Deloitte.

AI data readiness (AIDR)

Getting your data ready for AI adoption at scale

July 2024

Introduction

Data is becoming increasingly important for the success of a business as organizations adopt to changes in the business environment; become more digital, data-driven, and use data to influence decision-making; and become more responsive to customer needs. Data has historically been used to drive various aspects of business and has been an enabler for emerging technologies, including artificial intelligence (AI), which has been a game-changer in recent years.

AI, at its core, is a sophisticated and multifaceted concept, intricately woven from three fundamental components:

Business context | **Technique/ algorithm** | **Data**

Each of these elements plays a crucial role in the functioning and impact of AI and has specific risks and challenges that need to be mitigated through an effective set of implemented AI governance requirements.

What makes up AI?	Common risks/challenges for governing AI	
Business context The business context or problem intended to be addressed with the use of AI models/algorithms	• Purpose and value of AI • Accountability for AI use • Impact on people and ecosystem	• Operational controls • Human-in-the-loop • Response to unintended outcomes
Technique/algorithm Specific technique, technology, or combination of these that are used to address a specific use case or business problem (e.g., natural language processing (NLP), neural network)	• Applicability to use case • Obfuscation/explainability • Vendor/ platform dependency	• Life cycle controls • Performance indicators • Data and model drift
Data Datasets (internal or external) used to build and train AI models/algorithms, and their level of curation and fit -for-use (e.g., availability of vectors, weights, results)	• Data governance and standards • Data ethics and privacy • Data quality	• Data resiliency • Data movement • Data use/fit for purpose • Third-party data

Defining the business problem is a linchpin in maintaining a sharp focus on requirements throughout the creation of an AI model. This initial step serves as a compass, guiding the development process by helping to clearly articulate the business requirements (i.e., specific challenges and/or opportunities) that the AI model aims to address.

Data availability

Typical data-related challenges for organizations

- Is the required data associated with the business problem available within the organization?
- In case of availability of data, what are the insights into nuances of data availability that assist practitioners in making informed decisions regarding data collection, pre-processing, and augmentation?

Identifying the appropriate algorithm or technique is another critical step in implementing an AI solution, once the business requirements have been identified. This involves considering factors such as **scalability**, **interpretability**, and **computational efficiency**. Consequently, this step assists in laying the groundwork for subsequent phases of model development, including data pre-processing, feature engineering, and model evaluation, to reasonably ensure that the AI model is effective in addressing the targeted business challenge.

Data quality and fit for purpose

Typical data-related challenges for organizations

- Is the data on which the AI model is constructed, capable of providing meaningful insights or predictions?
- Are there potential challenges related to data quality or quantity requiring measures to address these issues throughout the model development process?

The effective implementation of AI hinges on adeptly managing various **data challenges**, especially in the context of heightened complexity in **data life cycle management** for AI applications. Typical challenges include:

- The **quality and availability of data**, with poor data quality potentially impeding AI system development.
- **Ethical considerations**, including **privacy** and **security**, which highlight the importance of **regulatory compliance**.
- **Data governance**, standards, regulatory compliance, and data resilience are emphasized to help minimize risks and reasonably ensure accountability in AI decision-making.

To mitigate the errors and inefficiencies, it is crucial to implement effective **data quality** processes, including data cleansing, validation, and monitoring. Data quality standards and practices are essential to reasonably ensure that the data used for training AI models is accurate, representative, and unbiased.

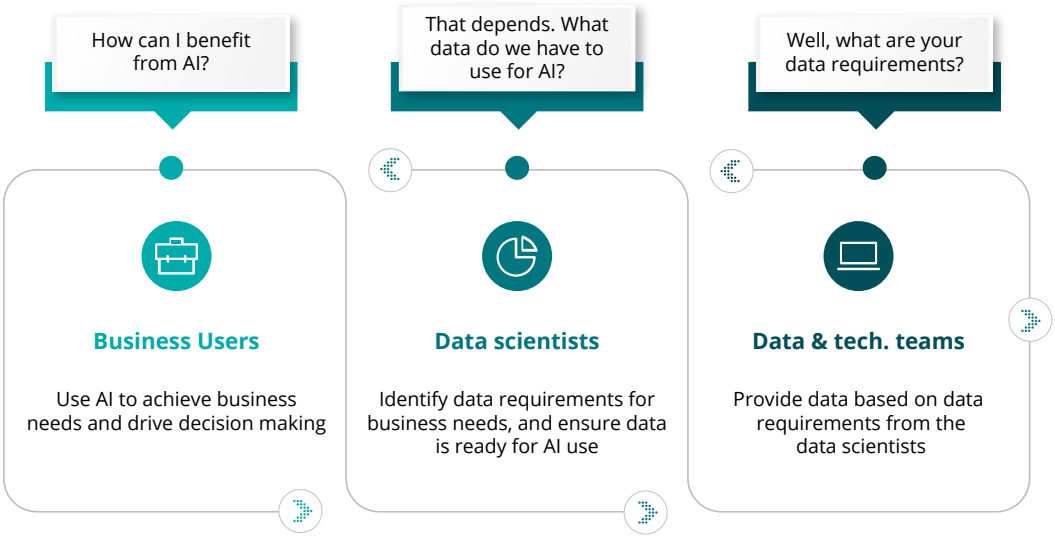
Use cases related to AI over the years and associated data concerns[1]

