MAB MAANDBLAD VOOR ACCOUNTANCY EN BEDRIJFSECONOMIE

Maandblad voor Accountancy en Bedrijfseconomie 94(3/4) (2020): 69–82
DOI 10.5117/mab.94.49425

Research Article

Improving trust in data and algorithms in the medium of AI

Aditya Vasan Srinivasan, Mona de Boer

Received 15 December 2019 | **Accepted** 17 March 2020 | **Published** 22 April 2020

Summary

Artificial Intelligence (AI) has great potential to solve a wide spectrum of real-world business problems, but the lack of trust from the perspective of potential users, investors, and other stakeholders towards AI is preventing them from adoption. To build and strengthen trust in AI, technology creators should ensure that the data which is acquired, processed and being fed into the algorithm is *accurate, reliable, consistent, relevant, bias-free, and complete*. Similarly, the algorithm that is selected, trained, and tested should be *explainable, interpretable, transparent, bias-free, reliable, and useful*. Most importantly, the algorithm and its outcomes should be auditable and properly governed.

Relevance for practice

The identification of essential trust factors of data and algorithms in the medium of AI, and the presentation of a trusted AI model incorporating such factors with detailed indicators, would be one of the prime contributions from this research. It would aid technology developers in assessing those trust factors upfront and thereby providing a seal of trust to potential users, stakeholders over the resulting AI solutions.

Keywords

Artificial Intelligence, Data, Data Quality, Algorithm, Trust

1. Introduction

Perhaps, the most transformative technology that is available today is Artificial Intelligence (AI). The resurrection of AI has been driven in recent times by the availability of more data, increased computing power and advances in algorithms. AI is claimed to ultimately transform every business in every industry (Marr 2019). It has started to become an important value for organizations, as it could help them move closer to their customers, enhance employee experience and capabilities, and decrease cost by automating tasks (IBM 2018).

The adoption of AI has, however, been met with skepticism from various stakeholders, as the technology is relatively new, untried and has posed several complications. One of the major impediments in adopting AI-based solutions is how to trust the particular technology. Imagine when two people meet for the first time, the first thing that would strike them is, questioning themselves: *Can the other person be trusted?* Similarly, in the context of AI, where investors, consumers or end-users encounter an AI-solution for the first time, what that typically strikes them *is whether the AI-solution can be trusted*. This question on trust has become a major concern in recent times. People are worried that AI won't behave as intended, that it makes mistakes, produces unethical results, shows bias, and through that harms society (Burkhardt 2019). Research from the Massachusetts Institute of Technology (MIT) has revealed that there are several issues which need to be addressed in order to trust AI (Davenport 2018). In one of the surveys done on U.S. consumers, it was found that around 41.5% of respondents don't trust any of the AI services like home assistants, financial planning, medical diagnosis, and hiring (Krogue 2017). In another survey conducted by PwC (2018) amongst business leaders, it was observed that most of the business leaders

Amsterdam University Press

are reluctant to adopt AI and holding a step back as they are concerned about biases, lack of transparency and governance along with the stakeholder's trust as a major risk (PwC 2018).

In the study of human-robot interaction, trust can be defined as "*willingness of the people to accept robot-produced information and follow robots suggestions, share tasks, contribute information, and provide support to the robot*" (Siau and Wang 2018, p. 49). Using this definition as a reference, trust in the context of AI can be defined as "*willingness of people to accept AI and believe in the suggestions, decisions made by the algorithm, share tasks, contribute information and provide support to such technology*". This view of trust is currently lacking from the perspective of technology investors, consumers, society. Take for instance, an e-commerce website using an AI-based algorithm to provide recommendations to customers, based on the history of orders purchased. The risk involved here would be relatively low as the impact on society is marginal, but what happens when an algorithm turns down bank loan applications without an explanation or what if AI flags a certain individual at the airport security checks with no apparent justification. The users or the leaders may not trust applications in such instances if they can't understand how they work and ideally, leaders won't be interested in investing in AI if they can't see evidence of how it made its decision (PwC 2018). The same problem persists when there is a need to audit such solutions and provide assurance to clients. *Investors therefore expect the solutions to be reliable, and mainly a stamp of trust from the technology creators and management (while implementing AI solutions) and from internal auditors (while assessing such AI solutions)*.

The problem statement of this research can, therefore, be described as:

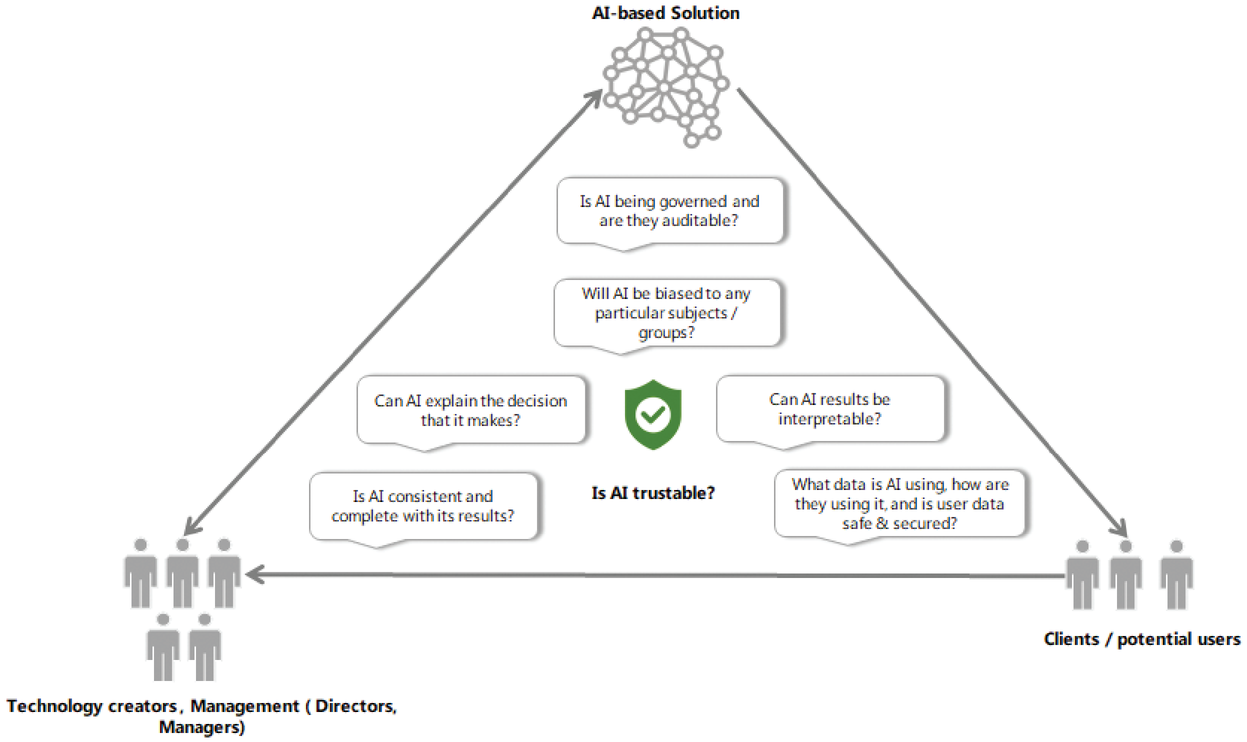
"*AI has the capability to transform the business model of every firm and industry, but the lack of trust from the investors, consumers and society as a whole is one of the major concerns and challenges that are preventing the adoption of such technology*".

As can be seen in Figure 1, trust can't be built on one set of factors alone. If an AI solution or an application is *fair* but can't resist an *attack*, it won't be trusted. If it is *secure* but can't *explain* the decision made, it won't be trusted. If it can *explain* the decisions but the *outcomes* are bad, it won't be trusted.

Siau and Wang (2017) cited that building trust is a dynamic process that involves movement from initial trust to continuous trust development. Likewise, in the context of AI, initial trust primarily depends on the quality of data and ensuring that data that is being fed to algorithms is completely trustworthy, while continuous trust depends on the performance and the purpose of algorithms.

In line with the above problem statement, the following section (2) defines the objective and scope of the research. Section (3) provides detailed analysis on the theoretical concepts of trust in terms of data and algorithms and it follows a sequential pattern. First, conscious of the value of data in AI, an overview of data quality and its dimensions is presented, followed by the identification of potential trust factors of data. Next, in the context of algorithms, potential trust factors of algorithms are identified based on studies conducted on trust factors in the field of automation and important themes laid by academic institutions and companies. Section (4) describes the

Figure 1. The actual problem in adopting AI.


