??????????

??????????

???Ahmer Inam



In the past few years, Nike has made three acquisitions in the data and analytics space, including a widely reported deal to acquire Datalogue, a startup that uses a proprietary machine-learning technology that automates data preparation and integration. This acquisition accelerates Nike's ability to develop digital consumer engagement platforms such as Nothing But Gold social commerce app. Nike is an example of how businesses are growing by using data more intelligently to understand customer buying habits and adapting their business operations and processes to meet consumer demand. In a rapidly evolving marketplace it is absolutely essential for organizations to grow their data readiness capability or be left behind by their competition.

What Is Data Readiness?

Data readiness consists of all the tasks a company needs to manage in order to ensure that its artificial intelligence (AI) platforms learn from reliable and relevant data sources. As the promise of AI/machine learning (ML) becomes realized in enterprises, the tech sector has responded with an increased focus on developing toolkits that enable building faster, smarter, and more automated models. Those efforts have typically focused on the downstream elements of the ML lifecycle. As a result, a gap exists in systematically understanding, evaluating, and verifying data before it is used in building ML models. Since data scientists typically need to spend 60%-80% of their time preparing data, it's important that enterprises understand and embrace data readiness to close the gap A model is only as good and relevant as the data that it is trained on. Data readiness systematically addresses this challenge and ensures that context-driven high quality is used in building ML models while fast-tracking the ML lifecycle.

At Centific, our approach to data readiness has following pillars that are powered by innovations in database engineering. This framework covers all stages of data readiness, ranging from data

sourcing to loading and publishing in order to expedite ML model development.



Why Is Data Readiness So Important?

There are a number of reasons why establishing data readiness has become essential to realizing the promise of AI/ML.

First off, businesses ranging from retailers to automakers are increasingly relying on AI to launch new services and products and deliver more personalized service faster than their competitors can. AI promises to help businesses deliver speed to insight, market, and value, which is essential at a time when customers want more personalized experiences. But AI is only as good – and inclusive – as the data it uses. And the task of preparing data is getting more complicated as unstructured data proliferates. Unstructured data consists of information ranging from social media posts to chat conversations that give businesses valuable information in real-time. Unstructured data yields the type of raw insight into a company's operations that formal research simply cannot answer. It's also prolific: according to IBM, as much as 80 percent of business data is unstructured. But unstructured data is messy. It doesn't present itself in any easily categorized format. So businesses need to invest more into better ways to collect and prepare that data, then train AI to train itself through machine learning. It's no wonder that Tesla's director of AI, Andrej Karpathy, recently said, "A lot of what I do is curating data sets. That's where all the engineering is. It's not people writing algorithms. It's people collecting data sets."

Another reason data readiness is becoming more important: businesses are under pressure to manage their first-party data more intelligently to deliver personalized experiences. That's because technology giants such as Google are taking steps to ensure the demise of third-party cookies, which are crucial for businesses to create personalized experiences across the web. For example, in 2020,

Google announced that it would stop supporting third-party cookies on its Chrome browser. Meanwhile, Apple has launched app tracking transparency (ATT), which stipulates that users of iPhones will now need to agree to allow a business to collect information about them – delivering another blow to third-party tracking. The writing is on the wall. Businesses need to look to their own websites to create personalized experiences. But creating personalized experiences is easier said than done. So they're turning to AI – which makes data readiness more important.

What Are Some Other Trends We're Seeing with Data Readiness?

One of the biggest trends we see is an increased emphasis on being inclusive. More businesses are realizing how data readiness is fraught with bias, which makes AI less inclusive. This problem must be fixed. We believe the answer is to take a human-centered approach. Being human-centered means, among other things, relying on a diverse, global team to manage crucial data readiness tasks such as data annotation. For example, at Centific, we have a pool of hundreds of thousands of people worldwide from 150+ countries. Our people speak 200+ languages and apply deep domain expertise as they curate data using on our data annotation platform, OneForma. Our diversity contributes to ensuring that we help our clients build inclusive and localized AI applications, such as voice applications accounting for languages, dialects, and accents.

What's an Example of Data Readiness?

We helped a global brand improve its computer vision model to 97 percent accuracy. Our client looked to enhance its computer vision model so that it could better recognize live images of objects and the text on the objects. The client's goal was to improve the user experience of its cloud-based image and video collection solution to help people easily navigate through thousands of stored pictures from the convenience of their mobile devices. The company approached Centific asking to collect and curate a high volume of high-quality images. Centific tapped into its global pool of hundreds of thousands of resources to collect live images and text in specified categories. Since quality was of high importance, Centific:

- Developed a customized collection tool to upload, store and classify these live images and a second customized tool to label the text and objects in the images.

- Trained a team of quality assurance and labeling experts to provide the highest quality deliverables.

Centific collected and curated images in 19 different categories, each with its own target volume and target specifications. We delivered over thousands of high-quality live images at a 97 percent accuracy rate in 12 weeks covering five continents.

Although this example does not cover the scope of data readiness, it's illustrative of how complicated just one aspect of data readiness – data curation – can be.

Contact Centific

To learn more about how we can help you meet your business goals with data readiness, contact Centific.

-
-
-
-
-