AI usage	Data concerns examples
1. Personalized customer service (e.g., AI powered chatbots and virtual assistants)	• Availability of historical transactional data • Accuracy of data and use of data that is fit for purpose
2. Real-time fraud detection and security	• False positives in data used to train the fraud detection model • Manual reviews vs. level of automation to validate data quality
3. AI powered robo-advisor	• Compliance to data privacy rules • Source/method for acquisition of data used for advanced analytics models
4. Credit risk assessment	• Processes to manage sourcing, evaluation, procurement, integration, and maintenance of third-party datasets • Bias in data uses for assessments and decisions

1. Todd Bigham et al., AI and risk management: Innovating with confidence, Deloitte, 2018.

What is AI data readiness (AIDR)?

An organization's preparedness in implementing strategies to help guide effective AI deployment by reasonably ensuring that its data is available, high quality, properly structured, and aligned with its AI use cases.



Considerations for AIDR:

- Understanding of the business context and objective for AI
- Structuring a process to identify and evaluate available data
- Articulating data readiness gaps and improvement opportunities
- Facilitating a common taxonomy across stakeholders for AI data readiness

What are the steps to implement and reasonably ensure the readiness of AI data?

Identifying the data scope, evaluating data readiness, and implementing improvements for data readiness are pivotal in creating an effective AI model.

Define the scope of data whether it focuses on a single, **specific use case or aims for an enterprise-wide transformation** to determine targeted and effective planning.

Utilize a **specialized Data Readiness Assessment Tool** to evaluate the readiness of in-scope data for the intended AI application and use.

Develop and execute a plan to improve data readiness through **combination of near and longer-term actions** to accelerate AI build and deployment.

Define data scope

Defining the data scope is a crucial initial step for financial institutions embarking on the journey into artificial intelligence. Specifically, the scope for AI data readiness involves evaluating risk tolerance, harnessing the insights of use-case owners through strategic collaboration, and ultimately identifying key characteristics to help articulate the problem or objective addressed by the AI model.

This scope can range from a focused application like fraud detection to a broader, enterprisewide embrace of artificial intelligence.

Several activities are involved in creating a well-defined data scope:

Activities	Considerations
1. Identify required data inputs	• Required data inputs designed to optimize AI model performance • Data availability • Specific data types, both structured and unstructured
2. Define data sources	• Data sources available and understanding how to apply them to the AI model • Identification and documentation of the data sources to be leveraged (e.g., internal databases, external APIs, third-party datasets, or acquired data)
3. Establish data collection and pre-processing requirements	• Data cleaning, normalization, feature engineering, and augmentation as required
4. Consider data privacy and security	• Adherence to data privacy regulations and safeguarding sensitive information • Access control to reasonably ensure that authorized personnel with specific roles can view or modify sensitive data
5. Define data scope boundaries	• Definition of time frame and scale of datasets • Limitations or exclusions to be imposed on the datasets

In the context of a fraud detection use case for a bank, establishing a precise data scope is paramount. The data scope for this scenario could include:

- **Financial transactions**: The primary focus may likely be on data related to financial transactions encompassing details such as transaction amounts, time stamps, and transaction types.
- **Customer behavior patterns**: Analyzing historical customer behavior is essential. This includes studying spending patterns, transaction frequency, and typical transaction sizes associated with each customer.
- **Geographical information**: Considering the geographic location of transactions can be crucial for identifying anomalies. Unusual transactions in locations not typically associated with the customer's behavior could be red flags.

Evaluate data readiness: Five dimensions

In the realm of financial services, where data is as valuable as currency, the readiness of this data for AI implementation is not just a technical requirement but a strategic imperative.

