International Journal of Economics, Commerce and Management
United Kingdom ISSN 2348 0386 Vol. X, Issue 9, Sep 2022

http://ijecm.co.uk/

ARTIFICIAL INTELLIGENCE AND NEW TECHNOLOGIES IN DATA MANAGEMENT

Elira Hoxha
Department of Statistics and Applied Informatics, Faculty of Economy,
University of Tirana, Rr. Arben Broci, Tirana, Albania
elirahoxha@yahoo.com

Blerina Vika
Department of Statistics and Applied Informatics, Faculty of Economy,
University of Tirana, Rr. Arben Broci, Tirana, Albania
blerina.vika@gmail.com

Abstract

This paper is a general overview of the latest trends in the field of data management, with a focus on artificial intelligence as a key for the improvement of several processes in this field. The literature review highlights many capabilities and benefits when applying artificial intelligence such as prediction, automation, optimization, etc., explained through examples in the different sections of the paper. Along with the benefits, several challenges related to the use of artificial intelligence in the data management lifecycle are shown, such as efficient real-time decision-making, challenges related to safety and security, and the need for implementing accurate data-processing algorithms that go in line with individuals' privacy rights, standards, and local legal requirements. In this study are listed also some other new approaches that empower data management professionals including dataOps, data governance, data fabric, self-service analytics tools, data marketplace, blockchain technology, the hybrid, intercloud, and multi-cloud approaches, etc.

Keywords: artificial intelligence, data management, augmented data management, data mining, data analysis, big data security

INTRODUCTION

In order to be a cut above in the global marketplace and retain a competitive edge, most companies must take an active role in managing and protecting their data throughout its lifecycle.

Data management involves the professional practice of collecting, organizing, protecting, and storing a company's data so it can become a constitution of intelligence for business decision-making. Actually, the majority of data used in companies are in digital form, and as the quantity of them increases continuously, a lot of efforts have been made on choosing the appropriate data management strategies. The amount of data that is produced and stored globally is extremely huge and it's not dwindling. If all these data are consistent, correct, complete, sign, and operational, then they can provide an essential part of a good organization's decision-making. Making these data manageable is one of the most important concerns in today's companies.

To become leaders, most companies are embracing new technologies to transform data management in a way that can help turn pinch points into gateways for deeper business insights and better decision-making processes. Artificial Intelligence techniques and models make it possible; these models work much faster than humans and don't need breaks.

In order to take full advantage of everything that artificial intelligence has to offer, enterprises must embed it at multiple levels, starting from the data level to ensure that it enables the full scope of the data management lifecycle, from ingestion to curation and discovery, as well as driving applications that are built to leverage that data (Aslett & Curtis, 2019).

This paper presents an overview of artificial intelligence features and other trends and tools that any enterprise needs to integrate into their data management process, as well as a study of different phases in the data lifecycle where these new approaches can be used to provide a better decision-making process.

ARTIFICIAL INTELLIGENCE CAPABILITIES, BENEFITS, AND CHALLENGES FOR DATA MANAGEMENT

Artificial intelligence, through its techniques and subfields, can be found almost in every application related to data management, data science, and business intelligence (Dennis, 2022). It enables these types of software used for data and analytics, to improve the time to value, through the following capabilities:

- **Prediction** of the required data in machine learning for a better and more accurate model. Artificial intelligence can also be used for the prediction of resource usage from different applications.

- **Security by identification** of anomalies and possible SQL injection queries originating from unauthorized sources, by monitoring the query logs. Artificial intelligence can also help in blocking these kinds of queries. It can also identify data that violate privacy issues and mask them automatically, in order to protect them from accidental disclosure.
- **Automation** of several tasks, which leads to a better performance in execution time and the elimination of manual work, especially for complex tasks. There can be mentioned many examples such as database diagnosis, monitoring and alerting done automatically by machine learning applications, or database modeling, scheduling and self-fixing anytime it is necessary. Artificial intelligence can also automate the decision-making process regarding resource usage, without the involvement of administrators, or automate the discovery and classifications of data. It can also improve productivity and performance by helping the users with simple interfaces containing voice commands or chatbots, or by adding automatic data preparation and interpretation, report generation and visual analysis, planning and forecasting, etc.
- **Optimization** in different directions of data management by applying new approaches and best practices in improving the processes, such as finding the best engines and infrastructure setup, understanding and rewriting queries, etc.

