

Review

Not peer-reviewed version

Re-think Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities and Challenges

[Abdulaziz Aldoseri](#) , [Khalifa N Al-Khalifa](#) , [Abdelmagid S. Hamouda](#) *

Posted Date: 1 June 2023

doi: 10.20944/preprints202305.1565.v2

Keywords: Data Strategies and learning approaches; Challenges and Opportunities; Artificial Intelligence



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Re-Think Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities and Challenges

Abdulaziz Aldoseri, Khalifa N. Al-Khalifa and Abdel Magid Hamouda *

Engineering Management Program, College of Engineering, Qatar University, P.O Box 2713, Doha, Qatar

* Correspondence: hamouda@qu.edu.qa; Tel.: +974-4403 4303

Abstract: 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. However, using data for AI is not without challenges. This paper comprehensively reviews and critically examines the challenges of using data for AI, including data quality, data volume, privacy and security, bias and fairness, interpretability and explainability, ethical concerns, and technical expertise and skills. This paper examines these challenges in detail and offers recommendations on how companies and organizations can address them. By understanding and addressing these challenges, organizations can harness the power of AI to make smarter decisions and gain competitive advantage in the digital age. It is expected, since this review article provides and discusses various strategies for data challenges for AI over the last decade, that it will be very helpful to the scientific research community to create new and novel ideas to rethink our approaches to data strategies for AI.

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].

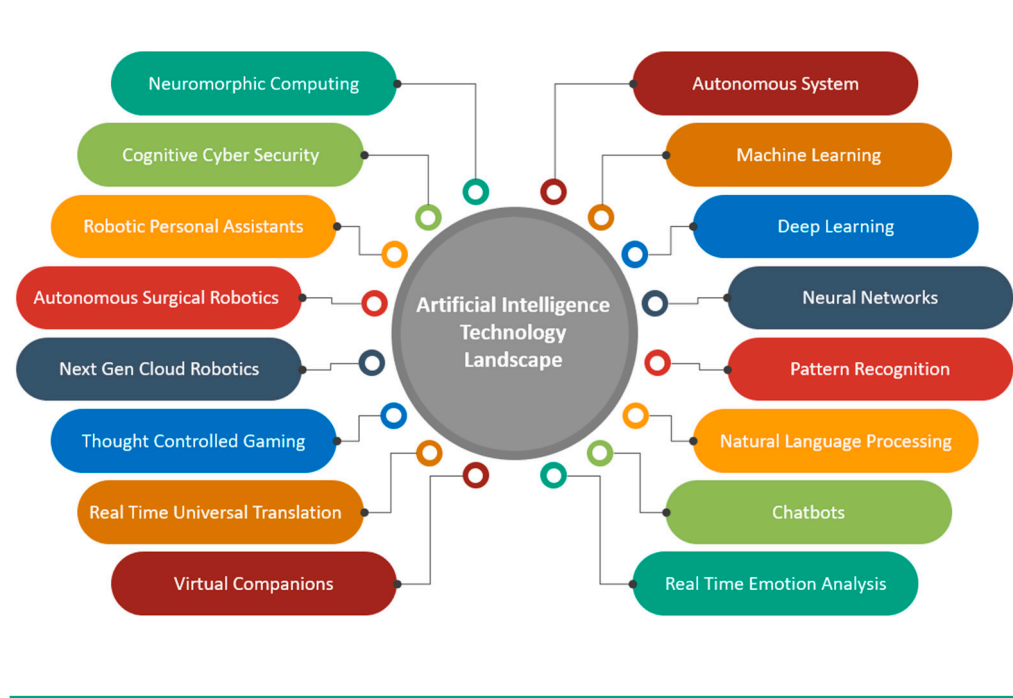


Figure 1. AI Technology Landscape.

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 is 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.

In this article, we address the challenges of using data for AI and offer recommendations for companies seeking to address them. To address these challenges, businesses and organizations need to develop strategies and frameworks that promote data quality, quantity, diversity, and privacy. This may involve implementing data cleaning and validation processes to ensure data quality, collecting and managing large quantities of diverse data, and implementing data privacy policies and procedures to protect the sensitive data. By focusing on these challenges, businesses and organizations can leverage the power of data to create accurate, effective, and fair AI applications that benefit society.

2. Materials and Methods

2.1. Data for AI

Data are critical for AI because they are the foundation upon which machine learning algorithms learn, make predictions, and improve their performance over time. To train an AI model, large

amounts of data are required to enable the model to recognize patterns, make predictions, and improve its performance over time.

2.1.1. Data Learning Approaches

AI algorithms require data to learn patterns and make predictions or decisions based on the data. AI machine learning techniques are algorithms that allow machines to learn patterns and make predictions from data without explicit programming [8]. These techniques are widely used in a variety of applications such as natural language processing, image and speech recognition, and recommendation systems. In general, the more data available for an AI algorithm to learn, the more accurate its predictions or decisions will be. There are several data-learning approaches to building AI systems [8,9], for comprehensiveness of the article, we include the following, as shown in Figure 2:

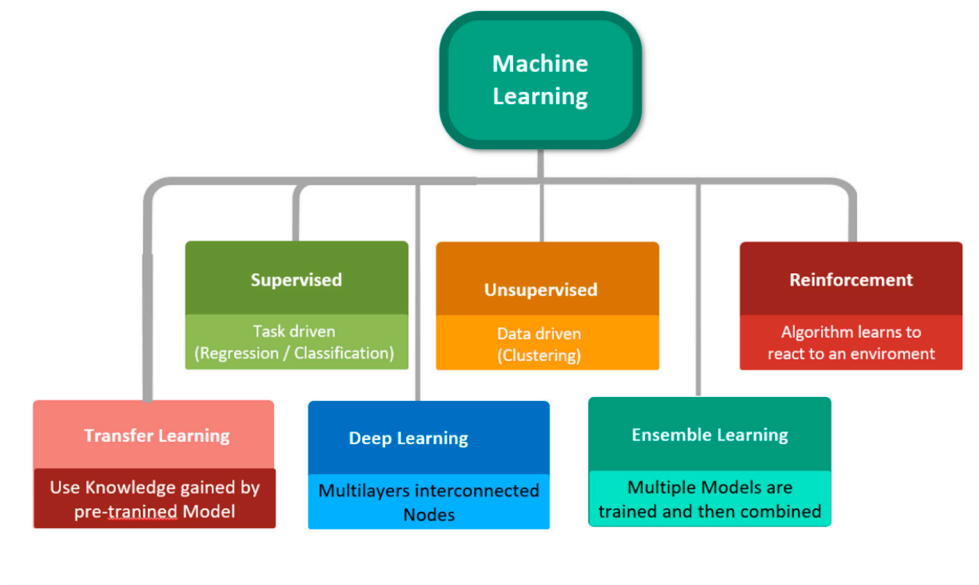


Figure 2. Machine Learning Approaches.

Supervised Learning: In supervised learning, an AI system is trained on a labeled dataset, where each data point is associated with a label or a target variable. The goal is to develop a model that can accurately predict the label or target variable for new data points. This approach is commonly used for tasks such as image classification, speech recognition, and natural language processing [10].

Unsupervised Learning: In unsupervised learning, an AI system is trained on an unlabeled dataset where there is no target variable to predict. The goal is to identify the patterns, relationships, and structures in the data. This approach is commonly used for tasks such as clustering, anomaly detection, and dimensionality reduction [11].

Reinforcement Learning: In reinforcement learning, an AI system learns to make decisions based on feedback from the environment. The system receives rewards or penalties based on its actions and adjusts its behavior accordingly. This approach is commonly used for tasks such as gaming, robotics, and autonomous driving [12].

Transfer Learning: In transfer learning, an AI system leverages the knowledge gained from one task to improve the performance on another related task. The system is pre-trained on a large dataset and then fine-tuned on a smaller dataset for a specific task at hand. This approach can help reduce the amount of data required to train an AI model and improve its accuracy and performance [13].

Deep Learning: Deep learning is a type of neural-network-based machine learning that is particularly effective for tasks involving large amounts of data and complex relationships. Deep learning models are composed of multiple layers of interconnected nodes that can learn increasingly complex representations of data. This approach is commonly used for tasks such as image and speech recognition, natural language processing, and computer vision [14].

Ensemble Learning: Ensemble learning is a technique in which multiple models are trained and combined to make predictions or decisions. Combining the predictions of multiple models can improve the accuracy and reliability of the final output [15].

Overall, the choice of the data learning approach depends on the specific task, data, and resources available. It is important to carefully evaluate the benefits and limitations of each approach and select the one that best fits the requirements of the AI application being developed.

2.1.2. Data-Centric and Data-Driven AI

Data-centric and data-driven are two related but distinct concepts in the world of data analysis and decision making. By leveraging data, organizations can gain a deeper understanding of their operations, customers, and markets and make more informed decisions based on data-driven insights. Data-centric approaches are commonly used in industries such as finance, healthcare, and retail, where accurate and timely data are critical for decision making. For example, in the healthcare industry, data-centric approaches are used to analyze patient data to improve outcomes, identify disease patterns, and optimize treatment plans. Data-centric and data-driven approaches are two approaches for building AI systems that rely on data [16,17].

Data-centric Approach: It refers to an approach in which data is the central focus of a system or process [16,18]. A data-centric approach involves a relatively fixed model that prioritizes the collection, storage, and analysis of high-quality data to train AI algorithms, improve their performance, and leverage data to inform decision-making and problem-solving processes [16,18,19]. This approach often involves using advanced analytics such as machine learning or artificial intelligence to uncover patterns, trends, or insights that may not be immediately apparent from the data [19]. The data-driven approach focuses on building robust and reliable data infrastructure that can support a wide range of AI applications. The goal is to create a centralized data repository that can serve as a single source of truth for all AI applications within an organization [20]. This approach is particularly useful when there is a large volume of data from different sources, or when the data are complex and difficult to work with.

In recent years, the rise of big data and advanced analytics has led to a growing emphasis on data-centric approaches across various industries from healthcare to finance to retail [21]. By adopting a data-centric approach, organizations can gain a competitive advantage by improving decision making, increasing efficiency, and reducing costs [22]. A data-centric approach is particularly important in the context of big data, where 7 V of big data (velocity, volume, value, variety, veracity, volatility, and validity) can make it challenging to extract meaningful insights [23]. AI algorithms must be designed to handle large volumes of data, which must be carefully curated to ensure accuracy and relevance [24]. A data-centric approach can lead to improved decision-making, increased efficiency, reduced costs, improved customer experience, competitive advantage, and risk mitigation [25,26]. It requires a strong data management infrastructure, skilled workforce, and advanced analytics and AI techniques to extract valuable insights from the data.

Overall, a data-centric approach is essential for effective AI decision-making and problem solving. By placing data at the center of the AI system and following best practices for data quality, processing, governance, and integration, organizations can unlock the full potential of AI and drive better outcomes.

Data-driven Approach: It focuses on building AI models that are specifically designed to make predictions or decisions based on data. This approach emphasizes the selection, processing, and analysis of data to identify patterns, relationships, and insights that can be used to improve the accuracy and performance of an AI model [27]. The goal was to develop an AI model that can learn and adapt to new data without being constrained by a predefined set of rules or assumptions. This approach is particularly useful when data are relatively homogeneous or when the goal is to automate a specific decision-making process [28]. A data-driven approach to AI involves the use of data as the primary source of information for training and improving AI models. In this approach, the AI system learns directly from the data rather than being programmed by humans [29]. Data-driven AI involves several key steps, as illustrated in Figure 3.

