LOCATION> will be <INSERT TEMPERATURE>."* Some companies have enhanced these systems to create slightly less generic sentences. For the purposes of clarity, we can call these slightly more advanced systems conventional NLG systems.

Challenges of conventional NLG systems:
1. Generates unclear, robotic text
2. Lacks the ability to summarize large datasets
3. Complex (not self-service)
4. Unable to install locally

These problems can result in text that is difficult to understand which, of course is not the purpose of an NLG solution. An effective solution could analyse a dataset with 10,000 records, identify the most important points, and summarize the analysis into an easy to read, simple narrative.

From a business perspective, these conventional NLG systems offer yet another problem: they are not self-service. This means businesses must send their data across the internet to a service provider who, with a team of PhDs, can build an application.

Ignoring for the moment the security concerns of transferring large volumes of data across the internet, there is still the issue of maintaining the application. The ability to maintain the product is in the hands of the service provider, making even the simplest changes to the application time consuming.
> *This makes conventional NLG systems not scalable, secure, or sustainable in the long-term.*

Next-Generation NLG: Combining Automation and Language Technology
Presented with the problems discussed above, conventional NLG systems need to improve, but how? The next generation of NLG unifies two divergent aspects of Artificial Intelligence technology**: automation of repetitive tasks** and **language generation**.

AI's automation of repetitive tasks is generally managed by what is known as an **inference engine**. A common example of an inference engine is the technology behind autopilots in commercial aviation. These powerful AI systems take best practices in the form of rules and reason at exceptional speeds. The highly sophisticated technology behind the inference engine provides value to the business with **speed of reasoning**, **speed of development**, and **ease of maintenance**. By marrying the inference engine technology with NLG, this gives us the Next-Generation NLG.

A Game Changer for NLG
The marriage between an inference engine and NLG technology brings about a new age in Next-Generation NLG. For the first time in history, machines can analyze data, extract insights, and explain what these insights mean— in plain language—so that anyone can understand. However, this new software can go even deeper. It can explain the drivers behind a dataset as well as next-step actions. From thousands of pieces of data, an example of Next-Generation NLG text in real time could look something like this:

"By the end of January 2018, the clear positive growth of total deposits (+4.9%) mainly stemmed from the rise of deposits made towards certificates of deposit, checking accounts, and savings accounts".

Next-Generation NLG means a fundamental shift in Business Intelligence. As we have discussed, the business problem in this industry is clear: too much data and the inability to express insight in real time. Companies either rely on dashboards or manually written reports. However, Next-Generation NLG software offers the answer with easy to use software that explains data analysis results in language that anyone can understand.

Automating Dashboard Analysis & Reporting
Whether you work in finance, retail, and marketing or anywhere across the service sector, chances are you spend a large amount of your day looking at dashboards. There is no shortage of data visualization companies on the market from **Qlik** and **Tableau** to **Power BI**, the list goes on and on. In fact, many companies use **Microsoft Excel** to both store data and represent it graphically. While each of these solutions have different strengths and weaknesses, there is one area where they all fall short: the ability to explain in English what exactly is happening with the data.

Business Intelligence managers today spend a large amount of their time designing dashboard to visualize data. While these dashboards may make immediate sense to the BI Managers, everyday business users may have a hard time drawing immediate insight from the dashboards.

Next-Generation NLG systems offer several advantages when compared with conventional systems.
1. Situational Analysis: A powerful AI engine allows the system to **compare** items or figures against each other. This means you can generate text that compares sales versus objectives or examine your performance against a competitor. In short, the majority of business reporting is situational awareness, and Savvy can do this automatically.
2. Summarizing: It is very easy to use a conventional NLG solution to write something about every data point, but if you have 4,000 data points you don't want 4,000 sentences! With narratives, **less is very often more**. With a click of a button, Next Generation NLG can summarize large volumes of data into an easy to read narrative.
3. Server Installation and Cloud Functionality: Originally, conventional NLG systems were unable to be installed on a client's server or PC. This can pose a problem as many industries have security concerns about sending sensitive data across the internet. Next Generation NLG can be installed locally on a server or even a laptop. This provides a secure solution for companies in highly regulated industries with sensitive data.

Automating Written Advice & Prescriptive Analytics
So far we've clearly identified the existing business problems in Business Intelligence and data analysis, and discussed the value of Next-Generation NLG in BI. We've shown how insight can be drawn from both datasets and dashboards, but we have not yet addressed the final step in the process: how to draw actionable advice from data in real time.

The final goal in any data analysis process is to drive decisions based on data. Technically, we are talking about what many call "prescriptive analytics". Prescriptive analytics is when you can not only diagnose what has happened, but you go further explaining what should be done next.

Today the next steps or actionable advice must be supplied by the end user. This works well in many industries, but what if you are in a highly regulated industry where bad advice can cost you millions? What if you want to disseminate the best practices from your HQ in the field? Or what if you want to speed up the time it takes to make decisions from real-time data streams to gain a competitive advantage?

