California Management Review

MANAGEMENT

The New Data Management Model: Effective Data Management for AI Systems

by Luca Collina, Mostafa Sayyadi, and Michael Provitera

Image Credit | New Data Services

As AI and data are critical to businesses, ensuring the data used in AI systems is accurate is challenging.

INSIGHT | FRONTIER 26 Mar 2024

The research presents the Data Quality Funnel Model to improve business decision-making and flexibility, by making data more accurate, reliable, and valuable data for AI systems. This model talks about the critical role of machine learning and predictive analytics. They can effectively enable business strategy and, thus, growth when companies can control the quality of the data that goes into them.

RELATED CMR ARTICLES

"Getting AI Implementation Right: Insights from a Global Survey" by Rebecka C. Ångström, Michael Björn, Linus Dahlander, Magnus Mähring, & Martin W. Wallin

"Organizational Decision-Making Structures in the Age of Artificial Intelligence" by Yash Raj Shrestha, Shiko M. Ben-Menahem, & Georg von Krogh

All companies have to deal with messy, fragmented data from different silos within their organization.[1, 2] However, prior studies indicate that most companies do not understand the economic impacts of bad data. For example, 60% of companies in the US did not grasp the effects of poor-quality data.[3] Inaccurate or incomplete information costs the US over $3 trillion per year. Poor data quality also costs large organizations an average of $12.9 million annually.[4] Therefore, the business costs of bad data are systemic and substantial.

The Data Quality Funnel Model is a new data management model that can improve the performance of machine learning and artificial intelligence (AI), so companies clean the data they train and operationalize machine learning to help them run actions faster and more informed. With machine learning, explainable AI, cloud computing, and robust data governance, executives take these advanced technologies to bring them to decision-making. Looking at the Data Quality Funnel, executives can see how technology innovation and the company's culture must join together. This funnel considers high-tech solutions to solve a genuine business need to get high-quality data that drives business growth and keeps companies ahead of others in the digital world.

The Potential Issues and Opportunities of Data Quality

Data quality should always be the initial point of consideration before any machine learning model implementation. Companies can implement data governance and management policies to more effectively handle information. Companies can then maintain data integrity while increasing output quality with such policies.[5]

Effetive Data Management for AI Systems

Data Pre-processing or Cleansing: Data cleansing is the critical first step in creating machine learning models. Data cleansing entails eliminating errors or inconsistencies from data to make it reliable for analysis; normalizing brings it all into a standard format to make comparison easier; integration brings in data from various sources in ways that make sense for analysis; finally, data fusion represents merging multiple sources into one coherent analysis.[6]

Data-as-a-Service (DaaS): Recent efforts and proposals attempting to ensure data quality from raw sources for Machine Learning and Artificial Intelligence have resulted in the concept of Data-as-a-Service (DaaS), where users receive data without knowing its source, hence requiring continuous Data Quality Management processes using Machine learning models for quality management.[7]

Synthetic Data: Synthetic data or pre-fabricated data is data that has been generated using a purpose-built mathematical model or algorithm to solve a (set of) data science task(s).[8] Synthetic data are meant to imitate real data and reuse it for privacy, ethics, and overall quality data. Several applications can be supported by synthetic data: Machine learning for training and privacy and internal business uses like software testing and training models.[9]

AI Trust and Governance

Explainable AI (XAI): A lack of clarity around AI can reduce trust in automated decisions.[10] Corporate leaders can use Explainable AI (XAI) to explain AI recommendations. Popular XAI methods like LIME quickly explain individual AI predictions via basic models. SHAP more accurately explains predictions using global data patterns. Companies must train all employees to understand AI outputs and explanations to fully benefit from XAI, empowering people to use AI more confidently.

Algorithms Governance: Studies are developing guidance for companies and governments to get AI's benefits while minimizing downsides.[11] Recent studies have been focused on healthcare and industry. However, simple processes for responsible AI governance are needed more broadly. This research area is still exploratory. Leaders need plain guidelines to govern AI development. A recent white paper released by HM Guidelines for AI indicates how generative AI requires governance to guarantee high-quality information, accountability, oversight, and privacy, which is a further step ahead.

We propose a specific structure that highlights roles with different levels of responsibility and accountability. A compelling proposal elaborates on the potential strategies to consider to validate the results of elaborations through algorithms, their processes, and XAI. Companies can create oversight to ensure artificial intelligence (AI) is used properly, specifically for algorithms.

Institutional Challenging: Institutions, by creating committees, including AI specialists and non-executive directors, may establish overarching rules to guide decisions with both artificial intelligence technology and human expertise.

Consultancy Challenging: These challenges may be tackled by external professionals who utilize critical assessment to produce more substantial and sustainable outcomes through independent and impartial opinions.