Having an AI data readiness approach allows for a structured process to evaluate the preparedness of a client's data landscape across five critical dimensions:

Dimensions	Capabilities for evaluation
1. Data Availability Data that's well-organized, structured, and easily accessible in a timely manner to **boost efficiency in storage, retrieval, and processing,** while promoting reusability along with abstraction.	• Data Management • Data Integration and Utilization • Advanced Analytics • Data Storage
2. Data volume & diversity Sufficient and diverse datasets (e.g., representing real-world scenarios) allow for AI solutions to **identify complex patterns** and deliver more **accurate predictions.**	• Historical Data • Data Sourcing • Data Diversity (Features)
3. Data quality & integrity By adopting and upholding leading data quality standards and processes, AI models can work with **accurate, consistent & fit-for-purpose data, leading to reliable and accurate outcomes.**	• Data Accuracy and Fitness • Standardization & Protocols • Metadata • Documentation & Reporting
4. Data governance Implementing a **robust governance framework for data and AI** can help **manage data throughout its lifecycle**, establish **policies, standards,** data **ownership,** and **set guidelines** around **use of data for AI.**	• Data Strategies • AI Governance and Documentation • Data Collaboration • Data Assessment
5. Data ethics & responsibility **Data ethical considerations incorporated in data policies** for use of data for AI to drive improvement in **accountability of AI-based decision-making, emphasize safety, and foster transparency around data usage** for AI processes.	• Regulatory Compliance • Data Protection & Access Monitoring • Use Case Specific Data Rules

Our AIDR questionnaire can help measure an AIDR score to facilitate a GO / NO-GO DECISION to move forward with the AI Build.

Each dimension is a pillar that upholds the integrity and efficacy of AI applications. The capabilities for evaluation listed above can reasonably ensure that these pillars are strong both individually and cohesively to support the overarching goal of AI-driven transformation.

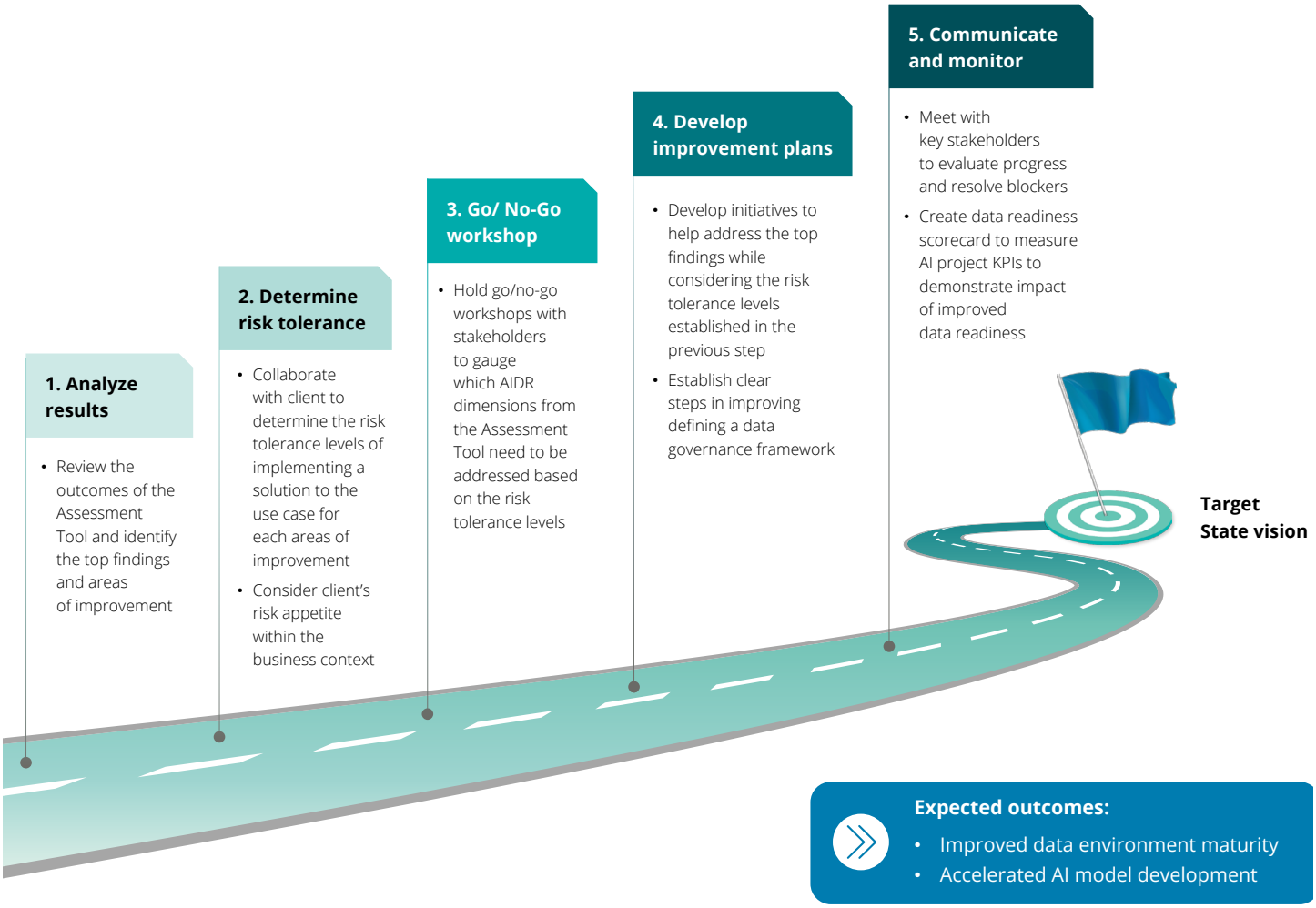
A highly effective strategy for achieving data readiness is to dedicate ample time to thoroughly analyze the existing landscape across each dimension outlined above. This process helps ensure the availability of many essential building blocks and that the necessary factors are considered for constructing the initial AI model.