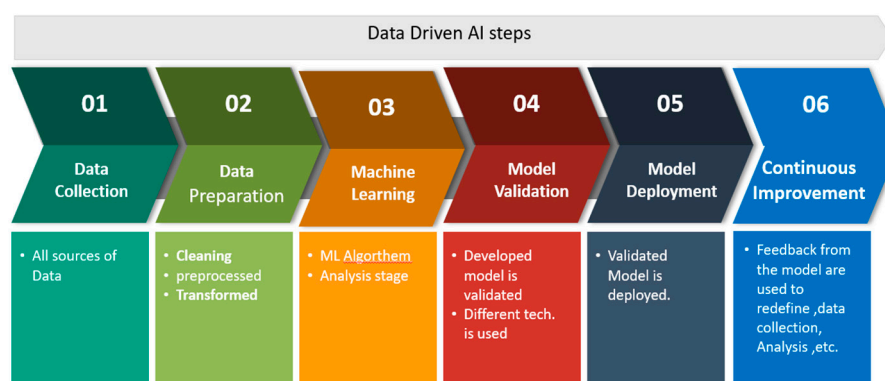


Figure 3. Data-driven AI Process.

Data Collection: Collecting relevant data from various sources is the first step in a data-driven approach. This may involve capturing data from internal systems, external sources, or user-generated content [30].

Data Preparation: Once the data are collected, it needs to be cleaned, preprocessed, and transformed to make it suitable for analysis. This may involve data cleansing, normalization, and feature engineering [31].

Machine Learning: Machine learning algorithms are applied to preprocessed data to develop predictive models that can be used to make decisions or automate processes [32].

Model Validation: The models developed through machine learning are validated using various techniques to ensure accuracy and reliability [33].

Model Deployment: Once models have been validated, they are deployed in production environments to automate the decision-making processes or provide insights [34].

Continuous Improvement: A data-driven approach involves continuous improvement, with feedback from the models used to refine the data collection, analysis, and decision-making processes [35].

A key advantage of a data-driven approach to AI is that it allows the AI model to adapt and improve over time as new data become available. This means that the model can continue to learn and refine its predictions and performance, leading to better outcomes over time. Data-driven AI has several advantages over other approaches, including the ability to learn from large amounts of data, detect complex patterns and relationships, and adapt to changing conditions [36]. However, it also requires careful attention to data quality, privacy, and ethical considerations.

Overall, a data-driven AI approach emphasizes the importance of data at every stage of the AI development and decision-making process, from data collection to model deployment. This approach can help organizations make informed decisions and improve the accuracy and effectiveness of their AI models.

Generally, both data-centric and data-driven approaches are important for building effective AI systems. A strong data-centric approach can help ensure that the data used to train AI models are of high quality, whereas a data-driven approach can help identify patterns and insights that can improve the accuracy and performance of the AI model. The choice between these two approaches depends on the specific needs and requirements of the organization and AI application being developed.

2.2. Dimensions of Data Challenges for AI

Several data challenges are associated with AI. Figure 4 depicts the dimensions of the data challenges for AI. This section covers the most important and essential and major ones, as illustrated in Figure 4.



Figure 4. Dimensions of Data Challenges for AI.

2.2.1. Dimension I: Data Quality

Data quality is a critical aspect of AI. The accuracy, completeness, and consistency of the data used for training and testing AI models directly affects the performance and effectiveness of the AI system. Low-quality data can lead to biased, inaccurate, or irrelevant results, negatively affecting decision-making processes based on AI outputs. Therefore, ensuring high-quality data are crucial for AI systems to produce reliable and valuable results. These include data cleansing, validation, enrichment, and management. AI applications require high-quality relevant, representative, and reliable data to produce optimal outcomes. AI systems also require ongoing monitoring and maintenance to ensure that data quality is consistent over time. The performance of AI systems is heavily reliant on the quality of the data used for training and validation [37]. Data quality is a multidimensional concept that encompasses factors such as accuracy, completeness, consistency, and timeliness [38]. Ensuring data quality is a challenging task given the vast amount of data generated daily and the inherent complexity of data structures [39]. Figure 5 shows the challenging elements of the data quality.



Figure 5. Challenging elements of data quality.

2.2.1.1. Challenging Measures of Data Quality and Implications on AI Systems

This section presents different measures to ensure the data quality for AI applications. Figure 6 represents the challenging measures of Data Quality.



Figure 6. Data quality measures.

Accuracy: Data accuracy is critical for AI to function effectively. Accuracy refers to the degree to which data are correct and error-free. In other words, it aims to describe the degree to which data correctly represent real-world phenomena [40]. AI systems require accurate data for training and validation to ensure accurate predictions and decisions [41]. Inaccurate data can lead to biased or erroneous outcomes, undermining the reliability and usefulness of AI systems [42].

Completeness: Completeness refers to the extent to which all relevant and sufficient data coverage is present in the dataset to provide insight and meaningful results for AI [37]. Incomplete data can lead to biased or unrepresentative AI models because the algorithms may not have sufficient information to learn the underlying patterns and relationships [43]. Missing data can be attributed to various factors such as data collection or data entry errors [44].

Consistency: Consistency refers to the uniformity of data representations and formats across a dataset [38]. In other words, this is the extent to which the data are free from conflicts, inaccuracies, or discrepancies when compared to other sources or systems. Inconsistent data can lead to confusion and misinterpretation by AI algorithms, resulting in suboptimal performance [45]. Ensuring consistency requires the standardization and harmonization of data formats, units, and terminologies [46].

Timeliness: Timeliness refers to the degree to which data is updated and relevant to the current context [37]. AI systems require timely data to adapt to dynamic environments and provide accurate predictions [47]. Outdated data may lead to poor performance and even harmful consequences because AI systems may not account for recent changes in the underlying phenomena [48]. Timeliness is particularly important in domains such as finance, healthcare, and transportation, where real-time insights offer significant advantages. For instance, if the data used to build a weather forecasting model are outdated, the model might not be able to make accurate predictions. Similarly, if the data used for training a stock market predictor are not timely, the model can make decisions based on outdated information that may not be relevant to the current state of the market. Therefore, data timeliness is an important dimension of AI data quality.

Integrity: Data integrity refers to the maintenance of data accuracy and consistency throughout its lifecycle, including during storage, retrieval, and processing [46]. In other words, it is the degree to which data are reliable and trusted to be correct. Compromised data integrity can result in AI systems making decisions based on corrupt or inconsistent data, leading to unreliable or flawed outcomes.

Relevance: Relevant data refer to the degree to which the data used for training and building machine learning models are appropriate and applicable to the task or problem being addressed [37]. This is directly related to a specific problem or task being addressed by an AI system. Irrelevant data can introduce noise or bias into a system, thereby reducing its performance and effectiveness [39].

By considering these dimensions of data quality and their implications for AI systems, organizations can better understand the challenges they face in maintaining high-quality data for AI applications. This understanding can inform the development of strategies and best practices to

address data quality issues, thereby ensuring that AI systems can deliver accurate, reliable, and valuable insights and outcomes.

2.2.1.2. Challenges in Data Collection, Pre-processing, and Management

Data Collection: Data collection for AI applications is often driven by the problem to be solved to assure relevant collected data as challenged because of the sheer volume of data, diversity of data sources, and need for representative samples [49]. Data collection to assure data quality is determined by data requirements and identifying the types of data needed for the application, data sources, and data quantity [50]. To ensure data quality in the data collection phase, the following considerations must be considered:

Data Pre-Processing: Data pre-processing is a crucial step in ensuring data quality, as it involves cleaning, transforming, and integrating the data to facilitate analysis [51]. Pre-processing can be time-consuming and resource-intensive given the need to handle missing values, outliers, inconsistencies, and other data quality issues [52]. Moreover, pre-processing decisions can have significant implications for AI model performance, as they influence the characteristics of the input data [53].

Data Quality Management: Effective data management is essential for maintaining data quality and ensuring that AI systems can access and process data efficiently [41]. Data management challenges include maintaining data storage and retrieval systems, implementing version control, and ensuring data security and privacy [46].

2.2.1.3. The Role of Data Governance in Ensuring Data Quality

Data governance plays a critical role in maintaining, ensuring, and enhancing data quality in organizations [54]. It encompasses the processes, policies, standards, and technologies that manage the availability, usability, integrity, and security of data [55]. Effective data governance helps organizations make better decisions, optimize operations, comply with regulations, and create a competitive advantage [56]. Implementing a comprehensive data governance framework is essential for addressing data quality challenges in AI [57].

Data governance includes several aspects such as data stewardship, data quality management, data privacy and security, and data architecture [58]. Data stewardship involves assigning responsibility and accountability for data quality to designated data stewards to ensure that data meets organizational standards [59]. Data quality management refers to the processes and tools used to measure, monitor, and improve data quality, such as data profiling, cleansing, and enrichment [60]. Data privacy and security are concerned with protecting sensitive information and ensuring compliance with relevant regulations [61]. The data architecture includes the design, organization, and management of data structures, storage systems, and data integration technologies [57].

Implementing an effective data governance framework requires a clear understanding of the organization's goals, data quality requirements, and existing data management practices [62]. Organizations must establish data quality metrics, set data quality targets, and monitor data quality performance regularly [63]. In addition, organizations should invest in data governance technologies such as data catalogs, data lineage tools, and data quality management systems to support the data governance process [54]. By adopting a robust data governance framework, organizations can significantly improve data quality and unleash the full potential of AI.

Furthermore, effective data governance is essential for fostering a data-driven culture within an organization. By promoting collaboration and communication between different departments and stakeholders, data governance helps break down data silos and facilitates the sharing of data assets [54]. This enables organizations to leverage their data more effectively and gain valuable insights into strategic decision-making [58].

Training and education are critical components of data governance [60]. Ensuring that employees have a solid understanding of data quality concepts, tools, and best practices helps create a shared vision and commitment to maintaining high-quality data. This can lead to more accurate and reliable AI models that drive innovation and create a competitive advantage [59].

In addition to internal data governance efforts, organizations should consider the importance of external data quality. As AI systems often rely on data from various sources, including third-party providers and public datasets, ensuring the quality of external data is crucial for the success of AI initiatives [57]. Collaborating with data providers and establishing data quality agreements can help mitigate potential data quality issues stemming from external sources [58].

In summary, implementing a comprehensive data governance framework is vital for addressing data quality challenges in AI. Organizations that prioritize data governance can enhance their decision making, optimize operations, and unlock the full potential of AI technologies. By fostering a data-driven culture, investing in data governance technologies, and ensuring both internal and external data quality, organizations can build a solid foundation for AI success.

2.2.1.4. Addressing Data Quality Challenges: Techniques and Strategies

This section proposes solutions to address the challenges in data quality. The results are summarized in Figure 7.

Proposed Solution for Data Quality challenges



Figure 7. Proposed solution to data quality challenges.

Data Cleaning: Data cleaning is an essential step in improving data quality for AI. This involves identifying and correcting errors, missing values, inconsistencies, and outliers in data. Techniques such as data profiling and validation can help identify areas of data that require cleaning.

Data Profiling and Data Preparation: Data profiling involves analyzing datasets to identify data quality issues such as missing data, duplicate records, and inconsistent values. Data preparation involves cleaning and transforming the raw data into a usable format for AI algorithms. These processes are essential to ensure that the data used to train the AI models are accurate, complete, and consistent.

Data Labeling: Data labeling involves tagging data with relevant metadata that describe its characteristics, which can help ensure that AI models are trained with high-quality data. For example, in image recognition, data labeling may involve identifying objects in the images and adding descriptive labels to the data.

Imputation Techniques for Missing Data: Missing data is a significant challenge in ensuring the data quality of AI systems. Various imputation techniques have been proposed to handle missing data, including mean imputation, regression imputation, and multiple imputation [54]. Advanced techniques, such as matrix completion methods, have been explored in recent years [55]. These

methods aim to provide reasonable estimates of the missing values and ensure the completeness of the dataset.

Feature Selection and Engineering: Feature selection and engineering play a crucial role in addressing data quality challenges as they involve identifying relevant features and transforming raw data into a format suitable for analysis [64]. Techniques such as Recursive Feature Elimination (RFE), LASSO, and principal component analysis can be employed to reduce the dimensionality of the data and eliminate noise [65]. Moreover, domain knowledge can be leveraged to create new features that better capture the underlying patterns and relationships.