Operational Challenging: These challenges are for the operations staff who watch directly how the AI systems work on tasks. They can run checks and raise issues about problems to rectify algorithms and improve them through an escalation process, but they don't intervene in modifying the algorithms.

There can also be high-level rules, outside audits, and day-to-day monitoring of the AI. Working together, these can help make AI accountable and catch problems early. The goal is to have people with different views in place to develop and use AI responsibly. Our proposed model requires integration between AI experts, managers, and executives. These responsibilities are diverse and different before and after the outcomes of AI's decision-making processes. The visualization of the possible roles following the algorithms' governance and auditing is shown in **Figure 1**.

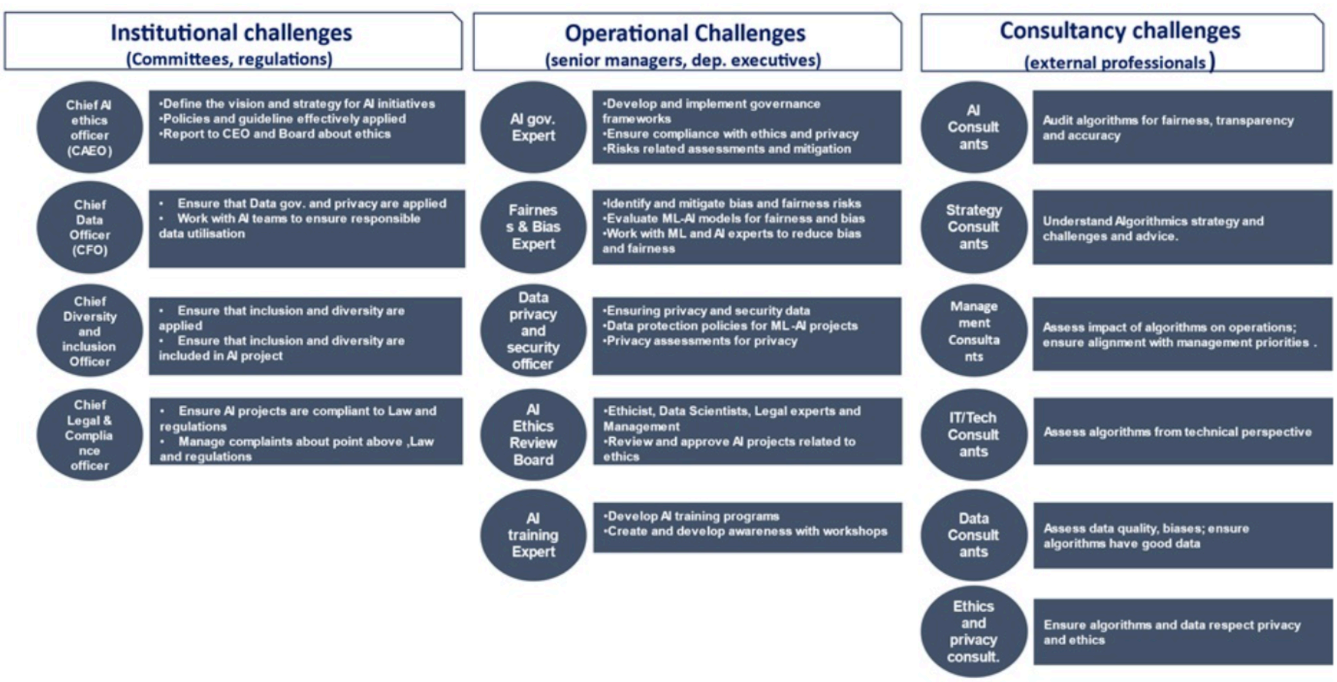

Figure 1: The Roles of AI Experts, Managers, Executives, and Consultants

The Moderating Factors

Data Culture and Leadership: Establishing a data culture within an organizational culture is vital in creating successful business strategies, particularly considering start-ups rely heavily on data from day one.[12, 13]

Trust in AI and Machine Learning Outcomes: Using AI and machine learning in business decisions has benefits and risks. AI can improve decision-making, especially regarding customers and marketing. However, AI could also damage value and privacy and models might expose private data, be unfair (show bias), or lack interpretability and transparency. These issues are severe in healthcare. More work is needed to make AI trustworthy and to balance accuracy, avoiding harm and bias while protecting privacy. Technology cannot just focus on performance; it needs collaboration to ensure systems are safe, fair, accountable, and compliant with regulations.[14]

XAI (Explainable Artificial Intelligence): There is no consensus on what makes an AI explanation valid or valuable. Some research suggests using logical, step-by-step approaches to build trust in explanations and objective ways to measure explanation quality.[15, 16] But critics say more work is needed so AI explanations are accurate, fair, and genuinely understandable to ordinary people. Overall, explainable AI lacks clear standards for defining and assessing explanations.

