ALTIMETRIK®

GenAI for HR Analytics

Leveraging the strengths and overcoming limitations of GenAI

WHITE PAPER

Executive Summary

Generative AI or Gen AI has certainly unlocked several possibilities in the world of analytics that can help business functions and their owners derive a lot of data-driven intelligence. This can make their decision-making process faster and more accurate. This becomes particularly useful for functions in a company where people are more business focused and less tech-savvy and grapple with complicated analytical tools.

In this whitepaper, we focus on the HR department where employee-centricity is the imperative to ensure effective operations and customer delight. Many GenAI use cases are just beginning to be explored by HR departments. However, Generative AI is not free of limitations and challenges. We are not yet at a place where we can trust GenAI a hundred percent to produce accurate, consistent and reliable outputs. This risk can be mitigated by using automation components that leverage the strengths of GenAI engines, while overcoming their limitations.

Let's dive deeper into the realm of GenAI for HR Analytics – it's limitations, challenges, and the ways to tackle them.

Introduction

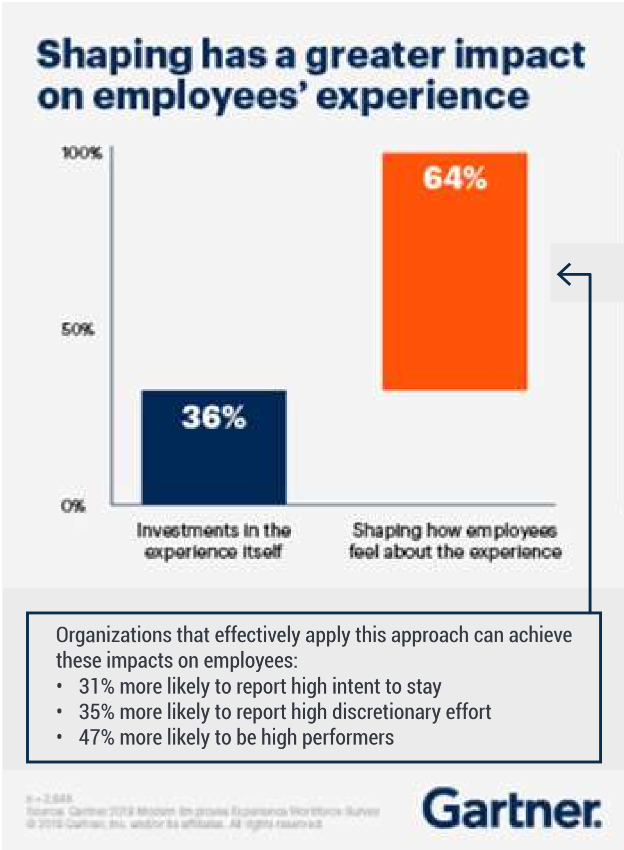
Peter Drucker and several other management gurus have long propounded the strong correlation between company profits and customer satisfaction. While that is definitely true, there is a need to peel one more layer and ask the question - what leads to consistent customer satisfaction?

Studies have shown that **Employee Satisfaction** is the single most important factor that drives **Customer Satisfaction.** The pivotal question is how does an employee feel about his/her role in the company?

Here's what Gartner research has to say about this topic:

> *"...two-thirds of the drivers of customer satisfaction are due to "feel factors," or how customers feel during and about their experience. HR leaders should take this same approach to employee experience, focusing on influencing and improving employees' feelings about their overall experience through the use of psychological, motivational and social principles. Improving the way the experience feels can lead to a boost in employee engagement and support a positive company culture. "*

The chart below provides some related statistics..



Employees Are Always Giving Feedback – But do you know exactly what?

A lot of companies today are clear about the priority of employee satisfaction in the present times. They are keen to know what their employees feel about the processes, benefits, the general work environment and how likely are they to stay with the job, how will they perform and so on. Which is why HR departments have set up various employee feedback tools/processes - third party tools, home grown applications, good old Excel files et al - for not just soliciting, but also eliciting feedback from employees. Plus there are external forums like Glassdoor, Ambitionbox etc. where employees voice their feedback about a company.