AI-based Solution
Is AI being governed and are they auditable?
Will AI be biased to any particular subjects / groups?
Can AI explain the decision that it makes?
Can AI results be interpretable?
Is AI consistent and complete with its results?
Is AI trustable?
What data is AI using, how are they using it, and is user data safe & secured?
Technology creators, Management (Directors, Managers)
Clients / potential users

interview process conducted with actors. One of the key aspects of this section is the analysis and findings from the interviews. Using the findings from interviews as a reference along with the analysis done on section (2), a trusted AI model is developed by recognizing the relevant phases in an AI development along with the associated trust factors with detailed indicators. The results therefore are presented in section (5). Additionally, section (5) discusses the value of the trusted AI model. In the final section (6), the research determines the essential trust factors of data and algorithms in order to realize the primary contribution to academic research, along with limitations and recommendations.

1.1 List of abbreviations

AI – Artificial Intelligence, ML – Machine Learning, DQ – Data Quality, IQ – Information Quality, GDPR – General Data Protection Regulation, SMEs – Small and Medium-sized Enterprises.

2. Research scope & objective

When it comes to autonomous systems, the level of human trust in such systems can be determined based on human characteristics, environment characteristics, and technology characteristics (French et al. 2018). One of the meta-analyses conducted on these trust-defining factors by (Hancock et al. 2011), found that the characteristics of the ‘robot’ (i.e. technology characteristics) were the major contributors to the development of human trust, followed by environment and human characteristics – having respectively moderate and little effect over the development of trust.

The remainder of this study focuses on the category ‘technology characteristics’. In the context of AI, technology characteristics relate to both the data used in the development of AI-solutions and that data’s quality (DQ), and the algorithms which are the end result of the development process. The study is an attempt to gain an initial understanding of the essential trust factors of data and algorithms in the medium of AI and a first move towards the development of a trusted AI model. Such model is aimed at guiding organizations and their internal auditors in the assessment of AI-solutions, with the overall objective to build trust in them. The objective of this research is, therefore, to *identify the essential trust factors of data and algorithms in the medium of AI and develop a trusted AI model with detailed indicators for each of the identified factors.*

2.1 Research approach

The research is exploratory in nature in terms of identifying the essential trust factors of data and algorithms in the medium of AI and uses a qualitative study with an inductive approach in order to generate valuable theory. The research consists of three phases: the first phase focuses on the identification of potential trust factors through a literature review, the second phase is aimed at validating the trust factors, identified in the first phase, with different actors involved in the development of AI-solutions, the third phase attempts to capture the views from the previous two phases in a trusted AI model with detailed indicators for each of the identified factors.

In the second phase, a clear distinction is made between the various types of actors involved in AI-development. Most of the research conducted on AI till date hasn’t necessarily focused on the actors or stakeholders involved in the AI development process. Phase two of this research starts with identifying the actors typically involved in these development processes. This overview is not intended to and does not cover all relevant actors in all relevant situations of AI-development, as when AI becomes more complicated and expansive, we may need to identify more actors than done in phase two of this research. Based on the engagement with AI experts and the examination of several white papers, the research has identified a number of probable actors, and these were the actors who were interviewed in order to find the essential trust factors of data and algorithms. The probable actors were AI experts, data scientists, data engineers, risk advisors, and auditors. These actors would be a representation for most of the large sized technology and auditing firms.

3. Literature study

A literature study was performed to identify potential trust factors of data and algorithms in the context of AI. The study examined literature on concepts of trust in terms of data quality dimensions, and factors influencing trust in automation. Although extensive studies have been conducted on the concept of trust and its factors from multiple dimensional aspects in scientific academia, no substantial research has been done to determine the factors that could possibly influence trust in emerging technologies, and AI in particular. For this study we have presumed that some of the trust factors that influence trust in automation would also hold key for influencing trust in AI, as automation is one of the dimensions of AI. Factors that could influence trust towards the data were recognized in terms of DQ dimensions. There is no single agreement on what dimensions constitutes DQ, as multiple dimensions of DQ could be found in the academic literature and they are subject to variation, depending on the context of the environment. As such, for this research, the chosen factors were ensured to: (1) be the most frequently recommended dimensions for DQ, (2) be relevant to the context of AI, and (3) be able to highly influence the trust in AI as a whole.

3.1 Potential trust factors of data

Data is considered to be good only when it conforms to relevant use and meets the basic requirements, and the best way to assess it is by DQ dimensions. DQ dimensions are the most commonly accepted and widely used method in the assessment of data, and they have been gaining a great deal of attention from researchers and data experts in recent years due to the abundance of (big) data and its influence in AI. It is commonly believed that most enterprises are moving towards data-driven decision making, using these technologies, and it is essential to have data that meets the required quality. That means, amongst other things, that all fields in the data should be present, that there are no duplicates in the data, that the data doesn't contain any special characters, etc. Data which is of low quality, outdated, incomplete or incorrect at different stages of data processing, would lead to poor prediction and assessment and, in turn, to biases, possibly resulting in the breach of fundamental rights of the individual, or purely incorrect conclusions and bad outcomes (FRA 2019). The use of data and its analysis are mainly reliant on the quality of data which is necessary for generating the value from the data (Cai and Zhu 2015). Professor Richard Y. Wang, who leads the data quality management group of MIT University and has conducted extensive research in the area of data quality, defines the concept as "*data that are fit for use by data consumers*" and "data quality dimension" as a "*set of data quality attributes that represent a single aspect or construct of data quality*" (Wang and Strong 1996, p. 6). Dimensions like accuracy, completeness, consistency, timeliness, relevancy, reliability, interpretability etc. contribute to the quality of data and these dimensions have been regarded as a basic set of DQ dimensions, which have constituted the focus of the majority of academic researchers in the data domain (Batini et al. 2009). Data which has a representation of a real-world state is referred to as 'accuracy' (Janssen et al. 2017). Likewise, Janssen et al. (2017) define 'relevancy' as addressing the customer needs through data. 'Completeness' means that there is no contradiction between the processed data and the actual data source, and 'timeliness' as the availability of data on time.

One of the central aspects of this research is the quest for dimensions that constitute data quality in the context of AI:

Accuracy – refers to the degree to which the data is correct, reliable, certified and free of error (Wang and Strong 1996). In the context of AI, where the algorithm would require a huge volume of data in order to learn and make decisions, accuracy would be the main factor. For accuracy, the data needs to reflect the actual state of what the user expects in terms of real-world representation during the acquisition and processing of the data.

Consistency – refers to the degree to which the data is presented in a format that is the same and compatible with previous data (Wang and Strong 1996). However, it can refer to several aspects of the data. For instance, with respect to the value of data: the value or entries in the data should be the same in all cases; with respect to the representation of the data: the entity types and attributes should have the same basic structure wherever possible. In the scope of AI applications, this means ensuring that the information (processed data) and the actual data source are still consistent and don't contain contradictions.

Completeness – refers to the degree to which a given dataset contains all the relevant data and mandatory attributes which shouldn't be null (Wand and Wang 1996). Similarly, in relevance to the context of AI, data is said to be complete only when main and mandatory entries in the data are not null and the data is reflecting all the possible states of the user population in order to avoid biases (Nelson et al. 2005).

Security – refers to the extent to which data received from third parties, client sources and other external sources is highly secured, especially when the data holds sensitive information.

Privacy – When the data holds user/customer information, assurances must be made that such data is not being used unlawfully and usage is in compliance with necessary data protection laws and EU General Data Protection Regulation (GDPR).

Relevancy – refers to the extent to which data obtained is relevant to the specific domain and purpose of interest in a given context (Bovee et al. 2003).

Interpretability – refers to the extent to which the data obtained is in proper language and metrics, meaningful, and the symbols used are understandable to the users working on the data (Pipino et al. 2002).

Reliability – refers to the extent to which data obtained from client sources or any external party is trustable and can be relied on to convey the right information (Batini and Scannapieco 2016).

Bias-Free or Objectivity – refers to the extent to which the data obtained and processed is unbiased, unprejudiced, and impartial (Wang and Strong 1996).