Data Augmentation: Data augmentation techniques can be used to address the data quality challenges related to limited or unbalanced datasets. These techniques generate synthetic data samples by applying various transformations, such as rotation, scaling, and flipping, to the original data [66]. Data augmentation has been particularly successful in improving the performance of deep learning models in computer vision and natural language processing tasks [67].

Active Learning: Active learning is an approach that can help address data-quality challenges by guiding the data-collection process. In active learning, AI models iteratively select the most informative samples to be labeled and added to the training set, thereby reducing the amount of required labeled data and improving model performance [68]. Active learning has shown promise in various applications, including text classification and object recognition [69,70].

Data Validation and Testing: Data validation involves checking data for accuracy and completeness, whereas testing involves assessing the performance of AI models using various metrics. These processes can help identify and address data quality issues that may affect the accuracy and effectiveness of AI models.

Algorithmic Fairness: Algorithmic fairness involves ensuring that AI models are not biased towards specific groups or individuals. This can be achieved by carefully selecting training datasets and implementing algorithms designed to reduce bias.

Data Bias Mitigation: Data bias can lead to inaccurate or unfair AI predictions and decisions. Mitigating data bias involves identifying and addressing bias in the training data through techniques such as dataset balancing, which involves adjusting the distribution of data to reduce bias.

Continuous Monitoring and Maintenance: AI models need to be continuously monitored and maintained to ensure that they remain accurate and effective. This involves ongoing data-quality checks, updating models with new data, and retraining models as needed.

Data Lineage: Data lineage involves tracking the history of the data and ensuring that it is used appropriately. This can help prevent issues such as data drift, where the quality of data changes over time.

2.2.2. Dimension II: Data Volume

2.2.2.1. Challenging elements of data volume

The data-volume challenge is a key aspect of AI research and application. Large datasets are critical for training AI models and continue to grow in size and complexity. Large datasets are critical for training AI models, and their size and complexity continue to increase exponentially. This growth brings with it some challenges that must be addressed to ensure the effective use of AI in various domains [71]. The challenging elements of the data volume are presented in Figure 8.

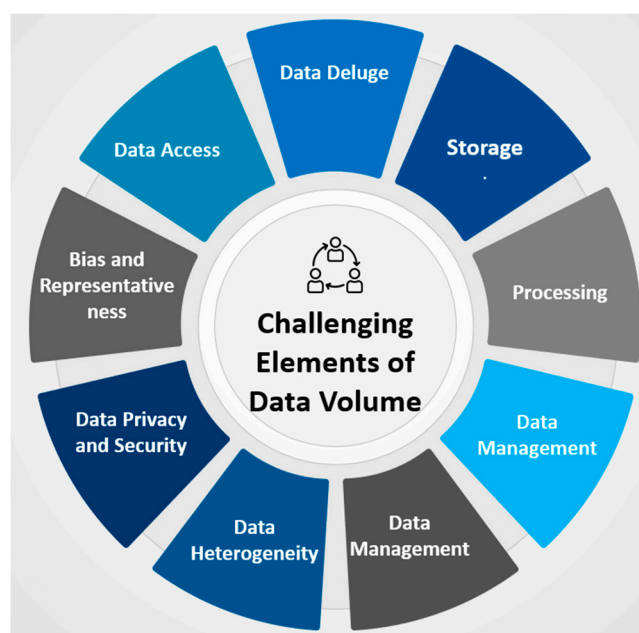


Figure 8. Challenging elements of data volume.

Data Deluge: A double-edged sword: Exponential data growth is the driving force behind the success of AI, particularly in deep learning techniques [72]. However, this massive amount of data poses several challenges, including storing, processing, and managing data [73].

Storage Challenges: The huge amount of data generated today requires more efficient storage solutions to support artificial intelligence applications [74]. Traditional storage architectures may not be able to meet the scalability, performance, and cost requirements of AI workloads [75]. New storage technologies, such as nonvolatile memory (NVM) and distributed storage systems, have been proposed as possible solutions [76].

Processing Challenges: AI models, particularly deep learning algorithms, require enormous computing resources to process large datasets [77]. This has led to an increased need for specialized hardware such as GPUs and TPUs to accelerate AI training and inference [78]. In addition, new techniques, such as model compression, pruning, and quantization have been explored to optimize AI models for more efficient processing [79].

Data Management Challenges: From perspective of big data volume, effective data management is critical for AI systems to handle massive amounts of data. This includes data cleaning, preprocessing, labeling, and curation [80]. Techniques such as active learning, weak supervision, and transfer learning have been proposed to alleviate the burden of manual data annotation [81].

Data Heterogeneity: Large datasets may contain data from multiple sources, which can be challenging to integrate and harmonize, particularly when the data are in different formats or structures.

Data Privacy and Security: Large data volumes can increase the risk of data breaches and privacy violations, particularly when sensitive data is involved. These issues need to be addressed as the amount of data increases [82].

Bias and Representativeness: Large volumes of data do not necessarily guarantee representativeness or lack bias, as they may still contain demographic, cultural, or other biases that can impact the accuracy of AI models.

Data Access: In some cases, organizations may have access to large datasets but may not be able to use them due to legal or regulatory constraints. Organizations must ensure that they have the necessary permissions and licenses to access and use data.

2.2.2.2. Mitigation Solutions for challenges of data volume

Several solutions have been proposed to address the data volume challenges in artificial intelligence, Figure 9 depicts the proposed solutions that includes:

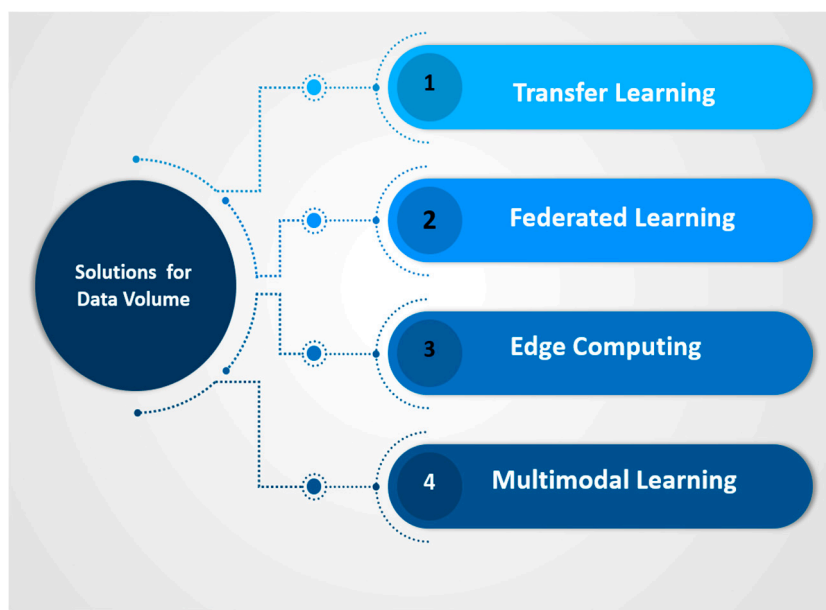


Figure 9. Proposed solutions for challenges of data volume.

- a) **Transfer Learning:** Transfer learning involves leveraging pretrained AI models to improve the performance of new models. Using pre-trained models, organizations can reduce the amount of training data required and improve the efficiency of the training process.
- b) **Federated Learning:** It enables collaborative model training across multiple devices without sharing raw data [83].
- c) **Edge Computing:** It brings data processing closer to the data source, thereby reducing the network latency and bandwidth usage [84].
- d) **Multimodal Learning** leverages multiple data sources to improve the model performance and reduce reliance on large datasets [85].

Federated Learning as proposed solution:

Federated learning was introduced as a solution to preserve user privacy while benefiting from the collective knowledge of multiple data sources [86]. In this framework, devices (also known as clients) train machine learning models using local data, and then share model updates with a central server. The server aggregates these updates, improves the global model, and distributes the updated model back to clients. This process is repeated iteratively until the model converges. The benefits of Federated Learning include several benefits that can help address data volume challenges in AI.

- a) **Privacy:** Because raw data remain on client devices, federated learning inherently provides a higher level of privacy compared to centralized approaches [83].
- b) **Reduced Data Transfer:** By sharing only model updates rather than raw data, federated learning can significantly reduce the amount of data that needs to be transferred over the network, reducing bandwidth and latency issues [87].
- c) **Scalability:** Federated learning can accommodate a large number of client devices, allowing the use of different data sources without overloading the central server [75].
- d) **Real-time learning:** By allowing clients to learn from local data, the federated learning enables real-time adaptation and improves model performance [82].

Challenges and Future directions: Despite its benefits, federated learning also presents some challenges that need to be addressed.

- a) **Heterogeneity:** The heterogeneity of client devices and data distribution may lead to an unbalanced contribution to the global model, which may affect convergence and model performance [75].
- b) **Communication Overhead:** The iterative process of exchanging model updates incurs significant communication overhead and may negate the benefits of reduced data transfer [87].
- c) **Security:** Federated learning is vulnerable to various security threats, including model poisoning, inference attacks, and Sybil attacks [88].

To overcome these challenges, researchers are exploring various techniques, such as weighted averaging to deal with heterogeneity [89], communication-efficient algorithms for reducing overhead [90], and differential privacy for enhancing security [91]. Continued research in these areas will be crucial to fully realize the potential of federated learning in addressing the data-volume challenge in AI.

- **Edge Computing as a Proposed Solution:** Edge computing is a distributed computing paradigm that aims to bring computation and data storage closer to the data source, or the "edge" of the network, where the data is generated [84]. By performing data processing on edge devices, such as smartphones, IoT devices, or edge servers, edge computing can reduce the amount of data that must be transmitted to the cloud or a centralized data center. This approach enables real-time data processing, reduces latency, and conserves the bandwidth.
- **Advantage of Edge Computing:** Edge computing offers several benefits that can help address the data-volume challenge in AI.
 - a) **Reduced Latency:** By processing data closer to the source, edge computing can significantly reduce latency and enable real-time AI applications [84].
 - b) **Bandwidth Efficiency:** Edge computing helps conserve bandwidth by reducing the amount of data transmitted over the network, which is particularly useful in situations where the network bandwidth is limited or expensive [76].
 - c) **Enhanced privacy and security:** Because data are processed and stored locally, edge computing can provide improved data privacy and security compared to centralized approaches [92].
 - d) **Scalability:** Edge computing can support many devices and applications, making it suitable for the growing demands of AI and IoT [84].

Challenges and Future Directions:

Despite its advantages, edge computing presents some challenges that need to be addressed.

- a) **Resource Constraints:** Edge devices typically have limited computational resources, which may hinder the performance of complex AI models [93].
- b) **Model Deployment and Management:** Deploying and managing AI models across a large number of edge devices can be challenging because it requires efficient model distribution, updates, and monitoring [94].
- c) **Heterogeneity:** The heterogeneity of edge devices in terms of hardware, software, and network connectivity can pose challenges for implementing consistent and efficient AI solutions [95].

Researchers have explored various techniques to overcome these challenges, such as model compression and hardware-aware neural architecture search for resource-constrained devices [79], edge-cloud collaborative learning for model deployment and management [96], and federated edge learning to address heterogeneity [97]. Continued research in these areas will be crucial to fully realize the potential of edge computing in addressing the data-volume challenge in AI.

2.2.3. Dimension III: Data Privacy and Security

The use of personal or sensitive data in AI can raise concerns regarding privacy and security. It is important to ensure that the data are stored and processed securely and that privacy regulations are followed. To address this challenge, businesses should implement data privacy and security policies and procedures such as data encryption and access control.