There in lies the challenge – so many channels of information that typically do not talk to each other. How does one get a unified view of:

- the topics of interest for the employees
- their sentiments around those topics - positive, negative, neutral
- the next level emotions tied to the topics - suggestion, compliment, complaint, gratitude, frustration etc.
- ...and so on.

The usual approach taken by companies is to pull data from all these different channels/sources into a central place, and use Analytics tools to get insights. That works to some extent, but not valuable enough. Why? Because a lot of the information in the reviews is in the form of unstructured (text) data. Some engine needs to extract the information dispersed in this broad swathe of text. This is where Generative AI (GenAI) engines like GPT and others come into play.

Here's an example:
Let's say the HR department wants to understand the **factors that influence** attrition (or, positively speaking, retention) in the organization, **how much** does each of these factors influence retention, and how to **predict retention/attrition.**

There are several predictors of retention/attrition – learning opportunities, organizational culture, manager, how challenging/interesting the work is, current salary, difference between the current salary and the market median and so on.

Many of the above factors (e.g. salary) are available as **structured** data. Traditional Machine Learning /AI engines are reasonably good at crunching such structured data to identify patterns, correlations, and use that information to predict the possibility of attrition.

At the same time, there are some critical factors like learning opportunities or work life balance which are subjective opinions/perspectives from the employee. These are present as **unstructured** (meaning regular text) data that are not easily consumed by traditional ML/AI engines. Large Language Models or LLMs can convert these unstructured data into structured data. For example, the LLM can be used to create a sentiment score per review from an employee, and aggregate that to get an overall sentiment score from the employee. This can then be added to the rest of the structured data (like salary etc), to be fed into the ML/AI engine. So, the **latent information potential** of such subjective data is effectively captured and fed into the prediction process.

To summarize, the Generative AI LLMs can be used not just on their own, but also in combination with other engines that can mutually strengthen each other's capabilities, thus providing a better toolkit for informed decision making.

Limitations and Practical Challenges with GenAI

While GenAI is an extremely useful tool to address some of the most complex challenges with data, there still are certain limitations that could be misleading and throw wrong interpretations of data. Let's dive into the challenges and see some examples.

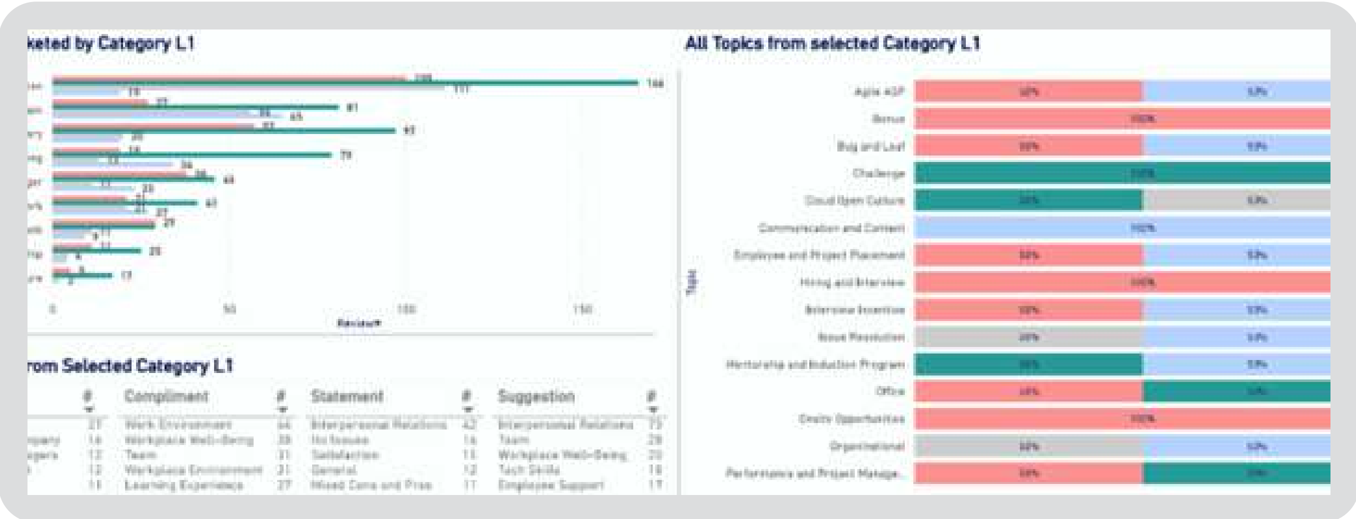
Challenge 1 Hallucination

Let's say some employee reviews collected from an organization are sent to an LLM. The goal is to extract what topics the employees care about, and what are the sentiments tied to those topics.