Usefulness – refers to the extent to which the information (processed data) is applicable to the actual context of the goal and useable for the task at hand (Wang and Strong 1996). In addition, the overall usefulness of the data is determined when other DQ dimensions such as reliability, bias-free, consistency, completeness, etc. of the data are validated.

3.2 Potential trust factors of algorithms

Since not many models of trust were proposed in academia with respect to the development of trust in AI, several researchers have identified multiple trust factors and proposed models in developing trust in automation. It is assumed that some of those factors could also influence trust in algorithms and its resulting outcomes in the context of AI, as automation is one of the dimensions of AI. One of the most widely accepted definitions of automation in the academic literature, which summarizes the whole process of automation, was proposed by Lee and See (2004):

> *"Automation is a technology that actively selects data, transforms information, makes decision, or control process"* (Lee and See 2004, p. 50).

The same researchers proposed a basic definition of trust, which is in line with their evaluation of trust as an attitude in automation: *the attitude that an agent will achieve an individual's goal in a situation characterized by uncertainty and vulnerability* and in this case, an agent could be an automated machine driven by AI. Muir (1994) initially developed a trust model in automation, and it was primarily based on the work done on trust in interpersonal relationships by Rempel et al. (1985), as the model closely ties with the theoretical basis of trust foundation. Dimensions like predictability, dependability and faith were recognized as a basis for trust formation. Predictability depends on the behavior of systems, the ability of humans to assess system predictability, while dependability refers to the degree to which humans can rely upon the system. Faith is based on future behavior of the trustee. In the early 2000s, Kelly et al. (2003) presented a model in automation, outlining the trust factors and their relationship between them. Three main references of trust were recommended: understanding, competence of the automation, and self-confidence. Understanding is based on predictability and familiarity, competence of the automation relies on reliability, dependability, usefulness and robustness. Furthermore, a trust model was designed in order to enable the inclusion of insights from behavioral trust theory in the design of automated systems (Hoffmann and Söllner 2014). The model basically considers the antecedents of three dimensions underlying the formation of trust: performance, process and purpose. The performance dimension includes antecedents like competence, information accuracy in terms of processed data, responsibility, and reliability over time, while the process dimension contains factors like dependability, understandability, control, and predictability. Finally, the purpose dimension includes motives, benevolence, and faith. Recently, Siau and Wang (2018) suggested some of the important factors of trust, which are crucial in developing initial and continuous trust in AI. Factors like representation, transparency, explainability, and trialability play a vital role in initiating trust, while factors like usability, reliability, security & privacy, and interpretability are necessary for developing continuous trust in AI.

The issue of trust in AI is top of mind for several academic researchers, technology developers and consumers, as they can exhibit vulnerabilities such as privacy and security concerns, lack of explainability and transparency, exposure to bias, etc. Academic researchers and experts are confronting such issues by laying down some foundational elements of trust as themes of AI in order to trust the technology. There has been no consensus or universally accepted elements on what contributes to the themes of AI, as several tech giants and research institutions have their own defined principles and themes of AI. A detailed summarization of these themes has been presented in Table 1.

Based on the studies conducted on the concept of trust in relation to automation and themes laid down by research institutions and tech giants, potential trust factors of the algorithm in the context of AI have been identified and detailed below:

Accuracy – refers to the ability of algorithms to make accurate judgments, predictions, and recommendations. Users and business leaders would expect a high level of accuracy from the AI based solutions, especially in critical environments that directly affect human lives (healthcare, financial sectors, etc.). This mainly depends on the choice of the algorithm used and the data that is being fed into the algorithm for training.

Consistency – refers to the ability of algorithms to exhibit the same behaviour without ambiguities, when repeated under same conditions.

Interpretability – refers to the ability of algorithms to be interpretable and to explain its actions (Siau and Wang 2018).

Bias-Free or Objectivity – refers to the ability of algorithms to exclude biases from the resulting AI solution. This by ensuring the solution is not being trained with biased data.

Reliability – refers to algorithms' ability to work appropriately with a range of inputs and a range of situations in order to prevent accidental harmful outcomes (European Commssion 2019). It is critical that the outcomes produced by algorithms are reliable, and performance is as desired, in order to place trust in the resulting AI solution.

Explainability – When AI gets more complicated, we speak of decisions being made by the so called 'algorithmic black box'. In order to place confidence in the outcomes produced by such technology and to cement the trust of the stakeholder, as they want to know why specific outcomes occur, it becomes necessary to know the rationale of how the algorithm arrived at its decision or recommendation.

Transparency – To trust AI solutions, one would expect AI-based solutions to be programmed and functioning according to the goal defined and containing evaluation metrics by which the likely behaviour of the algorithm can be inferred.

Auditability – refers to the ability for the system to undergo assessment, not only on the algorithm but

Table 1. Overview of the AI themes laid by leading technology giants, EU commission, and research institutions.

Themes	Description	Source
Fairness	Using training data and models that are free of bias in order to avoid unfair treatment of certain groups	Pillars of AI proposed by IBM – (Mojsilovic, 2018)
Robustness	AI systems should be safe and secure and not vulnerable to tampering or comprising that data that is being trained on.	
Explainability	Providing decision and suggestions that can be understood by their users or developers	
Lineage	Including the details of the development, deployment, and maintenance in order to be audited throughout the life cycle.	
Fairness	Making sure that AI systems are treating all people fairly.	Microsoft principles on AI – ("Microsoft AI principles," 2019)
Reliability & Safety	Making sure that AI systems are performing reliably and safely.	
Privacy & Security	Making sure that AI systems are secured & are respecting privacy.	
Transparency	Making sure that AI systems are understandable.	
Fairness	1.Using representative datasets to train and test the model.	Responsible AI practices by Google – ("Responsible AI Practices," 2019)
	2.Checking the system for unfair biases.	
	3.Analyzing the performance of the system.	
Interpretability	1.Designing a model that is interpretable.	
	2.Understanding the trained model.	
Privacy	1.Collecting and handling data responsibly	
	2.Safeguarding the privacy of AI/ML models.	
Security	1.Identifying the potential threats to the system	
Fairness	Whether the bias in the data and AI model is minimized and Are the bias been addressed when AI is being used?	Dimensions of Responsible AI by PwC – (*Building Trust in AI and Data Analytics*, 2018)
Interpretability	Can we explain how an AI model is making decisions and ensuring that the decisions made by AI models are accurate?	
Robustness & Security	Can we rely on the performance of the AI system and whether AI systems are vulnerable to attack?	
Governance	Who is accountable for the AI system and ensuring whether proper controls are in place?	
Technical robustness and Safety	Including resilience to attack and security, fall back plan and general safety, accuracy, reliability, and reproducibility	Some of the key requirements for building a trustworthy AI – (EUROPEAN COMMISSION, 2019)
Privacy and data governance	Including respect for privacy, quality, and integrity of data, and access to data.	
Transparency	Including traceability, explainability, and communication	
Bias evaluation	Understanding, monitoring, and documenting the biases in development and production	Some of the key machine learning principles by Institute of AI & Machine learning – ("The Institute for Ethical AI & Machine Learning," n.d.)
Explainability	Developing tools and process to improve transparency and explainability	
Trust by Privacy	Protect and handle data with stakeholders that interact with the system directly or indirectly	
Security risks	Develop and improve process to ensure data and model security are taken into consideration.	

also on the data and process encompassing it (European Commission 2019).

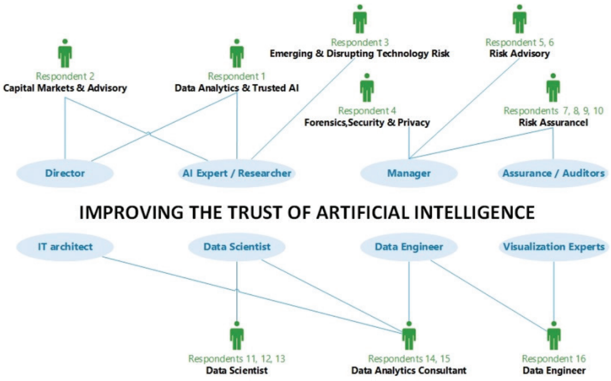
Governance – The term governance in the context of AI aims to close the gap that exists between accountability and ethics in technological development. Governance in AI involves identifying answers to the questions surrounding the safety of AI, identifying what legal and institutional sectors to involve, identifying who has the control and access to personal data, identifying what the roles are of moral and ethical institutions when interacting with AI, identifying control, and monitoring of the algorithms and the data (Rouse 2018).