Schmarzo and Caldwell (2021), propose going from the traditional data management seen as a passive and cost-avoidant set of technology functions, to the next generation data management solution defined as a proactive and data monetization business strategy, that emphasizes the real value of data. The integration of artificial intelligence and machine learning features in this solution adds a lot of benefits to data management strategies such as:

- Data discovery and data content automation, which enable faster and simpler discovery of relevant data from the users.
- Easier data exploration and data access.
- Faster experiments for data science.
- Better productivity from the user's perspective because of integrated insights and appropriate suggestions related to the data consumer issue context.
- Produce data and analytics progressively by highlighting issues and monitoring the performance of the machine learning model.
- Identify and trace the usage patterns of data consumers to enhance the productivity cycle in the data management ecosystem.

The role of artificial intelligence in data management and its positive impact on business can be viewed in different aspects. Primarily it is concerned with creating useful and accurate insights for helping decision-making and business growth. Companies can use artificial

intelligence on data to forecast stock prices for better trading opportunities (USDSI, 2022). They can also predict through historical information and transactional patterns, future expenses, the necessary budget, and expected revenues. Meanwhile, IBM based on a study over the 451 Research company evidenced the following impacts (IBM, 2019):

- **Better operational efficiency** through automatic queries management.
- **Better query performance and reliability**, leading to producing insights and making decisions in a shorter time.
- **Empower experts** through the results of machine learning models.
- **Better productivity** because of lowering the time of accessing and preparing the data through the integration of machine learning and developer tools.
- **Facilitation for database administrators** through the automatization of tasks such as database tuning and provisioning.

Apart from the benefits that come with the integration of artificial intelligence in data management systems, still there exist many challenges related primarily to the enormous amount of data that must be gathered, stored, analyzed, and visualized, often even in real-time. An example is shown on wireless sensor networks that should be able to capture, process, store, synchronize, and manage multiple data streams from large and dynamic sources (Diallo et al., 2012). Pike et al. (2019) mention technology, software, and social challenges, which include respectively the need for better efficiency and performance, the need for implementing algorithms and real-time decision-making solutions, and the challenges related to safety, security, and privacy. Ovalle et al. (2012) mention the need for intelligent behavior from the sensor network systems in order to better understand the gathered data and their analytics and highlighted the complexity of biomedical data due to privacy issues and heterogeneous types. Meanwhile Hassanalieragh et al. (2015) and Hung et al. (2015) focus on the challenges of real-time decisions making, raised by the integration of voluminous data streams and traditional biomedical information systems and from the need for developing accurate data-processing algorithms that comply with individuals' privacy rights and local legal systems and standards.

AUGMENTED DATA MANAGEMENT

Data lifecycle management is defined as an approach based on standards and policies, used to manage the data flow of information systems from creation and exploration to when it becomes obsolete and is deleted from the storage (Sheldon, 2022). Meanwhile augmented data management is the integration of artificial intelligence and machine learning techniques to help the automatization of some data management tasks such as data collection, integration,

preparation, cataloging, etc. throughout the data lifecycle (Karatas, 2021). It helps in the direction of:

- Better data quality as it makes automatic checks and data cleansing.
- Automated metadata management by collecting, organizing, and analyzing automatically structured and unstructured data.
- Unification of the data views by combining all types of data from several data sources.
- Automatic generation of master data as a single source of accurate and necessary information for all types of data consumers.

Foot (2021) interprets the augmented data lifecycle management as a capability of artificial intelligence and machine learning to enable self-configuring and self-turning data management by focusing mostly on these parts of the data lifecycle:

- Data documentation and governance through an effective automatic metadata management
- Data infrastructure management that helps the support department solve issues rapidly.
- Data analysis and presentation with improved capabilities in the phases of data ingestion, preparation, analysis, and exploration.

OTHER TRENDS RELATED TO DATA MANAGEMENT

Apart from machine learning, deep learning, and other techniques based on artificial intelligence, and used for the improvement of different aspects of data management, several other new approaches can be mentioned in this study as trends in data management:

The hybrid, intercloud, and multi-cloud approaches. In recent years, many companies have transferred their data and their applications to the cloud, using the benefits offered by this approach. Due to the covid-19 pandemic, many other companies were forced to join this strategy to communicate and coordinate more easily remotely. They have chosen cloud-based environments, tools, and solutions and they have implemented different data management strategies based on hybrid, intercloud, and multi-cloud approaches (Everett, 2022). This means that some companies use different cloud ecosystems for storing their resources and for accessing different tools to obtain the maximum potential of cloud-based technology, security, and financial benefits (Foote, 2022). The drawback of this approach is the complexity of the applications and data, which should be compatible and interoperable with the different public, private, and on-premises cloud environments.