This section provides a comprehensive review of the challenges associated with data privacy and security in AI, and discusses data collection and sharing, inference attacks, differential privacy, adversarial attacks, data poisoning, and model and data tampering. It also presents state-of-the-art mitigation strategies, such as privacy-preserving AI techniques, robustness and adversarial training, monitoring and anomaly detection, and compliance with data-protection regulations. The results are summarized in Figure 10.

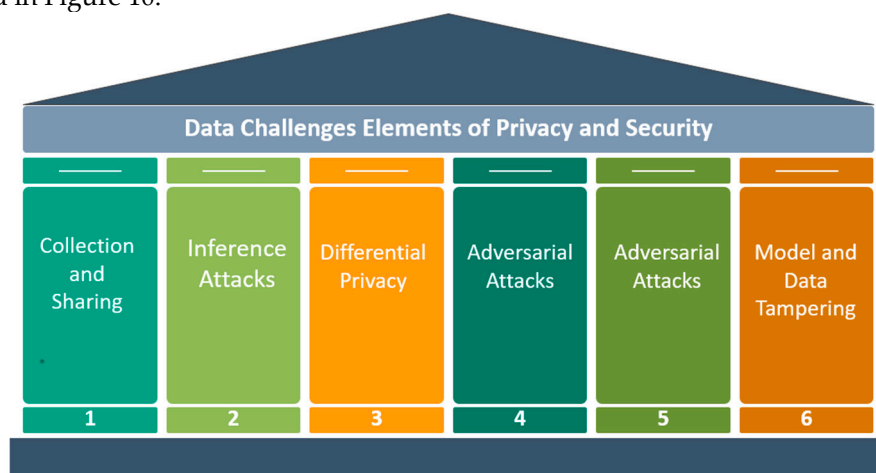


Figure 10. Challenges elements of data privacy and security.

2.2.3.1. Data Privacy Challenges in AI

Data Collection and Sharing: AI systems require large quantities of data to train effectively, often leading organizations to aggregate data from various sources, potentially exposing sensitive user information [98]. Data-sharing agreements and collaborative data analysis projects can exacerbate these concerns, especially when data are shared across international borders with differing privacy regulations [99].

Inference Attacks: AI models can inadvertently reveal sensitive information about training data, even when the data are anonymized [100]. For example, attackers can use model inversion or membership inference attacks to extract private information from a model's predictions or to learn whether a specific data point is included in the training set [101].

Differential Privacy: Differential privacy (DP) is a popular approach for preserving privacy during data analysis by adding controlled noise to the data [102]. Although DP provides strong privacy guarantees, it can be challenging to implement in practice, especially when balancing privacy protection and model utility [103].

2.2.3.2. Data Security Challenges in AI

Adversarial Attacks: Adversarial attacks, in which small perturbations are introduced to input data to deceive AI models, pose significant security risks [104]. These attacks can lead to incorrect predictions or classifications, undermining the reliability of AI systems in critical applications, such as healthcare, finance, and autonomous vehicles [105].

Data Poisoning: Data poisoning attacks involve tampering with training data to degrade the performance of an AI model [106]. These attacks can be difficult to detect because they often target a small subset of the training data and require minimal modifications to the poisoned data points [107].

Model and Data Tampering: Attackers can also target AI models and data directly by altering model parameters, weights, or the data itself [108]. Techniques such as backdoor attacks or Trojan

neural networks can introduce hidden malicious behavior into AI models, thereby posing significant security risks [109]. In backdoor attacks, an attacker injects malicious code into the model during the training process, causing the model to produce incorrect outputs or exhibit unintended behavior when triggered by a specific input pattern [110]. Trojan neural networks, on the other hand, involve embedding a hidden trigger within the model, which, when activated by a specific input, causes the model to perform unauthorized actions or provide incorrect predictions [111].

Model extraction attacks involve model tampering, in which an attacker seeks to replicate a target model by querying it and learning from the responses [112]. This can lead to intellectual property theft, or even the creation of duplicate models with malicious intent [113].

To defend against model and data tampering, organizations can employ various strategies such as model hardening, secure model storage, and input validation. Model hardening techniques, such as fine pruning, can help remove malicious components from a model while maintaining its overall performance [114]. Secure model storage using encryption and access controls can protect a model from unauthorized modifications [115]. Input validation can be employed to ensure that only legitimate inputs are processed by the AI system, mitigating the risk of triggering hidden backdoors or Trojan networks [116]. By implementing these countermeasures, organizations can enhance the security and trustworthiness of their AI systems in the face of model and data-tampering threats.

2.2.3.3. Mitigation Strategies for Data Privacy and Security Challenges

Several mitigation techniques are proposed as depicted in Figure 11.

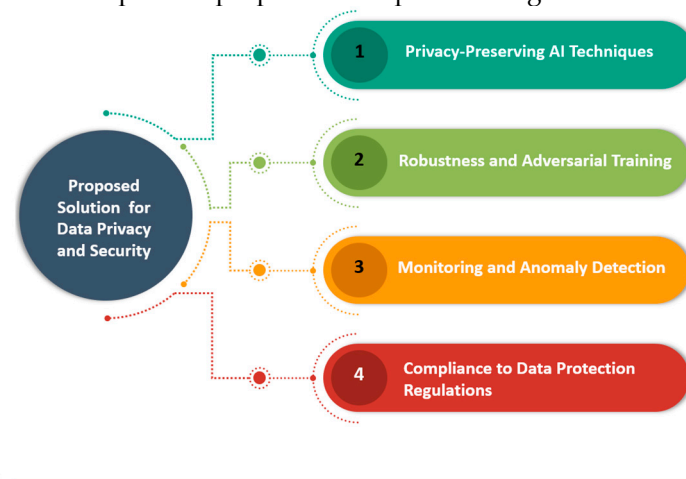


Figure 11. Mitigation strategies for data privacy and security.

Privacy-Preserving AI Techniques: To address privacy concerns, organizations can employ privacy-preserving AI techniques, such as federated learning, secure multi-party computation, and homomorphic encryption [117]. These methods allow organizations to train AI models on distributed data without sharing raw data between parties, thereby reducing the risk of data breach or leakage [118].

Robustness and Adversarial Training: To defend against adversarial attacks and improve model robustness, researchers have developed adversarial training techniques that involve augmenting a training dataset with adversarial examples [119]. By training the model on a combination of clean and adversarial data, it becomes more resilient to adversarial perturbations [120].

Monitoring and Anomaly Detection: To detect data poisoning and model tampering, organizations can employ monitoring and anomaly detection techniques to identify deviations from the expected behavior in model performance, training data, or model parameters [59]. Early detection can help prevent further damage to AI systems and provide valuable insights for improving security measures [121].

Compliance with Data Protection Regulations: Organizations should adhere to data protection regulations, such as the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States, to ensure that they collect, store, and process

data in a secure and compliant manner [122]. Compliance with these regulations can help minimize the risk of data breaches and protect user privacy [123].

Data privacy and security challenges in AI are significant concerns for organizations that develop and deploy AI systems. By understanding these challenges and implementing mitigation strategies, such as privacy-preserving AI techniques, robustness training, and compliance with data protection regulations, organizations can enhance the privacy and security of their AI systems. As AI continues to evolve and impact various industries, it is crucial for researchers, practitioners, and policymakers to work together to address these challenges and ensure that it serves the greater good without compromising user privacy and security.

2.2.4. Dimension IV: Bias and Fairness

AI has seen rapid advancements over the past decade, transforming many aspects of our lives. However, as AI systems have become more prevalent, concerns regarding data bias and fairness have emerged. The performance of AI models depends heavily on the quality of the data used for training, and biased data can lead to biased outcomes [124]. This section provides a comprehensive review of the challenges related to data bias and fairness in AI with a focus on recent research and solutions.

Bias, in the context of artificial intelligence and machine learning, refers to systematic errors in the algorithms' predictions or decisions resulting from skewed training data, flawed algorithms, or the influence of pre-existing assumptions. These biases can lead to unfair or discriminatory outcomes, impacting individuals or groups based on attributes, such as race, gender, age, or socioeconomic status. Addressing and mitigating bias in AI systems is crucial for ensuring the fair and ethical deployment of these technologies in various domains from healthcare and finance to criminal justice and social media.

Amazon's AI Recruiting Tool: In 2018, Amazon discontinued its AI recruiting tool after it was found to be biased against female candidates. The system was designed to review resumes and rank candidates based on their qualifications. However, the model was trained on resumes submitted to the company over a ten-year period, and predominantly belonged to male candidates. Consequently, the AI system preferred male candidates over equally qualified female candidates [125].

Google Photos' Racial Bias: In 2015, Google Photos was criticized for its image recognition algorithm, which mistakenly labeled African Americans as gorillas. This incident highlighted racial bias in the AI system, which was attributed to the lack of diversity in the training data. Google apologized for this mistake and worked on improving the algorithm to avoid such issues in the future [126].

Microsoft's Tay Chatbot: In 2016, Microsoft launched a Twitter-based AI chatbot called Tay. The chatbot was designed to learn from user interactions and to mimic human conversations. However, within 24 h of its launch, Tay started posting offensive and racist tweets, as it had learned from malicious users who intentionally fed biased and inappropriate content. Microsoft quickly took Tay offline and apologized for the incident [127].

Apple Card's Gender Bias: In 2019, Apple faced a backlash when its Apple Card, a credit card service powered by Goldman Sachs, was accused of gender bias. Several users reported that the credit limit offered to male applicants was significantly higher than that offered to their female counterparts despite having similar or even worse financial profiles. This issue raised concerns about the fairness and transparency of AI algorithms used for credit assessment [128].

COMPAS Risk Assessment Tool: The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool is an AI-based system used in the United States to assess the risk of recidivism in criminal defendants. A 2016 investigation by ProPublica revealed that the tool exhibited racial bias, with African American defendants being more likely to be incorrectly labeled as high-risk compared to white defendants with similar criminal records. This controversy led to increased scrutiny of AI-based risk assessment tools in the criminal justice system [129].

2.2.4.1. Types of Data Bias

Figure 12 represent the ten types of data bias which include:

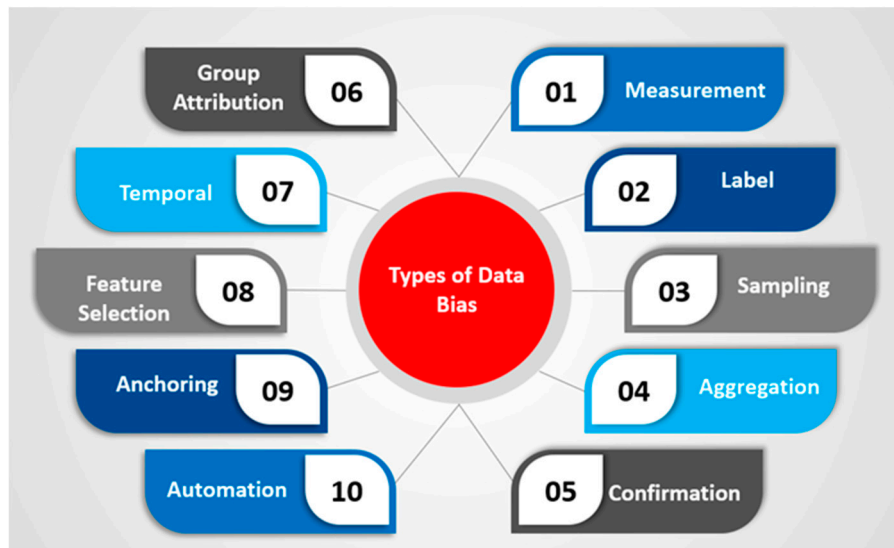


Figure 12. Types of data bias.

Measurement Bias: Measurement bias occurs when data collection methods systematically over- or under-represent certain features or aspects of data [130]. This can lead to AI models that generate biased predictions that are not representative of the true population.

Label Bias: Label bias arises when labels assigned to data instances are incorrect or unrepresentative of true outcomes. This can result from human errors, subjective judgments, or systemic issues during the labeling process [131].

