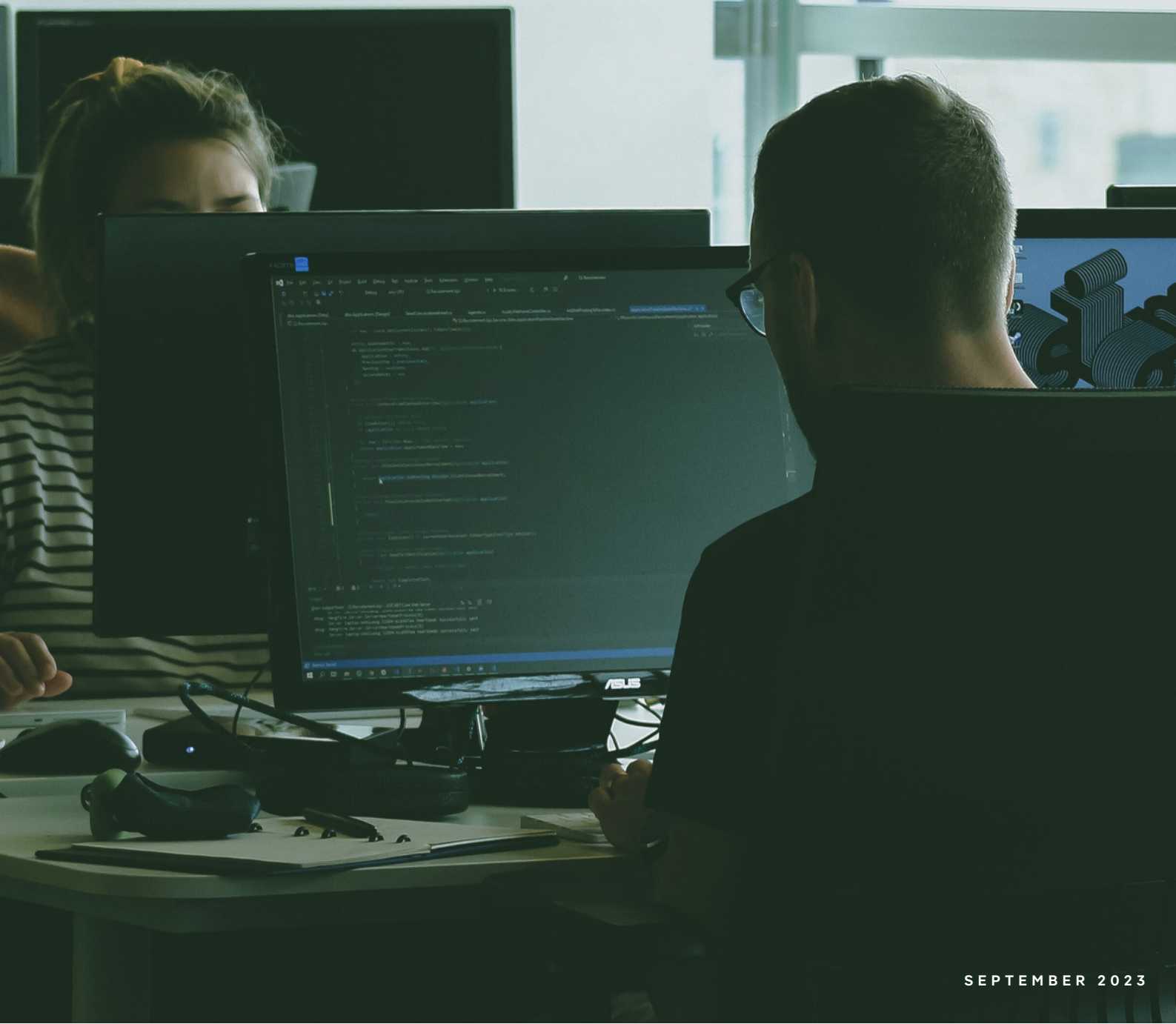


# Self-Fulfilling Innovation: The Impact of AI on Data Functions



**Artificial Intelligence (AI) is set to transform the world around us, changing the way industries operate and interact with their customers. But what impact will AI have on the data teams and functions that power AI capability? Kubrick consultants explore the automations, efficiencies, and improvements that AI will create for data and technology professionals to revolutionise their work, from meeting the demands of modern data governance to anticipating risk with chaos engineering.**

## INTRODUCTION

At this year's Databricks Data+AI Summit, Jacob Renn, co-founder at AI Squared, remarked at the 2023 that he was "surprised that [the conference] hasn't been renamed to AI+Data Summit"[1], conveying the increasingly accepted notion that the world of data is becoming further entangled with Artificial Intelligence (AI). The impact of AI technology is now inescapable - when was the last time you visited a website that didn't have a chatbot? AI has not only revolutionised customer-facing interactions, but is transforming the tech industry itself, changing how technology is developed, deployed, and maintained.