Usefulness – An algorithm is said to be only useful when it produces accurate results, has an explainable interface, has consistent results that should be reliable, and has transparency that helps the auditors assess the algorithm's behaviour.

4. Interview analysis: validating the trust factors

The list of potential factors identified from the academic literature was presented to the actors to make them familiar with the context of the issue. Furthermore, some potential use case examples were explained upfront to accentuate the importance of trust and factors influencing trust towards AI. The actors could then identify the factors from the visual aids that were presented beforehand, in order to identify the important factors from their perspective. The actors were expected to give a clear explanation when they identified the important factors from their perspective, as there could be factors which could be tied to both the data and algorithm. For example, accuracy can be considered as one of the main dimensions of DQ and it can also be tied to the algorithm as a trust factor in terms of the accuracy of results the algorithm produces. Figure 2 portrays the number of actors along with subject of expertise and roles.

Figure 2. Overview of the actors interviewed.



Actors such as data scientist, data engineer and stewards were interviewed to understand the value of data, its challenges, and its influence towards the algorithm.

Actors proficient in dealing with and mitigating the risk of emerging technologies were involved (risk advisors & auditors).

AI experts and specialists were also involved in this study.

Upon analysing interview transcripts using coding, it was observed that accuracy, auditability, bias-free, consistency, governance, explainability, privacy, reliability, security, transparency, and usefulness were identified to be prime trust factors from the perspective of the actors interviewed. Nevertheless, these factors must be clearly attributed towards the data and the algorithm, as there could be possibilities where the factors could mean crucial only to data or algorithms or both. For example, factors like consistency, accuracy, reliability, usefulness, interpretability, etc. can be tied to both data and algorithm. Based on further analysis and distinguishing the factors, it was clear that most of the actors with data backgrounds paid high attention to the data. To be more specific, actors like data scientists, data engineers and data analytics consultants recognized consistency, accuracy, completeness, bias-free, reliability, etc. as the important dimensions of DQ and governance as the most critical trust factor in order to be able to trust the entire data process and its lifecycle. Though these actors didn’t outline many of the trust factors towards the algorithm, it was believed that an algorithm which has accurate, reliable, consistent, bias-free data and is being trained with such data would also be reliable, consistent, and bias-free. Actors like risk advisors, auditors and managers recognized auditability, security, privacy, reliability, bias-free, transparency, explainability, governance, and usefulness as crucial factors in trusting the algorithm and its resulting outcomes. It was also believed that most of these actors would be involved either in the initial phase or at the final phase of developing an AI solution and they would mainly see the above factors as important in order to trust the algorithm and its results.

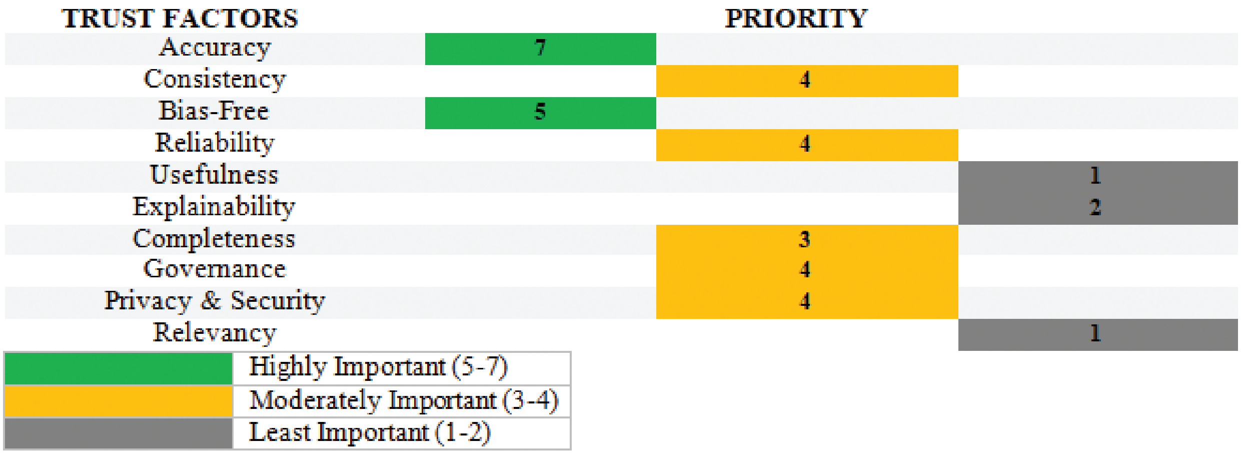
So, it can be inferred that DQ dimensions like accuracy, reliability, consistency, completeness, bias-free, etc. were highly important and indeed, many actors with a data background emphasized the same factors. Furthermore, risk advisors and auditors highlighted the same set of factors towards the data. Figure 3 shows the list of DQ dimensions pertaining to the data. This list has been categorized by priority based on the analysis from the interviews. At this point, it has to be noted that governance as a trust factor may not contribute to the dimensions of DQ, but it was assumed to be important in order to establish the trust around the data process. To clearly distinguish the trust factors, a conceptual map was derived, as shown in Figure 4, indicating the important dimensions of DQ based on the actor’s perspective. Similarly, with respect to the algorithm, factors that were identified from the analysis of the interviews were categorized according to their priority based on the actor’s perceptions towards those factors (as shown in Figure 5).

5. Building a trusted AI model

Using our literature research and the findings collected from the interviews with relevant actors as the main source, a trusted AI model was developed. We have not found any existing or concrete trustworthy AI models in academic research that have explicitly defined the prime trust factors towards AI. As such a new model needed to be developed, encompassing the phases involved in AI development and in those respective phases, relevant trust factors with indicators had to be identified.

It should be noted that there are no specific standard phases defined in scientific research with respect to AI development. This is possibly due to the variation’s dependent upon the context of every organization’s workflow. This research has used Guo’s (2017) study where he identifies several main phases involved in the development of machine learning (one of the dimensions of AI). His study was used as the main source of refer-

Figure 3. Categorization of trust factors of data in terms of priority.



TRUST FACTORS	PRIORITY		
Accuracy	7		
Consistency		4	
Bias-Free	5		
Reliability		4	
Usefulness			1
Explainability			2
Completeness		3	
Governance		4	
Privacy & Security		4	
Relevancy			1

🟩	Highly Important (5-7)
🟨	Moderately Important (3-4)
▨	Least Important (1-2)

Figure 4. Conceptual map showing the dimensions of DQ in the context of AI.