Sampling Bias: Sampling bias occurred when the collected data were not representative of the population of interest. This can lead to biased AI models as they learn from a non-representative sample [132].

Aggregation Bias: Aggregation bias emerges when data are combined from multiple sources with different characteristics or distributions. This can cause AI models to learn patterns that are not generalizable to the entire population [133].

Confirmation Bias: Confirmation bias emerges when data or information are selectively chosen or weighted to support pre-existing beliefs or expectations. This can inadvertently affect AI model outcomes because the training data may disproportionately represent certain aspects or patterns [134].

Group attribution bias: This bias arises when AI systems generalize, or stereotype individual behaviors based on the perceived characteristics of the group to which they belong. This can lead to biased predictions that do not accurately represent an individual's unique attributes [135].

Temporal Bias: Temporal bias occurs when AI models are trained on historical data that no longer reflects current trends or patterns. This can lead to biased predictions because the models fail to adapt to changes in the underlying data distribution over time [136].

Feature Selection Bias: Feature selection bias emerges when certain features are given greater importance or focus during the model development process, leading to biased outcomes. This can be a result of domain-specific biases or biases inherent to the algorithms used for feature selection [137].

Anchoring Bias: Anchoring bias occurs when AI models rely heavily on initial information or data points to make predictions. This can result in biased outcomes, as the models may not sufficiently consider other relevant factors or adjust their predictions based on new information [138].

Automation Bias: Automation bias refers to the tendency of humans to over rely on AI system outputs, even when they are flawed or biased. This can exacerbate the consequences of biased AI models because users may not question or scrutinize biased decisions or recommendations generated by these systems [139].

2.2.4.2. Consequences of Data Bias and Unfairness in AI

Discrimination: Biased AI systems can inadvertently discriminate against certain groups or individuals, thereby leading to unfair treatment. For example, biased facial recognition systems can misidentify individuals from minority groups at a higher rate than those from majority groups [140].

Misinformation: AI systems trained on biased data may propagate false or misleading information, exacerbating existing stereotypes and prejudices [141].

Legal and Ethical Implications: Biased AI systems can pose legal and ethical challenges as they may violate anti-discrimination laws or ethical guidelines [142].

2.2.4.3. Addressing Data Bias and Fairness in AI

This section proposes a solution to address data bias and fairness in AI. These results are shown in Figure 13.

Data Collection and Pre-processing: Collecting diverse and representative data is crucial for mitigating data bias [143]. Additionally, preprocessing techniques, such as resampling, reweighting, or data augmentation, can help reduce the bias in the dataset [144].

Algorithmic Fairness: Researchers have proposed various fairness-aware machine-learning algorithms that aim to minimize discriminatory outcomes (Friedler et al., 2019). These techniques typically incorporate fairness constraints into the model training process or post-process model predictions to ensure fairness [145].

Fairness Metrics: Developing appropriate fairness metrics is essential for quantifying and comparing the performance of AI models in terms of fairness [146]. Some commonly used metrics include demographic parity, equalized odds, and disparate impact ratios [147].

Explainable AI: Explainable AI (XAI) techniques can provide insights into the decision-making process of AI models, helping identify potential sources of bias and unfairness [148]. Researchers can develop more effective interventions to improve fairness by understanding the underlying reasons for the biased outcomes.

Interdisciplinary Collaboration: Addressing data bias and fairness requires the collaboration of experts from various fields including computer science, social sciences, and ethics [149]. Interdisciplinary efforts can help develop comprehensive strategies that consider the complex interplay between data, algorithms, and social contexts.

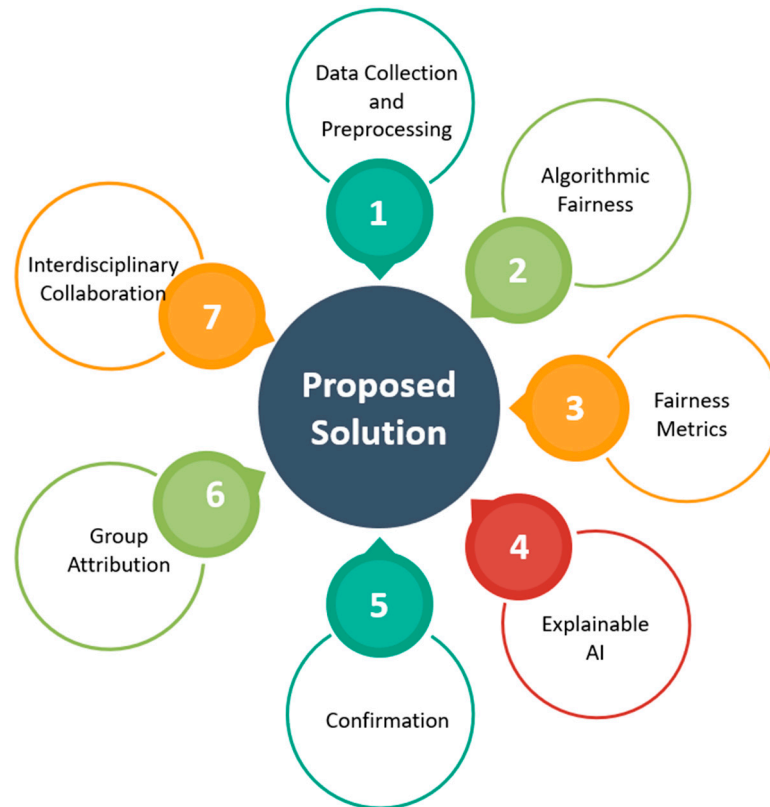


Figure 13. Proposed solutions to data bias and fairness challenge.

2.2.5. Dimension V: Interpretability and Explainability

AI models can be difficult to interpret and explain, which can make it difficult for organizations to understand how decisions are made. It is important to ensure that the AI models are transparent and explainable. To address this challenge, organizations should implement interpretability and interpretability controls such as feature importance analysis and model visualization tools. Figure 14 shows the challenging elements of interpretability and explainability.

Artificial intelligence has become an integral part of modern society, with its influence seen in various domains, including healthcare, finance, transportation, and many others [150]. The rise of AI has been largely driven by advances in machine learning and deep learning techniques, which have demonstrated impressive results in solving complex problems [151]. However, these techniques have also given rise to "black-box" models, characterized by their lack of interpretability and explainability [152].

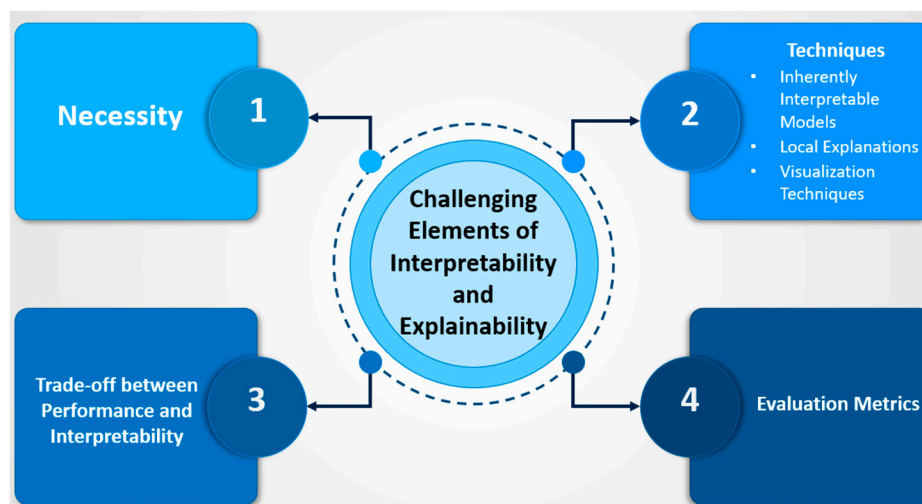


Figure 14. Challenging elements of Interpretability and explainability.

2.2.5.1. The Necessity of Interpretability and Explainability

The demand for interpretability and explainability in AI systems is driven by the need for trust, accountability, and ethical considerations [153]. Trust is essential for the adoption and successful integration of AI systems, as users need to understand and believe in the decisions made by these systems [154]. Accountability ensures that AI systems comply with legal and ethical standards and can be audited when necessary [155]. Ethical considerations call for AI systems to adhere to the principles of fairness, transparency, and nondiscrimination [156].

2.2.5.2. Current Techniques for Interpretability and Explainability

Various techniques have been proposed to enhance the interpretability and explainability of AI systems, ranging from inherently interpretable models to post-hoc explanations for black-box models [157]. Some of these techniques include the following:

Inherently interpretable models, such as decision trees, linear regression, and rule-based systems, are designed to be easily understood by humans, providing a direct relationship between input features and the model's output [158].

Local Explanations: Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive explanations (SHAP) are methods that explain individual predictions by approximating a complex model's behavior using simpler, interpretable models for a specific instance [159,161].

Visualization Techniques: Techniques such as t-distributed Stochastic Neighbor Embedding (t-SNE) and activation can provide visual representations of high-dimensional data processed by AI models, enabling human users to understand the relationships and patterns present in the data [162,163].

2.2.5.3. Remaining Challenges and Future Directions

Despite progress in developing techniques for interpretability and explainability, several challenges still need to be [150]:

Trade-off between Performance and Interpretability: Highly interpretable models often come at the cost of reduced predictive performance. Future research should focus on developing models that balance interpretability and performance [158].

Evaluation Metrics: The development of standardized evaluation metrics to assess the interpretability and explainability of AI models remains a challenge. Establishing universally accepted metrics will enable researchers to compare different techniques more effectively and drive further innovation [163].

Domain-specific Solutions: Certain application domains require specific interpretability and explainability techniques. For example, in the medical field, explanations must be tailored to the knowledge and understanding of both clinicians and patients [164]. Further research is required to develop domain-specific solutions that satisfy unique requirements.

Ethical Considerations: As explainable AI techniques become more advanced, there is a risk of generating explanations that may be misleading or biased, leading to potentially harmful consequences [165]. Future research should address the ethical implications of explainability and develop guidelines to ensure that the explanations are accurate, unbiased, and useful.

Thus, Interpretability and explainability are crucial components of the successful integration of AI systems into our daily lives. By addressing the challenges related to these concepts, trust can be fostered, accountability ensured, and ethical AI deployment promoted. The development of techniques to enhance interpretability and explainability remains an active area of research, with significant progress already being achieved. However, several challenges still need to be addressed, including balancing performance with interpretability, developing standardized evaluation metrics, creating domain-specific solutions, and considering ethical implications. By addressing these

challenges, we can bridge the gap between AI and human understanding, paving the way for a more transparent and trustworthy AI-powered future.

2.2.6. Dimension VI: Technical Expertise

Building and deploying AI models requires technical expertise, which can be challenging for companies that do not have the necessary skills in-house. To address this challenge, organizations can hire data scientists or partners with external vendors who provide the required expertise.

The growth of artificial intelligence (AI) has revolutionized multiple industries, including healthcare, finance, and manufacturing [166]. AI-driven systems have demonstrated exceptional performance in various tasks such as natural language processing, computer vision, and robotics [167]. However, the increasing complexity and sophistication of AI algorithms have resulted in new challenges in terms of technical expertise [168]. This section investigates these challenges and proposes potential solutions.

2.2.6.1. Scarcity of Skilled Professionals

One of the primary challenges in the AI domain is the scarcity of skilled professionals [169]. The rapid development of AI technologies has outpaced the growth of a workforce capable of handling the complexity and diversity of AI systems [170]. This talent gap can hinder further progress in AI and impede the adoption of AI technologies in various [168]. Potential solutions to address this talent gap include increasing investment in education and training, promoting interdisciplinary collaboration, and developing AI-driven tools for education [170].