One area where this innovation is felt most strongly is within the data functions that power AI. As the demand for AI-powered solutions grow, so does the need for better data governance and management to ensure AI is safe, ethical, and accurate. Here we explore how AI can enable data functions to drive faster, more efficient processes to govern, manage, and test the outputs of the data that will ultimately improve AI outputs, demonstrating the power of self-fulfilling innovation.

## 1 GOVERNING DATA WITH AI

Data is now a business-critical asset, which yields great reward but also risk; as a heavily regulated commodity under GDPR, the governance of data is essential for compliant data use. Traditional data governance (DG) involves the manual creation and dissemination of data policies for managing data across an organisation. The conventional approach has its pitfalls, being generally regarded as reactive, complex, and slow to keep up with the ever-changing data landscape. However, as businesses scale their data use, DG has been rightfully put under the spotlight and is now a priority for organisations across the world who want to wrangle the vast amounts of data at their disposal without risk. Implementing AI assisted tech functions in DG can increase operational

efficiency, enabling rapid detection of anomalies, ensuring compliance with regulatory requirements, and identifying potential risks, all in real-time. With AI supported tech functions, a business can transition towards a proactive approach, both addressing issues before they may arise and establishing protocols to best unlock the full value of their data assets.

Three key areas of governance where AI is adding value are data quality assessment, data classification, and data lineage analysis. AI-based data quality functions leverage machine learning algorithms to detect anomalies and errors in real-time, offering more options in how to respond and deal with issues. Data classification involves organising data into different categories based on sensitivity, confidentiality, and regulatory requirements. AI-powered data classification can automatically classify sensitive (including personally identifiable information) and non-sensitive data, ensuring compliance with regulations such as GDPR. Data functions that are assisted by AI in this realm are critical for keeping up with the continual change within regulatory policies while simultaneously reducing reliance on repetitive tasks by their human counterparts. Data lineage analysis is the process of understanding and visualising data as it flows across an organisation's system. AI can provide comprehensive tracking across entire systems, ensuring traceability, transparency, and accountability. Some of the technology vendors offering capability within the AI-powered DG space are Collibra DQ, which focuses on augmented data quality (DQ), and Microsoft Purview, which acts as a unified data governance platform to improve data loss prevention (DLP), discovery, and classification.

In their eBook 'From DG to AI Governance', Dataiku have emphasised the symbiotic relationship between AI and good DG.[2] They can reinforce each other's effectiveness and ethical implementation. DG can provide guardrails to protect against risk, and AI can leverage the well-managed data environment created by DG. Dataiku emphasised two new concepts that would present AI-assisted data functions to better fulfil DG, machine learning model management and responsible AI governance, both being crucial for proper AI implementation. Governance is often incorrectly considered to be a blocker to innovation, however, is not when implemented correctly. A balance is to be struck between governance and enablement, with space being provided for exploration and experimentation: governance should support and aid responsible innovation. Tech functions in this space seek to offer the correct balance between security of data and access to data to allow safe and stable innovation.

## 2

## MANAGING DATA WITH AI

Alongside data governance, AI in data management (DM) is not a new concept, as numerous organisations have successfully integrated AI into various aspects of this technology function. In fact, IBM having offered Hybrid Data Management since 2019, an approach to DM that leverages a mixture of human and AI processes. This allows businesses to manage, share, and analyse data across different storage types, including on-premises or in a private or public cloud provider, but

across structured and unstructured data types.[3] By leveraging AI capability within data management, organisations have seen improvements in data ingestion and query performance. IBM in collaboration with 451 Research, a leading technology research and advisory firm, reported that these significant advancements enable 'data engineers to accelerate data management and analytics projects and database administrators to focus on higher-impact tasks.'[4]

Within data management, one tech function in particular that is set for substantial growth with AI in 2023 is data observability.[5] Data observability concerns the health and performance of an organisation's data infrastructure. IBM has already demonstrated their support by acquiring databand.ai, an industry leader in data observability software, to further enhance their software portfolio in data and AI.[6] Through automated monitoring and the identification of pipeline failures (using historical trends to process statistics about the pipeline) poor-quality data can be minimised and therefore reduce downtime.[7] AI-assisted tech functions are of growing importance as organisations need to support an increasingly large volume of data; it is estimated that poor-quality data costs companies \$12.9 million per year.[8] Entities rely heavily on high quality and trustworthy data, and the implementation of AI in data observability will undoubtedly provide organisations with a competitive edge in maintaining a healthy and reliable data infrastructure.