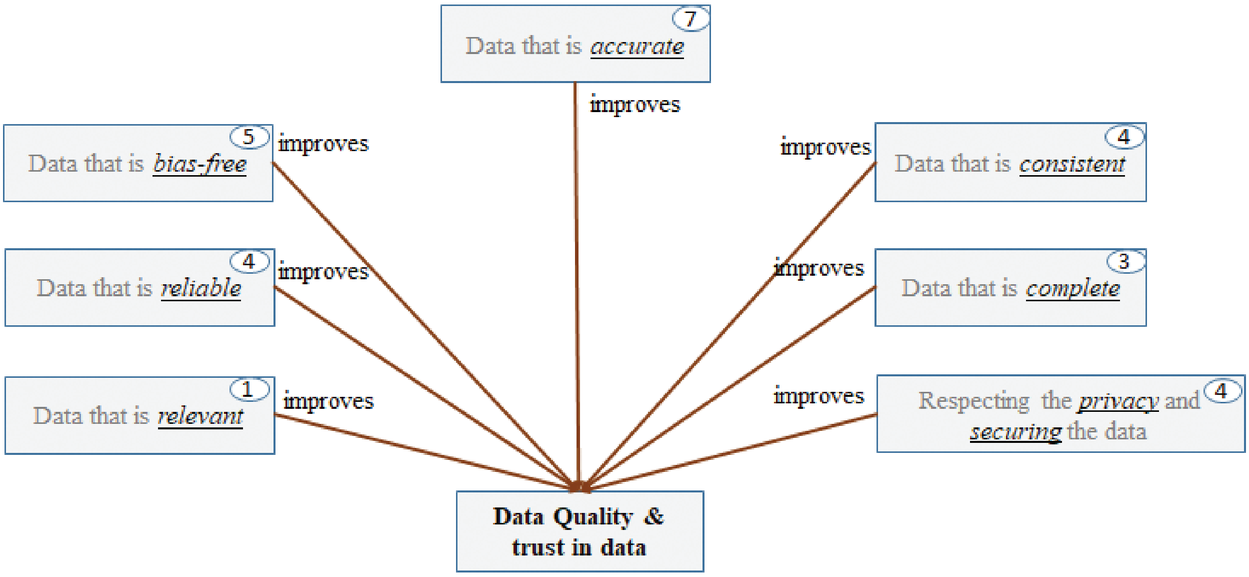
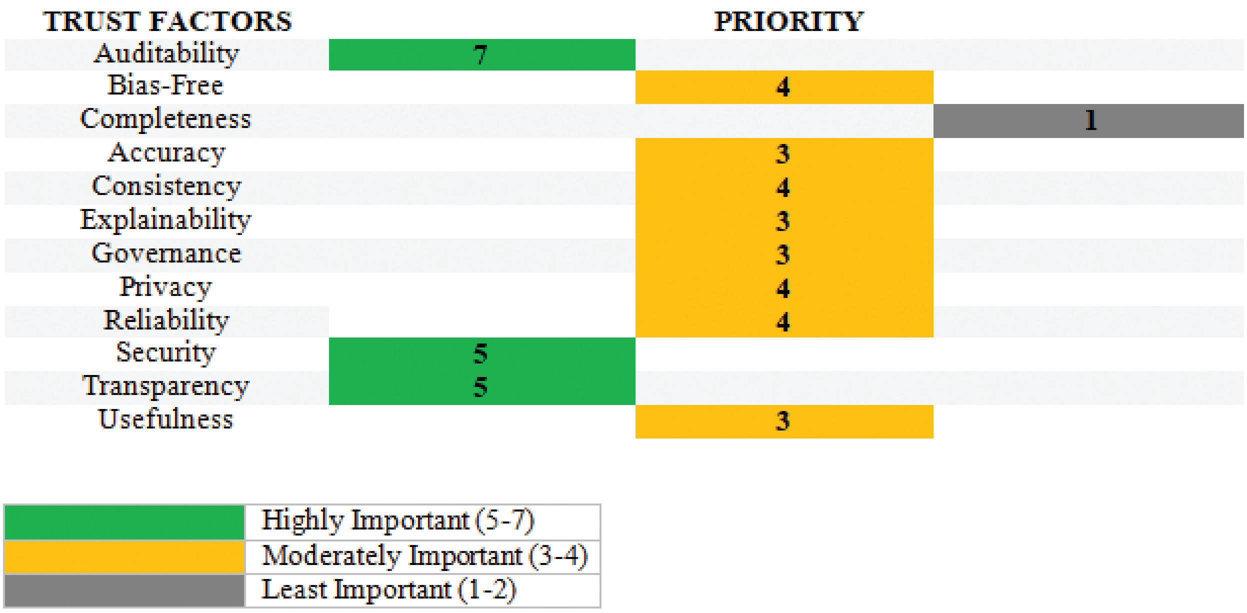


Figure 5. Categorization of trust factors of algorithm in terms of priority.



TRUST FACTORS	PRIORITY		
Auditability	7		
Bias-Free		4	
Completeness			1
Accuracy		3	
Consistency		4	
Explainability		3	
Governance		3	
Privacy		4	
Reliability		4	
Security	5		
Transparency	5		
Usefulness		3	

🟩	Highly Important (5-7)
🟨	Moderately Important (3-4)
⬛	Least Important (1-2)

ence in identifying the probable phases involved in AI development with regards to this research, along with the findings from the interviews of experts. Most of the phases like Data acquisition, Data preparation and Validation, Model Selection, Model Training and Testing, Model Validation and Deployment were retrieved from the study of Guo (2017). These identified phases were found to be very similar to many leading technology companies' process flows for developing AI-based products or solutions. So, the phases involved in AI development we defined are:

(1) Problem/Improvement Exploration, (2) Human-Centered Design, (3) Data Acquisition, (4) Data Preparation and Validation, (5) Feature Selection, (6) Model/Algorithm Selection, (7) Model/Algorithm Training and Testing, (8) Model/Algorithm Validation and Deployment, and (9) Model/Algorithm Monitoring.

It is important to recognize that phases like Problem/Improvement Exploration (1) and Human-Centered Design (2) won't be elaborated on in this research. The purpose of outlining those phases is to make the model complete and provide a clear view for the reader to help them recognize the actual phases involved in the development of AI. The research places its emphasis from the data acquisition phase on, as the research mainly presents the trust factors concerning data and algorithms. The associated factors for every phase were identified based on the engagement with the actors, along with the analysis done on literature and themes of AI laid down by research institutions and companies.

5.1 Main phases & associated trust factors of trusted AI model

Data Acquisition (*3*) – is one of the prime phases to establish initial trust. This is the phase where the developers of data such as data scientists, data engineers and stewards get involved and start gathering the data. There are several ways to acquire the data, which are dependent on the context of the client case. There can be instances where data can be obtained directly from the client or gotten from an external party or data provider or from public data sources. The importance of trust towards the data initiates from this phase where the dimensions of DQ serve as a base in assessing the quality of data and placing the initial trust over such data. Dimensions like *reliability, completeness and consistency* would be the prime trust factors that need to be assessed regardless of the context of the case. For instance, to assess *completeness* of the data, the developers working on the data must see whether all the possible states relevant to the user population are being represented in the stored data. Assessing the *accuracy* of the data at the first instance would be very hard, as the data received would be raw and unstructured most of the times.

Table 2. Dimension of DQ in Data Acquisition phase.

Factors	Indicators
Reliability	• The contents present in the data set are credible enough to process.
	• The data has been generated from a trusted source.
Completeness	• The mandatory and main attributes in the data are not null.
	• The data has all the possible states relevant to the user population.
	• The data is completely representative of the real-world state.
Consistency	• The source data is consistent even after a certain point in time.
	• The physical instance of the data is in accordance with some standard format.
	• The value and entries in the data are the same in all the case.

Data Preparation and Validation (*4*) – Once the initial trust is established through the data acquisition phase, the same level trust needs to be maintained in the phase of data preparation and validation. This phase involves a series of steps for processing the data. This is also the phase where the processed data can be split into two parts, where the first part is the training data that is used for training the model/algorithm, and the second part is the testing data that is used to evaluate the performance of the model/algorithm once the model is trained (Guo 2017). In either case, the dimensions of DQ need to be assessed in this phase. Having *bias-free* data would be one of the most crucial trust factors in this phase and it might not seem to be of prime importance in the previous phase, as it would be challenging to determine biases at the first sight when the data is raw, unformatted, and unstructured. Second, the actors working on the data should check for *consistency* of the data in order to ensure that there are no ambiguities with the actual data. Third, *accuracy* of the data at this stage has to be assessed to ensure that the data is now certified and error free. Finally, *relevancy* of data has to be looked upon by the actors, as there would be a course of processing like data cleansing, filtering, labeling, and visualization to get to the processed data and it needs be seen whether the processed data is still addressing the purpose of interest in a given context.

Table 3. Dimension of DQ in Data Preparation and Validation phase.

Factors	Indicators
Relevancy	• The processed data is relevant to the specific domain and purpose of interest in a given context.
	• Processed data contains the required variable in the right form and a representative of population interest.
Consistency	• The processed data and actual data source are consistent and representative, have no contradiction and are compatible with the previous data.
	• The attribute values in the data have no ambiguities after data processing.
Bias-Free	• The biased data identified during the data acquisition and processing phase has been removed.
	• Data contains all the possible representation of the subject, object and has no inclinations towards a specific object or a thing, i.e. no traces of biases, prejudiced, and partial data were found.
	• Determining the cause if any variance is detected so as to avoid biases.
Accuracy	• The processed data hasn't lost the structure during the data processing stages.
	• The processed data is certified and free of error.
	• The processed data is accurate, objective and is corresponding to a real-world context.