Cloud: Using machine learning and AI to make cloud computing more flexible for businesses has been researched and studied extensively. machine learning and AI can enhance resource management in cloud computing.

The Data Quality Funnel Model

Leaders must take responsibility for the AI technology their companies use, even if it is unclear who is accountable when machine learning causes harm. Rather than trying to force accountability despite messy data inputs, fixing problems earlier is more efficient. Carefully checking training data, removing errors, and standardizing inconsistencies builds trust in AI systems while avoiding extra work later. Putting good data practices naturally enables accountable AI systems down the road. Clean data flowing into algorithms pays forward accountability. Therefore, different ideas, good data management, and responsible AI reinforce each other.

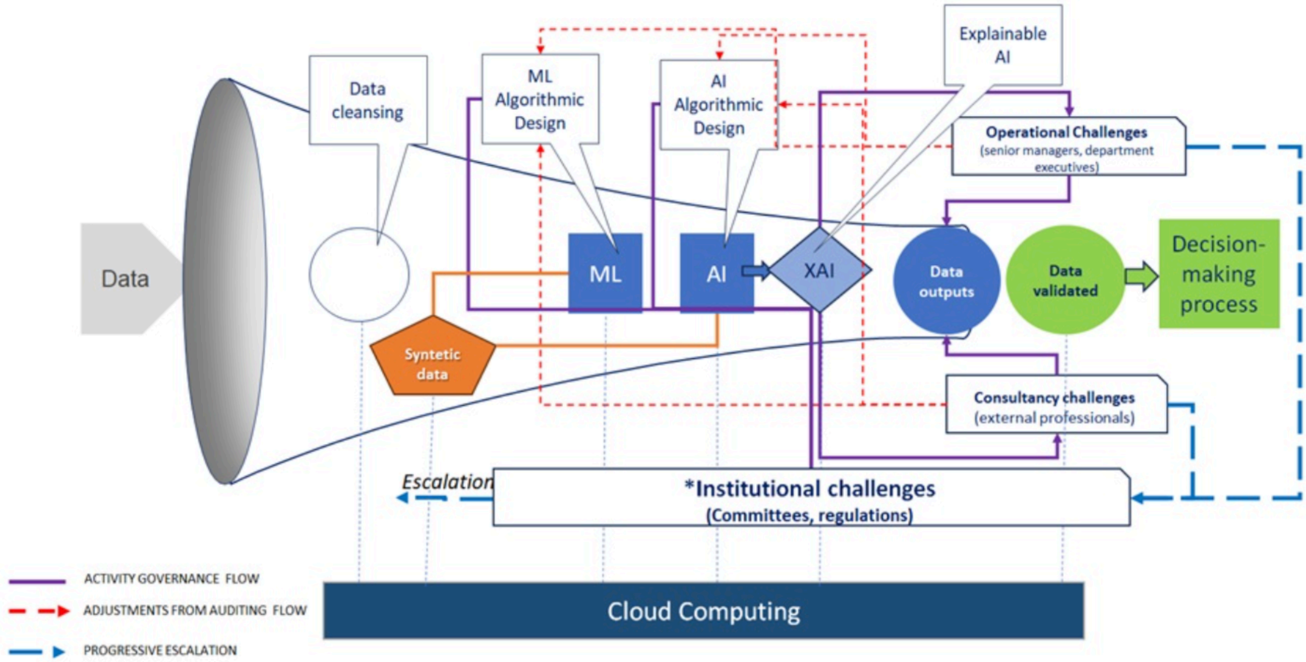

Figure 2: The Data Quality Funnel Model

In the following table, the integration between data quality and accountability is shown:

Accountability Framework		Data Quality Funnel		Integration explanation
Elements	Description	Elements	Description	
Impact Assessment	Pre-implementation: Assess the potential positive and negative impacts on internal and external stakeholders	**Infrastructure and Services**	Use cloud technology and machine learning platforms to manage data and improve model performance..	Integration is essential as assessing the impact of AI systems includes evaluating the infrastructure's ability to support ethical and accurate data management.
Risk Monitoring	Regularly measure fairness, security, safety, and transparency. Audit ML and AI algorithms design and XAI	**Data Management & XAI**	Include cleaning, preparing, combining data sources, and XAI to increase transparency.	Risk monitoring aligns with data management and XAI practices to ensure data integrity, security, and understandable AI decisions, which are critical for identifying and mitigating risks.
Incident Response	Have a plan to investigate, document, and fix problems. Involve legal teams and outside experts.	**Algorithms Governance**	Include the governance and the escalation process of algorithms and data output challenges	Incident response mechanisms match with the escalation process around the governance roles; there are benefits from robust AI trust and governance practices, to address and mitigate incidents transparently.
Accountability Mapping	Clarify who is responsible for monitoring the AI system and its impacts. Update policies as needed..	**Algorithms Governance**	Establishing guidelines and roles for responsible AI development, emphasizing oversight and privacy.	Accountability mapping and algorithm governance complement each other by ensuring clear roles and responsibilities, enhancing oversight, and promoting ethical AI practices.