### 3 ANTICIPATING RISK WITH AI

Would you let the most powerful AI in the world try to break your production application? The discipline of experimenting on a system to build confidence in its capability to withstand turbulent conditions is known as chaos engineering. An estimated 80% of organisations experienced outages in the last 3 years. In 2022, 60% of outages resulted in at least \$100,000 of losses, up from 39% in 2019. In the same time frame, the proportion of outages that cost upwards of \$1 million increased from 11% to 15%.[9]

Chaos engineering offers tech functions the ability to foresee catastrophic failures before they occur, protecting data in all facets of business. In Chaos Engineering, you define your system's steady state, design experiments to put the system under stress, and analyse the results to improve the system's fault tolerance. Instead of randomly injecting stress, or even consciously selecting where you believe it appropriate to inject it, you can have AI train a neural network with reinforced learning to develop a strategy in pinpointing where vulnerabilities lie and apply stress accordingly. AI can more accurately mimic a real-life outage, with one paper suggesting AI is 45% more performant than a traditional Chaos Monkey.[10] AI is not only assisting tech functions in their processes, but also offering resilience testing.

## 4 CONCLUDING THOUGHTS

AI is not a future risk; it is here today. Due to the technology's infancy (compounded with the rate at which it is advancing), there are still risks that should be addressed and a comprehensive framework will help address ethical concerns and mitigate risks (including bias and privacy breaches) pre-emptively. Consider the use of AI to group documents into sensitive and non-sensitive information in a hospital, if this was allowed to be performed on all data it may initially incorrectly classify some critical information which would breach both internal and government guidelines. In order to combat this, establishing an AI risk framework is a cornerstone to ensuring responsible development. A suitable risk framework should not only negate potential risks associated with AI but identify opportunities to maximise its positive impacts.

AI brings immense potential to tech functions, helping to create new-found value from efficiencies that turns operations within data governance and management from a mere compliance exercise to a source of rich data-driven insights. Firms should develop processes and tools to manage risk within the limits set by its culture and appetite (and appropriate regulations), from the beginning. Addressing ethical concerns, data privacy, security and technical limitations, workforce transformation, legal compliance and transparency are crucial. A thoughtful, proactive approach must be adopted to navigate the evolving AI landscape. However, embracing AI-driven innovations is the key to success in the modern era: engineering, governance, and management must evolve to reach unparalleled heights. More than just offering an edge to businesses, AI integration is incrementally becoming a requirement to operate in the data hemisphere.

## REFERENCES

1. <https://www.databricks.com/dataaisummit/speaker/jacob-renn/>
2. Howells, J. (2021) *From Data Governance to AI Governance: Pillars and Pitfalls*. (Accessed: 21 July 2023).
3. [www.ibm.com](https://www.ibm.com/downloads/cas/YD5R1XLB). (2018). *Hybrid Data Management Guide | Home*. [online] (Accessed 14 Jul. 2023).
4. Aslett, M. and Curtis, J. (2019). *Accelerating AI with Data Management; Accelerating Data Management with AI*. [online] IBM. Available at: <https://www.ibm.com/downloads/cas/YD5R1XLB> (Accessed 14 Jul. 2023).
5. Flynn, S. (2023). *5 Fast-Growing Data Management Trends in 2023*. Available at: <https://opendatascience.com/5-fast-growing-data-management-trends-in-2023/> (Accessed 14 Jul. 2023).
6. *IBM aims to capture growing market opportunity for data observability with Databand.ai acquisition* (2022) IBM Newsroom. (Accessed: 21 July 2023).
7. Monte Carlo Data. (2023). *What Is Data Observability? 5 Key Pillars To Know In 2023*. [online] (Accessed 14 Jul. 2023).
8. Wu, F. (2022) *Uptime Institute's 2022 outage analysis finds downtime costs and consequences worsening as industry efforts to curb outage frequency fall short*. Uptime Institute. (Accessed: 21 July 2023).
9. University, L.B.C.C. et al. (2020) *Human-ai partnerships for chaos engineering: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, ACM Conferences*. (Accessed: 21 July 2023).

**Authors:** Alex Fleetwood, Anya Dard, Blaydan Clish, Charlotte Harriman, Daniel Martin, Sam Thomson



Copyright 2023 Kubrick Group  
All rights reserved.



@kubrickgroup