Feature Selection (*5*) – At this point, there need to be checks to ensure that the most valuable information has been derived out of the data, as that can help the algorithm to produce the best results. So, selecting the right features from the data is paramount, as it reduces over-fitting of the algorithm, improves the accuracy of the results by discarding the data that is misleading, and reduces the training time by having only important data points that make the algorithm train faster (Shaikh 2018). To validate whether the chosen feature is addressing the actual context of the goal and is still representative of the entire population, *relevancy* and *completeness* of the feature have to be assessed.

Table 4. Dimension of DQ in Feature Selection phase.

Factors	Indicators
Relevancy	• The chosen features are applicable and addressing the actual context of the goal.
Completeness	• The chosen features are representative of the entire population.

Model/Algorithm Selection (*6*) – Once trust is established towards the data by mainly assessing the dimensions of DQ as trust factors in the last three phases where data has been the core driver, a relevant model/algorithm should be selected. Choosing an algorithm that can be easily *interpretable* with respect to the data and can be *explainable* for every certain output, should be the primary driver to establish continuous trust at this phase. In addition, it needs to be assessed whether the chosen algorithm can produce *accurate* outcomes if trained and it is not pre-trained to show any *biases*.

Table 5. Trust factors in Model/Algorithm Selection.

Factors	Indicators
Interpretability	• The result that an algorithm can produce is interpretable with respect to the data.
Accuracy	• The chosen algorithm can produce correct predictions, decisions, and recommendations results if trained properly.
Bias-Free	• The chosen algorithm is not pre-trained to show any biases and can work well if trained properly.
Explainability	• The algorithm can provide some explanation for every certain output.

Model/Algorithm Training and Testing (*7*) – Having selected the relevant algorithm based on the data analysis, the algorithm requires training and it needs to be trained based on the training data that was kept aside during the earlier phase of data preparation and validation. Once the algorithm has been trained, it needs to be tested. Test data at this stage would come into the picture and this test data should be different from the training data, so as to not give misleading results on the model's performance. One of the key trust factors at this stage is ensuring that the algorithm that is being trained and tested doesn't show any *biases or variances* and, if found, they have to be addressed. Second, to evaluate the *accuracy* of the algorithm, it needs to produce the same set of desired outcomes when trained and reflect the same outcomes when tested with new data that are bias-free. Also, it requires careful review and evaluation of training and test data with data subject expertise, and in order to improve the feasibility of auditing, it should be made *transparent* to the (internal) auditors.

Table 6. Trust factors in Model/Algorithm Training and Testing.

Factors	Indicators
Transparency	• The training and testing data is open to be reviewed by domain experts.
Usefulness	• Tuning has been made to improve the overall usefulness of the model.
Bias-Free	• The potential skews found during the training have been addressed.
	• Any skews observed during the testing were identified and addressed.
	• The training and testing data are free of biases.
Accuracy	• Testing data and training data has samples that represent all the targeted subjects, things, objects, etc.
	• The model can produce desired results based on the training.
	• The model can produce the same set of desired results when tested with new data (test data).

Model/Algorithm Validation and Deployment (*8*) – The algorithm that has been trained and tested needs to be validated and deployed. The main purpose for validation is to ensure that the algorithm has taken into consideration of the important themes like *explainability, reliability, and usefulness*, which serve as trust factors. An algorithm that has an *explainable* interface is one of the powerful tools that can help users detect major flaws, interpret the model behavior, ideally maximize the model performance, and make alterations based on the outcomes and rationales behind the outcome at this stage. It is also the phase where the actors have to make sure that the algorithm is actually working as intended, addressing the actual challenge and providing value in terms of *reliability* and *usefulness* of the outcomes.

Model/Algorithm Monitoring (*9*) – In this final stage, the algorithm needs to be closely monitored, especially when it is deployed in critical/high risk environments. The key to monitoring the algorithm is to understand the potential weaknesses as it would give a better indication of how the algorithm is performing, how it is interpreting the inputs and where it is failing, etc. when it is deployed in a real-time environment.

On top of these factors come the other core trust factors like *auditability* and *governance*, which need to play a central role through the development of AI. In terms of *auditability*, every change and assumption made with respect to the data and algorithm has to be documented in order to assess and verify the data process and handling of algorithms. This could improve the overall integrity and trust towards the data and algorithms. A detailed indicator for auditability has been presented in Table 8. Second, establishing a robust *governance* around the process of AI development will positively influence the overall trust in the solution, as governance is one such factor that acts as an outer shield to other influential trust factors like privacy and security of the data. *Privacy*, for instance, is gaining increasing attention across Europe with GDPR and other regulations currently in effect. It must be assured that the data and the algorithm as a whole are completely safe and at any cost, they are not vulnerable to tampering or comprising the data that is being trained and tested on. It must be noted that factors like privacy and security are not tied to any specific phase but instead, they must be assessed throughout the phases.

5.2 Value of trusted AI model

The development of a trusted AI model requires visualizing trust factors of the data in the form of DQ dimensions and algorithms and projecting those factors clearly in the respective phases of the AI development. This is one of the key merits to developing a trusted AI model. The presentation of the trusted AI model itself serves as a main and added value to the organization, next to the identification of DQ dimensions and trust factors of the

Table 7. Trust factors in Model/Algorithm Validation and Deployment.

Factors	Indicators
Explainability	• The algorithm can explain every decision, recommendation, and the prediction made.
	• The decision made by the system can be understood and traced by developers.
Reliability	• The algorithm is performing as intended.
	• The algorithm is working properly when a new range of inputs is being applied.
Usefulness	• The resulting AI solution had addressed the challenges and is providing value to the clients, society and the stakeholders involved.

Table 8. Trust factors in Model/Algorithm Validation and Deployment.

Factors	Indicators
Auditability	• Every change made with respect to datasets that have been documented is verified.
	• A process on how data was received, analyzed, cleansed, filtered, and labeled are documented and verified.
	• Assumptions made on the data to improve the usefulness of the data has been documented and reviewed.
	• Possible risks and opportunities that were encountered and how they were addressed is documented.
	• The results produced by AI systems are compared with the actual results.
	• Reviewing the output of the AI model and the meaning derived from those outputs.
	• Verify by interviewing the ones responsible for developing AI models that they can understand and explain the AI black box data.
	• Assessing the already existed data from the client or other third-party vendors to validate the usefulness of the data for requirements.
Governance	• A clear line of accountability, responsibility standards is clearly established.
	• Know the responsibility of the data & model owners at each and every phase involved in the creation of AI.
	• Ensuring that human is involved in every phase of the model as well as to oversee the overall activity of the AI application.
	• Data protocols outlining
	1. Who can access the data and under which circumstances?
	2. How the data are being handled and how is it being protected.
	3. Whether the data is in compliance with GDPR & data protection policies.
	4. What kind of biases were identified, how were they solved?
	5. Were there any measures been taken to inform the relevant stakeholders when major bases were notified? etc. should be put in place.
Under the governance, special attention must be paid to factors like privacy and security	
Privacy and Security	• Identify and protect the core strategic data assets.
	• Access to final processed data is restricted & secured from unauthorized access.
	• Processed data has been encrypted & anonymized if any personal or confidential information is present.
	• Processed data is used fairly and respects the privacy of the user's data.
	• The processed data is in compliance with GDPR and other data protection policies.
	• The chosen model can be resistant to malicious training.
	• The model is making fair predictions with the data being trained
	• The model is neither using the test data maliciously nor being fed with malicious data and is making fair predictions with the test data.
	• The data used by AI algorithm or model is not used unlawfully or unfairly against the users.
	• The AI algorithm are safe and secured and are not vulnerable to tampering or compromising the data that they have been trained on.
	• AI models are protected from being exposed to unexpected situations.
	• The AI models developed are limited to the context of the goal.
	• The AI algorithm respects the user privacy of information