The results come back from the LLM, and they look impressive at first sight. Upon closer inspection, it is noticed that certain topics identified by the LLM do not make any sense.

To make it real, here's the screenshot of a sample dashboard that was built using the results from the above exercise.

Among the topics pulled out from the reviews by the LLM engine, there's a topic called - 'bug and leaf'. That does not make any sense for the current context.



The next step is to read through the reviews from which the LLM picked up the topic 'bug and leaf', to double-check if there are any typos in the reviews. None. Screenshot of the reviews below.

Full review
Question: What's one thing from another company that you'd like to see implemented here? Answer: Nothing special i found to be implemented here as of now as i am a new joiner. But one suggestion from my side is to add sick leaves and earned leaves as separate.
Question: How can the organization support you in managing your personal and work life? Answer: due to work pressure unable to balance personal life and work life balance. No proper planning for deliverable - giving short notice for code delivery- due to that getting more defects in prod and UAT. No proper design in application. lack of functional knowledge.
Question: Is there anything else you'd like to share? Answer: hike also very disappointed

This is a clear example of what is called 'hallucination'. LLMs have a tendency to cook up topics that do not make any sense in the context.

There are other such topics in the above results, like 'Cloud Open Culture'.

Challenge 2 Duplication of Topics

Another issue that is seen in such LLM results, is that of duplication of topics. Here's an example set of results.

In the above screenshot, the LLM is considering topics like Technologies, Tech stack, Tech Integration, Tech Discussions, Tech Choice as **separate/distinct topics.** But practically they are all quite similar from the perspective of summarizing employee feedback. If an HR team was doing this exercise manually, they would simply group all of them into one topic to prevent the list of topics from getting diluted.

While this is obviously a problem, it can manifest itself in various damaging ways. For example, on the left side of the above screen, there's a prioritized list of topics that the employees are talking about, positively as well as negatively. Any company typically wants to take actions to improve the negative areas. There may be a critical topic, which does not get enough 'votes' just because the LLM distributed the votes for that topic across several similar topics.

This can lead to a wrong understanding of the feedback, thereby leading to wrong decisions/priorities.

Why does this happen? An online LLM like GPT sets a limit to the number of words per question. So, hundreds or thousands of employee reviews cannot be bundled and sent to the LLM engine in one shot. The next option is to send the reviews with multiple LLM calls. When that is done, the topic identified by the LLM in each call can vary, even if the reviews are talking about the same thing. And this is what is seen in the above screenshot.

Challenge 3 LLM ignoring key but less frequent topics

A third practical issue seen with LLMs, is the problem of glazing over topics which may be less frequent in the Foundation Models used by the LLMs, but relevant for the employees of a specific organization. LLMs typically skip such topics in the results sent back.

