datacentrix

Shedding light on dark data; a critical step before AI can deliver value

Shakeel Jhazbhay, General Manager: Digital Business Solutions at Datacentrix

With artificial intelligence moving rapidly moved from experimentation to strategic priority, today's businesses are collecting more information than ever before. However, many are overlooking a fundamental question during this process: is the information feeding these AI initiatives actually ready?

For years, businesses have focused on collecting and storing data. Information now exists across emails, contracts, collaboration platforms, shared drives, cloud repositories, enterprise applications and archives. However, volume alone does not create value. The reality is that much of this data remains unmanaged, inaccessible or underutilised. Often referred to as 'dark data', this growing pool of hidden enterprise information is becoming one of the biggest obstacles to meaningful AI adoption.

Research conducted by IBM suggests that more than 60% of organisations estimate at least half of their data is dark, while approximately one-third believe dark data accounts for 75% or more of their total information estate. These figures point to a significant disconnect between data accumulation and business value. This matters because AI does not create intelligence from nothing, but rather amplifies what already exists. If information is fragmented, duplicated, outdated or poorly governed, then AI outcomes reflect those same weaknesses.

Organisations are information rich but intelligence poor The root causes of dark data could include information silos across departments, ageing legacy platforms, a lack of governance, incomplete integration, compliance irony (when information is stored well beyond the mandatory period), or changing organisational priorities, where collection is prioritised over analysis.

“

Research conducted by IBM suggests that more than 60% of organisations estimate at least half of their data is dark, while approximately one-third believe dark data accounts for 75% or more of their total information estate.

Notably, says McKinsey, 70% of software used by Fortune 500 organisations is more than 20 years old, a statistic that illustrates the scale of modernisation challenges many enterprises still face. This also draws attention to the fact that, realistically, organisations cannot expect AI to compensate for decades of fragmented information practices. These factors all contribute to create a situation where organisations appear data rich on the surface but, in actual fact, they lack the ability to transform the information into actionable intelligence. The challenge becomes even more significant as AI initiatives expand.

> The businesses that ultimately derive the greatest value from AI are unlikely to be those deploying technology the fastest. Instead, success will belong to those creating visibility, governance and trust across their information environments.

What are the business consequences of dark data?

Poor information management is often treated as an IT issue, but its impact extends far beyond technology. For example, from an operational perspective, organisations experience reduced productivity as employees spend more time searching for information or recreating existing work. In terms of collaboration, poor version control and disconnected repositories make it harder for teams to work efficiently and confidently.

Security and compliance exposure also increase. Sensitive information may remain stored longer than necessary, retention requirements become difficult to enforce, and organisations risk falling short of regulatory expectations.

Customer outcomes are affected too. Delayed responses, repetitive requests and fragmented information ultimately reduce service quality and satisfaction. At the same time, businesses lose agility because they cannot confidently adopt automation or safely scale AI initiatives.

Information readiness foundational to AI success

One of the greatest misconceptions surrounding AI is that implementation begins with selecting a platform or deploying a use case. In reality, AI readiness actually begins with information readiness. This means organisations need to establish a structured approach to understanding and improving their information environments. But where do they start?

Ideally, the process should kick off with an audit of existing information assets to identify where content resides, who owns it and how it is being used. From there, businesses need to classify their data, differentiating between unstructured, semi-structured and structured assets. The third step is to address, or cleanse, what is commonly known as ROT data – information that is redundant, obsolete or trivial – to immediately reduce storage costs. Next, companies should establish stronger governance practices across repositories and workflows implement strict policies for data retention and destruction. Only once this foundation is in place should organisations accelerate AI deployment, introducing AI/ML tools to parse unstructured data and redact sensitive info.

Those that win with AI won't necessarily move first

The businesses that ultimately derive the greatest value from AI are unlikely to be those deploying technology the fastest. Instead, success will belong to those creating visibility, governance and trust across their information environments.

For organisations looking to unlock the next phase of digital transformation, illuminating dark data is no longer optional. It is becoming the prerequisite for turning information into usable intelligence.

Shakeel Jhazbhay, General Manager: Digital Business Solutions at Datacentrix

For more information contact Elzette du Preez at edupreez@datacentrix.co.za

www.datacentrix.co.za