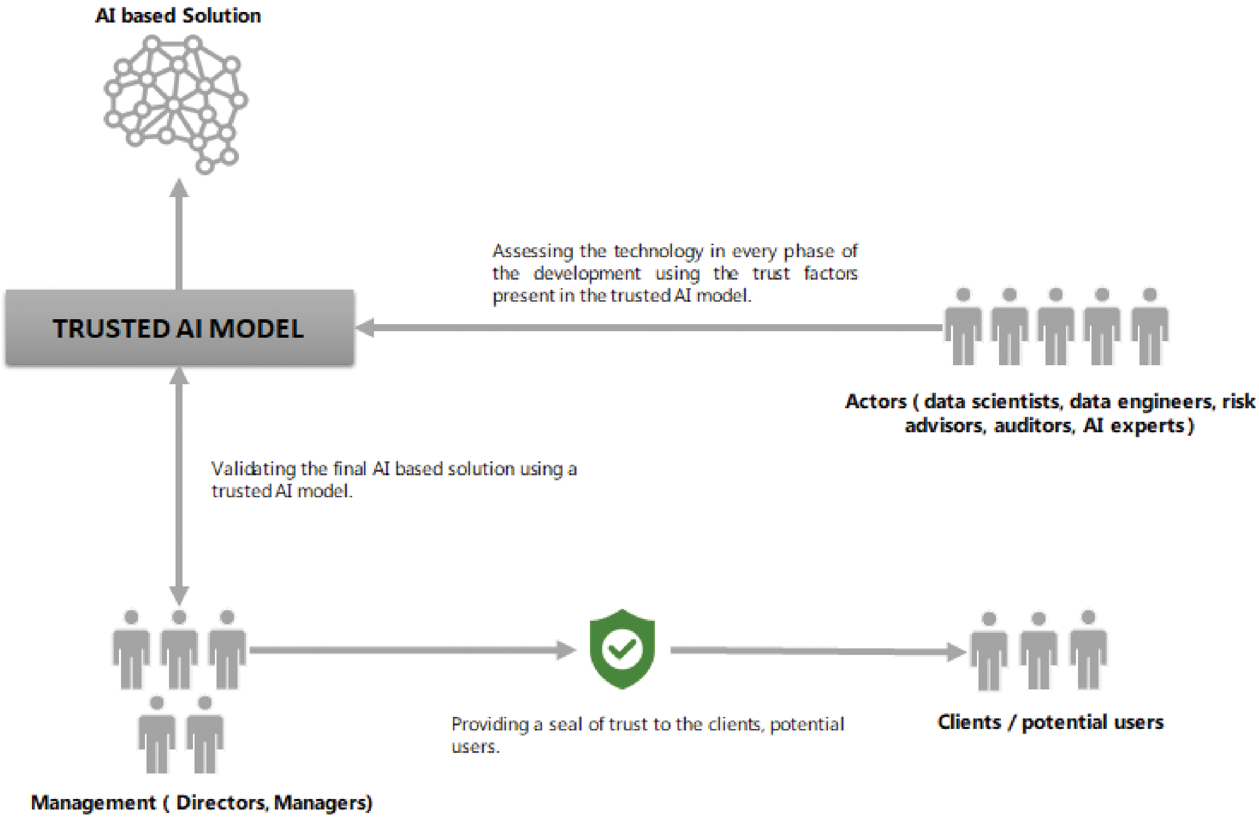
algorithm. *Firstly*, the trusted AI model can assist technology developers in assessing the technology critically using the trust factors identified in the respective phases of the AI development. *Second*, the top- and mid-management levels of the organization comprising directors and managers can use this model as a guide or an assessment tool to validate the AI solution at the end or at every phase of the development. The model not only portrays the trust factors present in every phase of the AI development, but also provides detailed indicators for every trust factor. Having these indicators available can guide the management to validate the technology and see what factors would require more attention when the indicators of respective trust factors don't meet the required expectations. *Third*, looking at scenario's where there would be a need to audit an AI solution, internal auditors – in such a situation – can use this model as a guide to critically assess the solution. At this stand, it must be observed that the model as a whole won't be efficient when it is being used by small and medium-sized enterprises (SMEs) or any AI startups. It is assumed that most of the startups and SMEs have a flexible workflow and they don't necessarily follow the same phase identified in the model. Nevertheless, the trust factor which is one of the prime aspects of the model would serve as major value provider to those enterprises. As can be seen in Figure 6, the model at the end (*1*) would serve as a useful purpose to the developers (actors working on the data and building the model) to assess the technology in every phase of the development using the trust factors, (*2*) aids the management to validate the final AI solution using the trusted AI model, and (*3*) would assist internal auditors in evaluating AI solution using the trust factors and indicators as an assessment check.

6. Discussion

AI has been gaining more traction over the recent years, especially in business. Several leading tech companies and start-ups have already invested in the development of AI by delivering data-driven AI-based solutions. These solutions are currently more focussed on performing a specific task i.e. more towards narrow AI, but when AI starts getting more complex and moves into the stage of enhanced and cognitive automation, the value of trust becomes increasingly important. Leading research institutions, scientific researchers, and tech companies have already started addressing the importance of trust towards AI. Some of the questions being asked upfront about AI are: Does AI explain the results, how is AI using my data, is AI being governed, etc. and all these questions lead to one main question – Whether AI can be trusted.

One of the major challenges during the development of AI are the trust issues that are encircled not only towards the outcome of an AI solution, but in the process of developing those solutions. Data has been the seed to the success of AI and as such the quality of data is an essential property that determines the value and trust towards the data. To trust the data, one needs to look for all rele-

Figure 6. Value of trusted AI model.


AI based Solution
TRUSTED AI MODEL
Assessing the technology in every phase of the development using the trust factors present in the trusted AI model.
Actors (data scientists, data engineers, risk advisors, auditors, AI experts)
Validating the final AI based solution using a trusted AI model.
Providing a seal of trust to the clients, potential users.
Clients / potential users
Management (Directors, Managers)

vant dimensions of DQ and needs to validate recursively at every phase of handling and processing the data.

Bad DQ can lead to bad information quality (IQ) and good IQ can imply good DQ. Good DQ however might not lead to good IQ.

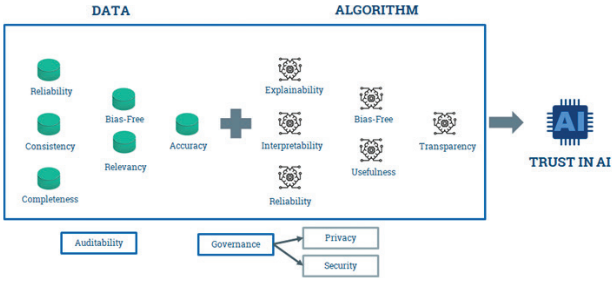
Assessing the quality of data using the dimensions is certainly not a once-off validation as there are multiple levels of dealing with the data in the light of AI. So, dimensions of DQ need to be assessed right from gathering the data to the phase where the data is processed and ready as information. Once such processed data is loaded to the algorithm, the trust needs to be shifted towards the algorithm's behaviour and its outcome.

The assumption is that realizing the trust factors associated with AI and assessing them upfront using a model can improve the overall trust of the technology and thereby provide a stamp of trust to the required stakeholders.

The identified factors, though they seem to be comprehensive in this paper, can be further boiled down to determine the essential set of trust factors of the data and algorithm in the medium of AI. Deriving the essential factors of the data and algorithm would be one of the prime opportunities for academic research. With respect to the data, it can be concluded that factors such as reliability, accuracy, completeness, consistency, bias-free, relevancy are the key factors that were identified to be indispensable in the context of AI, based on findings from the interviews conducted with relevant actors and from the studies done on the DQ dimensions, factors influencing trust in automation and themes of AI. So, the data that is received and processed needs to be consistent, complete, relevant, bias-free, accurate and reliable and these dimensions can't be ignored and must be vital regardless of the context of the case and the target environment. In the context of the algorithms, the most vital trust factors identified from this research were bias-free, explainability, interpretability, reliability, transparency, and usefulness of algorithms. Factors like bias-free, explainability and interpretability were highly stressed during the interviews and it was even emphasized by several AI research institutions, as these factors would ideally help to place stronger trust in the technology and help develop AI for good.

Furthermore, governance was identified as the most essential trust factor that needs to be in place during the whole process (mainly from data gathering to monitoring the final AI solution). Perhaps, governance would serve

Figure 7. Essential trust factors of data, algorithm in the context of ai.



as an umbrella for major influential trust factors like privacy, data security and other unexplored trust factors like accountability and ethicality, which might play a prime role in the resulting AI solution. Most interestingly, one of the trust factors, auditability happens to be highly important from the interview findings, as it was believed that examining the AI solution thoroughly can positively influence the trust in AI. The consensus was found to be that the AI solution developed has to be explainable, interpretable and transparent, as that would improve the feasibility of auditing such solutions. In such a scenario, factors like *explainability, interpretability, and transparency of the algorithms* would stand as the main prerequisites for internal auditors while assessing the resulting AI solutions.