2.2.6.2. Ethical Concerns

AI technologies have raised various ethical concerns that require careful consideration and technical expertise [171]. These concerns include bias, fairness, transparency, accountability, and potential misuse of AI technologies [172]. Addressing these ethical issues necessitates the development of robust AI systems that align with human values and ethical principles, as well as fostering interdisciplinary collaboration between AI researchers, ethicists, and policymakers [171] [173].

2.2.6.3. The Growing Demand for AI-Related Expertise

The increasing adoption of AI technologies across various industries has led to a surge in the demand for AI-related expertise [174]. According to the [175], AI and machine learning are among the top 10 emerging professions, with a projected growth rate of 41% between 2020 and 2025. This increasing demand for AI professionals is driven by the need for specialized knowledge in areas such as algorithm development, data analysis, and system integration [176]. Addressing this demand requires concerted efforts in education and training as well as fostering interdisciplinary collaboration to develop a workforce capable of tackling the complex challenges associated with AI technologies [170].

2.2.6.4. Key Disciplines in High Demand for AI-Related Expertise

As AI technologies continue to evolve, demand for expertise in various disciplines is expected to grow. Some of the key disciplines that are needed in the AI domain are:

- a) **Computer Science and Computer Engineering:** Professionals with skills in algorithm development, machine learning, deep learning, natural language processing, and computer vision are essential for designing, building, and maintaining AI systems [176].
- b) **Data Science and Analytics:** AI systems often rely on large volumes of data. Experts in data science and analytics are required to preprocess, analyze, and interpret data to generate actionable insights and improve AI models [174].

- c) Human–computer Interaction (HCI) and Cognitive Science: As AI technologies become more integrated into our daily lives, understanding how humans interact with these systems is becoming increasingly important. HCI and cognitive science experts can help design AI systems that are intuitive, user-friendly, and adaptable to human needs [177].
- d) Ethics, Philosophy, and Policy: The growing influence of AI technology raises several ethical and philosophical questions. Experts in these fields are needed to address issues related to fairness, transparency, and accountability and to develop policies and frameworks that ensure responsible AI development and deployment [171].
- e) Cybersecurity and Privacy: Protecting sensitive data and maintaining the security of AI systems is a critical concern. Professionals skilled in cryptography, secure multiparty computation, and privacy-preserving machine learning techniques are essential to ensure data privacy and security [178].
- f) Robotics and Autonomous Systems: As AI-powered robotics and autonomous systems become more prevalent, expertise in areas such as control systems, sensor fusion, and robotics software engineering will become increasingly valuable [179].

2.2.6.4. Collaboration between Humans and AI

AI systems are becoming increasingly more autonomous. As a result, there is a growing need for effective collaboration between humans and AI [179]. This collaboration requires developing AI systems that can understand and adapt to human preferences, communicate effectively, and support human decision making [179]. Technical expertise in human-computer interaction, cognitive science, and explainable AI is crucial for designing AI systems that can seamlessly integrate into human workflows [177].

The references used in this study to address the challenges of data in AI applications are listed in Table 1.

Table 1. References to challenges of data in AI.

Challenge	Reference	
Data Quality	Introduction	[37–39]
	Quality Measures	[37–47]
	Collection and Management	[41,46,49–53]
	Data Governance	[54–63]
	Proposed solutions	[54,55,64–70]
Data Volume	Introduction	[71]
	Data Deluge	[72,73]
	Storage Challenges	[74–76]
	Processing Challenges	[77–79]
	Data Management Challenges	[80,81]
	Data Privacy and Security	[82]
	Proposed Solutions	[75,76,79,82–97]
Data Privacy and Security	Data Privacy	[98–103]
	Data Security	[104–116]
	Mitigation Strategies	[117–123]
Bias and Fairness	Introduction	[124–129]
	Types of Data Bias	[130–139]

	Consequences of Data Bias and Unfairness	[140–142]
	Proposed Solutions	[143–149]
Interpretability and Explainability	Introduction	[150–152]
	The Necessity	[153–156]
	Current Techniques	[157–163]
	Remaining Challenges and Future Directions	[150,158,163–165]
Technical Expertise	Introduction	[166–168]
	Scarcity of Skilled Professionals	[169,170]
	Ethical Concerns	[171–73]
	The Growing Demand for AI-Related Expertise	[170,174–176]
	Key Disciplines in High Demand for AI-Related Expertise	[171,174,180–184]
	Collaboration between Humans and AI	[177,179,185–188]

3. Results

In our quest to understand the hurdles to data quality in terms to artificial intelligence, we stumbled upon a few noteworthy insights. Specifically, we noticed that data quality plays a vital role in AI systems, which span multiple dimensions with significant implications. Therefore, ensuring data quality requires not only appropriate governance but also adherence to best practices to ensure optimal results.

3.1. Dimension of Data Quality and Implication for AI systems

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

3.2. 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, which can 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 our 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 data, it is important to keep a close eye on its quality and maintain control.
- c) Quality data and holding ourselves accountable should be a culture we strive to create.
- 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 among different systems. The integration of various data types can be achieved using appropriate methods.
- e) Regulations and laws must be followed carefully to remain compliant.

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.

3.3. 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's 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 monitor and update 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 we 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 it. 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 a 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 are 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 our analysis. Data quality plays a critical role and should be a top priority for organizations that utilize AI.

4. Discussion

The comprehensive analysis of the data quality challenges for artificial intelligence presented in this article underscores the importance of understanding and addressing data quality issues in the development and deployment of AI systems. In this discussion section, we emphasize the broader

implications of our findings, highlight the limitations of the current study, and suggest future research directions.

4.1. Broader Implications

Our analysis has several broader implications for organizations, policymakers, and researchers involved in AI development and deployment. First, by identifying and understanding the key dimensions of data quality and their implications for AI systems, organizations can prioritize their efforts to address data quality challenges, ensuring that their AI systems deliver accurate, reliable, and unbiased results. Second, our analysis highlights the significance of data governance in maintaining and improving data quality, emphasizing the need for organizations to invest in robust data governance structures and processes. Finally, the best practices identified in this study can serve as a practical guide for organizations seeking to enhance their data quality management efforts and optimize the performance and reliability of AI systems.

4.2. Limitations

Despite providing a comprehensive analysis of the data quality challenges for AI, this study had several limitations. First, the scope of our analysis primarily focuses on the dimensions of data quality, data governance, and best practices. Additional factors, such as organizational culture and technical infrastructure, could impact data quality and AI performance. Second, although we have drawn upon a wide range of literature sources, there may be other relevant publications that were not considered in this study. Finally, our findings are largely based on a synthesis of the existing literature, and future empirical research is needed to further validate and expand these findings.

4.3. Real-time environment challenges

Real-time data pose unique challenges in AI applications because of the need to process and analyze data in near real-time or with minimal delays. Some key challenges associated with real-time data in AI applications are as follows:

- a) **Real-time Data Volume and Velocity:** Real-time data often come in high volumes and at high velocities, requiring efficient processing and analysis techniques. AI systems must handle incoming data streams and make timely decisions or predictions based on these data.
- b) **Latency and Response Time:** Real-time applications demand low latency and fast response times. AI models must be designed and optimized to provide quick insights and actions in real-time scenarios. High computational requirements and complex algorithms can hinder real-time performance.
- c) **Real-time Data Quality and Noise:** Real-time data can be noisy, incomplete, or contain outliers. Ensuring data quality becomes challenging as there is limited time for data validation and cleaning. AI models must be robust to handle such noisy data and make accurate predictions or decisions.
- d) **Scalability and Resource Constraints:** Real-time AI applications often require scalability to handle large volumes of incoming data. Scaling AI models and infrastructure to handle the increased workload can be challenging considering resource constraints such as computing power, memory, and network bandwidth.
- e) **Data Synchronization:** Real-time data may come from multiple sources and must be synchronized for accurate analysis and decision-making. Aligning and integrating data streams from various sources in real time can be complex, particularly when dealing with data in different formats or time zones.
- f) **Real-time Model Training and Adaptation:** Updating or retraining AI models in real-time can be challenging. Continuous learning and model adaptation may be required to incorporate new data and adjust model parameters to changing conditions. Balancing the model stability with the need for real-time updates is crucial.

- g) **Real-time Analytics and Visualization:** Effectively analyzing and visualizing real-time data to extract meaningful insights and support decision making is a challenge. Real-time analytics techniques and interactive visualization tools are required to process and present data in a timely and actionable manner.

Addressing these challenges requires careful design, optimization, and integration of AI algorithms, infrastructure, and data-processing pipelines. It often involves a combination of efficient data streaming techniques, distributed computing, real-time analytics frameworks, and scalable AI architectures to enable real-time decision making and insights.

4.4. Future Research Directions

Based on the limitations and findings of this study, we suggest several directions for future research:

- a) Investigate the role of organizational culture, leadership, and technical infrastructure in ensuring data quality for AI systems.
- b) Conduct empirical research to assess the effectiveness of different data governance practices and data quality management strategies in real-world AI applications.
- c) Examine the relationship between specific dimensions of data quality and AI performance across different industries and use cases.
- d) Develop novel AI and machine learning techniques to automatically detect, diagnose, and resolve data quality issues.
- e) Explore the ethical and legal implications of data quality challenges in AI, particularly in relation to privacy, transparency, and fairness.

By addressing these future research directions, we can further deepen our understanding of the challenges of data quality for AI, develop more effective strategies to overcome these challenges, and harness the full potential of AI technologies.

5. Conclusions

This work address the challenges in data for AI technology and applications that businesses and organizations are recommended to develop strategies and frameworks to meet and handles these challenges in different dimensions such as:1) Data Quality that covers accuracy, completeness, consistency, timeliness, integrity, relevance, data collection, pre-processing, management, data governance, data labeling, etc, 2) Data Volume that covers data deluge, storage challenges, processing challenges, data management challenges, data heterogeneity,data privacy and security, bias and representativeness, data access, etc, 3) Data Privacy and security that cover inference attacks, differential privacy, adversarial attacks, data poisoning, model and data tampering, privacy-preserving AI techniques, robustness and adversarial training, monitoring and anomaly detection, compliance with data protection regulations, 4) Bias and Fairness that cover measurement bias, label bias, sampling bias, aggregation bias, confirmation bias, temporal bias, feature selection bias, etc, 5)Interpretability and Explainability that cover local explanations, visualization techniques, trade-off between performance and interpretability, evaluation metrics, domain-specific solutions, ethical considerations, 6) Technical Expertise that cover computer science and computer engineering, data science and analytics, human-computer Interaction, ethics, philosophy, and policy, cybersecurity and privacy, etc. Technical expertise in AI is essential to address the challenges posed by the rapid development of AI technologies. By fostering interdisciplinary collaboration; investing in education and training; and promoting the development of ethical, secure, and human-centered AI systems, researchers and policymakers can overcome these challenges and pave the way for further advancements in AI.