Self-service analytics tools. The use of analytics and the support offered by many data management platforms are being embraced by many companies, because of the many benefits related to costs optimization, income increase, and better competitiveness. The analytics

expansion is enforced using augmented and cloud analytics, and also by self-service analytics tools (Foote, 2022). These are user-friendly tools, easy to understand, easy to learn, and easy to use, even for non-technical users. They can produce reports from specific data in real-time.

DataOps is an approach based on agile methodology and inspired by the DevOps principles (Ereth, 2018). It combines principles, processes, methods, and practices on a cultural, organizational, and technological level. DataOps' focus is to improve communication among the company members, by removing the barriers between small teams and promoting a culture of continuous improvement. Doing so also decreases the time of processes for data lifecycle and analytics.

Data governance. As defined by The Data Governance Institute (DGI, 2022), data governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods. In the era of big data, associated with privacy laws, standards, and regulations such as the GDPR (General Data Protection Regulation), data governance becomes very difficult. Practically it is related to developing systems, standards, rules, processes, and procedures to deliver to all the stakeholders of a company, consistent and secure data, intending to increase data quality, ensure process transparency, enable better decision-making, and increase the company efficiency, etc. A detailed and suitable architecture for data governance can improve business outcomes across an organization (Bartley, 2022).

Blockchain. Blockchain technology is used in data management strategies to store data in decentralized systems and to provide improved security, audit trails, and asset traceability. In this structure, the records cannot be changed, which implies better accuracy and authenticity for the data (Hunt, 2022). This technology is also promising for data management in the sixth-generation wireless networks (6G), which is a highly dynamic and heterogeneous environment. It complies naturally with two major requirements that are decentralization and transparency (Shen et al., 2022), but increases the complexity of data management in the direction of efficiency and privacy (Eberhardt & Tai, 2018).

Data fabric is an emerging approach for data management, functioning on a network-based architecture that enables data handling (analytics, insights, orchestration, etc.) from the data sources level (BasuMallick, 2022). It can be defined as a collection of hardware and software infrastructure, tools, and other services used to handle data operations on all data storage platforms, across the company. It includes all data activities such as data acquisition, transformation, storage, distribution, integration, replication, availability, security, protection, and analytics (McSweeney, 2019). This approach can be adopted by several sectors. An interesting

proposal comes from Kuftinova et al. (2022) that uses the data fabric technology as a system for gathering data from transport systems and events. They suggested adding blockchain technology and artificial intelligence for a secure, decentralized, and successful transport ecosystem.

Data marketplace is a capability to address the need for greater data access and sharing. It provides order management and delivery actions for a better and simpler data consumption experience in an online shop. Users can easily find the needed topic, put on data sets into their shopping cart, make the checkout process, and have data securely delivered (Everett, 2022).

Edge computing. These technologies have emerged as a way to support cloud computing and other traditional technologies 1for data processing, in the context of a modern and intelligent society, where huge volumes of data gathered from different smart devices cause issues such as poor security and privacy, large bandwidth load, data transmission delay, slow response speed, etc. Edge computing is a new paradigm for performing calculations at the edge of the network, which can be closer to both the user and the source of the data (Cao et al., 2020). This approach is suitable for real-time data analytics by providing speed, flexibility, agility, and autonomy. It can be used not only for analytical reasons but also during the phases of the creation and management of data. Some avant-garde companies are focused on obtaining new opportunities and benefits by combining the concepts of edge computing and artificial intelligence, which is also known as Edge AI (Vyas, 2021). This approach can help applications based on artificial intelligence to have better performances with lower operational costs and to overcome challenges related to the technology part.

Going from big data to small and wide data. Businesses use different approaches, such as composable analytics and data mining techniques to collect and analyze data from different sources, to gain valuable results, useful information, and insights. Nevertheless, processing all these data is time-consuming and needs a lot of effort. Tools and frameworks for composable data analytics can combine different analytics techniques to perform the processes of receiving, composing, messaging, and analyzing the data, with only a minimum of interaction with the user (Goh & Chua, 2018). This helps the companies to identify or predict customer behavior leading to more effective decisions. The composable data analytics approach improves the efficiency and agility of data management compared to the traditional tools. Many companies shortly will shift the focus from big data and pay attention to small and wide data to perform more intelligent decision-making (Vyas, 2022), in the direction of innovation, strategies, and effective changes.