Advice and recommendations, using your own best practices, is what differentiates you from your competitors. However, it is the speed at which you go from data to advice that enables you to beat your competitors. For the first time, Next Generation NLG allows you to automate the generation of written reports using your best practices to offer recommendations and advice.

Next Generation NLG Use Cases
Today, Next Generation NLG software such as Yseop Compose is used across industries, from leading companies in financial services media, marketing, and much more. Because the system can be fully tailored to your needs, this allows for a very broad usage of the technology. Here are some examples of how Next Generation NLG is helping users today.

Financial Services
In financial services, professionals are confronted with a threefold problem: too many regulations, too much data, and too little time. How can they quickly utilize their data assets all while following relevant regulations and unique best

practices? Next Generation NLG automates the generation of financial, risk analysis, compliance, and other reports taking into consideration relevant regulations and using unique best practices. Across the financial services industry, from New York to London and Tokyo to Sydney, Next Generation NLG is automating the generation of written reports for compliance reporting, back office and personalized reports for clients.

Big Data
Companies are investing millions in faster ways to share data. However, there's no change in the time it takes users to draw insight from complex data and data visualizations. The money spent on quicker data streams isn't realized if you still rely on someone to manually draw conclusions from data. Next Generation NLG tool can provide insight in real time using multiple languages. You can tailor the solution, leveraging expertise of your best data analysts and corporate best practices. Next Generation NLG scales your best practices and turns everyone into an expert.

Marketing
Marketing today is about making data-based decisions, but marketing executives aren't data analysts! Next Generation NLG allows you to automate your marketing reports. So whether you're looking at the competition, campaign management or sales analysis, NLG gives you the tools to analyse data quickly and accurately. NLG makes marketing insights accessible to the sales team on the floor. Any change in marketing message can be pushed to your teams in a matter of seconds, without any additional training.

What Does the Future Hold for Business Intelligence?
Predicting the future is a risky business, but there are professional analysts who make a living doing this sort of thing. Gartner wrote that by 2018 **"smart data discovery will (…) include self-service Natural Language Generation (NLG) capability."**[4] Essentially, Gartner argued that Business Intelligence companies need language generation capability to explain the results of data analysis. In the same research note, Gartner presented a bigger problem: whether you are analysing data, presenting the results of data discovery or even sharing search results, data visualizations are too imprecise. If you want a precise answer to a precise question, a graph just won't cut it.

CONCLUSION
This paper highlighted some of the major Business Intelligence and analytics challenges that we all face, but to summarize, here are the top reasons that companies are adopting Next-Generation NLG systems today:

1. It is too slow to manually draw insights from data visualizations
We are not all visual people. Sometimes graphs are more complicated than the data used to build them. Dashboards are helpful, but they usually require the skills of an expert data scientist to interpret them.

2. There aren't enough data scientists
An analyst can help with one dashboard from one dataset, but what if you have thousands of data points coming from multiple data sources? Even if you wanted to hire a data analyst to make sense of your dashboards, chances are it would be far too expensive, and you might not even be able to find the appropriate talent to help you.

3. Cut costs & expand capacity
Writing reports costs money, but according to existing clients, using NLG software has cut costs by 77% and increased levels of performance. Working with Fortune 500 companies on both sides of the Atlantic, NLG has transformed real-time data streams into written reports, enabling businesses to make informed decisions with data only seconds old.

4. Businesses want data analysed in real time
Over the last decade, businesses have invested millions of dollars in real-time data streams and data collection, including data storage, e-discovery, and data visualisation tools. While these investments have laid the groundwork for a means to draw insights from data, in and of themselves they really only help acquire and organize data faster. Ultimately, conventional models do not allow businesses to understand their data in real time.

ABOUT THE AUTHOR

Jeff McCrindle serves as Vice President of Sales at Yseop (pronounced "Easy Op"). Yseop is an industry leading provider of Natural Language Generation (NLG) solutions. Prior to Yseop he has served as Vice President of predictive analytics firms (ProModel, FlexSim) and healthcare simulation firms (Education Management Solutions, CHI Systems). His career experience spans defense, intelligence, and commercial markets such as healthcare, finance, communications, and logistics.

In 2010 Jeff McCrindle served as deputy chair of the healthcare track of the MODSIM conference and presented a paper discussing the "Best Practices in Medical Simulation".

Since 2008 he has taught graduate business analytics courses at Saint Joseph's University. He holds a Six Sigma Black Belt from Villanova University, MBA and BS Mathematics degrees from Saint Joseph's University, and has completed executive programs at Harvard Law School and the Wharton School of Business.

REFERENCES

1 Kahemann, Daniel "Thinking, Fast and Slow" Penguin Psychology, New York, 2012.
2 Sallam, Rita, Gartner, "Smart Data Discovery Will Empower Enable a New Generation of Citizen Data Scientists" June 2015.
3 Luhn, H.P. A Business Intelligence System, IBM Journal, October 1958.
4 Manyika, J., Chui, M., et. al. "Big data: The next frontier for innovation, competition, and productivity". McKinsey Global Institute, May 2011