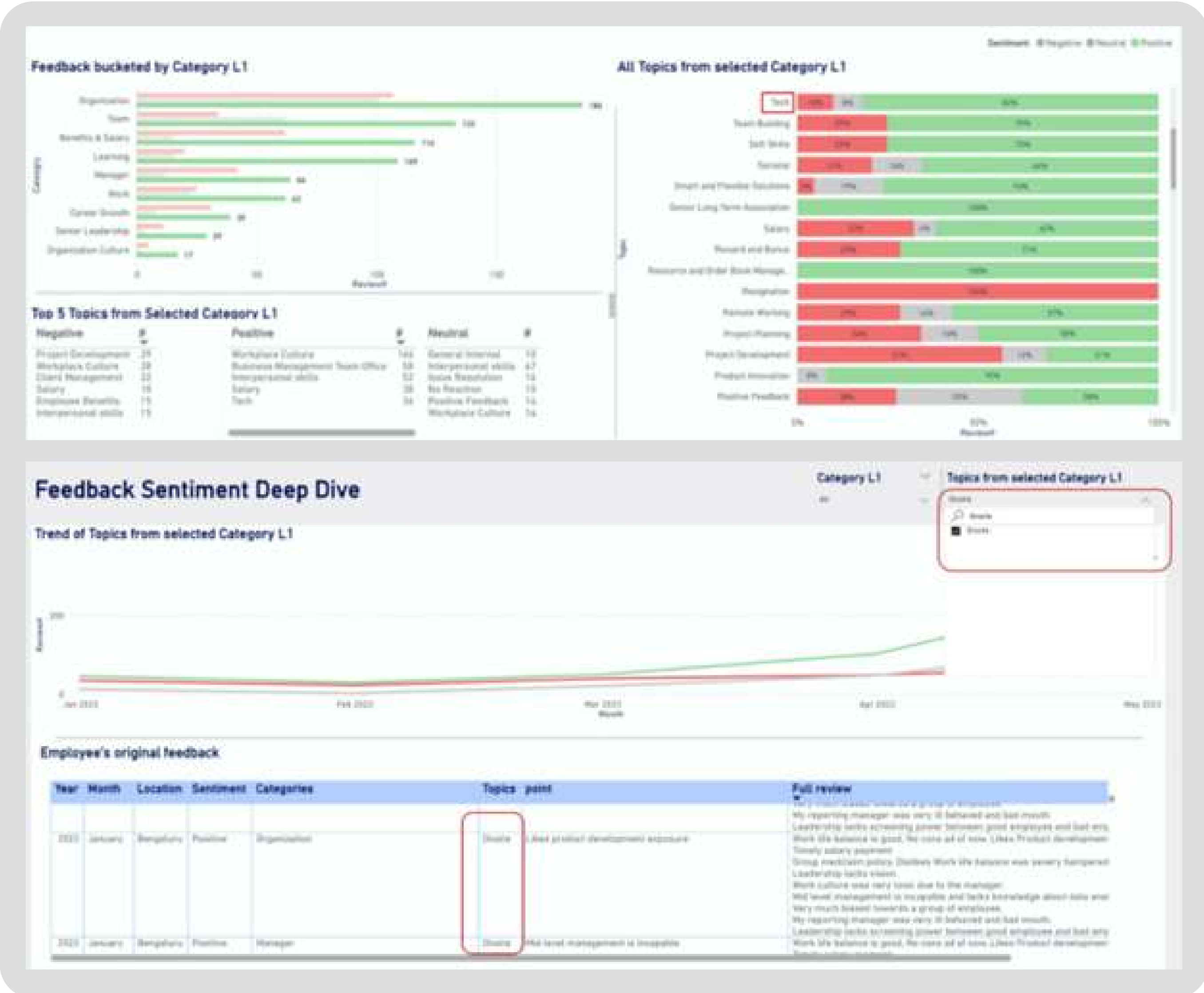
For e.g., in many companies, employees would love to work at their client's place - typically called 'onsite'. The below screenshot shows the results when HR department is searching for this topic in the results from the LLM.

The search does not yield any result, since this was not a key topic in the Foundation LLM model, and got left out of its results.

The Solution - Pre and Post-processing of the LLM Results

It is clear that a decent amount of work needs to be done before sending inputs to the LLMs, and also after receiving the results from the LLMs, to avoid some of the afore-mentioned problems. But it is obviously not feasible to do that manually on a regular basis. The solution is to have an automated pipeline/engine that takes care of the above problems, and others similar to these. This last phrase is critical, else the same problem will rear its head very frequently in different shapes and forms.

The below screenshots showcase how a well-designed engine can solve the above problems. The first screenshot below is for Challenge #2, and the second one is for Challenge #3 above. The results are highlighted in red.