Multidomain master data management. Master data is a concept used for connecting business transactions with business entities or domains, such as customers, materials, products, locations, suppliers, etc. A step further, the multidomain master data management

helps the companies in having a full view of their business, to improve and ensure the quality, integrity, timeliness, consistency, availability, and accuracy of the master data. This can be achieved through a proposed framework by Dahlberg et al. (2015), which provides several practical means for resolving master data drawbacks.

CONCLUSIONS

This paper is a literature review with a focus on augmented data management. There are considered the added values and capabilities from the usage of artificial intelligence in data management processes such as prediction, identification, automation, and optimization. There are evidenced benefits including data discovery and data content automation, simpler data exploration and data access, faster experiments, better productivity, and better performance for generating accurate analytics. It is shown the business impact that brings artificial intelligence in data management such as the help in decision-making and business growth, better operational efficiency, query performance, reliability, and the empowerment of experts. Subsequently, in the literature are evinced some challenges to be faced that come especially from the voluminous and complex data. Other challenges include security, privacy, performance, and real-time decision-making. Apart from the artificial intelligence approach, this study shows a general overview of the latest trends, techniques, and approaches, used by different researchers and companies for improvements in data management solutions, such as hybrid, intercloud, and multi-cloud approaches, self-service analytics tools, dataOps, data governance, blockchain technology, data fabric, data marketplace, edge computing, small and wide data, and multidomain master data management. The scope for further studies related to this topic is to analyze a case for a specific system, by considering quantitative and qualitative KPIs (Key Performance Indicators), to better evaluate the differences and benefits of applying artificial intelligence in data management.

REFERENCES

Aslett, M., & Curtis, J. (2019). Accelerating AI with Data Management; Accelerating Data Management with AI. Available at: <https://www.ibm.com/downloads/cas/YD5R1XLB> [Accessed 17 August 2022].

Bartley, K. (2021). Data Management Trends & Technology (2021) | Rivery. [online] Rivery. Available at: <https://rivery.io/blog/data-management-2021-trends-technology-that-will-define-the-year/> [Accessed 8 June 2022].

Basumallick, C. (2022). What Is Data Fabric? Definition, Architecture, and Best Practices | Spiceworks. [online] Spiceworks. Available at: <https://www.spiceworks.com/tech/big-data/articles/what-is-data-fabric/> [Accessed 7 August 2022].

Cao, K., Liu, Y., Meng, G., & Sun, Q. (2020). An Overview on Edge Computing Research. IEEE Access, 8, pp.85714-85728.

Dahlberg, T., Heikkilä, J., Heikkilä, M., & Dahlberg, T. (2011). Framework and Research Agenda for Master Data Management in Distributed Environments. Proceedings of IRIS 2011 Conference.

Dennis, A. L. (2022). Artificial Intelligence Augments Data Management - DATAVERSITY. [online] DATAVERSITY. Available at: <https://www.dataversity.net/artificial-intelligence-augments-data-management/> [Accessed 15 May 2022].

DGI - The Data Governance Institute. (2003). Definitions of Data Governance - The Data Governance Institute. [online] Available at: <https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/> [Accessed 9 July 2022].

Diallo, O., Rodrigues, J., & Sene, M. (2012). Real-time data management on wireless sensor networks: A survey. Journal of Network and Computer Applications, 35(3), pp.1013-1021.

Eberhardt, J., & Tai, S. (2018). ZoKrates-scalable privacy-preserving off-chain computations. Proceedings of 2018 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data; 2018 Jul 30–Aug 3; Halifax, NS, Canada; 2018. p. 1084–91.

Ereth, J. (2018). DataOps - Towards a Definition. Lernen, Wissen, Daten, Analysen Conference. Mannheim, 2018.

Everett, D. (2022). The Top Five Data Management Trends to Watch in 2022. [online] Informatica. Available at: <https://www.informatica.com/blogs/the-top-five-data-management-trends-to-watch-in-2022.html> [Accessed 8 June 2022].