Table 1: Data Quality and Accountability

In Conclusion

This article shows how vital good data is for companies making choices and plans in our tech world. As AI and data become more critical to businesses, ensuring the data used in AI systems is correct and secure is challenging. This paper gives a way to manage these issues - the Data Quality Funnel Model. This model lays out steps to check data is reliable, easy to access, and safe before using it to guide major choices. Clearly showing how to check data at each point helps avoid mistakes or problems. Using this model lets businesses apply AI well to keep up with the competition. The Data Quality Funnel Model fills a gap by showing companies how to handle data troubles posed by new tech. This model gives clear guidance on preparing quality data for strategy and choices that are current real business needs. By lighting the way for accuracy, our proposal displays a route for success in navigating the intricate, tech-driven business world today.

References

1. Fan, W., & Geerts, F. (2022). Foundations of data quality management. Switzerland: Springer Nature.

2. Ghasemaghaei, M., & Calic, G. (2019). Does big data enhance firm innovation competency? The mediating role of data-driven insights. Journal of Business Research, 104(C), 69-84.

3. Moore, S. (2018). How to Stop Data Quality Undermining Your Business. Retrieved 02 02, 2024, from **https://www.gartner.com/smarterwithgartner/how-to-stop-data-quality-undermining-your-business**

4. Sakpal, M. (2021). How to Improve Your Data Quality. Retrieved 02 02, 2024, from **https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality**

5. Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), pp. 148-152.

6. Allamanis, M., & Brockschmidt, M. (2021, 12 8). Finding and fixing bugs with deep learning. Retrieved from Microsoft Research Blog: https://www.microsoft.com/en-us/research/blog/finding-and-fixing-bugs-with-deep-learning/

7. Azimi, S., & Pahl, C. (2021). Continuous Data Quality Management for Machine Learning based Data-as-a-Service Architectures. International Conference on Cloud Computing and Services Science. 328-335.

8. Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C. & Weller, A. (2022). Synthetic Data – what, why and how?,. arXiv:, arXiv:2205.03257v1, 5.

9. James, S., Harbron, C., Branson, J., & Sundler, M. (2021). Synthetic data use: exploring use cases to optimize data utility. Discover Artificial Intelligence, 1, 15. https://doi.org/10.1007/s44163-021-00016-y

10. Tiwari, R. (2023). Explainable AI (XAI) and its Applications in Building Trust and Understanding in AI Decision Making. International Journal of Management Science and Engineering Management, 7(1), 1-13.

11. Nikitaeva, A., & Salem, A. (2022.). Institutional Framework for The Development of Artificial Intelligence in The Industry. Journal of Institutional Studies, 14(1), 108-126.

12. Antonopoulou, H., Halkiopoulos, C., Barlou, O., & Beligiannis, G. (2020). Leadership Types and Digital Leadership in Higher Education: Behavioural Data Analysis from University of Patras in Greece. International Journal of Learning, International Journal of Learning Teaching and Educational Research, 19(4), 110-129.

13. Denning, S. (2020). Why a culture of experimentation requires management transformation. Strategy & Leadership,(48), 11-16.

14. Strobel, M., & Shokri, R. (2022). Data Privacy and Trustworthy Machine Learning. IEEE Security & Privacy, 20(5), 44-49.

15. Ignatiev, A. (2020). Towards Trustable Explainable AI. Proceedings of the Twenty-Ninth International Joint Conference on Artificial IntelligenceJanuary, 5154–5158.

16. Yang, C., Sinning, R., Lewis, G., Kastner, C., & T., W. (2022). Capabilities for better ML engineering. Retrieved from ARXIV: https://arxiv.org/abs/2211.06409

Luca Collina Follow

is a transformational and AI Business consultant at TRANSFORAGE TCA LTD. York St John University awarded him the Business – Postgraduate Programme Prize and CMCE (Centre for Management Consulting Excellence-UK) for his paper in Technology and Consulting Research Prize. Author/External Collaborator of CMCE.

Mostafa Sayyadi Follow

Mostafa Sayyadi is an Associate Fellow at Australian Institute of Management. He works with senior business leaders to effectively develop innovation in companies, and helps companies—from start-ups to the Fortune 100—succeed by improving the effectiveness of their leaders.

Michael Provitera Follow

Michael Provitera is an Associate Professor at Barry University. He is an author of Level Up Leadership published by Business Expert Press.