Author Contributions: Conceptualization, A. A., A.M.H and K,N,A.; writing—original draft preparation, A.A., A.M.H and K.N.A.; writing—review and editing, A.M.H and A.A.; supervision, K.N.A and A.M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson Education Limited.
2. Lavanya Sharma; Pradeep Kumar Garg, *Artificial Intelligence: Technologies, Applications, and Challenges* by Publisher: Taylor & Francis, 2021.
3. Aguiar-Pérez, Javier M., et al. "Understanding Machine Learning Concepts." *Encyclopedia of Data Science and Machine Learning*. IGI Global, 2023. 1007-1022.
4. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 1(1), 4171-4186.
5. Gumbs, Andrew A., et al. "The advances in computer vision that are enabling more autonomous actions in surgery: a systematic review of the literature." *Sensors* 22.13 (2022): 4918.
6. Enholm, Ida Merete, et al. "Artificial intelligence and business value: A literature review" *Information Systems Frontiers* 24.5 (2022): 1709-1734.
7. Wang, Zeyu, et al. "Business Innovation based on artificial intelligence and Blockchain technology." *Information Processing & Management* 59.1 (2022): 102759.
8. Dahiya, Neelam, Sheifali Gupta, and Sartajvir Singh. "A Review Paper on Machine Learning Applications, Advantages, and Techniques." *ECS Transactions* 107.1 (2022): 6137.
9. Marr, B. (2018). *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*. John Wiley & Sons.
10. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc.
11. Liu, Xiaofeng, et al. "Deep unsupervised domain adaptation: A review of recent advances and perspectives." *APSIPA Transactions on Signal and Information Processing* 11.1 (2022).
12. Li, Yuxi. "Deep reinforcement learning: An overview." *arXiv preprint arXiv:1701.07274* (2017).
13. Zhuang, Fuzhen, et al. "A comprehensive survey on transfer learning." *Proceedings of the IEEE* 109.1 (2020): 43-76.
14. Pouyanfar, Samira, et al. "A survey on deep learning: Algorithms, techniques, and applications." *ACM Computing Surveys (CSUR)* 51.5 (2018): 1-36.
15. Sun, X., Liu, Y., & Liu, J. (2018). Ensemble learning for multi-source remote sensing data classification based on different feature extraction methods. *IEEE Access*, 6, 50861-50869.
16. Zha, Daochen, et al. "Data-centric Artificial Intelligence: A Survey." *arXiv preprint arXiv:2303.10158* (2023).
17. Ntoutsis, Eirini, et al. "Bias in data-driven artificial intelligence systems—An introductory survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.3 (2020): e1356.
18. Jarrahi, Mohammad Hossein, Ali Memariani, and Shion Guha. "The Principles of Data-Centric AI (DCAI)." *arXiv preprint arXiv:2211.14611* (2022).
19. Zha, Daochen, et al. "Data-centric AI: Perspectives and Challenges." *arXiv preprint arXiv:2301.04819* (2023).
20. Mazumder, Mark, et al. "Dataperf: Benchmarks for data-centric ai development." *arXiv preprint arXiv:2207.10062* (2022).
21. Miranda, Lester James. "Towards data-centric machine learning: a short review." *lvmiranda921*. github.io (2021).
22. Alvarez-Coello, Daniel, et al. "Towards a data-centric architecture in the automotive industry." *Procedia Computer Science* 181 (2021): 658-663.
23. Uddin, Muhammad Fahim, and Navarun Gupta. "Seven V's of Big Data understanding Big Data to extract value." *Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education*. IEEE, 2014.
24. O'Leary, Daniel E. "Artificial intelligence and big data." *IEEE intelligent systems* 28.2 (2013): 96-99.
25. Broo, Didem Gürdür, and Jennifer Schooling. "Towards data-centric decision making for smart infrastructure: Data and its challenges." *IFAC-PapersOnLine* 53.3 (2020): 90-94.
26. Jakubik, Johannes, et al. "Data-centric Artificial Intelligence." *arXiv preprint arXiv:2212.11854* (2022).
27. Li, Xiao-Hui, et al. "A survey of data-driven and knowledge-aware explainable ai." *IEEE Transactions on Knowledge and Data Engineering* 34.1 (2020): 29-49.

28. Ntoutsi, Eirini, et al. "Bias in data-driven artificial intelligence systems—An introductory survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.3 (2020): e1356.
29. Kanter, James Max, Benjamin Schreck, and Kalyan Veeramachaneni. "Machine Learning 2.0: Engineering Data Driven AI Products." arXiv preprint arXiv:1807.00401 (2018).
30. Xu, Ke, et al. "Advanced data collection and analysis in data-driven manufacturing process." *Chinese Journal of Mechanical Engineering* 33.1 (2020): 1-21.
31. Maranghi, Marianna, et al. "AI-based Data Preparation and Data Analytics in Healthcare: The Case of Diabetes." arXiv preprint arXiv:2206.06182 (2022).
32. Bergen, Karianne J., et al. "Machine learning for data-driven discovery in solid Earth geoscience." *Science* 363.6433 (2019): eaau0323.
33. Jöckel, Lisa, and Michael Kläs. "Increasing Trust in Data-Driven Model Validation: A Framework for Probabilistic Augmentation of Images and Meta-data Generation Using Application Scope Characteristics." *Computer Safety, Reliability, and Security: 38th International Conference, SAFECOMP 2019, Turku, Finland, September 11–13, 2019, Proceedings* 38. Springer International Publishing, 2019.
34. Burr, Christopher, and David Leslie. "Ethical assurance: a practical approach to the responsible design, development, and deployment of data-driven technologies." *AI and Ethics* (2022): 1-26.
35. Lomas, James, Nirmal Patel, and Jodi Forlizzi. "Continuous improvement: How systems design can benefit the data-driven design community." (2018).
36. Yablonsky, S. "Multidimensional data-driven artificial intelligence innovation." *Technology innovation management review* 9.12 (2019): 16-28.
37. Batista, G. E., & Monard, M. C. (2018). Data quality in machine learning: A study in the context of imbalanced data. *Neurocomputing*, 275, 1665-1679..
38. Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2018). Data quality assessment. In *Data and Information Quality* (pp. 219-253). Springer, Cham..
39. Halevy, A., Korn, F., Noy, N., Olston, C., Polyzotis, N., Roy, S., & Whang, S. (2020). Goods: Organizing Google's datasets. *Communications of the ACM*, 63(11), 50-57.
40. Redman, T. C. (1996). *Data quality for the information age*. Artech House, Inc.
41. Juran, J. M., & Godfrey, A. B. (2018). *Juran's Quality Handbook: The Complete Guide to Performance Excellence*. McGraw-Hill Education.
42. Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). TI-CNN: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749.
43. Barocas, S., Hardt, M., & Narayanan, A. (2021). Fairness and machine learning. *Limitations and Opportunities*, 1(1), 1-269.
44. Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.
45. Hassan, N. U., Asghar, M. Z., Ahmed, S., & Zafar, H. (2021). A survey on data quality issues in big data. *ACM Computing Surveys (CSUR)*, 54(1), 1-37.
46. Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188.
47. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
48. Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. (2018). Data quality in the Internet of Things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 124, 289-310.
49. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
50. Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., ... & Chuang, I. (2014). Privacy, anonymity, and big data in the social sciences. *Communications of the ACM*, 57(9), 56-63.
51. García, S., Luengo, J., & Herrera, F. (2016). *Data preprocessing in data mining*. Springer.
52. Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2018). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. MIT Press.
53. Guyon, I., Gunn, S., & Ben-Hur, A. (2004). Result analysis of the NIPS 2003 feature selection challenge. *Advances in Neural Information Processing Systems*, 17, 545-552.
54. Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148-152.
55. Otto, B. (2011). Organizing data quality management in enterprises. *Proceedings of the 17th Americas Conference on Information Systems (AMCIS)*, 1-9.
56. Weill, P., & Ross, J. W. (2004). *IT governance: How top performers manage IT decision rights for superior results*. Harvard Business Press.
57. Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *IEEE Computer*, 46(6), 32-38.
58. Panian, Z. (2010). Some practical experiences in data governance. *World Academy of Science, Engineering, and Technology*, 66, 1248-1253.

59. Laney, D. (2012). *Infonomics: The economics of managing, measuring, and monetizing information*. Gartner Research.
60. Thomas, G., & Griffin, R. (2015). Data governance: A taxonomy of data quality interventions. *International Journal of Information Quality*, 4(1), 4-17.
61. Begg, C., & Caira, T. (2013). Data governance: More than just keeping data clean. *Journal of Enterprise Information Management*, 26(6), 595-610.
62. Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.
63. Candès, E. J., & Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6), 717-772.
64. Chandrashekar, G., & Sahin, F. (2018). A survey on feature selection methods. *Computers & Electrical Engineering*, 66, 31-47.
65. Hastie, T., Tibshirani, R., & Wainwright, M. (2019). *Statistical learning with sparsity: the Lasso and generalizations*. Chapman and Hall/CRC.
66. Wong, S. C., Gatt, A., Stamatescu, V., & McDonnell, M. D. (2018). Understanding data augmentation for classification: when to warp? In *2018 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-8). IEEE.
67. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2018). Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501.
68. Yang, Y., Loog, M., & Hospedales, T. M. (2018). Active Learning by Querying Informative and Representative Examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10), 2436-2450. DOI: 10.1109/TPAMI.2017.2760833.
69. Li, Y., & Guo, Y. (2019). Adaptive Active Learning for Image Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, pp. 7663-7671. DOI: 10.1109/CVPR.2019.00786.
70. Siddiquie, B., & Gupta, A. (2019). Human Effort Estimation for Visual Tasks. *International Journal of Computer Vision*, 127(8), 1161-1179. DOI: 10.1007/s11263-019-01166-2.
71. Zhang, Y., Chen, T., & Zhang, Y. (2019). Challenges and countermeasures of big data in artificial intelligence. *Journal of Physics: Conference Series*, 1237(3), 032023.
72. Zhu, Y., & Lapata, M. (2020). Learning to attend, copy, and generate for session-based query suggestion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
73. Halevy, A., Norvig, P., & Pereira, F. (2020). The unreasonable effectiveness of data. In *IEEE Intelligent Systems*, 24(2), 8-12.
74. X., Li, Q., Dong, S., & Ye, S. (2021). Storage challenges and solutions in the AI era. *Frontiers of Information Technology & Electronic Engineering*, 22(6), 743-767.
75. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50-60.
76. Wu, Y., Liu, J., He, H., Chen, H., & Chen, J. (2021). Data storage technology in artificial intelligence. *IEEE Access*, 9, 37864-37881.
77. Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). *Automated machine learning: Methods, systems, challenges*. Springer Nature.
78. Sharma, H., Park, J., Mahajan, D., Amaro, E., Kaeli, D., & Kim, Y. (2020). From high-level deep neural models to FPGAs. In *Proceedings of the 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*.
79. Chen, Y., Wang, T., Yang, Y., & Zhang, B. (2020). Deep model compression: Distilling knowledge from noisy teachers. arXiv preprint arXiv:1610.09650.
80. Ratner, A., Bach, S., Ehrenberg, H., Fries, J., Wu, S., & Ré, C. (2019). Snorkel: Rapid training data creation with weak supervision. *Proceedings of the VLDB Endowment*, 11(3), 269-282.
81. Zhang, H., Wu, J., Zhang, Z., & Yang, Q. (2021). Collaborative learning for data privacy and data utility. *IEEE Transactions on Knowledge and Data Engineering*.
82. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2), 1-19.
83. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhang, Y. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1-2), 1-210.
84. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2020). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637-646.
85. Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423-443.
86. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*.

