Re-Thinking Data Strategy and Integration for Artificial Intelligence

Subjects: Engineering, Industrial
Contributor: Abdulaziz Aldoseri, Khalifa N. Al-Khalifa, Abdel Magid Hamouda

The use of artificial intelligence (AI) is becoming more prevalent across industries such as healthcare, finance, and transportation. Artificial intelligence is based on the analysis of large datasets and requires a continuous supply of high-quality data.

Keywords: Artificial Intelligence (AI); data strategies and learning approaches; challenges and opportunities

1. Introduction

Artificial Intelligence (AI) refers to the ability of machines to mimic human intelligence and perform tasks that typically require human intelligence, such as learning, problem-solving, decision-making, and natural language understanding [1]. Figure 1 depicts AI technologies including machine learning, natural language processing, robotics, and computer vision. Machine learning is a subset of AI that involves training computer algorithms to learn patterns in data and make predictions or decisions based on the data [2]. Deep learning is a type of machine learning that uses neural networks with multiple layers to process complex data such as images or speech [3]. Natural language processing is the ability of computers to understand, interpret, and generate human language, including speech and text [4]. Computer vision is the ability of computers to analyze and interpret visual information such as images and videos [5].

Figure 1. AI Technology Landscape.

AI is a rapidly expanding field with the potential to revolutionize the way we live and work. From healthcare to finance and transportation, AI has the potential to transform a wide range of industries, creating new opportunities for businesses and organizations. AI has been transforming various sectors, including healthcare, finance, and transportation, with significant advancements in machine learning and deep learning techniques [6,7]. The heart of this transformation is data, which are essential for training and testing the AI models. AI models rely on large datasets to identify patterns and trends that are difficult to detect using traditional data-analysis methods. This allows them to learn and make predictions based on the data on which they have been trained.

However, using AI data is challenging. Data quality, quantity, diversity, and privacy are critical components of data-driven AI applications, and each presents its own set of challenges. Poor data quality can lead to inaccurate or biased AI models, which can have serious consequences in areas such as healthcare and finance. Insufficient data can lead to models that
are too simplistic and incapable of accurately predicting real-world outcomes. A lack of data diversity can also lead to biased models that do not accurately represent the population they are designed to serve. Lastly, data privacy is a major concern, as AI models may require access to sensitive data, which raises concerns about data privacy and security.

2. Dimension of Data Quality and Implication for AI systems

Particularly pertinent to AI systems, the analysis disclosed the following dimensions of data quality: Accuracy, Completeness, Consistency, Timeliness, Relevance, and Integrity.

AI predictions and decisions are influenced by crucial dimensions of performance, reliability, and trustworthiness. The quality of data across these dimensions must be diligently maintained to minimize the hazards of distorted or prejudiced outcomes and optimize the efficiency of AI programs.

3. The Role of Data Governance in Ensuring Data Quality

Ensuring data quality is a crucial aspect of data governance. It involves managing and monitoring the data to maintain accuracy, completeness, and consistency. Data governance plays a vital role in this process by establishing policies and standards to regulate how data are collected, used, and shared. By implementing effective data governance, organizations can reduce the risk of errors and inconsistencies in their data that could lead to costly mistakes and poor decision making. Overall, data governance is essential for maintaining high-quality data and ensuring their usefulness and reliability for various purposes.

Maintaining the data quality for AI systems can be ensured through data governance, an aspect that the analysis has also highlighted. Organizational benefits from establishing a strong data governance system include the following:

(a) Quality standards and policies for data must be defined and put into action.

(b) Throughout the lifecycle of the data, it is important to keep a close eye on their quality and maintain control.

(c) Quality data and holding ourselves accountable should be part of a culture.

(d) The sharing, integration, and management of data can be enhanced through various means. The optimization of data management techniques should be prioritized. Improved data sharing is crucial for seamless exchanges between different systems. The integration of various data types can be achieved using appropriate methods.

(e) Regulations and laws must be followed carefully to maintain compliance.

Systematically addressing data quality challenges and minimizing the risks associated with poor data quality can be achieved by integrating data governance into the AI development process.

4. Best Practices to Ensure Data Quality for AI

AI systems’ data quality can be ensured by adopting the following best practices:

(a) Implementing an effective data management strategy that includes data curation and preprocessing before usage.

(b) Fostering transparency and accountability in the data collection process, including defining data sources and conducting regular audits.

(c) Conducting diversity checks on the collected dataset to avoid bias, and making sure that it is representative of the target population.

(d) Ensuring the security and privacy of the data by implementing the necessary security protocols and obtaining consent from the data subjects.

(e) Proactively monitoring and updating the dataset to maintain accuracy and relevance, especially when it comes to dynamic or constantly changing environments.

AI systems can provide reliable data quality if organizations implement certain best practices, and researchers have identified quality data frameworks and strategies that need to be developed and implemented. The structures and
processes for data governance must be established to ensure proper management. Governance data structures and processes provide accountability and responsibility for data management. Establishing governance structures and processes is critical for ensuring the proper use of data. To effectively manage data, clear guidelines and procedures must be in place for those responsible for handling them. Proper governance ensures that data are properly managed, and that rights permissions and access are granted. Without these structures and processes, data can be lost or misused, affecting an organization's overall performance. Regular assessments and audits must be conducted to ensure data quality. Do not forget to conduct these examinations sporadically. Enrichment tools, coupled with data cleansing and validation, should be used. Traceability solutions and data lineages are means of implementing change. Machine learning and AI-based solutions offer powerful tools for improving data quality; therefore, their adoption is highly recommended. The best data quality practices should be taught to employees through training sessions. To enhance the quality of the data, alliances can be formed with colleagues and associates. Organizations can improve their AI performance and reliability by embracing these best practices, which in turn will increase the quality of the data used. Implementing data governance and best practices across various dimensions is necessary to unlock the full potential of AI and to drive better outcomes. Organizations must address data-quality challenges to ensure the successful development and deployment of AI systems, as evidenced by a analysis. Data quality plays a critical role and should be a top priority for organizations that utilize AI.

References


Retrieved from https://encyclopedia.pub/entry/history/show/104588