Foot, C. (2021). Enterprise augmented data management benefits and growth. [online] SearchDataManagement. Available at: <https://www.techtarget.com/searchdatamanagement/feature/Growing-enterprise-benefits-of-augmented-data-management> [Accessed 17 July 2022].

Foote, K. (2022). Data Management Trends in 2022 - DATAVERSITY. [online] DATAVERSITY. Available at: <https://www.dataversity.net/data-management-trends-in-2022/> [Accessed 9 July 2022].

Goh, J., & Chua, F. (2018). Dynamic Composable Analytics on Consumer Behaviour. Computational Science and Its Applications – ICCSA 2018, pp.353-365.

Hassanalieragh, M., Page, A., Soyata, T., Sharma, G., Aktas, M., Mateos, G., Kantarci, B. & Andreescu, S. (2015). Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges. International Conference on Services Computing, IEEE, pp. 285-292, 2015.

Hung, K., Lee, C. C., & Choy, S. O. (2015). Ubiquitous health monitoring: Integration of wearable sensors, novel sensing techniques, and body sensor networks. Mobile health 2015, pp. 319-342. Springer, Cham.

Hunt, S. (2021). Top Data Management Trends & Technology 2022 | Datamation. [online] Datamation. Available at: <https://www.datamation.com/big-data/data-management-trends/> [Accessed 9 July 2022].

IBM, (2019). AI and Data Management. Delivering Data-Driven Business Transformation and Operational Efficiencies. [online] Ibm.com. Available at: <https://www.ibm.com/downloads/cas/MGJ2DLQV> [Accessed 07 August 2022].

Karatas, G. (2021). 4 Ways Augmented Data Management Changes Traditional Data Management. [online] AIMultiple. Available at: <https://research.aimultiple.com/augmented-data-management/#:~:text=Augmented%20data%20management%20automates%20data,decisions%20based%20on%20reliable%20data.> [Accessed 17 July 2022].

Kuftinova, N. G., Maksimychev, O. I., Ostroukh, A. V., Volosova, A. V., & Matukhina, E. N. (2022). Data Fabric as an Effective Method of Data Management in Traffic and Road Systems. Systems of Signals Generating and Processing in the Field of on Board Communications, 2022, pp. 1-4, doi: 10.1109/IEEECONF53456.2022.9744402.

McSweeney, A. (2019). Designing An Enterprise Data Fabric. Data Architecture. May 2019.

Ovalle, D., Restrepo, D., & Montoya, A. (2010). Artificial intelligence for wireless sensor networks enhancement. Smart Wireless Sensor Networks, InTech, 2010.

Pike, M., Mustafa, N. M., Towey, D., & Brusic, V. (2019). Sensor Networks and Data Management in Healthcare: Emerging Technologies and New Challenges. IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), 2019, pp. 834-839, doi: 10.1109/COMPSAC.2019.00123.

Schmarzo, B., & Caldwell, R. (2021). The Ugly Truth About Data Management & the Journey to Unleashing the Economic Value of Data. Whitepaper. [online] Delltechnologies.com. Available at: <https://www.delltechnologies.com/asset/en-us/products/servers/industry-market/the-ugly-truth-about-data-management.pdf> [Accessed 16 June 2022].

Sheldon, R. (2022). What is data lifecycle management (DLM)? - Definition from TechTarget. [online] SearchStorage. Available at: <https://www.techtarget.com/searchstorage/definition/data-life-cycle-management> [Accessed 17 July 2022].

Shen, X., Liu, D., Huang, C., Xue, L., Yin, H., Zhuang, W., Sun, R., & Ying, B. (2021). Blockchain for Transparent Data Management Toward 6G. Engineering, 8, pp.74-85.

USDSI, (2022). Unlocking Your Data with AI and ML. [online] USDSI Available at: <https://www.usdsi.org/data-science-insights/unlocking-your-data-with-ai-and-ml> [Accessed 07 August 2022].

Vyas, K. (2021). Edge AI: The Future of Artificial Intelligence and Edge Computing | ITBE. [online] IT Business Edge. Available at: <https://www.itbusinessedge.com/data-center/developments-edge-ai/> [Accessed 5 July 2022].

Vyas, K. (2022). Top 7 Data Management Trends to Watch in 2022 | IT Business Edge. [online] IT Business Edge. Available at: <https://www.itbusinessedge.com/business-intelligence/top-data-management-trends-2022/> [Accessed 9 July 2022].