Though there might seem to be an array of trust factors in terms of data and algorithms (as presented in Figure 7), it is essential that these trust factors are critically assessed and validated in the respective phases, as every trust factor plays a key role in the context of AI and certain trust factors concerning algorithms are mainly dependent on the data. For instance, trust factors with respect to the algorithms, like explainability and interpretability, can be assessed only based on the algorithm that is being chosen. While other trust factors of algorithms like reliability usefulness, and bias-free are primarily dependent on the data. The developers and AI should evolve in adherence to the concept of a parental-child relationship, where parents nurture their child and feed them with good values and norms. Likewise, creators of the technology and management need to ensure that the data that is being fed to the algorithm is *reliable, accurate, consistent, complete, relevant, and bias-free* in order to initiate and establish stronger trust over the data. Also, the algorithm that has been chosen needs to be *explainable, interpretable, transparent, reliable, useful, and bias-free*. More precisely, the algorithm and its resulting outcomes are actually auditable and properly governed. So, the essential trust factors would provide the main essence to organizations, irrespective of the organization size and process flow of the development. But it won't be appropriate to generalize these essential trust factors to all the subsets of AI, which would include deep learning, natural language processing, machine vision, etc. as the factors might be subject to variation. In the context of data, it is presumed that the dimensions of DQ would be identical in almost all the subsets of AI. Specifically, subsets like deep learning, natural language processing, etc. would require enormous data volumes and if such subset is used for building a solution, dimensions like accuracy, consistency, reliability, completeness, bias-free and relevancy would still be at the forefront as a factor to trust the data. In relation to the algorithm, further research would be required to make valid generalizations, as it would give a better clarity on what trust factors would really mean important to each and every subset of AI based on its purpose.

7. Glossary

Bias – Inclination of prejudice towards or against a particular person, subject, or object.

Blackbox Algorithm – Algorithm that makes decisions without explanation or without traces on how the algorithm has arrived at a decision.

Data Quality Dimension – Set of data quality attributes that represent a single aspect or construct of data quality.

Feature Selection – The process of selecting a subset of relevant features from a large pool of features present in the data.

Information Quality – Quality of the processed data.

Machine Learning – Process of enabling a machine to learn from the data without explicitly programming it with rules.

Narrow AI – AI that is designed to perform a repetitive and narrow task. Also, referred as Weak AI or Artificial Narrow Intelligence (ANI).

Test Data – Set of new-data (data never seen by algorithm) that is separate from the training data which is used to evaluate the algorithm post training.

Training Data – Set of data that is used to train the algorithm.

- **Mr. A.V. Srinivasan MSc** is graduate student at Delft University of Technology.
- **Drs. M. de Boer RE RA** is director data analytics at PwC and PhD researcher at the University of Amsterdam.

References

- Batini C, Cappiello C, Francalanci C, Maurino A (2009) Methodologies for data quality assessment and improvement. ACM Computing Surveys 41(3): 1–16. https://doi.org/10.1145/1541880.1541883
- Batini C, Scannapieco M (2016) Data and information quality. https://doi.org/10.1007/978-3-319-24106-7
- Bovee M, Srivastava RP, Mak B (2003) A conceptual framework and belief-function approach to assessing overall information quality. International Journal of Intelligent Systems 18(1): 51–74. https://doi.org/10.1002/int.10074
- Burkhardt M (2019) How-to build trust in artificial intelligence solutions. https://medium.com/omdena/how-to-build-trust-in-artificial-intelligence-solutions-a6d3c7ddf4c3
- Cai L, Zhu Y (2015) The challenges of data quality and data quality assessment in the big data era. Data Science Journal 14(2): 1–10. https://doi.org/10.5334/dsj-2015-002

- Davenport TH (2018) Can we solve AI's 'trust problem'? MIT Sloan Management Review. https://sloanreview.mit.edu/article/can-we-solve-ais-trust-problem/
- European Commission (2019) High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
- FRA (European Union Agency for Fundamental Rights) (2019) Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf
- French B, Duenser A, Heathcote A (2018) Trust in automation – A literature review. CSIRO Report. https://www.semanticscholar.org/paper/Trust-in-Automation-A-Literature-Review-French-Dünser/92f07d3d1356307decb6e97382ad884d0f62668d
- Guo Y (2017) The 7 steps of machine learning – Towards data science. https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e
- Hancock PA, Billings DR, Schaefer KE, Chen JYC, De Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human-robot interaction. Human Factors: The Journal of the Human Factors and Ergonomics Society 53(5): 517–527. https://doi.org/10.1177/0018720811417254
- Hoffmann H, Söllner M (2014) Incorporating behavioral trust theory into system development for ubiquitous applications. Personal and Ubiquitous Computing 18(1): 117–128. https://doi.org/10.1007/s00779-012-0631-1
- IBM (2018) Beyond the hype: A guide to understanding and successfully implementing artificial intelligence within your business. https://www.ibm.com/downloads/cas/8ZDXNKQ4
- Janssen M, Haryadi AF, Hulstijn J, Wahyudi A, Van der Voort H (2017) Antecedents of big data quality: An empirical examination in financial service organizations. Proceedings – 2016 IEEE International Conference on Big Data, Big Data 2016: 116–121. https://doi.org/10.1109/BigData.2016.7840595
- Kelly C, Boardman M, Goillau P, Jeannot E (2003) Guidelines for trust in future ATM systems: A literature review. European Air Traffic Management Programme, (November), 52. http://www.hf.faa.gov/hfportalnew/docsPages/DocPage.aspx?id=1129
- Krogue K (2017) Artificial intelligence is here to stay, but consumer trust is a must for AI in business. Forbes website: https://www.forbes.com/sites/kenkrogue/2017/09/11/artificial-intelligence-is-here-to-stay-but-consumer-trust-is-a-must-for-ai-in-business/#3248625e776e
- Lee JD, See KA (2004) Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society 46(1): 50–80. https://doi.org/10.1518/hfes.46.1.50.30392
- Marr B (2019) Why every company needs an artificial intelligence (AI) strategy for 2019. Forbes. https://www.forbes.com/sites/bernardmarr/2019/03/21/why-every-company-needs-an-artificial-intelligence-ai-strategy-for-2019/#7b2bfe568ea9
- Microsoft AI principles (2019) Microsoft website: https://www.microsoft.com/en-us/ai/our-approach-to-ai
- Mojsilovic A (2018) Factsheets for AI Services. https://www.ibm.com/blogs/research/2018/08/factsheets-ai/
- Muir BM (1994) Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37(11): 1905–1922. https://doi.org/10.1080/00140139408964957
- Nelson RR, Todd PA, Wixom BH (2005) Antecedents of information and system quality: An empirical examination within the context of data warehousing. Journal of Management Information Systems 21(4): 199–235. https://doi.org/10.1080/07421222.2005.11045823
- Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Communications of the ACM 45(4): 211–218. https://doi.org/10.1145/505248.506010
- PwC (2018) What to expect from artificial intelligence (AI) in 2018. https://www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions.html
- PwC (2018) Building trust in AI and data analytics. https://www.pwc.com/sg/en/publications/assets/building-trust-ai-data-analytics-122018.pdf
- Rempel JK, Holmes JG, Zanna MP (1985) Trust in close relationships. Journal of Personality and Social Psychology 49(1): 95–112. https://doi.org/10.1037/0022-3514.49.1.95
- Responsible AI Practices (2019) Google website: https://ai.google/responsibilities/responsible-ai-practices/?category=general
- Rouse M (2018) What is AI governance? https://searchenterpriseai.techtarget.com/definition/AI-governance
- Shaikh R (2018) Feature selection techniques in machine learning with Python. https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e
- Siau K, Wang W (2018) Building trust in artificial intelligence, machine learning, and robotics. Cutter Business Technology Journal 31(2): 47–53. https://www.cutter.com/article/building-trust-artificial-intelligence-machine-learning-and-robotics-498981
- The Institute for Ethical AI & Machine Learning (n.d.). https://ethical.institute/
- Wand Y, Wang RY (1996) Anchoring data quality dimensions in ontological foundations. Communications of the ACM 39(11): 86–95. https://doi.org/10.1145/240455.240479
- Wang RY, Strong DM (1996) Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems 12(4): 5–33. https://doi.org/10.1080/07421222.1996.11518099