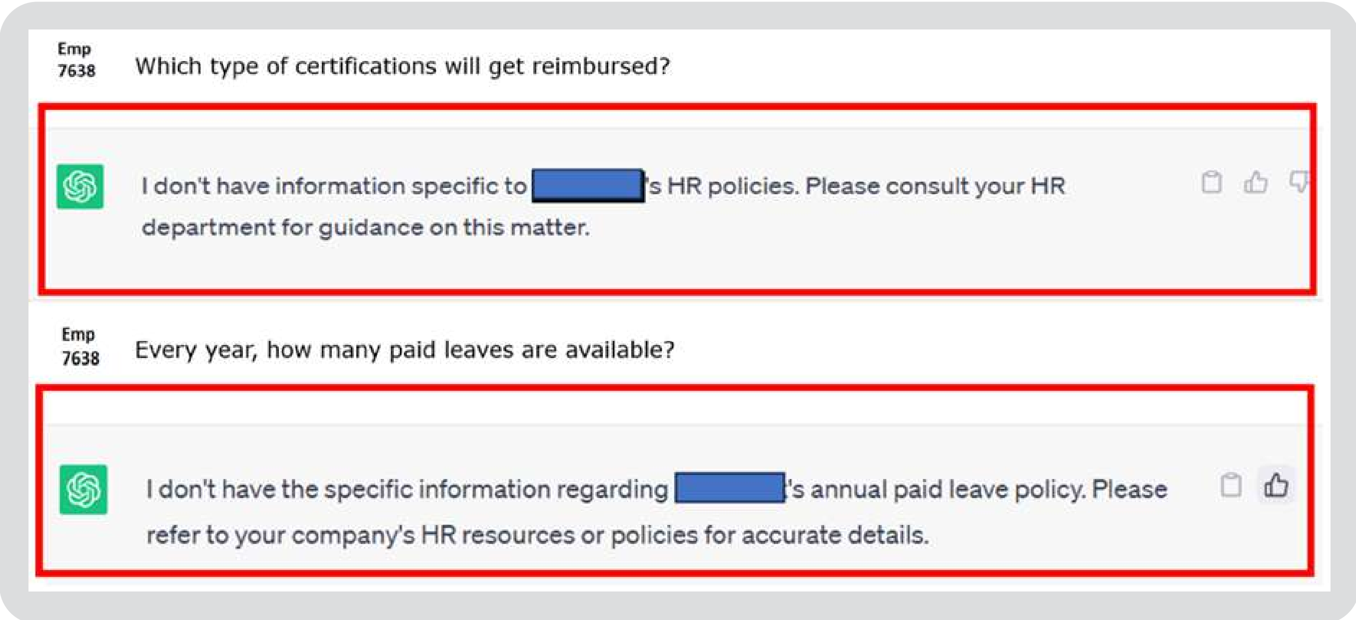


Challenge 4 Unaware of local context

Let's say there is a requirement for building a chatbot for Employee Queries. An employee can ask a question about HR policies, and get answers from the chatbot. That's the expectation.

Here's the catch though. ChatGPT sitting out there has no clue about any particular organization's HR policies. To be fair, it has never been trained on that. So, it may get lucky with the answers occasionally, but there is no way it can provide reliable answers in a reasonably consistent manner.

Here's a screenshot depicting one such interaction between an Employee and ChatGPT. The company name has been greyed out.



What is the solution to this particular challenge?

There are 3 options to be considered:

Option 1 Simple prompt engineering

Simply bundle the context in the prompt/question provided to ChatGPT.

Unfortunately, this option may not work for large HR policy documents, since there is a limit to the number of words that can be sent to ChatGPT for each question.

Option 2 LLM Fine-tuning

Large Language Models (LLMs) like GPT, which underpin applications like ChatGPT, are trained using a very large corpus of documents. Sure, those models can be trained further on the company's HR policy documents. In LLM parlance, this is called 'fine tuning'. That will make these models more aware of the company-specific HR policies, in turn making its answers more relevant. But there is a time and cost associated with this exercise. It also needs to be repeated every now and then, to include the latest policy, changes in policies etc. Moreover, it has been practically seen in the industry that such fine-tuned models, while better than the 'no fine tuning' option, are still found lacking in terms of relevance of answers. One reason is that the amount of text contained in a particular company's HR policy documents is still miniscule as compared to the humongous amount of data the models have been trained on, thereby limiting the 'influence' of these specific policy documents on the overall model results.