87. Sattler, F., Wiedemann, S., Müller, K. R., & Samek, W. (2019). Robust and communication-efficient federated learning from non-IID data. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9), 3400-3413.
88. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. In *Proceedings of the 2020 International Conference on Learning Representations (ICLR)*.
89. Yurochkin, M., Agarwal, N., Ghosh, S., Greenewald, K., Hoang, L., & Khazaeni, Y. (2019). Bayesian nonparametric federated learning of neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*. K
90. onečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*.
91. Truex, S., Baracaldo, N., Anwar, A., Steinke, T., & Ludwig, H. (2020). A hybrid approach to privacy-preserving federated learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security (AISeC)*.
92. Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, 78, 680-698.
93. Wang, S., Tuor, T., Salonidis, T., Leung, K. K., Makaya, C., He, T., & Chan, K. (2019). Adaptive deep learning model selection on embedded systems. In *Proceedings of the 3rd ACM/IEEE Symposium on Edge Computing (SEC)*.
94. Kumar, A., Goyal, S., Varma, M., & Jain, P. (2021). Resource-constrained distributed machine learning: A survey. *ACM Computing Surveys (CSUR)*, 54(5), 1-34.
95. Zhang, Z., Mao, Y., & Letaief, K. B. (2019). Energy-efficient user association and resource allocation in heterogeneous cloud radio access networks. *IEEE Journal on Selected Areas in Communications*, 37(5), 1107-1121.
96. Zhang, H., Wu, J., Zhang, Z., & Yang, Q. (2021). Collaborative learning for data privacy and data utility. *IEEE Transactions on Knowledge and Data Engineering*.
97. Liang, X., Zhao, J., Shetty, S., Liu, J., & Li, D. (2020). Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In *Proceedings of the IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*.
98. Chen, X., Zhang, W., Wang, X., & Li, T. (2021). Privacy-Preserving Federated Learning for IoT Applications: A Review. *IEEE Internet of Things Journal*, 8(8), 6078-6093. doi: 10.1109/JIOT.2021.3095079
99. Zhao, Y., & Fan, L. (2021). A secure data sharing scheme for cross-border cooperation in the artificial intelligence era. *Security and Communication Networks*, 2021, 1-12. doi: 10.1155/2021/5583546
100. Carlini, N., Liu, C., Erlingsson, U., Kos, J., Song, D., & Wicker, M. (2019). The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks. *Proceedings of the 28th USENIX Security Symposium*, 267-284. Retrieved from <https://www.usenix.org/system/files/sec19-carlini.pdf>
101. Jayaraman, B., & Evans, D. (2020). Evaluating Membership Inference Attacks in Machine Learning: An Information Theoretic Framework. *IEEE Transactions on Information Forensics and Security*, 15, 1875-1890. doi: 10.1109/TIFS.2020.2967275
102. Dwork, C., Roth, A., & Naor, M. (2018). Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation* (pp. 1-19). Springer. doi: 10.1007/978-3-319-90023-0_1
103. Truex, S., Xu, C., Calandrino, J., & Boneh, D. (2019). The Limitations of Differential Privacy in Practice. *Proceedings of the 28th USENIX Security Symposium*, 1045-1062.
104. Goodfellow, I., Shlens, J., & Szegedy, C. (2022). Explaining and Harnessing Adversarial Examples. *Communications of the ACM*, 65(1), 56-65. doi: 10.1145/3460113
105. Akhtar, N., & Mian, A. (2018). Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey. *IEEE Access*, 6, 14410-14430. doi: 10.1109/ACCESS.2018.2800435
106. Steinhardt, J., Koh, P. W., & Liang, P. (2018). Certified Defenses Against Adversarial Examples. *Proceedings of the 6th International Conference on Learning Representations*.
107. Zhu, M., Yin, H., & Yang, X. (2021). A Comprehensive Survey of Poisoning Attacks in Federated Learning. *IEEE Access*, 9, 57427-57447. doi: 10.1109/ACCESS.2021.3071097
108. Sun, Y., Zhang, T., Wang, J., & Wang, X. (2020). A Survey of Deep Neural Network Backdoor Attacks and Defenses. *IEEE Transactions on Neural Networks and Learning Systems*, 31(10), 4150-4169. doi: 10.1109/TNN
109. Gu, T., Dolan-Gavitt, B., & Garg, S. (2019). BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain. *Proceedings of the 28th USENIX Security Symposium*, 1965-1980.
110. Liu, Y., Ma, X., Ateniese, G., & Hsu, W. L. (2018). Trojaning Attack on Neural Networks. *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 27-41. doi: 10.1145/3243734.3243835
111. Gao, Y., Sun, X., Zhang, Y., & Liu, J. (2021). Trojan Attacks on Federated Learning Systems: An Overview. *IEEE Network*, 35(2), 144-150. doi: 10.1109/MNET.001.2000741

112. Tramèr, F., et al. (2016). Stealing Machine Learning Models via Prediction APIs. Proceedings of the 25th USENIX Security Symposium, 601-618. Retrieved from https://www.usenix.org/system/files/conference/usenixsecurity16/sec16_paper_tramer.pdf
113. Jagielski, M., et al. (2020). Model Theft and Out-of-Distribution Detection in Machine Learning. Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, 212-226. doi: 10.1145/3372297.3417299
114. Liu, Y., Chen, J., Liu, T., & Yang, Y. (2020). Trojan Detection via Fine-Pruning. Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, 1151-1168. doi: 10.1145/3372297.3417866
115. Abadi, M., et al. (2016). Deep Learning with Differential Privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308-318. doi: 10.1145/2976749.2978318
116. Xu, W., et al. (2021). Bridging the Gap Between Input Validation and Trustworthy AI. Proceedings of the 2021 IEEE Symposium on Security and Privacy, 1395-1412. doi: 10.1109/SP40000.2021.00081
117. Bonawitz, K., et al. (2019). Towards Federated Learning at Scale: System Design. Proceedings of the 2nd Workshop on Systems for ML at Scale, 1-6
118. Rai, S., et al. (2021). A Survey of Privacy-Preserving Machine Learning Techniques. *ACM Computing Surveys*, 54(2), 1-42. doi: 10.1145/3447391
119. Madry, A., et al. (2018). Towards Deep Learning Models Resistant to Adversarial Attacks. Proceedings of the 35th International Conference on Machine Learning, 297-306.
120. Yuan, X., et al. (2019). Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9), 2805-2824. doi: 10.1109/TNNLS.2019.2919365
121. Paudice, A., et al. (2018). MAMADroid: Detecting Android Malware by Building Markov Chains of Behavioral Models. Proceedings of the 27th USENIX Security Symposium, 1355-1372.
122. Chen, Y., et al. (2020). Data Poisoning Attacks on Machine Learning: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 32(4), 685-706. doi: 10.1109/TKDE.2019.2919365
123. Polonetsky, J., & Tene, O. (2018). GDPR and AI: Friends or Foes? *IEEE Security & Privacy*, 16(3), 26-33. doi: 10.1109/MSEC.2018.2806198
124. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. FairMLBook.org.
125. Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters.
126. Simonite, T. (2018). When it comes to gorillas, Google Photos remains blind. Wired.
127. Vincent, J. (2016). Twitter taught Microsoft's AI chatbot to be a racist in less than a day. The Verge.
128. Harding, S. (2019). Apple's credit card gender bias draws regulatory scrutiny. Forbes.
129. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.
130. Olteanu, A., Castillo, C., Diaz, F., & Kicman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data Science*, 2, 13.
131. Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., ... & Wang, W. Y. (2019). Mitigating gender bias in natural language processing: Literature review. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1630-1640.
132. Torralba, A., & Efros, A. A. (2011). Unbiased look at dataset bias. *IEEE Conference on Computer Vision and Pattern Recognition*, 1521-1528.
133. Zhao, Z., Wallace, B. C., Jang, E., Choi, Y., & Lease, M. (2021). Combating human trafficking: A survey of AI techniques and opportunities for technology-enabled counter-trafficking. *ACM Computing Surveys*, 54(1), 1-35.
134. Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.
135. Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, 27(3), 313-327.
136. Gupta, P., & Raghavan, H. (2021). Temporal bias in machine learning. arXiv preprint arXiv:2104.12843.
137. Gutierrez, M., & Serrano-Guerrero, J. (2020). Bias-aware feature selection in machine learning. arXiv preprint arXiv:2007.07956.
138. Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus, and Giroux.
139. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.
140. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on fairness, accountability and transparency, 77-91.
141. Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.
142. Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841-887.

143. Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2020). On fairness and calibration. *Advances in Neural Information Processing Systems*, 33.
144. Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. W. (2019). Gender bias in contextualized word embeddings. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 629-634.
145. Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Nagar, S. (2018). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *IBM Journal of Research and Development*, 63(4/5), 4-1.
146. Verma, S., & Rubin, J. (2018). Fairness definitions explained. *Proceedings of the International Workshop on Software Fairness*, 1-7.
147. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214-226.
148. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115
149. Hao, K. (2020). This is how AI bias really happens—and why it's so hard to fix. *MIT Technology Review*.
150. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
151. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
152. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160.
153. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 80-89). IEEE.
154. Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., ... & Eckersley, P. (2020). Explainable machine learning in deployment. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 648-657).
155. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841-887.
156. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
157. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), 1-42.
158. Rudin, C. (2019). Stop explaining black-box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215.
159. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
160. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (pp. 4765-4774).
161. Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
162. Carter, S., Armstrong, Z., Schönberger, L., & Olah, C. (2019). Activation atlases: Unsupervised exploration of high-dimensional model internals. *Distill*, 4(6), e00020.
163. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
164. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4), e1312.
165. Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 279-288).
166. Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Langhans, S. D. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233.
167. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Agarwal, S. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
168. Knight, W. (2021). The future of AI depends on a huge workforce of human teachers. *Wired*.
169. Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25.

170. Wang, S., Fisch, A., Oh, J., & Liang, P. (2020). Data Programming for Learning with Noisy Labels. *Advances in Neural Information Processing Systems*, 33, 14883-14894.
171. Crawford, K., & Calo, R. (2021). There is a blind spot in AI research. *Nature*, 538(7625), 311-313.
172. Mittelstadt, B., Russell, C., & Wachter, S. (2021). Explaining explanations in AI. *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 279-288.
173. Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. Nuffield Foundation.
174. Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., ... & Trench, M. (2018). Skill shift: Automation and the future of the workforce. McKinsey Global Institute.
175. World Economic Forum. (2021). Jobs of Tomorrow: Mapping Opportunity in the New Economy. http://www3.weforum.org/docs/WEF_Jobs_of_Tomorrow_2020.pdf
176. Bessen, J. E., Impink, S. M., Reichensperger, L., & Seamans, R. (2019). The Business of AI Startups. NBER Working Paper No. 24255.
177. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38.
178. Xu, H., Gu, L., Choi, E., & Zhang, Y. (2021). Secure and privacy-preserving machine learning: A survey. *Frontiers of Computer Science*, 15(2), 1-38.
179. Yang, G. Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., ... & Wood, R. (2020). The grand challenges of Science Robotics. *Science Robotics*, 3(14), 7650.
180. Andreas Holzinger, Katharina Keiblinger, Petr Holub, Kurt Zatloukal, Heimo Müller, AI for life: Trends in artificial intelligence for biotechnology, *New Biotechnology*, Volume 74, 2023, Pages 16-24..
181. Jawad, Khurram, et al. "Novel Cuckoo Search-Based Metaheuristic Approach for Deep Learning Prediction of Depression." *Applied Sciences* 13.9 (2023): 5322.
182. Yaqoob, Abrar, et al. "A review on nature-inspired algorithms for cancer disease prediction and classification." *Mathematics* 11.5 (2023): 1081.
183. Aziz, Rabia Musheer, et al. (2023) Modified Genetic Algorithm with Deep Learning for Fraud Transactions of Ethereum Smart Contract." *Applied Sciences* 13.2, 697.
184. Aziz, R.M., Desai, N.P. & Baluch, M.F. (2023) Computer vision model with novel cuckoo search based deep learning approach for classification of fish image. *Multimed Tools Appl* 82, 3677-3696
185. Aziz, Rabia Musheer; Baluch, Mohammed Farhan; Patel, Sarthak; and Kumar, Pavan (2022) "A Machine Learning based Approach to Detect the Ethereum Fraud Transactions with Limited Attributes," *Karbala International Journal of Modern Science: Vol. 8 : Iss. 2 , Article 13.*
186. Thayyib, P. V., et al. (2023) State-of-the-Art of Artificial Intelligence and Big Data Analytics Reviews in Five Different Domains: A Bibliometric Summary." *Sustainability* 15, 5, 4026.
187. :Saghiri, A.M.; Vahidipour, S.M.; Jabbarpour, M.R.; Sookhak, M.; Forestiero, (2022) A. A Survey of Artificial Intelligence Challenges: Analyzing the Definitions, Relationships, and Evolutions. *Appl. Sci.*, 12, 4054
188. Serey, J.; Quezada, L.; Alfaro, M.; Fuertes, G.; Vargas, M.; Ternero, R.; Sabattin, J.; Duran, C.; Gutierrez, S. (2021) Artificial Intelligence Methodologies for Data Management. *Symmetry*, 13, 2040.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the MDPI and/or editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.