This meticulous approach not only facilitates the ease of implementing subsequent models, but also reasonably ensures the ongoing health and performance of the initial AI model.

Improve data readiness

Improving AI data readiness is important because high-quality and well-structured data is one of the foundations of successful AI models and algorithms. By improving data readiness, organizations can unlock the full potential of AI and derive meaningful insights, as accuracy and reliability of data is crucial for training AI models.

The following are five tactical steps to improve AI data readiness, based on the evaluation of capabilities across each of the five domains .



Conclusion

AI has emerged as a transformative force in today's data-driven world. While readiness of data is critical for harnessing the potential of AI applications, several key takeaways have emerged, including challenges, prospects, and considerations for adoption. As organizations are investing in data infrastructure and formulating AI strategies, the continuous advancements, and a commitment to addressing data related concerns will assist in driving the future of AI applications.

AI data readiness will thus be the foundation for unlocking AI's full potential in a wide range of applications. Your data is not just information; it can be the key to your potential opportunities in leveraging AI solutions.

Are you ready to unleash the power of AI?

How can we help?

AI data readiness approach

Deloitte's AIDR approach is a tool for assessing data readiness in preparation for AI implementation.

AI Data Readiness Assessment Tool

The Data Readiness Assessment Tool is leveraged to evaluate the current state of a company's data environment in the five key dimensions of data readiness.

AI Data Readiness Score/Results

The AIDR Assessment Tool aggregates the responses from each question to show a score for each dimension, rolled up into an aggregate score.

Reach out to get started:

Our team is standing by to help and is excited for the opportunity to assist with your AI data readiness journey.

Vic Katyal
Risk & Financial Advisory
Principal and Chief Operating Officer
Deloitte & Touche LLP

Cory Liepold
Risk & Financial Advisory
Principal
Deloitte & Touche LLP

Satish Iyengar
Risk & Financial Advisory
Managing Director
Deloitte & Touche LLP

Ajay Ravikumar
Risk & Financial Advisory
Senior Manager
Deloitte & Touche LLP

Akiva Ehrlich
Digital Controls
Advisory Partner
Deloitte Israel & Co.

Shlomi Cohen
Artificial Intelligence and Data
Managing Director
Deloitte Israel & Co.

Amit Golan
Financial Industry Risk and Regulatory
Senior Manager
Deloitte Israel & Co.

Contributors:

Soumya Gollamudi, Risk & Financial Advisory, Manager, Deloitte & Touche LLP
Brian Nam, Risk & Financial Advisory, Senior Consultant, Deloitte & Touche LLP
Amanda Schwartz, Risk & Financial Advisory, Analyst, Deloitte & Touche LLP
Sahana Gurumurthy, Risk & Financial Advisory, Senior Solution Advisor, Deloitte & Touche Assurance & Enterprise Risk Services India Pvt. Ltd.

Deloitte.

About Deloitte.