Option 3 Take the middle path

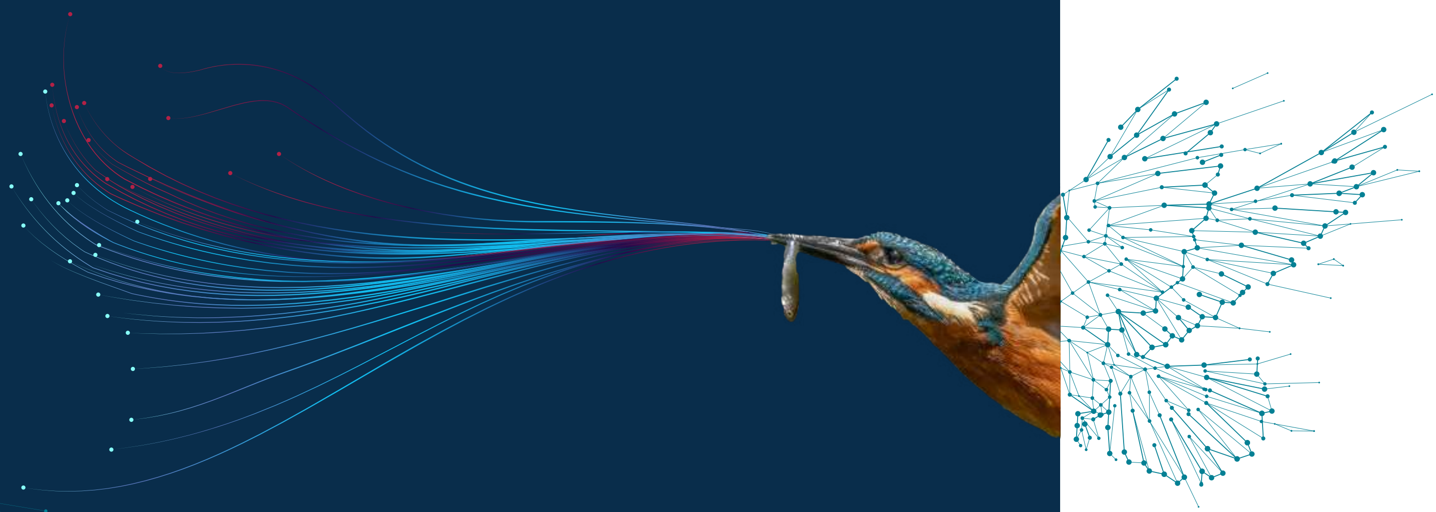
An underlying engine running within the company can filter the context related to the question, and send that context along with the question, to ChatGPT.

The results will look much better now. Building the above automated engine is a tricky yet effective part of the overall solution.

That way, **the local Context Intelligence within the company, is getting combined with the global Language Intelligence of the LLM.**

Empulse – a GenAI Framework for Deeper & Accurate HR Intelligence

Yes there are limitations, but there are ways to tackle them. And these are not expensive, time-consuming or prone to any risk. The Empulse framework by Altimetrik is a ready to use, deployable real code that can resolve all the challenges listed above and more. It has been developed by engineers and practitioners with deep expertise in the AI domain who have been closely watching the HR industry, its challenges, goals, and evolution for years. Empulse is posited to help HR departments leverage GenAI effectively and deliver deep insights into the employee pulse in the organization, and make decisions to positively impact retention/attrition.



Summary

Companies where employee welfare is not a mere lip-service but taken seriously can derive huge benefits from Generative AI and LLM tools. But there is a learning curve for everyone on how to make these intelligent technologies do the jobs we want them to. That's where the irreplaceable human intelligence kicks in to pack in the right techniques, strategies, and thoughts to ensure effective application of the technology. We will see more evolution and improvements in the capabilities of these technologies in the future; for now, applying the ways to overcome the limitations is the smartest way to go.

About Altimetrik

Altimetrik is a pure-play digital business and digital transformation company unlocking growth and opportunity with speed, scale, and consistency. We focus on delivering business outcomes with an agile, product-oriented approach. Our digital business methodology provides a blueprint to develop, scale, and launch new products to market faster. Our team of 5,500+ practitioners with software, data, and cloud engineering skills helps create a culture of innovation and agility that optimizes team performance, modernizes technology, and builds new business models. As a strategic partner and catalyst, Altimetrik quickly delivers results without disruption to the business.